NSDL Reflections Project

Comments on this essay? Please visit
nsdlreflections.wordpress.com

Implementation and Innovation in the NSDL

William Y. Arms
Cornell University
Revised: August 28, 2008

Alternative views of the NSDL

This essay is a personal reflection on how early decisions shaped the National Science Digital Library (NSDL) and how the program has evolved over the past decade. It draws heavily on observations from the planning studies in 1997 and 1998, and my experience as principal investigator of the Cornell University's part of the Core Integration team until 2005. The thoughts expressed here are purely my own.

The underlying theme of this essay is that the NSDL program has two missions: implementation and innovation. Confusion between these two missions goes back to the beginning of the program. The original concept was to implement a digital library for science education. But the NSF's principal goal is to support research and the NSDL program also gives grants for innovation in digital libraries and science education.

The first public discussion of a library for science, mathematics, engineering, and technology education (SMETE, as it was originally called) was a National Research Council workshop in August 1997 [1]. This workshop was only partially successful for a reason that has challenged the NSDL throughout its life. Some members of the workshop were genuinely wanted to build a library for scientific education, but the majority were researchers. They welcomed the NSF's interest in this area because it might provide funding for their personal research. The participants gave insufficient attention to the tough implementation questions. Would a digital library really improve the quality of scientific education? Is this a good way to spend the taxpayers' money?

The NSF tried again with a workshop in July 1998, which I chaired [2]. The report of the workshop concentrated on the narrow objective of how to build a digital library that would have an impact on science education. While urging that there should be an associated research program, the report emphasized that, "The SMETE Library provides a service; it is not a research project." It envisioned a central organization that would coordinate a federation of major partners. The importance of these partners was expressed in a sentence that proved prophetic, "The site to which the NSF lends its name and towards which it directs its primary marketing will be considered the central site, but it is unlikely to be the biggest or most heavily used."

The first solicitation

The NSF provided seed grants through the Digital Libraries Initiative in 1998 and 1999. The main program was announced in 2000, under the name of the National Science Digital Library (NSDL). Since then, the annual solicitations provide a record of how the NSF's view of the NSDL has evolved. Small alterations have been made every year, but the solicitations in 2000, 2004, and 2008 represented major changes of emphasis. The solicitation in 2000 was particularly important. It differed from the recommendations of the planning reports in several significant ways that have shaped the program ever since.

The most important change received little attention at the time. While the planning meetings had been asked to consider undergraduate education, the solicitation called for a library to support scientific education at all levels. This was a major change, since K-12 education is fundamentally different from undergraduate education in emphasizing curricula through highly structured class plans and state standards, and making limited use of primary materials. Failure to grasp this distinction was a serious problem in the early days of the NSDL and remains a challenge today, particularly as most of the funding has gone to university groups with expertise in undergraduate education but lesser understanding of K-12.

A second major difference was to mix the two missions of the NSDL in a single solicitation. The stated aim was to implement a digital library, but the funding program was more suitable for a broad program of research and development. In the early years of the NSDL, only twenty to thirty percent of the funding went to support the library building view of the program. The bulk of the money went to a range of smaller projects with labels such as, "services", "collections", or "research". Most of these early grants were to groups whose primary interest was in innovation. While the services and collection projects were expected to contribute to the goal of building a digital library for science education, the structure of the program provided few incentives for them to integrate their work into the broader NSDL.

The 2000 solicitation postponed the selection of the team that would build the central parts of the library. It funded six short-term projects, intended to create separate demonstration systems to show how the library might be built. The plan was that they would be followed by a single Core Integration project, with long-term funding, to combine the various activities of the NSDL into a coherent library. In the event, only the projects at Cornell University and the University of California at Berkeley actually built demonstration systems.

The Cornell demonstration project

Since the Cornell demonstration project, Site for Science [3], became the technical basis for the central NSDL system, it is interesting to compare what was envisioned in 2000 with what has developed since. Reading the Cornell proposal today, with the benefit of hindsight, I am impressed with how strong it was technically and how little emphasis it gave to other important areas.

Today's central system has many technical similarities with the Site for Science prototype, particularly in the central role of a metadata repository. The details have changed, e.g., a relational database has been replaced by a Fedora repository, but the proposal was explicit about the role of Dublin Core and the importance of compound digital objects. It had an interesting

section that anticipates personal information services of the type now known as Web 2.0. The prototype was an early adopter of the OAI-PMH.

Elsewhere, the demonstration project was less strong. There was no recognition of the differences between K-12 and higher education. The focus was on building a virtual library that will be a resource for everybody interested in science. In the tradition of higher education, the goal was to make primary materials available, with no distinction between the library provision for students, teachers, and researchers. There was no understanding that this model of a library needed to be modified for K-12 education.

The organizational issues that frustrated the early years of the NSDL were mainly ignored in the proposal. There was little consideration of how to combine the various small projects into a coherent whole, but it was assumed that there would be a federation with the NSF providing firm direction via the Core Integration team. There was an optimistic assumption that many of the major partners would be external projects that were not funded by the NSF.

The 2000 solicitation had a few sentences on how the NSDL might continue when the original NSF funding expired and the Cornell proposal had a short section on this topic, but only in the context of the central system.

Core Integration

During 2000-2001, with encouragement from the NSF, the six demonstration projects attempted to develop a joint proposal to create the Core Integration effort, but despite valiant efforts the six teams were unable to overcome technical and philosophical differences on how the NSDL should be developed and how the funding should be divided among the groups. Eventually, UCAR, Cornell, and Columbia broke away from the other three and submitted a proposal that was accepted by the NSF. This proposal included substantial subcontracts to three other universities.

Seven years later, the proposal still reads well, but some central themes have not passed the test of time. By avoiding making hard decisions, the group was able to present a united front and succeed in being funded (which is the purpose of any proposal), but it created problems that lingered for many years.

The first problem was that the Core Integration team promised to do too much. The team committed itself to extensive efforts in library building, technology, outreach, and evaluation. In the opening section of the proposal, three models for the NSDL were discussed with varying degrees of centralization; the proposal offered a framework that embraced all of them. The proposal used two appealing phrases, "one library, many portals" and "a spectrum of interoperability", but made little attempt to measure the effort needed. At least eight metadata standards were to be supported and there would be a broad set of authentication and authorization services. The proposal envisaged extensive work with commercial publishers.

With encouragement from the NSF, the proposal was over-reliant on consensus building among the NSDL-funded projects. "We envision a strong NSDL community that sees itself as owning

the program and having major influence on the character of the library." In retrospect it was naive to expect effective collaboration among hundreds of projects many of them having little in common except being funded by the NSF. External events aggravated the problems of collaboration. Because of the trauma of September 11, 2001 the NSF postponed an initial meeting of all the grantees, including the services, collections, and research projects. When they were brought together several months later, the fragmentation into a range of independent projects was well under way.

Even within the Core Integration team, little though was given to effective management. In practice, the universities who were subcontractors to the Core Integration team used the funding more to support their general research programs than to provide specific services to NSDL. Organizationally, the history of the Core Integration team since 2001 is a progressive tightening of the management structure. The loose oversight in the early days allowed conflicting visions of the goals, even within the Cornell team. For several years, I combined part-time leadership of the Cornell team with the usual academic duties of teaching and other research. When Dean Krafft replaced me as a full-time principal investigator in 2005, he immediately reorganized the work at Cornell to give it more focus.

Finally, the proposal was very weak on how to support the library in the long term. Basically it stated that sustainability is important and the team would think about it. Nobody had recognized the urgency of sustaining the smaller projects, which would be a serious concern within very few years.

The Pathways

The NSF solicitations from 2000 to 2004 gave almost eighty percent of the funds to small projects, chosen on their individual merit rather than their combined value as components of the NSDL library. As with any NSF program, many of these projects have been excellent and others have been disappointments. Some of the collections and services continued after the initial funding expired, but others failed to survive. Meanwhile progress was slow on the task of building a comprehensive digital library for education.

During 2003, several people suggested that the NSDL needed to put more of its resources into implementation. The strategy that emerged to support the equivalent of branch libraries, each with a focus in a specific area of scientific education. For example, one of the first ideas to be considered was the possibility of an undergraduate mathematics education, which is a discipline with a tradition of cooperation across universities.

These discussions led to the Pathways program, which was introduced in the 2004 solicitation. In the NSDL, Pathways are digital libraries that concentrate on specific areas of science education. Currently there are eleven of them. Compared with the smaller projects: they are funded at a higher level and for longer periods of time; they are expected to provide operational services; and they are charged with cooperation among themselves and with the Core Integration. Several Pathways concentrate on a specific level of education, such as K-12, middle school, or undergraduate.

I have been less involved with the NSDL in recent years, but the Pathways program appears to be a success. Usage statistics suggest that several of them are having considerable impact on education.

A curious feature of the Pathways is how loosely they are tied to the central NSDL web site operated by the Core Integration team [4]. The central site has a search service that indexes the content of all the NSDL projects, but currently it gives surprisingly little emphasis to the Pathways, beyond a web page that lists their separate web sites. Each Pathway acknowledges the NSDL on its home page, but the acknowledgements usually appear in lists of contributors, rather than as proclamations that they are branches of a single national digital library. One reason is that most of the Pathways receive support from several sources; their users value them as resources in their fields of interest, not because they are components of the larger NSDL.

The changes in 2008

In 2007, NSF's directorate of Education and Human Resources had a new assistant director. This led to a review of the NSDL program and the solicitation in 2008 had some important changes.

The goal of the NSDL program was subtly redefined. With minor changes, the annual solicitation from 2000 to 2007 had stated that the aim of the NSDL was "to found a national digital library that will constitute an online network of learning environments and resources for science, mathematics, engineering, and technology (SMET) education at all levels." In 2008, there was no longer any mention of founding a national digital library. The aim was simply, "to establish a national network of learning environments and resources..."

The solicitation continued to emphasize the importance of the Pathways track, with funding for new projects and continuation grants for the earlier ones, and emphasized the responsibility of each Pathways project to provide "a stewardship role on behalf of NSDL for the educational content and/or the services needed by a broad community of learners." A portion of the Pathways' budgets will support technical integration, but their explicit commitments to the integrated goals of the NSDL are minimal, simply a requirement to provide metadata to the central repository. In essence, the NSF appears to view the Pathways as a federation of independent digital libraries, each responsible for an area of scientific education, rather than tightly integrated branches of a single national library.

By encouraging earlier Pathways projects to apply for continuation grants, the NSF is recognizing the realities of sustainability. Despite numerous plans, meetings, and workshops, nobody has been able to propose a realistic financial model for educational digital libraries other than support from the governments or other external sources. The NSF is always reluctant to make commitments that imply it will support any activity indefinitely. This reluctance is part of its overall success, but sometimes it goes too far. The cost of the Pathways is small when compared with the amounts that are spent on educational resources by the federal and state governments, and by universities and other organizations. Educational technology has a history of excellent projects that are lost when initial funding expires. If the Pathways can collect and provide access to these resources over the long-term, the program will be an excellent investment.

Since the original Core Integration grant expired, the NSF has wrestled with the question of what comes next. With long-term projects, the NSF's culture is to give other organizations opportunities to submit proposals. There is no reason why UCAR, Cornell, and Columbia should run the Core Integration indefinitely. One option that the NSF explored was a "management entity" that would be responsible for the contractual aspects of the major NSDL grants and a focus for the sustainability efforts. There were several one-year extensions of the original Core Integration grant and during 2008 the NSF is winding up the original Core Integration team. In its place, the 2008 solicitation asked for separate bids for a Resource Center and a Technical Resource Network Services project. These two activities will continue much of the work that has been carried out at UCAR and Cornell respectively, but with a greater emphasis on the operational aspects of the services.

Observations

The NSF must be congratulated on persevering with the NSDL. It would have been easy to abandon the program after the first round of grants. If so, it would now be looked on as an overly ambitious venture that supported a number of isolated projects – some good, others less good – which in aggregate had little impact on education and did not survive after the initial funding expired.

While it is still too early to declare success, there are some hopeful signs. Here are some lessons that can be learned from the first eight years of the NSDL program.

First, the NSDL is a fascinating contrast with the usual way that implementation projects are managed in the federal government, which is to issue a request for proposal and choose a contractor to build the system. The NSF has followed its usual process of issuing open-ended solicitations, selecting projects on merit, and giving them remarkable flexibility. Sometimes this leads to waste, but it also leads to creativity and energy far beyond the ordinary.

Second, the decision to expand the NSDL from undergraduate education to all of science created a colossus without clear direction. The Pathways program rescued the program, by dividing the NSDL into a group of motivated projects each with its own emphasis and enthusiasm.

Third, sustainability has proved to be a chimera. The only effective way to support digital libraries for science education is continuing funding from external sources. This is no different from other libraries and from most scientific research.

Fourth, everybody over-estimated the willingness and ability of independent investigators to form a collective governance structure, even while directly funded by the NSF. With research grants, the NSF provides little direction beyond the words in the solicitation. An implementation program needs a more formal management structure.

The NSF is committed to fund the NSDL until 2012. It is providing substantial central funding for the Technical Resource Network Services and the Resource Center, but the bulk of its resources are for the Pathways. In 2007-08, the letters "NSDL" were redefined. The new name

of the program is the "National Science Distributed Learning". Hopefully, it will provide a continuing framework in which the two missions of the NSDL can both thrive: implementation combined with innovation.

References

[1] National Research Council, *Developing a national digital library for undergraduate science, mathematics, engineering, and technology*, August 1997. http://www.nap.edu/readingroom/books/dlibrary/appa.html

[2] *Report of the SMETE Library Workshop*. National Science Foundation, NSF 99-112, July 1998. http://www.dlib.org/smete/public/report.html

[3] William Y. Arms, et al. A spectrum of interoperability: the Site for Science prototype for the NSDL. *D-Lib Magazine*, 8(1), January 2002. http://www.dlib.org/dlib/january02/arms/01arms.html

[4] The NSDL web site is http://nsdl.org/.