



Support & Integration of Materials Research & Education: Beginning Steps toward Cyberdiscovery

Report from the NSDL Materials Digital Library Workshop held in Arlington, Virginia
September 26, 2006, sponsored by the National Science Foundation.

Workshop Organizers

Materials Digital Library Pathway:

Laura Bartolo, PI
Kent State University

Sharon C. Glotzer, Co-PI
University of Michigan

James A. Warren, Co-PI
NIST/MSEL

National Science Digital Library:

Kaye Howe, PI
UCAR

Dean Krafft, CoPI, Technology
Cornell University

WORKSHOP REPORT OUTLINE

A. Workshop Aims

1. NSDL & MatDL explore collaborations with MRSECs

B. Key Issues & Needs

1. Enhance materials research through innovative approaches to support how scientists want to work
2. Support effective materials data organization, exchange, dissemination, and reuse by processing information as it is generated in the lab

C. Group Discussion & Consensus

1. Education: Near term - Pilot collaborative, participatory activities
2. Shared Facilities: Near term - training; Long term - organize data
3. Research: Long term - DL to archive, organize, preserve data

D. Plan of Action Summary: Path Forward & Recommended Beginning Next Steps

- Invite MRSECs to participate in Soft Matter Wiki & MatForge
- Work with MRSECs to use Soft Matter Wiki in REU programs
- Use wiki technology with summer programs for high school teachers
- NSDL & MatDL participate in upcoming MRSEC Education Coordinators Meeting
- Establish a gallery of searchable experimental images on NSDL MatDL
- Investigate interest in joint MRSEC/NSDL/MatDL pilot virtual lab focused on Soft Matter integrating remote instrumentation for use with REU programs
- Explore combined standards effort among workshop participant groups to determine metadata and provenance information for materials data

Appendix A: Workshop Agenda

Appendix B: List of Participants



A. Workshop Aims

On September 26, 2006 MatDL hosted a small workshop at NSF to identify key issues and opportunities in the enhancement and integration of materials research and education. To help focus the discussion, workshop scope was limited to the materials sub-area of Soft Materials. Workshop participants included: 1) Directors of four Materials Research Science and Engineering Centers (MRSECs) with two representatives from each Center to speak about education, research, and shared facilities from: Cornell Center for Materials Research, MIT Center for Materials Science & Engineering, Northwestern Materials Research Center, and Princeton Center for Complex Materials; 2) PIs/CoPIs from NSDL Core Integration (NSDL CI) and its Materials Digital Library Pathway (MatDL) to talk about their missions and goals; and 3) Program Officers from NSF Division of Materials Research and Division of Undergraduate Education. The workshop had two main objectives:

- to introduce and provide background information to the MRSECs about the NSDL and its MatDL Pathway,
- to facilitate communication and partnerships among interested MRSECs, the NSDL, and MatDL to leverage NSF investments as well as facilitate integration of materials research and education.

NSF launched the NSDL in 2002 as a free, online library providing access to high quality resources for science, technology, engineering, and mathematics (STEM) education and research. Ten NSDL Pathway projects have been established to create community-centered, user-tailored access across all STEM domains and across all audiences. MatDL Pathway focuses on serving the materials community with a target audience that includes undergraduate and graduate students, faculty, and researchers. The current consortium of NSDL MatDL Pathway includes: Kent State University, National Institute of Standards & Technology (NIST), Massachusetts Institute of Technology, University of Michigan, Purdue University, and Iowa State University. As part of the NSDL, MatDL's mission is to:

- Implement an information infrastructure for the materials community,
- Disseminate information generated by government-funded efforts in materials.
- Provide content and services to support the integration of materials research and education

This document was authored by the Workshop organizers listed on Page 1 with feedback from and the support of the MRSEC workshop participants. The report provides an analysis of key issues and needs discussed as well as recommended beginning next steps. We view this work as outlining initial steps to shape near-term and long-term priorities that optimize the cost effectiveness of NSF investments through the complementary strengths of MRSECs, NSDL, and MatDL to augment positive impacts on materials research and education. The Workshop web page (http://matdl.org/workshop_2006.html) includes more detailed information: the workshop agenda, list of participants, copies of the presentations made during the workshop, and links to other relevant sites. We would like to thank the workshop participants who took time from their busy schedules to contribute to the productive discussions, as well as those who helped to organize the workshop, and NSF whose funding made this event possible.



B. Key Issues & Needs

Preparation for the NSDL MatDL Workshop was guided by several key workshops and reports concerning a new generation of scientific collaboration and communication. In particular, the *Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure* (<http://www.nsf.gov/od/oci/reports/toc.jsp>) laid out the vision of scientific data and tools freely available, housed in open access repositories, as a means of transforming “scientific and engineering research, practice, and allied education”. The *NSF’S Cyberinfrastructure Vision for 21st Century Discovery* called for action by NSF as the only agency within the U.S. government that funds research and education across all disciplines of science and engineering (version 7.1 <http://www.nsf.gov/od/oci/ci-v7.pdf>). In the August 2006 National Science Foundation Workshop: *From Cyberinfrastructure to Cyberdiscovery in Materials Science* (<http://www.dmrciw2006.org>), the materials science research and education community explored how and where a cyberinfrastructure can enhance the field’s research and educational needs. Following the lead of these works, the MatDL Workshop included presentations centered around two functional aspects to support transformative materials research and education through cyberinfrastructure:

1. How materials researchers work now – how they want to work

Enhance materials research through innovative approaches to support how scientists want to work.

Today materials scientists are highly collaborative, discussing and performing research with colleagues across multiple labs, departments, institutions, and countries. Despite increased interactions, effortless sharing and integration of data and related information pose considerable challenges. Yet ubiquitous information and communication technologies, like iTunes, MySpace, Flickr, Wikipedia, and del.icio.us, designed for nonscientific purposes provide novel ways of organizing data and communities. Such tools are needed for the materials community.¹ Scientists want to collect, archive, analyze, exchange and integrate their results into local and global knowledge bases with sophisticated tools based on a ‘new paradigm that will involve technologies where “the data finds the data” and “relevance finds the user”’.²

2. Expert-community guided standards for common, domain specific metadata and provenance information

Support effective materials data organization, exchange, dissemination, and reuse by processing information as it is generated in the lab

Information creation is exploding and encoding digital data with domain-specific metadata as it is generated can ensure better reuse by STEM communities to pursue research and learning goals. By enabling unobtrusive upfront capture of metadata and

¹ Glotzer, S.C. (2006) A Vision of Cybertools for Materials Research, Education and Collaboration, Invited talk at NSF Workshop on Cyberinfrastructure for Materials, August 2006.

² Jonas, Jeff (2006) Introducing Perpetual Analytics. Invited talk at Industrial and Government Applications Track, Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 20-23, 2006, Philadelphia, Pennsylvania, USA.



data provenance that is integrated into scientists' lab workflow, related research output (datasets, images, preprints) can be associated as Resource Description Framework (RDF) objects for easy discovery and access by science and engineering research and learning communities.

C. Group Discussion & Consensus

Breakout sessions were held following presentations that provided the MRSECs with background information about the NSDL and its MatDL Pathway. Each of the three breakout groups included MRSEC participants representing research, education and shared facilities as well as representatives from NSDL CI, and MatDL. Each group was asked to discuss: MRSECs' needs; complementary strengths of the MRSECs, NSDL, and MatDL; and funding needs. Breakout reports from each group were presented by the MRSEC Director participating in each group. Discussions led to a consensus that near term collaborative activities between MRSECs, MatDL, and NSDL would be easiest to initiate in the area of education. For example, while activities directly addressing the information needs of researchers were thought to be a long-term priority, integrating research data from participating MRSECs into collaborative educational efforts, as, e.g., in the MatDL Soft Matter Wiki, was considered an achievable near-term activity.

1. Education: Near term - Pilot collaborative, participatory activities

Group discussions on education were especially productive in generating many possibilities for near term collaborative activities. Consensus was achieved across groups that the NSDL and MatDL Pathway could provide MRSECs with a valuable dissemination route for their educational resources.

MRSEC participants recognized that collaborating with NSDL and MatDL could benefit their programs by more broadly disseminating MRSEC resources. They saw value in providing, through NSDL and MatDL, a general portal for educators which would offer more opportunity for serendipitous discovery. The dedicated technical information infrastructure and single or integrated access points provided by NSDL and MatDL were also viewed as positive attributes. Visiting the NSDL or MatDL sites would likely be more convenient for educators. They might also be more likely to retrieve relevant results more easily. NSDL and MatDL usage statistics could provide additional evidence that educators use and value MRSECs' resources. The user friendly tools offered by NSDL and MatDL were also viewed favorably and could benefit educators using MRSEC resources as well as MRSEC staff providing educational resources. For example, K-12 educators wishing to generate their own teaching materials would have access to NSDL tools which would facilitate the task of merging individual resources into coherent learning objects. MRSEC staff would have access to numerous NSDL tools to facilitate the publication process of a learning object, including tools for resource review and educational standards alignment. Tools offered by MatDL, such as the MatForge site for collaborative code development, could also be used to contribute to the MRSECs' commitment to education. Public domain code generated by the MRSECs could be made available in MatForge for educators to incorporate into computational teaching resources appropriate for undergraduate and graduate students.

Challenges associated with resource description, management, and preservation were considered more immediately tractable for educational resources, enabling discussion of numerous near-



term pilot activities targeting both K-12 and undergraduate audiences. Some pilots could span education and research audiences, such as a repository of polymer microstructures.

For K-12, possible pilot activities included facilitating high school and middle school teachers in creating age-appropriate teaching resources from MRSEC products by using NSDL tools and offering teacher training workshops. Subsequent dissemination of these teacher-generated resources through the NSDL would benefit the teachers by providing additional tangible evidence of their efforts to a wider audience of their peers.

For Undergraduate education, possible pilot activities included scaling up the current MatDL collection of Research Experience for Undergraduates (REU) papers to include participation from more MRSEC REU programs as well as using the Soft Matter wiki as a resource for REU students. Another suggested activity involved supplementing undergraduates understanding of instrumentation and interpretation of results through tutorials or virtual laboratories. For example, these resources could aid students in determining the best instrument(s) to use to answer a particular materials research question and to learn to distinguish artifacts from real data.

Several suggested pilot activities could span audiences: 1) assembling a searchable archive of experimental images for education, 2) making available sample databases of instrumentation data, and 3) using a wiki for community outreach. Participants were optimistic that a wiki could provide a low barrier path to achieve a number of educational goals. It could offer access to specific types of information such as news stories or course materials. It could also be used to interactively address gaps in student knowledge or promote inquiry learning.

2. Shared Facilities: Near-term – training; Long-term – organize data

Shared facilities participants expressed a need to increase awareness of available instrumentation appropriate for gathering specific measurements. They also expressed the need for aids that would help with the analysis of the data output. Not only might such activities help increase collaboration among MRSECs, they could also facilitate connections between research and learning goals if integrated into REU programs. Participants also expressed a willingness to make available sample databases of instrumentation data to support education. Establishing industry standards in data output was considered a long term goal. Given the enormous amount of data generated through shared facilities, data organization must also be addressed.

3. Research: Long term – DL to archive, organize, preserve data

Searchable digital archives of materials properties and data would greatly facilitate the work of materials researchers, but significant issues associated with resource description, management, and preservation must be resolved before this goal can be realized. Joint standards efforts among workshop participants were suggested as a way to begin to address these problems, although some participants pointed out that many standards efforts have been undertaken and the community may not be best served by waiting out these efforts before embarking on testbeds.



D. Plan of Action Summary: Path Forward & Recommended Beginning Next Steps

Workshop participants agreed that increased and sustained collaborations between the MRSECs, NSDL, and MatDL represent a significant opportunity to optimize the cost effectiveness of NSF investments by multiplying the impacts of two NSF initiatives in support of, as well as interactions between, materials research and education. Initiating focused pilot projects among the workshop participants would demonstrate to the larger body of MRSECs the range of possible ways the two NSF programs could benefit from working together. Participation in the workshop and review of this workshop report has led the participants to recommend the following beginning next steps:

- Invite MRSECs to contribute to the Soft Matter Wiki and MatForge efforts. Participants should be invited to contribute text, code projects, data, images, and teaching resources to the Soft Matter Wiki and MatForge. They would receive technical support and full branding recognition for sharing their data and documenting their contributions.
- Integrate Soft Matter Wiki into the MRSECs REU programs. Explore cooperation between NSDL, MatDL, and MRSECs by using the Soft Matter Wiki as a way to enhance existing MRSEC efforts as well as improve connections between the MRSECs, NSDL, and MatDL.
- Use wiki technology with summer programs for high school teachers
Investigate using the wiki technology in support of summer programs for high school teachers to give them a uniform way to create teaching resources for dissemination and use by other teachers.
- Participate in the next MRSEC Education Coordinators Meeting.
Establish ongoing interaction and communication between NSDL, MRSECs, and MatDL to promote further development and engender new collaborations. Explore extending participation to the MRSEC Directors meeting and the other MRSECs when appropriate.
- Establish a gallery of searchable experimental images on NSDL MatDL.
Support the development and deployment of the experimental images gallery as a means of linking dispersed resources together into a unified collection. Such a collection would contribute to MRSECs education and research efforts as well as serve as a testbed for automatic preservation metadata generation of archived digital objects' provenance. Focus initial efforts on polymer nanostructures.
- Support a joint MRSEC/NSDL/MatDL pilot virtual lab for use with REU programs.
Stimulate cooperation between MRSECs, NSDL, and MatDL, perhaps through designating funding to support a joint MRSEC/NSDL pilot virtual lab focused on Soft Matter integrating remote instrumentation for training and use with REU programs.
- Explore among workshop participants common metadata and data provenance standards in accordance with guidelines established by that community and in support of cooperative principles (e.g., maintaining open access to data and integrating learning and research goals).