**Workshop on the Convergence of Materials Research and Multi-Sensory Data Science - Summary**

Lehigh University, in partnership with The Ohio State University, hosted a National Science Foundation (Award #1821933) sponsored 2½-day, international “Workshop on the Convergence of Materials Research and Multi-Sensory Data Science” at Bear Creek Mountain Resort in Macungie, PA from June 11-13, 2018.  The purpose of this workshop was to bring together scientists from allied disciplines in the basic and social sciences and engineering to address many issues involved in multi-sensory data science as applied to problems in materials research.   It served to forge ongoing partnerships for the participants and to establish a common language and framework for continued dialogue among disciplines.  One outcome of this gathering was the creation of a roadmap that highlights the salient issues identified by the participants and sets out a list of milestones that must be met to move this effort of promoting multi-sensory data science forward.

The participants first considered the data analytical challenges that materials science research is, or will soon be, confronting.  As is commonly understood in this era of "Big Data," these challenges will include issues of volume, variety, velocity, and veracity, although not all elements will be present in all applications.  Focusing first on the data, the group noted that serious issues will arise from the need to cope with very large volumes of materials data, both in the form of microscopy images and from other sources.  Organizing and sharing data will need to move beyond the *ad hoc*, informal methods used today, which are often the personal decision of a single PI or perhaps a research group.  From these considerations, the following needs were inferred, namely: 1.) mechanisms for easy interchanges between data formats, 2.) common schemas and metadata, 3.) data that are deemed as important in different subfields, and 4.) the ability to augment schemas and metadata as new knowledge is developed.

The role of data curation was a topic of special concern to the workshop attendees as it was recognized that issues of data ownership in materials science and engineering center on the tension between privacy and community availability. Clearly, shared, federated data open new vistas for materials research, development and deployment.    As we strive for broad sharing of materials data, we must address a variety of challenges including annotation of data (metadata) and adherence or creation of policies to meet the community’s needs.  It was recognized that a particular challenge to community data centers around developing systems to value data appropriately. As materials researchers move towards valuing shared data, the need for data hosting and curation will likely grow.    Thus, one goal should be community adherence to agreed semantic standards and the FAIR principles, as exemplified in the NIST CDCS project, to create a new generation of more flexible data curation options.

The workshop participants also discussed the many aspects of the convergence of informatics, simulation and experimentation.  While some in the scientific community consider informatics to be a new, fourth “pillar” of scientific discovery, along with theory, experiment and simulation, the participants focused instead on the synergy of these approaches rather than on the creation of a new, independent pillar. Towards this end, several challenges were identified, including: 1.) leveraging informatics tools in both experimental or simulational contexts to explore large parameter spaces, 2.) employing data to inform and refine specific models, and 3.) extrapolating from a data set at a given point in space and time using simulation.

The importance of cognitive science in information processing was also recognized given that material scientists often employ sophisticated equipment, vast data sets, and complex scientific theories within their work environment. Indeed, the fields of cognitive science and psychology have long-standing research traditions in advancing the information processing enterprise, and existing research is currently available to help workers to understand the perceptual, motor, and cognitive demands confronting material scientists and limiting the speed of scientific advancement in the field. Throughout the course of the workshop, a number of specific areas of research expertise in human cognitive processing were discussed, with an emphasis on visual attention, visual search and spatial cognition.

Finally, the participants discussed the role of education in the convergence of materials research and data science and noted many associated challenges.  One central challenge is the development of new software and other tools to convey important concepts. In addition, it is essential to exploit technological innovations that permit students to learn in new ways. Exciting developments in this area include mixed realities, simulation and immersive visualization, and they may be employed to pre-educate students before they use expensive instrumentation (e.g., electron microscopes). Moreover, as artificial intelligence becomes more prevalent, there will be novel instructional opportunities to help students understand, integrate and navigate a machine’s learned intelligence in addition to their own. There are several benefits to this mode of instruction, including the encouragement of student experimentation with instrumentation and the development of an intuitive understanding of the workings of various instruments.

Many specific recommendations related to the topics summarized above were made in the roadmap. In summary, the participants recognized that the convergence of materials science with cognitive and computer science is a difficult undertaking, but there are also substantial benefits to all involved. These benefits include the following:

1. A powerful set of analytical tools that supplement those already in the materials science toolbox.
2. A greater appreciation on the part of materials and computer scientists of the role of cognitive processes (e.g., visual attention and task switching) in the interpretation of data.
3. The development of a common language for the interpretation of data that mitigates existing interdisciplinary barriers to understanding important concepts.
4. A better understanding of the importance of cross-cutting issues associated with data curation that will dictate the availability and utility of data to researchers in all fields.