Evaluating the National Science Digital Library (NSDL) 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 large open-access digital libraries for K-16 education depends upon the readiness of the collection to be placed in the path of users within diverse contexts and across a variety of learning applications. The NSDL has employed the concept of Learning Application Readiness (LAR) to assess the capacity of its collections to be deployed within a variety of contexts 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.

Categories and Subject Descriptors

H.3.7 [**Digital Libraries**]: Collection, Dissemination, Standards, System issues, User issues

General Terms

Performance, Design, Reliability, Standardization.

Keywords

Educational digital libraries, educational metadata, collaborative digital library practice, cyber-infrastructure for learning, Learning Application Readiness

1. INTRODUCTION

Educational digital libraries support teaching, learning, and assessment by providing instructional resources, user annotation, usage data, and tools that support customized use of its resources across a wide variety of educational contexts. In tandem, information technologies are a driving force for innovation in education, and educational communities are increasingly supported by a variety of technologies and platforms for learning, teaching, and assessment. These new technologies improve the quality of instructional materials available and provide for the integration of digital instructional resources into customized local

contexts.[1] 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 these diverse applications not only depends upon its technical sophistication and size, but also upon the relevance, reliability, usability and usefulness of its instructional materials within the local context of end users and the applications they increasingly work within [1].

In order to support the contextualization of NSDL resources across a variety of learning platforms, the NSDL has extended its assessment of the library to include the notion of Learning Application Readiness. As such, in addition to evaluating the collection for content quality and relevance, user experience, and system performance, NSDL now includes an assessment of the readiness of its collections for deployment across a variety of learning applications, such as learning management systems, educational resource registries, and customized curriculum services.

NSDL's approach toward building Learning Application Readiness into its collections has evolved within three phases that have paralleled the evolution of digital libraries in general, i.e., rapid early deployment via OAI-PMH protocol over a basic qualified Dublin Core metadata schema, accumulation of a large mass of resources and, then, realignment of the collection and corresponding metadata to meet user need within a wide range of local contexts [2]. NSDL phases 1 and 2 have occurred, and NSDL is currently in phase 3, described below.

1.1 NSDL Development Phases

Within the rapid deployment mandate of *Phase 1*, the NSDL concentrated on building architectures and production methods to build a repository of Science, Technology, Engineering and Mathematics (STEM) metadata objects to describe resources that were either educationally focused or subject research oriented. The goal was to develop the infrastructure to manage a very large repository (millions of object) with a very large user base (millions of users) [3]. *Phase 2* unfolded within a resource-centric paradigm that supported a curator-focused collection building

---

[1] In order to extend the usefulness of data and descriptions on educational resources, the NSDL has responded to an OSTP call to share paradata (use data about resources) in collaboration with STEM partners by building the STEM Exchange. See: http://nsdlnetwork.org/stemexchange/paradata

strategy, resulting in the creation of NSF-funded NSDL Pathway collections of stand-alone, audience-specific and/or topically oriented portals and communities 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 collections of records.[2] In *Phase 3*, the NSDL focus is on the contextualization of NSDL resources into a variety of learning applications to meet the needs of diverse learners with an outcome of demonstrable educational impact. This strategy leverages NSDL's K-16 digital resources and supporting educational metadata to provide for successful interoperability with emerging cyber-learning platforms. In order for the library to meet this goal, the resources in the NSDL must meet essential criteria of being usable, accessible, and interoperable [4]. NSDL has revised its collection policy, refocused the library collection, is normalizing its descriptive educational metadata, and gathering paradata on resources to support Learning Application Readiness.

2. LEARNING APPLICATION READINESS

This concept refers to how closely educational resources, collections, and their related metadata are aligned to educational goals, curriculum, or professional development needs of users [5] and how readily said resources and collections can be embedded in tools and services that educators and students use. 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.

2.1 Adjusting for Access and Enabled Use

In *Phase 3*, NSDL has embarked on a 4-step process to define, develop and support Learning Application Readiness within the library. **Step 1**: adjust the library scope and hone the suite of collections to emphasize educational and classroom ready 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 collection scope adjustment and weeding. **Step 3**: define criteria that determine if individual collections within the library are learning application ready and judge each collection against these criteria. **Step 4**: perform further metadata normalization and support cataloging efforts to provide greater consistency and quality to those using NSDL collections within learning applications.

2.1.1 Updating the Scope of the Library Collections

The *NSDL Collections Development Policy* underwent several updates from its inception in 2003, but it wasn't until 2010 that the collection scope 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 2010 policy established an NSDL Accessioning Board (NAB)

as the community review mechanism for incoming collections. [4] All collections in NSDL then underwent review to determine alignment to the new policy and recommendations for de-accessioning.

2.1.1.1 De-accessioning Outcomes

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

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

Results on the Collection:
- 2.1 million individual resource URLs (not metadata records) decreased to 115,692 resource URLs, a 95% change;
- Number of NSDL collections went from 170[5] to 113; and
- NSF-funded NSDL Pathways resources comprise 56% of the library, as opposed to only 2.74% before de-accessioning.

Results on Search Returns:
- NSDL Pathway resource returns more than doubled (27% to 71%);
- Results not accessible dropped (45% to 8%); and
- Results returning only a metadata record and not directly linked to a resource decreased by half.

Effect on Education Level Resource Returns for:
- Undergraduates rose slightly;
- Graduate level decreased threefold;
- High, middle and elementary school tripled; and
- General public and informal education doubled.

Effect on Resource Type Returns for:
- Learning resources, datasets, pedagogical and educational standards all doubled;
- Animations, videos, visualizations almost quadrupled;
- Articles, journal, books, abstracts, conference proceedings decreased by 38%; and
- University, corporate pages, lists of links increased by 50%.[6]

---

[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://onramp.nsdl.org/eserv/onramp:42/NSDL_Collection_Development_Policy.pdf

[4] http://onramp.nsdl.org/eserv/onramp:42/NSDL_Resource_Quality_Guidelines.pdf

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

[6] The increase of university/corporate homepages and lists of links was a surprise, showing the granularity of resources must be improved as top level resources do not support direct learning, per se, but can contain learning resources deeper in the website.

2.1.2 Guiding Questions and Collections Assessment

After such a large de-accessioning of resources, NSDL needed additional understanding of what remained in the library. Accordingly, the NSDL conducted a collections assessment in 2010 that aimed to answer[7]:

1. What is in the NSDL?
2. Which collections are providing what metadata?
3. What is the nature of growth of NSDL as a whole, and of individual collections?[8]
4. Which collections are Learning Application Ready?

2.1.3 Collections Assessment Results

It is recognized that education metadata quality evaluation is relatively unexplored [7], and that the challenges and implications surrounding the use of educational metadata are significant [8]. In order to begin to answer guiding questions 1 and 2, eight (8) metadata fields of the NSDL_DC metadata format were examined: access rights, audience, educational level, educational standards, language, mime type, resource type and subject.[9] These fields were chosen because: 1) they have direct use for education, 2) they had existing controlled vocabularies on which to build a foundation collection assessment benchmark term set, and 3) if there was not a NSDL vocabulary on the field, a benchmark set could be readily developed by analyzing actual metadata values in the records. The table below shows the metadata field name, the maximum number of unique terms discovered within NSDL records, the percentage of records with any value, the percentage of records that could be categorized with a benchmark term and, parenthetically, the number that could not be categorized.

Table 1: Term and percentage counts for analyzed fields[10]

Field	Number of Terms Used	% Records With Any Entry	% Records Reviewed and Categorized
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% (10133 not)
Total	**168,912**		

Field and term use reports were generated for the NSDL as a whole, and on each collection in the library.[11] General results are:

- 25% of the library has no education metadata (education level, resource type, or audience), but the significance of this figure is even greater as many records only had 1 of these 3 'education-necessary' fields completed.
- Education level is fairly balanced from elementary to higher education to informal learning settings
- Audience clusters around learners, and then around educators
- Resource type is dominated by text, reference material, instructional material, and then audio visuals
- Language is overwhelmingly English

After refocusing the collection, audience and resource type return results are encouraging toward supporting NSDL's *Phase 3* goals, but improvements still need to be made in providing complete audience and educational level data.

2.2 Learning Application Readiness Criteria

Refocusing the collection in 2009 and the 2010 collection assessment laid the foundation for determining which NSDL collections are learning application ready. In 2010, two teams worked independently to develop five (5) criteria to judge Learning Application Readiness.[12] Each team then, independently, used its criteria to select ten (10) collections most learning application ready. It is significant that the teams independently matched on 4 of 5 LAR criteria and identified the same nine (9) (out of 10 each) collections as learning application ready.[13] LAR resources have pedagogical value and adequate educational metadata. Specifically,

Learning application ready resources are:
- **Presented within 21st century contexts** and advance critical thinking, problem solving, collaboration, and recognize the interdisciplinary nature of knowledge [11];
- **Relevant and reliable** for STEM education, authored, meet pedagogical needs of educators and interests of learners;
- **Accessible** with rights, licenses, permissions, and technical requirements clearly stated.

Metadata for learning application ready resources are:
- **Complete**, including title, description, URL, educational level, resource type, audience, language, rights, access rights, contributors/creators, language, mime type and, if appropriate, educational standards;

---

[7] A necessary precursor to work that transforms metadata is an evaluation of what metadata fields are present, what percentage of the total number of records have each field, how consistent is the metadata within those fields, and what patterns can be detected [6].

[8] This process has now provided a routine and automated mechanism for regularly exposing the nature of NSDL content.

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

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

[11] As expected, educational metadata across the collections was diverse, and often very sparse. The perspectives and experiences

---

of cataloging and metadata professionals reveal a multitude of challenges encountered in the metadata application process [9] and semantic and syntactic errors, which are problematic locally, compound in a networked repository environment [10]. In the case of NSDL, the wide variety of metadata use could arise from local metadata requirements on the collection, mapping issues from a native metadata format to the NSDL_DC format, lack of familiarity with cataloging for K16 education, confusion about appropriate values for the field, or lack of manpower for cataloging.

[12] 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.

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

- **Accurate,** using NSDL vocabularies or appropriate values able to be mapped to NSDL vocabularies; and
- **Useful for providing direct access** to resources, with the URL directly and freely linking to an accessible learning resource and not to another metadata record.

The metadata requirement on LARs includes a strong emphasis on educational metadata, but it was not required that resources themselves always have explicitly stated educational/learning goals. Rather, resources were selected for their ability to support learning goals, a curriculum, or educator professional development. Thus, resources such as images, simulations, or animations may be considered very appropriate for learning applications. Collections broke down into three tiers:

1. *Most learning application ready*: both resource and metadata criteria are mostly met (32 collections)
2. *Possibly learning application ready*: resource or metadata criteria not met entirely but could be used in some applications, depending on search criteria (48 collections)
3. *Not learning application ready*: resource and metadata criteria are not met and collection cannot be recommended for learning applications (41 collections)

3. BUILDING LAR in the NSDL

Several NSF funded projects are utilizing Learning Application Ready resources and related educational metadata to support their projects. Notably, the interactive content in the Curriculum Customization Service, developed at the University of Colorado, Boulder, currently supports science instruction in the Denver Public School system, and utilizes educationally described LAR content from the NSDL to provide a range of materials for teachers to successfully customize instruction in an infrastructure that can be leveraged to support scalable customization [12].

The NSDL Stem-Exchange is working in collaboration with a range of STEM education partners gathering paradata upon NSDL LAR described resources that are contextualized in a variety of learning applications in California, Colorado, and Florida. This allows stakeholders, curriculum developers, instructional specialists, application developers, and teachers to consider usage information in their evaluation of resources they might use. As such, the NSDL Stem-Exchange is contributing to the technical platform for how materials are being disseminated, used and contextualized into instruction and in multiple platforms by practitioners and, thus, further increasing the LAR capacity of the NSDL and those who build upon its resources.[14]

4. CONCLUSION

Phase 3 of NSDL development is extending upon the major collection work of 2009 and 2010, using the educational metadata term sets outlined in Section 2.1.3 as a foundation for: 1) normalizing educational metadata to support NSDL resource use across a variety of learning applications and within diverse local contexts, 2) building educational resource description guidelines and metadata input tools with enhanced user feedback loops to build capacity for best practice for describing educational resources, and 3) information schemas to support technologies that collect resource usage data across multiple learning platforms. These steps ensure stakeholders that their efforts meet the information and instructional needs of teachers and learners, and provide data to demonstrate educational impact. As such, we are very encouraged that building the notion of Learning Application Readiness into the assessment of the NSDL has set the stage to increase its capacity to provide the nation's learning communities with relevant, quality STEM resources across multiple cyber-infrastructures and within diverse contexts.

5. REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] Margaryan, A., Milligan, C., & Douglas, P. (2007). *Structured Guidelines for Setting Up Learning Object Repositories*. CDLOR Deliverable 9. JISC, UK (17 pp.). http://academy.gcal.ac.uk/cdlor/documents/CD-LOR_Structured_Guidelines_v1p0.pdf.

[5] 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.

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

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

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

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

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

[11] 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.

[12] Sumner, T. (2010). Customizing science instruction with educational digital libraries. *10th Annual Joint Conference on Digital Libraries*. Queensland, Australia, June 2010. DOI: 10.1145/1816123.1816.

---

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