

NexusLIMS: Leveraging Shared Microscopy Resources for Data Analysis with a Configurable Laboratory Information Management System

Joshua Taillon¹, Raymond Plante², Marcus Newrock², June Lau² and Gretchen Greene²

¹National Institute of Standards and Technology, Boulder, Colorado, United States, ²National Institute of Standards and Technology, Gaithersburg, Maryland, United States

Researchers at electron microscopy (EM) facilities produce large amounts of data from many types of instrumentation, typically using proprietary software to collect and analyze their results. Often, due to finite funding and restrictive licensing terms, data analysis is restricted to a single computer or constrained functionality in an “offline” copy of the software, limiting individual user productivity, as well as the number of researchers capable of data analysis. Luckily, this situation has brought about a sharp increase in the publication of open-source scientific software tools, as evidenced by the success of the *Data Acquisition Schemes, Machine Learning Algorithms, and Open Source Software Development for Electron Microscopy* symposium at the 2019 Microscopy and Microanalysis Meeting [1]. Such tools greatly improve reproducibility and scientific integrity by illuminating all data processing steps, rather than relying on “black box” operations from vendor-supplied software.

While critical for reproducible science, the use of open source data analysis tools is hampered by the lacking data infrastructure at most EM facilities (until recently, NIST included). At a typical facility, raw data is written to hard drives on the microscope “support” computers, which are periodically cleared as space becomes limited, and the onus is on the user to ensure their data is copied regularly (typically via USB) and backed up elsewhere, such as on external hard drives in the user’s office. Perhaps most importantly, this data (and any associated metadata such as instrument configuration) is usually only viewable from within the commercial software packages, leading users to haphazardly embed metadata via individual file-naming conventions that vary from person to person. As a result, this data becomes unmanageable between users, hampering collaboration and data sharing. Eventually, over longer time spans or across guest-researcher or postdoc tenures, datasets are effectively abandoned and then forgotten after publication, even though that data may be useful to others (with the right context). Individually managed (meta)data based on arbitrary user preference fails to meet the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles [2], and limits the scientific productivity of multi-user facilities and institutions.

In an attempt to rectify this situation at NIST, the Electron Microscopy Nexus was chosen as a test bed for a custom-built Laboratory Information Management System (LIMS) [3]. The Nexus facility is an internal multi-user EM cooperative where tool owners agree to share a common reservation system and basic data management model (*i.e.* where data is saved). It encompasses 10 instruments, varying in capabilities (TEM, SEM, STEM, FIB, EDS, etc.), manufacturer, and age, which is typical of many multi-instrument facilities. This variety presents challenges, but also opportunities to design a modular LIMS that provides value to researchers and is flexible enough to handle a wide array of instrumentation and use patterns. The result of this design is NexusLIMS, a data management solution that automatically (without user input) captures experimental context and raw data while streamlining researchers’ abilities to search, explore, edit, and download data from any networked device, a vast improvement over the *status quo*.

NexusLIMS is comprised of two interacting systems (see Figure 1), known as the back- and front-ends. The back-end is responsible for gathering enough contextual information to build a “record” of a given

experiment. This experimental context is captured from multiple data sources (highlighted in green in Fig. 1), including instrument schedulers, metadata recorded by proprietary software, electronic laboratory notebook records, a database of session logs, and the collected data itself. Using this content, a structured eXtensible Markup Language (XML) record is built to fully describe the experiment and data, together with their relevant context. This record is automatically loaded into a customized instance of the Configurable Data Curation System (CDCS) (highlighted in blue) [4], linked to archival storage of data on a central file server. Experiments are displayed to users using customizable XML stylesheets, creating a central system for browsing, searching, accessing, and analyzing experimental records (see Figure 2). This entire process transpires with zero interaction from users, requiring no changes to their current workflows.

By extracting and collecting these experimental records, a centralized repository of experiment metadata and raw data from the EM Nexus Facility has been created, and is currently being used by over 50 research staff and managers at NIST. Users have been able to more easily share their data, search across multiple experiments or samples, or find related experiments using the powerful querying tools provided by CDCS. Project managers can easily find data collected by former interns and postdocs, and all data is automatically backed up without any intervention required by users. In the future (once more data has been collected), this system will enable “in the cloud” data analysis using centralized services (such as Jupyter Notebooks), as well as automated analysis of large amounts of data using AI/ML techniques.

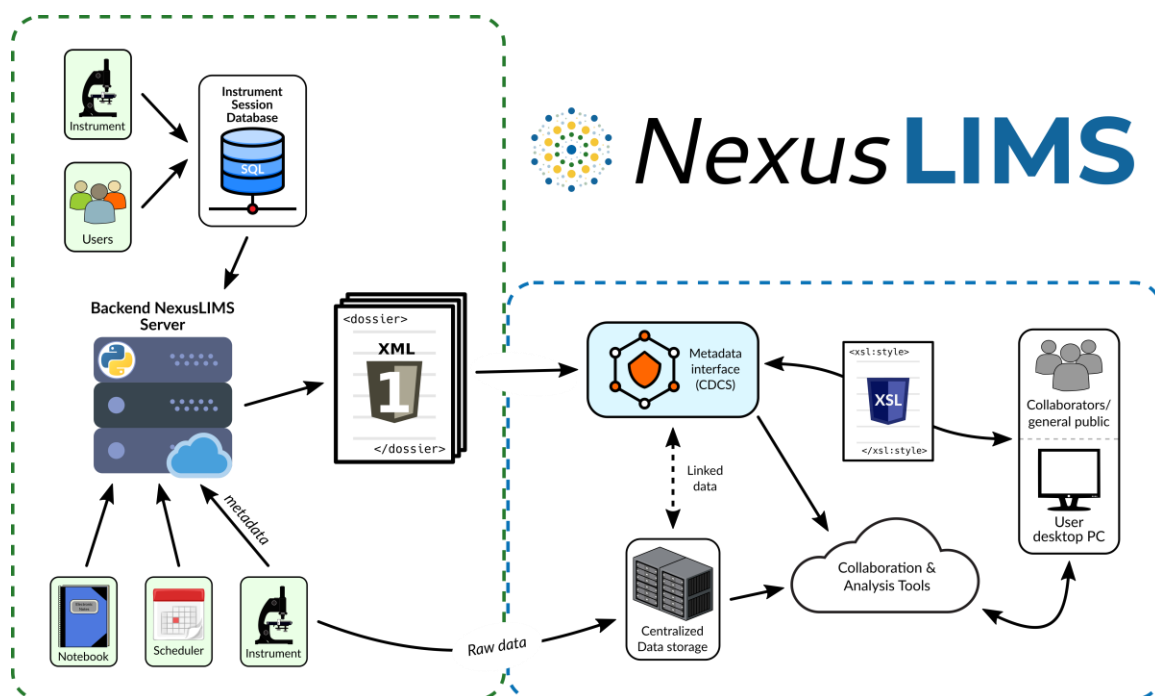


Figure 1. A schematic representation of the NexusLIMS system developed at NIST. The “back-end” experimental record building code is depicted on the left, with the “front-end” user-facing interactive portion depicted on the right. See the text for further details. Such a design enables easy access to data from any location via user interaction or API access, and promotes the FAIR principles.

The screenshot displays the NexusLIMS web interface for an experimental record titled "STEM of LLZO" by Joshua A. Taillon on February 06, 2020. The interface is annotated with colored rectangles highlighting key features:

- Summary (Blue Rectangle):** Provides contextual information including the date (2020-02-06), start/end times, session ID (173), sample name (LLZO), and sample ID (22f1f33e-f2e4-4435-827c-484b17350aa).
- Preview Images (Green Rectangle):** An interactive gallery showing a grid of images, with "Dataset 5 of 56" and "Activity 1 of 4" indicated.
- Setup Parameters (Brown Rectangle):** A table listing technical details:

Setup Parameter	Value
Start time	10:33:41.573654
Acquisition Device	DigiScan
Cs(mm)	1.0
Data Type	STEM Imaging
Detector	GIF CCD
GMS Version	2.32.000.0
Illumination Mode	STEM NANOPROBE
Imaging Mode	DIFFRACTION
Instrument ID	FEI-Titan-STEM-630901
Microscope	Titan90-300_D3094
- Metadata (Red Rectangle):** A JSON-like structure containing detailed acquisition parameters:


```

      {
        "nx_meta": {
          "Acquisition Device": "DigiScan",
          "Creation Time": "2020-02-06T10:35:41.573654",
          "Cs(mm)": 1,
          "Data Dimensions": "(2048, 2048)",
          "Data Type": "STEM_Imaging",
          "DatasetType": "Image",
          "Detector": "GIF CCD",
          "Field of View (µm)": 0.5867976271159243,
          "GMS Version": "2.32.000.0",
          "Illumination Mode": "STEM NANOPROBE",
          "Imaging Mode": "DIFFRACTION",
          "Indicated Magnification": 225000,
          "Instrument ID": "FEI-Titan-STEM-630901",
          "Microscope": "Titan90-300_D3094",
          "Name": "FEI Technical Remote",
          "Operation Mode": "SCANNING",
          "Operator": "JoshuaTaillon",
          "STEM Camera Length": 100,
          "Specimen": "STEM",
          "Stage Position": {
            "X": 77.62950000000001,
            "Y": 53.356300000000005,
            "Z": -55.5565,
            "a": 3.089038927090269
          },
          "Voltage": 300000,
          "warnings": []
        }
      }
      
```

Additional features include a search bar, a list of datasets with columns for Dataset Name, Creation Time, Type, Role, and Head, and a small thumbnail image of a specimen.

Figure 2. A typical display of an experimental record shown to the user via the CDCS-powered front-end. A few features are highlighted: the summary of the experiment (blue rectangle) contains contextual information about the experiment, extracted from the metadata collected when the instrument is reserved. On the right side (green rectangle), an interactive gallery of preview images allows a user to quickly page through the data contained within to gain insight into the experiment's content (often not possible with proprietary data formats). On the left (pink rectangle), a listing of the experiment's content provides context about how many different sets of data were collected during this session. Further below, the individual datasets (files) are clustered into "experimental activities", and for every file, the user can search instrumental setup parameters and metadata extracted from the proprietary data formats in both human-readable (brown rectangle) and machine-readable (red rectangle) formats. Links are also provided to quickly download the original data.

References

- [1] Microscopy and Microanalysis **25**, S2 (2019), pp. 122-245.
- [2] MD Wilkinson et al., Scientific Data **3** (2016), p. 160018.
- [3] J Taillon et al., Microscopy and Microanalysis **25**, S2 (2019), p. 140.
- [4] A Dima et al., JOM **68** (2016), p. 2053.