1082
doi:10.1017/S1431927621004074
Microsc. Microanal. 27 (Suppl 1), 2021

Challenges and opportunities for data management and collaborative analysis in shared electron microscopy facilities

Joshua Sugar

Sandia National Laboratories, Livermore, California, United States

The amount of data that one researcher can collect with an electron microscope has grown immensely in the last couple of decades. The purpose of this symposium is to discuss the challenges (and opportunities) that this explosive growth has created. When I started learning about electron microscopy, I developed negatives in a dark room and then digitized them with a scanner. At that time, the amount of data I could collect was not limited by microscope time, but rather the processing of the data into a useable digital form. Since the data did not initiate in a digital form, it made sense for each microscope user of the shared facility to manage the storage and digitization of their own negatives. As technology improved, data became "born digital", but the tradition of each user managing the storage and backup of their own data continued.

As the complexity of the electron microscope hardware increased, so did the number of proprietary software packages that we use to operate them. It is common to find one microscope that may need 3 or 4 (or more) proprietary software packages for data acquisition and analysis. Because the data acquisition time is so valuable on the microscope, there needs to be "offline" locations where users can analyze their data. Therefore, there is an opportunity for facilities to create automated systems for transferring, storing, and backing up data from the microscope to some cloud-based or other archival storage system. However, the details of keeping the data uncorrupted, version control, and providing access to multiple users may be complex and varied depending on the IT environment of each facility. It is also important to have a digital record of what was done for the collection of each piece of data that goes beyond the metadata stored in the proprietary format. A record with notes from the operator should be automatically stored in a non-intrusive and automated way with each data set. One example of an online job tracking tool with file location and sample notes is shown in Figure 1.

In addition, with the increasing complexity and dimensionality of data acquired, the trend of multiple researchers performing analysis on the same data set is common for large teams and research projects. The use of machine learning and artificial intelligence (AI) computing methods for data analysis may also require that teams of researchers have access to the same microscopy data to try different analysis routines. A unified system is needed that can provide multiple researchers simultaneous access to proprietary software for analysis or conversion to a more open-source platform to answer the relevant scientific questions. One possible solution is the virtual server shown in Figure 2. However, there is an opportunity for the community to create a better system that is backed up, accessible to multiple users, capable of handling any data format, and provides version control so that any user can always go back to the original data collected on the microscope. The goal of this symposium is to provide a platform for researchers to discuss potential solutions that solve these problems at the facility level and not for each individual researcher separately. I hope that the benefits and potential limitations of each solution will be demonstrated.

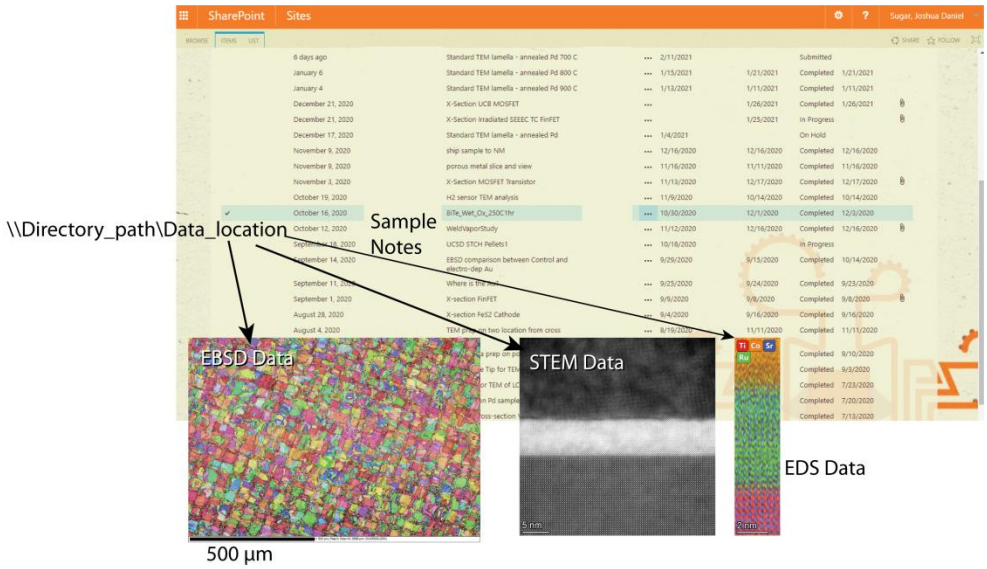

Figure 1. Figure 1: Example of a job tracking system that tracks the date, sample, data directory location, notes, and work done on a microscope. It points to a cloud-based directory where corresponding data from several different microscopy techniques can be stored.

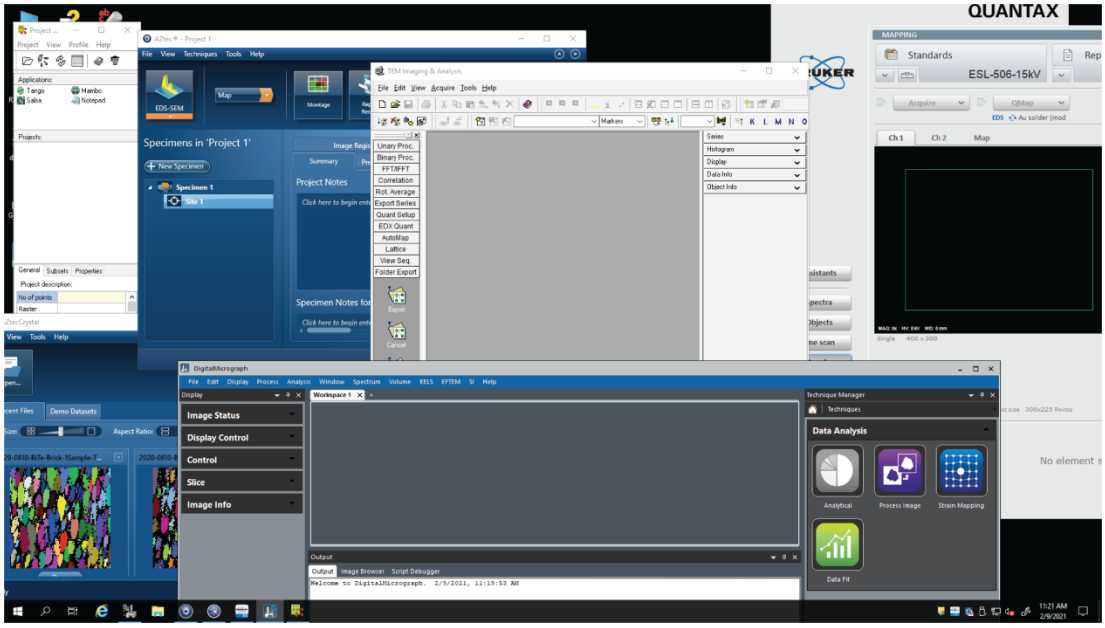

Figure 2. Figure 2: Example of a cloud-based virtual server that can be used as a multi-user offline analysis computer. Multiple users can use the system simultaneously, and multiple proprietary software licenses can be managed on one system. Several proprietary software packages are shown running at the same time. However, some needed software packages do not function in the virtual server environment because of graphics card limitations, so better solutions exist.

References

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

https://doi.org/10.1017/S1431927621004074 Published online by Cambridge University Press