Facilitating Wiki/Repository Communication with Metadata

Laura M. Bartolo[1], Cathy S. Lowe[1], Robert J. Tandy[2], and Poonam Songar[1]

[1]Kent State University, Kent OH
[2]Yale University, New Haven CT

Abstract

The National Science Digital Library (NSDL) Materials Digital Library Pathway (MatDL) has implemented an information infrastructure to disseminate government funded research results and to provide content as well as services to support the integration of research and education in materials. This paper describes how we are enabling two-way communication between a digital repository and open-source collaborative tools, such as wikis, to support users in materials research and education in the creation and re-use of compelling learning resources. A search results plug-in for MediaWiki has been developed to display relevant search results from the Fedora-based MatDL repository in the Soft Matter Wiki established and developed by MatDL and its partners. Wiki-to-repository information transfer has also been facilitated by mapping the metadata associated with resources originating in the wiki onto Dublin Core (DC) metadata elements and making the metadata and resources available in the repository.

Keywords: digital library; digital repository; Fedora; materials research and learning; metadata; wiki

1. Introduction

The National Science Digital Library (NSDL) provides a dynamic, organized point of access to science, technology, engineering, and mathematics (STEM) education and research resources as well as access to services and tools that enhance the use of this content in a variety of contexts (Zia, 2002). The NSDL Materials Digital Library Pathway (MatDL) is a consortium of organizations establishing an information infrastructure and assuming stewardship of significant content and services to support the integration of education and research in materials science (MS) (Bartolo, Lowe, Sadoway, Powell, and Glotzer, 2005).

This paper focuses on enabling communication between two of the numerous services that MatDL provides to the MS community as part of an open-source information infrastructure: the MatDL Repository (http://matdl.org) and the Soft Matter Wiki (http://matdl.org/matdlwiki). The Repository uses the Fez content management system (http://www.library.uq.edu.au/escholarship/) web front-end built on the Fedora (http://fedora.info) digital repository middleware. The Soft Matter Wiki is based on open-source MediaWiki (http://mediawiki.org) software. The Repository contains a range of MS resources, including a soft matter collection. The Wiki readily enables collaboration, communication, and dissemination (Rzepa, 2006) within the sub-domain of soft matter and facilitates expert community-driven development (Butler, 2006) of the publicly accessible site, ensuring that users have access to trustworthy, authoritative scientific information. Each service is stand-alone, but cross-service communication is needed so that users can be made aware of relevant resources in both services without visiting each separately. A plug-in has been developed which feeds information regarding relevant resources in the MatDL Repository into Soft Matter Wiki pages, alerting Wiki users to additional related content (Bartolo, Lowe, Krafft, and Tandy, 2007). The work described in this paper focuses primarily on enabling information transfer including metadata in the opposite direction, from the Wiki to the Repository.

2. Results

MatDL is beginning to integrate its services both by bringing relevant results from the MatDL Repository into the Soft Matter Wiki, and by making resources originating in the wiki available in the repository.

2.1. Search Results Plug-in

A search results plug-in was developed to bring MatDL Repository results into the Soft Matter Wiki. MediaWiki markup was extended to include a <fez> tag using the hooks provided to facilitate this task for extension programmers. This tag can be invoked with parameters indicating what to search for in the MatDL Repository and how to display the results (Bartolo et al., 2007). Search results may be included on any MediaWiki page, simply by inserting a tag like this: <fez showimage="true" showtitle="true" limit="4">Brownian</fez>. On the left, Figure 1 shows the results box returned by the command within a Soft Matter Wiki page. On the right, Figure 1 shows a more detailed view of the results box.
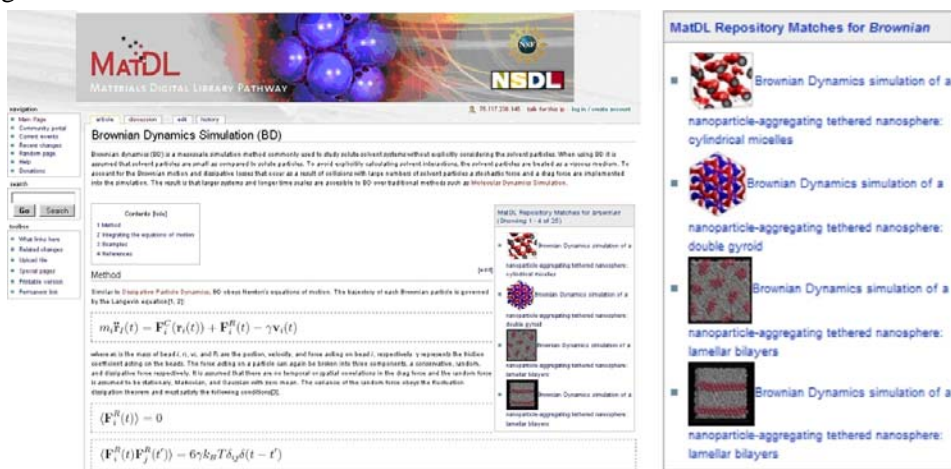

FIG. 1. Soft Matter Wiki page with results box generated by the plug-in. A more detailed view of results box is shown on the right.

Any desired search term can be substituted for the word, e.g. *Brownian*, inside the <fez> tag. The Repository search is accomplished using the search functionality provided by Fez and can be customized. If broad retrieval is desired, the search can target resources that have metadata matching the search term in any of several fields corresponding to title, keywords, or description. If a narrower focus is preferred, matches may be restricted to a single field such as keywords. The display can be easily customized by modifying the <fez> tag's arguments. The number in the limit= argument controls the maximum number of search results that are displayed in the results box. The showtitle= argument controls the display of the titles of search results. The showimage= argument controls the display of the thumbnail images associated with each of the search results. The MatDL Repository Matches link allows the user to view full search results list in the MatDL Repository. Links from titles of the individual search results lead directly to the resource in MatDL Repository.

2.2. Transmission of Wiki resources into the Repository

Recent work has enabled the transmission of Soft Matter Wiki resources, such as images, to the MatDL Repository in order to describe the component parts of the Wiki pages in more detail using DC metadata and to facilitate their discovery for possible reuse and repurposing.

The wiki2fedora process pulls metadata from a wiki installation, packages it for import into Fedora by converting to DC, then ingests the datastream into Fedora. Currently, MediaWiki is the only supported wiki engine that can give data to wiki2fedora, but in practice, any wiki engine that provides a structured way to query the database could be supported. The MediaWiki API

(http://www.mediawiki.org/wiki/API), is relatively new and is undergoing constant development, however the main features are relatively stable. An advantage of this API is its language independence; it listens for HTTP connections from clients, to which it will send back query responses in one of several formats. It sends a response formatted as XML, serialized php format, or attractively, in simplejson format, among others. Wiki2fedora is implemented in Python, so a platform- and language-neutral format is desirable. For this reason and its simplicity, simplejson is used for serializing data. As of this writing, only a subset of the information contained in MediaWiki's database is available for query through the API. This can be a limitation because key information, such as the real names corresponding to userids, is not available. In cases where this information is desired, the server side, MediaWiki API php files, can be edited to ensure that this information is published via the API interface, by adding the appropriate code to MediaWiki.

Wiki2fedora runs at scheduled intervals and is able to distinguish new or changed content from existing content. It maintains a small, local SQLite database, now standard with Python distributions, to keep track of what objects it has already seen and imported into Fedora. When a new object is encountered via its query of all relevant wikipages, it asks for the relevant metadata from MediaWiki (e.g., the title of the digital artifact, file name, size, URLs of referring wikipages, post date, userid, and real name of poster). This data is converted to DC, put into Fedora Object XML (FOXML) 1.0, and written to disk. Upon completion of the wiki scan, the data needs to be ingested into Fedora. For a process like this, it is not necessary to use the full fledged SOAP interface to ingest this data into Fedora. Instead it is possible to leverage the tools included in the Fedora distribution and have the FOXML files ingested serially with the command line ingest tool. Once the files have been ingested into Fedora, they can be reindexed into Fez, making them available for reviewing and editing before publication into openly viewable areas of the Repository.

Figure 2 shows a Wiki image and associated metadata uploaded by a Wiki user. Figure 3 shows the same image after it has gone through the wiki2fedora process and has been transferred into a review area of the Repository along with as much basic metadata as can be automatically mapped to the DC elements. In the most current stage of development: the filename is serving as a provisionary *dc:title*; comments are placed into *dc:description*; file upload date goes into *dc:date*; the *dc:rights* field is prefilled with the general rights statement that applies to all publicly viewable resources made available on the Soft Matter Wiki; wiki username is used to lookup the full name of the user for *dc:creator*; the wiki user's affiliation is used to determine *dc:publisher*. The record is manually reviewed and augmented with additional metadata, such as keywords, before it is published into publicly viewable areas of the Repository.
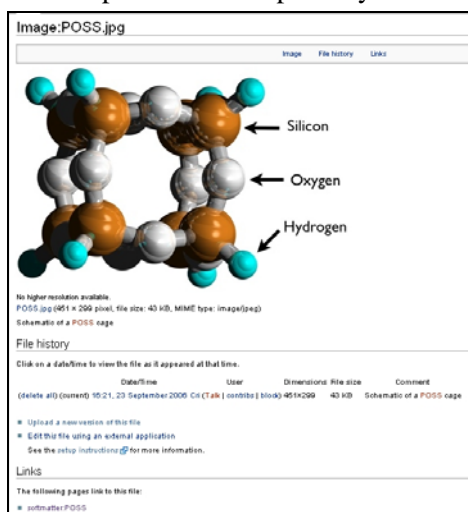

FIG. 2. File list page of an image uploaded onto the Soft Matter Wiki.

FIG. 3. MatDL Repository page showing image and metadata ingested from Soft Matter Wiki.

3. Discussion

The NSDL Materials Digital Library Pathway provides the materials community an integrated and evolving information infrastructure that supports the needs and interests of users in research and learning environments. We have described two methods in which MatDL is enabling joint communication between some of the services it provides. The first approach is a plug-in developed for bringing resources from the MatDL Repository into the collaborative context of MatDL's expert community-driven Soft Matter Wiki. The second enables information flow from the Soft Matter Wiki into the MatDL Repository, allowing the component parts of the Wiki pages to be more comprehensively described and facilitating their discovery for possible reuse and repurposing. Future directions will extend the integration of community-based tools and repositories to cyberinfrastructure platforms to bridge research, classroom teaching, and independent undergraduate learning.

Acknowledgements

References

Bartolo, L., Lowe, C., Krafft, D. and Tandy, R. (2007). NSDL MatDL: Adding context to bridge materials e-research and e-education. *Proceedings of the 11th European Conference on Research and Advanced Technology for Digital Libraries, 2007,* 499-500.

Bartolo, L., Lowe, C., Sadoway, D., Powell, A. and Glotzer, S. (2005). NSDL MatDL: Exploring digital library roles. *D-Lib Magazine. 11(3).*

Butler, D. (2006). Experts plan to reclaim the web for pop science. *Nature. 439(2),* 516-517.

Rzepa, H. (2006). Wikis and (meta)data rich environments: A model for scholarly publishing. *UKOLN Wiki Workshop 2006, Exploiting The Potential Of Wikis.* Retrieved April 10, 2008 from http://www.ukoln.ac.uk/web-focus/events/workshops/wiki-workshop-2006/talks/

Zia, L. (2002). The NSF National Science, Technology, Engineering, and Mathematics Education Digital Library (NSDL) Program. *D-Lib Magazine. 8(11).*