

## Report of the IMLS Digital Library Forum

### I. Introduction

In the spring of 2001, IMLS supported a Digital Library Forum to discuss the implementation and management of networked digital libraries, including issues of infrastructure, metadata, thesauri and other vocabularies, and content enrichment such as curriculum materials and teacher guides. Forum members include representatives from a range of libraries and museums who have been involved in digital library initiatives. Many, but not all, have been recipients of IMLS grants.

The Forum discussions proceeded from two initiatives to encourage interagency cooperation in the development of national information portals to support education and learning. An Presidential Memorandum issued in December 1999 directed the National Science Foundation (NSF), the Smithsonian, the National Park Service, and the IMLS to work together to establish a Digital Library for Education. Comprehensive "E-Government Act of 2001" bills introduced in both the House of Representatives (H 2458) and Senate (S 803) contain parallel provisions that direct NSF, the Library of Congress, the IMLS, the Park Service, and the Smithsonian to cooperate in the establishment of an Online National Library. The functions of the Online National Library include:

- providing access to an expanding database of educational resource materials;
- providing functional integration, so that a user may access the resources without regard to boundaries of the contributing institution; and,
- including educational resource materials across a broad spectrum of scientific, historical and cultural materials.

Because of the IMLS's focus on creating and preserving digital resources, it is important that IMLS and the museum/library community explore opportunities for bringing these rich collections into the Online National Library. One of the most advanced existing initiatives in this area is the NSF's National Digital Library for Science, Mathematics, Engineering, and Technology Education, otherwise known as the National Science Digital Library program (referred to here as the NSDL). Therefore, one of the initial tasks of the Forum was to meet with representatives from the NSDL to explore opportunities for integrating the NSDL initiative and IMLS-funded digital content, making this content available for use in educational settings. The initial questions addressed by the Digital Library Forum were: 1) What must IMLS-funded projects do to be compatible with the NSDL? and, 2) How can institutions and communities with an interest in building digital libraries be brought together in fruitful collaboration?

The IMLS and NSF "teams" met jointly twice and worked together by e-mail, conference call, and in subgroup meetings. A key to all discussions was the white paper "Pathways to Progress", which lays out the strategies, standards and structure for the NSDL.

This document reflects the contributions of the participants.

---

### II. The NSDL

The NSDL is an NSF program to build a national digital library for education in science, mathematics, engineering, and technology. According to Program Director Lee Zia, "the resulting digital library is intended to serve the needs of learners belonging to a broad user audience — K to 12, undergraduate, graduate, and life-long learning — in both formal and informal settings. Envisioned as the premier portal to current and future high-quality educational content and services, this virtual facility will enable seamless access to a rich array of interactive learning materials and resources, distinguished by the depth and breadth of the subject matter addressed, and valued for its authority and reliability."

The current NSDL participants are the recipients of roughly 76 NSF grants awarded under the NSDL

program or its predecessor. Together, representatives from these projects have been addressing issues of governance, pedagogy, collections, standards, services, and technical infrastructure. Dozens of projects are working to build specific portions of the digital library, such as collections of educational materials. At the same time, Cornell University and a suite of collaborators have been funded to develop "core integration services" for the NSDL, including a central portal, indexing and search capabilities, software and services to support collection building, and an intellectual property rights management system. The proposed technical architecture posits a single central metadata repository, and will be built around twin themes of a "spectrum of interoperability" and "one library with many portals".

Organizationally, the NSDL is a current and future group of projects who gain participant status by being awarded funds under the NSF program. However, NSDL can also be thought of functionally, as the set of collections, content and services that can be found through the NSDL portals, which is potentially more inclusive. The spectrum of interoperability is based on multiple levels of participation requiring different levels of commitment from collection owners. At the lowest level, collection pages may be gathered by spiders and descriptive metadata extracted automatically, requiring almost no effort from the collection site. At a middle level, metadata will be harvested according to the Open Archives Metadata Harvesting Protocol, requiring the source site to support this protocol and expose metadata in a Dublin Core-based format. A third level allows some collections to be even more tightly coupled with each other through federated search services using standards like the Z39.50 information retrieval protocol, in addition to contributing metadata to the central repository.

The levels of participation discussed above are based on the creation and contribution of metadata. However, it is also possible to think of levels of participation in terms of content creation and contribution. At the lowest level is the creation of digital collections of content relevant to the NSDL even if not primarily intended for it or for educational purposes. Examples range from the archival papers of a physicist to images of plant specimens to the documents of a state water quality agency. At the middle level, the raw content can be enhanced by the addition of educational context, such as links to standard curricula or incorporation into learning modules. At the highest level is the building of coherent collections of learning materials along with tools and services to support their use.

It is the opinion of the Digital Library Forum that it is appropriate for institutions wanting to participate at the highest levels in terms of either metadata or content contribution to seek funding from the NSF's NSDL program. However, the IMLS can and should fund alternative levels of participation by libraries and museums. Appropriate activities for funding include:

- creating new digital content relevant to the NSDL;
- making new and existing content available to the NSDL through the contribution of metadata within the identified subject areas of science, mathematics, engineering, and technology; and,
- enhancing existing content with educational context such as curriculum guides or lessons.

### **Recommendations for Projects**

Projects proposing to create topically relevant digital content should be encouraged to take steps to ensure that the content can be included in the NSDL. Specifically, these projects should:

1. Describe the potential uses of the collection for education in science, math, engineering or technology, and indicate how these uses are served by project decisions, such as the metadata elements used for description;
2. Create both collection-level and item-level metadata;
3. Map item-level metadata elements used by the project to the normalized NSDL metadata set (Dublin Core with DC-Ed extensions and the IEEE extensions);
4. Support the Open Archives Metadata Harvesting protocol to contribute metadata for harvesting;

5. Guarantee the accessibility of materials to NSDL users and the availability of materials for use for a stated period of time; and,

6. Follow procedures in the creation and documentation of objects and metadata that support interoperability, reusability, and persistence (see General Recommendations, below).

Funding should be provided to IMLS-funded digitization projects to take steps to map metadata elements to the NSDL set, upgrade systems to support the Open Archives Metadata Harvesting Protocol, and upgrade objects and metadata to meet appropriate benchmarks.

Note: IMLS will provide funding in 2002 for a limited number of projects to add value to already-digitized collections as a demonstration of interoperability with the NSDL. Eligibility is limited to eligible library applicants (and partners, if appropriate) with digital collections created with prior IMLS funding. See description under the Library Research and Demonstration category in the 2002 National Leadership Grant guidelines, which are available on the IMLS Web site at <http://www.ims.gov>. Applicants should follow all application procedures for National Leadership Grants for Libraries as well as the recommendations for projects in this document.

---

### III. General Recommendations

The primary intellectual criterion for inclusion in the NSDL is the relevance of the subject content of the material to education in science, math, engineering or technology. However, technical criteria for NSDL inclusion are not closely tied to either science or education. The Digital Library Forum considers these recommendations to be relevant to IMLS-funded projects generally.

1. Digital collections built with support of public funds can and should be held to standards that support interoperability, reusability, and persistence. All funded projects should meet reasonable benchmarks for the quality of the digital objects created or gathered, the metadata that describes them, and the collection interface through which they are accessed. This recommendation deliberately does not suggest the establishment of minimal standards of quality in any area, because there is great variability in what is appropriate or feasible under different circumstances. However, we do recommend that the IMLS help create, disseminate and maintain over time a framework of guidance to aid projects in understanding and establishing appropriate quality benchmarks. We submit with this report a draft document, A Framework of Guidance for Building Good Digital Collections that we believe can serve as the basis for such a framework.

2. IMLS should maintain its own registry of funded digital collections. Although various global registries exist (and projects should be encouraged to create collection-level entries in these as appropriate) it is also important to be able to identify the set of IMLS collections. This can aid grant applicants looking for models and practical examples of acceptable practice, can help further the sense of community among past and present awardees, and can provide a mechanism for identifying collections with various features (for example, those existing collections which might be appropriate for future inclusion in the NSDL).

Note: IMLS has established a new funding priority to develop a metadata agent in accordance with the Open Archives Initiative protocol. See the description under the Library Research and Demonstration category in the 2002 National Leadership Grant guidelines at <http://www.ims.gov>. Applicants should follow the application requirements in the guidelines except as noted in the supplementary document, "Request for Proposals to Develop a Registry and Metadata Repository for Digital Collections," available upon request from IMLS. Contact Martha Crawley at [mcrawley@ims.gov](mailto:mcrawley@ims.gov).

3. Because so much of the IMLS constituency consists of small and medium-sized organizations without sophisticated in-house technical support, the IMLS should also consider projects to develop infrastructure services that lower barriers to NSDL contribution by smaller organizations. Such initiatives could be undertaken at the state or regional level. These include:

establishing digitization service centers;  
establishing archival repository services;  
establishing mirror website services; and,  
establishing name resolution services.

There is no recommendation that IMLS provide ongoing operational support for such services, but grants might help fund one-time costs for such things as pilot studies, market research, the development of business plans, or initial capital investments.

4. The issue of persistence is perhaps the most difficult of the problems facing digital collections at this time. Although rapid progress is being made, as a community we have not yet achieved an understanding of the requirements, methodology and expense involved to ensure that digital objects are physically preserved and remain accessible and usable over time. For many projects this provides an excuse to delay decisions and investments related to archiving. IMLS should encourage the integration of an archiving component into every project plan by requiring a description of how data will be preserved. We also suggest that every proposal to build a digital collection include some funding earmarked either for in-house archiving or for third-party archiving services.

---

#### **IV. Opportunities for Collaboration**

The NSF and the IMLS are complementary in many respects. The NSF has close ties with the educational community and long experience applying content in academic programs. NSF computer scientists have the expertise needed to develop scalable infrastructure services and address many of the problems of working digital libraries. The culture of the NSF community supports communication, research, innovation, experimentation, and the rapid prototyping of applications, tools and services.

IMLS, on the other hand, has close ties to museums, libraries, and state governments. The IMLS constituency maintains a massive quantity of quality content with vast scope in terms of subjects, place and time. The staff of cultural heritage institutions bring curatorial interest and expertise in the selection, description and preservation of materials. The community honors standards, collaboration and public service, and is markedly user-oriented.

The digital collections built by libraries and museums can provide raw materials for educationally oriented products. Images of herbarium specimens might be incorporated into a classroom module on taxonomy, for example, or on geography, or biodiversity. Artworks have been used in the teaching of math. A module of the NASA Space Educator's Handbook uses rockets and robots from science fiction novels and comic books to teach space technology. To exploit these connections, educators need to be able to find appropriate materials, and, when found, evaluate their quality and verify their authenticity. They need to be able to identify the rightsholder to obtain permission for use. They need assurance that the digital object won't go away or change names. In short, the ability to repurpose materials for educational use requires that collections be built with attention to search interoperability, reusability and persistence.

In turn, the NSDL could develop services that would encourage libraries and museums to invest in building reusable collections. Maintaining a good, easily searchable, widely accessed registry of collections, for example, could increase web traffic to collection sites by making them more widely known, which in turn could be used to promote the collection or justify funding. The fact that a collection's resources have been incorporated into external learning modules also helps justify the collection. The development of a reporting service for linked-to items would allow organizations to quantify such uses their governance or funding sources.

Projects that combine the strengths of each community should be considered for funding or for co-funding with NSF. Examples of useful collaborations include:

projects to develop curriculum materials for the NSDL based on library and museum

collections;

projects to determine how certain categories of users (undergraduates, elementary school teachers) seek information; and,

projects to build tools to help automate the creation of quality metadata incorporating controlled vocabularies and classification schemes.

We close by thanking everyone involved in this initiative. The IMLS and NSF "teams" participating in the meetings leading up to this report found a mutual respect for each other's knowledge and views, enthusiasm for the assignment, and a shared belief in the promise of digital libraries. We all appreciate having had the opportunity to think about these issues together.

---

#### **IMLS Participants**

Liz Bishoff, Colorado Digitization Alliance

Priscilla Caplan, Florida Center for Library Automation (chair)

Tim Cole, University of Illinois Urbana-Champaign

Anne Craig, Illinois State Library

Daniel Greenstein, Digital Library Federation

Doug Holland, Missouri Botanical Garden

Ellen Kabat-Lensch, Eastern Iowa Community College

Tom Moritz, American Museum of Natural History

#### **NSF Participants**

Su-Shing Chen, University of Florida

Gerry Hanley, California State University

Karon Kelly, Digital Library for Earth System Education

Kathy Liu, National Health Museum

Cathy Manduca, Carleton College

Brandon Muramatsu, SMETE.org and University of California Berkeley

Kate Pittsley, MERIT Network

John Saylor, Cornell University

VivianLee Ward, National Health Museum