2930
doi:10.1017/S1431927622010996
Microsc. Microanal. 28 (Suppl 1), 2022


# Development of a FAIR Data Management Infrastructure

Sherjeel Shabih[1], Markus Kühbach[1], Markus Scheidgen[1], Lauri Himanen[1], Sandor Brockhauser[1], Benedikt Haas[1] and Christoph Koch[1]

[1.] Humboldt Universität zu Berlin, Institut für Physik & IRIS, Adlershof, Berlin, Germany.
* Corresponding author: sherjeel.shabih@hu-berlin.de

Recent developments in data science tools have shown how powerful repurposing of data can be, if enough of it is available. In electron microscopy, where modern detectors produce data at rates of multiple GB/s, most data is analyzed for only a single aspect out of the wealth of information recorded. In most cases, only a fraction of the extracted data is polished and presented in publications. The rest is lost on personal devices, laptops, and external hard drives, and the unrevealed statistical information of materials is never brought to the limelight. In the pursuit of making data available, we not only tackle the storage and accessibility of the raw data, but also reduce the barrier of entry into making sense of this data, by researchers who are not necessarily specialists in the technique.

Most instrument vendors do not provide open data formats that a researcher can use for long term archival or even short term sharing of data with colleagues, who may not have access to the vendor software. Even with access, much of the accompanying pieces of information that make the acquisition reproducible in a future session are not available. Most vendors store information that is required for the internal functioning of the acquisition software, and there exists no public documentation from the vendor or a community template for the vendor to provide data in accordance with. In this talk, I will present our approach to enabling FAIR (findable, accessible, interoperable, and reusable) data and processing.

The benefits of FAIR data processing and sharing motivated us to develop a framework to help the community design, test, and write conversion templates from their instrument's formats to an open hierarchical file structure. The most important aspect of this endeavor is the documentation that goes in the file along with the data stored. Anyone can read what a data element means in human language rather than inferring what it means through abbreviations or knowledge of the technique/instrument. We build these documentations with the NeXus file format [1]. This not only allows us to store this data for our own future use, but it allows commercial and community analysis software to interpret the data with less discrepancies.

One of the earliest steps we took was to add support in the microscope acquisition software to directly export data in an open format and store it on a server. In our case, this was possible to implement due to the open source nature of Nion Swift [2], the acquisition software. The central server runs web services that provide remote access to the data, avoiding the need of having to copy data onto hard drives and creating duplicates, other than the automatic backups of raw experimental data that are produced automatically.

The next step was to ensure this data does not need to be duplicated for analysis either, as this can be very slow and files may damage during copying. To achieve this, we moved the tools that we frequently use for data analysis from the desktop computer to the cloud. This way, the data is directly analyzed on the server and a log of all mutations applied to the data is stored. Storing this history makes it possible to

easily reproduce the workflow that was added with the rest of the methodology, for example, when writing a publication.

To practically enable this, we have set up and extended the free and open source electronic lab notebook eLabFTW [3] to let users access this data and utilize it online. Extensions of eLabFTW include the capability to edit JSON files and connect external tools such as JupyterLab, LiberTEM, and our in-house analysis routines (e.g. the focal series reconstruction FRWR [4,5]), containerized to be able to be used from a web browser. The initial eLabFTW-based system has now been replaced by NOMAD [6].

This framework allows us to easily maintain our data while reducing unnecessary steps in preparing the data for analysis and ultimately publication. More importantly this allows us to run these routines much faster due to the computational power of the central server hardware running the services as compared to user laptops. As a result, resources are shared efficiently and the entry barrier to setting up complex tools is minimized, making them easily accessible to all members of the group, as well as external collaborators.

NOMAD provides additional structural benefits with its collection of data conversion tools provided by the community and in its handling of metadata, allowing labs to create a documented version of their data. Labs can install local instances of NOMAD, called NOMAD Oasis, and store FAIR data of their own research locally. Uploading selected datasets to a publicly accessible NOMAD server enables researchers to not only archive the data locally but make it available for others, either along with their publication or as standalone datasets. NOMAD also comes with the data conversion framework baked in to read from various community readers.

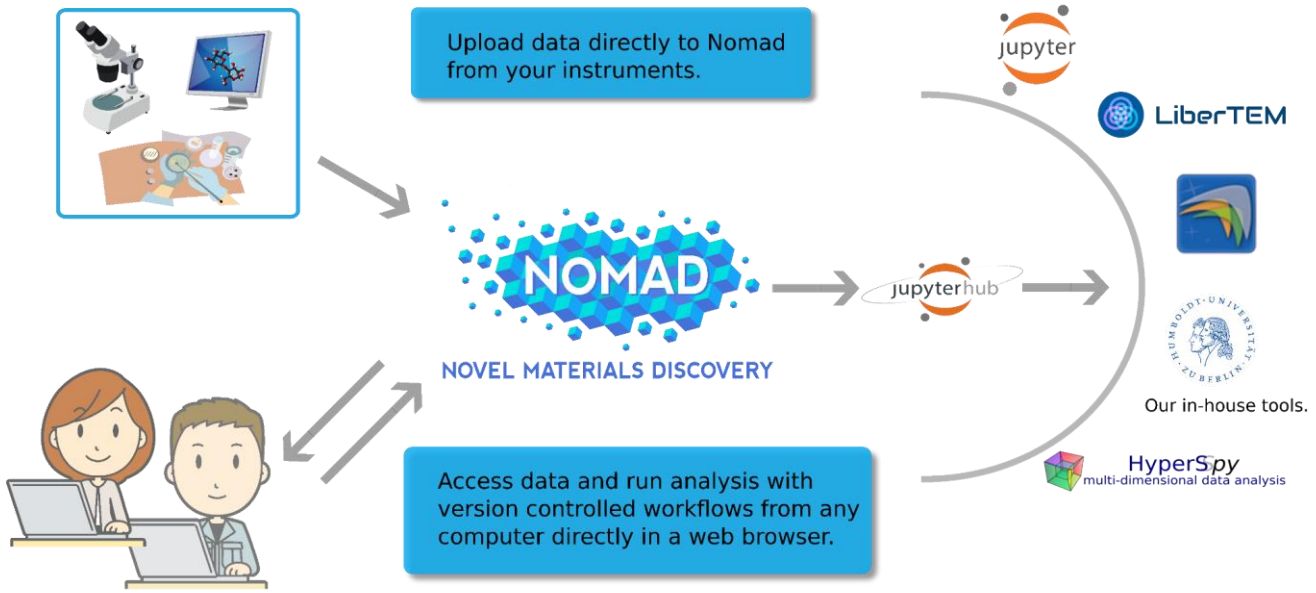


**Figure 1.** The overall workflow of how to upload your lab data to Nomad and process it on your own computer or, better, directly in the cloud using any of your favorite community tools.

https://doi.org/10.1017/S1431927622010996 Published online by Cambridge University Press

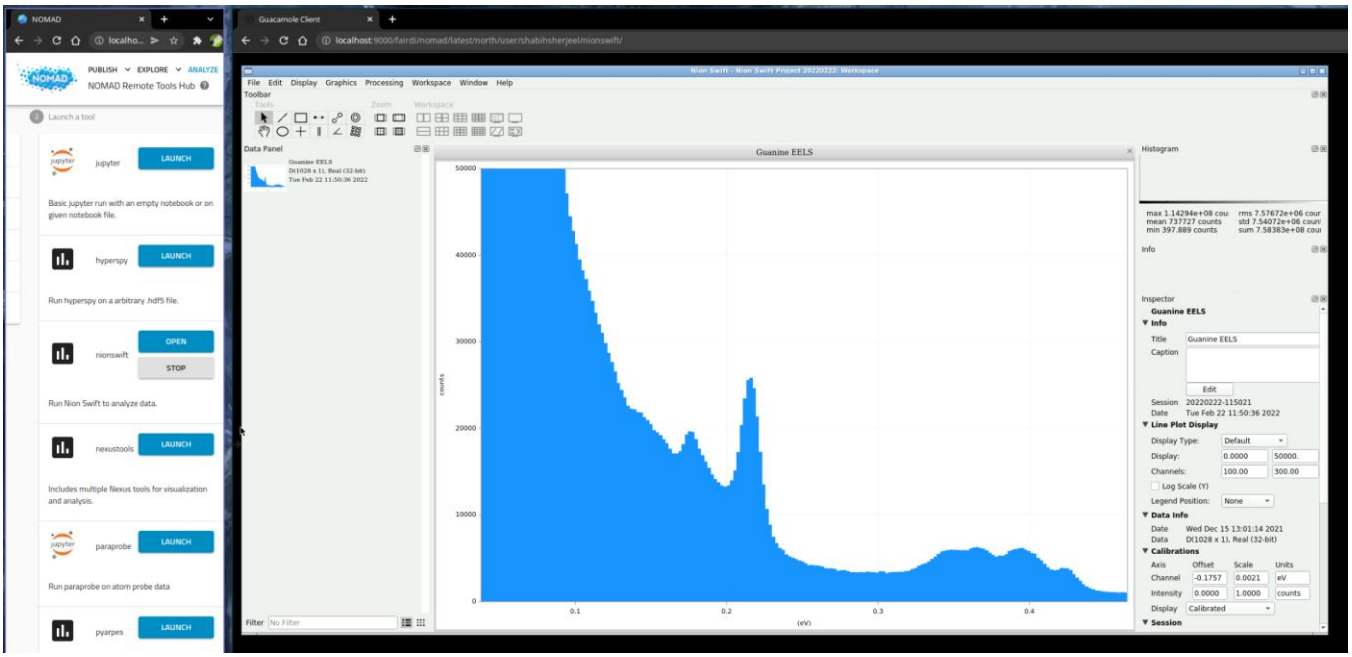


**Figure 2.** A screenshot showing Nion Swift, a desktop scientific image processing software, running in a browser window launched within Nomad to analyze an EELS data file.

References:

[1] M Könnecke et al., Journal of Applied Crystallography **48** (2015), p. 301. doi:10.1107/S1600576714027575
[2] C Meyer et al., Microscopy and Microanalysis **25**(S2) (2019), p. 122. doi:10.1017/S143192761900134X
[3] N CARPi, A Minges, and M Piel., J. Open Source Softw. **2**(12) (2017), p. 146. doi:10.21105/joss.00146
[4] CT Koch, Ultramicroscopy **108** (2008), p. 141. doi:10.1016/j.ultramic.2007.03.007
[5] CT Koch, Micron **63** (2014) p. 69. doi:10.1016/j.micron.2013.10.009
[6] C. Draxl, and M. Scheffler., MRS Bulletin, 43.9 (2018) p. 676-682. doi:10.1557/mrs.2018.208

https://doi.org/10.1017/S1431927622010996 Published online by Cambridge University Press