A Preliminary Investigation of Metadata Description Mechanisms for Materials Science

Laura M. Bartolo, Cathy S. Lowe,
College of Arts and Sciences, GREEN Digital Library, Kent State University, USA,
{lbartolo, clowe}@kent.edu

Abstract

Materials Property Data Markup Language (MatML) provides detailed materials property information necessary for the exchange of materials science digital resources among users. Dublin Core (DC) provides a consistent generic characterization of content important for resource discovery. The feasibility of reusing detailed information provided by MatML to supply DC metadata is explored. Preliminary data gathered from three examples of materials science resources indicates that MatML tags may be successfully mapped to DC elements. A prototype web-based authoring tool designed to assist authors in generating MatML as well as to map MatML information to DC elements is discussed.

Keywords: *Domain metadata, materials science, Materials Markup Language (MatML), Dublin Core (DC).*

1. Preliminary Investigation

Currently, there is a lack of rich description attached to scientific content available via the Internet, such as materials science resources as well as easy-to-use authoring tools to attach such description. Domain metadata can be integral to the Semantic Web [1] and to a cyberinfrastructure with its potential to "revolutionize what scientists can do, how they do it, and who participates" [2]. These developments necessitate new research on effective delivery of content to users. Previous work has shown that authors can generate high quality metadata, especially when equipped with tools to assist metadata production [3]. Information and computer scientists in collaboration with domain experts, can develop the missing authoring tools that incorporate domain expertise and address information needs.

This poster presents preliminary data of an investigation by computer, information, and material scientists at Kent State University and the National Institute of Standards and Technology applying two description schemas to create a valid and detailed metadata record. Three samples representing different types of materials science resources were characterized as materials scientists would describe them. Dublin Core (DC) was used to depict the three examples as individual discrete information resources. Materials Property Data Markup Language (MatML) was used to represent the information content within each resource. Selected MatML tags were mapped to DC elements to provide a common representation of the resources, as well as the domain information necessary for materials scientists. Integration of MatML into DC was handled through a prototype web-based DC/MatML generator designed for the end user.

2. Dublin Core

An all-purpose metadata schema, such as DC, provides consistent generic content description supporting discovery and retrieval [4]. DC's characteristic simplicity and aid for cross disciplinary retrieval offer an easy and effective way to convey general characteristics of resources that domain experts author. However, experts, such as materials scientists, require detailed domain metadata to describe the content of digital resources they generate as well as to find resources that match their information needs.

3. Materials Markup Language (MatML)

Standardized domain-specific markup languages represent the most important way that explicit domain semantics can be applied directly to natural language, multimedia resources, and specialized resources, such as microstructures (i.e., visual data commonly generated and used by materials scientists). *MatML Version 3.0 Schema* for the exchange and management of materials property data has been accepted, registered, and reposited at xml.org [5]. Its development grew out of the efforts of private industry, government laboratories, universities, standards organizations, and professional societies based upon the need to address the semantic and interoperability issues related to materials property data. Materials science resources are characterized by a unique set of information. This information must accompany a resource for it to be useful to a student, materials scientist, or engineer. A key issue in this investigation will be whether the combination of DC elements and MatML tags are sufficient in number and in depth to adequately describe research data.

4. Three Materials Science Examples

Three materials science examples, already tagged with MatML [5] were examined in order to identify MatML

elements that potentially could be mapped to DC elements (see **Table 1**). The first example, *Silicon Nitride (NCX-5102)* was originally obtained from an online database [6]. Free and fee-based online materials databases make accessible great volumes of materials property and processing information. The markup for this resource provides property information, such as flexular strength and Weibull modulus, which are key data for modeling and simulation. Preliminary results indicate that Name, Class, Subclass, Specification, Source, and Notes from the BulkDetails element are prime candidates for successful MatML-to-DC element conversion. In addition, under the Metadata Element , the Type attribute, as well as the Name and Notes tags from DataSourceDetails may map well.

Materials science handbooks supply property data similar to that found in databases. In addition to numerical information, handbooks provide classification systems that bring together similar substances as well as glossaries that define terminology used within the classification systems. Example 2, *Aluminum Alloy*, was taken from an ASM Handbook [7]. Examination of this item suggests that the NameAuthority attribute from the BulkDetails element, and the Name and Definition subelements of Term from the Glossary element may also be converted into DC elements.

Scientific journal literature comprises the largest, most authoritative source of information. Example 3, *Steel with TiC Coating*, originally appeared in a journal article [8]. Inspection of this item suggests that Name, ProcessingDetails, Geometry, and PropertyData from the ComponentDetails element may also be likely targets for MatML-to-DC conversion.

Table 1. Preliminary DC/MatML Mapping

DC	**MatML**
Title	<BulkDetails><Name>
Subject	<BulkDetails><Class> <BulkDetails><Subclass> <BulkDetails><Specification> [2]<Glossary> <Term><Name> [2]<Glossary> <Term><Definition> [3]<ComponentDetails><Name> [3]<ComponentDetails><ProcessingDetails> [3]<ComponentDetails><Geometry> [3]<ComponentDetails><PropertyData>
Contributor	[2]BulkDetails><Name authority = " ">, <BulkDetails><Source>
Description	<BulkDetails><Notes>
Relation	<Metadata><DataSourceDetails type = " "> <Metadata><DataSourceDetails><Name> <Metadata><DataSourceDetails><Notes>

[2] from Example 2, [3] from Example 3.

5. A Simple DC/MatML Generator

A prototype web-based authoring tool was developed to tag detailed information about materials scientists' resources through MatML, mapping selected MatML tags to DC, thereby creating a valid DC metadata record with detailed domain description. Access to authoring tools can enable domain experts to transmit preliminary critical description of their scientific output to intermediaries for subsequent resource description.

7. Acknowledgement

This work was supported by the National Institute of Standards and Technology 70NANB3H1079 and the National Science Digital Library Program.

References

[1] Greenberg, J., Sutton, S., & Campbell, D. (2003). Metadata: A fundamental component of the semantic web. *Bulletin of the American Society for Information Science and Technology, 29* (4), 16-18.

[2] Atkins, D., Droegemeier, K., Feldman, S., Garcia-Molina, H., Klein, M., Messerschmitt, D., Messina, P., Ostriker, J., & Wright, M. (2003). *Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure*. (p. ES 2). The National Science Foundation. Retrieved May 16, 2003, from http://www.cise.nsf.gov/evnt/reports/toc.htm.

[3] Greenberg, J., Pattuelli, M., Parsia, B., & Robertson, W. (2001). Author-generated Dublin Core metadata for Web resources: A baseline study in an organization. *Journal of Digital Information, 2* (2). Retrieved July 16, 2003, from http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Greenberg/.

[4] Hillman, D. (2001). *Using Dublin Core*. Dublin Core Metadata Initiative. Retrieved April 7, 2003, from http://dublincore.org/documents/usageguide/

[5] Begley, E. (2003). *MatML version 3.0 schema.* (NISTIR 6939). National Institute of Standards and Technology. Retrieved May 16, 2003, from http://www.matml.org/downloads/MatMLv30.pdf.

[6] Munro, R. & Begley, E. (2002). *Structural ceramics database.* National Institute of Standards and Technology. Retrieved May 16, 2003, from http://www.ceramics.nist.gov/srd/scd/Z00363.htm

[7] Kaufman, J.G. (1999). *Properties of aluminum alloys : tensile, creep, and fatigue data at high and low temperatures*. Materials Park, OH: ASM International.

[8] Agarwal, A. & Dahotre, N. (1999). Pulsed electrode surfacing of steel with TiC coating - microstructure and wear properties. *ASM Journal of Materials Engineering and Performance, 8* (4), 479-486.