Microsc. Microanal. 21 (Suppl 3), 2015
doi:10.1017/S1431927615002640 Paper No. 0185

Now What am I Going to do…..This is Going to be a Large Data Set.

C. A. Brantner[1]

[1.] GW Center for Nanofabrication and Imaging, The George Washington University, Washington, D.C., USA.

Many of us have gone digital in our microscopy labs. Our instruments now come with sensitive, sophisticated digital cameras. We can record enormous amounts of data from any given experiment. What are the things that we have to know and do to ensure that we obtain useful data?

The instruments/tools that we are using as well as the cameras have many settings that need to be optimized for the collection of data that can be mined for quantification, large area imaging or 3D processing. Our fluorescence/light microscopes, electron microscopes, flow cytometers, x-ray spectrometers, etc. all produce kilobits, megabits, even gigabits and terabits of image and metadata files. Once the data has been collected, now there are questions about how to transmit the files to storage or to an off-site collaborator. Speed and efficiency are required to move large data faithfully. In some cases the policies of the institutions where we work do not have the provisions or the capacities to handle transfer of large data sets from your instruments. It takes collaboration with IT departments and administration to set up the hardware and policies to use your large data sets.

Storage is another topic that we know to ask questions about, but what are the questions that are most important to ask? There may be a gigabit of data from today's experiment, which you have the capacity to store, but what of the experiment tomorrow or the next day or from others in the lab? Now there will soon be more data than the hard disc of the computer can handle. What is the next step? An external hard drive? A shared drive? SFTP sites? A cloud account? Large data sets need to be stored in close proximity to the acquisition device to improve the quality control of the transfer. Is an electron microscope facility/core responsible for storing data that is generated there? What policies do you have in place?

Data security can be another topic that causes much confusion and anxiety. Again there is the need to work with our institutions' information security people to understand what needs to be done with data as it relates to policy of access. Is encryption needed for the transmission or the storage of the data? Passwords? Firewalls? If so, how will you need to accomplish that?

How do you display the data so that you can extract meaningful trends from all of the gigabits that you are now storing? There are many software packages written to handle this. What are the pros and cons of such packages? How can you share data with an off-site collaborator? There are web-based solutions for making data widely available. These have been used for crowd-sourcing that some researchers are using to analyze data.

Our speakers in this symposium will tackle these questions and more before the end of the day.

References.
[1] Ideas from 2014 FOM FIG symposia brain-storming session.
[2] Ideas from 2012 NIST Big Data Workshop via discussion with Dr. Anastas Popratiloff.