529
doi:10.1017/S143192761500344X Paper No. 0265 Microsc. Microanal. 21 (Suppl 3), 2015

Evolution of Data Acquisition, Storage and Analysis in a Multi-user Facility

Nestor J. Zaluzec

Electron Microscopy Center, Nanoscience and Technology Division, Argonne National Laboratory, Argonne, Il, USA

When the principle output of electron optical instruments consisted of images recorded on film, data storage and archiving was a relatively simple task. Each image was logged in an appropriate paper notebook and the negative or positive print filed in cabinet. Forty years ago I kept track of some 10,000 negative recorded during my thesis research using that method. Although that sounds like a daunting task, that was only for the images I recorded for my own research it did not include information for other users of the EM facility at the University of Illinois Materials Research Lab that also used our core instruments. Even more importantly, back in the 70's we were just in the early stages of developing both x-ray and electron loss spectroscopy in the instruments. To organize that data, which also needed archiving it was necessary to develop a scheme to do a similar task, however, the challenge was formatting and storage protocols as something as simple as used for film did not exist. Now that forty years have passed, and technology has changed several questions are relevant: 1.) can that data still be accessed, 2.) do we have software that can access archival data, 3.) how do we preserve current data and the associated meta-data which documents today's experiments, and 4.) how do we overcome the issue of obsolescence of modern computing accessory hardware on instruments which far out live the life time of a typical operating system.

Obsolescence of peripheral computing hardware is one of the biggest challenges. Modern IT staff principally worry about keeping all systems up to date with the latest operating systems and off the shelf "office" software and stopping spam/virus attacks. They rarely if ever deal with scientific resources, those problems are relegated to the core facility staff. As the lifetime of major instruments can exceed a decade, I have found it is essential that core facilities maintain backup computational hardware and software resources to replace items which simply fail with age. Older hardware and software can be difficult to find and without it some instruments can become dead weight. Because custom software and hardware are integrated for controlling our complex, computationally mediated instruments it is not simply just the task of upgrading the hardware and software to keep up with the OS updates, we must insure that something as trivial as a $3K computer and a few hundred dollars of software takes down a multi-thousand/million dollar research too.

Movement of data to servers which are not only backed up but more importantly connected using secure communications protocols is essential. Users should never be allowed to directly connect portable storage devices to core instruments, but rather data on core instrument should be transferred to remote servers and users required to download their data from therein. Using secure protocols similar to SFTP [1] or Globus tools sets [2] for these jobs should be standard procedure.

Obsolescence of data formats is even more significance. Data (images, diffraction patterns, etc…) should never be stored in compressed image format (JPG, GIF) instead full raw data or at the least full bit depth files without compression should be used.

Manufacturer's formats are frequently used as substitutes, which is fine as long as software platforms are also archived that can read and process that data. If a software platform is about to be retired then translator program should be procured or developed to allow access to the data from the archive. Spectroscopic data should be handled no differently, and translators or alternative formats such as MSA or HMSA Spectral Image file formats should be considered [3-5]. Translators into public domain file formats also simplifies sharing of data with colleagues when commercial programs using proprietary formats are used to acquire data.

Record keeping also needs to shift along with modern data needs. While it is sometimes possible to incorporate meta-data into the image or spectroscopic files which we are now storing, this really does not fully document our experiments. The traditional paper notebook which once followed me into the laboratory has been replaced by an electronic notebook. In it I can fully annotate an experiment, using not only handwritten notes, but in addition (and with much better readability) typed text as well as photographs of experimental setups and/or screen captures. The added ability to convert and upload the pages of an eNotebook into an archivable PDF file and store it concurrently with the data files it documents is not only a logical but necessary protocol.

The last challenge is a simple monetary one. Namely whose responsibility is it to store and archive data. It the past, when we principally used film, all data was stored in the facility. Today that is not practical. Maintaining a reasonably sized file server of ~ 10-20 Tbytes is neither expensive nor difficult and that is a task which can be readily delegated to an IT group. The approach which is most reasonable to take, in a multi-user core facility, is that the individual user is responsible for the long-term archiving of their data sets. Temporary storage of data for a fixed period of time, the length of which can be determined by each local facility, on the facility servers is reasonable. Users are expected to use secure protocols to move data from the temporary storage facility to their own systems for off-line analysis and record keeping.

By the way, as I am careful, as well as a packrat as many of my colleagues at ANL will attest to, I can still read and process the images and data, which I recorded back in the late 70's, but then again I also wrote some of the original code to store and process that data.

References:

[1] STFP http://en.wikipedia.org/wiki/SSH_File_Transfer_Protocol
[2] Globus https://www.globus.org
[3] RF Egerton, *et al*, Proc. of the EMSA, San Francisco Press, (1991) p. 526.
[4] International Organization for Standardization, standard ISO 22029:2003.

[5] A Torpy *et al*, Microscopy and Microanalysis **19 S2** (2013), p. 830

[6] Work supported by the U.S. DoE, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357 at the Electron Microscopy Center, Nanoscience and Technology Division of Argonne National Laboratory

https://doi.org/10.1017/S143192761500344X Published online by Cambridge University Press