

Networked Data Storage and Analysis for the Wisconsin Regional Materials Network

Jon J. McCarthy¹, Paul M. Voyles², Julie A. Last³, Davis R. Bittner⁴, Ian W. Sadkovich⁴

¹. Wisconsin Materials Institute, University of Wisconsin, Madison WI, USA

². Materials Science and Engineering Department, University of Wisconsin, Madison WI, USA

³. Materials Science Center, University of Wisconsin, Madison WI, USA

⁴. Computer-Aided Engineering Department, University of Wisconsin, Madison WI, USA

Modern materials characterization instruments produce data in prodigious quantities which require sophisticated, often proprietary software to analyze. Making data accessible and providing access to specialized software is a significant challenge for central materials characterization facilities like the University of Wisconsin-Madison's Materials Science Center (MSC). The traditional MSC strategy for data is to leave it on the computers attached to this instruments for an unspecified amount of time, then delete it and trust that the users have made copies. The strategy for software is for MSC users to install software on their own computers when that is (rarely) possible, or, in the more common case of limited available licenses, for them to access software either on computers attached to instruments or on offline analysis computers maintained by the MSC in the Materials Science and Engineering building. Use of instrument computers for analysis is terribly inefficient, since it blocks the instrument for data acquisition, and the offline analysis computers are overused, out of date, and virus-ridden. The result is fewer people using the instruments at all and only some users taking best advantage of the available analysis tools. Our approach to overcoming these obstacles is to use a networked data analysis (NDA) system, briefly described here.

The NDA system provides delivery over the Internet of data analysis applications with access to a central data repository. Figure 1. shows the basic system configuration. Users acquire their data on the instruments, either in person or by Internet-enabled remote access. Data is copied off the instrument computers to a centralized data store. The data store is connected to an analysis software server. Users can connect to the analysis software server from any Windows or Mac computer using any Internet connection faster than 1 Mbit/sec to perform analysis. Raw data and Analysis results are stored on the server, and can be copied to the users' computers over the Internet, or be left in the data store.

Analysis software is installed on backend Dell servers running the application virtualization product Citrix XenApp [1]. Remote users are able to use the tools without the struggle of installing and configuring the software. In addition, the servers running the software are collocated with and have high-speed connections to the data to be analyzed. XenApp is currently used to by UW College of Engineering (CoE) Computer Aided Engineering (CAE) organization to deliver access to 18 different engineering applications for educational use by CoE undergraduates, graduate students, faculty, and staff. The XenApp client runs on the user's computer, but the application runs on a CAE-managed Windows server. XenApp delivery is legal for many of the end user license agreements (EULAs) for data analysis applications. User access to the analysis/data server is controlled by a configurable user group which can include University of Wisconsin engineering students, faculty, or staff, researchers from other disciplines, or from outside UW. Outside users cannot access other UW resources like journal subscriptions or email, or other computers on the CoE network. One XenAPP server is capable of supporting between 10 and 50 simultaneous users for typical engineering applications using 24 cores

on two processors and 256 MB of RAM. The initial data store consists of 4 TB of high-speed disk storage with redundant off-site backup. Initial storage was provisioned based on one year of data from tomographic (S)TEM, serial section reconstruction from a FIB, a spectral imaging XPS, and several XRD instruments. A typical tomogram is 3 GB of raw data and an additional 3 GB of analysis files, and up to 5 tomograms can be acquired per day. Estimating 100 days of heavy tomography use in a year yields 3 TB of data. XPS and XRD data generation is estimated at 1 TB per year. The system now has 10 XenApp licenses for concurrent users, and 200 licenses for CoE faculty, staff, and student users. Analysis software includes ImageJ, NIH IMOD, XPS Thermo Scientific Avantage package and several diffraction packages from Bruker [2,3,4].

References:

- [1] XenAPP, www.citrix.com/products/xenapp
 [2] C.A. Schneider, W.S. Rasband, and K.W. Eliceiri, *Nature methods* 9 (2012), p. 671.
 [3] J.R. Kremer, D.N. Kisielowski, and J. R. McIntosh, *Journal of Structural Biology* 116 (1996), p. 71.
 [4] The authors acknowledge The Wisconsin Materials Institute for financial support, and Thermo Fisher Scientific and Bruker for software license arrangements.

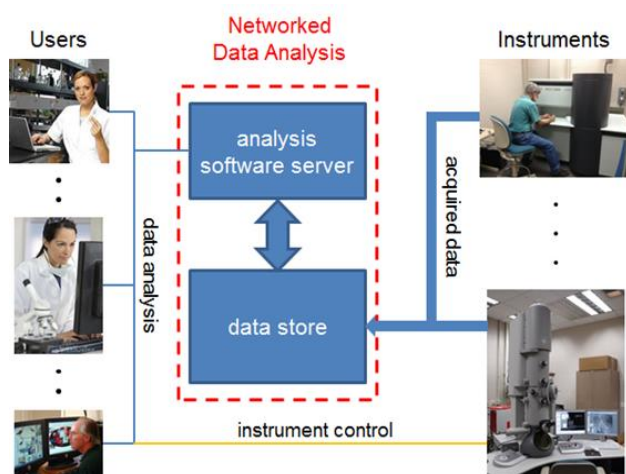


Figure 1. Schematic of the networked data analysis system. Data is acquired on the instruments, then moved to a central data store, connected to the analysis software server. Results can be kept on the data store and copied to the user's computers, as can raw data.

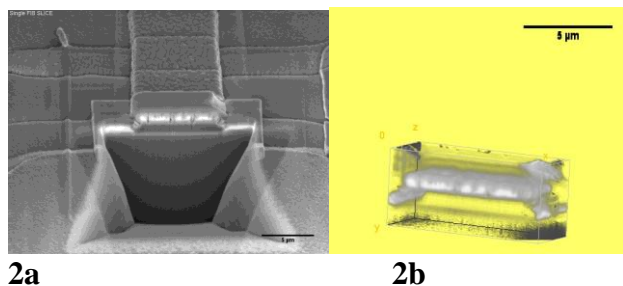


Figure 2. **2a)** A serial section of a MOS FET collected on a Zeiss 1540 Cross Beam. **2b)** 3-D volume view of the Gate section of this FET. Both images from data uploaded to the NDA system and processed with ImageJ2 via XenApp.