
The OCKHAM Library Network

Integrating the NSDL into Traditional Library Services

Introduction

Emory University, the University of Arizona, Virginia Tech, and the University of Notre Dame propose a two-year project to improve usage and access to the NSDL by learning communities through the existing national infrastructure of traditional libraries.

Statement of Need

The NSDL has assembled an extraordinary set of resources and services in support of scientific education. However, there has been limited integration and deployment of NSDL services and content into the existing service programs and infrastructure of the traditional library community.

We feel the opportunity exists to explore how the traditional library community can facilitate and expand access to the content and services of the NSDL. Traditional libraries are prominent and active channels of information dissemination in virtually every community in the United States. Libraries provide a range of familiar and stable services that add value to information. Learning communities would realize many benefits from a coordinated set of networked services for dissemination of NSDL resources through traditional library protocols.

Target Audiences

- 1) **Students and independent learners** who could use the content of the NSDL as it will be disseminated through library services enabled by this project.
- 2) **Professors and other teachers** served by libraries, who might potentially wish to incorporate NSDL content so it can be accessed and organized in the context of local information infrastructures into their teaching activities.
- 3) **Librarians**, who will be able to integrate and deploy NSDL collections through local library operations using the project's network of modular web services and reference models.

We believe that all users served by the library community will benefit from improved access to the NSDL through the activities of this project. Our goal is to leverage the resources and services developed via the NSDL through the infrastructure of traditional library operations. By targeting the users of libraries as a whole, we hope to provide significantly greater access and use of the NSDL, as well as new avenues of interoperability and sustainability for NSDL systems.

Background

The OCKHAM Initiative has been sponsored by the Digital Library Federation to analyze the ways that modular component-based digital library architectures, reference models of services, lightweight protocols such as the OAI-PMH, and open standards for data exchange can be used to enable better interoperability and affordability of digital library services. The initiative is termed OCKHAM intentionally, to make a statement about simplicity and the need to not replicate new technical standards unnecessarily, but rather to look for creative ways to re-use existing standards and systems wherever possible. We believe that our methodology can fruitfully be utilized to enable wider deployment and utilization of the NSDL through existing library operations.

Project Goals, Outcomes, and Deliverables

The project has four broad goals with associated outcomes and deliverables. In this context, a deliverable is a specific product of project work, and an outcome is the measurable impact of project activities on our target audience groups. Details of the evaluation plan for outcomes are provided in the section on Project Evaluation

Goal 1: Reference Model Development. The project team will study existing library services and the NSDL intensively, and develop a clearly articulated and generalized reference model for deploying and integrating the NSDL into existing library service offerings and systems. The project team will document how existing and emerging lightweight protocols can be used for wider deployment and integration of the NSDL resources, and will identify gaps where new protocols are needed. *Deliverable:* Reference Model Document. *Outcome:* The reference model will primarily impact information providers seeking to deploy the NSDL in libraries. This impact will be assessed primarily through qualitative feedback during project team meetings and focus groups conducted at both DLF and CNI meetings.

Goal 2: Middleware and Testbed Services Development. The project team will design and implement a set of testbed services that will enable a greatly enhanced ability to deploy the NSDL through traditional library service programs, and networked middleware that coordinates discovery and interoperation of the testbed services. *Deliverables:* A package of network middleware and associated set of testbed services, described below. *Outcome:* The network of testbed services will primarily impact faculty and students – to be measured qualitatively by focus groups of faculty and students, and quantitatively through surveys and the project log analysis system.

Goal 3: Evaluation. The project team will assess the utilization and benefits of the testbed services through specific usage metrics, usability studies, and focus groups. *Deliverables:* Focus group summary findings, project team analyses, survey results, and log analysis. *Outcome:* The project evaluation activities will primarily impact information providers seeking to understand how the OCKHAM project can make a difference for library NSDL-related services. The evaluation products will themselves be evaluated in the final project report by means of focus group discussions at DLF and/or CNI meetings.

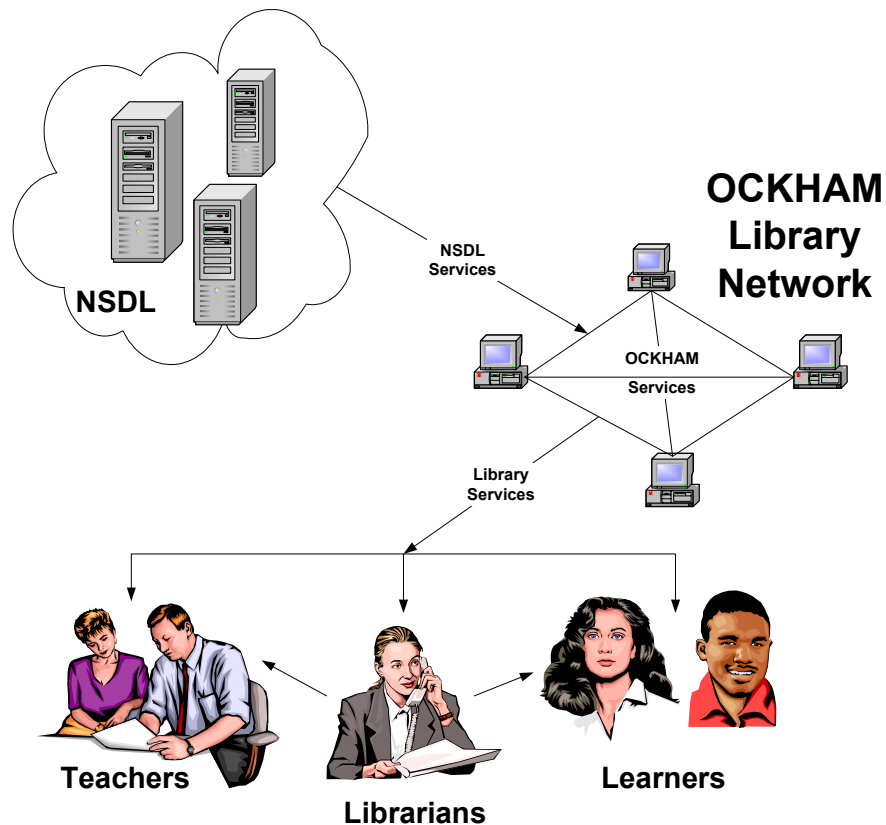
Goal 4: Dissemination and Networking. The project team will disseminate project results, and solicit additional participants in the OCKHAM Network to build up access to NSDL content through traditional library services. *Deliverables:* Published articles concerning the project results, conference/workshop/informal presentations, and additional institutional participants in the OCKHAM Network. *Outcome:* The number and prominence of articles and conference presentations will form one metric of project impacts in the field. Another more important metric will

be how many institutions decide to join the OCKHAM network during its development, and how many actually bring up OCKHAM server nodes in the OCKHAM P2P environment.

Project Design

This project will establish a reference model and functioning network of testbed services enabling traditional libraries to provide access to the NSDL through their local service programs. The network will initially include the four collaborating institutions of the project, and subsequently be expanded among a growing group of institutional partners, especially aided by DLF and OCLC.

Figure 1: Overview of the OCKHAM Framework



Several key design approaches form the foundation of this project, and make this group of collaborating institutions well situated to carry it out: A) our concept of *reference model* development as a means of guiding distributed development efforts during and after the project; B) our approach to *minimization of barriers to adoption*, including reduction of capital outlay barriers and other entry costs for new institutions entering the OCKHAM network; C) our *networked middleware design*, the fundamental architecture of our network; D) a core set of practical *testbed services* that will form the core NSDL utilities of the network and attract additional traditional libraries to join the network; E) Project phases for organizing this work; and F) a plan for *deployment at collaborating institutions*. Additional relevant information about the established working relationships of our group, complementary roles, and collaborative responsibilities of the project staff and institutions are provided later, in the section on Key Staff.

Reference Models

We define a *reference model* to be a document that provides a systematic understanding of relationships among the entities within an interacting environment, supplies a consistent vocabulary for discussions between potential collaborators, and supports the development of standardized protocols and specifications to enhance that environment. Reference models have been demonstrated to be a very effective means for mobilizing collaborative action in recent years, and the OCKHAM Initiative has embraced the concept for its work. An example of a successful reference model is the OAIS Reference Model, developed by NASA, and adopted by many other categories of organizations. [OAIS, 1999] Another example is the technical framework Open Archives Initiative Protocol for Metadata Harvesting. [Lagoze, 2001] Of the two examples, the OAI-PMH framework is slightly preferred, in that it provides a specification for a lightweight protocol that can easily be implemented in many modular contexts. The reference model that the OCKHAM project would create would be modeled after these examples, with specific focus on deploying and integrating NSDL collections into traditional library operations. The reference model will have sections covering scope, terminology, revision history, entity relationships, protocol specifications, and implementation discussion. The reference model also will include an explicit process for evolving the reference model over time through collaborative group activity.

Minimization of Barriers to Adoption

Various strategies for minimizing barriers to adoption have proven amazingly effective during the past decade in enabling open source software systems to rapidly spread and become widely adopted. Software such as the Linux operating system and the Apache web server have come to dominate a large proportion of learning communities because they are inexpensive (usually freely distributed), require minimal hardware configurations, and can be collaboratively developed as packages that are simple to install and maintain. [Morgan, 2002]

If it is to be successful, we believe that the OCKHAM Network must utilize all of these strategies to lower the barriers to adoption. Capital outlays and expertise required to bring up a node on the network must be essentially trivial. The fundamental nodes comprising the OCKHAM network could be low-cost Linux servers, each with ***a standardized package of easily installed open source software*** that provides access to all of the network's NSDL-to-library testbed services. Special attention will be devoted to making the OCKHAM server package simple to implement and maintain, with minimal hardware requirements. All of the project principals have experience in systems integration and design of interoperable, low-cost library software packages, gained during a variety of projects ranging from the Open Digital Library (ODL) project [Suleman, 2003] to the LOCKSS project. [LOCKSS, 2001]

OCKHAM Networked Middleware Design

The OCKHAM Initiative has been studying approaches for creating effective networks of federated digital library services interoperable with existing library systems. The strategy that we believe is most effective is establishing an easily adopted framework for discovering and accessing services provided by entities on the network. We will accomplish this through a peer-to-peer network of services capable of communicating with user interface modules in selected learning applications:

Peer-to-Peer (P2P) Networked Middleware: A peer-to-peer (P2P) network is an architecture in which all nodes (i.e., peers) have equivalent capabilities and responsibilities. In the OCKHAM P2P network, each participating institution (library, academic institution, or NSDL project) is a peer site. Each site would have the ability to a) provide advertisements describing the services that site is making available, and b) be able to search the network for services advertised by other peer sites.

Peers on a P2P network locate advertisements by querying other known peers on the network; in turn, those known peers pass on the query to additional peers that they know (but may have been unknown to the peer originating the query). In this manner the query traverses the P2P network to collect all relevant answers. This type of approach has been proven effective in music-sharing communities. It is our belief that this approach promotes service-sharing.

The most common method of providing discovery and access information about services is through a registration service. In the business world, an emerging registration standard is the Universal Description, Discovery, and Integration (UDDI) protocol. The core integration project of the NSDL provides a registry service for NSDL collections. The difference between these approaches and a peer-to-peer approach is that in the P2P model, the act of registering available services is managed and maintained locally, instead of centrally. Each peer site is in effect its own registrar – the P2P functionality then brings together the various locally registered services to appear as a virtually centralized registry. This approach has the potential to reduce cost barriers for discovering available services. It also can reduce the cost of registering and maintaining an accurate list of available services, because these are things that now can be done locally for each peer site.

We propose building a P2P network for the purpose of providing the ability for peer sites to advertise (i.e., register) their available services, and to provide the ability for available services to be discovered via searching the advertisements (i.e., service metadata records). This is described further under the 'interoperation service' in the testbed services section. There are currently a number of P2P frameworks that are freely available (among them Gnutella, Freenet, and JXTA). [Wilson, 2002] We will examine the current set of P2P frameworks for a 'best fit' for the OCKHAM P2P network, and adopt the appropriate framework for specific use with the NSDL and library services.

Testbed Services: Our testbed services will utilize lightweight protocols. Wherever possible, we make these available through web services. The web services approach to enterprise-to-enterprise communication through XML, SOAP, and WSDL standards promises to be a key means of interoperability in coming years. Web services build on the success of XML as an open data exchange standard, and will comprise the main communication meta-standard for our network. Thus, current library systems can be augmented to utilize and interoperate with the NSDL's broad range of quality services and content.

Because they are based on lightweight protocols, OCKHAM web services can be incorporated within existing portal applications relatively easily. We will implement end-user interfaces to our testbed services in a variety of portal environments at the four collaborating institutions. Our primary focus will be on the popular library portal software MyLibrary, which one of the senior personnel (Morgan) developed, although we also will explore development of end-user interfaces to some of the testbed services for the Blackboard courseware system, and the uPortal campus portal software. MyLibrary is a user-centered database application with a Web front-end. [Morgan, 2000] It turns the description of information on its head by describing information resources in terms of users instead of the materials themselves. A MyLibrary end-user interface will be developed for all OCKHAM testbed services. By combining the customizable features of MyLibrary with the rich content of the NSDL, learners will be able to access and use NSDL content in the context of established library service programs. Based on the content of the MyLibrary database the system can generate static and/or dynamic sets of HTML pages for the purposes of creating a library's: home page, pathfinders, directories of resources, and searchable indexes. The content can be syndicated to external applications such as campus-wide information systems or news feeds. This is accomplished by saving metadata in the database and repurposing it for various venues. The creation of customizable interfaces to the collection, portals, is one of the more popular features of the system. This is done by gathering rudimentary information about individuals, displaying prescribed sets of resources for individuals with particular characteristics,

and allowing individuals to refine the prescriptions to meet their specific needs. Consequently the system allows patrons to create their own personalized library. MyLibrary is written in Perl and distributed as open source software under the GNU Public License. It is in production in more than two-dozen major libraries across the world, and has been used as a model for other "MyLibrary" services in many places such as the Los Alamos National Laboratory and the University of Michigan. For more information about the MyLibrary software, see <http://dewey.library.nd.edu/mylibrary>.

Testbed Services

Analytic work to date by the OCKHAM Initiative on needed functionality for libraries seeking to integrate NSDL collections leads us to propose the creation of a set of initial testbed services during this project, summarized as follows, with lead development institutions indicated in each case:

- 1) **Interoperation Service (Arizona)**: This is the primary service for discovering and coordinating access to all other services made available on the OCKHAM P2P network. Users (in this case, NSDL projects, academic institutions, and libraries) will be able to locally register their available services for advertisement on the P2P network. Additionally, users will discover the breadth of available services, and obtain the necessary information to interoperate with those services. Initially, we will develop an application that can be easily downloaded and deployed so users can easily manage their service advertisements, as well as search and discover other existing services. This application will utilize a standard metadata schema for service advertisements.
- 2) **OAI-PMH-to-Z39.50 Searching Service (Arizona)**: This service will provide the capability to search OAI-PMH compatible repositories through the ZNG/Z39.50 protocol. Many libraries currently utilize the Z39.50 protocol to provide a centralized method of searching across a distributed set of collections; most libraries do not currently have a method of providing search mechanisms into standard OAI-PMH compliant repositories. By providing a service that will allow standard OAI-PMH compliant provider systems to be natively searchable via ZNG/Z39.50, many more libraries will be able to incorporate NSDL collections into their own portals and services. In order to build this service, we will use the Yaz toolkit (<http://www.indexdata.dk/yaz/>), an open source toolkit available for creating Z39.50 clients and servers. We also will make use of Zebra, an open source structured text indexing and retrieval engine (<http://www.indexdata.dk/zebra/>). Zebra will enable us to take the harvested metadata from a repository, index that metadata for searching purposes, and then connect to Yaz for Z39.50 access.
- 3) **Alerting Service (Notre Dame)**: This testbed service will proactively alert learners, scholars, and educators of the existence of new materials in the NSDL. The alerting service will work by first gathering information about an individual's learning needs and interests, saved as one (or more) user profiles. On a regular basis these profiles will be applied (searched) against sets of newly available content metadata exposed via the OAI-PMH. Some of this exposed metadata will come from NSDL services, and some will come from metadata aggregation services. Some of this content will come from simple lists of newly acquired materials from various library catalogs. Finally, the results of these searches will be provided to end-users via portals such as MyLibrary. The lightweight protocol for this service will provide the capability for the alerts to be transmitted to a user's portal account or optionally sent to the individual via email. We will be able to proactively alert learners, scholars, and educators of the existence of new materials and make the process of finding and getting information from digital libraries easier and more transparent.
- 4) **Browsing Service (Notre Dame)**: This testbed service would provide end-users with the ability to find resources similar to known items within the NSDL. The browsing or "find similar" service will facilitate serendipitous discovery of related content. A common approach to finding information in libraries is to first identify a known item of interest, and locate other items similar to the known item – often by specifying one or more pieces of metadata from the item of interest and finding other materials with the same characteristics. For example, an item might be of interest because of its

author and keywords. We can find other materials written by the same author and containing the same keywords. Further, the browsing service will allow users to choose an item of interest and then prompt the system to find other similar items. The system would then return a list of metadata values used to describe the item of interest, asking the user to select the qualities they prefer for browsing. Based on this input, the system will return sets of other items with the selected characteristics and present them in ordered lists for browsing. Ideally, the browsing service might recommend information resources based on previous usage patterns. We will experiment with algorithms for organizing browse responses using both source repositories and Dublin Core metadata fields.

- 5) **Conversion Service (Emory):** This testbed service would function solely to automate the conversion of data between different formats used by the P2P network. The service would have an associated protocol simply specifying a particular conversion among those registered on the server. Such ad hoc conversion scripts would specify unambiguously an incoming format stream and a desired response stream, much as an XSLT process for XML, but additionally operating on MARC transport format records. This would provide extensible functionality for converting between existing and ad hoc data formats utilized by the network during the project. The service would at a minimum be designed to convert unqualified Dublin Core records to MARC transport records, enabling the records from OAI-PMH providers to be converted for use in library catalogs. Many such conversion scripts would be stored, accessible to other OCKHAM nodes, and “tweakable” for the idiosyncrasies of particular library online catalog systems or library portals (such as MyLibrary). The service would be used for both the cataloging and pathfinding testbed services.
- 6) **Cataloging Service (Emory):** This testbed service would provide OCKHAM servers in the P2P network with convenient automated means of generating, transmitting, and editing cataloging records for NSDL resources. The service would utilize both the conversion and OAI-to-Z39.50 testbed services of the project. The cataloging service will be built on a lightweight protocol for the following operations: A) automatically generating and temporarily saving a repository of either MODS or MARC format transport records from either the OAI DC records in an OAI-PMH provider or the pages crawled from a website, B) overwriting saved records with edited records, and C) transmitting saved records to a specified receiving URL destination. Operations could be invoked on an entire saved repository, a set of records in a saved repository, or a single record in a saved repository. Depending on the request, the Cataloging Service would generate either a single collection-level record for the entire target site or item-level records for the individual DC records or pages crawled. An associated web interface will enable catalogers to edit the temporary repositories of records created by this service on the local OCKHAM server, and apply classification to trees of records in such repositories, using the Catalogers Desktop software from the Library of Congress. Interoperation of this service with OCLC’s Connexion service will be explored.
- 7) **Pathfinding Service (Emory):** This testbed service would provide a means of generating, transmitting, and editing standardized “pathfinders” for NSDL resources. Pathfinders are library guides created by subject experts for learners who are seeking to use library resources for specialized topic research. Pathfinders produced by subject experts knowledgeable about highly technical scientific topics are relatively rare and labor-intensive to produce, and are therefore a highly desirable commodity worth sharing among libraries for the benefit of their learning communities. This testbed service would have an associated lightweight protocol providing a means of rapidly distributing and discovering such specialized guides between libraries in the P2P network. The service would specify a variety of interchangeable encoding formats for pathfinders, and enable the automated conversion from one format to another via the conversion testbed service. This will enable libraries to collaboratively share pathfinders interchangeably, without imposing a unified format for all pathfinders on all participating institutions. The pathfinding service will be important for building up a body of reference and training materials for library users seeking to use NSDL resources.

Deployment at Collaborating Institutions

Access to NSDL content will be deployed through the library services of the four core institutions (Emory, Arizona, Notre Dame, and Virginia Tech) using the services of the OCKHAM Network. All four institutions would deploy the interoperation service. Emory will focus on deploying the cataloging, pathfinding, and conversion services in regular library service operations. Arizona will focus on the Z39.50-OAI-PMH searching service. Notre Dame will focus on the Alerting and Browsing services. In addition, each institution will deploy and test as many of the other services as possible. Deployments of the testbed services will conclude with guides prepared by the relevant institution on successes, failures, and implementation notes for other institutions seeking to use the OCKHAM Network for these respective services.

Project Phases

The project would have three overlapping phases: A) analysis and documentation, B) testbed development and deployment, and C) assessment and dissemination.

The Analysis Phase will focus on the development of the reference model, and will include focus groups with librarians and users. Request for Comments (RFCs) and open meetings held at meetings of the Digital Library Federation and other professional groups will allow public vetting of draft reference models and protocol recommendations. The NSDL Core Integration team will be engaged in consultation for guidance regarding the underlying NSDL architecture. We will begin this analysis informed by previous theoretical work undertaken at Virginia Tech on the 5S model of digital library functions, but with the intent of carefully studying NSDL applications from the perspective of experienced librarians and the NSDL Core Integration team.

The Development Phase will focus on the creation of the networked middleware and testbed services. This phase will begin in any of its parallel parts when enough guidance is accumulated from the analysis phase to effectively begin that portion of testbed development. Testbed programming efforts will be undertaken at the four primary collaborating institutions, Emory, Notre Dame, Arizona, and Virginia Tech. The Virginia Tech effort will build upon our work with the Open Digital Library (ODL) software, engaging advanced graduate students. We will create both a network middleware package and a sampling of services. This will benefit both our individual campus communities and the wider academic community through dissemination of results and inter-campus service offerings.

The Assessment Phase will measure the utility, usage, and impacts of the local library testbed services in the participating campus communities by means of the evaluation plan, and will include analysis of web logs, focus groups, and usability studies. Results of the analysis and testbed efforts will be disseminated through presentations and publications in public meetings such as the DLF Forum, CNI Task Force Meetings, ARL symposia, the JCDL, and the ECDL conferences.

Key Staff - Roles and Responsibilities

The vision for the OCKHAM Network is the result of a yearlong planning effort by a group of librarians and computer scientists. The core project team consists of four individuals representing complementary institutions and backgrounds, with a strong personal commitment to working together to apply OCKHAM concepts to the NSDL for the benefit of learning communities. Following are summaries of the roles, responsibilities, and qualifications of the core team.

Martin Halbert (Director for Library Systems, Emory University). Halbert will serve as the lead PI, with responsibility for overall project coordination. He will organize and chair project meetings, which will lead to development of the reference model. He will serve as editor responsible for producing and revising project documents. Halbert will personally supervise two programmers, and will be responsible for design, implementation, and testing of three of the testbed web services. He will be responsible for project management tasks and NSF reporting. He will have responsibility for coordinating dissemination and network expansion activities in the traditional library community. Halbert is currently principal investigator on two of the Mellon metadata harvesting initiative projects, and serves as executive director of the MetaScholar Initiative, a consortium of thirty institutions working to aggregate metadata for scholarly portal services. Halbert serves as the chair of the LOCKSS sub-committee on Institutional Access Integration (see Supplementary Documents), and has there studied issues of low-cost library server networks and associated integration issues. He has served as editor of several library publication projects, and currently supervises a university library division of fourteen professional staff.

Ed Fox (Director of the Virginia Tech Digital Library Research Laboratory and the CITIDEL Project). Fox will serve as a Co-PI, with responsibility for project evaluation and systems architecture issues. He will supervise a graduate research assistant undertaking the project evaluation activities. Fox will be the lead technical advisor on the project. He has worked extensively with Halbert on the MetaScholar metadata project. He also will have responsibility for coordinating dissemination activities in the digital library research community. Professor Fox has served as PI or co-PI on over 80 research grants. He led one of the first NSF-funded DL projects in 1991. He and Gail McMillan co-founded the Electronic Thesis and Dissertation (ETD) project at VT, which has since grown to national and international dimensions. He has directed the Digital Library Research Laboratory at VT since 1998, and is Director of CITIDEL, the Computing and Information Technology Interactive Digital Educational Library, part of the National Science Digital Laboratory (NSDL). He is the Chairman of the Policy Committee of NSDL, and is co-PI for the GetSmart (<http://collab.dlib.vt.edu/runwiki/wiki.pl?GetSmart>), NSDL project.

Jeremy Frumkin (Metadata Systems Librarian, University of Arizona). Frumkin will serve as a Co-PI, and will have responsibility for implementing the core P2P interoperation service, as well as the OAI-PMH-to-Z39.50 search service. He will supervise one programmer. Frumkin currently works with implementing and building digital library tools and services, and serves as the chair of the Open eBook Forum's Metadata and Identifier's Working Group. He is currently exploring the concept of peer-to-peer networks as related to library functions in the metadata3 project (<http://md3.org>), funded through a small grant given by the University of Arizona Library. Frumkin is also the technical project manager for continued development of POLIS (<http://polis.arizona.edu>), the University of Arizona's homegrown courseware system.

Eric Lease Morgan (Head of the Digital Access and Information Architecture Department, University of Notre Dame). Morgan will serve as project senior personnel, and will be responsible for developing user interface modules for the project. He will be responsible for designing, overseeing implementation, and testing several of the OCKHAM testbed services described below, with special attention to integration with the MyLibrary software. He has been a long-time member of the LITA Top Library Technology Trends group, and is noted for his Infomotions consulting service. Morgan developed the MyLibrary software, which has been implemented as a library portal service at many major academic libraries nationally.

Additional consultants and involved parties include:

David Seaman (Director, Digital Library Federation). Seaman will oversee DLF involvement with the OCKHAM project, providing a senior coordinating role with members of the DLF community, the premier research libraries in the country. By providing venues for project presentations, Seaman will assist with the dissemination and networking aspects of the project.

Seaman has extensive experience in the area of library e-text centers, and has coordinated many collaborative library endeavors.

Lorcan Dempsey (Vice President for Research, OCLC). Dempsey will participate as a consultant advising the OCKHAM project on interoperability issues with existing library bibliographic utility functions provided by OCLC, including name authority, knowledge organization, and other OCLC functional components operating as web services. OCLC has indicated that they are interested in exposing these services within whatever framework OCKHAM develops for test purposes. Dempsey has a long-time international background in library network services, having previously been Director of the UK Office for Library and Information Networking (UKOLN), founding co-Director of the Resource Discovery Network (RDN), and Director of the Joint Information Systems Committee's Distributed National Electronic Resource (JISC DNER).

Intellectual Merit of the Proposed Activity

This project will advance the dialog between librarians and researchers through the creation of an extensible framework for networked interoperation between the NSDL and traditional libraries. Building on a year long planning effort by a group of systems librarians and computer scientists, we will advance the state of the art in digital libraries by integrating research on interoperability, reference models, peer-to-peer networked services, and componentized systems – with a focus on the needs of teachers and learners who will use NSDL, aided in whole or in part by an enhanced university library infrastructure.

Broader Impacts of the Proposed Activity

The OCKHAM Library Network proposal has been designed to maximize the broader impacts desired by the NSDL. The following summaries outline these broader impacts.

Advance Discovery and Understanding while Promoting Teaching, Training, and Learning:

This project will enable many new categories of students, teachers, and learning communities to use the NSDL through traditional library services. Libraries and librarians are key intermediaries in bridging research and education activities in higher education and society in general. Libraries provide the fundamental paired functions of stable access and preservation to scientific information for both academic and public learning communities. The NSDL will greatly benefit from closer integration with service programs in traditional libraries. Libraries devote extensive resources to the maintenance of catalogs and other information retrieval systems that are core information access mechanisms for learning communities. Librarians provide systematic training programs orienting learners to the use of the vast range of information resources that government and higher education makes available for the public good, and assist learners with practical aspects of using these information resources. The broad range of reference services provided at every library is “high-touch” and personalized through individual interactions with students, researchers, teachers, and the public alike. Library services are familiar to virtually all educated persons, and moreover are perceived as stable channels of information. Integration with traditional library services through the bridge infrastructure of the OCKHAM Library Network will significantly leverage the investment in NSDL resources for both research and educational purposes.

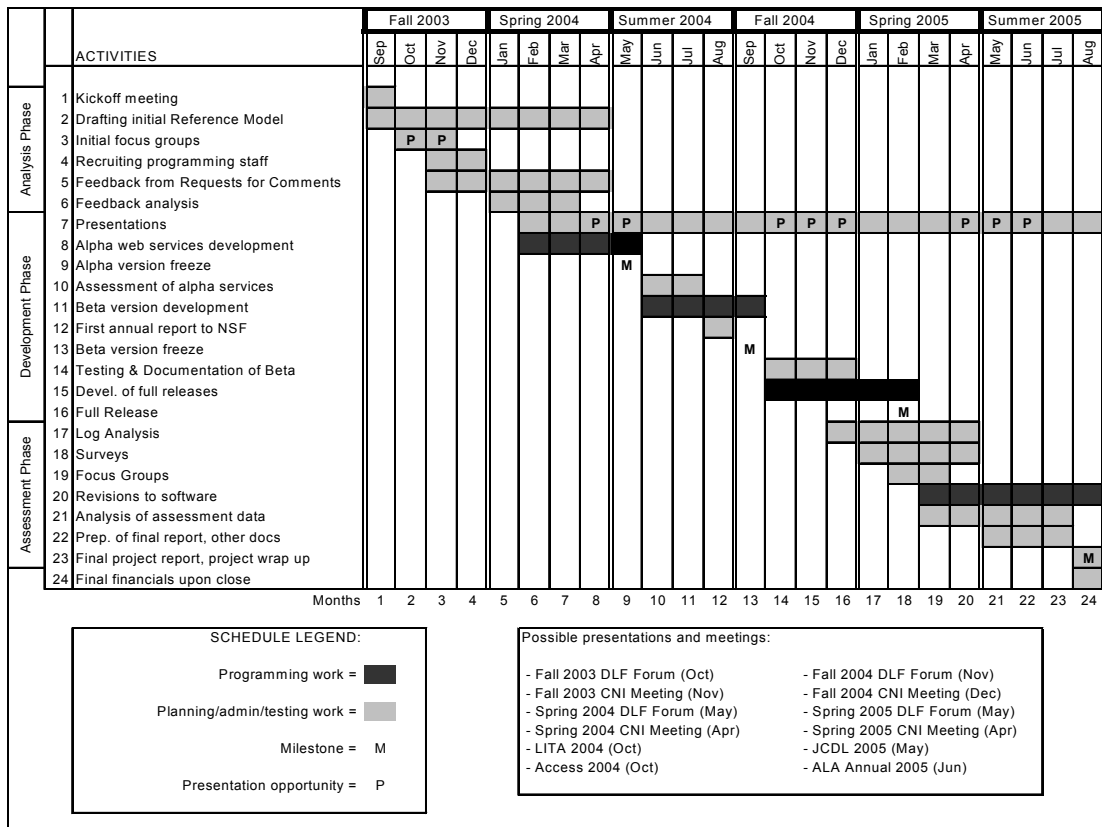
Enhance Infrastructure for Research and Education: Establishing the OCKHAM network effectively expands the infrastructure of the NSDL through the addition of the extensive apparatus of traditional library services. The extensive infrastructure of existing library catalogs, reference services, bibliographic utilities, and customized faculty services becomes available to the NSDL through the interoperable framework of the OCKHAM network. OCKHAM represents a range of new and powerful institutional alliances between the agencies of the NSF NSDL and the

tremendous range of standard libraries and library consortia. Further, because the OCKHAM network has been designed for extensibility in terms of new services and participating institutions, and with a minimum of entry barriers and cost outlays entailed in local implementations, the network is architected to spread relatively rapidly through the traditional library community. This will provide the NSDL with a growing and evolving presence in a widespread infrastructure supporting research and education.

Broad Dissemination to Enhance Scientific and Technological Understanding: The core mission of traditional libraries is providing access and preservation to information of relevance to the communities they serve. Since broad dissemination of information is at the core of their organizational priorities and operations, libraries are an excellent conduit and means for NSDL content to propagate to appropriate learning communities. The OCKHAM network will provide effective mechanisms for enabling the content of the NSDL to be disseminated through traditional libraries.

Timeline

We anticipate three overlapping phases for the project: the Analysis Phase in months 1- 8, the Development Phase in months 6 – 18, and the Assessment Phase in months 16 – 24. The initial Reference Model will be produced by month 6. The first working version of the network (alpha version) will be in operation by month 9. The second working version of the network (beta version), will be in operation by month 13, and will be used for initial tests of the evaluation systems. The full release will occur by month 18 and will be the primary evaluation target, though formative evaluation will proceed from day one. All assessment and evaluation activities will be complete by month 21. See the schedule of completion chart below for itemized details.



Management Plan

The project will be initiated by a kickoff meeting in month 1. This meeting will produce an outline of the reference model, initial drafts of some of the reference model sections, the detailed work breakdown for the project, and a plan for immediate steps to complete the model document. Weekly conference calls will provide a mechanism for coordinating work subsequently by the project team. Project meetings will subsequently be held opportunistically as members of the project team attend professional meetings. A project portal will be created to coordinate and archive working documents and code. This portal will be linked to the project's SourceForge site during the Development Phase. Status updates regarding testbed service development will be recorded on a bi-weekly basis. The schedule of completion (see Supplementary Documents) will be used to plan project activities, milestones around deliverables, and subsequent project meetings.

Dissemination Plan

The project will undertake several forms of dissemination and networking activities. The OCKHAM reference model, networked middleware, testbed services, project reports, and other outcomes will be widely shared with the traditional library community and the digital library research community. The dissemination plan and networking for the project includes several strategies:

Dissemination, Publications, Presentations

Conference presentations concerning the reference model, software toolkit, and web services will be timed to coincide with major project milestones. Initial versions of the reference model will be reported at the Spring 2004 meetings of the Digital Library Federation (DLF) and the Coalition for Networked Information (CNI). The release of the networked middleware and testbed services will be reported at the fall meetings of DLF and CNI, as well as the LITA National Forum and the Access 2004 conference. Finally, the assessment report and deployment model will be presented at the spring meetings of DLF and CNI, as well as the 2005 Joint Conference on Digital Libraries and ALA Annual 2005. The co-principal investigators and those providing support letters are frequent presenters at many of these meetings. The Annual NSDL Meeting will be a key event for sharing project results and progress. All of these presentations will provide public awareness of the project, as well as forums for garnering feedback concerning the network and reference model.

The OCKHAM reference model and network of testbed services will be publicized in the form of one or more peer-reviewed journal articles. These submissions will be made in year 2 of the project.

An open-source software site on Sourceforge.Net will provide an ongoing mechanism for distribution of the basic OCKHAM networked middleware software package, testbed service software packages, and documentation generated by the project. The Sourceforge.Net site will be updated throughout the project, and will provide a means of continuing software development and soliciting additional participation from others after the end of the project. The Sourceforge.Net site will be linked to an official project site with additional information concerning the software and deployment model.

Networking Activities

Server nodes at the four collaborating institutions will initially form the OCKHAM Network of Emory, Arizona, Notre Dame, and Virginia Tech. These four servers will be used to test the network architecture and local library applications of the OCKHAM web services.

Additional institutional participants will be carefully solicited during the project, and as many as possible will be selected on the basis of their understanding of the OCKHAM reference model, their ability to engage in the development process as implementers and testers of early version software, the degree to which they represent a good cross-section of different types of libraries, and on the basis of their ideas for local library services based on OCKHAM access to NSDL content.

One metric of success of the project (see below) will be the number and quality of additional OCKHAM node institutions that join the network by the end of the project.

Project Evaluation

Virginia Tech will engage in two types of evaluation. First, we will evaluate the impact of OCKHAM efforts in the NSDL. The aim will be to assess degree of deployment and of integration of NSDL into existing US library service offerings and systems. The focus will be on the utilization and benefits, at institutions participating in the proposal, of the reusable component-based testbeds that we develop. An impact measure will be defined and computed to indicate: how relevant the system is to the potential users, how many potential users know about it, and how many potential users actually incorporate it into their teaching and learning. Second, we will evaluate the services exposed as a part of the OCKHAM P2P network, both in terms of their functionality and ease of use. Services such as the alerting service and the find-similar service implemented as web services against the NSDL collections will be evaluated on the basis of the quality of content provided, and regarding their integration with the varied NSDL services and collections. We also will evaluate the manner in which the web services components (as instantiations of the reference models) interoperate among one another and among other web services.

Evaluation Process

Evaluation will take several forms. First, users will be surveyed using an online instrument. Users will be asked to assess both the aims of the work and the degree to which the project has achieved each of the aims. They also will be asked to comment on the services that are exposed, especially to describe if those are easy to use and if each service adequately addresses users' needs (i.e., regarding utility).

Second, comments and advice will be provided twice each year from our advisory group, representing DLF and OCLC members. In an advisory group meeting, we will summarize the goals and accomplishments of the project. Feedback will allow us to revise goals as needed to reflect user interests and needs.

Third, we will study effects in a small number of system testbeds (applications of the project technology to established NSDL resources). One system testbed will be CITIDEL, launched through an NSDL collection project award, serving the computing and information technology community (www.citidel.org). A second system testbed will be the NSDL portion of VT library services, aimed to support teaching and learning in STEM areas, that will be developed in keeping with generalized reference models. As other partner sites generate web service components, these will be deployed through the Virginia Tech library testbed. Thus, the aim of this testbed will be to establish and integrate the web services that are exposed project-wide as a part of the P2P

model. Similar testbeds will be established by OCLC and by other partners. The system testbeds established will be evaluated by a combination of usability studies, software trials, critiques of functionality, as well as through focus groups. Evaluation reports for all testbeds will be prepared and disseminated via public presentations, as well as through appropriate publications and web sites. We will use special technology to gain detailed knowledge regarding usage. Part of this will involve deployment of the remote usability technology developed by H. Rex Hartson at Virginia Tech. [Hartson, 1998] The rest will be based on logging, as explained below.

Log-Based Evaluation Method

Log analysis can serve as a primary source of knowledge about how DL patrons use DL systems and services and how systems behave while trying to support user information seeking activities. Thus, log recording and analysis facilitate evaluation and assessment. They open opportunities to improvements and enhanced services. We have defined an XML based digital library log format [Gonçalves & Luo, 2002] [Gonçalves 2003] using 5S as a guide [Gonçalves & Fox, 2003] for organizing the log structure and for defining the semantics of the DL components whose behavior is characterized. This standard captures a rich, detailed set of system and user behaviors supported by DL services. Our log component tool can be plugged into any DL system (e.g., the MARIAN DL system [Gonçalves & France, 2000]) thereby facilitating interoperability, reusability, and completeness. Thus, we will collect and analyze DL logs for the CITIDEL system and for all of the abovementioned testbeds. We will seek to integrate the logging capability into all the component web services exposed by the OCKHAM framework so that all aspects of behavior of the web service components can be made visible for analysis and evaluation.

Standardization of the logs has led to a standardization of their processing/analysis. Accordingly, we will deploy several of these analysis modules and tools as a part of the P2P network. A separate web service logging component will collect the log information generated by other components and provide analysis and reporting capabilities configurable to the needs of an individual system. As the design of the log analysis tool is highly modular, with little or no coupling between modules, this makes modification of current modules and creation of new modules very easy. In the case where a novel statistic is required or in the case that (recording regarding) a new feature is added, a new module can be built and connected to the already existing set of modules, thereby providing better quantitative results for evaluation.

Sustainability

To ensure the benefits of this project extend to multiple institutions well beyond the grant period, Emory University and the other core four members of the OCKHAM Network project will commit to maintaining the web services created during the grant project and their respective OCKHAM servers for at least two years beyond the end of grant-funded activities. The long-term success and viability of the network after this point will depend on several issues.

The sustainability of the OCKHAM Network is critically related to the degree to which the network attracts additional members and develops a community of practitioners. The degree to which the network grows is dependent on several factors: high utility in the services the network provides, minimal barriers to adoption, and reliability of the network. The project has taken steps in each of these areas to maximize the chances of success.

Another sustainability factor is the degree to which the OCKHAM Network is able to engage major library service vendors such as OCLC in promoting access to NSDL content. OCKHAM is fortunate in having OCLC represented in the analysis process through Lorcan Dempsey.

Project Criteria for Completion and Success

The following constitute the completion criteria for the project (how we will know when we're done), and the success criteria for the project (how we will know if we've succeeded).

Completion Criteria

All of the following are required for project completion.

- Criteria #1:** The reference model must be completed. It must contain the following sections: A) purpose and scope discussion, B) discussion of major concepts of the reference model, C) terminology section, D) detailed sub-models of the different parts of the reference model, E) scenarios for implementation, and F) a discussion of the process by which the reference model will evolve over time and what entities are responsible for maintaining the reference model.
- Criteria #2:** The OCKHAM P2P Network must function as a whole, demonstrating the main properties described in this proposal. The networked middleware and testbed services must all be completed and must function as specified in the proposal. All associated software must be accessible for public download via the SourceForge site.
- Criteria #3:** The evaluation plan must be completed, and all results documented.
- Criteria #4:** The dissemination plan must be completed. At least one article must be accepted for publication in a peer-reviewed publication.

Success Metrics

- Metric #1:** Degree of favorable survey responses. Success in this metric will be defined by a statistically relevant number of responses affirming utility of services enabled.
- Metric #2:** Degree of favorable feedback from Advisory Group. Both DLF and OCLC members affirming utility of services enabled will define success in this metric.
- Metric #3:** System testbed evaluation results. Success in this metric will be defined by positive results in the majority of software trials, usability studies, critiques of functionality, and focus groups.
- Metric #4:** Log analysis results. Success in this metric will be defined by patterns of usage that can be demonstrated to indicate utility of the OCKHAM network for end-users.
- Metric #5:** Growth of the network through additional institutions. Doubling the size of the network by the end of the project through the addition of at least four (or 14!) additional participant institutions as new nodes on the OCKHAM network will define success in this metric.