Evaluating the National Science Digital Library for Learning Application Readiness

Kathryn Ginger
University Corporation for Atmospheric Research
P.O. Box 3000
Boulder, CO 80307
+1-303-497-8341
ginger@ucar.edu

Letha Goger
University Corporation for Atmospheric Research
3908 Ben Hur Road
Mariposa, CA 95338
+1-209-202-3005
lgoger@ucar.edu

ABSTRACT
The usefulness of open-access digital libraries for K-16 education depends on the readiness of the collection to be placed in the path of users within diverse contexts across a variety of learning applications. The National Science Digital Library (NSDL) employs the concept of Learning Application Readiness (LAR) to assess the capacity of its collections to be deployed into applications such as learning management systems, educational resource registries, and customized curriculum services. This paper describes a multi-year evaluation of the NSDL and the notion of collection assessment as it relates to Learning Application Readiness. It then outlines steps that have been taken to increase NSDL capacity for contextualization into a wide array of K-16 educational applications.

Keywords
Digital libraries, educational metadata, cyber-infrastructure for learning, Learning Application Readiness

INTRODUCTION and BACKGROUND

Educational digital libraries support teaching and learning by providing instructional resources, user annotation, usage data, and tools to support customized use across a variety of educational contexts. In tandem, educational communities are increasingly supported by technologies and platforms that improve the quality of instructional materials available and provide for the integration of digital instructional resources into customized local contexts. Within the diverse educational community, project scope and scale, community of practice, stakeholder requirements, educational objectives and desired learning outcomes, as well as local, state and federal initiatives, define the unique nature of each learning application and the demands upon its users. As such, the utility of a digital library to support diverse applications not only depends upon its technical sophistication and size, but also upon the relevance, reliability, usability and usefulness of its materials within the local context of end users and the applications they increasingly work within (Margaryan & Littlejohn, 2008).

The National Science Digital Library (NSDL) is funded by the National Science Foundation and supports K-16 science, technology, engineering and mathematics (STEM) education. To support the contextualization of its resources across a variety of learning platforms, the NSDL has extended evaluation of the library to include the notion of Learning Application Readiness (LAR). As such, in addition to evaluating the collection for content quality and relevance, user experience and system performance, NSDL now includes an evaluation of the readiness of its collections for deployment across various applications, such as learning management systems, educational resource registries, and customized curriculum services.[1]

NSDL's growth toward LAR evolved in three phases that parallel evolution of early, large digital libraries, in general: 1) rapid early deployment via OAI-PMH protocol over a qualified Dublin Core metadata schema, 2) accumulation of a mass of resources, and 3) realignment of collection to user need in a wide range of local contexts (McArthur & Zia, 2008). NSDL phases 1 and 2 have occurred, and NSDL is currently in phase 3, further described below.

NSDL Development Phases

Within the rapid deployment mandate of *Phase 1*, NSDL built architectures and production methods for a repository of STEM metadata objects that were educationally focused or subject research oriented. The goal was to develop

[1] To further extend usefulness of data on educational resources, NSDL recently responded to a call from President Obama's Office of Science and Technology Policy to share usage data (paradata) in collaboration with STEM partners, building the STEM Exchange. http://nsdlnetwork.org/stemexchange/paradata

infrastructure to manage a large repository (millions of objects) for a large user base (millions of users) (Lagoze, Krafft, Cornell, Dushay, Eckstron, & Saylor, 2006).

Phase 2 unfolded in a resource-centric paradigm to support collection building by NSF-funded NSDL Pathway partners of audience-specific or topically-oriented portals that contribute metadata objects to the library. As such, the NSDL library is a collection of collections and the quality of educational metadata varies between contributors.[2]

In *Phase 3*, the NSDL focus is on the contextualization of library resources into emerging cyber-learning platforms to meet the needs of diverse learners with an outcome of demonstrable educational impact. In order to meet this goal, the resources in the NSDL must meet essential criteria of being usable, accessible, and interoperable (Margaryan, Milligan, & Douglas, 2007). NSDL has revised its collection policy, refocused the collection, is normalizing descriptive educational metadata, and gathering usage data (paradata) to support Learning Application Readiness.

LEARNING APPLICATION READINESS

This concept refers to how closely resources, collections, and their related metadata are aligned to educational goals, curriculum, or professional development needs of users, and how readily said resources and collections can be embedded in tools and services that educators and students use (Sumner, 2010). For this context, a learning application generally uses frameworks that characterize resources by subject, education level, resource type, audience, and educational standards, among other elements.

Adjusting for Access and Enabled Use

In *Phase 3*, NSDL has embarked on a 4-step process to support Learning Application Readiness in the library. **Step 1**: Adjust the library scope to emphasize educational materials, as opposed to subject specific academic research materials. **Step 2**: Perform a detailed collections assessment to understand the nature of the materials remaining after scope adjustment and weeding. **Step 3**: Define criteria that determine if individual collections within the library are Learning Application Ready. **Step 4**: Perform metadata normalization to provide consistency and quality to those using NSDL collections within learning applications.

Updating the scope of the library collections

In 2010, the *NSDL Collection Policy* changed from 'support teaching, learning, and research at all levels of science, technology, engineering, and mathematics (STEM)' to 'collect resources designed for teaching, learning, and conducting research relevant to STEM education.'[3] The policy established an NSDL Accessioning Board as the community collection review mechanism.[4] All collections then underwent review to determine alignment to the new policy and recommendations for de-accessioning.

De-accessioning outcomes

A before and after comparison of end user search results at NSDL.org was developed. The goal was to select searches that yielded in-scope resources (i.e. how do cells make proteins?) that could be compared before and after de-accessioning. For the period September 2008 through May 2009, the top 50 search terms and questions that began with how, what, where, when and why were used to create a reference set of 15 search phrases. For each search phrase, the first 10 resources returned were classified with a resource type and education level, along with a note about whether the resource came from an NSF-funded Pathway collection and whether the resource was accessible, broken, or led to a metadata record instead of to an actual resource.

Following this pre-evaluation, a large de-accessioning of out-of-scope collections occurred from July through November of 2009, with the following notable results:

Collection: 2.1 million unique resource URLs (not records) decreased to 115,692, a 95% change; NSDL collections decreased from 170[5] to 113; NSDL Pathways comprised 56% of the library, from 2.74% before de-accessioning.

Search returns: NSDL Pathway returns more than doubled (27% to 71%); results not accessible dropped significantly (45% to 8%); results not directly linked to a resource decreased by half.

Education level resource returns: Undergraduate rose slightly; graduate level decreased threefold; K-12 tripled; general public and informal education doubled.

Resource type returns: Learning resources, datasets, pedagogical and educational standards all doubled; animations, videos, visualizations quadrupled; articles, journal, books, abstracts, decreased 38%; university and corporate websites, and lists of links increased by 50%.

Collections Assessment

NSDL then conducted a collections assessment to answer: 1) What remains in the NSDL? 2) Which collections are providing what metadata? 3) What is the nature of growth

---

[2] In-depth discussion of the challenges of building standard metadata vocabularies across multiple collections is beyond the scope of this paper. Understanding these challenges and addressing issues of normalization of educational metadata is a critical focus of NSDL phase 3 efforts so resources can be used in a variety of contexts and for a broad spectrum of purposes.

[3] http://nsdlnetwork.org/sites/default/files/NSDL_Collection_Development_Policy.pdf

[4] http://nsdlnetwork.org/sites/default/files/NSDL_Resource_Quality_Guidelines.pdf

[5] Prior to de-accessioning, 5 'mega-collections' accounted for 70% of the library.

of NSDL as a whole, and of individual collections? 4) Which collections are Learning Application Ready?

Collections assessment results

Education metadata quality is relatively unexplored (Park, 2009) and implications around use of educational metadata are significant (Diekema, 2009). To answer guiding questions #1 and #2, eight (8) NSDL_DC metadata fields were examined: access rights, audience, educational level, educational standards, language, mime type, resource type and subject.[6] These were chosen because: 1) they have direct use for education, 2) they had NSDL controlled vocabularies on which to build a foundation collection assessment benchmark term set, and 3) if there were no NSDL vocabulary on the field, a benchmark set could be developed from metadata values present in the record.

Table 1 shows the metadata field, maximum number of unique terms discovered within NSDL records, percentage of records with any value, percentage of records that could be categorized with a benchmark term and, parenthetically, number of records with terms that could not be categorized.

FIELD	Number of Terms Used	% Records With Any Entry	% Records Categorized (# not)
Access Rights	36	9.4%	100%
Audience	157	41.3%	99.1% (562 not)
Education Level	82,951	55.7%	99.2% (2798 not)
Ed Standards	1,078	3.78%	94.5% (304 not)
Language	60	75.4%	99.9% (34 not)
Mime Type	1,345	48.7%	94.0% (4410 not)
Resource Type	565	78.3%	99.7% (397 not)
Subject	82,722	81.0%	91.9% (10,133 not)
Total	**168,912**		

Table 1: Term and percentage counts for analyzed fields[7]

Field and term use reports were generated for the NSDL as a whole, and on each collection in the library.[8] Results were that: 25% of the library had no educational metadata; education level was balanced from elementary to higher education and informal learning; audience clustered around learners and then educators; resource type was dominated by text, reference material, instructional material, and audio visuals; and, language was overwhelmingly English.

LEARNING APPLICATION READINESS CRITERIA

The 2010 collection assessment laid the foundation for determining NSDL collections that were Learning Application Ready. In 2010, two teams worked independently to develop criteria for Learning Application Readiness and to select ten (10) collections most Learning Application Ready.[9] It is significant that the independent teams matched on 5 of 6 LAR criteria and identified the same 9 out of 10 each collections as Learning Application Ready.[10] Their criteria required that resources have *pedagogical value* and *adequate educational metadata*:

- **Learning application ready resources** are *presented within 21st century contexts*, advancing critical thinking, problem solving, collaboration, and the inter-disciplinary nature of knowledge (Borgman, 2008); *relevant and reliable*, authored, meeting pedagogical needs of educators and interests of learners; and *accessible,* rights, licenses, permissions, technical requirements clearly stated.

- **Metadata for learning application ready resources** are *complete,* with title, description, URL, educational level, resource type, audience, language, rights, access rights, contributors/creators, language, mime, creation date and, if appropriate, educational standards; *accurate,* using NSDL vocabularies or values able to be mapped to NSDL vocabularies; and, *useful for direct access* to resources, with the URL directly and freely linking to an accessible learning resource and not to another metadata record.

NSDL collections broke down into three tiers: 1) *Most learning application ready,* both resource and metadata criteria being mostly met (32 collections); 2) *Possibly learning application ready*, resource or metadata criteria not met entirely, but could be used in some applications (48 collections); and 3) *Not learning application ready,* as both resource and metadata criteria are not met (41 collections).

BUILDING 'LAR' IN THE NSDL

In April 2011, NSDL used the benchmark term sets to provide collection builders with automated, web-based

[6] http://nsdl.org/collection/metadata-guide.php

[7] September 2010 analysis, encompassing 142,600 metadata records, 131,342 unique URLs, across 121 NSDL collections.

[8] As expected, educational metadata across collections was diverse and often sparse. Perspectives and experiences of cataloging and metadata professionals reveal a multitude of challenges in the metadata application process (Park & Childress, 2009) and semantic and syntactic errors, which are problematic locally, compound in a networked repository environment (Hillman, 2008). In the case of NSDL, the wide variety of metadata use could arise from local metadata requirements on collections, mapping issues from a native metadata format to the NSDL_DC format, lack of familiarity with cataloging for K16 education, confusion about values for a field, lack of cataloging manpower.

[9] Each team had educational expertise in developing resources for educators and strong library science expertise in cataloging, metadata issues, controlled vocabularies, information design, and managing large educational libraries other than NSDL.

[10] www.nsdlnetwork.org/sites/default/files/collection-assessment-public.pdf

graphs that identify distribution of resources across their collection and gaps in educational metadata. NSDL is now able to assist collection developers to identify and complete missing educational metadata in order to build LAR in the NSDL. These benchmark term sets are also used to normalize NSDL metadata. In May 2011, NSDL hosted a working group of collection developers, librarians, and software engineers to further develop and refine LAR criteria and its application across the NSDL.[11]

LAR Collaboration and Dissemination

Several NSF funded projects utilize NSDL Learning Application Ready resources and metadata objects. Notably, the interactive content in the *Curriculum Customization Service*, developed at the University of Colorado, Boulder, for the Denver Public Schools, utilizes educationally described NSDL LAR content to provide materials for teachers to tailor instruction in an infrastructure that can be leveraged to support scalable customization (Sumner, 2010). In early 2011, NSDL LAR content was embedded into two California statewide learning resource and professional development content portals, *Brokers of Expertise* and *CTE Online*.

The NSDL Stem Exchange[12] is collaborating with *Brokers of Expertise, CTE Online, Curriculum Customization Service, Instructional Architects*, and *ICPalms* to return annotative and usage data on NSDL LAR resources that are contextualized into their learning applications. These online communities of educational practitioners can integrate customizable data streams about resources from NSDL directly into their user platforms. Reciprocally, data about resource use is fed back into NSDL resource profiles to assist user discovery and utilization, and to enhance resource providers' understanding of how their materials are being contextualized and used by teachers and learners.

CONCLUSION

Phase 3 of NSDL development extends upon the major collection work of 2009 and 2010, using the educational metadata term sets as a foundation for: 1) normalizing educational metadata to support resource use in diverse contexts across multiple learning applications, 2) building educational resource description guidelines and metadata tools with user feedback loops to build capacity for cataloging best practice, and 3) developing information schemas for technologies that collect resource use data. These steps ensure stakeholders that their efforts meet the instructional needs of teachers and learners, and provide data to demonstrate educational impact. As such, building the notion of Learning Application Readiness in the NSDL has increased its capacity to provide the nation's learning communities with relevant, quality STEM resources across multiple cyber-infrastructures and in diverse contexts.

[11] https://nsdlnetwork.org/LAR/LAR-workshop-May2011

[12] http://nsdlnetwork.org/stemexchange

ACKNOWLEDGMENTS

This work is supported by the National Science Foundation, *National Science Digital Library*, grant NSF #0840744.

REFERENCES

Borgman, C. (Chair) (2008). *Fostering Learning in the Networked World: The Cyberlearning Opportunity and Challenge. A 21st Century Agenda for the National Science Foundation*. NSF Task Force Report on Cyberlearning. http://www.nsf.gov/pubs/2008/nsf08204/nsf08204.pdf.

Diekema, A. (2009). Implications and challenges of educational standards metadata. *Journal of Library Metadata, (9)*3, 239-251. DOI: 10:1080/19386380903405157.

Hillman, D. (2008). Metadata quality: From evaluation to augmentation. *Cataloging & Classification Quarterly, (46)*1, 65-80. DOI: 10.1080/01639370802183008.

Lagoze, C., Krafft, D., Cornell, T., Dushay, N., Eckstron, D., & Saylor, J. (2006). Metadata aggregation and automated digital libraries: A retrospective on the NSDL experience. 'Vannevar Bush Best Paper, *Joint Conference on Digital Libraries*, June 2006, Chapel Hill, North Carolina, USA. DOI: 10.1145/1141753.1141804.

Margaryan, A. & Littlejohn, A. (2008). Repositories and communities at cross-purposes: issues in sharing and reuse of digital learning resources. *Journal of Computer Assisted Learning, (24)*4, 333-347. DOI: 10.1111/j.1365-2729.2007.00267.x.

Margaryan, A., Milligan, C., & Douglas, P. (2007). *Structured Guidelines for Setting Up Learning Object Repositories*. CDLOR Deliverable 9. JISC, UK. http://academy.gcal.ac.uk/cdlor/documents/CD-LOR_Structured_Guidelines_v1p0.pdf.

McArthur, D., & Zia L. (2008). From NSDL 1.0 to NSDL 2.0: Towards a comprehensive cyberinfrastructure for teaching and learning. *Joint Conference on Digital Libraries,* June 2008, Pittsburgh, Pennsylvania, USA. DOI: 10.1145/1378889.1378902.

Park, J-R. (2009). Metadata quality in digital repositories: A survey of the current state of the art. *Cataloging & Classification Quarterly, (47),* 213-228. DOI: 10.1080/01639370902737240.

Park, J-R. & Childress, E. (2009). Dublin Core metadata semantics: An analysis of the perspectives of information professionals. *Journal of Information Science, (35),* 727-739. DOI: 10.1177/0165551509337871.

Sumner, T. (2010). Customizing science instruction with educational digital libraries. *10th Annual Joint Conference on Digital Libraries*. Queensland, Australia, June 2010. DOI: 10.1145/1816123.1816178