NSDL 2004 All Projects Meeting Panel, 17 November 2004

Use of Scientific Markup Languages in the NSDL

Agenda:

- Opportunities & Roadblocks (Tim Cole)
- Case Study: MatML (Laura Bartolo)
- Recommendations from the workshop (Mike Wright)
- Discussion & Comments from the Audience

Use of Scientific Markup Languages in NSDL: Opportunities & Roadblocks

Timothy W. Cole (t-cole3@uiuc.edu)
Mathematics Librarian &
Professor of Library Administration
University of Illinois at Urbana-Champaign

Panel Presentation at NSDL Annual Meeting
17 November 2004

The Vision for Scientific Markup Languages

A standard, well-structured way to disseminate & use data & results

- Provide support for processing information at finer granularity
- Facilitate interactive use of scientific content
- Enhance search & discovery of resources
- Enable interoperability, reuse & repurposing of information through XSLT transformations, XPath selections, …

Domain-specific & generic XML-based markup languages for scientific content & raw data represent promising approaches for addressing tension between static data exchange standards and the dynamic nature of science, science research, and scientific data

Key Issues

- Need to avoid a 'Tower of Babel'
  - Plethora of domain-specific MLs introduced since 1998
- Need to avoid reinventing the wheel in each language
  - Shared semantics, structures for common elements
  - Common approaches to data modeling, sharing raw data
- Need to develop essential supporting infrastructure
  - Vocabularies, data dictionaries, ontologies
  - Transformations & cross-walks
  - W3C Compound Document Formats Initiative
- Must do more design, development, test & evaluation
  - NSDL offers unique testbed for scientific markup languages

Engaging Intermediaries

- Publishers
- Teachers & other educators
- Librarians, archivists, curators
- Other aggregators & service providers

- Demonstrate potential benefits:
  - Facilitate re-use & development of recombinant content
    - Robust transformations & extraction of content selectively
    - Easier interchange of content between applications
  - Better archiving & preservation of content (sustainability)
  - Facilitate analysis & enrichment of aggregated content

Engaging End Users

- Authors & other creators of content
- Students & other consumers

- Demonstrate potential benefits:
  - Robust, easy-to-use authoring & editing tools
  - Enhanced search & discovery (e.g., search on mathematics, material properties, chemical structures, …)
  - More interactive use of content, including content from other sources

For Example

- MathML Examples
  - http://www.w3.org/Math/testsuite/
  - http://www.dessci.com/en/products/webeq/interactive/default.htm
- CML-SVG Examples
  - http://www.adobe.com/svg/demos/cml2svg/html/index.html
- Related NSDL-affiliated workshops:
  - Scientific Markup Language Workshop
    - http://scimarkuplang.comm.nsdl.org
  - Enhancing the Searching of Mathematics (IMA Hot Topic)
    - http://www.ima.umn.edu/complex/spring/searching.html

Materials Digital Library

MatDL.org

NSDL

NIST National Institute of Standards and Technology

Using MatML with Software Applications for E-learning

Laura M. Bartolo[1], Cathy S. Lowe[1], Adam C. Powell IV[2], Jorge Vieyra[2], Kyle Stemen[1]

[1] Computer Science, College of Arts & Sciences, Kent State University,

[2] Materials Science and Engineering, MIT

NSDL Annual Meeting 2004

Panel: Use of Scientific Markup Languages in the NSDL

KENT STATE UNIVERSITY

Pilot

Materials & MatML Grapher

MIT

Teaching Resources
- Transport Phenomena
- Modeling & Simulation
- Materials Processing

MatDL.org Repository

University of Michigan

Teaching
- Computational Nanoscience of Soft Matter

Research
- Soft Matter Repository

MSEL/NIST
- Phase Diagrams
- MALDI
- CTCMS
- Combinatorial Methods
- Green's Functions

University of Colorado at Boulder
- Metis Workflow Technology

Kent State University

Authoring Tools:
- MatML grapher

Background: MatML

MatML

- XML-based language
- Exchange format of materials property data
- Initiated by MSEL/NIST (Ed Begley, Project Leader)
- Coordinated int'l effort (academe, gov't, industry)
- Tags based on ASTM materials terminology
- Currently: MatML 3.1 Schema (http://www.matml.org)

Summary of Pilot

- Benefit for MSE education
  - "Real" data
  - Web accessible application
  - Interactive learning
  - Training for next-generation researchers
- Demonstration of viability (reasonable effort)
  - Tag data
  - Parse data
  - Use in software application

Pilot: Property Data

Materials Property Data Sheet
Titanium, Ti

Subcategory: Metal; Nonferrous Metal; Pure Metallic Element; Titanium Alloy

Thermal Properties

Heat of Fusion	435.4 J/g	187 BTU/lb	
CTE, linear 20°C	8.9 µm/m-°C	4.94 µin/in-°F	over the range 20-100ºC
CTE, linear 1000°C	10.1 µm/m-°C	5.61 µin/in-°F	
Heat Capacity	0.528 J/g-°C	0.126 BTU/lb-°F	
Thermal Conductivity	17 W/m-K	118 BTU-in/hr-ft²-°F	
Melting Point	1650 - 1670 °C	3000 - 3040 °F	

Pilot: Property Data

MatML for Ti, Heat of Fusion

```xml
<Material>
    <BulkDetails>
        <Name>Titanium, Ti</Name>
        <PropertyData property="pr13">
            <Data format="float">435.4</Data>
        </PropertyData>
    </BulkDetails>
</Material>
<Metadata>
    <PropertyDetails id="pr13" type="thermal">
        <Name>Heat of Fusion</Name>
        <Units name="J/g" description="Joules per gram">
            <Unit description="Joules">
                <Name>J</Name>
            </Unit>
            <Unit power="-1">
            </Unit>
        </Unit>
    </PropertyDetails>
</Metadata>
```

Parsing MatML

Currently:

- PHP script extracts data from MatML using XPath
- parses data and generates graph for selected properties

In future:

- Cache selected properties in relational database

Software Application

Generate scatterplot comparing properties across materials

- Currently:
  - 80 possible materials
  - prescribes 4 types of property comparisons
- In future:
  - more balanced list of materials and properties

Pilot: Application

MatML Grapher

Selected Material Information

Name	Titanium, Ti
Heat Capacity * Density (J/g-Â°C * g/cc)	2.376
Thermal Conductivity (W/m-K)	17

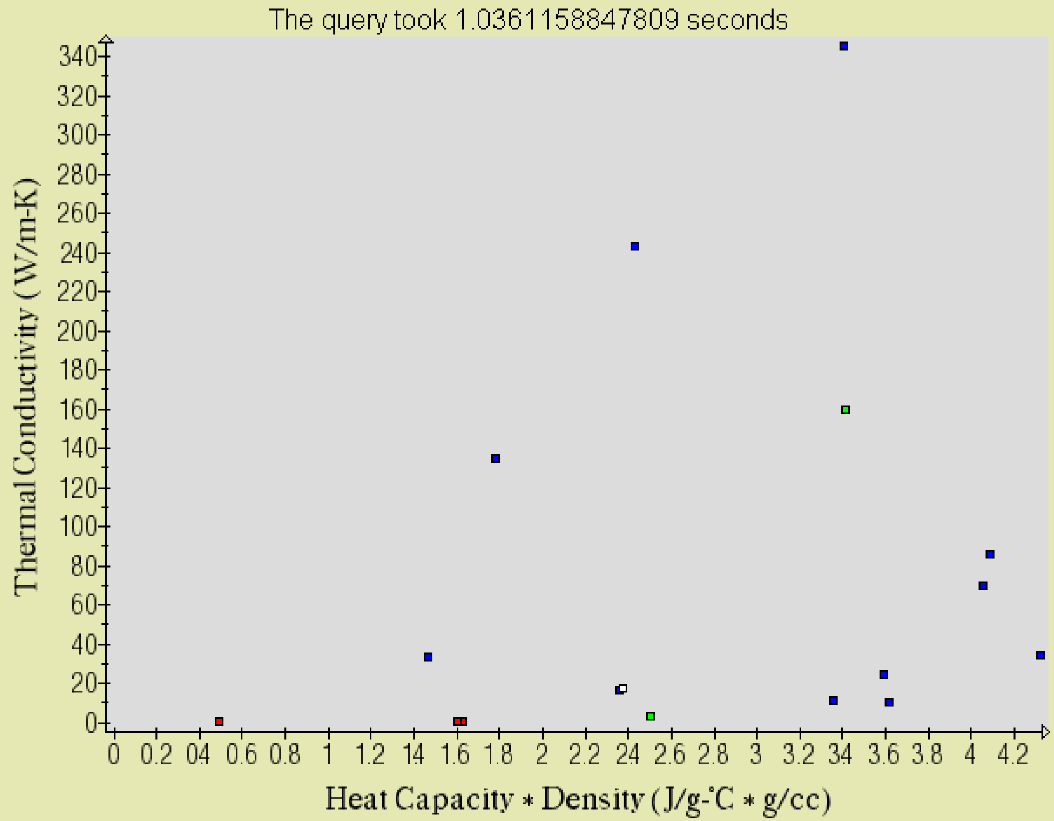
Plot type

- ○ Young's Modulus vs Density
- ○ Young's Modulus vs Yield
- ○ Young's Modulus / Density vs Yield / Density
- ◉ Thermal Conductivity vs Density * Heat Capacity

Categorization

- ○ None
- ◉ Metals, Polymers, and Ceramics

[Change Plot]

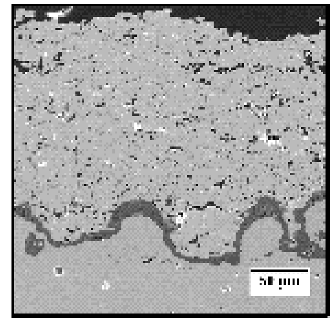


Future Work

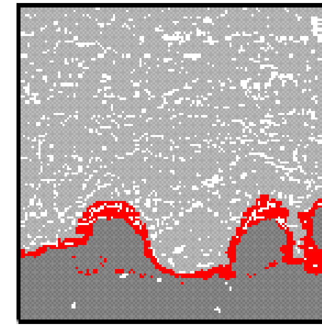
MatML with Modeling & Simulations

- Figures courtesy of Stephen Langer at MSEL/NIST (OOF)

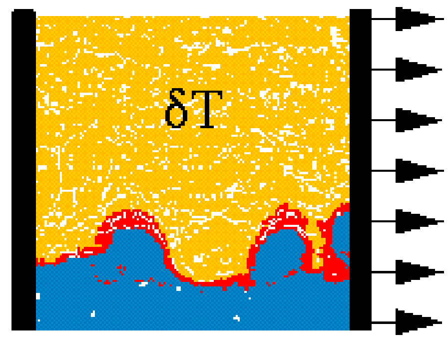
*



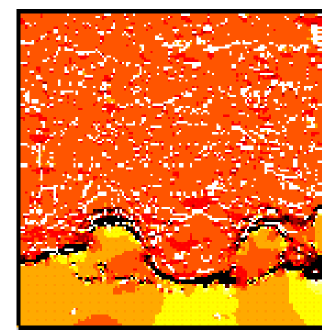
Start with a micrograph...



... select features in the image and assign material properties to them...



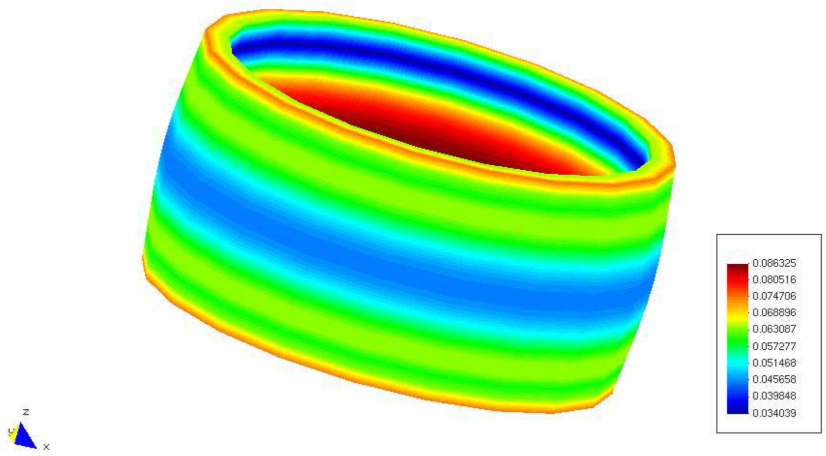
... then perform a virtual experiment...



... and measure and visualize the internal stresses.

Future Work

For example, with MatML file…



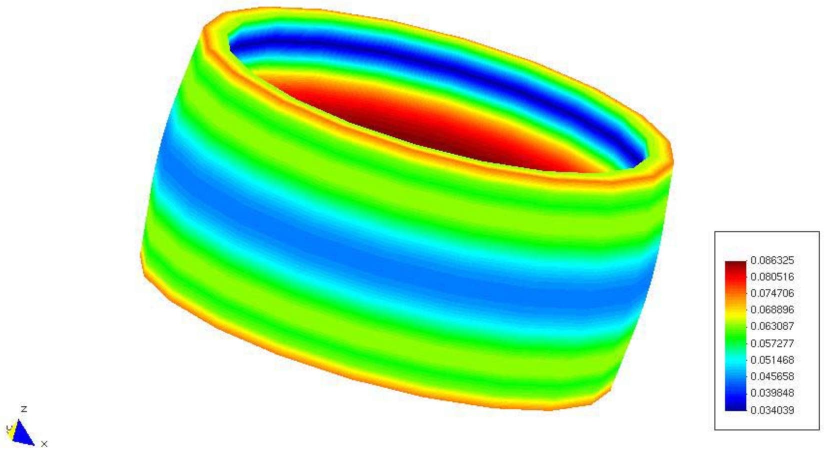
- *MathML* namespace: describe the functional relationship between yield stress & temperature, composition
- CAD software
- choose material for performance across temperature range

OR

- optimize the composition for a particular function

Future Work

Proof of Concept



*

- Intuitive user interface to specify function across range of data available
- MatML-tagged documents
- Parser to handle namespaces (MatML/MathML compound document)
- Software application to choose best material

OR

- Optimization algorithm to select across range of compositions in systems permitting it

* Figures courtesy of Tochnog (Dennis Roddeman).

Scientific Markup Languages Workshop Outcomes

Workshop held June 14-15

NSF, Arlington VA.

Organizing Committee

- Laura Bartolo, Kent State University (Chair)
- Howard Burrows, Autonomous Undersea Institute
- Stuart Chalk, University of North Florida
- Tim Cole, University of Illinois Urbana Champagne (Co-chair)
- Ben Domenico, UCAR
- Sam Dooley, Integre Technical Publishing Co., Inc.
- Sarah Giersch, Association of Research Libraries
- John Saylor, Cornell University
- Mike Wright, UCAR

Workshop Goals

- To assess and document scientific disciplines' work on markup languages

- To begin articulating a vision for the future evolution and implementation of markup languages in support of a cyberinfrastructure for research and education, with a particular focus on using markup languages in the context of the National Science Digital Library (NSDL).

Workshop Outline

- Keynotes
  - “Making the web safe for Intelligent Agents”
    - *Tim Finnin, University of Maryland*
  - “The Dynamics of Data Standards”
    - *John Rumble, Information International Associates*
- Domain sciences view points and discussion
  - Math (MathML), Chemistry (ChemML), Materials (MatML), Earth Science (ArcXML, ESML, GML, NcML)
- Cross-cutting discussions, development of themes
- Workshop draft report available

Theme 1: Demonstrating the value of markup languages

- While there was a clear consensus (albeit largely intuitive and qualitative) that markup languages can be of significant benefit in scientific research and science education, it was clear that
  - Benefits are spread across many different classes of markup language users,
  - Need to make a better case that the benefits of a markup language outweigh start-up and ongoing implementation costs to potential funders in order to stimulate broader adoption of scientific markup languages

- *Recommendation 1*: Support assessing the potential benefits of markup languages

- *Recommendation 2*: Continue support for cross-domain community interaction

Theme 2: Creating and disseminating the pre-requisite tools

- Better tools, both technically and in the form of broader, more robust ontologies, would facilitate and speed the adoption of scientific markup languages.

- *Recommendation 3*: Conduct an environmental scan of scientific markup language tools and ontologies

- *Recommendation 4*: Support applied research to produce needed tools and ontologies

Theme 3: Mediation of markup languages

- "Mediation" covers the concept of tools and services that provide a translation interface between representations in different markup languages, or that provide access to information in a single markup language to a wide variety of users.

- *Recommendation 5*: Support research on mediation services and tools between markup languages

- *Recommendation 6*: Support research on services and tools that mediate between markup languages and end users in education

Theme 4: Identifying challenges to maturation of markup languages

- There are cultural and market-related challenges to sustaining an attenuated consensus-building process around scientific markup languages.
  - tension between reflecting the dynamic nature of science (and thus the ever-changing landscape of the scientific markup language) while supporting a certain amount of standardization

- *Recommendation 7*: Support the next stage of scientific markup language standardization and implementation

- *Recommendation 8*: Fund targeted needs assessments to identify audience(s) for scientific markup languages

Theme 5: Vision

- Motivating the development of markup languages that are built on XML is the belief that by providing a means to exchange information, or data, in a structured form enables
  - colleagues across scientific domains to discover, understand and use, scientific research
  - Through common interoperability mechanisms, NSDL can support the exchange of information between the sciences and provides a framework for markup languages to be extended even further as they are tested and applied in science education settings.

- *Recommendation 9*: NSDL should play a central role in organizing cross-domain work on markup languages

Specific action items from the workshop include

- Continuing the workshop listserv to support ongoing cross-domain discussions
- Develop a registry of scientific markup languages
- Plan a follow-on workshop in 2005 (JCDL 2005)

Recommendations

1. Support assessing the potential benefits of markup languages
2. Continue support for cross-domain community interaction
3. Conduct an environmental scan of scientific markup language tools and ontologies
4. Support applied research to produce needed tools and ontologies
5. Support research on mediation services and tools between markup languages
6. Support research on services and tools that mediate between markup languages and end users in education
7. Support the next stage of scientific markup language standardization and implementation
8. Fund targeted needs assessments to identify audience(s) for scientific markup languages
9. NSDL should play a central role in organizing cross-domain work on markup languages

- Draft Report URL: **http://scimarkuplang.comm.nsdl.org**

End

A final report will be available in the near future. The draft report is avaialbe at http://scimarkuplang.comm.nsdl.org

Key Points

- Tension between static data exchange standards and the dynamic nature of science, science research and scientific data
- For MLs to move forward in a discipline, adoption and development must occur among communities of scientists, publishers and vendors, and end-users simultaneously
- Despite the potential to benefit several science and research applications, markup languages' value in those contexts remains unproven. Their broadest implementation to date occurs in processes that are virtually invisible to most users
- Better tools, both technically and in the form of broader, more robust ontologies, would facilitate and speed the adoption of scientific markup languages

Key Points

- There are cultural and market-related challenges to sustaining an attenuated consensus-building process around scientific markup languages
- Motivating the development of markup languages that are built on XML is the belief that by providing a means to exchange information, or data, in a structured form that colleagues across scientific domains can read, understand and use, scientific research and discovery can be moved forward. Through common interoperability mechanisms, NSDL supports the exchange of information between the sciences and provides a framework for markup languages to be extended even further as they are tested and applied in science education settings