January 2003

A Survey of Digital Library Aggregation Services

Martha L. Brogan
University of Pennsylvania, brogan@pobox.upenn.edu

NOTE: At the time of publication, the author Martha L. Brogan was an Independent Digital Library Researcher and Consultant. Currently June 2007, she is Associate University Librarian for Collection Development and Management at the University of Pennsylvania.

A Survey of Digital Library Aggregation Services

Abstract
This report provides an overview of a diverse set of more than thirty digital library aggregation services, organizes them into functional clusters, and then evaluates them more fully from the perspective of an informed user. Most of the services under review rely wholly or partially on the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), although some of them predate its inception and a few use predominantly Z39.50 protocols. In the opening section of this report, each service is annotated with its organizational affiliation, subject coverage, function, audience, status, and size. Critical issues surrounding each of these elements are presented in order to provide the reader with an appreciation of the nuances inherent in seemingly straightforward factual information, such as "audience" or "size." Each service is then grouped into one of five functional clusters:

- open access e-print archives and servers;
- cross-archive search services and aggregators;
- from digital collections to digital library environments;
- from peer-reviewed "referratories" to portal services;
- specialized search engines.

NOTE: At the time of publication, the author Martha L. Brogan was an Independent Digital Library Researcher and Consultant. Currently June 2007, she is Associate University Librarian for Collection Development and Management at the University of Pennsylvania.

A Survey of Digital Library Aggregation Services

2003

Martha L. Brogan

Digital Library Federation

Washington, D.C.

About the Author

Martha Brogan is an independent library consultant with two decades of experience in academic libraries. Ms. Brogan has served as associate dean of libraries and director of collection development at Indiana University-Bloomington; as a social sciences librarian at Yale University; and as a western European library specialist and assistant to the provost and vice president of academic affairs at the University of Minnesota. In 2001, Ms. Brogan was a fellow in the Frye Leadership Institute sponsored by the Council on Library & Information Resources, EDUCAUSE and Emory University.

Based on a survey of Digital Library Aggregation Services conducted in summer 2003.

ISBN 1-932326-12-X
ISBN 978-1-932326-12-3

Published by:

The Digital Library Federation
Council on Library and Information Resources
1755 Massachusetts Avenue, NW, Suite 500
Washington, DC 20036
Web sites at www.diglib.org and www.clir.org

Additional copies can be purchased from the Digital Library Federation's Web site at www.diglib.org.

The paper in this publication meets the minimum requirements of the American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials ANSI Z39.48-1984.

Contents

Acknowledgments

I wish to acknowledge the exchange of ideas and invaluable feedback that I received from Jian Liu, Reference Librarian, Indiana University, in the early formulation of this study.

1.0 Executive Summary

This report provides an overview of a diverse set of more than thirty digital library aggregation services, organizes them into functional clusters, and then evaluates them more fully from the perspective of an informed user. Most of the services under review rely wholly or partially on the *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH), although some of them predate its inception and a few use predominantly Z39.50 protocols. In the opening section of this report, each service is annotated with its organizational affiliation, subject coverage, function, audience, status, and size. Critical issues surrounding each of these elements are presented in order to provide the reader with an appreciation of the nuances inherent in seemingly straightforward factual information, such as "audience" or "size." Each service is then grouped into one of five functional clusters:

- open access e-print archives and servers;
- cross-archive search services and aggregators;
- from digital collections to digital library environments;
- from peer-reviewed "referratories" to portal services;
- specialized search engines.

After a brief discussion of difficulties in attempts at categorization, each cluster is discussed at greater length through a closer examination of the purpose and functionality of individual services. A summary of overarching issues is provided for each cluster along with observations about disciplinary or national differences. The report concludes with observations about current practices and future directions. A list of major Web sites cited, a bibliography of cited works and further reading, and an appendix with scope notes round out the report.

The services under review are evolving and improving quickly—many are experimental or under development—so any attempt to describe or evaluate them must be undertaken with caution. The report is best viewed as a snapshot at a particular point in time seen

through the lens of an informed user, looking at a moving target.

Most of the published literature is project-specific and authored by those involved in developing and implementing the service. The 2003 special issue of *Library High Tech* focusing on the Open Archives Initiative merits special attention as an excellent state-of-the-art review of significant successes and challenges in creating OAI aggregators written by principal investigators [Cole 2003a]. The European Union's *Open Archives Forum* survey is exceptional in its effort to review broadly the organizational and technical characteristics of its member's archives [Dobratz and Matthaei 2003]. Meanwhile, the papers from the June 2003 "Wave of the Future: NSF Post-Digital Library Futures Workshop" give a fascinating picture of the challenges ahead [NSF 2003].

This report offers preliminary observations and points to future comparative studies—both broad-based and focused—that are necessary to sharpen and deepen our understanding of digital library aggregation services. Overall, it finds reason for optimism about open archives initiatives, especially given the relative youth of the OAI-PMH. However, it also points to the lack of information that users have about these services and their lack of knowledge about how they fit into the larger landscape of information seeking, resource discovery, and scholarly collaboration.

Many of the services are still in their first stage of development—collection and constituency building—where a primary concern is to increase the size of their holdings to achieve a critical mass, while continuing to assure quality control. As a second stage, some are beginning to provide coherent pathways through vast quantities of information by offering personalization and customization services. Most still have a long way to go in building extended services such as systems of annotation and collaboration. There is growing attention to a third phase of development, which is based on more flexible approaches to re-purposing resources for varied audiences and uses. Protocols for digital rights management and reliable digital preservation solutions will help to assure that these services reach their full potential.

2.0 Charge

This report, commissioned by the Digital Library Federation (DLF), reviews digital library aggregation services typified by Open Archive Initiative sites such as the *National Science Digital Library* (NSDL) or *OAIster*. The survey relied on a core list of 28 online digital libraries, federations, and OAI services provided to me by the DLF. The original annotated list of Web sites was arranged into four broad categories:

- Science, Technology and Medicine;
- Cross-Discipline;
- Humanities;
- Open Archive Initiative services—General.

As outlined in the section on "Scope Notes" (*Appendix 1*), I refined this list by removing some services and adding others.

More specifically, I was charged to evaluate these services based on their type, size, and function, while addressing the following questions:

- Do they cluster into sub-groups by function as well as by discipline?
- What broadly characterizes their scope and operation?
- What range of audiences do they purport to serve? How successful are they, in your opinion and in the opinion of any prior published assessments?
- What characterizes the experience of using these sites?
- Are there distinct differences in approach according to the discipline or nation that has produced the service?

3.0 Methodology

I conducted the review during July and August 2003, relying primarily on perusal of the Web sites, sample searches, and follow-up e-mail correspondence with many of the service providers. In addition, phone interviews were conducted with Los Alamos National Laboratory's (LANL) librarian regarding *Flashpoint*, the National Science Digital Library's (NSDL) communication director and the director of collection development, and ARL's Scholars Portal project manager. Site visits were made to OCLC in Dublin, Ohio and to the ENC Online (Eisenhower National Clearinghouse), headquartered at Ohio State University.

Due to the constraints of time and the diversity of services represented, a formal survey or questionnaire was not devised, although the review was informed by selected literature about digitized collections, subject gateways, portals, digital libraries, and open archives, especially in the broad areas of selection criteria and best practices; evaluation schemes; and most problematic of all—conceptual or organizational frameworks.

The European Union's Open Archives Forum's survey of "Open Archives Activities and Experiences in Europe" [Dobratz and Matthaei 2003] provides an excellent overview of a wide range of services in Europe. Although there is a growing body of literature about such digital library services worldwide, there are few examples of evaluations that compare resources or usability across multiple services. Ultimately, much of the literature is derived from the Web sites of these services themselves, most of which contain useful reports and studies.

I approached this review from the perspective of an "informed user" whose interest in technical issues is largely circumscribed by a desire to understand, in general terms, how technical decisions or restrictions affect the "scope and operation" of any given service, especially in terms of the "collections" covered or "items" retrieved. Given the recent literature on holistic approaches to digital library

evaluation, which take into account the expectations of diverse users—individually and collectively—with diverse needs, I acknowledge that my experience reflects a single stakeholder only.

4.0 Survey Overview

Although I suggested that the review be limited to those digital library aggregation services that rely solely on the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), ultimately a broader range of services was considered for several reasons:

- There are numerous exemplary hybrid services which include a mix of OAI-compliant and other resources.
- OAI-compliancy is evolving rapidly—some services, while not presently OAI-compliant, will be tomorrow.
- Non-OAI-compliant services can provide useful comparisons—especially in terms of purpose and functionality, both inferior and superior—to many fully compliant sites.

Table 1 provides an overview of all sites included in this review.[1] Each site is annotated by: organizational affiliation, subject coverage, function, audience, status, and size. A summary of findings and major issues in each of these categories follows.

4.1 Organizational affiliation

This category identifies the host institution, agency or consortia, along with selected funding highlights.

Critical issues:

Concerns about organizational affiliation are closely tied to issues of quality assurance, economic viability, and long-term sustainability. Virtually all of the sites under review are sponsored by institutions of higher education or by governmental agencies. Many are promoted by a handful of key individuals; few are fully integrated into a broad-based organizational structure. Many address "R&D" issues related to digital libraries and have not yet transitioned to full production. Almost none of the services make readily known their business plan, although some rely on community-based input and collaboration with varying degrees of formal governance structures. Most were developed with external funding support. Governmental agencies supporting the Digital Library Initiatives Phase 2 include:

- National Science Foundation (NSF) Digital Libraries Initiative http://www.dli2.nsf.gov/
- Defense Advanced Research Projects Agency (DARPA) Information Technology Office http://www.darpa.mil/ito/

[1] Table 1 is located in Appendix 2.

- National Library of Medicine (NLM) Extramural Programs http://www.nlm.nih.gov/ep/
- Library of Congress (LC) Digital Library Initiatives http://lcweb2.loc.gov/ammem/dli2/
- National Endowment for the Humanities (NEH) Digital Library Initiative http://www.neh.gov/html/guidelin/dli2.html
- National Aeronautics & Space Administration (NASA) http://www.nasa.gov/

In Partnership with:

- National Archives and Records Administration (NARA) http://www.nara.gov/
- Smithsonian Institution (SI) http://www.si.edu/
- Institute of Museum and Library Services (IMLS) Projects http://www.imls.gov/closer/cls_po.htm

The Andrew W. Mellon Foundation stands out among private foundations in its support for digital library initiatives. Waters [2001] describes Mellon's support for seven metadata harvesting projects. Cross-disciplinary services and R&D projects have also been supported by the Coalition for Networked Information and the Digital Library Federation.

Published literature on organizational issues, including business models, is scarce but growing:

- As noted above, Dobratz and Matthaei [2003] survey the landscape in Europe for the Open Archives Forum. The OAF published an "Interim Review of Organizational Issues" in November 2002 and will release its final report in early September 2003. It considers two taxonomies of business models [Rappa 2001 and Timmer 1999], commenting on their applicability to the Open Archives Initiative in its European context.

- Greenstein and Thorin [2002] consider three stages of digital library development within the context of research libraries: from aspiration to "skunk works;" rolling projects into programs; and from integration to interdependency.

- Chien [2002] in "Whither Digital Libraries? The Case of a 'Billion-Dollar' Business" considers the changing vision of a digital library to render them "sustainable (technologically, socially and economically) at the Internet scale." In particular, he draws on examples from digital government to turn it into "a business partner and research investor…making e-contents accessible, useful and profitable," with references to the European Commission's Information Society eEurope (http://europa.eu.int/information_society/eeurope/index_en.htm) and Japan's e-Japan Priority Policy Program http://www.kantei.go.jp/foreign/it/network/priority/).

- In "Business Models of News Web Sites: A Survey of Empirical Trends and Expert Opinion," Schiff [2003] examines differences among eight business models and summarizes them "in terms of three cross-cutting characteristics: (A) features that differentiate the online medium from print, broadcast and cable media; (B) key variables or components that affect business operations; and, (C) the maximizing or optimizing behavior that guides management strategy and measures their performance." The eight models include: Advertising revenue; Online traffic; Infant industry profits and stock values; Digital content delivery; Continuous breaking news; Information retrieval and storage; Portal conduit; and, Interactive networking.

- In "Ghosts in the Machine: People and Organization Level Issues in Distributed Libraries," Nicholson [2003] argues that: "Size matters where cooperation and collaboration are concerned. Even in a country as small as Scotland, a loose, nationally coordinated hierarchy of relatively small sectoral, regional, and special interest groups is the key to success. Where interoperability between people is concerned, small is beautiful."

- Zorich's [2003] "A Survey of Digital Cultural Heritage Initiatives and Their Sustainability Concerns" summarizes the organizational types, governance structures, business models, and sustainability concerns of thirty-three organizations or projects and five funding agencies or foundations.

- At the NSF's June 2003 "Wave of the Future: Post Digital Library Futures Workshop," Waters [2003b], in a paper entitled, "Beyond Digital Libraries: The Organizational Design of a New Cyberinfrastructure," recommends a new program of research on "organizations and organizational design," arguing that Advanced Cyberinfrastructure Program Centers "would need to be informed by current expert understandings and additional targeted research regarding organizational factors such as mission, leadership, governance, organizational structure, legal arrangements for intellectual property and financing…" He further recommends "an apparatus for incubating and supporting new organizations" that are responsive to disciplinary contexts but also "economize on the costly duplication of services" by creating a "family (or families) of efficiently run organizations" that "take responsibility for providing a set of common services, such as accounting, human resources, board governance and legal advice."

- Also at the NSF 2003 workshop, Van de Sompel [2003] laments the "lack of impact of the DL field … at the level of defining essential building blocks for the evolution of the Web infrastructure" and proposes the creation of "centers of excellence" as a partial answer. He also envisions a new digital library "ecology based

on distributed service provision: nodes specialize in specific tasks and *exchange* their services for those of nodes with other specializations" rather than the pre-digital era library model wherein "a library is an island [*peninsula*] that provides each and every service."

- Finally, Lynch [2003] highlights "the entire area of the stewardship, preservation and curation of information, discourse, knowledge, data, and culture. There are tremendous technical, economic, legal and political problems here; much progress has been made in mapping these problems, but much less in developing solutions." He suggests that these also need to become public policy goals and be examined in relation to "national security, or the protection of a nation's cultural heritage."

4.2 Subject coverage

Services are broadly categorized by subject as: cross-disciplinary; cultural heritage; science; humanities; and language resources.

Critical issues:

Given the funding streams, it is not surprising that major initiatives cluster around the mission of those governmental agencies supporting the Digital Library Initiatives Phase 2—predominantly in the sciences and cultural heritage. The participation of scholarly societies and commercial publishers is evident in the sciences. Meanwhile, the cultural heritage services bring together the museum, library, archive, and special collections sectors. Communities of practice are also forming around disciplines (e.g., geosciences); audiences (e.g., K-12 educators); media (e.g., sheet music, images); software (e.g., eprint.org, DSpace, Arc); or philosophies (e.g., preserving endangered languages, open access to scholarly communication).

Disciplinary differences in scholarly communication have been studied by Kling et al. [2000, 2002]; they argue that "it's not just a matter of time" before all branches of the sciences join the preprint movement. Brown [2002, 2003] and Lawal [2002] also survey differences in the adoption of preprints in various scientific disciplines. Articles about subject-based digital aggregation services are only beginning to appear in disciplinary journals [Johnston 2003; Lundmark 2003]. Much of the literature is produced by those who have created various services, e.g., Cole and other principal investigators in the special issue devoted to OAI of *Library Hi Tech* [2003]. Both the *Public Library of Science* and *DSpace* have been the subject of recent mainstream newspaper coverage, focusing in part on the economic dynamics of the open access movement.[2]

---

[2] For a summary and useful links refer to Open Access News: http://www.earlham.edu/~peters/fos/fosblog.html or the September 4, 2003 issue of ARL's *SPARC Open Access Newsletter:* https://mx2.arl.org/Lists/SPARC-OANews/Message/97.html

4.3 Function

Function is extracted in large part from the descriptions at each site as it relates to the service's primary mission. The categories are:

- open access e-print archives and servers
- cross-archive search services and OAI aggregators
- from digital library collections to digital library environments
- from peer-reviewed referratories to portal services
- specialized search engines

Selected relevant published literature is discussed in each section along with disciplinary and national differences. Grouping services by function is impeded by the issues outlined below and is discussed at greater length in the report.

Critical issues:

- the conflicting and overlapping definitions of concepts such as digital libraries, virtual libraries, portals, etc.
- the complexity of many services which don't lend themselves readily to solitary functional "encapsulation"
- the dynamic and innovative nature of these services which fuels their capacity to change functionality or scope
- the way in which successful data providers attract multiple new services, creating new levels of aggregation and customized functionality

4.4 Audience

Audience identifies the primary targeted users as: academic community, research community, educators, digital library developers or "interested public" although the latter could be attached to virtually of them.

Critical issues:

Two counter prevailing trends: serving multiple audiences for multiple uses versus serving a specialized audience for restricted uses. Many services purport to serve multiple audiences although they are primarily designed by and for the scholarly community. Others expect to serve a broad and diverse set of constituents, such as the *NSDL*, which has three audiences: users, content developers, and supporters (financial and political). Moreover, *NSDL* aims to serve users who are predominantly educators as well as users who have an interest in science in general. *NSDL* aims to provide the technical space, training, and tools for each constituency to use its collections appropriately. As the concept of reusable or repurposed digital assets gains acceptance, digital libraries may routinely support multiple user communities for multiple uses.

The counter-trend is services that are tailored to the particular needs of specialists and where some (or all) resources may only be available to members or subscribers.

4.5 Status

The services' "status" is noted as: experimental; pilot; under development; or, established. However, the latter term is used advisedly and is probably better conveyed as "evolving" because even the longest-lived sites adapt in response to new technology or may have unstable funding.

Critical issues:

Status is a moving target: *Arc* is stable in terms of its technical underpinnings, but as a cross-archive search service, it is experimental and its financial base is uncertain. *OAIster*, initially grant-funded, continues to improve its search functionality "as time permits." *Perseus*, created more than a decade ago, describes itself as "evolving." *DLESE* states that it is funded through 2007. These examples are characteristic of the overall ambiguous status of most of these services.

4.6 Size

Size is expressed in varying ways contingent on what was most readily available at the site, but including such measures as the number of institutional members, archives groups, or records. Size is difficult to measure and interpret for the reasons outlined below.

Critical issues:

- Size can change rapidly and growth or reduction must be interpreted with care. For example, although *OAIster* attempts to "de-dup" records among overlapping services, it harvests data from the *Open Language Archives Community* (OLAC) aggregator as well as from some of the individual repositories that comprise *OLAC*, such as *Ethnologue* and *Talkbank*. As a result, it is difficult to determine the actual number of "distinct" repositories covered by *OAIster*. This overlap also results in duplicate records when searches are performed. (OAIster is by no means an isolated example of these problems.)

- Close examination may reveal that a handful of archives account for the preponderance of records. For example, when OCLC provided OAI-compliant data from an extract of WorldCat's theses and dissertations (*XTCat*), it suddenly made available 4.3 million records for harvesting. As a result, any service provider (such as *Arc*) which has harvested all of these records grew tremendously in size. Meanwhile, *OAIster* limits its harvest of *XTCat* to the subset of 8,259 full-text items representing electronic theses and dissertations.

- The *UIUC Gateway to Cultural Heritage Materials* presents another interesting case study in changes in size. At its peak, the *Gateway* contained about 3.5 million metadata records, provided by a total of 39 metadata providers (both OAI-compliant and surrogates). However, the majority of these records described non-digital content resources (i.e., print and artifacts). Moreover, it included

metadata that was made available via other means than OAI-PMH, most notably 2.4 million Dublin Core records derived from EAD finding aids. *UIUC* decided to refocus its effort on metadata records describing digital resources and those derived from OAI-compliant metadata providers only. It removed all EAD collections, which they had broken apart into multiple item-level descriptors. The 8,500 EAD (Encoded Archival Description) records generated more than two million item-level records, which were removed from the database.[3] And when CIMI ("museum intelligence" consortium) shut down its testbed of 185,000 OAI-compliant museum records in early 2003, UIUC's coverage was further reduced. UIUC now harvests from 25 institutions and its repository contains 413,563 records.[4]

- The paradox of size: Critical mass is important. As repositories grow in size, they become more valuable; however by "being large and general, they are less easily tailored to individual uses" [Borgman 2003]. So at the same time that the *NSDL* is pushing to increase its size, it is also creating specialized portals to help different constituents filter its resources. Wiederhold [2003] refers to the "crucial task" of reducing the "available information to actionable information, i.e., the specific information that will cause a change in behavior, a reduction in further work, or the making of decisions" and describes the technologies to filter information that are "rapidly moving to harder and more speculative tasks."

  Meanwhile Schatz [2003] purports: "In the future, online information will be dominated by small collections maintained and indexed by small groups." He argues that "the Net has already made the transition from data transmission to information retrieval" and that it is "in the process of making the transition from information retrieval to knowledge management." Whereas the Grand Challenge in the 1990s was posed as "semantic interoperability across digital collections," Schatz proposes that the Grand Challenge in the 2000s will be "conceptual navigation across community repositories."

Bearing these issues in mind, *Table 1* is offered as a summary of key characteristics of each service. (See Appendix 2.)

5.0 Identifying Clusters By Function

5.1 Challenges to Categorization

After identifying the stated purpose or core function of these services, the greatest challenge lay in attempting to devise a broader framework which would cluster them and help to inform a comparative analysis. This exercise was hindered by several factors including:

---

[3] For further information about EAD refer to: http://www.loc.gov/ead/

[4] Information based on e-mail correspondence with UIUC's Timothy Cole and Sarah Shreeves on July 28, 2003.

- The conflicting and overlapping definitions (or lack thereof) of concepts such as digital libraries, virtual libraries, portals, gateways, archives, repositories, e-print archives, collections, digital objects, digital assets, and learning objects;

  To cite a few examples: *AARLIN (Australian Academic Research Library Information Network)* refers to itself as a "collaborative library service," a "research portal," and a "national virtual research library system." Meanwhile, the *Open Language Archives Community* is "an infrastructure for distributed archiving of language resources," a "worldwide virtual library," and a "network of language archives conforming to the Open Archives Initiative." *SMETE: Science, Math, Engineering & Technology Education Library* is a "dynamic online library and portal of services." Labeling itself a "digital library," *SMETE* is a "collection of collections" and a "community of communities."

  As far as labeling the data providers embedded within aggregators: *Arc* refers to them as "archives groups"; OAIster calls them "institutions"; *Cyclades* and *UIUC's Digital Gateway to Cultural Heritage Materials* refer to them as "collections." The Open Archives Initiative's registry of OAI service and data providers refers to both the aggregators and their component entities as "repositories." The *National Science Digital Library (NSDL)* refers to digital library services as "collections."

- The complexity of many services which don't lend themselves readily to solitary functional "encapsulation";

  The *National Science Digital Library (NSDL)* aims to serve three broad constituencies: the generally curious (interested in science and research information), the *NSDL* developer community and partners, and funding agencies and supporters.[5] A forthcoming reorganization of "nsdl.org" in October 2003 is intended to better reflect the needs of different constituents. *NSDL* seeks to provide the technical space, training, and tools for each of these audiences to use its resources appropriately. Meanwhile, *NSDL* covers 199 "collections," comprising 301,702 items derived from both NSF-funded projects and *NSDL*-selected sites. In addition, it has "services" available to help developers create digital content or to assist educators in evaluating and selecting digital resources. *NSDL* will also prototype several specialized portals to satisfy different sub-groups, e.g., middle school science teachers. As a result, *NSDL* serves different core functions for different audiences.

- The dynamic and innovative nature of these services which fuels their capacity to change functionality or scope;

  In July 2002, the *NASA Technical Reports Servers (NTRS)* launched the new version of its site, changing its architecture from distributed searching to metadata harvesting. Nelson et al. [2003] discuss the impact of this change; however, from a user perspec-

[5] Phone conversation and e-mail correspondence with Carol Terrizzi on August 4, 2003.

tive, several factors about the transition are noteworthy. The former collection constituted approximately 4.5 million abstracts and 300,000 full-text publications. As of August 4, 2003 the number of records in *NTRS* was 553,921, of which slightly more than half are full-text. However, NASA-agency full-text records number fewer than 15,000 and it is the newly introduced non-NASA archives (*Aeronautical Research Council*, UK; *arXiv*; *BioMed Central*; and OSTI's *Energy Citation database*) that account for almost all of the full-text content. Populating the new service with NASA-agency full-text data remains a priority. The new architecture makes it possible to search all the contents of *NTRS* by default. In addition, it offers both simple and "advanced search" functions. In the "advanced search" function, a search can be limited to specified NASA or non-NASA agencies. Two final points: (1) while widely regarded as an "e-print archive," *NTRS* is only about 50% full-text, mostly harvested from external non-NASA agencies and (2) with the inclusion of non-NASA archives, *NTRS's* subject scope has broadened.

The Perseus Digital Library originally concentrated on the development of collections, tools, and services to support classicists. However, after fifteen years of experience, it now comprises both third-party collections and those created for experimental purposes in seven different subject areas. Its future research agenda will focus on designing services that work with diverse collections and audiences.[6] It serves as a research bridge between cultural heritage digital libraries and the *NSDL*. With Johns Hopkins University, it received NSF funding starting in January 2003 to build a service for managing authority lists for customized linking and visualization for *NSDL*, based on tools already in use at *Perseus*.[7]

- The way in which successful data providers attract multiple new services, creating new levels of aggregation and customized functionality;

  The highly successful e-print archive and server for physics (and related disciplines), *arXiv*, (an OAI-registered data provider) now forms the core of the *Citebase* repository along with two other major e-print archives—*Cogprints* and *BioMed Central*. Meanwhile *Citebase* is registered as both an OAI data and service provider. *Citebase* offers an experimental search service that includes impact analysis and reference and citation linking. *ArXiv* also figures prominently in *Archon*, which identifies itself as a "digital library that federates physics collections with varying degrees of metadata richness." *Archon* will provide a unified search interface to diverse collections in physics, with sponsorship from LANL, Arc, the American Physical Society, the CERN Document Server, OAI, and Old Dominion University. *Archon* is a "collection" within the *NSDL*.

---

6 Crane et al. [2003]: 80.

7 The NSF award abstract is located at: https://www.fastlane.nsf.gov/servlet/showaward?award=0226304

Despite these difficulties in categorization, services were grouped together according to my understanding of their core value and mission. While clues were taken from how the services described themselves, I didn't always adhere to their self-analysis due to their overlapping use of terms as noted above, or because even though they referred to themselves as "digital libraries" or "portals" they didn't exhibit the same characteristics as other entities falling within that rubric.

The resulting categories are all open to debate. The differences among categories are subtle—a matter of nuance or interpretation—and their boundaries are fluid. Ultimately, the framework can best serve as an isolated effort to organize similar services together so I could use them to exemplify certain characteristics and trends. Depending on the user's perspective and needs, any given service could fall into a different category. In general, all the services under consideration are acknowledged to be exemplary, strive to excel, and offer quality assurance to users both in terms of the authority of their content and the qualifications of their producers.

5.2 Categories

5.2.1 Open Access E-Print Archives and Servers

This category includes scientific open access repositories that aim to provide access to full-text preprints, post-prints, technical reports, or other research output. The three examples represent a range in purpose from rapid dissemination of research findings without peer review, to public dissemination of scientific technical reports, to a publisher-based journal archiving system intended to preserve access to digital copies of articles. They aim to enhance open access to scientific scholarly communication and support the concept of "self-archiving" whether initiated by the author, the institution, or the publisher. This category of services has been cited in the mainstream news media for its efforts to challenge prevailing economic and publishing traditions.

5.2.2 Cross-Archive Search Services and OAI Aggregators

This category includes three broad-based, interdisciplinary cross-archive search services, one of which is a European collaborative that has introduced extended services layered on top of the repository. It then considers a set of more focused OAI metadata harvesting services and aggregators, grouped into either community-based or subject-based categories. The three community-based aggregators each have a different approach to building communities of practice, but they all aim to develop an organized federation of data providers who agree to adhere to certain philosophical and technical principles. These repositories serve, in large part, as union catalogs, providing a unified search interface to data at various levels of granularity. Finally, four examples of subject-based aggregators are discussed—one in cultural heritage and three in the sciences. Those in the sciences illustrate a narrowing of subject focus along

with increasingly sophisticated functionality—ranging from a basic repository of scientific e-prints and journals, to a testbed for e-prints with the potential for citation analysis and linking, to a federated collection that aims to serve as an authoritative physics "digital library" with extended services. Given the short history of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) on which these services are based, it is not surprising to find that most of these services have been created over the past few years and that many are experimental.

5.2.3 From Digital Collections to Digital Library Environments

This section considers a set of services that is evolving over time to greater complexity, starting with two examples from the cultural heritage sector that were created to heighten the use and visibility of primary resources and whose foundation is based on digitized collections. It continues with an example of how a discipline-focused digital resource is evolving over time into a broader testbed for research to improve digital library functionality. *The Perseus Digital Library* thus serves as a research bridge between cultural heritage digital collections and scientific digital libraries, which are examined more closely. It covers the *National Science Digital Library (NSDL)*—an extraordinarily rich and complex service—that strives to become a comprehensive digital library for the sciences, as well as four of its component "collections," which are independent and highly sophisticated digital libraries in and of themselves. These four are targeted for educators in the sciences at various academic levels.

The services in this category typically have more fully developed infrastructures—including evolving governance structures and policies for collection development, contributing data, or privacy of use. Many require users to register in order to obtain full benefits. They also offer a range of services such as conferences, workshops, or professional development opportunities; e-mail news alerts services; and opportunities to personalize services or to identify potential colleagues—characteristics that are also associated with "portals" as discussed below. Most represent a trajectory that moves beyond digital collections to digital library environments, as characterized by Lynch [2002]. Some also begin to cross the boundaries between digital library environments and digital learning environments as advocated by McLean and Lynch [2003].

5.2.4 From Peer-Reviewed Referratories to Portal Services

This category explores a set of Web or resource directories with different approaches to achieve quality-controlled content for an academic clientele.[8] They are labeled as "referratories" because they don't develop content or collections of their own, but rather refer the user to other sources of information. *MERLOT* has developed an advanced, distributed, national peer review system overseen by editorial boards and focused on expert-selected "learning materials" for college and university educators. Two other multidisciplinary, academically-ori-

[8] Several other examples of resource directories have been excluded from the discussion for reasons discussed in "Scope Notes" (see Appendix 1).

ented resource directories—one developed by a network of U.S. librarians and the other undertaken by disciplinary-based consortia in the UK—use a combination of expert-selected and machine-generated selections (including OAI harvesting) to build their "collections." Exhibiting features for personalization and collaboration, these services become "portals" to selected Internet resources and beyond, as exemplified by the UK projects.

AmericanSouth, a project under development by a network of Southern research institutions, based at Emory University, shows the potential of an OAI-based repository to become a "portal" for a specific user community through the implementation of customized services that facilitate scholarly communication. It is considered in this category rather than as an OAI aggregator, because its content will also draw on other sources such as local institutional catalogs and Internet resources.

Finally, two examples of research library portals under development, respectively in the U.S. and in Australia, are examined. Both of them are concentrating first on developing single search capability across licensed databases and local library catalogs relying on Z39.50 technology. In time, they may also develop the capacity to gather OAI-harvested data into their searches. They feature personalization services characteristic of portals.

5.2.5 Specialized Search Engines

This category includes three examples in the sciences: one is a proprietary in-house system for federated searches conducted primarily across locally loaded licensed databases and the local OPAC, and two are focused Web crawlers, capturing data from OAI-compliant and other Internet sources. These are presented as alternatives to generic, but hugely popular search engines such as *Google* or *AltaVista*.

Table 2: Overview of Core Functions and Services

CORE FUNCTION	SERVICES
OPEN ACCESS E-PRINT ARCHIVES AND SERVERS ❏ Open access to full-content via the Internet ❏ Typically author or institutional self-archiving ❏ Include: o Journal articles o Preprints & post-prints o Technical reports o Book chapters o Conference papers o Research output, including theses and dissertations ❏ May or may not be refereed. [Warner 2003 based on Pinfield et al. 2002] http://www.ariadne.ac.uk/issue31/eprint-archives/intro.html	**PHYSICS: AUTHOR SELF-ARCHIVING W/OUT PEER REVIEW FOR RAPID DISSEMINATION** arXiv **TECHNICAL REPORTS** NASA Technical Reports Server **VOLUNTARY PUBLISHER-BASED JOURNAL ARCHIVE OF PEER-REVIEWED ARTICLES** PubMed Central

CORE FUNCTION	SERVICES
CROSS-ARCHIVE SEARCH SERVICES & AGGREGATORS ❏ OAI metadata harvesting services and aggregators ❏ Search/discover gateways ❏ Information retrieval systems ❏ Indexes w/ unified search & browse features ❏ Function like union catalogs w/ enhancements ❏ Current status predominantly experimental ❏ Mix of collection-level and item-level access	**GENERAL OAI SERVICE PROVIDERS** Arc OAIster ❏ w/ EXTENDED SERVICES Cyclades **COMMUNITY-BASED ARCHIVES** ❏ THESES & DISSERTATIONS NDLTD Union Catalog (Networked Digital Library of Theses & Dissertations) (XTCat) Electronic Theses/Dissertations OAI Union Catalog based at OCLC [NDLTD] ❏ LANGUAGES Open Language Archives Community (OLAC) ❏ SHEET MUSIC Sheet Music Consortium **SUBJECT-BASED AGGREGATORS** ❏ CULTURAL HERITAGE UIUC Digital Gateway to Cultural Heritage Materials ❏ SCIENCES SCIENCE & ENGINEERING ARCHIVE Grainger Engineering Library at UIUC SELECTED E-PRINT REPOS W/ CITATION AND IMPACT ANALYSIS + REFERENCE & CITATION LINKING SERVICE Citebase FEDERATION SERVICE FOR PHYSICS ARCHON

CORE FUNCTION	SERVICES
FROM DIGITAL COLLECTIONS TO DIGITAL LIBRARY ENVIRONMENTS ❏ Collection of tools that make content alive ❏ Help user find content, manipulate, analyze, annotate, and comment on it ❏ Attract, create, define a community ❏ Collaboratories where active group annotation, analysis, and creation of new knowledge happens ❏ Enable and facilitate implicit communication (e.g., recommender systems) ❏ Sum greater than its parts [Lynch 2002]	**CULTURAL HERITAGE COLLECTIONS** American Memory Colorado Heritage (Colorado Digitization Program) **HUMANITIES** The Perseus Digital Library **SCIENCES** National Science Digital Library ❏ **FEDERATION** SMETE Digital Library (Science, Math, Engineering & Technology Education Digital Library) ❏ **K-12 TEACHER SUPPORT** ENC Online (Eisenhower National Clearinghouse for Mathematics and Science Education) ❏ **BIOLOGY NODE** BEN: A Digital Library of the Biological Sciences for Biology Teaching ❏ **EARTH SCIENCES NODE** DLESE: Digital Library for Earth System Education
FROM RESOURCE DIRECTORIES & REFERRATORIES TO PORTAL SERVICES ❏ Quality-controlled subject gateways ❏ Resource selection, discovery, annotation **PORTAL SERVICES** ❏ Collaborative information research service **Elements** ❏ Intuitive and customizable Web interface ❏ Personalized content presentation ❏ Security and Authentication ❏ Communication and collaboration **Components** ❏ Single-search interface ❏ User authentication ❏ Resource linking ❏ Content enhancement [Boss 2002]	**PEER REVIEWED LEARNING RESOURCES** Merlot (Multimedia Educational Resource for Online Learning & Teaching) **EXPERT & MACHINE-GATHERED INTERNET RESOURCES** ❏ ALL DISCIPLINES InfoMine *Scholarly Internet Resource Collections* ❏ DISCIPLINARY HUBS UK: Subject Portals Project of the Resource Discovery Network **SCHOLAR-DESIGNED OAI PORTAL** AmericanSouth **RESEARCH LIBRARY PORTALS W/ ACCESS TO PROPRIETARY DATABASES** ❏ U.S.: ARL Scholars Portal ❏ AUSTRALIA: AARLIN: the Australian Academic and Research Library Network

CORE FUNCTION	SERVICES
SPECIALIZED SEARCH ENGINES ❏ Information retrieval system ❏ Multidatabase search tool ❏ Filters ❏ Finds ❏ Searches ❏ "Niche" Search Engines	**SCIENCES** LANL FEDERATED SEARCH IN-HOUSE PROPRIETARY + SELECTED PREPRINTS + LIBRARY CATALOG Flashpoint COMPUTER SCIENCE WEB CRAWLER W/ REFERENCE LINKING, CITATION ANALYSIS, & RECOMMENDER SYSTEM CiteSeer (aka ResearchIndex) ELSEVIER WEB CRAWLER: SELECTED OAI REPOS + PROPRIETARY + WEB Scirus

6.0 Comparative Review By Function

6.1 Open Access E-Print Archives and Servers[9]

OPEN ACCESS E-PRINT ARCHIVES AND SERVERS ❏ Open access to full-content via the Internet ❏ Typically author or institutional self-archiving ❏ Include: o Journal articles o Preprints & post-prints o Technical reports o Book chapters o Conference papers o Research output, including theses and dissertations ❏ May or may not be refereed. [Warner 2003 based on Pinfield et al. 2002]	**PHYSICS: AUTHOR SELF-ARCHIVING W/OUT PEER REVIEW FOR RAPID DISSEMINATION** arXiv **TECHNICAL REPORTS** NASA Technical Reports Server **VOLUNTARY PUBLISHER-BASED JOURNAL ARCHIVE OF PEER-REVIEWED ARTICLES** PubMed Central

As Warner [2003] points out, definitions of e-prints vary widely from general meanings—*an e-print is a collection of digital documents*—to more restricted interpretations—*author self-archived preprints only*. Warner uses the term e-print to "group together many forms of

[9] Hitchcock [2003] has compiled a "core metalist" of open access e-print archives. This is not a list of individual archives, but rather an attempt to annotate and categorize by type other lists of individual e-print archives. In so doing, Hitchcock hopes to give a broad overview of the structure, size and progress of full-text open access e-print archives. Hitchcock's work should be consulted for a more comprehensive view.

scholarly literature for which there is open access to the full content via the Internet. E-prints may include: journal articles; preprints; technical reports; books; theses; and dissertations."[10]

It is appropriate to begin this survey with e-print archives because the Open Archives Initiative grew "from the 1999 Santa Fe Universal Preprint Service meeting [Ginsparg et al. 1999] and the Santa Fe Convention [Van de Sompel and Lagoze 2000], with the intention of improving scholarly communication through improved interoperability between e-print archives."[11] Moreover, e-print repositories represent a significant percentage of OAI data providers.[12] According to a survey conducted by Warner in October 2002, 54% of registered OAI data providers include metadata about e-prints.[13]

6.1.1 Physics: arXiv

ArXiv is the earliest, largest and most successful example of a subject-based e-print archive. What began in the early 1990s "as an experimental means of circumventing recognized inadequacies of research journals" quickly became the "primary means of communicating ongoing research information in formal areas of high energy particle theory."[14] *ArXiv* is based on a process of author self-archiving without peer review. It was widely accepted by this research community because of the pre-existing "preprint culture," which recognized the need for the rapid dissemination of research results without awaiting the time delays involved in peer review and formal publication. In addition to physics, *arXiv* now also covers mathematics, nonlinear science, and computer science, all overseen by advisory boards. In 2002, there were over 20 million full-text downloads from *arXiv*.[15] In the past year, monthly submissions to *arXiv* average from 3,000 to 3,500. (*ArXiv* is one of few repositories to make usage statistics readily available at its site.) *ArXiv* currently has about 230,000 items, all of which are full-text articles, technical reports, or theses.

In a 2003 submission, *Can Peer Review be Better Focused?*, Ginsparg [2003] discusses the characteristics of *arXiv* that account for its continued success: "From the outset, a variety of heuristic screening mechanisms have been in place to ensure insofar as possible that submissions are at least of *refereeable quality*… These mechanisms are

---

[10] Warner [2003]: 152. Nonetheless, Warner's survey of selected OAI data providers with a significant fraction of metadata about e-prints (as of October 2002), makes clear that some of the largest e-print archives have "limited time open access" or "restricted" access to full-text. See Warner: 154.

[11] Warner [2003]: 151.

[12] Identification of OAI Data and Service Providers comes from the OAI registry and is based on the following definitions: "The Open Archives Initiative Protocol for Metadata Harvesting (referred to as the OAI-PMH in the remainder of this document) provides an application-independent interoperability framework based on metadata harvesting. There are two classes of participants in the OAI-PMH framework: Data Providers administer systems that support the OAI-PMH as a means of exposing metadata; and Service Providers use metadata harvested via the OAI-PMH as a basis for building value-added services." http://www.openarchives.org/OAI/openarchivesprotocol.html

[13] Warner [2003]: 154.

[14] Ginsparg [1994].

[15] Ginsparg [2003].

an important—if not essential—component of why readers find the site so useful: though the most recently submitted articles have not yet necessarily undergone formal review, the vast majority of the articles can, would, or do eventually satisfy editorial requirements somewhere."[16] He goes on to suggest how various impact measures could be used in the preprint environment to bring greater efficiency to the full peer review process by focusing it on a smaller subset of submissions, but also one with a higher likely acceptance rate.

It is possible to search *arXiv* by date and sub-group within physics or by date and group for math, non-linear science, and computer science. It supports searches by author, title, full record, comments, journal-reference, subject-class, or report number with Boolean operators. Help and examples of search functions appear directly on the search page. There is also a forms-based interface to searching that permits different views (new abstracts, last update, recent, etc.). A "catch-up" function allows users to review new records, with or without abstracts, within the dates specified. There are several options to download files. The "Help" feature contains general information as well as information about browsing and instructions for those submitting papers. There is a FAQ. "What's new" informs users of changes to the site, e.g., on July 6, 2003: "A new and more sophisticated author registration system has been put on-line. It provides greater administrative flexibility and better user support, including user ability to maintain past submissions."

6.1.2 Technical Reports: NASA Technical Reports Server (NTRS)

NASA's technical reports server, ***NTRS***, seeks to collect, disseminate, and archive the "unclassified, unlimited" NASA-authored scientific and technical literature related to aeronautics. As discussed above, *NTRS* exemplifies some of the difficulties in making the technological transition from distributed searching to metadata harvesting. Populating the database, in particular with NASA-authored data, remains a priority. The new *NTRS* version does provide a unified search interface to reports from ten NASA agencies and four non-NASA agencies, receiving about 6,000 to 7,000 searches monthly. Among the 555,358 records, NASA-authored reports account for less than half of the total and only about 50% of all items are available in full text, of which NASA reports account for less than 5%. Search results, which include extensive abstracts, clearly indicate if a digital version is available, and when it is not, provide information about ordering it. Four key NASA agencies are not part of *NTRS* and must be searched separately. *NTRS* has a useful update feature where it is possible to search the records added to all of the archives or to specific archives on a weekly basis (up to the past four weeks) or by entering a specific date stamp. *Table 3* summarizes the contents as of August 19, 2003.[17]

---

[16] Ginsparg [2003].

[17] Full-text count and number of monthly searches were provided in email with Michael Nelson on August 4, 2003.

Table 3: NTRS Contents

NASA ARCHIVES	NUMBER OF METADATA RECORDS
GENESIS (NASA Jet Propulsion Laboratory)	All full text: 27
NASA Ames Research Center	Metadata indexed but full text quarantined because they haven't been reviewed. Records: 354
NASA Center for AeroSpace Information (CASI)	100 full-text documents out of 256,637
NASA Goddard Institute for Space Studies	All full text: 1,335
NASA Goddard Space Flight Center	Metadata indexed but full text quarantined because they haven't been reviewed. Records: 11
NASA Johnson Space Center	All full text: 128
NASA Kennedy Space Center	Metadata indexed but full text quarantined because they haven't been reviewed. Records: 82
NASA Langley Research Center	All full text: 3,948
NASA Marshall Space Flight Center	All full text: 498
NASA Stennis Space Center	Metadata indexed but full text quarantined because they haven't been reviewed. Records: 39
National Advisory Committee for Aeronautics (NACA)	All full text: 7639
RIACS (NASA Ames Research Center)	All full text: 61
NON-NASA ARCHIVES	**METADATA RECORDS**
Aeronautical Research Council (UK)	All full text: 2,647
arXiv Physics Eprint Server	All full text: 243,707
BioMed Central	All full text: 17,507
Energy Citation Database (OSTI)	7,000 full-text articles out of 20,738
PREVIOUSLY INCLUDED BUT NOT IN THE NEW OAI NTRS	**RELATED WEB SITES**
NASA Astrophysics Data System: (1) Astronomy & Astrophysics, (2) Physics & (3) Geophysics, Space Instrumentation or available via: (4) *The Astrophysics Data System (ADS) is a NASA-funded project which maintains four bibliographic databases containing more than 3.3 million records: Astronomy and Astrophysics, Instrumentation, Physics and Geophysics, and preprints in Astronomy. The main body of data in the ADS consists of bibliographic records, which are searchable through our Abstract Service query forms, and full-text scans of much of the astronomical literature which can be browsed though our Browse interface.*	Searchable at: (1) http://adsabs.harvard.edu/abstract_service.html (2) http://adsabs.harvard.edu/physics_service.html (3) http://adsabs.harvard.edu/instrumentation_service.html (4) http://adswww.harvard.edu/
NASA Dryden Flight Research Center	Searchable at: http://www.dfrc.nasa.gov/DTRS/
NASA Glenn Research Center	Searchable at: http://gltrs.grc.nasa.gov/
NASA Jet Propulsion Laboratory	Searchable at: http://gltrs.grc.nasa.gov/

6.1.3 Voluntary Publisher-based Journal Archive: PubMed Central

Launched in February 2000, ***PubMed Central*** **(PMC)** is a digital archive of life sciences journal literature maintained by the National Center for Biotechnology Information (NCBI) at the U.S. National Library of Medicine. It provides free and unrestricted access to some 100,000 full-text articles from over 130 journals. *PMC* contains all peer-reviewed primary research articles from every participating journal; other content is made available at the discretion of the journal editor (e.g., letters, essays, and reviews). It strives to provide open access to this literature in perpetuity. Journals may deposit the full text of articles with *PMC* and release it immediately upon publication or delay its release for a specified period. *Participation in PubMed Central (PMC) is voluntary and open to any life sciences journal that either is covered by one of the major abstracting and indexing services such as MEDLINE, Agricola, Biosis, Chemical Abstracts, EMBASE, PsycINFO or Science Citation Index, or (if a new journal) has at least three members on its editorial board who currently are principal investigators on research grants from major funding agencies (such as NIH) in the U.S. or abroad.*

PMC provides unified search capability across more than 75 life science journals and all 57 core journals published by *BioMed Central (BMC)*. *PMC* allows journals to maintain their distinct identity by supplying the journal logo at the top of each page (with a link to the journal's own site) and by running the journal's "watermark" the length of each page. At present *PMC's* coverage is limited to English-language journals. All articles in *PMC* are also indexed in *PubMed*, the online index and abstracting service of the National Library of Medicine, which includes *Medline*.[18]

As explained by Edwin Sequeira of NCBI in a 2003 article: "The standard PMC search technique is labeled 'SmartSearch,' reflecting the fact that it is based on an automated analysis of the title, abstract, and full text of each article. SmartSearch is intended to increase the relevance of one's search results. It includes intelligent phrase recognition and does not search every word in an article as a simple full-text search would do (although it is also possible to do the latter if one wishes)." *PMC* also offers extensive search features, including automatic term mapping that matches unqualified terms against a MeSH (Medical Subject Headings) Translation Table, a Journals Translation Table, a Phrase List, and an Author Index. Terms can be qualified using search field tags and date ranging. It is possible to limit your search to specific search fields, to preview the search results before displaying the citations and to refine your search. Results can be sorted according to various options. You can save a text file of citations on your computer with results up to a maximum of 10,000

---

[18] PMC is closely affiliated with both BioMed Central (BMC) and the Public Library of Science. As explained at PMC's FAQ: *BioMed Central (BMC) is a commercial publisher of online biomedical journals, which provides free access to articles at its site. BMC also deposits its articles in PubMed Central as they are published. The Public Library of Science (PLS) was created by an independent group of researchers who seek to ensure that all life science literature becomes freely accessible to the public within six months of publication. PLS views PubMed Central as an appropriate vehicle through which to distribute scientific content.*

items. There are options to print or e-mail results from the clipboard that holds up to 500 citations. These and other search features are explained at length at the site's Help page: http://www.ncbi.nlm.nih.gov/entrez/query/Pmc/pmchelp.html.

PMC is also extending its content through a systematic scanning program of back runs of journal articles. Sequeira reports that about a year ago: "NLM offered to scan any issues of a PMC journal that are not already available in electronic form, in return for permanent rights to archive and distribute the scanned material freely. Almost all the current PMC journals that have pre-electronic issues are participating in the project, as are the 20-plus specialist journals of the BMJ Publishing Group, whose current content will be added to PMC later." The back issue digitization project is described more fully at the PMC Web site: http://www.pubmedcentral.gov/about/scanning.html.

OAI access to *PMC* is anticipated by mid-September 2003, including access to much of the *BMC* content, as well as to the new *PLoS Biology* journal and any other open access journals.[19]

6.1.4 Summary of Issues

All three of these e-print archives support the concept of open access and self-archiving (by author, by agency, or by publisher). *ArXiv* and *NTRS* are registered OAI data providers and OAI access to *PMC* is anticipated by mid-September 2003. While they also all aim to speed up access to research findings, each of them also illustrates a different purpose: *arXiv* serves primarily for the rapid dissemination of research findings without peer review based on author self-archiving; *NTRS* aims to distribute scientific and technical literature quickly and widely through agency-based archiving; and *PubMed Central* promotes publisher-based archiving in order to preserve life sciences journal articles in electronic form. Both *arXiv* and *PubMed Central* provide access to full content only. *NTRS* on the other hand, is about 50% digital full text—most of it from *arXiv*—with many NASA reports requiring a purchase in hard copy or microfiche.

These three services also highlight disciplinary differences. While physics has a tradition of distribution of preprints without peer review, acceptance varies even among its sub-fields [Brown 2002]. Lawal [2002] discusses some of the underlying reasons for varying rates of adoption by researchers in nine scientific disciplines including chemistry, biological sciences, engineering, cognitive science and psychology, mathematics and computer science, physics, and astronomy. She found widest adoption in physics, followed by mathematics, and the least in chemistry. Publishers' policies are a primary factor in chemistry's non-use of preprint archives. Brown [2003] surveyed authors of e-prints appearing in the *Chemistry Preprint Server (CPS)*, operated by Elsevier and the editors of top chemistry journals about their acceptance of *CPS* e-prints. She notes that while authors found *CPS* "a convenient vehicle for dissemination

[19] Based on email correspondence with *PMC* on August 18, 2003.

of research findings and for receipt of feedback before submitting to a peer-reviewed journal, reception of CPS e-prints by editors of top chemistry journals is very poor." At the same time, she reports that "32 percent of the most highly rated, viewed and discussed e-prints eventually appear in the journal literature, indicating the validity of the work submitted to the CPS." Meanwhile, the two dominant publishers in Chemistry—Elsevier and the American Chemical Society—in 2003 announced an even closer collaboration:

> Elsevier and two divisions of the American Chemical Society—Chemical Abstracts Service (CAS) and Publications—have announced that they have agreed to provide linking between their services for scientists. Under the agreement:
>
> 1. Users of Elsevier products and services (such as Science Direct®, MDL® databases and ChemWeb) will be able to link directly to ACS scientific journals
>
> 2. Users of CAS products and services (SciFinder, STN®, and others) will be able to link, via ChemPort, directly to Elsevier scientific journals.[20]

Researchers in the life sciences adhere to the tradition of peer review prior to dissemination of research papers, but readily deposit genetic sequences into GenBank®, the National Institute of Health's annotated collection of all publicly available DNA sequences.[21] The life sciences are also vigorously promoting open access to peer-reviewed literature and taking advantage of technology and new pricing models (institution-based article-input fees as opposed to subscriber fees) to speed up dissemination.[22]

Although no examples of institutional archives were part of this review, MIT's *DSpace* has received attention for spurring the self-archiving movement. As reported in the *New York Times* it "will have 5,000 items archived by this fall, and plans call for adding 7,500 theses later this year. MIT estimates that its free software has been downloaded 3,400 times and says it is aware of 100 research institutions that are evaluating DSpace with an eye toward archiving their own faculty's publications."[23] The United Kingdom has also announced its plans to develop a national archive of e-print papers available from OAI-compliant repositories provided by UK universities and colleges. According to the UK plan:

> Metadata will be harvested using the OAI protocol into a single database hosted by UKOLN at the University of Bath, and

---

[20] For full press release of August 18, 2003 see: http://www.elsevier.com/homepage/newhpgnews/production/cas/links/link1.htm, accessed on August 23, 2003.

[21] For more information see about GenBank, see Benson et al. [2003].

[22] BioMed Central has started to publish a newsletter "Open Access Now," with the inaugural issue of July 14, 2003. See: http://www.biomedcentral.com/openaccess/pdf/OpenAccessNow_1.pdf, accessed on September 3, 2003.

[23] Vivien Marx, "TECHNOLOGY; In DSpace, Ideas Are Forever" in EDUCATION LIFE SUPPLEMENT, Section 4A, Page 8, Column 1 *New York Times*. (August 3, 2003): available for purchase at http://www.nytimes.com/2003/08/03/edlife/03EDTECH.html, accessed on August 19, 2003.

> will then be passed to external web services—OCLC and the University of Southampton—where the records will be enhanced with subject classification, name authority, and citation analysis. The enhanced records will be returned to the central database from where they may be harvested by institutions or academic subject gateways. The project is funded by the JISC FAIR program. [See: http://www.rdn.ac.uk/projects/eprints-uk/]

Day [2003] reviews the status of institutional and subject-based repositories in the UK using Eprints.org software and corroborates the assertion of Pinfield [2003] that more effort now needs to focus on actually *populating* repositories:

> Setting up an institutional repository and designing collection management policies are relatively straightforward; populating the repository is not. The content of institutional repositories needs to come largely from researchers within the institution, and persuading them to submit this content is a major challenge. Self-archiving requires a cultural change amongst researchers that can only be achieved through significant advocacy activity, and even then it will probably happen only gradually.[24]

While the e-print and self-archiving movement may be gaining momentum, there are still obstacles to overcome, namely acceptance by authors in sufficient numbers to develop repositories of sufficient size to be of interest, and finding efficient ways to manage copyright issues. Turning again to the United Kingdom, the Joint Information Systems Committee (JISC) funded a project (through August 31, 2003), RoMEO (Rights Metadata for Open archiving), to investigate the rights issues surrounding the "self-archiving" of research in the UK academic community under the OAI-PMH. According to RoMEO's Web site:

> It will perform a series of stakeholder surveys to ascertain how 'give-away' research literature (and metadata) is used, and how it should be protected. Building on existing schemas and vocabularies (such as Open Digital Rights Language) a series of rights elements will be developed. A demonstrator system will then be created to show how rights metadata might be assigned, disclosed, harvested, and displayed to end users via the OAI Protocol for Metadata Harvesting [http://www.lboro.ac.uk/departments/ls/disresearch/romeo/index.html].

Meanwhile, Van de Sompel stated in a 2003 interview that the Open Archives Initiative expects to set up a technical committee soon in collaboration with the JISC RoMEO project, "in the realm of expressing rights statements about metadata and content in the OAI framework."

---

[24] Pinfield, cited by Day, p.8-9.

6.2 Cross-Archive Search Services and Aggregators

CROSS ARCHIVE SEARCH SERVICES & AGGREGATORS

- ❏ OAI metadata harvesting services and aggregators
- ❏ Search/discover gateways
- ❏ Information retrieval systems
- ❏ Indexes w/ unified search & browse features
- ❏ Function like union catalogs w/ enhancements
- ❏ Current status predominantly experimental
- ❏ Mix of collection-level and item-level access

GENERAL OAI SERVICE PROVIDERS

Arc
OAIster
- ❏ w/ EXTENDED SERVICES

Cyclades

COMMUNITY-BASED ARCHIVES

- ❏ THESES & DISSERTATIONS

NDLTD Union Catalog (Networked Digital Library of Theses & Dissertations)
(XTCat)
Electronic Theses/Dissertations OAI Union Catalog based at OCLC
- ❏ LANGUAGES

Open Language Archives Community (OLAC)
- ❏ SHEET MUSIC

Sheet Music Consortium

SUBJECT-BASED AGGREGATORS

- ❏ CULTURAL HERITAGE

UIUC Digital Gateway to Cultural Heritage Materials
- ❏ SCIENCES

SCIENCE & ENGINEERING ARCHIVE
Grainger Engineering Library at UIUC

SELECTED E-PRINT REPOS W/ CITATION AND IMPACT ANALYSIS + REFERENCE & CITATION LINKING SERVICE
Citebase

FEDERATION SERVICE FOR PHYSICS ARCHON

This category consists of three general OAI service providers—*Arc*, *OAIster* and *Cyclades*; three examples of community-based aggregators—*Networked Digital Library of Theses & Dissertations*, *Open Language Archives Community*, and the *Sheet Music Consortium*, and four examples of subject-based repositories, one for cultural heritage and three in the sciences—*UIUC Digital Gateway to Cultural Heritage Materials*, *Grainger Engineering Library at University of Illinois at Urbana-Champaign*, *Citebase*, and *Archon*. All of these services federate metadata of "varying degrees of richness" from heterogeneous sources, relying on the OAI-PMH, and provide unified search and browse interfaces. They are all established or experimental in nature, typically with support from external funding agencies. Most cover materials in multiple languages and formats. They represent

a mix of approaches to collection-level and item-level access. Several of the services have established special metadata standards. Other differences are apparent in the level of sophistication of their search capabilities and their post-result processing features.

6.2.1 General OAI Service Providers: Arc, OAIster, Cyclades

Arc, developed by Old Dominion University's Digital Library Research Group, is one of the first federated searching services based on the OAI Protocol.[25] It serves as a technology demonstrator by harvesting from all OAI repositories without any limitations by the type or subject of their holdings. As a result, it is the largest OAI service provider included in this review, currently harvesting from 163 archives, comprising a total of 6.4 million records, 4.3 million of which are derived from OCLC's XTCat (theses and dissertations extracted from WorldCat). *Arc* serves as a testbed where *Arc* and other OAI service providers can experiment with the resulting federation. (For example, at present *Arc* is conducting a study on accession growth rates.) *Arc* is also "experimental" in the sense that it has no base funding to support a sustainable federation service. From a technical standpoint, however, *Arc* is well established and its "code" for federated searching has proven to be extremely stable, robust, and error free. Its harvesting and indexing software is available to download as Open Source software at sourceforge.net.[26] *Arc*'s developers are committed to improve OAI services and are currently working on a major upgrade of the Open Source version of *Arc* to upgrade services for its community of users.[27]

Arc offers simple keyword searching with Boolean operators where the user can specify how to group (by Archive, Discovery Year, or Subject) and sort the results (by Relevance Ranking or Discovery Date). Advanced searches permit Author, Title, and Abstract searches where the user can indicate if all or any instances of the specified terms should be retrieved. Advanced searches can also be filtered by Archive, Subject, Date Stamp, or Discovery Date. The subject filter includes an interactive feature where the user can input a term and receive a listing of related subjects and their archive group affiliation, making it possible to further refine the subject search. *Arc* also offers a "browse" feature that lists records in alphabetical order by archive group; however, browsing by subject or year returns incomplete results. All search results can be displayed in summary or detailed views. Following the links on the detail page, lead the user to the particular document, residing at the local host site. When there are multiple pages of returns, the user can traverse them.[28] The

---

[25] Information about ODU's Digital Library Research Group is located at: http://dlib.cs.odu.edu/

[26] The following OAI service providers are known to use the Arc search engine: The Resource Discovery Network's Resourcefinder: http://www.rdn.ac.uk/resourcefinder/; MetaArchive Initiative http://www.MetaArchive.org and its affiliate AmericanSouth.org http://www.AmericanSouth.org; Archon, a federated search service for physics: http://archon.cs.odu.edu; Networked Computer Science Technical Library: http://www.ncstrl.org; SNEL Digital Library (serving academic interests in the Sudan): http://www.snelonline.net/snel/index.jsp

[27] Information about *Arc* is based, in part, on email communication with Xiaoming Liu and Kurt Maly in July and August 2003.

[28] See Liu [2002] for further information about search functionality.

"Help" page provides information about how to search the site with an e-mail address for additional questions.

In practice, I encountered a number of problems in conducting searches.

- There is no overall collection policy or statement about the scope of coverage. If the user clicks on the "browse" feature, it is possible to view all the "archive groups" in the left frame with their individual records appearing in the main body of the page. However, the left frame listing of "archives groups" includes two consecutive alphabetical listings because those beginning with capital letters are filed separately from those in lower case letters. As a result, at first blush, the user would think that "arXiv" is *not* included in an *Arc* search.
- The names of the "archives groups" are typically cryptic and most of them don't carry any meaning for the general user. For example, only specialists would know that "AIM25" provides collection-level descriptions of archives in London or that "CPS" is the Chemistry Preprint Server. *Arc* makes some effort to provide the fuller names of these repositories delivered via mouseovers to the list of abbreviated identifiers, but it still leaves the user without many clues.
- By going to the "Administration" page, which is scarcely an intuitive choice, the user will also find a list of all the existing archives, their "identifiers" and full names, along with the date on which they were last harvested. From this administrative page, it is possible to link to the Web site of each of the archives, where the user can get an understanding of their scope and coverage.
- *Arc* continues to harvest from both the current and prior versions of the OAI Protocol, resulting in duplicate archive groups. For example, results are returned for two separate archives groups identified as "arXiv" (214,215 records) and "arXiv.org" (240,164 records).
- Many returns don't actually link to full content, even at the host site. For example, the item "Harlem nocturne" only leads to a description of Indiana University's DeVincent Sheet Music Collection without any direct link to the site. Even when the user goes to this site and searches the database, there is no full content available, only the bibliographic record.
- It is not possible to revise a search.
- Searches return duplicate "hits" when an item is recorded by more than one repository or within a repository when it is still represented in two versions (e.g., OAI-PMH 1.x and OAI-PMH 2.0).

OAIster, a project of the University of Michigan Digital Library Production Services, originally funded through a grant from the Andrew W. Mellon Foundation[29], represents another broad-based OAI search service, but unlike *Arc*, the information resources that the metadata describe must have a corresponding Web-based digital rep-

[29] For further information about seven Mellon-funded OAI metadata harvesting initiatives refer to Waters [2001].

resentation (e.g., records from Indiana's "Harlem nocturne" would not be retrieved from this site because this piece of sheet music itself is not available in digital form.) As a result of this requirement, *OAIster*'s coverage is narrower than *Arc*'s—as of August 28, 2003, *OAIster* included over 1.5 million items from 197 "institutions." (This is an increase over the July 3rd harvest of 1.4 million records from 189 institutions.) All searches result in links to digital objects. *OAIster* is exemplary in its efforts to provide context and elaboration about its scope and operation. The annotated listing of institutions from which *OAIster* harvests offers the user basic information about each repository, along with the number of records harvested. Search functions are unified into a single search page that permits varying degrees of refinement from basic keyword searching with Boolean operators to searching within particular fields (Title, Author/Creator, Subject, Resource type). This latter category is especially useful in that it permits the user to limit the search to all types or to text, image, audio, or video formats. Results can be sorted by: title, author/creator, date descending or date ascending, and by hit frequency or weighted hit frequency.

Like *Arc*, *OAIster* displays results counts by institution and makes it possible to link to a specific institution's results in the left frame. An immediate full view of each result avoids the double-clicking to "more information" required in *Arc*. Other strengths of *OAIster* search capability are that it:

- provides the total number of returns
- permits users to revise the search
- permits post-search (re)sorting of results according to different criteria
- highlights the search term within the results
- offers ample "help" opportunities
- prominently acknowledges and explains the "duplicate records" problem

At this juncture, neither *Arc* nor *OAIster* offer post-result services such as printing, book marking, downloading, or incorporating the digital object into another document or file, although *OAIster* notes that these are desired improvements.[30]

Some problematic notes about *OAIster*:

- For the benefit of regular users, when updates are made to *OAIster*, it would be helpful if a "What's new" column informed users of institutions added or removed, along with the number of items associated with these changes. Right now this information is purged from the database on a monthly basis so there's no record of changes.
- The listing of "institutions" is somewhat problematic since the organization of the list is sometimes by the name of the service rather than the institution, e.g., Theoretical and Applied Linguis-

[30] For further information about OAIster's development, including user survey results, refer to Hagedorn [2003] and Wilkin et al. [2002].

tics (TAAL) Eprints Archive, University of Edinburgh, files under "Theoretical" not "University of Edinburgh."
- Individual archives within aggregators lose their identity and are not specified in the annotations by institution. For example, the *Open Language Archives Community (OLAC)* has 25 registered archives. These are covered by *OAIster* collectively through *OLAC*, but the user needs to go to the *OLAC* site to determine what the 25 archives are. Meanwhile, in some instances the individual archives are also covered separately by *OAIster*, e.g., *Talkbank* and *Ethnologue*.

OAIster invites participation from potential data providers, encouraging them to make their collections better known through *OAIster* and informing them about services *OAIster* will offer them if they need assistance in making their metadata OAI-enabled.

Cyclades, a registered OAI service provider, is a system designed to provide an "open collaborative virtual archive environment," which supports individual users and communities of users with the ability to conduct searches across large, heterogeneous, multidisciplinary OAI-compliant archives. It features value-added services including ad hoc or profile-based user query and browse functions; mechanisms to build meaningful collections dynamically; filtering and recommendation services; and community work areas to support collaborative work. It also provides personal document and collection storage space. Users need to read the "quick start" instructions, then register and log in to use *Cyclades*. *Cyclades* is a R&D project, sponsored by the IST (Information Society Technologies) Programme of the European Commission from November 2000 through August 2003. An impressive group of European research agencies have been involved in its development. An article by Renda and Straccia [2003/2004] about *Cyclades* is forthcoming in *Information Processing & Management* (Elsevier) but not yet available. The Web site has links to conference presentations. While *Cyclades* may point the way to the future in terms of creating and managing large personal or collaborative digital library collections, the service requires a dedicated and serious user to take full advantage of its capabilities. It currently has a user questionnaire posted at its Web site.

6.2.2 Community-based Aggregators: NDLTD Union Catalogs, OLAC, Sheet Music Consortium

Theses and Dissertations: NDLTD Union Catalogs

The *Networked Digital Library of Theses & Dissertations (NDLTD)*, founded in 1996, is a federation of more than 190 *NDLTD* members, comprised of 160 universities, 6 consortia, and 24 other institutions around the world. *NDLTD* promotes the creation, archiving, and distribution of electronic theses and dissertations. ETDs (electronic theses and dissertations) constitute a significant fraction of e-print archives and *NDLTD* has devised a new standard for metadata specific to ETDs (ETDMS), which it encourages (but does not require)

member institutions to use. *NDLTD* has also adopted use of the OAI protocol for metadata transfer. As Suleman and Fox explain [2003], *NDLTD* has developed the *NDLTD Union Archive* to function as both a provider of services (harvester) and a provider of data. Two service providers that harvest from the *NDLTD Union Archive* are the official *NDLTD Union Catalog* (hosted by VTLS) and the experimental *Electronic Theses/Dissertations OAI Union Catalog* based at OCLC. In a project known as *XTCat*, OCLC has extracted 4.3 million records of theses and dissertations (of which 8,264 are full-text) from its WorldCat database and made them available to export as an OAI data provider. (Information about *XTCat* is available at: http://alcme.oclc.org/index.html or to view *XTCat* see: http://alcme.oclc.org/ndltd/SearchbySru.html).

OCLC's experimental *ETD OAI Union Catalog* includes only those full-text records from *XTCat* plus the records of twenty other participating institutions (see site list at: http://alcme.oclc.org/ndltd/servlet/OAIHandler?verb=ListSets). The various browsing and search options are accessible at: http://www.ndltd.org/browse.html. The VTLS-hosted union catalog offers no information about its scope or coverage. The OCLC-based version indicates the institutions/sites included and the number of records each provides. Neither search interface has optimal features or "help" pages. Overall the *NDLTD* Web site contains a surprising amount of out-of-date information (e.g., notable dissertations, usage statistics, community activities). This neglect is probably temporary…as the phenomenon of ETDs is growing along with *NDLTD*'s influence.[31]

Languages: OLAC

OLAC (Open Language Archives Community) is an international partnership of institutions and individuals who are creating a worldwide virtual library of language resources by: (1) developing consensus on best current practice for the digital archiving of language resources, and (2) developing a network of interoperating repositories and services for housing and accessing such resources. *OLAC* strives to create a community of practice and it has developed a well-articulated governance structure, which includes an advisory board and a council (named in August 2003). *OLAC*'s initial development was informed by a user survey, the results of which are posted at its Web site.[32]

OLAC offers the following definitions that guide its collection policy:

> A language resource is any kind of DATA, TOOL or ADVICE pertaining to the documentation, description or analysis of a human language. Texts, recordings, dictionaries, annotations, field notebooks, software, protocols, data models, file formats, newsgroup archives and web indexes are some examples of such resources. OLAC metadata can be used to describe any

[31] See Hagen et al. [2003] report on the 2003 ETD conference. The 2003 Conference Web site is located at: http://www.hu-berlin.de/etd2003/

[32] Survey: http://www.ldc.upenn.edu/exploration/survey.html and Results: http://www.ldc.upenn.edu/exploration/survey/

> kind of language resource. Language resources may be digital or non-digital, published or restricted. A language archive is any collection of language resources and their resource descriptions. (documents/fog.html)

In operation since 2001, *OLAC* now comprises 25 archives and approximately 20,000 records, representing a range of resources from texts to software. OLAC is governed by three standards:

- the "OLAC Process standard," which defines the governing ideas of *OLAC* (its purpose, vision, and core values), the organization of *OLAC* (coordinators, advisory board, participating archives and services, etc.), and the operation of *OLAC* (how documents are generated and progress from development to proposals, to testing, to adoption, to retirement);
- the "OLAC Repositories standard," which *specifies the requirements on participating data providers, permitting their content to be successfully harvested by OLAC service providers*; and
- the "OLAC Metadata standard," which *defines the format used by OLAC for the interchange of metadata within the framework of the Open Archives Initiative* (including recommended metadata extensions).

Participating archives vary widely in size—seven of them contribute only one record and five others account for almost 80% of the content. The largest, *Ethnologue*, constitutes more than one-third of the total records. A standard template of information about each participating archive includes helpful information such as its size, the name of the institution and its "curator," a synopsis of its scope, notes about "access" (public, Web-accessible, etc.), and the date last harvested. The template of information about *Ethnologue* is reproduced below.

OLAC Template for "more information" about participating archives

Ethnologue:	**Languages of the World**
Size:	7148
RepositoryName:	Ethnologue: Languages of the World
Institution:	SIL International
ArchiveURL:	http://www.ethnologue.com
Curator:	Raymond G. Gordon, Jr.
Location:	7500 W. Camp Wisdom Rd., Dallas, TX 75236, U.S.A.
Short location:	Dallas, USA
Synopsis:	The Ethnologue data provider gives a metadata record for every language entry in the Web edition of the Ethnologue. The latter provides basic information about each of the 7,000+ modern language of the world (both living and recent ly extinct).
Access:	Every resource described by the Ethnologue data provider is a public Web page that may be accessed without restriction. Reuse of material on the site is subject to the Terms of Use that are posted.
Administrator:	gary_simons@sil.org
Base URL:	http://www.ethnologue.com/oai2.asp
Repository ID:	ethnologue.com

OAI version: 2.0
OLAC MS version(s): 1.0
Explore: Visit archive with the Repository Explorer
Last harvested: 2003-08-21

[For a complete list of participating archives see:
http://www.language-archives.org/archives.php4]

The Linguist List serves as the host of the *OLAC Union Catalog* and as the *OLAC repository editor*. The *OLAC Union Catalog* can be searched by basic keyword (searches title, description, and subject language) or by an advanced search (searches by keyword in "all" or selected archives via pull-down menu). The advanced search option has additional delimiters, at least two of which must be selected: Title; Creator; Subject Language (via a menu of language options); and Type (via a menu of types including: Annotation Tools, Datasets, Grammars, Image Data, Lexicon, Semantic & Pragmatic Analysis, etc.). Results are returned with a title and description along with a link to a fuller description. Direct links to the source site are provided from the full description. There are no post-result processing functions such as refining the search, sorting results, saving, or e-mailing results.

OLAC is exemplary in several ways: the technical and social infrastructure that it has developed to support its community of contributors, based on shared principles and standards; the resources that it provides at its Web site about its purpose, scope, history, tools, news, and events; and the efforts of its two leaders—Gary Simons and Steven Bird [2003a, 2003b, 2003c]—to articulate the challenges, analyze the options, and recommend possible solutions to their community of contributors in order to improve *OLAC*. With the formal appointment of an Outreach Working Group and its other efforts to accommodate small archives that lack technical support, *OLAC*'s content and influence is likely to grow.

Sheet Music: Sheet Music Consortium

*The **Sheet Music Consortium (SMC)** is a group of music libraries working with digital library programs in their respective institutions toward the goal of building an open collection of digitized sheet music using the Open Archives Initiative: Protocol for Metadata Harvesting (OAI: PMH).* (/OAIProject.html). Launched as a registered OAI service provider in September 2003, its founding members include Indiana University, Johns Hopkins, and UCLA. In addition, Duke University and the Music Division of the Library Congress (LC) are also data providers. The *SMC* holds nearly 100,000 records. All of LC's (47,528) and UCLA's (2,173) records have associated digital images. Johns Hopkins (11,590) and Duke (17,698) both have some digital images, while Indiana (17, 417) has none at present.

The service permits browsing (all or specified collections by title with date delimiters and sorting by title or date), basic keyword, and advanced searching. Searches can be limited to digitized sheet music only. *A keyword search retrieves text from all elements in the sheet*

music data: song titles, subjects, composer or lyricist, date, and publisher. Advanced searches provide options of combining these fields. Search tips appear directly on the search pages, facilitating use. There is also a Help page with more extensive search tips.

Results are returned with a brief record and the option to: access online, obtain more information, or add to a "Virtual Collection" with or without a note.

> *Virtual Collections is an application that allows any user to save personal collections of sheet music and attach notes to the records. The notes do not change the original record, nor do they become a part of it, except within the collection made by the user.*

Sample Search Result from the Sheet Music Consortium:

• Title: A Maiden sang to the rising moon /

Creator : Estabrooke, H. M..
Publisher : Boston: Richardson, Geo. W. 1880
Collection : Library of Congress

[access online] [more info] [Add or add with a note - add to virtual collection]

You can make a Virtual Collection without registering or signing in, but if you register with an identity and password, the Sheet Music Consortium site will store and display the collections you have saved in a separate list. As a registered user you will also be able to:

1. *lock your collection*
2. *protect the notes you have written*
3. *choose whether or not to make the contents accessible to other users*

It is possible to save collections for group use with password protection. Results saved in Virtual Collections can also be e-mailed.

6.2.3 Subject-Based Aggregators

The University of Illinois at Urbana-Champaign (UIUC) is a registered OAI data provider. It currently has four main metadata harvesting projects underway, but no planned generic gateway to its holdings.[33]

- ***UIUC Digital Gateway to Cultural Heritage Materials*** (described below)
- ***Grainger Engineering Library at University of Illinois at Urbana-Champaign***, aggregation for science and engineering (described below)
- ***IMLS Digital Collections and Content (DCC)***, a three-year effort at the University of Illinois to build a national infrastructure for adaptable, interoperable, and sustainable digital collections, which includes using OAI-PMH to harvest metadata from current and past National Leadership Grant (NLG) awardees with digital collections (up to about 100 potential providers). Launched

[33] Information from e-mail correspondence with Timothy Cole on July 28, 2003.

in January 2003, this cooperative agreement prohibits including the NLG repository in a general aggregator at this time, although some individual awardees have been given separate permission to include their metadata in UIUC's cultural heritage aggregation. See: http://imlsdcc.grainger.uiuc.edu/

- A collaboration with ten member libraries from the Midwest Committee on Institutional Cooperation (CIC), which will result in an OAI-PMH metadata harvest hosted at UIUC. The CIC collaboration will involve looking at a variety of metadata harvesting issues of primary interest to the consortium, including the use of restricted access/CIC-licensed metadata. Some sets from CIC providers are also included in UIUC's cultural heritage or science and engineering aggregations.

Cultural Heritage: UIUC Digital Gateway to Cultural Heritage Materials

The ***UIUC Digital Gateway to Cultural Heritage Materials***, like *OAIster*, received its initial funding from the Andrew W. Mellon Foundation. It currently consists of the holdings from 25 OAI-compliant metadata providers and contains over 400,000 records. The site provides an annotated list of the collections covered, organized both by the name of the collection and by the type of material contributed (images; text, sheet music, Web sites; and museums and archives.) These "types" also correspond to browsing options and search-display options. At present, texts represent over 260,000 of the records, with images and video accounting for about 80,000, and museums and archives for 12,000. Shreeves et al. [2003] describe at length their experience in aggregating metadata records that originate from different communities, and describe heterogeneous collections of resources. They include a discussion about normalizing metadata to enable users to get consistent and predictable results, especially by type of materials and dates. Ultimately, the reduction of type of material to three broad categories, while necessary in order to traverse diverse collections, is crude in comparison to the options provided by many of its component collections, e.g., the Library of Congress *American Memory* service has more options to search by format (three document types—manuscripts, printed texts, or sheet music; maps; motion pictures; photos/prints; and sound recordings) as well as by "user format" (hear, read, or view).

The UIUC aggregation, *unlike* OAIster, includes metadata for resources with collection-level descriptions only as well as for some analog items. In order to distinguish between collection-level records and actual digital objects, two different labels are used in the results display. Collection-level records are labeled with a link to, "Learn more about this item," whereas those with a direct link to the digitized object are labeled, "View Item." Records without links don't have any labels. For example, if the user searchers for "Harlem nocturne" at the UIUC site, it retrieves two hits: one to Indiana University's DeVincent Sheet Music Collection, which links to that collection site via "learn more about this item" and the other to a print copy

in the UIUC's library collection without any link. Users can limit their search to "online primary sources" (when applied to the search above, only the Indiana University (IU) record is retrieved). While this is a valuable feature of particular relevance to cultural heritage materials, in this particular instance, it is somewhat misleading because IU's version is also a print copy with only the bibliographic record available online.

Like *OAIster*, UIUC's site provides simple search and advanced search functions from a single page. Advanced searches permit terms in "any field" or limited by author/artist and/or title/subject. Again, important to its particular community of users, the UIUC site offers "Date range delimiters" as well as limits by type of material. There is a "search history" function, but no "revise search" capability. Results default to a short display with a link to the full record. As noted above, results can be sorted by "type" of materials. The *UIUC Gateway* does *not* return results by collection with a left frame navigation bar. Unlike *Arc* and *OAIster*, the *UIUC Gateway* makes it possible to save and/or download records with a "bookbag" option. Help is available at a separate Web page, and provides further information about searching, viewing, and saving results. It also gives a helpful table outlining how different fields of data are indexed. An e-mail address is provided for additional questions.

Sciences: Grainger Engineering Library at UIUC, Citebase, Archon

UIUC's other metadata-harvesting service covers science and engineering resources via the ***Grainger Engineering Library at UIUC***. This service aggregates data from 12 repositories and contains 443,131 records as of August 7, 2003. (The status of the latest OAI harvests, which are performed frequently, is easily available via a link at the bottom of the site's search page.) *ArXiv* constitutes more than half of the record count with the Institute of Physics journals (IOP) accounting for another 25% of the holdings. This site predominantly provides access to scientific e-prints, technical reports, theses and dissertations, and e-journals collections. At present this service is intended primarily for local institutional use. As a result, it does not provide any context or documentation about its mission, scope of operation, or collection policy. Its search interface and functionality has many of the basic features of the cultural heritage gateway, but modified for its clientele. It is possible to search the database by author/editor, title/subject/abstract (collectively or separately), report number/journal source, publisher, date, or language. The search can be limited to "all" or specified individual collections. There is a pre-search option to sort by relevance or collection. It has a post-search "modify search" function as well as the ability to save or download records into a "bookbag." There is no "Help" page although an email address is provided for comments.

In contrast to Grainger's reliable and up-to-date, but no-frills, utilitarian service, two experimental science aggregators are under development that feature extended services: *Citebase* and *Archon*.

Citebase, an OAI registered service provider, is under development by the University of Southampton (originator of GNU eprints.org Open Source software, which was first supported by *Cogprints* and is now used by many other e-print archives).[34] *Citebase*, along with its companion Open Citation Project (OpCit), which supports reference linking and citation analysis, is being developed to facilitate the self-archiving movement.[35] Citebase will search across multiple archives—presently these include *arXiv*, *Cogprints*, and *BioMed Central*—with results ranked according to various criteria, including citation (author or paper), date (created or updated), or hits (author or paper).

Archon, funded by NSF as part of the *NSDL*, is a collaborative project of Old Dominion University, the American Physical Society, and Los Alamos National Laboratory in concert with the *CERN Document Server* (http://cds.cern.ch/). *Archon* identifies itself as a "digital library that federates physics collections with varying degrees of metadata richness." In its present state it is considered here as a cross-archive search service and aggregator rather than as a full-service digital library environment. *Archon* presently federates holdings from five archives groups:

- *arXiv*
- *Physical Review D* from the American Physical Society (http://prd.aps.org)
- selected records from *CERN* (a service with over 550,000 bibliographic records, including 220,000 full-text documents related to particle physics)
- NASA's *NTRS* (technical reports)
- from the *Emilio Segre Visual Archive* (http://www.aip.org/history/esva/), historical images in the history of physics

Archon supports both the DP9 Gateway (open source gateway service that allows general search engines, like Google, to index OAI-compliant archives)[36], and Vac Gateway (in progress, a gateway service to harvest non-OAI collections into OAI-compliant repository). *Archon* uses an enhanced version of *Arc*'s harvester and search engine, with added functionality for equations-based and formulae searches that are important to physicists. It also supports extended services such as cross-reference linking and citation ranking. Even simple search results give the user an option to link to "show equations," "similar subjects," or "citations." There are a variety of post-result processing options including the capability to: re-organize the result set by grouping (by archive, date, or subject) or sorting (by archive, date, subject, or title); refine the result set by author/subject/title, or abstract; or refine the result set by discovery date. *Ar-*

---

[34] 72 archives are using GNU e-prints.org software worldwide. For more information and a listing see: http://www.software.eprints.org, accessed on August 7, 2003.

[35] Information about OpCit is located at: http://opcit.eprints.org/, accessed on August 7, 2003. A FAQ on the self-archiving movement is located at http://www.eprints.org/self-faq/, accessed on August 7, 2003.

[36] For further information about DP9 refer to: http://egbert.cs.odu.edu/dp9/, accessed on August 7, 2003.

chon demonstrates how effectively *Arc* can be refined and enhanced to serve a particular constituency, while also leveraging the research of *Citebase* and OpenURL.[37] Maly et al. [2002] give an account of *Archon*'s technical architecture as well as its future development plans.

6.2.4 Summary of Issues

All of these services have been created in the last few years since the inception of the first version of the OAI-PMH, although the *NDLTD* federation itself predates it. Outside of conference presentations and reports to sponsoring agencies, published articles are only starting to appear. In virtually all cases, these are presented or written by those involved in the projects, most often concentrating on technical issues, such as ways in which the metadata has been "normalized" to create more effective searches or design and functionality issues, often informed by user surveys or user testing. *Archon*, *Cyclades*, and *OLAC* have strong ties to, and impressive support from, their respective research communities. Steven Bird's and Gary Simons (*OLAC*) research projects and publications in the areas of linguistic annotation, digital archives, and language documentation demonstrate how digital tools and resources are interconnected and made manifest to support linguists worldwide.[38] Apart from this complex and multi-faceted effort, most other disciplinary differences are revealed primarily in special search features or fields, e.g., *Archon* permits searching for formulae; *UIUC's Digital Gateway* limits searches by type of media; the *Sheet Music Consortium* allows searching by composer or lyricist, and *OLAC* offers searches by language.

Several of these services, as they are further developed, point the way towards creating personalized digital libraries. Turning again to Europe, the *TORII* prototype designed and implemented as part of the European Union's IST program, shares much in common with *Archon* in terms of content and future aspirations.[39] *TORII*, a registered OAI service provider, is designed to serve as a single environment from which the following open archives can be accessed: *arXiv*, *BioMed Central*, the Mathematics, Computer Science and Chemistry preprint servers maintained by Elsevier, and the CERN Document Server. *TORII*, like *Cyclades*, and at a more rudimentary level, the *Sheet Music Consortium*, demonstrates the potential to create personal collections. After registering, *TORII* makes available more advanced features including personal folders to store documents, defining a profile of interests, which the system uses to return search results in relevance order according to user preferences. As a first step to implement a community network of quality control tools, *TORII* permits registered users to evaluate any of its documents.[40]

---

[37] For further information about OpenURL refer to NISO: http://www.niso.org/news/releases/pr-OpenURL.html

[38] See: Birds' Research project at: http://www.ldc.upenn.edu/sb/home/projects.html and Publications: http://www.ldc.upenn.edu/sb/home/publications.html and Simons' selected Publications at: http://www.ethnologue.com/show_author.asp?auth=Simons%2C+Gary+F%2E

[39] Access TORII at: http://torii.sissa.it

[40] A guide explaining TORII is located at: http://tips.sissa.it/docs/booklet.pdf, accessed on August 23, 2003.

Most of these services are too new or experimental to have received much attention from regular users. Since they are all evolving and improving rapidly, it is imprudent to be too critical. With the exception of *Cyclades*, where users need to read a tutorial and register before using the service, all of these services can be readily accessed and searched. From my perspective, there are three overarching concerns:

- Having sufficient data to make the service worthwhile to use.
- Providing the user with sufficient information so they understand the scope and currency of coverage. For example: "What results will be retrieved: links to the source collection-level only, direct links to digital objects, links to analog objects, links to resources available to restricted users?"
- Providing the user with a "context" in which to understand the items retrieved, i.e. items are detached from their richer original-source native environment. From what original collection is the item derived and how can it be accessed?

A selection of the best features from this suite of services.

Informative and user-friendly home page

OAIster
http://oaister.umdl.umich.edu/

OLAC
http://www.language-archives.org

Service put into context

OAIster
http://oaister.umdl.umich.edu/o/oaister/description.html

OLAC
http://www.language-archives.org/documents.html

UIUC Gateway
http://oai.grainger.uiuc.edu/

Most organizational members

NDLTD
http://tennessee.cc.vt.edu/~lming/cgi-bin/ODL/nm-ui/members/index.htm

Largest number of harvested archives and most records

Arc
http://arc.cs.odu.edu:8080/oai/admin.jsp

Extensive information about the mission, vision, and governing structure of the service

OLAC
http://www.language-archives.org/organization.html

Tools and services for potential contributors

OLAC
http://www.language-archives.org/tools.html

Descriptions of participating archives or collections

OAIster for its annotations about collections and indication of the number of records
http://oaister.umdl.umich.edu/o/oaister/viewcolls.html

OLAC for its template of information about each archive (see "more information")
http://www.language-archives.org/archives.php4

Frequent harvesting of metadata

Grainger Engineering Library
Check the status of latest OAI harvests
http://g118.grainger.uiuc.edu/engroai/LastHarvest.asp

Explaining duplicate records

OAIster for its explanation to users
http://oaister.umdl.umich.edu/cgi/b/bib/bib-idx?c=oaister;page=simple

Distinguishing bibliographic records about collections or items from actual digital objects

UIUC Gateway for search feature that permits limiting to "primary online sources" only and for returning other results as "view this collection" versus "view this item"
http://nergal.grainger.uiuc.edu/cgi/b/bib/bib-idx

Sheet Music Consortium for search feature that permits limiting to "digital sheet music"
http://digital.library.ucla.edu/sheetmusic/librarian?SEARCHPAGE&Search

Search tips on the search screen

Sheet Music Consortium
http://digital.library.ucla.edu/sheetmusic/librarian?SEARCHPAGE&AdvSearch

Help page

OAIster
http://oaister.umdl.umich.edu/o/oaister/help.html

Advanced Search filter and display options

Archon including interactive subject selection
http://mercury.seven.research.odu.edu/archon/advanced_search.jsp#

Post-results processing

Archon for its options to link to equations, similar subjects, or citations, as well as ability to reorganize, sort, or refine result sets

Explanation of search improvements

OAIster: Search Improvements
http://oaister.umdl.umich.edu/o/oaister/phase2.html

"Using OAI-PMH to Aggregate Metadata Describing Cultural Heritage Resources," by Timothy W. Cole, University of Illinois at Urbana-Champaign, ALA/CLA Annual Meeting, June 22, 2003, Toronto. Includes description of how search functions were changed based on pilot study with 23 Curriculum & Instruction student teachers.
http://dli.grainger.uiuc.edu/Publications/TWCole/ALA2003OAI/ALA2003_OAI.ppt

Building personal or group collections

Cyclades (must log in to see system)
http://www.ercim.org/cyclades/index.html

Sheet Music Consortium (Virtual Collections tutorial)
http://digital.library.ucla.edu/sheetmusic/help.jsp#virtual_collections

Extended search services

Archon for cross-reference linking and citation ranking

Citebase for its ranking options

Cyclades for its recommendation services and ability to create collections
http://www.ercim.org/cyclades/overview.html

Sheet Music Consortium for ability to annotate and save records
http://digital.library.ucla.edu/sheetmusic/help.jsp#virtual_collections

Potential to transform scholarly collaboration

Cyclades ("Quick Start" tutorial)
http://www.fit.fraunhofer.de/projekte/cyclades/quickstart/

Documents about the service (progress reports, standards, publications)

Archon
http://archon.cs.odu.edu/publications.html

Cyclades
http://www.ercim.org/cyclades/pub.html

OAIster
http://oaister.umdl.umich.edu/o/oaister/reports.html

OLAC
http://www.language-archives.org/documents.html

UIUC Gateway
http://oai.grainger.uiuc.edu/presentations.htm

6.3 From Digital Collections to Digital Library Environments

FROM DIGITAL COLLECTIONS TO DIGITAL LIBRARY ENVIRONMENTS	CULTURAL HERITAGE COLLECTIONS
❏ Collection of tools that make content alive ❏ Help user find content, manipulate, analyze, annotate, and comment on it ❏ Attract, create, define a community ❏ *Collaboratories* where active group annotation, analysis, and creation of new knowledge happens ❏ Enable and facilitate implicit communication (e.g., recommender systems) ❏ Sum greater than its parts [Lynch 2002]	American Memory Colorado Heritage (Colorado Digitization Program) **HUMANITIES** The Perseus Digital Library **SCIENCES** National Science Digital Library ❏ **FEDERATION** SMETE Digital Library (Science, Math, Engineering & Technology Education Digital Library) ❏ **K-12 TEACHER SUPPORT** ENC Online (Eisenhower National Clearinghouse for Mathematics and Science Education) ❏ **BIOLOGY NODE** BEN: A Digital Library of the Biological Sciences for Biology Teaching ❏ **EARTH SCIENCES NODE** DLESE: Digital Library for Earth System Education

This category spans the scope from digitized collections to digital library environments. These services are typically involved in content creation, originally with digitized collections at their core, but they also represent an increasingly sophisticated suite of functions and services in support of users. In comparison to the previous category, they have a collections-driven focus and many of them predate the inception of the OAI-PMH. As a result they represent a mix of OAI and non-OAI-compliant metadata. Although four of the services listed here are components of the *National Science Digital Library (NSDL)*, all of them represent sophisticated independent services in their own right. Some of them start to cross the boundaries between

digital library environments and digital learning environments. They frequently include such features as: online newsletters, e-mail alert services, online reference assistance, tools to analyze data or use collections, and opportunities for collaboration or professional development. Lynch [2002] describes the trajectory from digitized collections to digital libraries. This report also draws on the definition of "digital libraries" offered by the Joint Committee on Digital Libraries:

> JCDL encompasses the many meanings of the term "digital libraries", including (but not limited to) new forms of information institutions; operational information systems with all manner of digital content; new means of selecting, collecting, organizing, and distributing digital content; and theoretical models of information media, including document genres and electronic publishing. Digital libraries are distinguished from information retrieval systems because they include more types of media, provide additional functionality and services, and include other stages of the information life cycle, from creation through use. Digital libraries also can be viewed as a new form of information institution or as an extension of the services libraries currently provide. [Retrieved from: http://www.jcdl.org/about-jcdl.shtml, accessed on September 4, 2003]

6.3.1 Cultural Heritage: American Memory and Heritage Colorado

Begun in 1995 after a five-year pilot project, ***American Memory*** is a corpus of electronic versions of the Library of Congress's (LC) archival collections related to the nation's cultural heritage. From its inception, *American Memory* intended to make collections not only accessible, but also *useable*. It conducted an extensive user evaluation in the early 1990s in order to determine its core audience: students, researchers, and educators.[41] Its selection of materials is based on cultural and educational value, expected demand, input from the NDL (National Digital Library) Advisory Committee, and the ability of current technology to capture the content. It currently consists of more than 100 collections and over 7 million digital items. From 1996 through 1999, LC ran a competition funded by Ameritech to create digital collections of primary resources from other libraries, museums, historical societies, and archival institutions nationwide. Twenty-three collections received these awards and became integral to *American Memory*. Collections cover a wide spectrum of types of media and also vary greatly in their scope—from a digital version of a World War I newspaper to a gateway of resources in women's history.

At its core, *American Memory* has three main features—a "Collection Finder" that describes all collections, a "Search" function that extends across all or selected collections, and a "Learning Page" that connects collections to ideas for teaching and learning. Through the Collection Finder, users can select one of fourteen broad topics by

---

[41] Final Report of the American Memory User Evaluation (1991-1993) http://lcweb2.loc.gov/ammem/usereval.html

which to limit a search or link directly to that collection. The Collection Finder also identifies collections by "User's format" (e.g., hear, read, or view), time period, place, LC Library Division, or digital format (e.g., jpeg, pdf, QuickTime, RealMedia, etc.). The collections themselves have a common screen design that presents information under the categories of: "Understanding the Collection," "Working with the Collection," and "from The Learning Page." The Learning Page provides contextual material, search help, sample lesson plans and activities, special presentations, and descriptions of the digital collections for K-12 school teachers and media specialists. There is a "community center" that features monthly thematic live discussions about using primary resources and a subscription-based e-mail update service. As explained by "What American Memory resources are included in this search?":

> Searches that begin from the American Memory Collections: Search All Collections search page or from any Collection Finder search page include detailed bibliographic records about most items. The full text for items in some collections is also included.
>
> Within individual collections, additional options are available for searching or for browsing lists of names, places, or subjects (as appropriate for each collection). To use these features, follow links from the collection's home page.
>
> Not included in any American Memory search are the collection Home Pages, background texts and illustrations, and the Learning Page texts. These may be searched, along with other Library of Congress texts, through the Library of Congress Search/Browse page. "Today in History Archive" can be searched separately.

There is a notable list of exceptions to these guidelines, including two collections that are not searchable at all (except within the collections themselves) and an explanation of why searches for specific format types (e.g., photographs, maps) may not always find all of the items a user is seeking or may return items in a variety of formats. This provides the serious user with clues about how to tailor her search to overcome these variances. Many textual collections give the option of searching for bibliographic records only or for the full text. Results from full-text searches can be ranked in two ways. A "Search Tips" link explains search functions in general. Meanwhile, due to the increasing complexity of this service, more focused search guides, such as the one related to women's history materials, are necessary to fully exploit *American Memory*. [42] There is no functionality to save, e-mail, or download search results. *American Memory* is exemplary in its coherent design which gives a common "look and feel" to its contents and keeps the user within the "context" of its resources.

LC has registered its OAI server for *American Memory* collections

[42] See the guide at: http://lcweb2.loc.gov/ammem/awhhtml/awsearcham.html

as a data provider and RLG's *Cultural Materials*, *OAIster*, *UIUC's Digital Gateway to Cultural Heritage Materials*, the *Sheet Music Consortium* and *Perseus Digital Library* all harvest records from *American Memory*. As of August 2003, more than 136,000 item-level digital representations were made available through LC.[43] Caroline Arms's [2003] discussion of LC's experiences with OAI-PMH is valuable because she compares how other services—in particular, *Perseus* and RLG's *Cultural Materials*—have used LC's records to enhance their services. Arms points out the special features of RLG's (proprietary) service that allow "users to switch easily between different structural views of the current result set enabling different browsing strategies and different approaches to successive refinement of a search." She also notes: "every item is represented by a thumbnail for visual browsing" and "explicit modeling of parent-child relationships facilitates navigation from collection records to item records and vice versa." She suggests—quite rightly—that users may respond to RLG's enhanced interface and features more favorably than to the traditional approaches of *OAIster*, the *UIUC Gateway* and even *American Memory* itself. She wonders aloud if thumbnails will become a standard component of metadata records.[44] At the same time, she acknowledges the costs and trade-offs involved in balancing quality and quantity—recognizing the need for collaboration among service providers.

The Colorado Digitization Program (CDP), ***Heritage Colorado***, was established in 1998 to provide the people of Colorado with online access to cultural, historical, and scientific resources through the collaborative effort of Colorado's archives, historical societies, libraries, and museums. The *CDP* operates with a board of directors and five working groups (for collection development, digital audio, metadata standards, scanning standards, and scanning centers) that coordinate and guide the implementation of its projects. Participating organizations that apply for membership receive a number of benefits including a reduction in various service fees. Membership fees are based on the size of the organization's operating budget and range from $200 to $2500 annually.[45] The *CDP* is also funded by the Colorado Department of Education in partnership with the Colorado Virtual Library and with additional financial support from the Institute of Museum and Library Services (IMLS) and other granting agencies.

Heritage Colorado has an exemplary collection development policy that covers its guiding principles, defines its audience (five categories of users), its subject matter, and formats.[46] In addition, the policy spells out who can contribute, the criteria for adding resources, criteria local sites should consider when starting a digitization project, ownership issues, and accuracy of data. It is one of few sites

---

[43] A list of *American Memory's* OAI-compliant collections (last updated March 10, 2003) is available at: http://memory.loc.gov/ammem/oamh/

[44] C. Arms [2003]: 137.

[45] Membership information and applications are available at: http://www.cdpheritage.org/about/project_membership.html

[46] The Collection Development Policy is available at: http://www.cdpheritage.org/about/policy_collection.html

reviewed to include a statement about the grounds for removal of a site and the ensuing appeal process.

"Market segments and their information needs" further defines the *CDP*'s five user categories—general/casual user, student and lifelong learner, hobbyist, scholar/researcher, and business community—and specifies their content interests, along with their respective design and retrieval preferences.[47] This definition of audience and their needs guides the development of *Heritage Colorado*.

To have resources included in *Heritage Colorado*, participating institutions must not only demonstrate their commitment to the principles of the *CDP,* but also contribute metadata to the *CDP Union Catalog* and have a plan for the ongoing sustainability of the collection. Although *CDP* is based on a model of distributed images and centralized metadata, it has devised an Image Storage Policy whereby it encourages participating institutions to store Master images with the *CDP* in order to manage data migration and upgrades. [48]

Heritage Colorado's major collections consist of "Western Trails," "Colorado Main Streets," and projects by region. They can be browsed in nine different subject categories or searched. Sample searches are provided for each category. Advanced searches permit the combination of terms by a variety of fields including keyword, author, title, subject, language, and project. Searches can be renewed or refined and records can be saved or e-mailed. Like *American Memory*, *Heritage Colorado* features a special section for educators, which includes lesson plans, workshops, and tools to use the collections.

Access is provided via a Z39.50 compliant system to more than 150,000 digital objects with metadata hosted on two systems, the *Heritage Colorado* database and the Denver Public Library (DPL) system. The *Heritage Colorado* system with about 20,000 metadata records is OAI-compliant, but the DPL is not.[49] The *UIUC Digital Gateway* and *OAIster* harvest metadata from *Heritage Colorado*.

6.3.2 Humanities: The Perseus Digital Library

The Perseus Digital Library, launched in 1995 with antecedents dating to the mid-1980s, describes itself as an "evolving digital library of resources for the study of the humanities." It is a non-profit enterprise located in the Department of the Classics at Tufts University and funded by the Digital Libraries Initiative Phase 2, the National Endowment for the Humanities, the National Science Foundation, private donations, and Tufts University. According to its "FAQ,"

> Perseus is funded to perform research on developing tools to provide users with improved access to various types of materials. Past work has focused on building and linking together collections. Current work considers ways of developing and

[47] "Market segments and their information needs" is located at: http://www.cdpheritage.org/resource/reports/rsrc_users.html

[48] The "Image Storage Policy" is located at: http://www.cdpheritage.org/about/documents/policy_imagestorage_2001.pdf

[49] Based on e-mail correspondence with CDP's Executive Director Liz Bishoff of July 28, 2003.

> refining tools for presentation of the materials in the Perseus DL. We are primarily a research project, although we do incorporate services for our audience. [50]

Its original scope—to construct a large, heterogeneous collection of materials, textual and visual, on the Archaic and Classical Greek world—has expanded to other areas such as the Renaissance and the history of London. In addition to gathering materials, *Perseus* builds specialized searching and indexing tools to facilitate the exploration of its collections.[51]

Perseus is one of few resources to register as both an OAI data and service provider. As such, it has harvested collections on California, the Upper Midwest, and the Chesapeake from *American Memory,* and is experimenting with automatically generated maps and timelines as a means to visualize the contents of these collections.[52] A search of the full text of these documents also provides access to thumbnail images through the *Perseus Image Browser*. Caroline Arms [2003] describes how *Perseus* links highlighted words or phrases from *American Memory* to other reference texts within *Perseus*.

Crane [et al. 2003] describes how *Perseus* has evolved over the past fifteen years to serve more diverse audiences and to develop specialized services to meet their needs. He notes: "The emerging challenge for digital libraries seems to be multisource, customized summarization: a DL system should be able to determine what supporting information a particular user would require to understand a particular piece of information."[53] Crane then enumerates various basic services (document chunking and navigation services, visualization tools, citation linking, etc.) required for such a system.

Together with Johns Hopkins University, Tufts was awarded a NSF grant, effective January 2003, which will extend some of the link generation tools of *Perseus* and apply them to support all levels of reading in the *NSDL*. As outlined in the proposal abstract, "Services for Customizable Authority Linking Environment" (SCALE) will automatically bind keyword and phrases to supplementary information. To elaborate:

> Much of the work at the Tufts University Perseus Digital Library Project (http://www.perseus.tufts.edu/) and the Johns Hopkins Digital Knowledge Center (http://dkc.mse.jhu.edu/) has already focused on exploiting various kinds of authority lists (gazetteers, biographical dictionaries, dictionaries, glossaries of technical terms, and name authority files) for the automatic generation of hypertext links and for visualizations such as automatically generated dynamic maps and timelines. Such link generation complements the current practice of automatic identification

---

[50] *Perseus* maintains a bibliography of research articles written by its staff, located at: http://www.perseus.tufts.edu/Articles/index.html

[51] These tools, e.g., the Art & Archeology Browser, the Atlas Tool, the Lookup Tool, the Greek Vocabulary Tool, are listed at: http://www.perseus.tufts.edu/cgi-bin/perscoll?collection=Perseus:collection:PersInfo&type=interactive+resource

[52] For more information about this and other collaborations see: http://www.perseus.tufts.edu/collab.html

[53] Crane et al. [2003]: 78.

and aggregation of citations. The current project augments and transfers the existing technology for managing authority lists, converting this from a research effort to an institutionalized service serving a wider community.[54]

6.3.3 Sciences: NSDL, SMETE, ENC, BEN, DLESE

NSDL (National Science Mathematics Engineering & Technology Education Digital Library) is a digital library of exemplary resource collections and services, organized in support of science education at all levels. Starting with a partnership of *NSDL*-funded projects, *NSDL* is emerging as a center of innovation in digital libraries as applied to education, and a community center for groups focused on digital-library-enabled science education.

The *NSDL* arose from the recommendations of a 1996 National Science Foundation (NSF) report as a way to improve undergraduate education in science, mathematics, engineering, and technology (so-called "SMET" education).[55] After a series of national workshops and prototype projects that help to build a technological, disciplinary, and community base exemplified by its precursors—*DLESE* and *SMETE*—the *NSDL* was lodged in NSF's Division for Undergraduate Education (DUE) and began its first formal funding cycle in 2000.[56] To date, it has made 121 awards for projects in four areas: collections (66 projects), services (35 projects), targeted research (11 projects), and core integration (9 projects). Despite its programmatic affiliation with undergraduate education, *NSDL* aims to reach a "K to gray" audience and to serve all those with an interest in improving science literacy—a community aggregated from disciplinary groups, educational groups, technology and information science groups, special interest groups (policy-makers, journalists, commercial sector), and learners of all kinds (from students to citizens-at-large).[57]

NSDL is now a complex network of libraries within libraries. It provides access to a wide array of collections and user services, while also supporting the needs of developers by providing the workspace and tools for digital library development through its "Communication Portal." Maintained by the "Core Integration" team, the "Communication Portal" links to information about the *NSDL*'s governance structure and working committees, including their activities, reports, tools, and news. The *NSDL*'s monthly newsletter, "Whiteboard Report," is accessible from this portal or you can subscribe to receive it via e-mail. The "Collaboration Finder," developed in partnership with *SMETE*, *MERLOT*, and the Merit Network, is a useful tool to identify the individual projects funded by the

---

[54] The NSF award abstract is located at: https://www.fastlane.nsf.gov/servlet/showaward?award=0226304

[55] See NSF, "Shaping the Future" [1996].

[56] See "Key Reports and Background Materials" about the *NSDL* at the NSF site: http://www.ehr.nsf.gov/ehr/DUE/programs/nsdl/reports.asp, accessed on August 27, 2003.

[57] See the *NSDL's* first year report, "Pathways to Progress: Vision and Plans for Developing the NSDL," March 20, 2001, p. 13; at http://doclib.comm.nsdlib.org/PathwaysToProgress.pdf, accessed on August 27, 2003.

[58] Collaboration Finder is located at: http://www.smete.org/smete/nsdl/collabfinder/

NSDL program and to learn more about their specific activities.[58] For example, you can search for all "physics" projects funded under the "collections" track in 2002. From the results' list, it is possible to link to specific projects and view a template of basic information about the project, as well as review its specific activities, progress reports, and the status of its deliverables. Although intended for developers, the "Collaboration Finder" is helpful to users in identifying *NSDL* collections or ascertaining the status of *NSDL* service projects. The "Document Library" in the Communication Portal contains key reports, access to information about the governing structure, and a spreadsheet of all *NSDL*-funded projects.[59]

Returning to the main *NSDL* site, it is currently available in its "initial version," with changes expected in October 2003. According to *NSDL*'s communication director, the new version will look very different and have an updated search engine. In its present state and with a collections policy only released in a draft form in 2003, it is difficult to ascertain the size, scope, and coverage of the *NSDL*.[60] *NSDL* only has two broad "filters" that serve as criteria for inclusion:

- Relevance to any aspect of Science, Technology, Engineering, and Mathematics Education at all levels of learners.
- Basic integrity of the resources in the collection. (Does it function reasonably? i.e., no blatant technical failures of the digital resource.)

The site's online glossary provides the following definition of Collections:

> Similar to museum and library collections NSDL collections are organized arrangements of items. An NSDL collection may have been organized by a person or organization, or may be collected automatically by the NSDL. Meanwhile "Items" are defined as, "an item is a unit of a collection. It may be large or small and it may itself contain parts or smaller units. Every item in the NSDL has an association with a collection."

According to *NSDL* staff, as of mid-August 2003, *NSDL* comprised 199 collections of which 42—about 18 of them NSF-funded NSDL "collections" projects—have individually analyzed item records. This translates into some 301,702 items with full content or direct links to digital objects with 204,888 derived from one source—*arXiv*. It is noteworthy that until the 2003 cycle of NSF funding, there was no requirement that collections had to be OAI-harvestable. NSF does have a Metadata Primer available to contributors and provides tools for automated ways to provide metadata using OAI.[61] Other

---

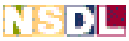
[59] Document Library is located at: http://doclib.comm.nsdlib.org/cgi-bin/wiki.pl

[60] Draft *NSDL Collection Policy:* http://content.comm.nsdlib.org/doc_tracker/docs_download.php?id=452

[61] *NSDL Metadata Primer* is located at: http://metamanagement.comm.nsdlib.org/outline.html and the Collection Metadata Form is available at: http://metamanagement.comm.nsdlib.org/collection_form.html Both are accessible from the Metadata Management page: http://metamanagement.comm.nsdlib.org

collections are obtained by harvesting OAI-2 compliant records outside the NSF-funded initiatives, either upon recommendation of the *NSDL* collections working group or gathered via Web-crawler technology—as illustrated by the "collection" entry that follows:

Collection Entry retrieved by Browsing by Topic:

> *Science, Mathematics, Technology and Engineering Resources Gathered by the National Science Digital Library* [No link available]
> A collection of materials gathered via web-crawler technology, for which not much is known regarding quality or level of appropriateness.
> 

Collections selected by *NSDL* are marked with its logo, whereas collections that are funded by or officially part of *NSDL* carry their own branding. At present, these distinctions are not intuitively obvious to users—in fact, users might assume that the *NSDL* logo indicates a collection that was funded by NSF, instead of the reverse. Results of searches are also returned with these collection logos. Collections can be browsed by topic, organized according to The Gateway to Educational Materials (GEM) main topics, and subcategories. It may come as a surprise that sources related to health, nutrition, and medicine are part of *NSDL*. At present, some categories have no entries.

The site features basic searches by keyword with the ability to limit by type of resource (collections, items, news, exhibits, collections with reviews, items with reviews) or by format (text, image, audio, video, interactive, data), as well as by Boolean operators limited to keyword anywhere, keyword in content, title, author/creator/contributor, subject, and format/genre. It is not possible to limit by audience or grade level. Results identify "resource format" and the collection in which it was found. "More information" provides an annotated record with descriptors.

There have been a number of improvements since the new search engine was launched in July and more refinements are anticipated in October. A few of the problems currently encountered:

- There are broken links.
- There are duplicate records.
- There is no explanation of search protocols, e.g., for phrase searches.
- There is no capability to sort results.
- Some featured collections link to sites where users must pay to obtain information—they must be authenticated or register to access resources—inhibiting or slowing down navigation through various services.
- Sending a message to feedback@nsdl.org, as suggested in the "help" menu, obtained the following reply: "Your feedback has been received by the staff of the National Science Digital Library.

> Thank you for taking the time to send us your comments. Please note that we are not able to answer individual questions via this feedback mechanism."

User registration and login are optional, but required for access to certain services, including the ability to access *AskNSDL*. According to the site, *NSDL* Registration and Login are the equivalent to applying for and receiving a library card. Users register and login within the *NSDL* Access Management System, enabling them to be "recognized" by the *NSDL* and its associated services. Later releases of the *NSDL* will include customization and personalization options available only to logged-in users. Users can register or login at any point in a session.

Some of the anticipated user enhancements include a "My Preferred Collections" service that will bookmark collections from *NSDL* searches and limit searches to those collections. The "My Site" service will guide users through entering, storing, retrieving, editing, and publishing personal information pages. Users will enter text in a simple web form adding text and images to their pages.

Clearly the *NSDL* is a vast and ambitious undertaking. According to the five-year planning targets for the scale of the *NSDL*, it aims to have 1 million users, 10 million digital objects, and 10,000 to 100,000 collections by 2007.[62] Its future success may hinge on how successful it is at devising specialized portals—it has several under development—to meet the needs of targeted audiences. It now defines its audience as: the generally curious (interested in science and research information), the NSDL developer community and partners, and funding agencies and supporters. Each of these has widely different needs and expectations. Some of these concerns will be addressed when the redesigned site debuts in October.

D-Lib Magazine, which is funded by NSF, regularly carries progress reports about the NSDL.[63] Williams Arms [2003] discusses the NSDL architecture for metadata harvesting in a 2003 issue of *Library High Tech*, devoted to the Open Archives Initiative. Roy Tennant devoted his March 15, 2003 column in *Library Journal* to "Science Portals," in which he discusses *NSDL* and *Science.gov*. Dean Johnston reports on his experience in using the *NSDL* and some of its affiliates, including *MERLOT* and *DLESE* in the July 2003 issue of the *Journal of Chemical Education*. He concludes:

> The official National Science Digital Library… has the potential to become a one-stop site for a wide range of educational science material. The advanced search tools are quite detailed, but at this early time the site suffers from a lack of content. Clearly as more digital library collections come online, this will become an invaluable tool for science educators at all levels. [64]

---

[62] Arms, W. et al. [2002].

[63] See articles of October 2000, March 2001, November 2001, January 2002, and November 2002, searchable at *D-Lib Magazine*: http://www.dlib.org

[64] Johnston [2003]: 733.

Although the next four resources—*SMETE, ENC, BEN* and *DLESE*—are all multi-faceted independent services in their own right, they are considered here primarily in the comparative context of their affiliation with the *NSDL*. They all aim to support science education.

Federation: SMETE

SMETE is an "open federation community" built with funding from NSF's *NSDL* program that aims to serve as an "integrative organization" and "gateway to a comprehensive collection of science, math, engineering and technology (SMET) educational content." Presently an unincorporated entity, *SMETE* is a membership organization that includes more than forty partners, such as the American Association for the Advancement of Science (AAAS), the Coalition of Networked Information (CNI), and OCLC, as well as other digital libraries dedicated to science education, including *BEN*, *ENC*, *DLESE* and *MERLOT*. *SMETE* identifies itself as "a collection of collections and a community of communities." Users can search its online catalog to find science-related learning resources, browse its collections, search for items in its "partner collections," or look for books and articles in the California Digital Library. An option to limit searches to peer-reviewed items is under development. Registered users can create a profile and save resources in a workspace. The profile and downloaded items also serve as the basis for *SMETE* to identify other members of the community who have similar interests or to recommend additional similar learning resources. *SMETE* is funded as both a "Collection" and "Core Integration" project with *NSDL*. To learn more about the status of its NSF award, "Enhancing Interoperability of NSDL Collections and Services," use *NSDL*'s Collaboration Finder (a tool developed by *SMETE*) and search "Agogino" as the principal investigator at: http://www.smete.org/smete/nsdl/collabfinder/.

K-12 Teacher Support: ENC Online

Established in 1992, the **ENC (Eisenhower National Clearinghouse for Mathematics and Science Education)** is funded in part by the U.S. Department of Education and located at The Ohio State University. *ENC*'s mission is to identify effective curriculum resources, create high-quality professional development materials, and disseminate useful information and products to improve K-12 mathematics and science teaching and learning. With a staff of 65, *ENC* acquires and catalogs math and science curriculum resources, provides a selection of quality resources on the Internet, supports teachers' professional development, and collaborates with the National Network of Eisenhower Regional Consortia and many other organizations across the nation to promote education reform. *ENC Focus: Magazine for Classroom Innovation* has a circulation of 125,000 subscribers to its printed edition; access is also provided online via the *ENC* Web site. In addition, the *ENC* Web site features a "Classroom Calendar" (with *entries that contain background information, ready-to-go activities, and*

other suggested curriculum materials related to math and science topics), the "Digital Dozen" (a monthly selection of quality Web resources), "Lessons & Activities" (access to Web sites with lesson plans and activities organized by sub-topic in math and science), and "Ask ENC" (submit questions to reference librarians).

At its core, *ENC* is a national repository of more than 25,000 resources collected from federal and state agencies, commercial publishers, professional organizations, local school districts, and individuals. *The collection includes print materials, software and CD-ROMs, kits and manipulatives, along with thousands of Internet sites. This information resides in a searchable database found in the Curriculum Resources area* of the *ENC online* Web site. *ENC* provides unique and comprehensive catalog records—with more than 20 fields of information—for all resources in its collection. In addition to standard bibliographic information, records include fields designed to meet the needs of educators, such as grade level, table of contents, a descriptive abstract, research and reviews, and product information. Teachers can annotate records based on their experience in the classroom. A sample annotated catalog record, which includes user comments, can be viewed by linking to "Research and Reviews" for the entry—Touchmath Computation Set.[65] The Z39.50 online catalog permits searches refined by: Resource type (lessons & activities; standards & frameworks, professional development), Media type (only Web sites, excluding Web sites), Grade level (intervals from pre-K to post-secondary), and Cost (less than $50, including free Web sites). Searches can also be limited to resources with quality indicators: "evaluated resources," those that are "*ENC Focus*" features, or have won "Digital Dozen" recognition.

About 10 percent of *ENC*'s resources are OAI-compliant (or represent digital objects), however, *ENC* is a registered OAI data provider and is actively involved with the NSF in digital library development. *NSDL* as an aggregator site, providing access to the *ENC* collection, can't match the level of search filtering and processing available from *ENC* site itself. However, with *NSDL* funding, *ENC* is engaged in *creating a uniform semantic base for science metadata for K-12 science education based on the National Science Education Standards and combining existing metadata sets for various types of scientific resources to form a consistent scheme covering all objects relevant to K-12 science*.

With *NSDL* funding, *ENC* is also developing *FERL: The Federal Education Digital Resources Library, an archived collection of outstanding, Federally-supported, digital Science, Technology, Engineering, and Mathematics (STEM) resources, cataloged at a high level of granularity and richly described using the IEEE Learning Object Metadata Standard*.[66] This collection is available through *NSDL* and *ENC*.

ENC has also been a partner in other NSF-funded projects that are collections within the *NSDL* aggregation:

---

[65] The direct link is located at: http://enc.org/resources/records/contents/0,1240,025104,00.shtm

[66] For information about IEEE's Learning Technology Standards Committee (LTSC) and LOM see: http://ltsc.ieee.org/, accessed on September 5, 2003.

- *ICON: Innovative Curriculum Online Network*, a partnership with the International Technology Education Association to develop a digital library to promote K-12 technological literacy. http://icon-techlit.enc.org
- *The Learning Matrix*, peer-reviewed electronic resources for teaching future mathematics, and science teachers and information about best practices in undergraduate teaching and assessments, course syllabi, and interactive learning materials. http://thelearningmatrix.enc.org/
- *GSDL: The Gender and Science Digital Library*, collaboration with the Gender and Diversities Institute at the Education Development Center to promote gender-equitable science education. http://www.gsdl.org
- *EDL: The Ethnomathematics Digital Library,* designed to preserve and affirm the rich cultural and mathematical heritage of indigenous cultures, and to ensure worldwide access to this heritage.[67] http://www.ethnomath.com/

Finally, *ENC* is currently engaged with *NSDL* to develop a portal focusing on the needs of middle school science teachers:

> *The purpose is to build a practical portal that supports standards-based science teaching, while creating a general model and technology framework for future development and integration of other specialized capabilities and libraries into NSDL.*
>
> *This project will build on the ENC experience that teachers do not need more information, but a trusted advocate who will clear a path through an overload of information for teachers. This implementation will allow discovery of learning resources, enable reuse of those resources, and promote community conversations about developing useful resources… The initial library will be available Fall 2003, including such capabilities as searching, browsing, news, calendars, and tutorials.* [From http://about.nsdl.org/xhtml/portals/MiddleSchool.php]

Biology Node: BEN (BiosciEdNet)

The ***BEN (BiosciEdNet)*** "portal" provides access to learning resources from *BEN* Collaborative partner organizations and is managed by the American Association for the Advancement of Science (AAAS). The Collaborative is composed of 15 professional societies and coalitions for biology education. Funded by its individual partners and through a NSF-*NSDL* grant, the *BEN* online catalog has over 1,000 reviewed resources covering 51 topics in the biological sciences derived from the AAAS/Science's STKE (Signal Transduction Knowledge Environment), the *Association for Biology Laboratory Education* (ABLE), the American Physiological Society's *Archive of Teaching Resources*, the American Society for Microbiology's *MicrobeLibrary*, Ecological Society of America's *EcoEdNet*, and the *Society of Toxicol-*

[67] These projects are described by Roempler [2002a and 2002b].

ogy. Resources from other partners are added when cataloged. Registration is required to use the search, advanced search, and browse services. Users can browse by resource type (more than thirty categories, ranging from images to teaching strategies and guidelines) or by subject (51 categories ranging from bacteriology to physiology). There are direct links from the online entries to the resource, some of which require additional log-on to view. In addition to standard bibliographic information, education information such as audience and pedagogical use is supplied along with copyright or usage restrictions and technical information such as file type and size. Many of these fields are also available as filters in the advanced search (e.g., users can limit searches by type of resource, subject, grade level, or pedagogical use—assess, learn, research, plan, teach). As *BEN* grows it expects to implement community services to assist faculty users in networking with each other based on profiles established upon registration. *BEN* is currently in its second cycle of NSF-*NSDL* "collection" track funding (through 9/30/04). An article about *BEN*, one of few articles to appear in a disciplinary-based journal, was published in *BioScience* this July [Lundmark 2003].

Geosciences Node: DLESE

The ***DLESE (Digital Library for Earth Systems Education)*** is conceived as an information system dedicated to the collections, enhancement, and distribution of materials that facilitate learning about the Earth system at all educational levels. It is being built as a community effort; collections, services, and tools will be developed and maintained by numerous partners that reflect the broadest possible participation from the Earth system educational community.

DLESE predates *NSDL* by one year, but they have worked closely together from the outset in articulating a vision for a national digital library for science.[68] *DLESE* serves as the geoscience "node" of the *NSDL*, and both communities benefit from a synergystic exchange of intellectual capital, social innovation in understanding and doing distributed development on a large scale, and technological innovation.[69] (*DLESE*'s Program Center and *NSDL* also share a central office space.) *DLESE* is an OAI-registered data provider and also makes available its open source software in support of other OAI data providers and harvesters.[70]

DLESE distinguishes itself as a grassroots, community-based organization, complete with "Articles of Federation" and a Strategic Plan.[71] The *DLESE* Web site thoroughly documents its evolving governance structure, which is becoming more formalized. In 2002, it appointed a management council that is composed of the principal

---

[68] See previously cited, "Pathways to Progress: Vision and Plans for Developing the NSDL," at http://doclib.comm.nsdlib.org/PathwaysToProgress.pdf

[69] See *DLESE* and *NSDL* for more background about this partnership: http://www.dlese.org/about/dlese_nsdl.html

[70] See *DLESE interoperability and OAI* for details including links to its software documentation: http://www.dlese.org/libdev/interop/

[71] Articles of Federation: http://www.dlese.org/documents/policy/art_of_fed1-19-01.html
Strategic Plan: http://www.dlese.org/documents/plans/stratplanver12.html (last updated on May 7, 2003).

investigators of *DLESE* core-funded projects, and it operates under the leadership of a newly appointed executive director, reporting to its steering committee.[72] *DLESE* has also developed an outstanding set of "policies" pertaining to collections, services, governance, and intellectual property.[73]

Some of *DLESE*'s core collections and exemplary library services were developed with initial funding through the *NSDL*. *DLESE* maintains two primary collections: a "Broad Collection" of non-reviewed resources and a "Reviewed Collection" of resources that have been reviewed according to a required set of criteria. The criteria are: scientific accuracy, pedagogical effectiveness, ease of use, clarity and completeness of documentation, ability to motivate learners, robustness, and significance of content. The resource and metadata attributes required for designation as a broad or reviewed collection are clearly articulated. The types of collections, collection requirements, appropriate metadata framework, and examples are summarized in a "Collection Information Sheet."[74] Although intended for contributors, this information helps the user understand search results and the ways in which they are "branded." *DLESE* estimates that the "Reviewed Collection" comprises about 5 to 10 percent of the *DLESE* Collection.

The *DLESE Versioning Document* charts the development of its library and Web site from 2001 with projected targets through 2006.[75] In August 2003, with the release of Version 2.0, *DLESE* permits users to locate educational resources aligned with the National Science Education Standards and the Geography for Life Standards. This version also incorporates the Community Review System, which allows library users to contribute peer reviews and teaching tips about DLESE resources, and incorporates multiple collections. The Community Review System is a pathway into the DLESE Reviewed Collection, which combines Web-mediated feedback from educators who have used the resource with real learners and peer review by specialists selected by an editorial review board.[76]

In addition to the extensive "Community Review System," *DLESE* has some other unique and very helpful features:

- "View all resources" provides bar graphs indicating the number of items by subject, grade level, or resource type. At a glance the user can compare the size of the collections by sub-category, e.g., geology, atmospheric science, environmental science, and space science have the largest number of resources. From here, the user can then link to annotated listings of specific collections in each sub-category. http://www.dlese.org/documents/bibliographies/

---

[72] See "Governance and Organization" under "About DLESE," located at: http://www.dlese.org/about/about_gov.html

[73] Also accessible under "About DLESE," see "Policies" at: http://www.dlese.org/documents/policy/index.html

[74] Collection Information Sheet: http://www.dlese.org/Metadata/collections/collection-type-info.doc

[75] Summary table is located at: http://www.dlese.org/documents/plans/versioning.html

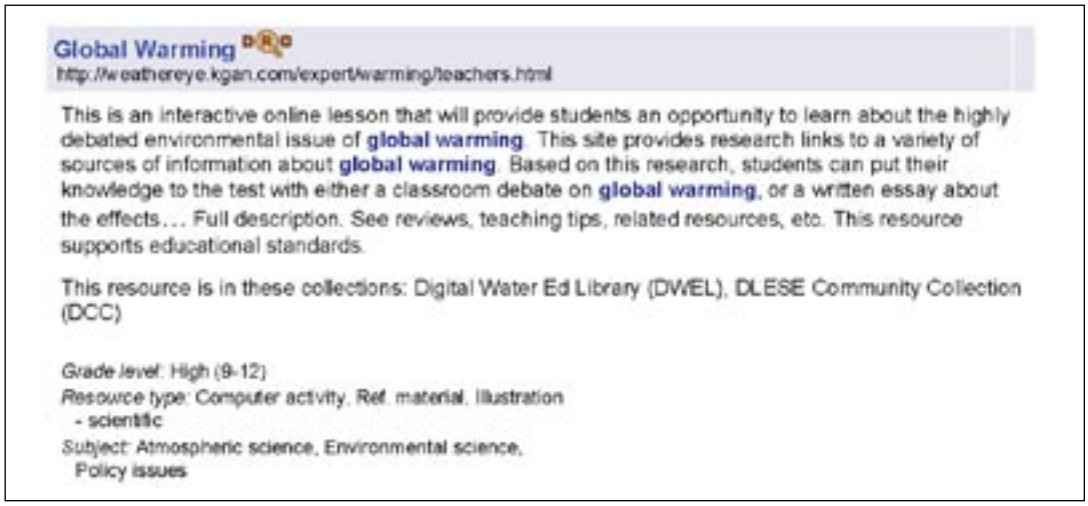
Graphical version is located at: http://www.dlese.org/documents/plans/versions_files/slide0001.htm

[76] The Community Review System is explained at: http://crs.dlese.org/

DLESE_bibliography.html

- Reviewed collections are clearly marked with a DRC (*DLESE Reviewed Collection*) icon.
- You can filter a search by educational standards made available via a drop-down menu of options.
- There are no duplicate records. Instead, the entry for the record indicates: "This resource is in these collections " as illustrated below:

Entry from DLESE:



Projected for release in 2005, Version 3.0 *will support the discovery and classroom integration of spatially and temporally referenced resources, such as data, maps, and images*. *DLESE* has stable funding through August 2007. *DLESE* maintains a bibliography of publications and presentations by members of its community.[77]

6.3.4 Summary of Issues

These services represent the two most influential sectors of digital library services: cultural heritage and scientific information. The cultural heritage sector forms its community base around the worldwide network of institutions (museums, library, archives, historical societies) that are creating digital collections. Although both examples considered here were created with users and audiences in mind, the cultural heritage sector, in general, focuses on creating digital content or the raw materials, which then often "find their own unexpected user communities" [Lynch 2002]. The cultural heritage sector has a growing cadre of trained specialists with some consensus on "good practices" promulgated through national organizations, such as the IMLS and NINCH.[78] There are also many excellent examples from Australia, Canada, and Europe of coordinated large-scale me-

[77] *DLESE* bibliography of publications and presentations: http://www.dlese.org/documents/bibliographies/DLESE_bibliography.html

[78] IMLS [2001a] *A Framework of Guidance for Building Good Digital Collections*. *The NINCH Guide to Good Practice in Digital Representation and Management of Cultural Heritage Materials* [2002].

dia-based digitization programs. A few are noted here:

- *Minerva eEurope: Ministerial Network Valorising Activities in digitization* (network of Member State's Ministries) http://www.minervaeurope.org/home.htm
- *National Library of Australia: Digitisation of Traditional Format Library Materials*: http://www.nla.gov.au/digital/program.html
- *Picture Australia*: http://www.pictureaustralia.org/
- *Music Australia*: http://www.musicaustralia.org
- *Australia Dancing*: http://www.australiadancing.org
- *National Library of Canada*: http://www.imagescanada.ca

The cultural heritage and scientific sectors share "good practices" and collaborate on digital library design,[79] as N*SDL*'s Arms attests:

> A particularly fruitful relationship has been developed between the NSDL and the Institute for Museum and Library Services (IMLS). This relationship has produced two documents on interoperability. The first provides guidance on building good digital collections [IMLS 2001a]. The second addresses collaboration between IMLS and the NSDL [IMLS 2001b].[80]

The June 2003 NSF invitational workshop "Wave of the Future: NSF Post Digital Library Futures" also included speakers from a full spectrum of stakeholders and many of the presentations suggest opportunities for cross-sector collaboration—especially in the areas of cross-cultural and multi-language applications. In contrast to the cultural heritage sector, the sciences are building digital libraries with purpose *and* a disciplinary-based audience in mind. In addition to a collection base, the initiatives in the sciences have a community base—and most of them expect to construct value-added services on top of the "collection" to facilitate communication and collaboration within that community. Eventually, the digital *library* environment may evolve into a digital *online community* focused on teaching or research.

Salient features of these services, worthy of emulation:

- The common "look and feel" across *American Memory*'s many collections give users a coherent visual, organizational, and content schema to follow. Unlike most of the other services, users can easily stay within the "context" of the site. (*ENC* handles the transfer to external sources adeptly by inserting an intermediary screen to notify the user that they are leaving the main *ENC* site.) *American Memory* also provides effective access to different types of media ("hear, read, view").
- *Heritage Colorado* has outstanding documentation about governance, policies, and recommended practices that may serve as a model for other large-scale cooperative digitization programs.

---

[79] A report about joint NSDL/IMLS forums was issued by IMLS [2001b]. See: http://www.imls.gov/pubs/natscidiglibrary.htm

[80] Arms et al. [2002].

- *Perseus*'s research tools for the automatic generation of hypertext links and for visualizations (dynamic maps and timelines) are starting to cross the divide between the cultural heritage and scientific communities.
- *NSDL*'s "Collaboration Finder," built in partnership with *SMETE* and *MERLOT*, is a tool of tremendous potential and value that could be developed for other sectors or comprehensive aggregator sites.
- *ENC* has a cohesive collection and user focus. Its extensive catalog records, emerging system of annotation, connection to educational standards, and work with the IEEE LOM (Learning Objects Metadata) standard are noteworthy.
- *BEN* has successfully attracted an influential number of partners from its disciplinary community.
- *DLESE* might win the "best in show" award for putting into practice many of the desired features—extending from its strategic plan to its documentation and from its search functionality to its management of duplication, and its system of community peer review.

Overarching issues that need attention:

- Organizational sustainability of these initiatives, with increasing attention paid to governance structures and the need for business plans;
- Management and preservation of data or data "curation"—assigning long-term responsibility;
- Managing comprehensive "collections" or "libraries" while providing subsets of users with organized pathways through the content and services tailored to their needs;
- Figuring out how to make digital representations reusable for different purposes by different constituents; and
- Transitioning from digital libraries to digital learning environments, with more attention on users and uses.

6.4 From Peer-Reviewed Referratories to Portal Services

FROM PEER-REVIEWED REFERRATORIES TO PORTAL SERVICES ❏ Quality-controlled subject gateways ❏ Resource selection, discovery, annotation **PORTAL SERVICES** ❏ Collaborative information research service **Elements** ❏ Intuitive and customizable Web interface ❏ Personalized content presentation ❏ Security and Authentication ❏ Communication and collaboration **Components** ❏ Single-search interface ❏ User authentication ❏ Resource linking ❏ Content enhancement [Boss 2002]	**PEER-REVIEWED LEARNING RESOURCES** Merlot (Multimedia Educational Resource for Online Learning & Teaching) **EXPERT & MACHINE-GATHERED INTERNET RESOURCES** ❏ ALL DISCIPLINES InfoMine *Scholarly Internet Resource Collections* ❏ DISCIPLINARY HUBS UK: Subject Portals Project of the Resource Discovery Network **SCHOLAR-DESIGNED PORTAL** AmericanSouth **RESEARCH LIBRARY PORTALS W/ ACCESS TO PROPRIETARY DATABASES** U.S.: ARL Scholars Portal AUSTRALIA: AARLIN: the Australian Academic and Research Library Network

This section considers a set of services that range from "referratories" or "subject gateways"[81] of quality-controlled Internet resources to "portals" that provide customized access to user-selected content. The referratories discussed here represent different forms of peer review as well as different methods of gathering resources—from expert-selected to hybrid expert/machine-selected approaches. They each offer opportunities to contribute or customize content, bridging the boundary between search engine, Web directory, and portal. Of particular interest is the example of a scholar-designed portal that overlays an OAI repository. Finally, two research library portals are considered that are concentrating primarily on access to proprietary licensed databases thus far.

Like "digital libraries," definitions for "portals" are plentiful and evolving. For the purposes of this discussion, I offer the UK's Joint Information Systems Committee definition:

> …a network service that brings together content from diverse distributed resources using technologies such as cross searching, harvesting, and alerting, and collates this into an amalgamated form for presentation to the user. This presentation is usually via a web browser, though other means are also possible. For users, a portal is a, possibly personalized, single point of access where searching can be carried out across one or more than one

[81] For definitions of subject gateways refer to Koch [2000].

> resource and the amalgamated results viewed. Information may also be presented via other means, for example, alerting services and conference listings or links to e-prints and learning materials. [From: JISC Portals FAQ at: http://www.portal.ac.uk/spp/, accessed on September 2, 2003.]

Indeed, as the JISC Subject Portals Project explains its development strategy, it is possible to understand how various categories of services discussed in this report merge into a unified "portal" application:

> The project is committed to using open source products wherever possible, and our development strategy has been to create areas of functionality in modular "portlets" which can be embedded in a portal framework. It is therefore an aim of the project to explore the feasibility of embedding the portlets within alternative third party portal environments, such as institutional portals and virtual learning environments, and to make this technology open source. [From: JISC/RDN's Subject Portals Phase II at: http://www.portal.ac.uk/spp/, accessed on September 2, 2003.]

6.4.1 Peer-Reviewed Learning Resources: MERLOT

MERLOT, the Multimedia Educational Resource for Learning and Online Teaching is a community of educators in higher education who collaborate to develop and disseminate high quality online resources for faculty to incorporate into their courses. The California State University developed the prototype for the national *MERLOT* project in 1997, and continues to play a key role in the project's technical design, implementation, and user evaluation. *MERLOT* has five membership categories representing higher education organizations: disciplinary professional societies and digital libraries; individual campus institutions of higher education; content publishing companies, academic technology companies, and technology companies; sponsors; and multiple-campus institutions of higher education. Crossing the bounds from digital libraries to e-learning environments, among *MERLOT*'s notable members are *SMETE*, the IMS Global Learning Consortium, and the National Learning Information Initiative. *MERLOT* has been awarded three *NSDL* grants since 2000:

- Peer Review of Digital Learning Materials: Critical Service for Digital Libraries http://taste.merlot.org/projects/nsdl/peer_review/
- The NSDL Collaboration Finder: Connecting Projects for Effective and Efficient NSDL Development http://taste.merlot.org/projects/nsdl/collaboration_finder/
- Scaling the Peer-Review Process for National STEM Education Digital Library Collections http://taste.merlot.org/projects/nsdl/scaling_peer_review/

MERLOT has implemented a peer-review process for its collection of more than 9,500 learning materials. The peer-review process

includes: evaluation standards, peer-review procedures, collections policies, a rating system, and training of reviewers. Its collection focuses on 14 disciplinary communities, each of which oversees a subset of the *MERLOT* collection, and is curated by an Editorial Board.

Each disciplinary community can tailor the *MERLOT* Evaluation Criteria and Collection Development Guidelines to meet its specific needs. Overall, *MERLOT* has three broad Evaluation Criteria, each of which is further defined:

- Quality of Content
- Potential Effectiveness as a Teaching-Learning Tool
- Ease of Use

[See: http://taste.merlot.org/projects/peer_review/criteria.php]

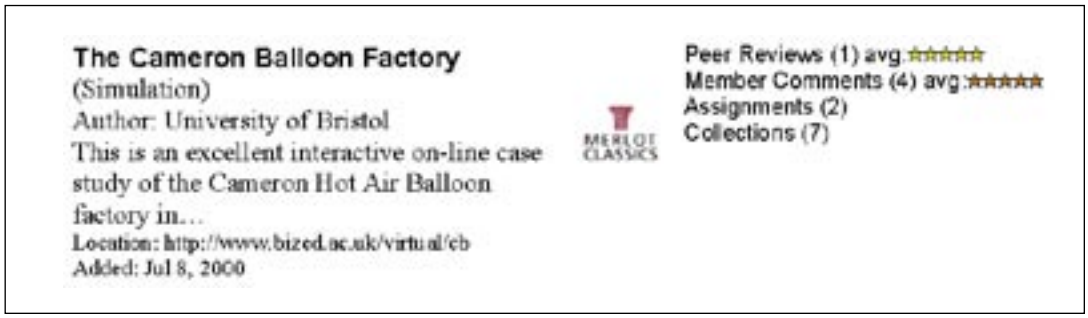
The *MERLOT* collection is guided by the following principles:

- Users should be able to find the best available learning material on *MERLOT*.
- Users should be able to expect searches of *MERLOT* to provide quality material and information regarding the quality of what is found.
- The process of finding material should be simple but also flexible enough to meet the diverse needs and searching styles of users.
- *MERLOT* should be as broad as possible to cover the needs of diverse users.
- Materials in *MERLOT* are not necessarily designed to be stand-alone learning materials. It is expected that some guidance on using some materials would be provided.

[See: http://taste.merlot.org/policies/collection_development.php][82]

Users can browse or search its collections; registered users can submit items for consideration by the editorial board. Registered users can also create personal annotated collections. As illustrated below, entries offer links to: Peer Reviews (with ratings), Member Comments (with ratings), Assignments, and the number of Personal Collections to which they belong.

Entry from MERLOT:



[82] The complete document," MERLOT Collection Development Guidelines" (includes policy for removing materials from MERLOT) revised January 24, 2003, is located at: http://taste.merlot.org/documents/policies/MERLOT-collec_dev_guidelines-012403.pdf

MERLOT's advanced search features permit users to restrict their queries to items that have been peer reviewed or have user comments as well as to those that have received specified minimum ratings. Users can also search by material type (e.g., animation, simulation, case study), technical format (e.g., Flash, Shockwave, Audio), audience level, language, copyright restriction, cost, and other qualifiers.

6.4.2 Expert and Machine-Gathered Internet Resources: INFOMINE and UK's Subject Portals Project

All Disciplines: INFOMINE

INFOMINE: Scholarly Internet Resource Collections is a librarian-built virtual library of Internet resources relevant to faculty, students, and research staff at the university level. It provides access to over 100,000 resources across all subjects and of all types. It is also a flexible and collaborative system that allows other institutions to develop Internet resource directories by providing them with resource discovery and content building, editing, and maintenance tools.

INFOMINE created the iVia open source virtual library system, which is intended to scale well with burgeoning Web content by using on a hybrid expert-selected/machine-identified approach to collection creation and management. It relies on an expert-created, first-tier collection (currently about one-third of its content), augmented by a second-tier collection of Internet resources that are automatically gathered and described. It supports the following standards: OAI Protocol for Metadata Harvesting (OAI-PMH), Dublin Core, MARC (Machine-Readable Cataloging), Library of Congress Subject Headings (LCSH), and Library of Congress Classifications (LCC). Headquartered at the Library of the University of California, Riverside, *INFOMINE* was developed with funding from an IMLS National Leadership Grant and from the Fund for the Improvement of Post-Secondary Education (FISPE).[83]

Searches can be limited to "expert-selected" or "robot-selected" entries. Advanced searches can be restricted by field (author, title, etc.) or by subject/type category. Users can browse expert-selected records by Library of Congress subject classification. Users can comment on resources, select resources, and receive e-mail news alerts about *INFOMINE*.

Disciplinary Hubs: Subject Portals Project

The UK's ***Resource Directory Network*** is a collaboration of more than 70 educational and research organizations, including the Natural History Museum and the British Library. In 1998, JISC (Joint Information Systems Committee) funded the development of RDN's *Subject Portals Project. A subject portal, for the purposes of this project therefore, is a tailored view of the web within a particular subject area, with access to high-quality information resources made easier for the user*

[83] See Mitchell et al. [2003] for a history and description of *INFOMINE* and its open source software "iVia."

through aggregated cross searching; streamlined account management; user profiling; and the provision of additional services.[84]

Now in Phase 2, September 2003 through August 2004, this project builds on the earlier work of the RDN and is developing portal functionality for five subject hubs. Hubs are typically consortia of prominent library, academic, research, and professional organizations in the UK with the expertise and subject knowledge to oversee the selection and evaluation of resources. The five subject hubs under development are:

- ***BIOME*** for health and life sciences
  http://biome.ac.uk/

- ***EEVL*** for engineering, math, and computer sciences
  http://www.eevl.ac.uk

- ***HUMBUL*** for the humanities
  http://www.humbul.ac.uk

- ***PSIGate*** for the physical sciences
  http://www.psigate.ac.uk

- ***SOSIG*** for the social sciences, business, and law
  http://www.sosig.ac.uk

> *The project is committed to using open source products wherever possible, and our development strategy has been to create areas of functionality in modular "portlets" which can be embedded in a portal framework. It is therefore an aim of the project to explore the feasibility of embedding the portlets within alternative third party portal environments, such as institutional portals and virtual learning environments, and to make this technology open source.*[85]

The subject hubs are developed within a common framework of collection policy guidelines and evaluation criteria promulgated by RDN, however, each varies in its presentation, functionality, and specific features.[86] They all warrant closer examination. A feature about EEVL appeared in the August 6, 2003 "Search Day" column at SearchWatch.com [Price 2003]; *SOSIG* was reviewed in the July/August 2003 issue of *C&RL News* [Roberts and Drost 2003].

SOSIG features high-quality Internet Resources, selected by experts according to well-articulated criteria; it can be searched or browsed by subject via the *SOSIG* "Internet Catalogue."[87] The Z39.50

---

[84] Subject Portals Project Phase II: http://www.portal.ac.uk/spp/, accessed on September 1, 2003.

[85] Subject Portals Phase 2: http://www.portal.ac.uk/spp/, accessed on September 1, 2003.

[86] The excellent "RDN Collections Development Framework" (version 1.2, July 2002) is available at:
http://www.rdn.ac.uk/publications/collections/cdframework3.doc, accessed on September 4, 2003.

[87] *SOSIG's* "Selection Criteria" are located at: http://www.sosig.ac.uk/desire/ecrit.html, accessed on September 4, 2003.

catalog offers controlled vocabulary searching with three different thesauri. In addition, it covers a full spectrum of types of materials, as specified in the list reproduced below. Search results return these "resource types" in a left-hand frame, making it possible for users to narrow their search to particular formats.

SOSIG Resource Types:

Articles/Papers/Reports (collections)	Collections of materials as opposed to individual documents. These may be articles, working paper series, conference proceedings, pre-prints, or other collections of materials. Papers may or may not be available as full-text. Does not include government publications - see *Government Publications*.
Articles/Papers/Reports (individual)	An online document (paper, article, report, etc.) available as full-text. Does not include government publications - see *Government Publications*.
Bibliographic Databases	Databases of bibliographic information; including library OPACs.
Bibliographies	Individual lists of bibliographic information, not contained within a database.
Books/Book Equivalents	Either online versions of printed books or else Web sites that provide access to original content held locally, created by a single author or corporate body, and relating to a single topic. Does not include reference books - see *Reference Materials*.
Companies	Links to individual company Web sites.
Company information	Resources providing data about companies (usually financial).
Data	Primary data, usually stored in online databases; including statistics, socio-economic data, etc.
Documents - Digests	Online indexes and compilations of case law and/ or legislation summaries with commentary and subject guidance.
Documents - Law Reports	Online texts and collections of case reports, judicial decisions, opinions and judgments from law courts or tribunals.
Documents - Legislation	Online texts and collections of primary and secondary legislation, including acts, ordinances, statutes, constitutions, rules, regulations, orders and statutory instruments proposed and passed by parliaments around the world.
Documents - Treaties	Online texts and collections of bilateral and multilateral treaties and international agreements between nation states, and agreements relating to inter-governmental and international organisations.
Educational Materials	Online materials designed for teaching and learning.
FAQS	Frequently-Asked Question lists, providing commonly requested answers on a particular topic.
Government Publications	Online documents published by government bodies. May be individual documents or collections.
Governmental Bodies	Web sites produced by governments and government bodies, including the European Union.
Journals (contents and abstracts)	Information on individual or lists of serial titles, where the full-text of the articles is not available. Includes all serial types, from refereed journals to newsletters (except newspapers - see *News*). May also refer to titles where the full-text of articles is only available via a subscription.
Journals (full text)	Online, full-text serials, from refereed journals to newsletters, not including newspapers - see *News*.
Mailing Lists/Discussion Groups	Information about email lists and newsgroups, including mailing list archives.
News	Online news services, including newspapers.

Organisations/Societies	Web sites providing information about organisations, societies, or professional associations.
Reference Materials	Dictionaries, directories, encyclopedias, etc.
Research Projects/Centres	Web sites providing information about individual research projects or centres.
Resource Guides	Sites which collate links to other Internet resources, relating to a particular topic or topics.
Software	Software available via the Internet; for downloading or for use online. May require payment.

If users want to expand their search, they can turn to the "Social Science Search Engine," a database of over 50,000 Social Science Web pages harvested via a focused Web crawler and including OAI-compliant sites.

Users who sign up for a *SOSIG* account can set up a personal, customized Web page on *SOSIG*, with channels of their own choice, receive e-mail current awareness alerts, post details about conferences or events, post their CVs, and locate "like-minded colleagues" via the "Grapevine." The Grapevine is an online center for information about professional development opportunities.

6.4.3 Scholar-Designed Portal: AmericanSouth

The ***AmericanSouth*** is sponsored by the *MetaScholar Initiative*, based at Emory University in partnership with the Association of Southeastern Research Libraries (ASERL) and funded by the Andrew W. Mellon Foundation. It "seeks to create a definitive scholarly portal for Southern history and culture," by "layering portal services on top of a central metadata harvester that would aggregate information from cooperating partner libraries."[88] A "Scholarly Design Team," composed of five senior scholars from different disciplines, is responsible for the intellectual organization of the site, recommending content, and identifying and testing the types of contextual and interpretive tools needed to access the content, and to facilitate communication among scholars. This includes the development of contextual tools such as subject guides, thematic articles, commentary, and Web site annotations. [89] Ten institutions are currently participating in the project: Auburn University, Emory University, Louisiana State University, the University of Florida, the University of Georgia, the University of Kentucky, the Kentucky Virtual Library, the University of North Carolina at Chapel Hill, the University of Tennessee at Knoxville, and Vanderbilt University. *AmericanSouth* uses *Arc* for its central harvesting infrastructure and is "creating a metadata harvesting network of OAI provider systems installed and maintained at partner research libraries."[90] It relies on an open source software system with portal and content management features (*PostNuke*) that

[88] Halbert [2003]: 184.

[89] See the FAQ at the site for further information about the Scholarly Design Team.

[90] Halbert [2003]: 186.

supports Web site annotation, threaded commentary, and topic discussion forums.[91]

AmericanSouth will make its official debut later this Fall. From the current home page, the purpose, audience, and collection scope of *AmericanSouth* are not clearly stated. The main body of the page includes lengthy texts, which, to the uninitiated, appear to be in the form of threaded e-mail discussions. In the forthcoming release, developers plan to create distinct sections within the site: commissioned scholarly subject guides, peer-reviewed articles, previously published encyclopedia articles, and lightly moderated threaded discussion forums.[92] At present, the first-time user might not even notice the search box in the upper right-hand corner: "Search Archives." There is a "Help" button close at hand, but there is no description of the content to be searched and the participating archives aren't described. There are plans to post a collection development policy in the future. The site includes metadata for both print and online collections; criteria are based on scholarly value, not format or media accessibility. As of early September 2003, *AmericanSouth* comprised 18 archives, totaling nearly 30,000 records. The Senator John Tower papers from Southwestern University constitute more than half of the total collection, some 18,000 records. As a registered user you can post comments, send news, have a personal box on the homepage, customize comments, select different themes and take advantage of other customized features.

6.4.4 Research Library Portals: ARL Scholars Portal (U.S.) and AARLIN (Australia)

U.S.: ARL Scholars Portal

Launched in spring 2002, the ***ARL Scholars Portal*** is a collaborative project of seven ARL libraries—University of Southern California, University of California, San Diego, Dartmouth College, University of Arizona, Arizona State University, Iowa State University, and the University of Utah—with Fretwell-Downing Inc., which relies on Z39.50 technology.[93] The project has two overarching goals: (1) to provide meta-search capability—single-search access to information resources, and (2) to offer advanced linking—connecting the user to the resource from the bibliographic metadata. Each institution has its own customized implementation of Fretwell-Downing's ZPORTAL product, but participating members are cooperating to establish a cohesive pool of resources, with the initial focus—based on user demand—on licensed databases. Databases are collaboratively configured to become Z39.50 compliant, in priority order agreed upon by the group, starting with resources in the categories of Literature,

---

[91] See the PostNuke Web site at: http://postnuke.org

[92] Information about *AmericanSouth* is based on email correspondence with Michael Halbert on July 29, 2003 and from his (unpublished) PowerPoint presentation to the Association for Computing and the Humanities on June 1, 2003.

[93] Information about the *ARL Scholars Portal* is based on email correspondence and a phone interview with Krisellen Maloney, Team Leader, Digital Library and Information Systems Team, University of Arizona, Tucson, on August 28, 2003.

Environmental Studies, and "panic"—readily accessible full-text databases that typically meet the "must have it now" demands of last-minute, late-night student research. Next in line are databases related to Engineering, General Reference, Social Sciences, Biomedicine, History, and Nursing.

The project's developers have encountered a great deal of resistance from database vendors to Z39.50 compliancy. They estimate that only 25% of licensed databases are Z39.50 compliant and compliance to the standard does not guarantee interoperability. There is no universally accepted format for citation-related fields. As a result, the process of writing new scripts can be very time-consuming—taking anywhere from one hour to eighty hours per database. To date about eighty resources have been configured for use by participants. Information about the configuration of each of the resources is stored at a site maintained by the University of Utah.

The University of Arizona and Iowa State University (ISU) have launched their systems, with others anticipated for release in fall 2003. Resources are grouped into "profiles" from which users can select or deselect specific databases when initiating a search. In Iowa State's deployment, the Basic Search Profile currently searches across four databases: Expanded Academic ASP, ISU's Library Catalog, Science Direct, and WorldCat. Users can create and save their own customized profile and context-sensitive Help is available. Called "Find It," ISU's service is viewable at its Library Web site.[94]

Future work may address the integration of the system with courseware, advanced manipulation of results (relevancy ranking), and more intelligence at the front-end of the search to recognize the needs of individual users. Participants in this project have the tools to access other types of information—Utah is actively pursuing the inclusion of locally developed resources. A Z39.50-to-OAI mapper is also under development; however, this is not a high priority at present. Overall, full development of the *ARL Scholars Portal* is anticipated to take three years. Meanwhile, there are early indications that many users are more than willing to accept the trade off of a less elegant search for the convenience of executing a single search across multiple resources. Even in its "Beta-search" version, the portal is the fifth most frequently used information resource at the University of Arizona.

Australia: AARLIN

The ***AARLIN (Australian Academic Research Libraries Network)*** has undertaken a very similar project on behalf of its members, using Ex Libris' *Metalib* software. The project is currently in Phase 2 (2002-2004), having successfully completed a pilot and received a government grant to develop a framework to facilitate implementation across participating academic libraries in Australia. In addition to the rollout of portal software, Phase 2 aims:

---

[94] "Find It" is directly accessible at: http://pollux.lib.iastate.edu:8080/zportal/zengine?VDXaction=ZSearchSimple, accessed on September 4, 2003.

- *To develop an administrative structure that ensures cost-efficiencies and sustainability of the AARLIN system.*
- *To create a legal framework that will encompass issues such as copyright, and intellectual property; and development streams such as e-commerce.*
- *To devise and implement a business plan.*

In contrast to ARL's project, the AARLIN portal is being developed centrally. It reports difficulties similar to those encountered by ARL, with even longer estimates of database configuration—on average, one week.

6.4.5 Summary of Issues

All of the services considered in this section are targeted for academic users, and aim to provide them with a customized search that has sifted through a larger body of information and filtered out unwanted or less reliable resources. With the exception of *MERLOT*, they all rely on a combination of expert- and machine-driven protocols—although these differ by degree and method. They explicitly or implicitly introduce systems of rating or ranking resources and give users ways to customize access. *MERLOT*, *RDN's Subject Portals*, and *AmericanSouth* all aim to build scholarly communities and have mechanisms to identify like-minded colleagues. Along with the previously considered digital library services, they begin to support functions for collaboration. *MERLOT* is widely regarded as a model for building a community-based peer-review system.

There are also fine examples of nationally coordinated subject gateways and research portals in Australia and Germany. Australia's Subject Gateways Forum (ASGF) site,[95] sponsored by the National Library of Australia, tracks the development of Australian subject gateways and lists each gateway's approach to: software, metadata, interoperability, thesauri, quality assurance, usage statistics, partners, milestones, and contact detail [Schmidt et al. 2003]. Launched in 2003, *Vascoda*[96] is an interdisciplinary research portal created by German libraries and information centers that will form the nucleus of a German Digital Library [Pianos 2003].

[95] Australian Subject Gateways: http://www.nla.gov.au/initiatives/sg/

[96] Vascoda: http://www.vascoda.de

6.5 Specialized Search Engines

SPECIALIZED SEARCH ENGINES	SCIENCES
❏ Information retrieval system	LANL FEDERATED SEARCH IN-HOUSE PROPRIETARY + SELECTED PREPRINTS + LIBRARY CATALOG
❏ Multi-database search tool	Flashpoint
❏ Filters	
❏ Finds	COMPUTER SCIENCE WEB CRAWLER W/ REFERENCE LINKING, CITATION ANALYSIS, & RECOMMENDER SYSTEM
❏ Searches	CiteSeer (aka ResearchIndex)
❏ "Niche" Search Engines	ELSEVIER WEB CRAWLER: SELECTED OAI REPOS + PROPRIETARY + WEB
	Scirus

Search engines play a critical role in helping users cope with the growing mass of diverse information resources. These examples illustrate three ways in which diverse sets of resources can be accessed through a unified search interface. Each of them is tailored to a specific audience within the scientific community. Both ***CiteSeer*** and ***Scirus*** use advanced, focused Web-crawling techniques to retrieve targeted relevant information from a vast array of resources.

6.5.1 Sciences: Flashpoint, CiteSeer, Scirus

Los Alamos Federated Search Engine: Flashpoint

Flashpoint is a proprietary, in-house multi-database search tool devised primarily for users of LANL's Research Library. It provides a unified search interface to twelve distinct databases: BIOSIS, DOE Energy, Engineering Index, INSPEC, ISI Proceedings, MathSciNet, Nuclear Science Abstracts, PubMed, Science Server, SciSearch, Social SciSearch, and LANL's Library Catalog. All but MathSciNet and PubMed are locally loaded. *Flashpoint* can be searched by specific database (user-selected) or by subject (the system selects relevant databases). It is possible to filter the search to "LANL research only." Search results show at a glance which database contains the most matches to a query. As you click on specific results, you enter the native environment of each database. You can then view records, go to full-text documents online, mark records, and download or e-mail search results. Subject coverage includes primarily: Astronomy; Biology/Genetics; Bioinformatics; Chemistry; Computer Science; Environment; Engineering; Earth Sciences; Library & Info Science; Mathematics; Nuclear Information; and Physics. Because access to this service is restricted, it cannot be evaluated further; Mahoney and Di Giacomo [2001] published an article about its early development. Overall, the LANL Research Library provides access to 5,860+ journals online, 5 million+ full-text articles, 55,000+ electronic technical

reports, and 73 million+ citation records. Additional information is available from http://lib-www.lanl.gov/lww/flashpoint.htm.

Computer Science Web Crawler: CiteSeer

CiteSeer, also known as *ResearchIndex*, is a database of computer science literature that is built via Web crawling, using data mining and intelligent search functions. Because *CiteSeer* is based on algorithms, techniques and software that can be used to develop other specialized collections, it is considered here as a "niche search engine" rather than as a *digital library of scientific literature*.[97] *CiteSeer* allows for keyword searching but also indexes all of its documents by citation (via autonomous citation indexing).[98] Its other features include reference linking, citation context, awareness, and tracking, locating related and similar documents.[99] Results can be sorted in various ways including by citation, date, or usage. Users can view or download results and also rate and submit comments about them. *CiteSeer* also permits users to submit documents, links, and content updates. According to one of its developers, Lee Giles [2003], as of May 2003, *CiteSeer* had cataloged some 500,000 papers, adding some 10,000 papers monthly and receiving 100,000 visits per day. Publications by *CiteSeer*'s developers on digital libraries and citation indexing, and on Web analysis and Web search, are available at the site.[100]

Elsevier's Web Crawler: Scirus

Winner of Search Engine Watch's 2001 and 2002 award for best specialty search engine, ***Scirus*** is a Web search engine for scientific information launched by Elsevier in 2001. It relies on a focused crawler from FAST™ (Fast Search & Transfer™), which targets a combination of science-specific Web pages, including relevant OAI sources, and of access-controlled proprietary information sources, (including 4.5 million full-text articles from Elsevier's *ScienceDirect*.) Per its Web site, as of late-August 2003, *Scirus* covers:

- 45 million .edu sites
- 14.8 million .org sites
- 5.5 million .ac.uk sites
- 18 million .com sites
- 4.7 million .gov sites
- over 40 million other STM and university sites around the world

In addition to Web pages, *Scirus* indexes from the following journal and e-print sources: MEDLINE, Science Direct, U.S. Patent Office, Beilstein abstracts, arXiv, NTRS, Cogprints, BioMed Central, and Elsevier's three OAI preprint servers for Mathematics, Chemistry,

---

[97] Giles uses the term "niche search engine" in his interview with David Pacchioli [2003]. The *CiteSeer* Web site, refers to itself as a "digital library for scientific literature."

[98] Information about autonomous citation linking see: http://www.neci.nec.com/~lawrence/aci.html, accessed on September 4, 2003.

[99] For more information about these and other features see: http://www.neci.nec.com/~lawrence/researchindex.html, accessed on September 4, 2003.

[100] Lawrence's papers are located at: http://www.neci.nec.com/~lawrence/papers.html, accessed on September 4, 2003.

and Computer Science.[101] *Scirus* is a registered OAI service provider.

Scirus indexes all sources by subject and information type, making it possible to limit searches to twenty different subject areas or a range of document types, including abstracts, articles, books, patents, or scientists' home pages. All searches can be limited by "content source," separating journals from Web sources. Results are clearly displayed with the number of "hits" and their source (total number from journals or Web sources), and can be sorted by relevance or date. Each search result links to "more hits from" the same source or to "similar results." Searches can be refined easily by a dynamically created list of keywords that appears in the right-hand frame. Advanced searches can also be restricted to a specified date range or by file format (e.g., html or pdf). *Scirus* gives users the option to set and save their search preferences, including the number of results displayed per page, and the option of displaying results by opening a new browser, clustering results by domain, and automatically rewriting search queries to improve results. Search results can be saved or e-mailed.[102]

Scirus uses FAST (Fast Search & Transfer) software, which markets itself as a "3rd generation" search engine that uses both algorithmic and rule-based techniques to become an "information management platform." FAST is used by a number of sophisticated commercial and public databases, including Lexis-Nexis and FirstGov.gov. In 2003, FAST announced a partnership with the University Library of Bielefeld "to become a test bed for the use of enterprise search technology in the academic digital library market." [103] For more information about FAST's vision of the future of search engines, refer to CEO Lervik's [2003] presentation at the 2003 European Conference on Digital Libraries (ECDL).

6.5.2 Summary of Issues

Both *CiteSeer* and *Scirus* offer solutions to help the scientific community find and retrieve relevant information, using dynamic data-mining techniques to extract items "hidden" in the Web. Both offer users a better level of quality assurance than relying on general search engines, such as Google or AltaVista. From my perspective, *Scirus's* search functionality is unsurpassed. At the same time, the capabilities of general search engines are improving rapidly, and their popularity and influence, even within the academic community, is undeniable. For example, the Institute of Electrical and Electronics Engineers (IEEE) announced in mid-August that its technical papers will soon be indexed by Google.

> ***IEEE Xplore Indexed by Google***
> Researchers will soon be able to locate technical papers published by The Institute of Electrical and Electronics Engineers (IEEE)

---

[101] For more information about the scope of coverage see: http://www.scirus.com/about/#sources

[102] For more information about how *Scirus* works refer to the white paper at: http://www.scirus.com/about/scirus_white_paper.pdf

[103] http://www.fastsearch.com/us/news_events/press_releases/2003/fast_and_the_university_of_bielefeld_form_strategic_partnership_to_promote_the_use_of_enterprise_search_for_digital_libraries__1

> by using the Google search engine. Google is currently indexing the abstract records for all online IEEE technical documents and standards available through the *IEEE Xplore* online delivery platform (http://www.ieee.org/ieeexplore). Starting sometime in September, Google users will see the linked content in search results. Abstracts are free and full-text will be available for purchase.
>
> Google users can view abstract records when linking from a Google search into *IEEE Xplore*. Abstract records will contain the document's bibliographic information and abstract summary, wherever available. Guests can continue to browse tables of contents to locate and purchase articles of interest. IEEE Members and users at subscribing institutions continue to have access to complete abstract records containing index terms, download citation links, linked references (backward links), "documents that cite this document" links (forward links), and CrossRef links.
>
> IEEE has more than 380,000 members in approximately 150 countries. The IEEE publishes 120 technical journals, magazines, and transactions, and has developed more than 900 active industry standards. The organization also sponsors or co-sponsors more than 300 international technical conferences each year.
> [Announced on August 18, 2003 at NewsBreaks Weekly News Digest: http://www.infotoday.com/newsbreaks/wnd030818.shtml]

Also, as discussed previously when reviewing *Archon*, Old Dominion University's Digital Library Group in partnership with LANL is developing an open source gateway service, DP9, that allows general search engines to index OAI-compliant archives. *DP9 does this by providing a persistent URL for repository records, and converting this to an OAI query against the appropriate repository when the URL is requested. This allows search engines that do not support the OAI protocol to index the "deep web" contained within OAI-compliant repositories*.[104] DP9 is an OAI-registered service provider.

To keep up with search engine developments, readers should consult the SearchEngineWatch.com Web site, or subscribe to Price's "ResourceShelf" (www.resourceshelf.com) weekly news briefing that covers search engine and other e-resource news.

7.0 Conclusions

Given the diversity of these services and their stages of development, the following generalizations and conclusions are offered with some caution.

---

[104] As described at: http://egbert.cs.odu.edu/dp9/

7.1 Current Practice

Overall, there is reason for optimism about the future development of OAI-based services, in particular, and aggregated digital libraries, in general. Given the relative youth of OAI-PMH—first introduced in January 2001—the number, variety, and scope of data providers, and to a lesser degree, service providers, is remarkable. In a guest editorial to *Library Hi Tech* devoted in its entirety to the *Open Archives Initiative Metadata Harvesting*, Timothy Cole rightly asserts:

> The challenge of shedding light on the hidden Web is daunting, but experience so far with OAI-PMH gives cause for optimism. Clearly important and useful work is ongoing, and technologies and standards like OAI-PMH are making the job of sharing digital information resources easier and more tractable.[105]

The "theme articles" in *Library Hi Tech* are written by luminaries in the field—starting with the progenitors of OAI-PMH, Lagoze and Van De Sompel—and extending to the principals "out front" in the development of such resources as *American Memory*, *NASA's NTRS*, the *UIUC Gateway*, *OAIster*, *AmericanSouth*, *NDLTD's Union Catalogs*, *OLAC*, and *NSDL*. Taken together, they provide an outstanding overview of OAI-based initiatives and they give readers an understanding of the challenges and opportunities.

The network of communication and collaboration among researchers, developers, and implementers, if informal, is nonetheless strong, multidisciplinary, and international in scope. The list of presenters at the June 2003, "Wave of the Future: NSF Post Digital Library Futures Workshop," reads like a veritable "who's who" in digital libraries. There is a well-known cadre of "visionaries" and —although there is not unanimity among their views—it is heartening to note the depth and breadth of their engagement. Clearly, digital library futures are in the best hands and minds.

Moreover, the entire "open access movement" is now achieving much more widespread notice with rapid developments including the introduction of the *Public Access to Science Act* (June 26, 2003), numerous articles in the mainstream media, the launching of *PloS Biology,* and the advent of ARL's *Open Access Newsletter*.[106] As this discussion moves into the realm of public discourse and policy-making, it will help to fuel the further development of open access tools and services, such as those discussed in this report.

At the same time, there are numerous practical, technical, and philosophical impediments to the full realization of OAI-based services, in particular, and to digital library aggregations, in general. Many of these have already been discussed in this report. It may be useful to conduct a formal, broad-based survey, such as the one undertaken in Europe by the Open Archives Forum, to achieve a more definitive overview of the landscape in the United States.[107]

---

[105] Cole [2003]: 116.

[106] For a summary and useful links refer to "Open Access News": http://www.earlham.edu/~peters/fos/fosblog.html

[107] As reported by Dobratz and Matthaei in *D-Lib* [January 2003].

Highlighted below are concluding observations:

- There is no required registry of either OAI data or service providers, and it is difficult (at best) for users to know the extent of services available. The OAI's voluntary registries are useful starting points, along with the Open Archive Forum's "Information Resource Database," which encourages registration by European repositories and also attempts to identify services, projects, software, protocol, metadata schemes, and organizations.[108] There is overlap between these registries but each also has unique listings. Neither listing is comprehensive for either the United States or the European Union, however, this is also difficult to determine because of the way in which collections or libraries are aggregated within larger aggregations. Some smaller data providers may forego registering because a larger aggregation, which harvests their data, is registered. On the other hand, even small entities may want the exposure independent of their "parent" site, and so they may be registered separately. Moreover, the registries are really intended for use by system developers or implementers, not by users. The OAI's "Repository Explorer" is primarily for interactive exploration and technical validation, although users can select a repository and link to its originating site. There aren't any user-friendly comprehensive registries geared towards users. Instead, users must rely on a combination of the service providers themselves (e.g., *Arc*, *OAIster*), perusing longer lists of data providers (or using the Repository Explorer), or accessing tools like NSDL's "Collaboration Finder."

- Creating and exposing OAI-compliant metadata—to meet minimal, let alone quality, standards at either the collection- or item-level—may not figure among the top priorities of busy digital library developers. It is, for example, very difficult to "harmonize" the list of "provisional metadata sources" from DLF projects with object-level records in OAI union catalogs. To wit, listed among the DLF projects are the metadata for some 2,800 titles in the Lyle Wright bibliography of American Fiction being digitized by CIC institutions. Indiana University and the University of Michigan, in cooperation with OCLC, made available the full set of MARC electronic records for this collection in Fall 2002. As a result, these titles all appear in OCLC's *WorldCat* as well as in many library online catalogs worldwide. Because of my former connection with this project, I was especially eager to "discover" these resources via OAI services. There is a collection-level record that can be retrieved via *Arc*, however, the expert-created record in *INFOMINE* is far superior. Moreover, in neither instance is there title-level access because that metadata has not yet been made available. This is by no means an isolated case. On a much larger scale, *NSDL* illustrates the huge gap between collection–level information and

[108] Open Archives Forum "Information Resource Database" is located at: http://www.oaforum.org/oaf_db/index.php, accessed on September 4, 2003.

direct access to full digital content. In summary, users need not only to understand the various levels of granularity of resources represented in the aggregation, but also the relationship of the resource to its originating source.[109] Users need to know how "collections" are defined and what types of resources—and at what level of granularity—they can expect to find. As discussed previously in this report, many of the aggregations under review need to amass more object-level data.

- In general, the aggregators can't provide the "context," or match the level of refinement of either the originating source "database" or of their proprietary counterparts. Although the comparisons may be unfair, using *UIUC's Digital Gateway* to discover and retrieve images is primitive in comparison to going directly to *American Memory*; both pale in comparison to *RLG's Cultural Materials*. Similarly, users with access to Elsevier's *Science Direct* are unlikely to turn to *Scirus* to identify articles from 2003 in their field. *NDLTD's Union Catalogs* are not a substitute for *Dissertation Abstracts*. It should come as no surprise—as both the ARL and AARLIN scholar portal projects make clear—that access to proprietary databases is the highest priority among users. Many of the open access services under review in this report—with a few notable exceptions like *arXix*—aren't even on users' radar screens. Users need to understand the purpose and function of these services in order to know when to turn to them in preference to tools with which they are already familiar. To this end, more targeted comparative studies are needed to understand how users seek and find information across a variety of open access and proprietary sources. In short, for most users, it is not yet clear where these new tools fit into their search and discovery strategies, nor have most imagined building a personal digital library, or collaborating with colleagues in virtual workspaces.

7.2 Future Directions

7.2.1 More Attention to Users and Uses

Although many of the services under review have been informed by user studies of various types (e.g., *a priori*, focus group, iterative, continuous feedback), broader and deeper studies are needed. Borgman [2002a, 2002b], Fuhr et al. [2001], and Van House [2003] all provide conceptual frameworks, emphasizing holistic approaches to digital library evaluation that take into account *users* and *uses* within specific contexts. Moreover, this concern is international in scope and cuts across all sectors of digital library development. A few examples:

- In early September 2003, Helsinki University Library and The National Library of Finland are sponsoring an international conference: "Toward a User-Centered Approach to Digital Libraries."

[109] *DLESE* "Resource Granularity" document for catalogers: http://www.dlese.org/Metadata/cataloging/resource.htm, accessed on September 5, 2003.

[See program at: http://www.lib.helsinki.fi/finelib/digilib/programme.html]

- "Primarily History: Historians and the Search for Primary Sources" is a large-scale research project being conducted in the UK and the U.S. to: discover how historians are searching for and locating primary source materials; how they are teaching/advising their students to do so; and how archivists and other cultural heritage curators can best facilitate such information discovery.
  [See project description and questionnaires at: http://www.hatii.arts.gla.ac.uk/research/historians/primarily_history.htm]

- In a paper delivered at NSF's "Wave of the Future" workshop, "End-User Issues Should Have First Class Status," Terrence Smith [2003] exhorts:
  The time has come to treat both end-users and knowledge about end-users as first class entities in the development of electronic information environments that support research and learning. First class status in this case implies that they are as much an object of research and development as the information technology itself… A systematic and applicable understanding of how researchers and learners in any scholarly environment discover, learn, and apply information is surprisingly scarce, given the enormous literature on human perception, cognition, and behavior. [Paper available at: http://www.sis.pitt.edu/~dlwkshop/paper_smith.html]

7.2.2 Finding Solutions to Digital Rights Management and Digital Content Preservation

Solutions are needed for managing digital rights and for preserving digital content, if the services under review are expected to grow and flourish. Many promising initiatives are underway; a sample of more 2003 reports and developments follows.

Rights Management

"Open Archives and Intellectual Property: Incompatible World Views?", a report issued by the Open Archives Forum in November 2002 that discusses the relationship between open archives and Intellectual Property Rights (IPR). It explains IPR, the issues of copyright and its protection on the network, IPR in metadata and in resources, attitudes of stakeholders in IPR and open archives, and makes some initial recommendations. There is ultimately no conflict between Open Archives and Intellectual Property - but open archives must work within the framework of Intellectual Property law as outlined here. http://www.oaforum.org/

RoMEO: Rights Metadata for Open Archiving: the Open Archives Initiative expects to set up a technical committee soon in collaboration with the JISC RoMEO project, in the realm of expressing rights statements about metadata and content in the OAI framework. http://www.lboro.ac.uk/departments/ls/disresearch/romeo/

RoMEO Studies 4: An Analysis of Journal Publishers' Copyright Agreements http://www.lboro.ac.uk/departments/ls/disresearch/romeo/RoMEO%20Studies%204.pdf

IEEE's Learning Technology Standards Committee (LTSC): *Recommended Practice for Digital Rights Expression Languages (DRELs) Suitable for eLearning Technologies* http://ltsc.ieee.org/wg4/index.html

Besek [2003] reports on: *Copyright Issues Relevant to the Creation of a Digital Archive: A Preliminary Assessment*. http://www.clir.org/pubs/abstract/pub112abst.html

Preservation

Friedlander [2002] reports on "The National Digital Information Infrastructure Preservation Program: Expectations, Realities, Choices and Progress to Date."
http://www.dlib.org/dlib/april02/friedlander/04friedlander.html

Smith [2003] surveys the landscape of "New-Model Scholarship: How Will it Survive? http://www.clir.org/pubs/abstract/pub114abst.html

Beagrie [2003] reports on "National Digital Preservation Initiatives: An Overview of Developments in Australia, France, the Netherlands, and the United Kingdom and of Related International Activity."
http://www.clir.org/pubs/abstract/pub116abst.html

Jones [2003] gives an up-to-date account of UK digital preservation plans that focus on four major categories of digital content: deposited material, Web sites, digitization, and digital materials purchased for the provision of services.
http://www.ifla.org/IV/ifla69/papers/129e-Jones.pdf

Announced at IFLA 2003: The *ERPAePRINTS Service* is an Open Archive set-up for the Electronic Resource Preservation and Access Network (ERPANET) in conjunction with DAEDALUS, to provide an eprints' preservation and access facility for the cultural and scientific heritage community. http://daedalus.lib.gla.ac.uk/

7.2.3 Building Personal Libraries and Collaborative Work Spaces

A number of the services included in this report illustrate the potential for building personal libraries and collaborative workspaces (e.g., *Cyclades, Sheet Music Consortium, NSDL, AmericanSouth*). However, there is still a long way to go before these functions are fully supported. Borgman's 2003 NSF workshop paper on "Personal digital libraries" describes the limits of current practice and points to future directions. Also at the NSF workshop, Gennari et al. [2003] offer a framework of functions supporting collaboration systems that

identifies services (e.g., document management, calendaring-scheduling) and their features at "basic" and "extended" levels.

7.2.4 Putting "Digital Libraries in the Classroom" and Digital Objects in the Curriculum

There is growing evidence of communication and collaboration between the digital library and digital learning communities. This is necessary as a next step to put digital library collections and objects to use in "external" (non-library-centric) environments, including the classroom. Several examples of developments from 2003 follow:

- McLean and Lynch [2003] outline the challenges in a white paper from the Coalition for Networked Information and IMS Global Learning Consortium: "Interoperability between Information and Learning Environments—Bridging the Gaps." http://www.imsglobal.org/DLims_white_paper_publicdraft_1.pdf

- "Digital Libraries in the Classroom" is an international collaboration between the UK's JISC and the NSF, funded through 2006 "to bring about significant improvements in the learning and teaching process, through bringing emerging technologies and readily available digital content into mainstream educational use." http://www.jisc.ac.uk/index.cfm?name=programme_dlitc

- "COLIS: Collaborative Online Learning & Information Services" is a consortium of Australian universities with research support from OCLC, funded by the Australian government, that aims to develop a scalable standards based model for institutional interoperability, which enables the seamless sharing of online learning and scholarly information resources. It is conducting research on harvesting metadata for learning objects, and communicating and transferring the metadata to different computer systems that support online learning environments.[110] http://www.colis.mq.edu.au/index.html

- "Digital Culture: DigiCULT" has a compilation of links to sources about Learning Objects. http://www.digicult.info/pages/links.php?t=11

7.2.5 Promoting Excellence

In closing, the following initiatives hold promise for the future development of aggregated digital libraries.

- Digital Libraries Phase 2: NSF and its many partners http://www.dli2.nsf.gov/
  - o DLI2: International Projects http://www.dli2.nsf.gov/intl.html
  - o *Wave of the Future NSF Post Digital Libraries Futures Workshop*, June 15-17 2003 http://www.sis.pitt.edu/%7Edlwkshop/

[110] Extracted from COLIS Web site and from *OCLC Newsletter*, October 2002, p. 16.

- JISC Strategic Activities (The Joint Information Systems Committee, UK) http://www.jisc.ac.uk/index.cfm?name=about_strategic
- DELOS: Network of Excellence on Digital Libraries (European Union) http://delos-noe.iei.pi.cnr.it/

8.0 Major Web Sites Cited

All URLs were active as of date of publication.

AARLIN: the Australian Academic and Research Library Network
http://www.aarlin.edu.au/index.html

Advanced Library Collection Management Environment (OCLC)
http://www.oclc.org/research/projects/archive/alcme.htm

American Memory: Historical Collections for the National Digital Library, Library of Congress
http://memory.loc.gov/ammem/

AmericanSouth.org (Emory University with ASERL)
http://www.americansouth.org

APS Archive of Teaching Resources (American Physiological Society)
http://www.apsarchive.org/main/index.asp

Arc: A Cross Archive Search Service
http://arc.cs.odu.edu

ARCHON
http://archon.cs.odu.edu/

ARL Scholars Portal
http://www.arl.org/access/scholarsportal/

arXiv.org
http://arxiv.org

Association for Biology Laboratory Education (ABLE) (affiliated with BEN)
http://www.zoo.utoronto.ca/able/

Australia Dancing
www.australiadancing.org

Australia Subject Gateways
http://www.nla.gov.au/initiatives/sg/

BEN: A Digital Library of the Biological Sciences for Biology Teaching
http://www.biosciednet.org/portal

Biome, Health and Life Sciences (JISC, UK)
http://biome.ac.uk/

BioMed Central (BMC)
http://www.biomedcentral.com

BioMoleculesAlive.org (American Society for Biochemistry and Molecular Biology) (affiliated with BEN)
http://www.biomoleculesalive.org/

A Celebration of Women Writers (University of Pennsylvania)
http://digital.library.upenn.edu/women/

CERN Document Server
http://cdsweb.cern.ch/

Citebase
http://citebase.eprints.org/cgi-bin/search

CiteSeer (aka ResearchIndex)*
http://citeseer.ist.psu.edu/

Coalition for Networked Information (CNI)
http://www.cni.org

COLIS (Collaborative Online Learning & Information Systems, Australia)
http://www.colis.mq.edu.au/

CPS: Chemistry Preprint Server (Elsevier)
http://www.sciencedirect.com/preprintarchive

Cornucopia (UK)
http://www.cornucopia.org.uk

Crossroads, Discovering West Midlands Collections (UK)
http://www.crossroads-wm.org.uk/

Cyclades (European Union)
http://www.ercim.org/cyclades/

DELOS: Network of Excellence on Digital Libraries (European Union)
http://delos-noe.iei.pi.cnr.it/

DLI2: Digital Libraries Initiative Phase 2 (NSF)
http://www.itrd.gov/pubs/blue00/digital_libraries.html

Digital Library Federation
http://www.diglib.org

DLESE (Digital Library for Earth System Education)
http://www.dlese.org/dds/index.jsp

DP9, An OAI Gateway Service for Web Crawlers
http://arc.cs.odu.edu:8080/dp9/index.jsp

DSpace (MIT)
www.dspace.org

EcoEdNet (Ecological Society of America) (affiliated with BEN)
http://www.ecoed.net/

EDL: The Ethnomathematics Digital Library (ENC and NSDL affiliated)
http://www.ethnomath.com/

EEVL, Engineering, Mathematics and Computing (RDN/JISC, UK)
http://www.eevl.ac.uk

ENC (Eisenhower National Clearinghouse for Mathematics and Science Education)
http://www.enc.org

EPrints.org (projects and open source software; Joint Information Systems Council, UK)
http://www.eprints.org/

ERPAePRINTS Service
http://daedalus.lib.gla.ac.uk/

FAST (proprietary software used by Scirus)
http://www.fastsearch.com/

Flashpoint (Los Alamos National Laboratory)
http://lib-www.lanl.gov/lww/flashpoint.htm

GILS (Global Information Locator System)
http://www.gils.net

GSDL: The Gender and Science Digital Library (ENC & NSDL affiliated)
http://www.gsdl.enc.org

Grainger Engineering Library at University of Illinois-Urbana-Champaign
http://g118.grainger.uiuc.edu/engroai

Heritage Colorado (Colorado Digital Project)
http://www.cdpheritage.org

HUMBUL Humanities Web (RDN/JISC, UK)
http://www.humbul.ac.uk/

ICON: Innovative Curriculum Online Network (ENC affiliated)
http://icontechlit.enc.org

Images Canada
http://www.imagescanada.ca

InfoMine, Scholarly Internet Resource Collections
http://infomine.ucr.edu/

IMLS (Institute of Museum and Library Services)
http://www.imls.gov

iVia (INFOMINE's open source software)
http://infomine.ucr.edu/iVia/

JISC, Joint Information Systems Committee of the Higher Education Funding Councils, UK
http://www.jisc.ac.uk/

The Learning Matrix (ENC affiliated)
http://thelearningmatrix.enc.org/

The Linguist List (OLAC affiliated)
http://www.linguistlist.org/

MacquarieNet (Australia)
http://www.macnet.mq.edu.au

MathSciNet (American Mathematical Society)
http://www.ams.org/mathscinet

MERLOT (Multimedia Educational Resource for Learning and Online Teaching)
http://www.merlot.org

MetaArchive Initiative (Emory University with ASERL)
http://www.metaarchive.org/

MetaScholar Initiative (Emory University with ASERL)
http://www.metascholar.org

MicrobeLibrary (American Society for Microbiology)
http://www.microbelibrary.org/

MINERVA eEurope: Ministerial NetwoRk for Valorising Activities in digitization
http://www.minervaeurope.org/home.htm

MPS: Mathematics Preprint Server (Elsevier)
http://www.sciencedirect.com/preprintarchive

Music Australia
http://www.musicaustralia.org

NASA Technical Reports Server (NTRS)
http://ntrs.nasa.gov

National Library of Australia: Digitisation of Traditional Format Library Materials
http://www.nla.gov.au/digital/program.html

National Science Digital Library (NSF)
http://nsdl.org

NDLTD (Networked Digital Library of Theses & Dissertations)
http://www.ndltd.org

NDLTD Union Catalog (VTLS)
http://zippo.vtls.com/cgi-bin/ndltd/chameleon

NDLTD Union Catalog (OCLC)
Electronic Thesis/Dissertation OAI Union Catalog based at OCLC
http://rocky.dlib.vt.edu/~etdunion/cgi-bin/OCLCUnion/UI/index.pl

OAIster
http://oaister.umdl.umich.edu/o/oaister

OLAC: Open Language Archives Community
http://www.language-archives.org

Online Books Page
http://digital.library.upenn.edu/books/

Open Access News
http://www.earlham.edu/~peters/fos/fosblog.html

Open Archives Forum (EU-funded, partners: University of Bath-UKOLN (United Kingdom), Istituto di Scienza e Tecnologie della Informazione-CNR (Italy) and Computer- and Media Service (Computing Center) of Humboldt University (Germany).
http://www.oaforum.org/

Open Archives Initiative
http://www.openarchives.org/
OAI Registered Data Providers
http://www.openarchives.org/Register/BrowseSites.pl

OAI Registered Service Providers
http://www.openarchives.org/service/listproviders.html
Open Archives Initiative—Repository Explorer
http://oai.dlib.vt.edu/cgi-bin/Explorer/oai2.0/testoai

The Perseus Digital Library
http://www.perseus.tufts.edu/

Picture Australia
http://www.pictureaustralia.org

PostNuke (open source software used by AmericanSouth)
http://www.postnuke.org

PSIGate: Physical Sciences Information Gateway (RDN/JISC, UK)
http://www.psigate.ac.uk/

Public Library of Science (PLoS)
http://www.publiclibraryofscience.org

PubMed Central
http://www.pubmedcentral.gov

Resource Discovery Network (JISC, UK)
http://www.rdn.ac.uk/

ResourceShelf, Resources and News for Information Professionals
http://www.resourceshelf.com

RoMEO (Rights MEtadata for Open archiving, JISC Project)
http://www.lboro.ac.uk/departments/ls/disresearch/romeo/

Scirus (Elsevier)
http://www.scirus.com/srsapp/

SearchEngineWatch.com
http://www.searchenginewatch.com

Sheet Music Consortium
http://digital.library.ucla.edu/sheetmusic/

SMETE: Science, Math, Engineering and Technology Education Library
http://www.smete.org/smete

Social Science Information Gateway (RDN/JISC, UK)
http://www.sosig.ac.uk/

Society of Toxicology (affiliated with BEN)
http://www.toxicology.org/

SPARC Open Access Newsletter
http://www.earlham.edu/~peters/fos/index.htm

STKE, Signal Transduction Knowledge Environment (affiliated with BEN)
(American Association for the Advancement of Science)
http://stke.sciencemag.org/

Subject Portals Project (RDN/JISC)
http://www.portal.ac.uk/spp/

TORII (International School for Advanced Studies, Trieste, Italy)
http://torii.sissa.it

UIUC Digital Gateway to Cultural Heritage Materials
http://nergal.grainger.uiuc.edu/cgi/b/bib/bib-idx

UIUC Open Archives Initiative Metadata Harvesting Project
http://oai.grainger.uiuc.edu/

U.S. States Implementing GILS
http://states.gils.net

Vascoda (Multidisciplinary Subject Gateways in Germany)
http://www.vascoda.de/

Voice of the Shuttle: Web site for Humanities Research
http://vos.ucsb.edu

XTCat (ALCME/OCLC and NDLTD)
http://alcme.oclc.org/ndltd/SearchbySru.html

9.0 Bibliography of Cited Works and Further Reading

Arms, Caroline R. (2003). **Available and Useful: OAI at the Library of Congress**. *Library Hi Tech* 21, 2: 129-139.

Arms, William, et al. (2002). **A Spectrum of Interoperability: The Site for Science Prototype for the NSDL.** *D-Lib Magazine* 8, 1 (January) at: http://www.dlib.org/dlib/january02/arms/01arms.html, accessed on September 5, 2003.

Arms, William Y., Naomi Dushay, Dave Fulker, and Carl Lagoze (2003). **A Case Study in Metadata Harvesting: The NSDL**. *Library Hi Tech* 21, 2: 228-237.

Beagrie, Neil [2003]. **National Digital Preservation Initiatives: An Overview of Developments in Australia, France, the Netherlands, and the United Kingdom and of Related International Activity.**

Washington, DC: Council on Library & Information Resources (April); at http://www.clir.org/pubs/abstract/pub116abst.html, accessed on September 6, 2003.

Benson, David A., et al. (2003). **GenBank**. *Nucleic Acids Research* 31, 1: 23-27; at http://nar.oupjournals.org/cgi/content/full/31/1/23, accessed on September 8, 2003.

Borgman, Christine L. (2002a). **Challenges in Building Digital Libraries for the 21st Century**. In: Lim, E-P.; Foo, S.; & Khoo, C. (eds.). (2002). *Digital Libraries: People, Knowledge & Technology: Proceedings of the 5th International Conference on Asian Digital Libraries* (ICADL 2002), Singapore. December 12-14, 2002. *Lecture Notes in Computer Science* 2555: 1-13; Table of contents and abstract accessible at: http://www.springer.de/comp/lncs/index.html

_____ (2002b). **Final Report to the National Science Foundation**. Fourth DELOS Workshop. *Evaluation of Digital Libraries: Testbeds, Measurements, and Metrics*. Hungarian Academy of Sciences, Computer and Automation Research Institute (MTA SZTAKI), Budapest, Hungary, 6-7 June 2002. Grant IIS-0225626; at http://www.sztaki.hu/conferences/deval/presentations/final_report.html, accessed on August 5, 2003.

_____ (2003). **Personal Digital Libraries: Creating Individual Spaces for Innovation**. *Wave of the Future: NSF Post Digital Library Futures Workshop*, June 15-17, Cape Cod; at http://www.sis.pitt.edu/~dlwkshop/paper_borgman.html, accessed on August 18, 2003.

Boss, Richard W. (2002). **How to Plan and Implement a Library Portal.** *Library Technology Reports*, (Nov-Dec.): 1-61.

Brown, Cecelia M. (2002). **The Coming of Age of E-prints in the Literature of Physics**. *Issues in Science and Technology Librarianship* 31 (Summer); at http://www.istl.org/01-summer/refereed.html, accessed on August 19, 2003.

Brown, Cecelia M. (2003). **The Role of Electronic Preprints in Chemical Communication: Analysis of Citation, Usage, and Acceptance in the Journal Literature**. *Journal of the American Society for Information Science and Technology* 54, 5: 362-372.

Cameron, Jasmine (2003). **Die Barieren abbauen: Neue Richtlinien fuer die Nationalbiblioghek von Australien**. paper delivered at World Library & Information Congress, 69th IFLA General Conference and Council, 1-9 August, Berlin; at http://www.ifla.org/IV/ifla69/papers/027g_trans-Cameron.pdf, accessed on September 7, 2003.

Carpenter, Leona, et al. (2002). **Interim Review of Organisational Issues**. Open Archives Forum, IST-200132015 Project No.; draft D.3.1 (November): 20p. at http://www.oaforum.org/otherfiles/oaf_d31_organisational1.pdf, accessed on August 5, 2003.

Chien, Yi-Tzuu (2002). **Whither Digital Libraries? The Case of a "Billion-Dollar" Business**. (October 31); at http://www.sis.pitt.edu/%7Edlwkshop/supplement/YT_Chien.ppt, accessed on August 17, 2003.

Cole, Timothy W. (2003a). **Open Archives Initiative Metadata Harvesting**. Special issue of *Library Hi Tech* 21,2: 111-228. (Individual articles listed separately as well.)

_____ (2003b). **Using OAI: Innovations in Sharing of Information**. Guest editorial for special issue on "Open Archives Initiative metadata harvesting" *Library Hi Tech* 21, 2: 115-117.

Crane, Gregory, et al. (2003). **Towards a Cultural Heritage Digital Library**. In *2003 Joint Conference on Digital Libraries*, Houston, Texas, (June): 75-84; at http://www.perseus.tufts.edu/Articles/jcdl2003.pdf, accessed on August 26, 2003.

Day, Michael (2003). **Prospects for Enstitutional E-print Repositories in the United Kingdom**. ePrints UK supporting study, no.1, version 1.0, (28 May): 1-19; at http://www.rdn.ac.uk/projects/eprints-uk/docs/studies/impact/, accessed on August 5, 2003.

Dobratz, Susanne, and Birgit Matthaei (2003). **Open Archives Activities and Experiences in Europe.** An Overview by the Open Archives Forum. *D-Lib Magazine* 9, 1 (January); at http://www.dlib.org/dlib/january03/dobratz/01dobratz.html, accessed on August 5, 2003.

European Commission (2002). **Coordinating Digitisation in Europe**. Progress report of the National Representatives Group: coordination mechanisms for digitization policies and programmes 2002. European Commission: The Information Society Directorate-General, 243p.; at http://www.minervaeurope.org/publications/globalreport.htm, accessed on August 5, 2003.

Friedlander, Amy [2002]. **The National Digital Information Infrastructure Preservation Program: Expectations, Realities, Choices and Progress to Date.** *D-Lib Magazine,* 8, 4 (April); at: http://www.dlib.org/dlib/april02/friedlander/04friedlander.html, accessed on September 6, 2003.

Fuhr, Norbert, Preben Hansen, Michael Mabe, Andras Micsik, and Ingeborg Solvberg (2001). **Digital Libraries: A Generic Classification and Evaluation Scheme.** "Research and Advanced Technology for Digital Libraries: 5th European Conference, ECDL 2001, Darm-

stadt, Germany, September 4-9, 2001, Proceedings" in *Lecture Notes in Computer Science* 2163 (January): 187-199.

Gennari, Jeffrey, et al. (2003). **Preparatory Observations Ubiquitous Knowledge Environments: The Cyberinfrastructure Information Ether**. *Wave of the Future: NSF Post Digital Library Futures Workshop*, June 15-17, Cape Cod; at http://www.sis.pitt.edu/~dlwkshop/paper_spring.html, accessed on August 18, 2003.

Ginsparg, Paul (2003). **Can Peer Review Be Better Focused?** Final version of 13 March 2003 at http://arxiv.org/blurb/pg02pr.html accessed on August 5, 2003.

_____ (1994). **First Steps Towards Electronic Research Communication**, adapted from *Computers in Physics* 8, 4 (Jul/Aug): 390-396; updated version of April 1995 available at: http://arxiv.org/ftp/hep-th/papers/macros/blurb.tex, accessed on August 8, 2003.

Ginsparg, Paul, Rick Luce, and Herbert Van de Sompel (1999). **The Open Archives Initiative Aimed at the Further Promotion of Author Self-Archived Solutions**, Universal PrePrint Service (UPS) Meeting, available at: http://www.openarchives.org/meetings/SantaFe1999/ups-invitation-ori.htm, accessed on August 8, 2003.

Greenstein, Daniel, and Suzanne E. Thorin (2002). **The Digital Library: A Biography**. Washington, DC: Digital Library Federation and Council on Library Resources, 69 pp.

Hagedorn, Kat (2003). **OAIster: A "No Dead Ends" OAI Service Provider**. *Library Hi Tech* 21, 2: 170-181.

Hagen, John H., Susan Dobratz, and Peter Schirmbacher (2003). **Electronic Theses and Dissertations Worldwide: Highlights of the ETD 2003 Symposium**. *D-Lib Magazine* 9, 7/8 (July/August); at http://www.dlib.org/dlib/july03/hagen/07hagen.html, accessed on August 20, 2003.

Halbert, Martin (2003). **The Metascholar Initiative: AmericanSouth.org and MetaArchive.org**. *Library Hi Tech* 21, 2: 182-198.

Halbert, Martin, Sandra Nyberg, John Burger, and Kate Nevins (2002). **AmericanSouth.Org: A Collaborative Project to Improve Access to Digital Resources on Southern History and Culture**. Interim Report to the Andrew W. Mellon Foundation, (August): 23 p.; at http://www.metascholar.org/docs/reports/InterimReport-Amsouth-Final-Revised.doc, accessed on September 4, 2003.

Hitchcock, S. (2003). **Metalist of Open Access E-print Archives: The Genesis of Institutional Archives and Independent Services**. *ARL*

Bimonthly Report 227(April); at http://www.arl.org/newsltr/227/metalist.html, accessed on August 5, 2003. Updated **Core metalist of open access e-print archives** at http://opcit.eprints.org/explorearchives.shtml, accessed on August 5, 2003.

Institute of Museum and Library Services (2001a). **A Framework of Guidance for Building Good Digital Collections.** (November 6); at http://www.imls.gov/pubs/forumframework.htm, accessed on August 5, 2003.

_____ (2001b). **Report of the IMLS Digital Library Forum on the National Science Digital Library Program** produced jointly with representatives of the National Science Foundation's Science, Math, Engineering and Technology Education digital library project (otherwise known as the National Science Digital Library, or NSDL program) (October); at http://www.imls.gov/pubs/natscidiglibrary.htm, accessed on September 4, 2003.

Jennings, Simon (July 2002). **RDN Collections Development Framework**. Version 1.2 at; http://www.rdn.ac.uk/publications/collections/cdframework3.doc, accessed on September 4, 2003.

Johnston, Dean H. (2003). **News from Online: Untangling the Web-The National Digital Libraries Initiative**. *Chemical Education Today*. 80, 7 (July): 733-734.

Jones, Maggie (2003). **Digital Preservation Activities in the U.K. – Building the Infrastructure**. paper delivered at World Library & Information Congress, 69th IFLA General Conference and Council, 1-9 August, Berlin. at http://www.ifla.org/IV/ifla69/papers/129e-Jones.pdf, accessed on September 7, 2003.

Kling, Rob, and Geoff McKim (2000). **Not Just a Matter of Time: Field Differences and the Shaping of Electronic Media in Supporting Scientific Communication**. *Journal of the American Society for Information Science*, 51: 1306-1320.

Kling, Rob, Lisa Spector, and Geoff McKim (2002). **Locally Controlled Scholarly Publishing Via the Internet: The Guild Model**. *Proceedings of the 65th Annual Meeting of the American Society for Information Science and Technology*, 39: 228-238.

Koch, Traugott (2000). **Quality-Controlled Subject Gateways: Definitions, Typologies, Empirical Overview**. Manuscript of the article published in the Subject gateways special issue of *Online Information Review* 24, 1; at http://www.lub.lu.se/tk/publ/OIR-SBIG.html accessed on September 7, 2003.

Lagoze, Carl, and Herbert Van de Sompel (2003). **The Making of the Open Archives Protocol for Metadata Harvesting**. *Library Hi Tech* 21, 2: 118-128.

Lawal, Ibironke (2002). **Scholarly Communication: The Use and Non-Use of E-print Archives for the Dissemination of Scientific Information.** *Issues in Science and Technology Librarianship* (Fall): 15 p.; at http://www.istl.org/02-fall/article3.html, accessed on August 5, 2003.

Lervik, John M. (2003). **Digital Libraries: What Should We Expect from Search Engines?** Presentation at ECDL 2003 (August); at http://freepint.com/gary/lervik_fast.pdf, accessed on August 21, 2003.

Liu, Xiaoming (2002). **Federated Searching Interface Techniques for Heterogeneous OAI Repositories**. *Journal of Digital information* 2, 4; posted 21 May 2002 at http://jodi.ecs.soton.ac.uk/Articles/v02/i04/Liu/, accessed on August 6, 2003.

Lundmark, Cathy. (2003). **BEN: The Biology Branch of the National Science Digital Library**.*BioScience*, 53, 7 (July): 631.

Lynch, Clifford (2002). **Digital Collections, Digital Libraries and the Digitization of Cultural Heritage Collections**. *First Monday* 7, 5 (May); at http://www.firstmonday.dk/issues/issue7_5/lynch/index.html, accessed on August 5, 2003.

_____ (2003). **Reflections Towards the Development of a "Post-DL" Research Agenda**. *Wave of the Future: NSF Post Digital Library Futures Workshop*, June 15-17, Cape Cod; at http://www.sis.pitt.edu/~dlwkshop/paper_lynch.html, accessed on August 18, 2003.

Mahoney, Dan, and Mariella Di Giacomo (2001). **Flashpoint @ LANL.gov: A Simple Smart Search Interface**. *Issues in Science and Technology Librarianship* (Summer); at http://www.library.ucsb.edu/istl/01-summer/article2.html, accessed on August 23, 2003.

Maly, K., et al. (2002). **Archon – A Digital Library that Federates Physics Collections**. Posted 1 October 2002 at: http://kepler.cs.odu.edu:8080/testgroup/cache/oai.ODU_DLPublications.archon.pdf, accessed on August 7, 2003.

Marx, Vivien (2003). **TECHNOLOGY; In DSpace, Ideas Are Forever**. in EDUCATION LIFE SUPPLEMENT, Section 4A, Page 8, Column 1 *New York Times*. (August 3, 2003): available for purchase at http://www.nytimes.com/2003/08/03/edlife/03EDTECH.html, accessed on August 19, 2003.

McLean, Neil, and Clifford Lynch (2003). **Interoperability Between Information and Learning Environments—Bridging the Gaps.** A Joint White paper on behalf of the IMS Global Learning Consortium and the Coalition for Networked Information. Draft version of June 28, 2003: 13p.; at http://www.imsglobal.org/DLims_white_paper_publicdraft_1.pdf, accessed on August 5, 2003.

Mitchell, Steven, et al. (2003). **iVia Open Source Virtual Library System.** *D-Lib Magazine* 9, 1 (January); at http://www.dlib.org/dlib/january03/mitchell/01mitchell.html, accessed on August 31, 2003.

National Science Foundation (2003). **Wave of the Future: NSF Post Digital Library Futures Workshop**, June 15-17, Cape Cod; at http://www.sis.pitt.edu/~dlwkshop/index.html, accessed on August 18, 2003.

_____. Division of Undergraduate Education (2001). **Pathways to Progress: Vision and Plans for Developing the NSDL**. A white paper resulting from the collaborative efforts of the NSDL community workgroups during 2000-2001. (March 20); at http://doclib.comm.nsdlib.org/PathwaysToProgress.pdf, accessed on August 27, 2003.

_____. Division of Undergraduate Education (1996). **Shaping the Future: New Expectations for Undergraduate Education in Science, Mathematics, Engineering, and Technology**. A Report on the Review of Undergraduate Education from the Committee for the Review to the National Science Foundation Directorate for Education and Human Resources; at http://www.ehr.nsf.gov/ehr/due/documents/review/96139/start.htm, accessed on August 27, 2003.

Nelson, Michael L., JoAnne Rocker and Terry L. Harrison (2003). **OAI and NASA's Scientific and Technical Information**, *Library Hi Tech*, 21, 2: 140-150; at: http://techreports.larc.nasa.gov/ltrs/dublincore/2003/jp/NASA-2003-lht-mln.html, accessed on August 5, 2003.

Nicholson, Dennis (2003). **Ghosts in the Machine: People and Organization Level Issues in Distributed Digital Libraries**. *OCLC Systems & Services* 19, 1 (2003): 17-22.

The NINCH Guide to Good Practice in Digital Representation and Management of Cultural Heritage Materials (2002). by the Humanities Advanced Technology and Information Institute (HATII), University of Glasgow, and the National Initiative for a Networked Cultural Heritage (NINCH); at http://www.nyu.edu/its/humanities/ninchguide/, accessed on August 5, 2003.

Pacchioli, David (2003). **Smart Search**. *Research Penn State* 24, 2 (May): 6 pp.; at http://www.rps.psu.edu/0305/search.html, accessed on August 5, 2003.

Pianos, Tamara (2003). **Vascoda – a Portal for Scientific Resource Collections created by German Libraries and Information Centres**. paper delivered at World Library & Information Congress, 69th IFLA General Conference and Council, 1-9 August, Berlin. at: http://www.ifla.org/IV/ifla69/papers/055e-Pianos.pdf, accessed on September 7, 2003.

Pinfield, Stephen (2003). **Open Archives and UK Institutions: An Overview**. *D-Lib Magazine* 9, 3 (March); at http://www.dlib.org/dlib/march03/pinfield/03pinfield.html, accessed on August 5, 2003.

Pinfield, Stephen, Mike Gardner, and John MacColl (2002). **Setting Up an Institutional E-Print Archive**. *Ariadne* 31 (April); at: http://www.ariadne.ac.uk/issue31/eprint-archives/intro.html, accessed on August 5, 2003.

Price, Gary, *guest editor*, (2003). **A Search Haven for Engineers**. *SearchDay* column of Augut 6, 2003 for SearchEngineWatch.com; at http://www.searchenginewatch.com/searchday/article.php/2235331, accessed on August 7, 2003.

Renda E. M., and U. Straccia (forthcoming 2003/04). **A Personalized Collaborative Digital Library Environment: A Model and an Application.** *forthcoming in Information Processing & Management*, Elsevier; as announced at http://www.ercim.org/cyclades/pub.html, accessed on August 24, 2003.

Roberts, Joni R. and Drost Carol A., eds. (2003). **Internet Reviews: Social Science Information Gateway (SOSIG).** *C&RL News* 64, 7 (July/August): 477-478.

Roempler, Kimberly S. (2002a). **Building an Infrastructure to Support STEM Digital Library Collections**. Paper submitted to the Joint Conference on Digital Libraries 2002 National Conference submission and available as a word document; at http://thelearningmatrix.enc.org/documents/RoemplerJCDL2002.doc, accessed on August 31, 2003.

_____ (2002b). **ENC Partners: National Science Foundation**. *ENC Focus* 9, 2: 6-7; at http://www.enc.org/features/focus/archive/across/document.shtm?input=FOC-002767-index, accessed on August 31, 2003.

Schatz, Bruce R. (2003). **Navigating the Distributed World of Community Knowledge**. *Wave of the Future: NSF Post Digital Library Futures Workshop*, June 15-17, Cape Cod; at http://www.sis.pitt.edu/~dlwkshop/paper_schatz.html, accessed on August 18, 2003.

Schiff, Frederick (2003). **Business Models of News Web sites: A Survey of Empirical Trends and Expert Opinion**. *First Monday* 8, 6 (June); at http://firstmonday.org/issues/issue8_6/schiff/index.html, accessed on August 17, 2003.

Schmidt, Janine, Anne Horn, and Barbara Thorsen (2003). **Australian Subject Gateways, the Successes and Challenges**. paper delivered at World Library & Information Congress, 69th IFLA General Conference and Council, 1-9 August, Berlin. at: http://www.ifla.org/IV/ifla69/papers/166e-Schmidt_Horn_Thorsen.pdf

Sequeira, Edwin (2003). **PubMed Central--Three Years Old and Growing Stronger**. *ARL Bimonthly Newsletter*, 228 (June): 5-9; at http://www.arl.org/newsltr/228/pubmed.html, accessed on August 21, 2003.

Sherman, Chris, and Gary Price (2001). **The Invisible Web: Uncovering Information Sources Search Engines Can't See**. Medford, NJ: Information Today, Inc.

Shreeves, Sarah L., Joanne S. Kaczmarek, and Timothy W. Cole (2003). **Harvesting Cultural Heritage Metadata Using the OAI Protocol**. *Library Hi Tech* 21, 2: 159-169.

Simons, Gary, and Steven Bird (2003a). **Building an Open Language Archives Community on the OAI Foundation**. *Library Hi Tech* 21, 2: 210-218.

Simons, Gary, and Steven Bird (2003b). **Seven Dimensions of Portability for Language Documentation and Description** *Language* 79 (2003): 557-582; at http://www.ldc.upenn.edu/sb/home/papers/0204020/0204020-revised.pdf, accessed on August 23, 2003.

_____ (2003c). **The Open Language Archives Community: An Infrastructure for Distributed Archiving of Language Resources**, 10 p. *Literary and Linguistic Computing* 18(2): 117-128, special issue on "New Directions in Humanities Computing"; at http://arxiv.org/ftp/cs/papers/0306/0306040.pdf, accessed on August 21, 2003.

Smith, Abby (2003). **New-Model Scholarship: How Will It Survive?** Washington DC: Council on Library & Information Resources, (March); at http://www.clir.org/pubs/abstract/pub114abst.html, accessed on September 6, 2003.

Smith, Terence R. (2003). **End-User Issues Should Have a First Class Status**. *Wave of the Future: NSF Post Digital Library Futures Workshop*, June 15-17, Cape Cod; at http://www.sis.pitt.edu/~dlwkshop/paper_smith.html, accessed on August 18, 2003.

Stoklasova, Bohdana (2003). **Short Survey of Subject Gateways Activity**. paper delivered at World Library & Information Congress, 69th IFLA General Conference and Council, 1-9 August, Berlin. at: http://www.ifla.org/IV/ifla69/papers/152e-Stoklasova.pdf, accessed on September 7, 2003.

Suleman, Hussein, and Edward A. Fox. (2003). **Leveraging OAI harvesting to Disseminate Theses**. *Library Hi Tech* 21, 2: 219-227.

Tennant, Roy (2003). **Science Portals**. *Library Journal* 128, 5: 34.

Van de Sompel, Herbert (2003a). **Developing New Protocols to Support and Connect Digital Libraries.** *OCLC Newsletter* 261 (July); at http://www.oclc.org/news/e-newsletter/n261/interview.htm, accessed on September 3, 2003.

_____ (2003b). **Roadblocks**. *Wave of the Future: NSF Post Digital Library Futures Workshop*, June 15-17, Cape Cod; at http://www.sis.pitt.edu/~dlwkshop/paper_sompel.html, accessed on August 18, 2003.

Van de Sompel, Herbert, and Lagoze, Carl (2000). **The Santa Fe Convention of the Open Archives Initiative**. *D-Lib Magazine* 6, 2; at http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html, accessed on August 8, 2003.

Van House, Nancy A. (forthcoming 2003). **Digital Libraries and Collaborative Knowledge Construction**, To appear in: Ann. P. Bishop, Barbara P. Buttenfield, and Nancy A. Van House, eds. *Digital Library Use: Social Practice in Design and Evaluation*, MIT Press; at http://www.sims.berkeley.edu/~vanhouse/van_house_book_chapter.htm, accessed on August 5, 2003.

Warner, Simeon (2003). **E-prints and the Open Archives Initiative**. *Library Hi Tech* 21, 2: 151-158.

Waters, Donald J. (2001). **The Metadata Harvesting Initiative of the Mellon Foundation**. *ARL Bimonthly Report* 217 (August): 4 p.; at http://www.arl.org/newsltr/217/waters.html, accessed on August 5, 2003.

_____ (2003). **Beyond Digital Libraries: The Organizational Design of a New Cyberinfrastructure**. *Wave of the Future: NSF Post Digital Library Futures Workshop*, June 15-17, Cape Cod; at http://www.sis.pitt.edu/~dlwkshop/paper_waters.html, accessed on August 18, 2003.

Wiederhold, Gio (2003). **Increasing the Information Density in Digital Library Results**. *Wave of the Future: NSF Post Digital Library Futures Workshop*, June 15-17, Cape Cod; at http://www.sis.pitt.edu/~dlwkshop/paper_wiederhold.html, accessed on August 18, 2003.

Wilkin, John, Kat Hagedorn, and Mike Burek (2002). **Creating an Academic Hotbot: Final Report of the University of Michigan OAI Harvesting Project**. (January): 1-18; at http://oaister.umdl.umich.edu/o/oaister/mellon-harvesting-final.doc, accessed on August 5, 2003.

Zorich, Diane M. (2003). **A Survey of Digital Cultural Heritage Initiatives and Their Sustainability Concerns.** Washington, DC: Council on Library and Information Resources (June), 47pp.

FURTHER READING

Atkins, Daniel E., et al. (2003). **Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue Ribbon Advisory Panel on Cyberinfrastructure**. (January); at http://www.communitytechnology.org/nsf_ci_report/ accessed on August 18, 2003.

Berners-Lee, Tim, James Hendler, and Ora Lassila (2001). **The Semantic Web**. *Scientific American* (May 17); retrievable by searching "semantic web" at http://www.scientificamerican.com, accessed on September 8, 2003.

Borgman, Christine L. (1999). **What are Digital Libraries? Competing visions**. *Information Processing and Management* 35: 227-243.

Brand, Amy, Frank Daly, and Barbara Meyers (2003). **Metadata Demystified: A Guide for Publishers**. The Sheridan Press & NISO Press, 19 pp.; at http://www.niso.org/standards/resources/Metadata_Demystified.pdf, accessed on September 8, 2003.

Chen, Ching-chih (2003). **Toward a Global Digital Library.** *Wave of the Future: NSF Post Digital Library Futures Workshop*, June 15-17, Cape Cod; at http://www.sis.pitt.edu/~dlwkshop/paper_chen_ching.html, accessed on August 18, 2003.

Downie, Stephen J.(2003). **Thoughts on the Present and Future of DL Research and Funding**. *Wave of the Future: NSF Post Digital Library Futures Workshop*, June 15-17, Cape Cod; at http://www.sis.pitt.edu/~dlwkshop/paper_downie1.html, accessed on August 18, 2003.

Hodge, Gail (2001). **Metadata Made Simpler**. Bethesda, MD: NISO Press, 16 pp; at http://www.niso.org/news/Metadata_simpler.pdf, accessed on September 8, 2003.

Lagoze, Carl (2003). **NSF DL Position Paper.** *Wave of the Future: NSF Post Digital Library Futures Workshop*, June 15-17, Cape Cod; at http://www.sis.pitt.edu/~dlwkshop/paper_lagoze.html, accessed on August 18, 2003.

Lesk, Michael (2003). **The Future of Digital Libraries.** *Wave of the Future: NSF Post Digital Library Futures Workshop*, June 15-17, Cape Cod; at http://www.sis.pitt.edu/~dlwkshop/paper_lesk.html, accessed on August 18, 2003.

Lynch, Clifford (2003). **Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age**. *ARL Bimonthly Report* 226 (February); at http://www.arl.org/newsltr/217/mhp.html, accessed on August 5, 2003.

_____ (2001). **Metadata Harvesting and the Open Archives Initiative**. *ARL Bimonthly Report* 217 (August): 12 p.; at http://www.arl.org/newsltr/217/mph.html, accessed on August 5, 2003.

MacNeil, Jane Salodof (2003). **Molecular Databases Grow, and Grow … and Grow**. *The* Scientist, 17, 15/40 (July 28); at http://www.the-scientist.com/yr2003/jul/lcprofile_030728.html, accessed on August 21, 2003.

Marchionini, Gary (2000). **Evaluating Digital Libraries: A Longitudinal and Multifaceted View**. *Library Trends* 49, 2 (Fall): 304-333.

Peters, Thomas A., Issue Editor (2000). **Assessing Digital Library Services**. Special issue of *Library Trends* 49, 2 (Fall); 221-390.

Prom, Christopher J. (2003). **Reengineering Archival Access through the OAI Protocols**. *Library Hi Tech* 21, 2: 199-209.

Seaman, David (2003). **Deep Sharing: A Case for the Federated Digital Library**. *EDUCAUSE Review* (July/August): 10-11; at http://www.educause.edu/ir/library/pdf/erm0348.pdf, accessed on August 18, 2003.

Stewart, M. Claire, and H. Frank Cervone (2003). **Building a New Infrastructure for Digital Media: Northwestern University Library.** *ITAL: Information Technology & Libraries*. 22, 2 (June): 69+.

Sun Microsystems, Inc. (2003). **E-Learning Framework**. Technical White Paper (February) 36 pp.; at http://www.sun.com/products-n-solutions/edu/whitepapers/pdf/framework.pdf, accessed on September 8, 2003.

APPENDIX 1

Scope Notes

Some resources were eliminated from further review because they seemed beyond the scope of this study. These are briefly discussed below.

- **MacquarieNet**, a fee-based subscription reference tool, which serves primarily as an "online encyclopedia" for and about Australia, with access to other international reference works, is targeted for use by primary and secondary school students and teachers. While it has some clever features, including icons to differentiate formats of materials (e.g., photos, Internet links, news articles, sound, etc.), daily news feeds from the Australian Associated Press, and teacher support services, such as downloadable lesson plans, homework assignments, and activity worksheets, it is a commercial database that cannot be accessed without a subscription. As a result, it is not only difficult for the non-subscriber to evaluate, but also largely outside the primary interests of the DLF constituency.

Another set of services were excluded because they function primarily as **"Resource Directories"**—presenting resources as a subject guide, in a hierarchical order or some other organizational scheme, such as by place or by type of publication. They do not have content of their own, only records (e.g., cataloging records, annotated records, records with a summary or description, or pointers to external content). Although they may be supported with a database backend or some content management system, they are largely sustained by volunteer specialists, who serve as editors and gatekeepers, where manual intervention is still required for some aspects of their operations. Typically they are useful for browsing and provide some level of subject searching.[111] In the long-term, if these sites don't automate more functions, their survival is at stake. Scalability is a critical issue—with ever-growing Web content, they will confront the Sisyphean task of keeping up with resource identification and linking. These include *The Online Books Page* and the *Voice of the Shuttle*:

- ***The Online Books Page***, in existence for over ten years and continuously edited by one person, relies on a network of volunteer contributors. The site permits searching and browsing more than 20,000 full-text books (and serials), freely available on the

[111] I am indebted to Jian Liu, reference librarian *extraordinaire* at Indiana University, for defining the characteristics of resource directories.

Internet. Three features include: *A Celebration of Women Writers*, *Banned Books Online*, and *Prize Winners Online*. *A Celebration of Women Writers* comprises 4,000 titles or 20% of the books online and has been registered as an OAI data provider with the Open Archives Initiative since March 2001. Both *UIUC's Digital Gateway to Cultural Heritage* (416 records) and *OAIster* (204 records) harvest from *A Celebration* a small number of locally-held full-text books. According to *The Online Books Page*, more implementation of the OAI-PMH is anticipated in the coming year.

- ***Voice of the Shuttle: Web site for Humanities Research*** has existed since 1994 under the auspices of the English Department at UC Santa Barbara. It figures prominently among a handful of *Forbes' Favorites* in the category of "Academic Research" where it is referred to as a "premier online destination for the humanities and social sciences, for casual surfers and die-hard researchers alike." *VoS* was rebuilt in 2002 as a database that serves content dynamically over the Web, but it continues to rely on human intervention to approve and edit links. Users who sign up for an account may contribute links and in the future will be given editorial privileges to maintain a file of contributed links along with the rights to edit them. Immediate access to suggested links is provided in the category of "Unvetted Submissions." *VoS* is also planning to activate group accounts that will enable classes, organizations, and conferences to build subsets of *VoS* resources, which will appear both on the regular *VoS* pages and on a special page set aside for the group (e.g., "English 130," "History 186," or "Conference 2001" *VoS Resources Page*). According to their Web site: "*VoS* will thus be an open platform serving the needs of both general and specific communities of users."

- ***Cornucopia, discovering UK Collections*** serves as an entry point for collection-level information about museums in the United Kingdom. It, too, may be construed as a "Resource Directory," although it operates with a more stable organizational infrastructure than the two preceding examples, and is sponsored by Resource: The Council for Museums, Archives & Libraries in the UK. Also, unlike the two preceding examples, its scope is comparatively static, namely the collections of 1,800 UK museums, so maintaining the database is not as daunting a task as keeping up with dynamic changes on the Web, although it still requires regular oversight and updating of records. *Cornucopia* has developed a template for collection-level descriptions (CLDs) that presents information in a well-organized fashion; however it predates the establishment of the Collection Description framework developed by UKOLN in support of the Research Support Libraries Programme (RSLP).[112] *Cornucopia's* searchable database contains

[112] Information about Collection Description Focus a national post, jointly funded by the British Library, the Joint Information Systems Committee (JISC), the Research Support Libraries Programme (RSLP) and Resource is available at: http://www.ukoln.ac.uk/cd-focus/ and RSLP's Collection Description project is available at: http://www.ukoln.ac.uk/metadata/rslp/

partial records from 1,800 UK museums with full records for 500+ museums in the SW and West Midlands region. Among its advanced search features, only the "subject" function is operable. Moreover, *Cornucopia* relies on proprietary software (Index+) and is currently very limited in its ability to manipulate and transfer information.

Meanwhile a promising pilot project is underway to evaluate *Crossroads*' architecture using *Cornucopia*'s data. *Crossroads* is another Resource-funded project to provide collections-level descriptions for pottery collections around the West Midlands region. A full project description is located at Cornucopia's Web site. *Crossroads* advantages include:

- Based on the RSLP Schema for collections-level description, the project uses open-source software to deliver information through an online search interface.
- Because it is based on open standards and has been fully tested, use of the *Crossroads* architecture as a mechanism for delivering *Cornucopia* means that it will be possible to deliver a functional and flexible online search with a greater degree of interactivity than is currently possible.[113]

Last, the ***U.S. States Implementing GILS*** was also removed from closer examination because the GILS standard (Government Information Locator Service)[114], as used by federal agencies, is very closely tied to Z39.50 and represents much more than a metadata schema. Although a few state libraries looked at using GILS to help make state government information more discoverable, in part with funding from IMLS, according to UIUC's Timothy Cole, they didn't have adequate resources to fully implement it and ended up focusing on subject classification trees and the descriptive metadata aspects of GILS.[115] He reports:

> A related schema, essentially a subset of the federal GILS schema, evolved as part of this work, and several state libraries now (e.g., Illinois) have focused on getting state agency Webmasters to at least embed metadata (expressed in this abbreviated GILS schema) in their HTML pages. Some states even created utilities to help them do this (http://www.finditillinois.org/metadata/index.html).

---

[113] The *Cornucopia* and *Crossroads* project description is available at: http://www.cornucopia.org.uk/xroads_spec.htm

[114] "The Government Information Locator Service (GILS) is an effort to identify, locate, and describe publicly available Federal information resources, including electronic information resources. GILS records identify public information resources within the Federal Government, describe the information available in these resources, and assist in obtaining the information. GILS is a decentralized collection of agency-based information locators using network technology and international standards to direct users to relevant information resources within the Federal Government." Retrieved on July 25, 2003 from the GPO Access Web site: "What is GILS": http://www.access.gpo.gov/su_docs/gils/whatgils.html)

[115] Information about the Illinois-state project is available at IMLS: http://www.imls.gov/pubs/wbws01cp7.htm, accessed on August 7, 2003.

> The question now becomes how to search this metadata embedded in HTML Web pages. Generally it is not Z39.50 accessible, and is unlikely to become Z39.50 accessible in the near future. Individual states are trying various approaches, but one that's attractive, in part because of the potential for multi-state interoperability, would be to put such state GILS metadata in statewide OAI-PMH metadata provider repositories. This is likely easier to do than it would be for each state to put the metadata in a Z39.50 accessible site. At UIUC we've been experimenting with making Illinois GILS metadata OAI-PMH accessible as part of a larger effort to harvest and archive Illinois state agency Web sites over time. If this were to be done in a more stable production way and by multiple states, then one could easily imagine need for a search and discovery portal that worked well and simultaneously with both Z39.50 and OAI-PMH.[116]

Given the current fiscal crisis facing many state governments, the implementation and coordination of this effort is tenuous at best. As a result, it is worth noting as an experiment, but excluded from closer consideration at this juncture.

---

116 E-mail correspondence with Timothy Cole, UIUC, on July 28 and July 29, 2003. For more information see: Cole's Powerpoint Presentation at the 5th Annual State GILS Conference posted at his personal Web site: "OAI: What it is and what it could mean for GILS projects" http://dli.grainger.uiuc.edu/Publications/TWCole/GILS2003/ . Retrieved on July 25, 2003.

APPENDIX 2 Table 1

RESOURCE	ORGANIZATIONAL MODEL	SUBJECT	FUNCTION	AUDIENCE SERVICE LEVEL	STATUS	SIZE	SOURCE of SIZE
AARLIN: the Australian Academic and Research Library Network	National membership organization of Australian academic & research libraries	Cross-disciplinary	One-stop resource-discovery tool; single search system across multiple proprietary and local databases	Academic community	Under development in Phase 2 of pilot open only to members	22 Australian libraries	Web site
Advanced Library Collection Management Environment (OCLC)	OCLC	Cross-disciplinary	Suite of open source tools to build a distributed library collection management system	Digital library developers	Experimental basis w/ preliminary applications of Electronic Theses/Disserations OAI Union Catalog and XTCat	38,000 in ETD OAI Union Catalog and 4.3 million in XTCat	Web sites
American Memory: Historical Collections for the National Digital Library (Library of Congress)	Library of Congress in public-private partnership	Cultural heritage: Americana	Gateway to rich primary source materials relating to the history and culture of the United States.	Interested public and educators	Established	100 collections and over 7 million digital items of which some 136,000 OAI-compliant images & texts	Web site and Arms (2003)
AmericanSouth.org	Emory w/ regional association and foundation funding	Cultural heritage: American South	Scholar-designed portal with collaboratively created digital collection	Interested public	Under development for fall 2003 release	18 archives and 28,775 records	Web site
Arc	Old Dominion University w/out base funding	Cross-disciplinary	Cross archive digital search service harvests OAI compliant repositories	Research community	Experimental research service	163 distinct archive groups and 6,449,515 records of which 4,372,940 from xCat	email from X. Liu on 7/23/03 plus Web site
ARCHON	Old Dominion University, Los Alamos Nat'l Lab, American Physical Society, CERN, Am. Inst. Of Physics	Science: physics	A digital library that federates Physics collections with varying degrees of metadata richness	Research community	Under development	5 distinct archive groups and 327,363 records of which 229,076 from arXiv	Web site
ARL Scholars Portal	National institutional membership organization; pilot project at 7 ARL libraries	Cross-disciplinary	Portal with single search interface to multiple databases	Academic community	Under development	7 ARL libraries in initial stage	ARL Web site
arXiv.org	Originally LANL, now Cornell w/ NSF support	Science: physics, math, non-linear science, computer science	Automated e-print archive server; distribution system without peer review	Research community	Established	estimated 230,000 records; usage stats also at Web site	Web site plus ARCHON
BEN: A Digital Library of the Biological Sciences for Biology Teaching	Collaborative sponsored by the American Association for the Advancement of Science and other professional organizations	Science: biological sciences	Portal to digital libraries for teaching & learning in the biological sciences	Educators	Established	over 1,000 reviewed resources covering 51 biological science topics	Web site
Citebase	University of Southhampton	Science: physical, mathematical, computer science, psychology, neuroscience, and biomedical	Online services, resources and tools to support self-archiving movement.	Research community	Experimental research service, "not ready for evaluation"	estimated 202,300 research papers from arXiv, Cogprints and BioMed Central	Web site

Table 1, continued

RESOURCE	ORGANIZATIONAL MODEL	SUBJECT	FUNCTION	AUDIENCE SERVICE LEVEL	STATUS	SIZE	SOURCE of SIZE
CiteSeer (aka ResearchIndex)	NEC Research Institute, Inc.	Science: computer science	Computer science Web crawler w/ reference linking, citation analysis, recommender system	Research community	Experimental	500,000 papers; 100,000 visits daily	Pacchioli [2003]
Cornucopia	Resource: The Council for Museums, Archives, & Libraries	Cultural heritage	Comprehensive database about UK museum collections.	Interested public	Under development (in beta test)	1800 UK museums w/ basic records & 500+ w/ full records from two regions	Web site
Cyclades	European Commission, IST Programme	Cross-disciplinary	Digital Library Environment: Open, collaborative virtual archive service environment	Research community	Under development	Not available	
DLESE: Digital Library for Earch Systems Education	Community-based organization w/ NSDL funding	Science: geosciences	Information system to facilitate learning about the Earth system at all educational levels	Academic community; Educators--all levels	Under development	View "All Resources" for bar graphs w/ records per collection	Web site
Electronic Theses/Dissertations OAI Union Catalog	OCLC in collaboration w/ NDLTD	Cross-disciplinary: ETDs	Union catalog built for OAI harvesting of electronic theses and dissertations.	Academic community	Experimental	38,940 records from 20 institutions including 8,264 from XTCat (Worldcat extract)	Web site
ENC Online: Eisenhower National Clearinghouse for Mathematics and Science Education	Ohio State University under contract with the U.S. Department of Education	Science: math & science	Online math and science K-12 resource center	Educators: K-12	Established w/ annual federal funding	26,000 catalog records of which 2,500 describe electronic resources	On-site interview with Lightle and Simutis on 7/30/03
Flashpoint (Los Alamos National Laboratory)	LANL	Science	Multi-database search tool for in-house use only	Research community	Established	5,814 journals online, 3.4 million full-text articles, 68 million citation records	Phone conversation w/ Irma Holtkamp on 7/28/03
GILS (Global Information Locator System)	Public/private partnership	Cross-disciplinary	Access to government information	Interested public	Established	Not available	
Grainger Engineering Library at University of Illinois-Urbana-Champaign	University of Illinois w/ Mellon Foundation funding	Science: engineering, computer science, physics	OAI metadata harvesting aggregator in sciences	Research community	Established	12 repositories w/ 443,017 records	Web site
Heritage Colorado (Colorado Digitization Program)	Funded by Colorado Dept. of Education w/ Virtual Library of Colorado partner & ILMS support	Cultural heritage: all aspects of Colorado history, culture, government and industry	Collaborative effort of Colorado's archives, historical societies, libraries, and museums to make digital collections available to people of Colorado	Interested public: Coloradians	Established	51 participating partners; 18,813 OAI-compliant records from 17 instns.	Web site and email from Bishoff 7/28/03
INFOMINE, Scholarly Internet Resource Collections	UC-Riverside and national network of libraries w/ IMLS funding	Cross-disciplinary	Referratory of expert and machine-gathered scholarly Internet resources	Academic community	Established	105,126 academically reliable resources	Web site
MacquarieNet	Fee-based subscription service w/ Australian perspective	Cross-disciplinary	Web database: Online "encyclopedia", school reference tool w/ teacher support services	Educators & students: K-12	Established	Not available	

Table 1, continued

RESOURCE	ORGANIZATIONAL MODEL	SUBJECT	FUNCTION	AUDIENCE SERVICE LEVEL	STATUS	SIZE	SOURCE of SIZE
MERLOT	Community-based with free open individual or partner membership	Cross-disciplinary	Multi-media Education Resource for Learning & Online Teaching; peer reviewed Internet resources	Academic community	Established	9,500 learning materials	Web site
NASA Technical Report Server (NTRS)	NASA	Science: aerospace (writ large)	Technical Reports Servers to collect, archive and disseminate scientific papers	Research community; Interested public	Under development	553,921 records of which est. 284,000 full-text and less than 15,000 from NASA agencies; 6,000-7,000 searches per mth	Web site and email from Nelson on 8/04/03
NDLTD Union Catalog	NDLTD (incorporated as a non-profit organization) with VTLS	Cross-disciplinary: ETDs	Union catalog of ETDs	Academic community	Under development	Same as Electronic Theses & Disserations OAI Union Catalog above	Interview w/ Young & Hickey on 7/31/03
Networked Digital Library of Theses & Dissertations (NDLTD)	Federation of member institutions & organizations that publish ETDs	Cross-disciplinary: ETDs	Aims to improve graduate education by developing accessible digital libraries of theses and dissertations	Academic community	Established	190 NDLTD members: 160 member universities, including 6 consortia, 24 institutions	Web site
NSDL: National Science Digital Library	NSF	Science: science, technology, engineering and mathematics	A digital library of exemplary resource collections and services, organized in support of science education	Educators; Digital library developers; Interested public; Funding/policy partners	Under development w/ new release slated for October 2003	199 collection records and 301,702 item records of which 204,888 from arXiv	Email from Terrizzi on 7/30/03; phone interview with Saylor on 9/5/03
OAIster	University of Michigan in partnership w/ U of Illinois and Mellon initial funding	Cross-disciplinary	Collection of freely available, difficult-to-access, academically-oriented digital resources that are easily searchable	Academic community	Established w/ uncertain funding	195 "institutions" and 1,538,431 item records	Web site
OLAC: Open Language Archives Community	International partnership of institutions and individuals	Language resources	Network of language archives conforming with the Open Archives Initiative; Virtual library	Academic community	Established	25 archives comprising 19,879 records	Web site
Online Books Page	University of Pennsylvania	Cross-disciplinary	Web site that facilitates access to online books free on the Internet	Interested public	Established	20,000+ books indexed with 400+ OAI-compliant full-text	Web site
Open Archives Initiative	DLF, CNI, NSF	Cross-disciplinary	Develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content	Digital library developers	Established	Unknown but est. 108 registered data providers and 11 registered service providers (includes aggregators, e.g. OAIster)	Web site
Open Archives Initiative Metadata Harvesting Project	U of Illinois w/ U of Michigan and Mellon initial funding	Science (see Grainger Engineering Library) & Cultural Heritage (see UIUC Digital Gateway)	Create and implement a suite of OAI-based metadata harvesting services, search services, and tools to facilitate discovery & retrieval of scholarly works	Academic community	Under development	See Grainger & UIUC Digital Gateway to Cultural Materials; no centralized repository	Email from Timothy Cole on 7/28/03

Table 1, continued

RESOURCE	ORGANIZATIONAL MODEL	SUBJECT	FUNCTION	AUDIENCE SERVICE LEVEL	STATUS	SIZE	SOURCE of SIZE
Perseus Digital Library	Tufts University, Classics Dept. w/ NEH, NSF, & other public-private funders	Humanities	Evolving digital library of resources for the study of the humanities	Interested public	Established	Not available	
PubMed Central	U.S. National Library of Medicine	Science: life sciences	Volunary publisher-based archiving of life sciences journal literature	Research community	Established	100,000 full-text articles from 130 journals	Complete journal list at Web site
Scirus (Elsevier)	Elsevier	Science	Science-specific search engine	Research community	Established	Crawls over 135 million science-related pages, consisting of 120 million Web pages, as well as 17 million records from both proprietary & OAI-compliant sources	Web site
Sheet Music Consortium	UCLA, Indiana University, Johns Hopkins, Duke, LC	Humanities: Music	OAI aggregator of sheet music	Interested public	Under development	100,000 records	Web site
SMETE: Science, Math, Engineering and Technology Education Library	Open Federation, voluntary membership w/ partners and affiliates funded by NSF & other public/private agencies	Science: science, technology, engineering and mathematics	Collection of collections and community of communities	Educators: all levels	Established	Not available	
Subject Portals	Resource Discovery Network a cooperative of UK institutions	Cross-disciplinary: five subject hubs: bio-medical sciences; engineering, maths and computer sciences; humanities; physical sciences and social sciences	Portal: network service that brings together content from diverse distributed resources using technologies such as cross searching, harvesting, and alerting, and collates this into an amalgamated form for presentation to the user.	Academic community	Under development	Varies	
U.S. States Implementing GILS	Self-selected state governments implementing GILS	Cross-disciplinary	Web service using GILS standard to define search and locate state government information	Interested public	Established	Not available	
UIUC Digital Gateway to Cultural Heritage Materials	U of Illinois w/ U of Michigan and Mellon initial funding	Cultural heritage	OAI aggregator of cultural heritage materials	Academic community	Under development	413,563 records from 25 OAI-compliant metadata providers	email correspondene with Shreeves on 7/28/03
Voice of the Shuttle: Web site for Humanities Research	UC-Santa Barbara, English Dept.	Humanities	Database that serves content dynamically on the Web for humanities research	Academic community	Established	Not available	
XTCat NDLTD/NDLTD Union Catalog	OCLC w/ NDLTD	Cross-disciplinary: ETDs	OAI-compliant database of 4.3 million bibliographic records of theses & dissertations extracted from WorldCat	Academic community	Experimental	4.3 million records of which circa 8,000 are full-text	OCLC Web site