



Big Data sessions bring variety of opinions

Building from a series of features published in *MRS Bulletin* and *Materials360® Online*, the Materials Research Society (MRS) sought to expand community dialogue on “Big and Open Data for Materials Research” through a panel of experts and diverse perspectives at a lecture (Symposium X session) at the 2013 MRS Fall Meeting in Boston.

Big Data and open data are topics of increasing interest and discussion in the materials community and beyond. Advances in computational modeling as well as experimental output are introducing a shift in how research is conducted and how data are analyzed and shared. Broad access to data holds the promise of advancing the speed to new discoveries. However, this also raises questions relating to quality, validation, and reproducibility of data, and intellectual property issues.

In the session, Tia Benson Tolle of The Boeing Company and current MRS President, served as moderator. She asked if Big Data is critical to materials research and advances.

“Materials science has existed and accelerated without big data,” said Mary Galvin of the National Science Foundation Division of Materials Research, one of the panelists. “The critical part is the cultural shift that has to occur. We must integrate data mining, theory, and computation with experimentation, which is the real key and where we can begin to accelerate materials science. The infrastructure of creating the database is just one component of that.”

Another comment was that capabilities are needed to combine data from different sources, experiments, and times, and, in many cases, that is going to involve the merging of several data sets that have complementary information. “We need to ensure that can be done seamlessly and with appropriate statistics so we can extract the maximum scientific value out of the experiments,” said panelist J. Michael Simonson of Oak Ridge National Laboratory.

“We are starting into the data extensive research cycle. It doesn’t mean that theory, experiments, and computation aren’t relevant. But data-intensive science combined with these mechanisms of exploration are really going to change the scientific method,” said Jim Pinkleman of Microsoft Research. “Big data today will be small data 10 years from now. Big Data will absolutely change science, whether we decide for it to happen or not.”

The open data policy memorandum issued by the White House Office of Science and Technology Policy last year has had mixed reactions. Some see the policy as a threat, others as an opportunity. Nicola Marzari of École Polytechnique Fédérale de Lausanne, Switzerland, is in favor of the open source movement. “By making tools publicly available, you empower scientific communities all over the world. This will significantly change the panorama in the next decades.” It can help to verify information by having members of the community discover if data or tools are correct. It also allows reproducibility and empowers other researchers to build from information.

After the Symposium X session, two separate focus groups, comprising Early Career MRS Members and International (non-US) MRS Members, were convened to discuss opportunities and challenges connected to Big Data and data sharing.

From the international perspective, the members felt political (government and industrial) challenges would be much more difficult to work through than scientific challenges. They want to resolve this by learning from other research communities that face similar obstacles and by creating better bridges among the scientific, industrial, and government communities.

The group saw potential for big/open data to preserve previously unutilized information, predict behavior and structure, provide a better starting point for



research, and leverage the power of data mining.

Discussed challenges included the potential loss of intellectual property, the need to develop data standards and to ensure data integrity, the need for open code as well as open data, and changes to the competitive landscape for industry and participating nations.

Participants were also concerned about research funding hinging on data being published, and nations moving at different speeds to enforce open data mandates.

The early career group saw the potential to accelerate science and obtain better materials to improve the quality of life; engage in public knowledge development, possibly increasing the interest in materials science; and provide decision makers with better data. Foreseen challenges were the amounts of data to manage and reliability, the behavior of participants and issues of fairness, as well as the cost of organizing the data.

To view the OnDemand® video of the Symposium X session, visit www.mrs.org/big-data-video. For further discussion on Big Data, Tutorial YY on “Recognizing and Addressing Big Data Problems” is being offered at the 2014 MRS Spring Meeting in San Francisco, Calif., Monday, April 21, 1:30–5:00 p.m. in Moscone West.

Lori A. Wilson