Section C. Project Description

Table of Contents

C.1 The NSDL Vision and the Core Integration Strategy

C.1.1 The NSDL Vision

Though still in its infancy, the National Science Digital Library (NSDL) shows promise of becoming one of the major technological Federal investments in American education. While the funds invested in the NSDL are modest, relative to the total outlay for science education, the NSDL can provide a focal point for many public, commercial, and foundation efforts, thus returning significant multiples of the base investment. Momentum over the past year towards this long-term goal is evident. The NSDL projects funded by the NSF represent a broad range of scientific and education expertise, and the range of innovation demonstrated among them is remarkable [Zia 2001]. A vibrant and motivated NSDL community has emerged, shown by the enthusiastic participation of

funded projects and external partners at the first two annual All-Projects meetings. The result is excellent progress [Arms 2002] toward a December 2002 operational release.

Following the initial release, the NSDL challenges become increasingly educational: to improve how science, technology, engineering, and mathematics are learned across all sectors of society. The Third International Mathematics and Science Study—Repeat validates this need for change. TIMSS-R gives a disappointing assessment of American educational practices in these fields. Though our fourth-graders rank well among their counterparts in other countries, they clearly lose their edge within four years: eighth-grade students from fourteen other countries show *significantly* higher mathematical and scientific skills than do our own [NCES 2000]. Of particular importance from this study—especially for NSDL—are data showing that the lag occurs despite significantly better access to computers and the Internet by the U.S. students and their teachers.

Fortunately, there exists excellent pedagogical research to mitigate this news about the ineffectiveness, so far, of technology in schools. As we describe briefly in Section C.3.3, there are solid results showing that computers and networking can significantly enhance learning, when combined with pedagogy that is informed by current understandings about cognition, knowledge creation, i.e., constructivism, and the dynamics of collaborative learning [Jonassen 1996]. Thus a central NSDL challenge is to create not just a rich repository but an *intellectual commons*, where students and educators interact and are stimulated to change the way they teach and learn. This challenge is both technical and pedagogical. The CI team is not positioned to take on all the challenge of transforming educational practices by itself. However, by leveraging our own efforts on those of others, and vice versa, we believe the NSDL can have a major impact in making good our schools' and governments' investments in technology and educational research.

The key concept is leverage. We propose—by establishing a strategic set of partnerships and building an organization that supports parallel themes of *rich resources* and an *empowered community*—to build an NSDL that is a catalyst in support of the Nation's education agenda [PCAST 1997]. The overall NSDL Program budget, about $25 million per year, is modest compared with the total national expenditure for scientific and technical education, and the 60+ projects within the program cover a tiny fraction of the resources that are needed. Yet the NSDL potentially can have impact many times greater than that investment would seem to yield. The obstacles are impressive, but the potential benefits are huge. This proposal defines how the Core Integration (CI) team will build on the initial NSDL work, providing the organizational, technical, and operational leadership necessary for moving the NSDL toward its full and ambitious potential.

C.1.2 The Core Integration Strategy

The CI team was selected in summer 2001, and our concept of integration remains true to the original proposal [Fulker 2001], though that document focused primarily on the first year of effort. This proposal emphasizes longer-term aspects of integration, broadly defined, with greater focus on the educational effects of NSDL. Our strategy is built around two themes, which share the concept of leverage: multiplying the CI efforts by the contributions of others, and using our effort to further the goals of others.

The first theme is to build a *comprehensive* library of digital resources for all aspects of education in science, technology, engineering and mathematics. In our plan, the word *resource* is a very broad term. Students learn best when they engage in experiments, discover things for themselves or with others, exchange ideas with peers and mentors, and generally become involved with the subject matter. NSDL must support such activities with a very rich set of resources. To achieve this degree of completeness, we must develop methods whereby a small team multiplies the efforts of thousands of other groups and individuals. We are fortunate in being able to build on ten years' research in digital libraries, much of it funded by the NSF. However, the NSDL presents significant new challenges related to scale and breadth of audiences, resources, and services. The NSDL must respond to these challenges in a cost-effective manner or it will become too expensive to maintain. Technology to automatically find, organize, and manipulate resources will help accomplish this goal.

The second theme is to capitalize on *communities* that are involved with scientific and technical education, both traditional and non-traditional. We will do so by building on the strengths of educationally focused NSDL projects [Hunter 2001], engaging additional public- and private-sector partners [Wittenberg 2001], and publicizing NSDL in suitable forums. Our task is not simply to build a library. We must stimulate networks of institutions and individuals (educators, librarians, learners, publishers, parents, etc.) to become stakeholders in the NSDL, utilizing its resources, enhancing its quality, and finding it to be a place for fruitful discourse on education and learning. The payoff will be great—such stakeholders will feel less threatened by the impact of technology, and they will help ensure that NSDL does not overlook subtle social factors that are crucial to the creation of knowledge. (See *The Social Life of Information* by John Seeley Brown and Paul Duguid [Brown 2000].)

Much of this proposal is about practical details: reporting progress; characterizing our efforts to design the library and grow the collections; depicting the services we will offer; and describing our role in the organization of the NSDL program as a whole. But these prosaic details are only the means, not the end. Every detail should be seen in the context of the vision: to build a superb educational resource that will multiply the efforts of others and have a major, long-term impact on education in America.

C.1.3 Multiple Audiences

In an ideal world, all of the audiences that are implicit in the NSDL's broad mission would be served equally and at once. However, for pragmatic reasons of budget, scale, and readiness, we will target these multiple audiences in phases. The initial release will have a focus on serving teachers, librarians, and other professionals, since they are the primary agents of change. Inevitably, the resources that are available first will emphasize those disciplines that are best represented in the NSDL program. The projects currently funded by the NSF have a good mix of K-12 and undergraduate education, so that the NSDL immediately will serve both communities. However, the current projects are strong in some disciplines (e.g., earth sciences), but weak in others (e.g., chemistry).

In subsequent phases, the NSDL must become a tool that is adopted by—and even owned by—all students and teachers, at all levels of sophistication, in all disciplines, regardless

of their cultural backgrounds, physical limitations, or geographic locations. As discussed in Section C.3.3, we see the library as a cognitive tool, fostering the active creation of knowledge by both teachers and students. In the case of teachers, the knowledge gained will pertain to effective modes of pedagogy, in addition to subject knowledge about the worlds of science, technology, engineering, and mathematics.

Fitting the NSDL to the needs of different user groups is made possible by the technical strategy of *one library, many portals*. Each portal has a specific educational objective, but a portal is much more than a tailored view of NSDL contents; a portal can support the emergence of learning networks, communities of practice, and other groupings of people who engage in collaborative knowledge creation. Though teachers and learners are the primary audiences of the NSDL, we envision it engendering other vibrant communities. For example, we hope that creators of educational content—traditional publishers as well as newer members the field—will view the NSDL as a laboratory for experimentation and deployment of new learning tools. Similarly, we think the NSDL will provide a test bed of unparalleled scale and richness for digital-library researchers.

C.2 Primary Core Integration Roles

C.2.1 Progress Toward the Initial Release

The Core Integration team began work in October 2001. An immediate priority was to build the organizational and technical framework of the NSDL, in preparation for the first release in December 2002. As described in the previous proposal [Fulker 2001], this work has three aspects: engaging the community, providing technology, and operating core services. Here is a brief summary of the progress in the first year.

Engaging the community. The first act of the CI team was to establish a communications portal for library developers [NSDL 2002]. The CI team has organized the NSDL National Visiting Committee, the Policy Committee and its subcommittees, the annual meetings of the Principal Investigators, and several ad hoc meetings. The 60+ projects funded by the NSF have been augmented by groups such as ERIC and the Institute of Museum and Library Services (IMLS). The CI has created a Project Relations team that works with library builders and holds primary responsibility for integrating the NSF-funded projects into the initial release.

Providing technology. The CI team has established an architectural framework and is assembling the core technology needed to support the central services and to integrate resources created by others. Particular effort has been placed on:

- Building a metadata repository (MR), which holds native and standardized metadata records for each collection and item known to the NSDL.
- Implementing interfaces (primarily OAI protocols) by which the MR is populated and by which its contents may be accessed to construct various library services.
- Interfacing and testing fundamental library services, discussed below.
- Building the primary portal from user-interface *components* that can be reused in a broad array of additional portals.

Operating core services. The organization to support a robust production service is now in place. The first library-wide services that will be operated by the CI include the main NSDL portal, a comprehensive search and discovery service, and an initial authentication and authorization service. First versions of the metadata repository and the search service already are functioning, and a prototype portal built from the interface components is on schedule for testing in June. This will be followed by a busy period of integrating NSF-funded collections (those ready for production) and testing for the release in December.

C.2.2 Building a Comprehensive Library for Education

The initial release of NSDL will be a spare framework, consisting of basic services and resources from a relatively small set of collections. Our emphasis in the next phase will be on growth: expanding the collections, adding new services and partners, and—above all—encouraging use for education. In each area, the CI strategy is one of leverage: using our limited resources to stimulate broad community participation.

Growing the collections. A major goal is to make the NSDL a comprehensive library, covering all areas of scientific, technical, engineering, and mathematics education. This requires a vast expansion in the number of collections, far beyond those funded by the NSF. Our strategy has three tracks: working with other digital libraries, partnering with publishers, and building collections automatically from the Web. A Collections Development Board, formed in consultation with the NSDL Policy Committee (see Section C.4.2) will advise the CI team, developing guidelines, identifying gaps, etc. To integrate resources from the three tracks, the CI team will utilize technologies and support systems developed this year, but refined methods are certain to be needed for addressing intellectual property concerns (especially on the publisher track) and quality concerns (on the automatic track).

Rich services. The framework that we have established enables us to support potential service providers with expertise, standard interfaces, metadata, and tools. We know of several exciting services that others are proposing to the NSF and look forward to working with them. The CI budget will be sufficient to add a small number of general-purpose services, including preservation services and a generalized advice system. A key service that we will provide is a much expanded educational communications portal, designed to support interaction, collaboration, and sharing of information among educators and learners.

Educational services. Building the NSDL is only the beginning. This proposal describes how we will engage the educational community through several forms of outreach and leverage. We will continually evaluate the use of the library and augment the collections, portals, and services based on those evaluations. We will provide online forums so that library users —including teachers, librarians, and students—can also be contributors. Success in this activity depends on partnerships: with national organizations, such as the AAAS; with discipline-based societies, such as the Mathematical Association of America; and with professional organizations, such as those for school librarians.

C.3 Managing for Long-Term NSDL Success

C.3.1 The Core Integration Operating Divisions

To support a shift in emphasis from the initial release to a comprehensive, fully operational NSDL, we have established a management structure that will provide the needed organizational, technical, and operational leadership. This structure has six operating divisions, shown in Table 1. As indicated by the multi-site composition of the associated teams, our approach resembles matrix management, with some members serving on multiple teams to ensure effective interaction across divisions.

Table 1: The operating divisions of the NSDL Core Integration team

Central Office *Executive Director:* Dave Fulker (UCAR)	*NSDL Executive:* Arms (Cornell), Wittenberg (Columbia), and all Division Directors
Manages the CI team, builds an effective relationship between the CI team and the community governance model/committees, and leads the NSDL, as a whole, toward mission fulfillment.	
Educational Services *Director:* Kaye Howe (UCAR)	*Senior Members:* Hoffman (EMU), Marlino (UCAR), Sumner (Colorado)
Ensures that NSDL is beneficial across all of science, technical, engineering, and mathematics education, through strategic partnering, provision of high-impact services, targeted publicity, and continual evaluation of how well NSDL meets the needs of learners.	
Technology Design and Development *Director:* Carl Lagoze (Cornell)	*Senior Members:* Allan (U. Mass), Frew (UCSB), Hoehn (Columbia), Janée (UCSB), Jesuroga (UCAR), Moore (UCSD)
Designs and evolves the technical architecture for the NSDL, including the selection or development of critical interfaces, protocols, and essential underlying services.	
Library Services and Operations *Director:* Diane Hillmann (Cornell)	*Senior Members:* Jesuroga (UCAR), Krafft (Cornell), Ludaescher (UCSD), Millman (Columbia), Lankes (Syracuse)
Ensures that NSDL meets user needs for collections (breadth, depth, and depiction), reliability, availability, responsiveness, and core services, including general-purpose and tailored portals, search/browse capabilities, user identifications/profiles, access management, and help desks.	
Project Relations *Director:* Susan Jesuroga (UCAR)	*Senior Members:* Hoehn (Columbia), Saylor (Cornell), Lankes (Syracuse)
Coordinates and supports the efforts of collection/service developers—particularly those funded by NSF—to comply with standards, integrate their work into NSDL, and form a community, through problem solving and the provision of collaboration environments.	
Publisher Relations *Director:* Kate Wittenberg (Columbia)	*Senior Members:* Hoffman (EMU), Howe (UCAR), Millman (Columbia), Saylor (Cornell)
Builds relationships with publishers and private-sector entities to maximize the synergy between NSDL and traditional disseminators of high-quality learning materials. Develops IP policies and a business model to support these relationships.	

C.3.2 Executive Director and Central Office

The CI team is located at several different organizations, but is organized in a single management structure under Dave Fulker, the Executive Director. The CI Executive consists of the five divisional directors, who are listed in Section C.3.1, plus the three Principal Investigators, Dave Fulker, William Arms and Kate Wittenberg.

The CI Central Office has two major functions: management of the CI team and overall leadership/guidance for the whole of NSDL. The Central Office coordinates outreach activities as well as interactions with the governance committees, described in Section C.4, allowing the NSDL to benefit from the energy and vision of many people.

C.3.3 Educational Services

C.3.3.1 The National Education Agenda

Ultimately, the success of the NSDL will be judged by its impact on science education. Indeed, the NSDL initiative is [Marlino 2002]:

> *...in part, a response to the larger national science education agenda put forth by the NSF, the American Association for the Advancement of Science [AAAS 2001], and others [NRC 1996; NSF 1996; NSF 1997; NSF 1998; NSF 1999a; NSF 1999b]. This agenda calls for a variety of reforms, including inquiry-based science learning, and emphasis on the process of "doing science," and the integration of scientific research into education. Digital libraries promise to be a powerful tool in realizing the goals of these reform initiatives.*

The scope and the complexity of this call are such that we cannot have *direct* interaction with the millions of students and teachers we hope to serve. Therefore, our strategy for maximizing the educational impact of the NSDL is one of leverage, with five primary emphases—formative and summative evaluation, focused projects, a portal for educational communication, educational partnerships, and public relations—described in the subsections that follow. This strategy is informed by a belief that the NSDL, rich in resources and services, is both a cognitive tool [Jonassen 1996] and an intellectual commons. As stated in *How People Learn* [Bransford 2000]:

> *What has not yet been fully understood is that computer-based technologies can be powerful pedagogical tools – not just rich resources of information, but also extensions of human capabilities and contexts for social interactions supporting learning.*

The technology base of the NSDL, while the means of its ubiquity and inclusiveness, is still only the means. *Use* of the library by educators and students, i.e., its vitality as an intellectual commons, is what will cause positive educational change. The NSDL will provide pedagogical tools, an environment for inquiry and standards based learning, a locus for networked communities of practice, and a commons for those who, as Aristotle wrote, "desire to know."

There is still much to learn about how digital libraries may fulfill their promise as powerful cognitive tools [Marlino 2002]. Though proven tools, such as hypermedia authoring systems, semantic networks, expert systems, and data analysis capabilities, already are available for use *within* some digital libraries, the meta-value of the library

itself remains to be understood and exploited. We anticipate that the key will be in how NSDL helps users see and define *relationships among resources* and how NSDL plays out as an *intellectual commons*. Effectiveness along these lines will permit educators and learners to enlist the library as an aid in constructing their own knowledge about relationships among pedagogy, science, technology, engineering, and mathematics.

C.3.3.2 Formative and Summative Evaluation

The overall goals of the evaluation process are to inform the development and evolution of the library, to optimize its technologies and its implementation, and to determine the educational impact. Evaluation challenges in digital libraries are exacerbated by their technological complexity, the variety of content, uses, and users, and the lack of agreed-upon evaluation methods and metrics [Borgman 2000]. Because of these challenges, and the nature of NSDL as a distributed effort, we anticipate that evaluation activities will be conducted in a distributed fashion, and on many different levels by studying both the library as a whole and its individual components. The evaluation model for this effort is a comprehensive one, and intended to examine processes, technologies, uses, and outcomes [Fuhr 2001]. The role of CI in this effort is to coordinate community evaluation efforts, support the Evaluation Working Group (see Section C.4) in its distributed evaluation activities, and initiate targeted efforts to examine the impact of the NSDL on classroom culture and student learning outcomes.

Work is currently underway to develop an initial evaluation strategy for the distributed NSDL with respect to library building processes, library usage, collections growth and quality, and the usefulness and usability of library services. This is a joint project of the CI and the Evaluation Working Group. It is based on a six-month evaluation pilot study to develop analysis guidelines and metrics for the NSDL. The strategy will be designed to work with distributed portals, and individual libraries, collections and services, as well as the CI portals. Additionally, formative evaluation information will be collected to determine better the technical, social, and cost issues inherent in this large-scale distributed effort. A preliminary evaluation plan, informed by the pilot, will be distributed to the larger NSDL community for review and refinement. This plan will not be a final stage in the process, but will evolve and be refined over the four-year period of the project, in response to community activities and proprieties.

The leaders of the evaluation effort are Tamara Sumner, chair of the working group and Assistant Professor of Engineering at the University of Colorado (under contract) and Mary Marlino, Director of the DLESE Program Center. Professor Sumner will serve as a usability and collaborative learning consultant and we will support a Ph.D. student to work with Cornell University's Human Computer Interaction Group to study the usability of the interfaces to the library.

C.3.3.3 Focused Projects

During 2000/2001 the NSF invited the TeacherLib Project to work with the CI team in supporting K-12 education. Ellen Hoffman, leader of TeacherLib, is now joining the CI team as a subcontractor to develop outreach and library services with a K-12 orientation. To have significant impact in this huge arena, Professor Hoffman will engage *networks* of K-12 teachers and librarians, and her understanding of these communities will give us

access to their field-based experience and insights. A major responsibility will be to help build the educational communication portal and assure its clarity and usability. She also will assist in the development of a portal that serves children and can be adapted to the relevant federal, state and district regulations.

Accessibility and participant-diversity are important facets of the NSDL. Among the NSDL projects are several—including Digital Library Services for American Indians; WGBH; Internet Scout; and the Gender and Science Digital Library—that bring relevant work and experience, including means to address the Americans with Disabilities Act.

We are contracting with Syracuse University to adapt and operate its Virtual Reference Desk (VRD) Project for an *AskNSDL* service [VRD 1998]. Within the Information Institute of Syracuse, VRD operates the ERIC/IT, AskERIC, and GEM services, which are widely used by teachers, parents, administrators, researchers, and students. VRD hosts the definitive conference on digital reference, attended last year by over 450 international librarians and information professionals. The next of these conferences will include museum and education professionals, complementing the library perspective. The Educational Resources Information Center (ERIC) has 650 partner organizations for outreach, training, publications, and user services. In 2001 AskERIC responded to 34,181 e-mail-based questions, to over 1,000 live questions (in a chat environment), and received some 5 million hits per week on its web site. The Gateway to Education Materials (GEM) has 25,000 professionally evaluated records; received 12 million hits on its web site, and has over 300 organizational members in its Consortium.

These projects have contacts and databases that are of value in our outreach and evaluation efforts. Furthermore, Syracuse can help us build relationships with educators and librarians, as we position the NSDL to be the resource of choice for pedagogy and curriculum-development in science, technology, engineering, and mathematics education.

C.3.3.4 The Education Communications Portal

To support the emergence of learning networks and communities of practice, we will provide an educational communications portal that will encourage teachers and students to discover knowledge bases, discuss concepts, engage in collaborative knowledge creation, and contribute that knowledge to the NSDL. Currently, the CI is supporting a single communications portal, using the SourceForge software system. This has proved very effective for developers and other regular users, but overly complex for occasional users. The education communications portal will have a distinct emphasis: simplicity of use with basic services (e.g. discussion forums, collaboration tools, and newsgroups). The developers' portal has fostered a technical discourse and the sharing of resources; we envision the education communications portal fostering a similar education-oriented network of practitioners. We intend to conduct usability studies with our numerous audiences to ensure usefulness and use.

C.3.3.5 Educational Partnerships

Our goal is to be the resource of choice for disciplinary, curricular and pedagogical issues. Hence, NSDL should be a partner and major asset for every significant initiative in education for science, technology, engineering and mathematics. Teachers and

librarians have the most immediate impact on the quality of education. NSDL will fail if we do not develop a constant dialogue with these two groups. To accomplish this, we first must raise the visibility and knowledge of NSDL among opinion leaders in the educational and library communities.

Initially, we are focusing our efforts on developing strategic partnerships with a key set of professional organizations, including the National Science Teachers Association and the American Library Association. Within the ALA, we aim to develop partnerships with the American Association of School Librarians, the Association for Library Services to Children and the Public Library Association as well as the Association for Library Collections and Technical Services.

The outcome of these partnerships for NSDL will be an intimate knowledge, through consultation and feedback, of the pedagogical and resource needs of teachers and librarians as they face the daily realities of American classrooms and schools. One of our major agendas must be to meet those needs and empower these influential educational networks. The outcome for teachers and librarians should be a reduction of isolation, the establishment of communities of practice and the opportunity for a continually evolving mastery of their professions. Over time, we expect that a culture of contribution will emerge, in which teachers and librarians themselves will contribute to the library, share best practices, and help us refine library priorities.

The next phase is to involve discipline-specific organizations, including professional and learned societies. Our aim is to engage a group of partners who cover all disciplines and all levels of science education. The societies can help us understand how adequately NSDL resources represent the most current resources and research in their fields. We will also work to incorporate NSDL in their existing professional development activities. We know that without adequate proficiency in its use, a resource like NSDL may never have the impact it should. The incorporation of NSDL within the professional development work of these societies will not only spread knowledge and use of NSDL but also give us valuable information about usability and effectiveness.

The American Association for the Advancement and of Science, especially its Project 2061 [AAAS 2002], is a particularly valuable strategic partner in multiplying the value of NSDL to education. The recipient of a recent grant of $5.9 million from NSF, Project 2061 is "creating and testing an outreach campaign designed to build public support for science literacy." The project also notes that the grant will support "the development of new tools for teachers, curriculum developers, and textbook authors and publishers." As the national library for scientific education, we want to become the distributor of these educational resources. Specifically, we will encourage partnerships, technical developments, and pilot projects that provide scaffolding to the standards-based curricular reform and assessment efforts that are emerging from AAAS, in addition to those of the National Science Education Standards (NSES) published by the National Research Council [NRC 1996].

C.3.3.6 Public Relations

NSDL has strategic resources in its National Visiting Committee, Policy Committee and the growing group of NSDL project investigators. We are a partnership of partnerships, and the presence of people like Jeanne Narum of PKAL, Len Simutis of the Eisenhower National Clearinghouse, and Gerry Hanley of MERLOT on the Policy Committee give us constant access to the experience and networks of these and other groups.

The National Visiting Committee brings skill and experience in private foundation funding, alternative business models, publishing and private business. The Core Integration PI at Columbia, Kate Wittenberg, will be working with them on publisher, education, and technology business relations so that we understand the competitive environment we are working in and develop partnerships within it.

In preparation for the first release of NSDL, we are developing a suite of marketing materials, plans for media outreach, coverage of major conferences and meetings, the acquisition of the domain name nsdl.org from the Nova Scotia Drama League, designs for the portals of the library, and the evaluation activities that will support the continuous improvement of the library from the outset.

C.3.4 Technology Design and Development

From its inception the CI technical effort has built on two guiding principles: *a spectrum of interoperability* and *one library, many portals*. The first reflects a belief that success for the NSDL rests on the breadth of its collections and depth of its services. Achieving this depends on a technical infrastructure that accommodates rather than prescribes, allowing providers to participate at a level that fits their resources, technical expertise, and requirements. The second principle addresses the reality that the NSDL must serve a diversity of users, who vary in language, culture, educational background, and physical ability. Together, these goals require a layered architecture with open and data rich interfaces that facilitate customization of presentation and varying interface functionality.

C.3.4.1 The Phase I Architecture

Phase I of the CI technical architecture is scheduled for public release in December 2002. Figure 1 shows its principal features, reflecting the explicit recognition that the NSDL cannot dictate a single set of standards or demand that detailed metadata accompany all resources in its collections. We have to accommodate our services to what is available. Phase I includes the following features.

- *Metadata Repository* – The Metadata Repository is the central record of all resources known to the NSDL, providing robust storage of the metadata provided by participating projects as well as metadata gathered and automatically generated from open-access Web resources. The Metadata Repository exports the metadata to services such as search and browse, in both raw and normalized forms, via the Open Archives Initiative protocol for metadata harvesting [OAI 2002].
- *Search and Discovery* – Because the metadata provided by collections vary dramatically in formats, quality, and comprehensiveness, information discovery in the NSDL is particular challenging. The service combines indexing of metadata from the Metadata Repository with indexing of full text where available. We are using the InQuery search

engine [Allan 2000] and the SDLIP search protocol [Paepcke 2000]. This work is being carried out by the Center for Intelligent Information Retrieval at the University of Massachusetts, Amherst.

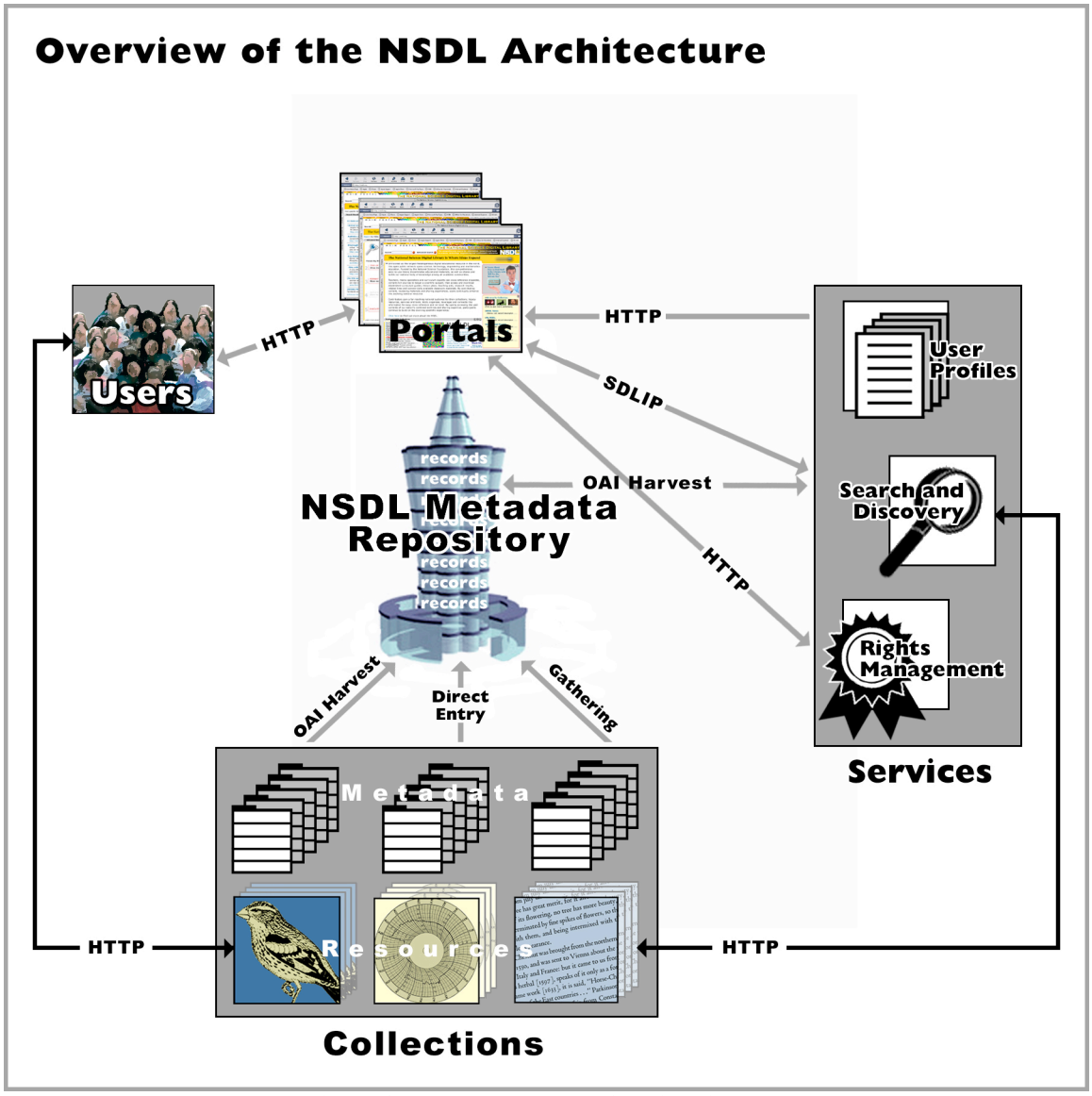

Figure 1. The Phase I Technical Architecture

- *Access Management* –The core access management system relies on the Shibboleth protocol [Shibboleth 2002] to distribute identity verification (authentication) and cohort membership (authorization) to the administrators of distinct communities of users. Federated communities performing user identity and capability management can easily tie-in to this system using standard protocols (e.g., Kerberos and LDAP).
- *User Interface and Portals* – The CI team is creating an initial set of portal components and several portals. They include a main library portal based on uPortal, which is a free, open-source, portal framework being developed collaboratively by a group of higher-education institutions [uPortal 2002].

C.3.4.2 Phase II: Scalability

The initial release is the first step in the long-term evolution of the NSDL architecture. The period immediately following the initial release will be dominated by the task of integrating a vast influx of new collections and establishing operational procedures for maintaining a robust and coherent library with minimal staff intervention. As a result, the Phase II release, planned for mid 2004, will not reflect substantial new technical concepts, but will incorporate necessary refinements to accommodate scaling issues.

A parallel task will be to extend the spectrum of interoperability. Many of the most valuable educational resources are dynamic, multimedia, interactive, or collaborative. To embrace such resources, the NSDL must offer additional technical options. The choice of standards and mechanisms is a complex process in which the needs of the NSDL, the availability of the technology, the costs, and the wishes of our partners are all important. To a small extent we can lead by example, but mainly we have to interpret trends and adopt the technical approaches that will best stimulate the growth of the NSDL. We will make these choices in conjunction with the full NSDL communities. Members of the CI team are actively involved in the World Wide Web Consortium, Internet2, and the NSF-funded digital library projects, with good industry contacts.

C.3.4.3 Extending the Architecture

Because the NSDL is the largest and most diverse digital library to date, it is well positioned to become the premier context for the deployment of new digital libraries technologies. Currently we have special relationships as early adopters with the Open Archives Initiative, uPortal, and Shibboleth developers. For Phase II and beyond we will need to identify the next set of technology partners. The list below describes some of the new functional areas, the current best practices and possible partnerships.

Describing and Integrating Services. The Phase I architecture lays a basis for extended services by collecting and storing metadata in various forms in a central repository. However, a more powerful multi-service framework requires standards for describing services, to permit interactions among them and access to them by users and agents. We expect to leverage recent developments in Web services, including the Web Services Description Language (WSDL) work within the W3C [W3C 2001] and perhaps the .net work from Microsoft [MS 2002]. (Our partners at the University of California, Santa Barbara are developing such services for geospatial information [Janée 2002].)

Creation and Management of Federated Ontologies. The NSDL will encompass resources from many disciplinary domains. Providing a coherent end-user view will require developing and establishing conceptual relationships among distributed resources. Work by the Semantic Web community [W3C 2002a] and some of our partners (UMass, UCSB, and UCSD, in particular) may prove useful in meeting this need.

Creating, Packaging and Accessing Complex Content. The resources available in the NSDL will be rich and complex. Dynamic and multimedia content requires access methods and digital object structuring standards that extend far beyond those now available via Web standards. Good work is coming out of the MPEG community, the METS work in the Library of Congress [METS 2002], and the FEDORA digital object

architecture [Payette 1998], but nothing has yet reached the level of acceptance where we would be confident to select it for the NSDL.

Scientific Data. Of particular value in science education are numerical data sets and other digital objects that are of little use except in conjunction with appropriate tools. For example, any student could make direct use of a traditional map room, but the digital equivalent requires use of Geographic Information System (GIS) or comparable tools. Similar problems arise in using other types of scientific data or using mathematical expressions, represented digitally. The CI team has direct experience in a number of these areas and will track advances in others.

Preservation of Digital Resources. The utility of the NSDL depends on the integrity of the resources, notably their longevity and stability. There is no one solution to the problems of digital preservation and longevity. Instead, we expect to develop a suite of approaches, by exploiting a number of active research efforts, including the persistent digital archives work at the University of California, San Diego [Moore 2000], the LOCKSS work at Stanford [Reich 2001], and the digital asset risk management at Cornell [Vernon 2002].

The NSDL is a natural laboratory to test new methods of preservation at a national level. This is currently the focus of a major planning study being led by the Library of Congress. Several of the leaders of the CI team are experts in this field and the Supercomputing Center at the University of California, San Diego operates one of the world's largest and most advanced preservation services. As a first step for the NSDL, the center will preserve static versions of NSDL-funded content and metadata, but this is just a beginning.

Data Provenance. Another factor related to integrity of digital resources is their record of origin. This is especially true for scientific data, where origin and derivation is critical to determining the veracity and utility of the data. We plan to leverage current research work in this area, notably the Data Provenance project at University of Pennsylvania, which is part of the NSF's Digital Library Initiative [Buneman 2002].

Annotation and Review Frameworks. The Metadata Repository has flexible means for annotating resources. This is intended to support educational services by linking reviews of resource quality and appropriateness, and encouraging collaborative reviews. On-line retailers such as Amazon and services like Epinions have demonstrated the high value of volunteer review systems and systems that review the reviewer. We plan to exploit work that extends these systems beyond commercial applications, using technology such as the Annotea work [Annotea 2002] coming out of the W3C and automated scholarly review systems being developed in the Digital Libraries Initiative project at the University of California, Berkeley [Riggs 2001].

Authentication of Digital Objects. Communities of authors, publishers, librarians, and consumers have consistently emphasized the importance of document authenticity. In the digital realm, comprehensive document authentication is in only early stages of development, but considerable progress can be made with simple tools. *Author Prestige*

Accounting. For authors, an important value in NSDL will be to understand how their work is being used (by whom and how often, e.g.) in order to receive the same level of academic credit for digital citations as for print publication. The planned NSDL Access Management infrastructure is uniquely positioned to supply this information to an author without compromising the privacy of the user. We plan to develop this technical capacity explicitly, as a component made available to collections and new services.

Collaborative Learning. Over the past decade, theoretical advances in distributed cognition [Hutchins 1995, Salomon 1996] and social constructivism [Vygotsky 1989, Lave 1991], coupled with the emergence of the Web, have contributed to rapid developments in the field of computer-supported collaborative learning. As demonstrated at a recent conference, hosted by the University of Colorado-Boulder, researchers and educators are pursuing a diverse array of technical and curricular approaches [Stahl 2002], many funded by the NSF. In the K-12 arena, collaborative knowledge construction environments, such as Knowledge Forum [Scardamalia 1994] and the Web-based Inquiry Science Environment (WISE) [Linn 2000], are being deployed in formal classroom settings and demonstrating significant learning outcomes. In the undergraduate arena, computer-mediated communication tools (e.g., WebCT [WebCT 2001], Blackboard [Blackboard 2002]) and collaborative authoring tools (e.g., CoWeb [Guzdial 2000]) are being widely used to support a variety of course management and collaborative learning activities. Collaboratories, enabling researchers and scientists to collaboratively collect and analyze real-time scientific data, have been shown to have positive impacts on the practices of scientists located at smaller, or less well-funded institutions [Olson 1998].

Supporting collaborative learning is of central importance to NSDL. It is beyond the resources and expertise of the CI to try to recreate or supplant the rich array of collaborative learning systems currently underway. Instead, the CI will take a two-pronged approach. First, we will shortly release a first-generation education communications portal for supporting educational discourse, using standard computer-mediated communication systems such as mailing lists and threaded discussion services. Second, we will work with the collaborative learning community to begin the process of identifying interoperability standards for integrating collaborative learning systems with digital library substrates, recognizing that, as yet, there are no widely accepted standards for enabling the diverse array of systems to communicate with each other or with information repositories.

C.3.5 Library Services and Operations

The Library Services and Operations division supplies the central library functions of collection development, metadata, portals, user support, and the integration of collections and services provided by partners, including the search and authentication services. The users see a production quality digital library, but the library developers are often treading untrodden paths. Many of these services are far from routine; no digital library has accommodated such diverse materials, with such a wide variety of metadata, or no metadata at all; the small staff size constrains every activity.

Our services and operations plan is designed for scalability and thus exploits automation. Every task is a candidate for automation, and every facet of every service is planned for robustness and enormous growth. The plan positions the CI team to be a multiplier for the NSDL program, and hence for scientific and technical education everywhere.

C.3.5.1 Collection Development

The purpose of collection development is to identify the categories of resources that are needed in the NSDL, find those materials, and work with providers to integrate them. This responsibility also involves other CI teams (see Sections C.3.6 and C.3.7). Already, we are working extensively with existing digital libraries and other organizations with resources of value in science and technology education. As discussed previously, a Collections Development Board will advise on this process, especially in determining needs for resources. Evaluation results also will inform this activity. Eventually, automated methods will augment human collection-building efforts.

Given their diverse sources, the quality of NSDL materials is a very complex issue. The expected size of the NSDL means that it will be impossible to select every item individually. Furthermore, the concept of the CI (or any other group) defining what constitutes good science and what resources are suitable for students at various levels runs counter to the traditions of our society. As a practical matter, requiring CI staff to review items individually would condemn the NSDL to a small size. Like any large library, the NSDL cannot guarantee that all resources are correct, or are suitable for the audience. Yet quality is very important. A crucial aspect of collection development is to develop the guidelines by which collections of materials are acquired. Because the mission of the NSDL is broad, definitions of relevance and appropriateness will necessarily be equally broad. As far as possible, we will rely on review and annotation services, plus the imprimatur of respected collections, to identify the resources of highest quality and suitability for specific contexts. Moreover, we will ensure that users know what criteria have been the basis for selecting the collections, so that they can make informed decisions about which resources are appropriate to their needs.

In creating a portal specifically to serve children (Section C.3.3.3), we will combine the aforementioned measures of quality and suitability with the flexibility of our search service and the power of our portal-building components. In this way, we will create a view of the library that garners high confidence among teachers, parents, and school boards without compromising the richness of the NSDL for other audiences.

C.3.5.2 Metadata Repository

The Metadata Repository is the keystone for building educational and library services that remain robust and coherent as the library grows to very large scales. It is a central source of information about NSDL collections and the items within them, and it makes that information available to service providers. In many respects, the NSDL Repository is like the central catalog of any large library, but it breaks new ground in several areas, supporting greater diversity of content and metadata than conventional libraries, and accompanied by software and processes built specifically to support NSDL goals.

Our strategy is described in the NSDL Metadata Primer [Hillmann 2001] and is based on the understanding that metadata are too expensive for the CI team to create in large quantities. Hence, the NSDL will reply on existing metadata or metadata that can be generated automatically. The strategy is to gather all the metadata available from collections, recognizing that they vary greatly, from none to extensive. Novel features of this metadata strategy include a choice of native metadata formats, normalization to a standard format, and the exposure of metadata to other services for harvesting.

In addition to collection- and item-metadata, the repository will support *annotation* metadata about those resources, as well as metadata for exhibit and news aggregations. Taken as a whole, these various kinds of metadata allow services to provide many different views of the NSDL, for many communities of users.

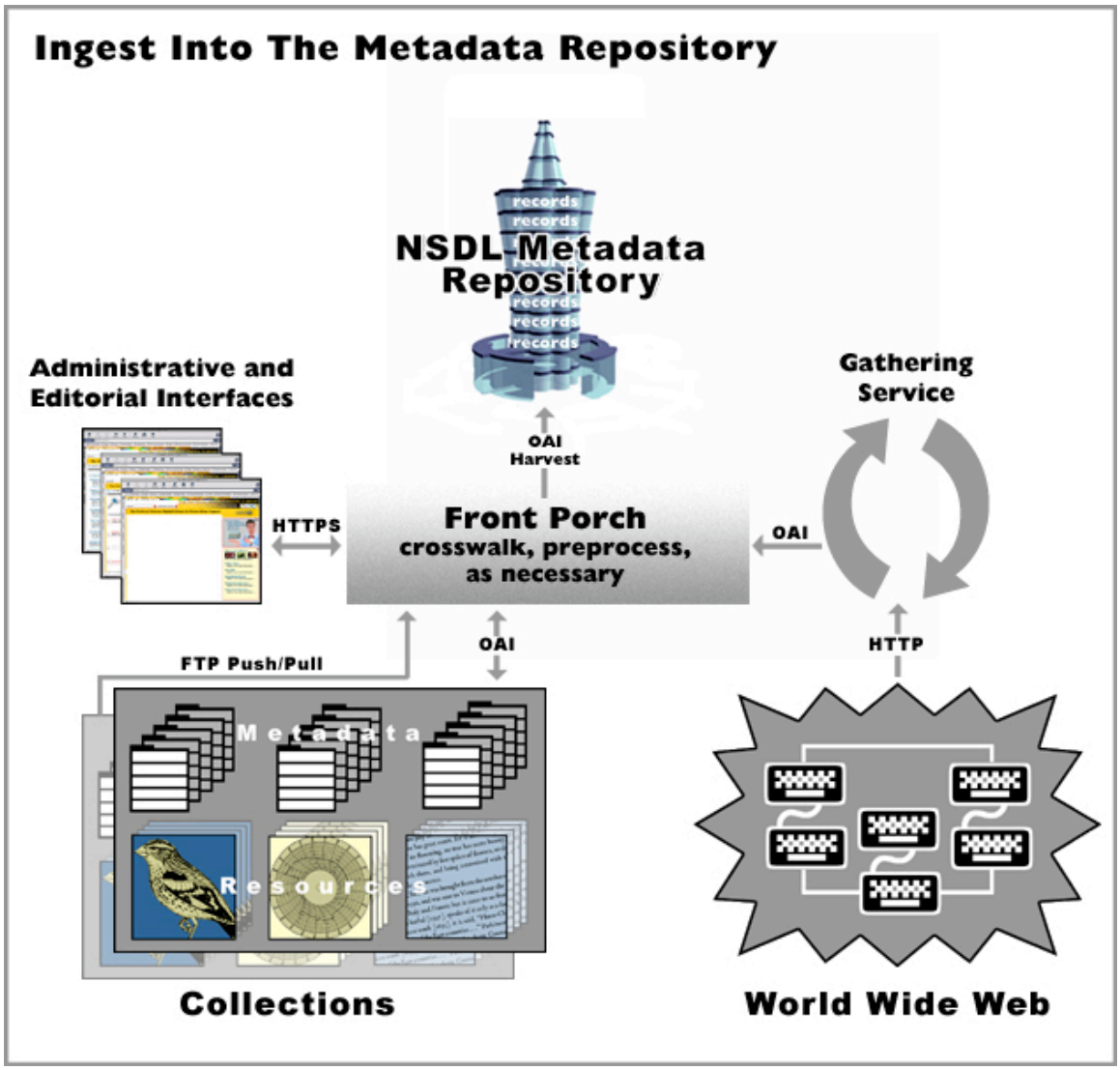

Figure 2. Metadata Flows into and from the Metadata Repository

Figure 2 shows how metadata are brought into the metadata repository and made available to service providers. Technically, the Metadata Repository is a relational

database system, wrapped with standard interfaces and protocols that hide the underlying implementation and present all metadata in XML format. Various protocols are offered for both ingest and export of metadata, with the OAI metadata harvesting protocol being the first choice. As with all the CI software components, the processes by which the Metadata Repository ingests records and makes them available are designed for automation. Realistically, much human effort will be needed for collection ingest in the beginning, because each is quite unique. But as protocols mature and experience is gained, addition of collections should become more routine and increasingly automated.

C.3.5.3 User Interface and Portal Development

The goal of *one library, many portals* is another area where the CI team is breaking new ground. The December 2002 release of the NSDL will include a main portal and a small number of special portals. We are currently working with colleagues who are developing a specialized portal on Using Data in the Classroom. In addition, we are creating a user configurable portal system (tentatively called MySite), which enables untrained users to create simple portals of their own. All portal developments are guided by the Federal standards for accessibility by people with disabilities, and recognize the wide variety of equipment, networking and expertise of different users.

Behind the user's view of the NSDL we are implementing—in conjunction with the Technology Design and Development Division—an architecture that enables skilled individuals to build data-driven user interfaces, which can grow to a very large scale without the need for creating individual Web pages by hand. Our goal is that subject experts, with some training, can, for example, create a chemistry portal for undergraduates or an online science museum for grade school children.

Initially, we have made two important decisions. The first is to use the uPortal framework and toolkit for the main portal. The second is to build a set of components that can be included in any portal, including interfaces to the search service, a browse service, logon and user profiles, a help service, news and exhibits. The components are designed to be usable with other portal technologies in addition to uPortal.

C.3.5.4 User Support

For reasons of scalability, our user-support and help services will be enhanced via automation—the CI team cannot provide direct personal assistance to individual users. However, users will be aided in gaining assistance from one another, e.g., from more experienced colleagues. As discussed above, we will engage Syracuse University under contract to provide its Virtual Reference Desk (VRD) and a modified form of its AskA system as central parts of the NSDL [VRD 1998]. We believe this will help establish a culture of sharing and a true sense of community among NSDL users and contributors.

The VRD service is e-mail based; questions are distributed to a large group of experts, primarily reference librarians, in previous applications of the system. Responses are not intended to be immediate in this model, as they are in the traditional face-to-face reference interaction. While some academic research libraries are experimenting with chat-style reference services, we believe the VRD approach is the only one that can scale to meet NSDL needs, under the current funding model.

Finally, a basic design principle is to minimize the number of operational problems that require user support, e.g., hardware and browser problems, by utilizing careful design processes and usability testing at the Cornell usability laboratory.

C.3.6 NSDL Project Relations

The initial task of the Project Relations division is to integrate all NSDL-funded collections and service projects into the library. The longer-term goal is to support the integration of projects far beyond those funded by the NSF. It is critical to engage all current and potential contributors in a way that builds community, and provides a sense of the larger library and its potential impacts on education. The Project Relations division and the Library Services and Operation division work together to integrate collections and services into the NSDL.

C.3.6.1 Project Support

The CI team will continue to establish personal relationships with all NSDL-funded projects, and with other strategic partners not funded under the NSDL umbrella. The Project Relations division will oversee these project relations, serving as the point of first contact for projects that require technical help, and providing NSDL information to other interested parties. The division maintains the CI database of project information, including contacts, outcomes, integration information, and major milestones, and uses this to track progress, and refer projects to others with similar needs, outcomes, or interests. Trouble-shooting assistance—linking project staff with appropriate experts, tools, and documentation—is a key role.

To broaden the NSDL's scope, the Project Relations team works with the Collections Development Board, NSDL committees, and the community to discover and engage outside groups with relevant collections, services, or research agendas, and who would gain from NSDL collaboration. The team works closely with the CI development teams to enable smooth integration of projects into the library and ensure that the evolution of the core NSDL infrastructure utilizes community development efforts to their fullest.

C.3.6.2 Integrating Collections and Services

The resources that will be incorporated in the NSDL are selected for their value in education, not for their technical virtues. Many splendid collections have limited technical support, e.g., they lack the expertise to set up OAI servers to expose their metadata. Some lack any metadata at all. To support integration, we work with some collections and services individually; for others, with more in-house expertise, we are building tools and documentation, so that they can help themselves. We also make available tools and lessons learned from other NSDL projects. This is a satisfying task, but it is never-ending. As discussed in Section C.3.4, the plan is to incorporate resources of ever-increasing sophistication over the next few years, particularly dynamic materials for experimentation, collaboration and interacting with information. Each advance will require more handholding and more supporting tools.

Services are even more difficult to integrate because they are potentially more varied and complex, and because the computing industry has not reached consensus on standards. Most of the early efforts to integrate services have concentrated on the basic library

services that users expect to see, at least in skeletal form, in the first release. Examples of these services harvest metadata from the repository, perhaps combine it with content drawn from collections, and provide an interactive interface through the portal components. Over the next few years, we anticipate integrating complex Web services, but cannot yet predict our role in supporting them.

C.3.6.3 The Developers' Communication Portal

The developers' communications portal provides community tools, including collaborative workspaces, email list and forum management, and source code repositories. Along with these tools, the portal contains technical documentation describing core infrastructure standards, protocols, programming interfaces, and other information to help projects understand how best to contribute to the library and work with the CI team. The portal will also continue to support a publishing mechanism to circulate the latest news of interest to the community. In addition, project principals have access to a community database and planning tool (a collaboration with smete.org) to find potential collaborators, perform gap analyses, and help develop consensus on community goals and values that shape NSDL.

C.3.6.4 Other Activities

Other project-relations activities include:

- The annual All-Projects meeting, bringing the community together for governance and technical discussions and workshops focused on collection- and service-builders' needs.
- Focused interactions in workspaces on the developers' portal, to encourage community discussions on such subjects as accessibility and diversity, sharing lessons learned from current projects, and other topics pertinent to growing the library and its services.
- Teleconferences and virtual "brown bag lunches", to encourage discussions and dissemination of information among interested community members on specific topics such as building new portals, evaluation strategies, and project integration issues.
- Communications and documentation support for the NSDL Policy Committee and its associated committee structure.

C.3.7 Publisher Relations

For the NSDL to become a resource of choice, used often by a broad range of teachers and students on a national scale, it is necessary to engage the interest and participation of the broad scientific publishing community. This includes both non-profit and for-profit organizations that control a substantial percentage of the high-quality educational science materials currently being produced by scientists and science teachers.

The NSDL must engage this community in a collaborative and productive manner, so as to insure that the NSDL becomes a strong and valued partner rather than a competitor to the traditional science publishing community. Science publishers possess a number of assets that will become critical to the future success of the NSDL, including an efficient and stable mechanism for acquiring and peer-reviewing high quality content from scientists; an effective system for editorial development, design, and production of this content; established models for contracts, licenses, copyright, and intellectual property management, and a reliable system for marketing and revenue production.

Clearly the NSDL community has much to gain from, but also much to contribute to the publishing community, including advances in digital design, interactivity, functionality, usability, information architecture, and pedagogy in science. By establishing a formal and stable means to engage the science publishing community, we will ensure that the NSDL reaches its full potential as a functional, valued, and highly used resource.

We therefore plan to hire a manager for publisher and partner collaboration. Responsibilities include: outreach to science publishers and the business community to inform them about NSDL and encourage their participation in appropriate ways; management of partner participation, including developing and negotiating partner agreements and terms, arranging business plans, revenue distribution, and rights management systems with partners (in collaboration with the CI team); advising on the selection and formatting of content for submission to the NSDL; and managing the editorial, production, and business aspects of ongoing publisher partnerships within the NSDL. Priorities for approaching publishers will be set in consultation with the Collections Development Board discussed previously. This staff person will have significant background in the field of educational, professional, and/or scholarly science publishing, including the negotiation of contract agreements, experience working with publishers and libraries in content development and business planning, and experience with market research and book and journal production.

C.4 Leadership and Community

C.4.1 National Visiting Committee

The NSF has appointed a National Visiting Committee for the NSDL, with thirteen members. It provides high-level strategic advice and counsel to CI as well as continuing evaluation of the project to NSF. Committee members also serve as advocates of the library to a larger public. The inaugural two-day meeting was in February, and the committee plans to meet twice this year and once yearly thereafter.

At its inaugural meeting the committee was briefed by CI on its activities to date, with a focus on the preliminary rollout of the library in December 2002. After discussion, the committee concentrated its advice on the accessibility of the library for all users and the centrality of learners overall; the need to understand the sophisticated and competitive work going on in publishing, education, and technology companies that relates to and impacts NSDL; possible alternative business and management models for the longer term; and the need to understand the special responsibility of a tax payer funded public initiative of the scope of NSDL to live up to its transformational educational mission.

C.4.2 Community-Based Governance

The NSDL Policy Committee represents the NSDL Assembly, which consists of all the NSDL-funded partners. The committee, which is selected annually, is the main instrument of community governance. The Policy Committee has created five standing committees: Community Services, Content, Educational Impact, Sustainability and Technology. A member of the CI team acts as liaison to each committee, so that policy and implementation are coordinated. This structure was established under the Interim

Governance Agreement approved at the December 2001 All Projects meeting. As the NSDL expands, we expect this structure to remain suitable, though additional members may be needed to represent teachers and non-NSF-funded developers.

The CI Central Office provides administrative support for the Policy Committee and its standing committees, including meeting arrangements, phone and travel support. With the new budget, the Central Office will continue such support and provide additional administrative assistance, especially with tools for communication and collaboration.

C.4.3 The Future Structure of the NSDL

Legally, the CI is a collaboration among several not-for profit organizations, some funded directly by the NSF and some as subcontractors. Currently, almost all funding comes from the NSF. Our long-term vision sees a combination of continued government funding, with substantial contributions from other sources including institutional subscription fees, sponsorship, and philanthropy.

Looking to the long term, the National Visiting Committee has asked us to look at various organizational models. The overall educational effectiveness of NSDL clearly depends upon the sustainability of its infrastructure, collections, and services. Is a separate non-for-profit organization the best way to achieve this goal? We plan to work with the committee on this topic, in particular, to study whether a separate organization would be better able to seek for other sources of funding. Our current thinking is that a separate organization would be effective, however we will rely very heavily on the NSF's judgment. In any event, the strong ties that we have to our respective universities and other organizations will not be jeopardized.

C.5 Timetable for Progress

From one perspective, the NSDL is a complex software project. As such, it has and will continue to be managed in accordance with established software engineering practices: requirements and design documents will be posted on the developers' portal for public review; new collections and services together with software upgrades will be released in batches, approximately every six months, with major versions every 18 months. However, precise technical milestones over the long term of the project are difficult to establish, because much of the technical progress responds to opportunities available through leveraging the work of other NSDL-funded projects as well as the digital library and web information community as whole.

Furthermore, the technical aspects of the NSDL are ancillary to the larger user-focused and educational tasks. In a distributed, community-engaged entity such as the NSDL, the introduction of technology is meaningless without a thorough understanding of the organizational context. Rather than a technology-driven process, we follow a needs-driven process, whereby the CI team engages communities of concern, understands the potential impact of the NSDL, and then evaluates available and necessary technologies to address their needs. This complicates the task of establishing technical milestones since the needs of the many NSDL communities are difficult to predict as they evolve over a five-year period. The remainder of this section, therefore, outlines a project timetable with a bias toward communities of focus rather than standalone technical milestones.

C.5.1 Version 1. The Initial Release

The first phase of the Core Integration work began in October 2001 and will end with the release of Version 1 of the NSDL in December 2002. During this phase the emphasis has been on building a robust foundation, both organizational and technical. The major parts of the organization are in place: the National Visiting Committee, the NSDL Assembly with its Policy Committee and standing committees, and the CI team with its central office and five operating divisions. The core technical architecture has been established and development of several major components will be completed by December: the metadata repository, search service, the main portal and several portal components, the authentication service and user profiles, the developers' portal, and an advanced help service. A broad program of evaluation is being created.

The limited scale of the collections and services in the initial release provides opportunity to evaluate the stability of the core technical components. More significantly, this phase will establish core user communities, most probably technically skilled educators, students, and content providers, who will help guide the planning for future releases.

C.5.2 Version 2. Large-scale Growth

Version 2 of the NSDL is scheduled for Summer 2004. By the end of this phase, the NSDL should be a very large, production-quality library, with substantial richness of educational resources, and a rapidly growing community of users and contributors. We will continue our community support and evaluation activities.

The emphasis of this version will be growth. The number and variety of collections will be greatly increased by the integration of all NSDL-funded collections, partnerships with publishers, and collaboration with other digital libraries and other resources providers. Partnerships in the educational and library communities will be established and leveraged to improve library usage and services, and further the goals of educational enhancement.

In parallel with this growth in resources, we plan a major emphasis on educational outreach, building partnerships with groups of educational users, including discipline-specific groups, sharing resources with them, providing services via the NSDL, understanding their needs and seeking for resources that meet them. One vehicle for expanding the base of users will be the development of more portals into the library, especially those targeted at the primary and secondary student community.

During this phase we plan to exploit the NSDL service framework in two directions. First, the CI team will create several new exemplar core services, notably a preservation service for NSDL collections and an authorization service for publishers. Second, we will work to spawn a vibrant community of service providers from the NSDL-funded projects and the wider commercial and research community. The growth of this community is one key to the growth of the functionality of the library.

An additional technical focus of this release will be the definition and preliminary deployment of a *service* architecture. Whereas the earlier release will define a limited number of core services, essentially hard-wired into the architecture, Version 2 will build on advances in web services [W3C 2002b] and deploy an NSDL infrastructure that

supports open integration of rich services. The effect will be a technical infrastructure that will support our broad range of target communities through focused services.

C.5.3 Future Releases

The earlier releases of the NSDL will establish the basis of community involvement, service infrastructure, and content corpora that can serve as the springboard for future growth. The challenge for future releases is to transition the NSDL from a digital library – defined strictly as an aggregation of resources and services to discover and provide access to those resources – to a rich learning environment.

Fundamental to this growth, in Version 3 and beyond, is an increased emphasis on creating communities of educators and learners, and feedback mechanisms to facilitate parent-teacher-student interaction. The growth will include more powerful and more varied services that serve larger and more diverse communities, with greater emphasis on novel types of resource for interactive experimentation, computation, visualization and other aspects of learning.

Two technical themes will dominate these future phases and enable this richer functionality. The first is automation, which is a central theme of all releases. To support an ever-expanding library with a fixed size of CI team, every activity must be automated to the greatest possible extent. Much collection development and most services, including end-user support, will be largely automated. The second theme is to develop a technical framework, for sharing and managing semantic information, to support the sharing of learning concepts across diverse communities and the creation of new learning communities within the NSDL rubric.

C.6 Recommendations Addressed in this Proposal

The NSF asked the NSDL/CI National Visiting Committee (February 2002) and a Reverse Site Visit panel (April 2002) to review the CI work and make recommendations thereon. This appendix lists the recommendations, with links to the sections of the proposal that address them. Additional comments are in bold-faced type.

C.6.1 National Visiting Committee Recommendations

Background and Overview

- Who are the intended users of the NSDL? **C.1.3**
- What is the sustainability of NSDL? **C.4.3**
- How big is this? Are there sufficient resources (funds and personnel) to handle the scope? **We believe that the budget in this proposal is adequate for the program proposed.**
- How can user support services be made scalable? **C.3.5.4**
- Is the budget adequate to let NSDL to be successful? How is this project budgeted compared to other ventures (e.g., commercial publisher)? **See the third bullet in this section.**
- What is the partnership strategy for NSDL? What might those partnerships be? What would the ground rules be? **C.2.2**

Educational Outreach

- How should the community be expanded beyond the "builders" to include the users? **C.3.3**
- How will the NSDL and NSF work with other groups (e.g., commercial entities) to broaden the project? How would those types of relationships work? **C.2.2, C.3.3, C.3.7, C.4.1**
- How does the project move beyond integration of the digital library and content components to ensure there is a significant educational impact? **C.3.3 plus the addition of Eastern Michigan University, the University of Colorado, and Syracuse University to the CI team**
- Who handles the educational impact? This isn't the core competency of CI team. In fact, is CI funded to do this? **See the preceding recommendation**
- How much of the ultimate educational value of NSDL is in the content? How much is in the tools? Can those tools be reusable? **C.3.3.2, C.3.4.3**
- For this to work, must NSF fund implementation projects? Demonstration projects? **(Note: this is not an explicit CI recommendation, but we would work with the NSF if it chooses to fund such projects.)**
- Should the CI team develop a network of collaborators that handle how to implement, train and use the material? What should the CI's role be in terms of educational impact? **C. 1.2, C.2.2**
- Syllabi and curricula should become part of the collections. Collect best practices about how to use things that are in the collection. This will improve the educational outreach. **C.3.3, C.3.5.1**

Technology Architecture

- There were a number of questions related to metadata and specific technology questions. **C.3.4**

Marketing and Outreach

- Need to know who the market is. **C.1, C.2.2, C.3.3, C.3.7**
- When naming the ultimate NSDL, should education be in the title? **We consistently are using "National Science Digital Library" plus a tag line that stresses education.**
- During the process of naming, remember that you'll need to communicate with many audiences (e.g., the academic community, NSF, teachers, kids). What would kids respond to? (Science Education Exchange?). May want to use a tag line to help define the name. **C.1.3**
- The domain name is important to consider. **We recently acquired nsdl.org.**
- Conclusion: can't really decide on a name right now. Simple is better than complex (i.e., science vs STEM). Need more work on what the product is and who the audience is. **C.1.3, C.3.3**
- Other companies are aggregating content. What is different here? Look for what the breakthrough areas are (e.g., unusual tools). (Remember, the CI team really isn't generating a product; they work with the products that come from the 64 other projects.) **C.1, C.3.7**

Intellectual Property

- The proposed profile service (based on Shibboleth) works fine for HE but there is really no model for K-12. Who vouches for the 2nd grade teacher? **C.3.4.3, C.3.7, but not fully addressed in this proposal.**
- What is the value add of the proposed mechanism? **Not addressed in this proposal.**
- What is the right solution? Is it Shibboleth? Or are certificates a better answer? Is Shibboleth a short/mid-term solution? So how do you define mid-term? **C.3.4**
- How do you figure out "prestige accounting"? Can you tell authors that X people at higher education institutions are accessing your resources? Sort of like a citations index. Use for promotion/tenure. Note that measuring access as a proxy for quality is dangerous. **C.3.4.3, C.3.7**
- How do you motivate the people who have the best stuff to put it in the NSDL? **C.3.3, C.3.4.3, C.3.6, C.3.7**

Sustainability

- Do we know what it costs to run the organization, steady-state? **C.3.4.2, C.3.5**
- Is it premature to think about going independent? Perhaps NSF should commit to fund this until it is "hatched". Although it may be premature to decide, it is not premature to talk about it. **C.4.4**
- What is the value proposition? **C.1, C.3.7**
- What is the product? **C.1, C.3.7**
- Is an option to fund like the American Memory project? Federal funding that is contingent on X among external support. **C.3.7, C.4.4**
- NSDL doesn't need to think about this as a monolithic business model; there could be several revenue streams. **C.3.7, C.4.4**
- Talk about a beta test in a year. Involve the various publics you think you might want to serve. Establishing December 2002 as a beta test is consistent with NSF's language of operationalizing the project. Note: A beta test is a limited access test but not limited functionality test. **C.1.3, C.2.1, C.3.3.2, C.3.5, C.5.1**
- What are the principles that could make this an ongoing venture rather than one that stops? State the rules that would have to be in effect. Communicate those to the other projects (e.g., you can't withdraw your content). **C.3.5, C.3.6, C.4.3**
- Is an option to have various institutions host portions of NSDL? **This proposal manifests distributed hosting.**
- What about the international dimension? **Not addressed in this proposal.**
- Consider a range of markets (e.g., AP, SmartForce, Saba, teacher preparation institutions). **C.3.3.6, C.3.3.7, C.4.1**
- Note that it doesn't sound as though most of the projects have committed to develop any specific product on any specific schedule. Will they produce something that goes into NSDL? The CI team is operating as though they have line responsibility to the NSF to create a viable library. Need to ensure there is the same kind of feeling of responsibility into the other projects. This should be NSF's responsibility. **(Note: this is not a CI responsibility, but we support the recommendation.)**
- What about the teacher preparation institutions? **C.3.3.5, C.3.3.6**

Management and Governance **(Note: these are not CI recommendations.)**

- An option for the policy committee is to invite members of the NVC to the "all hands" meeting then they can see and hear from NVC. **Done.**
- Consider inviting the policy committee (or a representative) make a report to the NVC. **Planned.**

General Questions

- What are the potential benefits to people to provide their material to the collection? What's in it for them? **C.3.3, C.3.4.3, C.3.6, C.3.7**
- What is the NSDL? Is it a "Google optimized for science education"? Is it a collection? Is it content? Is it tools? **C.1.1, C.1.2, C.3.3**

C.6.2 Reverse Site Visit Recommendations

1) Core team needs to take a stronger leadership role in driving NSDL as a program among opinion leaders in distribution and creation of knowledge (scientists, educators, publishers). **C.2.2, C.3.3, C.3.7, C.4.1**
2) Management role of the team is made difficult by distributed governance role—should use recommendations of NSF to take a stronger role. **C.4**
3) Need to clarify the distributed management structure among three institutions—an organizational chart that clearly reflects responsibilities and specific personnel is needed. **C.3, especially C.3.1**
4) Stronger focus on education, educators, learners—must get teaching/learning perspective on the management team. How is NSDL going to transform education? Work with the educational community now to create the vision for the educational breakthrough to occur in 5 years. **C.3.3, plus the addition of Eastern Michigan University, the University of Colorado, and Syracuse University to the CI team**

   If K-12 is the focus of the strategic roll-out, how do you address that audience of teachers, media specialists and learners? Need to relate more strongly to other that focus on education. (Since you are focusing on K-12 teachers, which represent an adult audience, then you should not dismiss the young adult audience of undergraduate students.) **C.3.3.2, C.3.3.3, C.3.3.5**
5) Collections issue—how will you establish what goes in and what does not go in? The NSDL stamp of approval should mean high quality. Need to establish editorial boards with scientists and educators as reviewers. **C.2.2, C.3.5.1**
6) How do you interact with other players at the national level who are working on digital libraries? **C.3.4, C.3.6, C.3.7, C.4.1**
7) Sustainability is a critical issue—concentrate serious effort on getting expertise needed to develop several alternative business models and to better understand the NSDL value proposition. **C.3.7, C.4.1, C.4.3**
8) Must form better ties with professional science organizations and professional educator societies. **C.3.3, C.4.1**
9) Take the time to define how the NSDL is different and better for educators that Google. **C.1.1, C.1.2, C.3.3**
10) Have you thought about the impact that the constant web updates and plug-ins available to be downloaded over the Internet will have on the NSDL? **C.3.4.3**
11) What is the defined set of deliverables for the NSDL all-projects meeting in December? **C.2.1, C.5.1**

12) We found the NSDL NVC report to be very thorough, and we support in principle all the questions they asked. Your concerted effort in answering those questions will produce a strong build-out proposal. **See above.**

C.7 Results of Prior NSF Support

Following the NSF letter of invitation to submit, the current proposal extends "the capabilities and functions of the NSDL's core integration capacity following the trajectory that has been laid out in work to date." For all PIs and several Co-PIs, the most relevant prior support is the collaborative award to UCAR, Cornell University, and Columbia University described (first) below. Separate results are shown for Co-PIs Hoffman, Lagoze, Marlino, Smith, Sumner, and several of our senior personnel.

Award #: EHR/DUE-0127298, 0127308, 0127520 *Dates:* 01-Oct-01 to 31-Dec-02
Title: **Core Integration of the National SMETE Digital Library**
Investigators: UCAR – David W Fulker (Principal), James M Allan, Cathryn A Manduca, Mary R Marlino, Terence R Smith
Cornell University – William Arms (Principal), Dean Krafft, Carl Lagoze, John Saylor, Sarah Thomas
Columbia University – Kate Wittenberg (Principal), David Millman
Amount: $5,000,000

As precursor to the current proposal, this project began the operational integration of NSDL. Primary categories of effort are: engaging the community, providing technology, and operating core services. As discussed with the National Visiting Committee and an NSF Committee of visitors, the project is on track, with an operational release of NSDL scheduled for late this fall. Though the project is new, the results already include:

- Prototype Core Integration specifications and components, some of which have served their purpose as demonstrations, while others are being refined for incorporation into the first operational release of NSDL.
  - Metadata standards for collections, with a corresponding metadata handbook.
  - Source-Forge-based communications portal, with workspaces for a variety of NSDL-construction activities.
  - Harvester- and harvested-side implementations of the OAI protocol spec.
  - Preliminary, operating version of the Metadata Repository (MR), populated with a small number of records from NSDL collections.
  - Collections development, as manifest in the MR contents.
  - Preliminary implementation of a search service built on the MR.
  - Preliminary definition of the NSDL architecture (published by Lagoze), plus progress toward a long-term definition with a *formal framework for services*.
- All Projects Conference (December 2001), involving 60+ research teams, and yielding a community-based governance framework for NSDL.
- Various panel discussions and invited conference presentations, plus one instance of testimony before an NRC advisory committee.

Selected Publications:

[Arms 2002], [Hillmann 2001], [Janée 2002], [Lagoze 2002], [Wittenberg 2001]

Award #: IIS-9817432 *Dates:* 01-Sep-99 to 31-Aug-04
Title: **Alexandria Digital Earth Modeling System (ADEPT)**
Investigators: Terence R. Smith (Principal), James Frew, Reagan Moore, Mike Goodchild, Richard Mayer, Christine Borgman, Divyakant Agrawal
Amount: $5,400,000

ADEPT is a distributed set of collections and services that support construction and use of *personalized* digital libraries containing geospatially referenced information. The prototype has been useful to undergraduate teachers and students, offering: multimedia information for education in physical, human, and cultural geography; user interfaces that employ the "Digital Earth" concept; and services for creating and managing collections. The Alexandria Digital Library (an NSF-funded digital library, by several of the same investigators) forms much of the foundation for ADEPT.

Selected Publications:

Gilliland-Swetland, A.J. and G. Leazer, "Iscapes: Digital Libraries Environments for the Promotion of Scientific Thinking by Undergraduates in Geography," Proceedings, First Joint Conference on Digital Libraries (JCDL '01), Roanoke, VA, June 24-28, 2001.

Hill, L., R. Dolin, M.A. Rae, L. Carver, J. Frew, and M. Larsgaard, "Alexandria Digital Library: User Evaluation Studies and System Design," Journal of the American Society for Information Science (Special issue on Digital Libraries), 2000.

Smith, T.R., G. Janée, J. Frew, and A. Coleman, "The Alexandria Digital Earth Prototype System," Proceedings, Joint Conference on Digital Libraries (JCDL '01), Roanoke, June 2001.

Award #: #GEO/DUE-0085600 *Dates:* 01-Oct-00 to 30-Sep-02
Title: **Digital Library for Earth System Education: Implementing the DLESE Community Plan**
Investigators: Mary Marlino (Principal), David Fulker, Cathryn Manduca, Dave Mogk, Tamara Sumner
Amount: $3,855,876

Building on the Geoscience Digital Library project, this focuses on DLESE infrastructure and developing the library community. A key result was the release of Version 1.0 in August 2001, which now has some 2000 carefully selected educational resources. It has features for searching by grade level, resource type, and keyword. Formative evaluation activities informed the design, including workplace interviews with educators, focus groups, formal usability studies, and ethnographic observations of users engaged in library use and design activities. The library also contains a web-based Collection Builder tool, and community oriented services such as discussion forums for working groups and posting tools for Geoscience education opportunities and announcements.

Selected Publications:

[Marlino 2002]

Marlino, M., T.R. Sumner, D. Fulker, C. Manduca and D. Mogk. The Digital Library for Earth System Education: Building Community, Building the Library. *Communications of the ACM, Special Issue on Digital Libraries (May 2001)*, pp: 80-81.

Sumner, Tamara, Melissa Dawe, and Holly Devaul (2002). "Creating Reusable Educational Components: Lessons from DLESE." *Journal of Geoscience Education* 50(1): 25-30.

Sumner, Tamara, Mary Marlino, and Terence Smith (2001). *Task-oriented digital library interactions: An end-user perspective on interoperability*. Proceedings of the Third DELOS Network of Excellence Workshop on Interoperability and Mediation in Heterogeneous Digital Libraries, Darmstadt, Germany, 8-9 September.

Sumner, Tamara and Melissa Dawe (2001). *Looking at Digital Library Usability from a Reuse Perspective*. ACM/IEEE Joint Conference on Digital Libraries (JCDL '01), Roanoke, Virginia (June 24 -28), pp: 416-425.

Award #: EHR/DUE-0085866 *Dates:* 01-Aug-00 to 31-Jul-02
Title: **TeacherLIB—Digital Community and Collections for Science and Mathematics Teacher Education**
Investigators: Eastern Michigan University – Ellen Hoffman
Merit Network, Inc. – Marcia Mardis
Amount: $799,864

The TeacherLIB project has reviewed issues and barriers to deploying NSDL in the K-12 arena, including potential impacts re economics, policy, and school systems and culture. Additional TeacherLIB activities and collections address K-12 deployment issues in communities, including legacy collections with quality content not originally designed for digital libraries. Hoffman has been an active participant in the NSDL community, contributing to the overall development of NSDL. The project maintains a Web site with information at http://www.teacherlib.org/.

Publications:

[Hoffman 2002], [Hunter 2001]

Hoffman, E. (2002). Can research improve technology planning? Paper presented at American Education Research Association Annual Meeting 2002. ERIC Clearinghouse in Assessment and Evaluation (in press).

Mardis, M. and E. Hoffman (2002). The National STEM Digital Library: Process, promise, progress. Proceedings of the 2002 Society for Information Technology in Teacher Education (SITE) Annual Meeting, pp. 2321-2323.

Hoffman, E. (2001). Building a national science digital library: Challenges and possibilities. Proceedings of the 16th Annual Conference on Infrastructure for e-Business, e-Education, and e-Science. L'Aquila, Italy, Scuola Superiore G. Reiss Romoli.

Manduca, C. and E. Hoffman (editors) (2001). NSDL Governance. NSDL Governance Working Group White Paper. Available online at http://comm.nsdlib.org/

Mardis, M. (2001). Uncovering the hidden Web I: Finding what the search engines don't. ERIC Clearinghouse on Information & Technology Digest EDO-IR-2001-02.

Mardis, M. (2001). Uncovering the hidden Web II: Resources for the classroom. ERIC Clearinghouse on Information & Technology Digest EDO-IR-2001-03.