SHARED METADATA AND MARKUP TOOLS FOR CHARACTERIZING MATERIALS SCIENCE DIGITAL RESOURCES

Laura M. Bartolo
Kent State University
P.O. Box 5190 Kent OH 44242-0001 USA

Cathy S. Lowe
Kent State University
P.O. Box 5190 Kent OH 44242-0001 USA

ABSTRACT
There is currently a lack of rich description attached to materials science content available on the Web as well as user-centered tools to attach such description. Dublin Core (DC) metadata and Materials Markup Language (MatML) are suggested as a means of providing materials science content information important for resource discovery and exchange. Domain specific tools are proposed for allowing authors to more easily supply this information. Such tools are expected to greatly increase the likelihood that appropriate metadata and markup will be attached to resources. Resource Description Framework (RDF) and Metadata Encoding and Transmission Standard (METS) are offered as possible mechanisms for representing multiple descriptive views.

KEYWORDS
Materials science, Dublin Core, Materials Markup Language, metadata, markup languages, authoring tools.

1. INTRODUCTION
Advances in information technology, the Internet, and the World Wide Web have provided the opportunity to create and display rich, complex scientific content. Due to the success of these improvements, expert users need better tools to describe and exchange the useful scientific content they are generating within multidisciplinary scientific teams. (Berners-Lee et al., 2001, Berners-Lee, 1998, Greenberg et al., 2003) With the development of such tools, collaboratories can play a vital role in the construction of a cyberinfrastructure with its potential to "revolutionize what scientists can do, how they do it, and who participates" (Atkins et al., 2003 p. ES 2). The challenges of sharing pertinent and high quality scientific information are directly tied to the lack of robust description of digital resources and the tools by which the necessary description can be attached at the time of resource creation. The dilemma of exchanging pertinent digital resources that affects many science disciplines is typified by the difficulties experienced in materials science, a multidisciplinary area of scientific study. For example, new approaches in materials science are enabling scientists to design materials, assembling molecules from the bottom up. Microstructures of these materials can provide visual data to a researcher helping him or her gain ideas to design structures with specific properties and functionalities. However, how does the scientist generating the microstructures describe and attach the necessary information so that other researchers can locate the image and understand how the structure was assembled? Information and computer scientists are in a strategic position to develop these missing domain specific description tools by incorporating the domain expertise and information needs of the materials science community.

Based upon recent advances in XML and web-based technologies, this paper reports on research in progress investigating effective delivery of materials science digital resources among a multidisciplinary
team of scientists in order to reduce the barriers of time and distance. Partners taking part in the project include computer scientists, information scientists, and materials scientists at Kent State University (KSU), the Materials Science and Engineering Laboratory at the National Institute of Standards and Technology (MSEL/NIST) and the University of Michigan (U-M). The paper discusses two research issues:

- What domain specific characterization do materials scientists need to describe comprehensively their digital resources?
- What common technological and conceptual tools are needed so that experts, as resource providers and information seekers, can easily contribute, find, and retrieve useful materials content?

Current practice for capturing scientific data generated within laboratory settings does not meet researchers’ requirement to describe, exchange, and integrate data in real time or for long term use. Furthermore, researchers in multidisciplinary enterprises need to be able to collaboratively gather and describe experimental analyses and results from raw data to finished scientific papers. The project described herein employs a user-centered approach for capturing and exchanging digital scientific data capitalizing on recent developments, such as digital libraries and the Open Archives Initiative (Van de Sompel and Lagoze, 2002), as non-traditional ways to assemble scientific information. These developments make it possible for information to be disseminated quickly and easily within the scientific community. While these major initiatives have bridged heterogeneous information streams through national and international standards, they have not linked directly to the end user, such as a materials science researcher generating scientific data in a laboratory as part of a multidisciplinary, multi-institutional effort.

2. MATERIALS SCIENCE DIGITAL RESOURCES

The approach employed within this paper draws upon the relationship of digital resources, use by materials scientists, and domain specific description tools. The digital resources examined within this preliminary study are drawn from MSEL/NIST and U-M focusing on the characterization and processing of materials. The scientists at MSEL/NIST and U-M use the resources as part of their collaborations with other academic and industrial partners. The scientific team members are working with computer scientists and information scientists at KSU and at MSEL/NIST to develop domain specific description tools to emphasize the important materials science components of generated data. The evaluation of the tools will examine whether the availability of the mechanisms enabled experts, as creators of materials science content, to easily and fully characterize their contributions, as materials scientists would describe them. In addition, storing the metadata records in XML based format as resources are generated, can support future generation of other materials sciences aids, such as automated knowledge organization, glossary development, and thesaurus construction.

2.1 Nanostructures: Examples of Materials Science Resources

Soft matter represents one area of materials science research conducted collaboratively at MSEL/NIST and U-M. Soft materials such as polymers, liquid crystals, colloids, DNA, proteins, and connective tissue play a central role in advanced materials and hold promise of great potential application for nanotechnology and biotechnology. The scientific team uses molecular and mesoscale simulation tools to predict morphologies of nanostructured and nano-assembled soft materials where a nanostructure or microstructure is the main output of the simulation. Each nanostructure is characterized by a unique set of information that must accompany the nanostructure for it to be useful to a materials scientist or engineer. For example, to fully describe a materials simulation requires information on the specific simulation method (e.g., molecular dynamics) and any parameters (such as for a thermostat or barostat), force fields between nanoparticles, polymers, and solvent, approximations (such as cutoffs in the range of the force field), geometric parameters of the nanoparticle and polymer (such as nanoparticle size and aspect ratio, or polymer stiffness), temperature and other thermodynamic variables, volume fraction of each species, initial condition, cooling schedule and equilibration time, be attached to the microstructure so that others can replicate that structure. Given these detailed characteristics, research results are currently difficult to capture and exchange effectively.
2.2 Metadata and Markup Languages: Related Work and Tool Design

Accessibility of digital resources in a collaboratory is closely tied to the characterization of scientific content through standardized metadata schemas, such as Dublin Core (DC) (Hillman, 2001), and domain science markup languages, such as Materials Property Data Markup Language (MatML) (Begley, 2003). DC has been specifically designed to address the problems of resource discovery and exchange on the Internet. It is a general approach comprised of 15 elements, which may be used optionally or repetitively to describe a resource, and can be used with all data formats and all domains (Hillman, 2001). Its design is intentionally simple so that someone who is not an expert in cataloging or characterizing information resources can independently create and maintain metadata records. Numerous generic DC metadata creation templates are currently available (DCMI, 2000) including DC-dot (Powell, 2000) and the Nordic DC metadata creator (Koch and Borell, 1998).

While DC provides a consistent generic characterization of content, researchers require detailed domain specific content description in order to exchange resources matching a user’s particular needs. Domain-specific markup languages, such as the MatML, can represent an integral component in a collaboratory. MatML Version 3.0 schema (Begley, 2003) is an extensible markup language (XML) for the management and exchange of materials property data on the World Wide Web that has been accepted, registered, and reposited at xml.org. Development of MatML, as a common data exchange format for materials data, arose from the lack of interoperability between Web documents and computer applications. While MatML tools have not yet become available, numerous tools have been developed for other XML-based scientific markup applications, such as Mathematics Markup Language (MathML) and Chemical Markup Language (CML). Many open access as well as commercially available tools are listed on the respective markup application websites (MathML, 2003, CML, 2003).

The preliminary study investigated the design of a domain specific description tool which incorporates both MatML and DC. The goal of the domain specification description tool is to enable a materials scientist to describe a structure in terms of the characterization and processing treatment of a material. A key issue to evaluate is whether the metadata elements will be sufficient in number and in depth to accomplish the researcher’s requirements for describing the research data. Using MatML, the tool will enable novice and expert users to view a user-friendly graphical representation of a document and will be designed to have the following features, at a minimum:

- Consistent user interface that runs on all major platforms such as Unix/Linux, Windows, and Mac
- Provision of accessing major functionalities through shortcut keys for advanced users
- Capabilities to allow users to view, update, insert, append, and delete MatML documents
- Ability to display the content of the included files hierarchically, achieved by using a tree view
- Capacity to prompt authors for domain specific DC metadata through an HTML form pre-filled with information gathered from the MatML tags

2.3 Multiple Descriptive Views

RDF (Resource Description Framework) (Lagoze, 2001, Lassila and Swick, 1999) and the Metadata Encoding and Transmission Standard (METS, 2003) are being investigated as sophisticated mechanisms for representing multiple descriptive views based upon standard metadata schemas, such as DC and standardized domain-specific markup languages, such as MatML. RDF uses XML namespaces to effectively allow RDF statements to reference a particular RDF vocabulary or schema, such as DC, and MatML. (Lagoze, 2001, Lassila and Swick, 1999) Using RDF, a collaboratory can “borrow” and use elements from each schema to describe a resource provided that the schema associated with each element is unambiguously identified. In the METS, standard four major sections compose a METS document: descriptive metadata, administrative metadata, file groups, and a structural map (METS, 2003). Using the METS approach, the file group is able to host the identification of various markup documents derived from an original document. Another research issue is to evaluate the effectiveness of METS, given the multidisciplinary nature of materials science.
3. CONCLUSION

This paper examines the research issues involved in the development of a domain specific description tool for material science digital resources based on existing metadata and markup language schemas such as DC and MatML respectively. In addition, approaches, such as RDF and METS, are particularly important in the Web environment with its emphasis on reusability of information objects. Resource providers, such as material scientists, will benefit from tools to describe and assemble their content as other materials scientists would. The creation of such tools will help to eliminate roadblocks to the development of new knowledge and commercialization in the characterization and processing of new materials. Collaboratories need better tools to link novice and expert users alike, thereby dramatically reducing the transition delay of new knowledge from the laboratory to society.

ACKNOWLEDGEMENT

This work was supported by the National Institute of Standards and Technology 70NANB3H1079.

REFERENCES

All URLs were accessed July 8, 2003.


