doi:10.1017/S1431927619001430
Microsc. Microanal. 25 (Suppl 2), 2019

Harvesting Microscopy Experimental Context with a Configurable Laboratory Information Management System

Joshua A. Taillon[1*], Rachel F. Devers[2], Raymond L. Plante[1], Marcus W. Newrock[1], June W. Lau[1], and Gretchen Greene[1]

[1.] Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg MD USA
[2.] Department of Electrical and Computer Engineering, University of Maryland, College Park, MD USA
* Corresponding author: joshua.taillon@nist.gov

Like those in many scientific fields, electron microscopy (EM) researchers generate data from a myriad of instruments, often utilizing expensive and proprietary software to collect and analyze their results. This software is commonly seat-limited, meaning the researcher is required to perform data analysis on a specific computer. If they are lucky, the researcher may have access to an "offline" license of such software, allowing them to use their personal workstation rather than the one attached to a microscope. In multi-user facilities this can sometimes result in researchers abusing software licenses just to examine their data. More critically, this data (and associated metadata) can typically only be viewed from within the commercial software packages, meaning users are left to their own devices to curate their personal data collection, relying on notes/memory, basic file metadata, and naming conventions to identify the significance of each file (Fig. 1). Such methods are unmanageable over long time spans, or when collaborating with other researchers, often resulting in "abandoned" datasets that are forgotten after publication. Needed instead is a centralized and automated laboratory information management system (LIMS) built on the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles [1].

In an attempt to modernize NIST's EM data infrastructure, an open-source LIMS was piloted previously, built on the 4CeeD platform [2, 3]. This pilot program allowed users to quickly browse through microscopy data (stored in proprietary formats) from any computer and search through thousands of images/spectra based on metadata extracted from the data itself. While useful, the system revealed limitations arising from the variability of input data and inflexibility of the underlying platform, requiring a more customized solution. To this end, a LIMS system was designed (see Fig. 2) based off the idea of an EM session *dossier*. This dossier is a high-level summary of an EM session that places equal importance on the data and experimental context (*i.e.* sample information, experimental purpose, instrument metadata, etc.). The dossier is a structured eXtensible Markup Language (XML) document created automatically without researcher intervention, and is linked to centralized archival storage of that session's data. These records are stored in an instance of the Configurable Data Curation System (CDCS, formerly MDCS) [4], which provides user-friendly interfaces for browsing, querying, and disseminating information from the database, built using XML stylesheets. This system is built on the FAIR principles and greatly facilitates open and reproducible science by promoting data discoverability.

References:

[1] MD Wilkinson et al., Scientific Data **3** (2016), p. 160018.
[2] RF Devers et al., Microscopy and Microanalysis **24** (2018), p. 564.
[3] 2017 17th IEEE/ACM Int. Symp. Clust. Cloud Grid Comput. (2017) p. 11.
[4] A Dima et al., JOM **68** (2016), p. 2053.

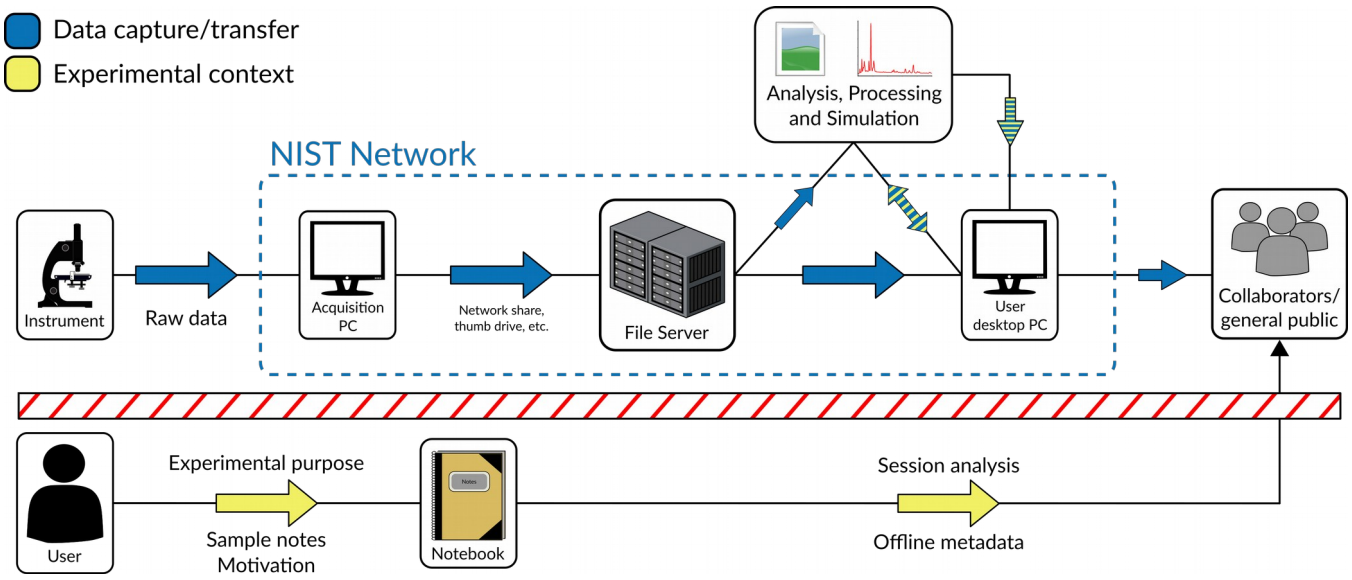

Figure 1. Current typical workflow of a microscopist at a shared-use facility. The flow of data and context are represented in blue and yellow, respectively. Experimental context is only recorded (if at all) manually by the user in a notebook or final publication. A stark divide exists between this context and the data/metadata collection, storage, and analysis processes that are controlled by instrumentation and infrastructure. As a result, massive amounts of data are abandoned on individual researchers' computers (or deleted), violating the FAIR data principles, and preventing the results from being analyzed further.

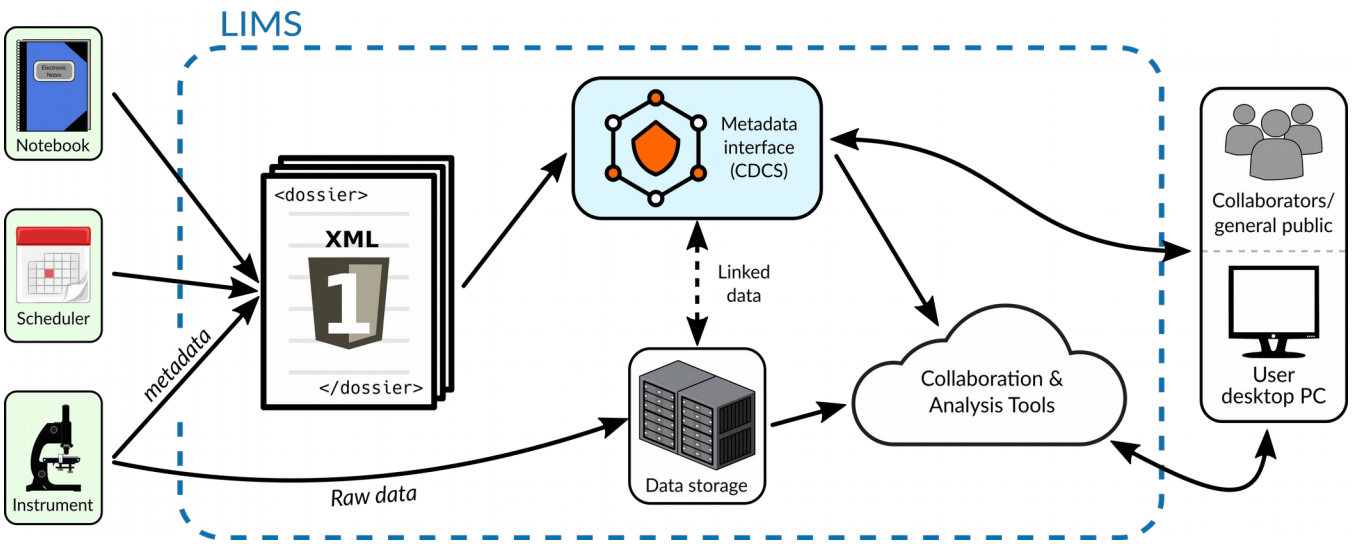

Figure 2. Schematic of the microscopy Laboratory Information Management System (LIMS) developed at NIST. Experimental context is captured from multiple data sources (highlighted in green), including instrument schedulers, metadata recorded by proprietary software, electronic laboratory notebook records, and the collected data itself. Using this content, an XML dossier is built to fully describe the experiment and data, together with their relevant context. This dossier is loaded into the Configurable Data Curation System (CDCS) (highlighted in blue), linked to archival storage of data on a central file server. Experiments are displayed using individually customizable XML stylesheets, creating a central system for browsing, searching, accessing, and (in the future) analyzing experimental records. Such a design enables easy access to data from any location, and promotes the FAIR principles.

https://doi.org/10.1017/S1431927619001430