2912
doi:10.1017/S1431927622010923
Microsc. Microanal. 28 (Suppl 1), 2022
© Microscopy Society of America 2022

Lessons Learned in Building a Modern Microscopy Data Infrastructure at NIST

Joshua A. Taillon[1*]

[1.] Material Measurement Laboratory, National Institute of Standards and Technology, Boulder, CO, United States.
* Corresponding author: joshua.taillon@nist.gov

In the past decade, the electron microscopy (EM) community has witnessed a paradigm shift in nearly every aspect of the experimental workflow. Hardware advancements and innovative data collection methods continually push the limits of performance in imaging and spectroscopy [1], while novel analysis techniques adopted from the data science community have led to the emergence of "data-driven" microscopy [2], where previously hidden insight arises from statistical analyses of large amounts of data. These advancements have given rise to new challenges in how we store, manage, and process the ever-increasing amounts of data produced, requiring the adaptation of EM researchers and collaboration with data scientists to extract the most insight.

The "data problem" in electron microscopy is multifaceted and impacts all portions of the research lifecycle, from initial experimental planning, through the data collection and analysis, to eventual data publication and maintenance, necessary to ensure data remains available even long after a manuscript has been published. Adequately addressing these challenges is needed to guarantee research data meets the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles [3], which promote the maximal use (and reuse) of research data. Efforts like these (seen perhaps by some as a quixotic pursuit of unrealistic scientific ideals) have real utility, and can revolutionize what sorts of science are possible when embraced *en masse* by a scientific community. For example, the adoption of open standards and common data models by the X-ray crystallography community has enabled automated experimentation together with wide-scale sharing and reusing of data [4]. Such an ecosystem in the electron microscopy community does not yet exist, but ongoing efforts towards the vision outlined by Spurgeon *et al.* [2] and those discussed in this symposium will help to revolutionize the way electron microscopy research is performed.

Like many institutions, the National Institute of Standards and Technology (NIST) has a wide range of research areas addressed by many independent researchers using a multitude of various commercial and home-built instruments. This heterogeneous environment makes it difficult to develop "one size fits all" solutions to the data challenges presented previously, but improvements in NIST's data infrastructure can enhance even disparate research workflows. In this talk, I will discuss some of the ways in which the electron microscopy community at NIST is navigating this transition to a data-driven microscopy paradigm and present a number of lessons we have learned during the process.

Specifically, I will describe improvements that have been made to both physical and digital infrastructure to safely network potentially obsolete or vendor-controlled instrument control computers [5, 6], as well as a multi-division effort to automatically harvest, backup, and index research data files from across the institution to a centralized network location, including automated and customizable metadata extraction. I will also present ongoing developments with (and lessons learned from) NexusLIMS, an open source laboratory information management system built in support of the EM community at NIST [7]. Finally, I will review the "Open Access to Research" framework [8], which

https://doi.org/10.1017/S1431927622010923 Published online by Cambridge University Press

enables the dissemination and public access to data resulting from research at NIST. While each of these infrastructure developments have been in response to specific needs of NIST, they have been made in collaboration with and consultation with external community organizations to ensure these efforts will be compatible with emerging community standards and expectations. The requirements, lessons, and approaches to solutions presented in this talk should be generalizable to nearly any organization, and will aid those undertaking similar efforts.

References:

[1] Z. Chen et al., Science **372** (2021), pp. 826–831. doi: 10.1126/science.abg2533
[2] S. R. Spurgeon et al., Nature Materials, **20** (2021), pp. 274–279. doi: 10.1038/s41563-020-00833-z
[3] M. D. Wilkinson et al., Scientific Data **3** (2016), p. 160018. doi: 10.1038/sdata.2016.18
[4] I. Bruno et al., Data Sci. J., **16** (2017)**,** p. 38. doi: 10.5334/dsj-2017-038
[5] M. Helu and T. Hedberg, Procedia Manuf., **1** (2015), pp. 86–97. doi: 10.1016/j.promfg.2015.09.066
[6] J. H. J. Scott, Microsc. Microanal., **21** S3 (2015), pp. 373–374. doi: 10.1017/S1431927615002664
[7] J. A. Taillon et al., Microscopy and Microanalysis **27**, (2021), pp. 511-527. doi: 10.1017/S1431927621000222
[8] G. Greene, R. Plante, and R. Hanisch, Data Sci. J., **18**, 30 (2019). doi: 10.1017/S1431927621000222

https://doi.org/10.1017/S1431927622010923 Published online by Cambridge University Press