



## Foreword

The fifth year of Harvard’s Library Digital Initiative (LDI) was marked by the continued expansion of the University’s digital library infrastructure and by a significant increase in the use of that infrastructure by our users—Harvard’s students, faculty, researchers, and staff. The year also proved notable for a long-planned review of LDI’s efforts to date by a panel of experts both inside the University and beyond its walls. The review charts the challenges of the future for LDI, and it provides highly supportive and detailed analysis of the Initiative’s first five years. I quote: “LDI has created an extraordinary resource and infrastructure. LDI technology is sound, robust, and scalable. The approach and implementation represents the current best thinking. Harvard is viewed by the national and international community as being a leader in this area of work and is envied by its peer institutions. It serves as a model for others.”

### Contents

**3 Introduction**

**4 Highlights**

**12 Collections  
and Resources**

**21 Systems  
and Services**

**27 Public Forums**

The Library Digital Initiative was established in 1998 to address three key aims:

- to make it easier for Harvard’s libraries to maintain their collections and services in the digital era, without obligating each library to acquire the expertise and systems needed to support digital resources on an individual basis;
- to create a coherent environment as digital collections grow. One objective was to avoid having individual schools create individual digital environments—environments unlikely to interoperate or to provide an organized or consistent view of resources available to the Harvard community; and
- to integrate digital resources with Harvard’s existing physical library collections: LDI is intended to provide users with integrated access to Harvard’s rich resources, regardless of format.

In short, LDI was established to recognize the growing necessity of comprehensive digital collections and to provide a robust, integrated, and University-wide framework that would address the questions of acquisition, access, authenticity, and preservation that digital collections present.

## F o r e w o r d

When the President and Provost convened their LDI Review Board (*see page 28*), it was part of a planned analysis to determine LDI's achievements toward stated goals and to recommend options for the future. The board, comprising a nationwide group of academic and IT-industry leaders, noted that, through its careful investment in LDI, Harvard's program had become a "leader among library digital initiatives in only a few years, thanks to careful planning and policy development."

The report of the LDI advisory board is an extremely gratifying response to the University's initial investments in LDI—and to all of the accomplishments of Harvard's library community that LDI represents. The report is equally challenging in its description of the complex and vital work that lies ahead of us in two major areas:

- sustained, cross-faculty development of comprehensive, multifaceted digital resources and collections; and
- integration of LDI with related efforts in research and instruction—most notably with the iCommons program in the Office of the Provost.

Like LDI, iCommons is a collaborative project involving numerous Harvard faculties. Coordinated by the Office of the Provost, iCommons has brought together academic software developed around the University and created an integrated platform for those software "modules" that can provide the Harvard faculty with support for teaching. The common ground between LDI and iCommons is clear: to meet the needs of Harvard's students and faculty. In the coming years, LDI and iCommons will grow together, pursuing distinct tactical goals but joining forces to serve this one great academic community.

Obviously, the years ahead will be filled with challenge and reward—and with new and growing benefits to our students. It will be my privilege to keep you apprised of this vital work.



Sidney Verba  
Carl H. Pforzheimer University Professor and  
Director of the University Library  
July 2003

## Introduction

This report for the fifth year of the Library Digital Initiative outlines the program's major directions, technical challenges, and most recent accomplishments; enumerates digital collections and resources developed through LDI along with grant projects that have tested and proven the LDI infrastructure; and summarizes the systems and services created through the Library Digital Initiative that are in use today.

Reflecting LDI's shift in focus from the development of systems to the needs of users, here are several examples of work that Harvard users can do today:

- Through the online finding aid in OASIS, ethnomusicologists can peruse the Laura Boulton Collection of Byzantine and Orthodox Musics at HCL's Eda Kuhn Loeb Music Library and listen to recordings made in the churches, monasteries, seminaries, and convents of the Greek and Eastern Orthodox world, looking at accompanying logbooks, as well as hand- and typewritten notes.
- Historians and educators interested in Harvard and Radcliffe archival materials are able to view online the annual reports of these institutions dating back to 1826; narrative histories; facts and figures; founding documents; Harvard music, newspapers, and magazines; and a list of buildings by street address.
- A researcher looking for material on women's salaries can search in the EconLit database, find a citation of interest, and be connected to the licensed, full text in the journal *Industrial Relations*—with the click of a Harvard SFX button.
- Users looking for pictorial representations of India can search online in VIA, Harvard's union catalog for visual materials, and browse through an array of colorful, diverse, and compelling images from the Harvard University Art Museums, the Fine Arts Library, the Harvard-Yenching Library, the Baker Library, the Arnold Arboretum, and more.
- Visitors to the Arnold Arboretum site, "South Central China and Tibet: Hotspot of Diversity," are able to virtually explore modern and historical botanical expeditions to South China and Tibet, including those of plant explorer Joseph Rock in the 1920s. Researchers and students can view digital images of botanical specimens collected by Rock, his hand-drawn map, correspondence, a gazetteer of place names from the expeditions, and links to living specimens at the Arboretum. Users can easily link from the botanical specimen to the Harvard Geospatial Library to create a customized map showing the location where the specimen was found. The Arboretum site is an excellent example of the true potential of LDI in that it brings together materials from multiple Harvard faculties and uses almost every part of the LDI infrastructure developed during these past five years.

## Highlights

FY 2003 was notable for significant new services to the Harvard community, including:

- an updated Harvard Libraries web site, which serves as the online gateway to the University's extraordinary library resources;
- the SFX system, which, with a single click, can move the user from a citation to the full text of an article;
- cross-catalog search, which enables Harvard users to conduct simultaneous searches in HOLLIS and other online library catalogs;
- the Templated Database service, which can improve access to small and/or specialized collections that may not fit easily into existing online catalogs;
- online availability of audio files held in Harvard's Digital Repository Service; and
- a new facility in VIA, the visual collections catalog, through which instructors can export images with descriptive metadata into a digital slide carousel tool jointly offered by the Faculty of Arts and Sciences and iCommons.

## Major Directions

### Integration with Educational Technology at Harvard

Libraries play an important role in providing resources for instruction—traditionally in such areas as developing undergraduate library collections, managing course reserves, and providing collections of teaching slides. The parallel growth of digital library collections and the development of course management systems, as a way of providing access to instructional resources, offer new opportunities for libraries to play a role in instructional support.

At Harvard, the iCommons project (<http://icommons.harvard.edu>), coordinated by the Office of the President and Provost, is bringing together academic software “modules” developed across the University and creating an integrated platform to support teaching. LDI will work with iCommons in such areas as:

- making reserve materials increasingly available through course web sites;
- increasing the array of relevant digital library resources presented to students in the context of their courses;
- making the Harvard libraries' instructional and reference services accessible from course web pages; and
- using Harvard's digital library infrastructure to preserve materials created specifically for courses and to ensure access to those materials over time.

For LDI and iCommons, developing ways of interoperating will be key activities over the next few years. In FY 2003, for example, LDI and iCommons collabo-

rated on a new facility in VIA. Instructors and librarians can now export images (along with descriptive metadata) into a digital slide carousel tool. At a more general level, discussions are now under way, both in the larger educational environment and specifically at Harvard, about the interrelation of digital library systems with a full range of instructional and research tools.

### Integration with the Larger Digital Library Environment

The Harvard community now has access to a wealth of resources provided through a wide range of systems and services across the Internet. The downside of this exciting and enormously useful array is complexity: there are many systems and interfaces employed to find and use digital resources. One of the challenges for LDI is to lead users to the appropriate resources and to simplify the navigation of this complex environment.

The ability to integrate diverse resources in ways that simplify use is an increasingly important development in the larger information technology environment. Many systems and services now integrate tools and data into people's working environments in a way that does not depend upon users knowing *where* those tools or data originate. Digital libraries, with their enormous range of diverse and distributed resources, will benefit greatly from developments of this sort. The integration of distributed resources and the removal of barriers to their use will be key themes in digital library developments over the next decade.

The FY 2003 implementation of SFX is one step in creating a more efficient, better integrated environment for Harvard library users. With SFX, for example, a user can navigate easily from a book or article citation in one system to the text of the cited book or article in another. This navigation is tailored to the Harvard information environment, so that users are led only to materials to which they have free access.

Another type of interoperation is the integration of Harvard's local resources with the larger digital library environment. Examples of this include:

- making information about Harvard-produced resources available through outside databases;
- making Harvard catalogs accessible to other portal systems; and
- ensuring that digital resources created at Harvard follow standards and that they are usable in systems beyond the University.

In FY 2004, LDI will implement a new resource portal that will allow for simultaneous searches across a variety of systems, databases, catalogs, and other resources, rather than requiring users to go to and search each system separately.

## Technical Challenges

### Digital Preservation

As increasing amounts of digital content are produced at Harvard and stored in the LDI Digital Repository Service (DRS), the importance of ongoing preservation activities cannot be overstated. Digital materials are inherently fragile and dependent for their long-range viability on technologies that change continuously. To protect Harvard's digital resources for the future, LDI is developing expertise in the underlying digital formats of objects accepted into DRS and in the extensive technical metadata about these objects. By closely monitoring the technological environment and the digital formats stored in DRS, LDI can initiate digital preservation activities that ensure the future of digital resources.

For LDI and for the University as a whole, digital preservation is a priority that is reflected in several areas of progress in FY 2003. Many of these activities are collaborative, demonstrating that digital preservation issues are most effectively addressed through national and international efforts of large-scale cooperation:

- As a follow-up to the FY 2002 Mellon Foundation-funded e-journal archiving project, LDI has collaborated with the National Library of Medicine (NLM) to produce an open-source archiving and interchange XML Document Type Definition (DTD). The DTD is designed to increase the ease of interchange between publishers and archives for article-level content. Without this DTD, the structure of e-journal content can vary widely, requiring costly human intervention and multiple parallel workflows within archival repositories. The DTD was designed after extensive document analysis in many subject domains to ensure that it does not reflect the bias of any particular academic discipline. Based on open standards, the DTD features a modular structure that allows customizing and that should be an easy target of transformation from existing XML- or SGML-encoded content. In addition to being used by NLM for the PubMed Central archive, this DTD is well positioned to become a standard format for the transfer and archival storage of scholarly literature.  
<http://dtd.nlm.nih.gov>
- Adobe's Portable Document Format (PDF) has rapidly become a de facto standard for the dissemination and presentation of electronic documents on the web. Unfortunately, the feature-rich nature of PDF permits tremendous variability in the internal structure of documents. Further, it allows documents to be dynamically composed at the time of their display from disparate external resources, which leads to significant difficulties in ensuring their long-term viability. In order to address these concerns, a multinational effort

has been established within the International Organization for Standardization (ISO) standards framework to produce a constrained version of PDF suitable for archival preservation, to be known as PDF/A. Stephen Abrams, LDI's digital library program manager, is the project leader and editor of the ISO Joint Working Group developing PDF/A.

<http://www.aiim.org/standards.asp?ID=25013>

- LDI is collaborating with JSTOR to produce a tool, called JHOVE, for automating format-specific validation of digital objects. The JSTOR/Harvard Object Validation Environment, or JHOVE (pronounced “jove”), is an extensible software framework for performing format identification, validation, and characterization of digital objects. This tool, which will be made publicly available under an open-source license, can be particularly useful for the validation of digital objects submitted for deposit into a digital repository such as DRS. In addition, JHOVE has facilities to extract important technical characteristics of digital objects from the objects themselves. To ensure future use of digital objects, it is important to verify that a format and its characteristics have been correctly identified. The initial deployment of JHOVE will provide validation for the PDF and TIFF formats, including recognition of many specific format profiles or named constrained subsets.

<http://hul.harvard.edu/jhove>

- Preservation activities depend upon extensive knowledge of the formats in which digital objects are manifested. Since this same information is useful to all institutions interested in preserving their digital assets, great economies of scale can be achieved from a central repository for this format information. LDI has been instrumental in organizing an ad hoc international group of interested stakeholders, including representatives of national libraries and archives, and academic research libraries, who have met to discuss the technical and policy issues surrounding the creation and sustainable operation of a global digital format registry. Stephen Abrams, LDI's digital library program manager, co-authored a paper on this topic that was presented at the 2003 IFLA conference. The paper is available on the IFLA site.

[http://www.ifla.org/IV/ifla69/papers/128e-Abrams\\_Seaman.pdf](http://www.ifla.org/IV/ifla69/papers/128e-Abrams_Seaman.pdf)

- To avoid duplication of digital reformatting efforts, LDI is actively participating with the Digital Library Federation (DLF) in plans for a national digital registry of born-digital materials and digitally reformatted books and journals. By consulting the registry before digitizing, an object owner can determine if an appropriate digital version already exists.

<http://www.diglib.org/collections/reg/reg.htm>

### Extended Character Set Support

The resources supported by LDI's infrastructure encompass many languages and script systems. While the use of Unicode (a character-encoding standard) provides the basis for uniform storage of electronic text, challenges still remain to ensure the consistent creation and display of non-Latin text resources.

The base technologies underlying most LDI systems—the Oracle RDBMS and the Tamino XML DB—have been configured to use Unicode as their internal character representation. This allows systems such as HOLLIS, the Full-Text Search Service (FTS), the Page Delivery Service (PDS), and the Templated Database System (TED) to properly store text using most of the world's living and historical languages. Re-implementations of VIA and OASIS in FY 2004 will provide those systems with the same capability. All LDI systems are also converging on the use of a common set of text normalizations—devoid of punctuation, case distinction, and diacritical marks as defined by the library community's long-standing Name Authority Cooperative Program (NACO)—prior to indexing and search operations, so that patrons can expect similar search behavior throughout different systems. Despite these advances in handling character sets, substantial challenges remain in the areas of uniform text display and input.

While text storage is controlled through the LDI infrastructure, text display is dictated by desktop browsers and commercial client-side applications—such as those used for the OLIVIA and HOLLIS cataloging systems—which are outside of LDI control. LDI is working to provide appropriate user support for text display in these systems through the configuration of local workstations.

An additional challenge remains in the creation of non-Latin text resources. No effective general mechanism currently exists to generate non-Latin characters using a standard desktop keyboard. Potential options for the creation of non-Latin text include transliteration (used, for example, in HOLLIS for Chinese) and Input Method Editors (IMEs), which are desktop applications that provide a visual interface to language-specific “virtual” keyboards. The difficulty with these options is that they tend to be tailored to specific systems, and LDI is striving to provide a common set of behaviors and a consistent look-and-feel across its infrastructure. LDI will continue its ongoing analysis of this problem in an effort to provide a uniform and effective solution to generation of non-Latin text.

## Library and User Services

### Harvard Libraries Web Site

On June 25, 2003, the Harvard University Library Office for Information Systems (OIS) launched a revised “portal” page for the Harvard Libraries site. The portal is an online gateway to the extraordinary library resources of Harvard University that serves as an important research tool for Harvard’s students, faculty, staff, and researchers who hold Harvard IDs and PINs. The site also provides practical information on each of the more than 90 libraries that form the Harvard system.

The major goals of this revision were to improve design and usability, to increase flexibility, to simplify maintenance, and to provide a short-term solution until the planned introduction late in 2004 of a new and more powerful library research portal based on MetaLib software from Ex Libris.

The revised Harvard portal page contains two new informational links: “Finding Materials at Harvard” and “FAQ for Visitors,” and it provides a stronger visual and informational connection to the University’s home page, which is found at *www.harvard.edu*.

<http://lib.harvard.edu>

### Implementation of SFX

SFX is a significant new research tool from Ex Libris that was implemented in the Harvard libraries on January 8, 2003. The tool uses resource-linking technology based on the OpenURL standard to allow users of external research databases to link directly from an article citation or abstract to a variety of related resources determined by the local library or institution. With the click of a button, SFX can provide access to the full text of an article (if available) or to local holdings in the HOLLIS catalog. It permits context-sensitive and dynamic linking between web-based resources in which the actual links are customized to reflect licensed digital resources available to users affiliated with Harvard.

During the academic year, the use of SFX neared 2,000 hits per day.

<http://hul.harvard.edu/ois/systems/sfx>

With the implementation of SFX, two related products were launched during this past year—Citation Linker and EJ2: Supplementary List of E-Journals. With Citation Linker, users can input an article citation to locate the licensed full text. Citation Linker is available on the e-resources menu of the Harvard Libraries web site at *http://lib.harvard.edu*. There are approximately 10,000 titles included on the EJ2 list, over half of which currently have no other point of access on the Harvard Libraries web site or in the HOLLIS catalog.

[http://sfx.hul.harvard.edu:82/sfx\\_local-e-collection/e-journals-A.html](http://sfx.hul.harvard.edu:82/sfx_local-e-collection/e-journals-A.html)

### Harvard Cross-Catalog Search

On November 6, 2002, Cross-Catalog Search service was made available from the Harvard Libraries web site as a demonstration system that would gauge the public's reaction to federated searching across multiple Harvard catalogs. It was developed using a subset of an early version of the MetaLib software from Ex Libris. This service is a high-level resource discovery tool that allows the user to search simultaneously across five of Harvard's catalogs, including HOLLIS, Baker, VIA, OASIS, and HGL. From November through the end of the academic year, the number of searches totaled over 17,000, with approximately 6,000 sessions recorded.

Feedback from both staff and patrons using Cross-Catalog Search indicated a strong interest in simultaneous searches of research databases and other external resources together with Harvard's library catalogs. With this in mind, OIS began to look seriously at the new version of the MetaLib software, which offers federated searching and personalization features not now available on the Harvard Libraries web site. A recommendation to pursue the analysis and implementation of MetaLib software for a new research portal was approved by the University Library Council (chaired by the director of the University Library and comprising the librarians of each of the Harvard faculties) in March 2003, with full implementation targeted for mid-2004.

<http://crosscatalog.harvard.edu>

### Templated Database Service (TED)

TED is a powerful new system designed and developed in FY 2003 to provide an online home to the myriad of small, specialized collections catalogs that do not fit within the scope of existing Harvard catalog systems. TED can provide web-based access to data that might otherwise be hidden in boxes of cards or desktop computers across campus. TED is intended to function without requiring an extensive OIS implementation effort or the need for high-level programming skills.

Any number of unique databases can be created with TED in order to satisfy the needs of many individual projects. With assistance from the LDI metadata analyst, collection managers can create an XML schema, select field names, and define the interface for their own database. Data can be imported from an existing database or created online using the TED maintenance system. Because each database can be built on the same framework, centralized system support can be provided for software upgrades and data migrations.

The first collection cataloged through TED is the Biomedical Image Library (<http://nrs.harvard.edu/urn-3:hul.eresource:bioimlib>), which is a set of digital micrographs produced in support of basic biological research. A second TED-based catalog for the Milman Parry Collection of Oral Literature is scheduled to be available online late in 2003.

<http://hul.harvard.edu/ois/systems/ted>

### Digital Audio

Digital audio represents the most complex type of digital resource that LDI now supports. This year, working in conjunction with David Ackerman, the audio preservation engineer in HCL's Loeb Music Library, LDI developed specifications for the deposit to DRS of audio works. Deposits consist of multiple versions of digital audio files, including high-resolution archival and production masters and lower-resolution use copies. In addition, deposits include a wealth of metadata to capture the technical properties of the audio files, their processing histories, and the structure and relationships among these various components. LDI is now developing a new desktop application called DMART that will automate the complicated packaging of these components. DMART, along with an upgrade to the DRS data model and new loading procedures to accommodate digital audio works, will be available late in 2003.

The listening versions of audio works will be delivered by RealAudio™ through the new Streaming Delivery Service (SDS) developed by LDI this year. SDS uses Access Management Service (AMS) to control access to audio materials restricted to the Harvard community, and it supports use logs to meet the legal requirements imposed by copyright holders of digitized material. The archival and production master versions of audio files can be retrieved from DRS (by authorized owners) using the WebAdmin interface and Asynchronous Delivery Service (ADS).

## **Collections and Resources: Digital Acquisitions**

The Digital Acquisitions Program (DAP) supports the shared purchase and licensing of commercially available digital resources for Harvard's libraries. Program services include the organization of prospective and ongoing product evaluation, license negotiation, access implementation and administration, and vendor relationship management. Consulting assistance is also offered to libraries that negotiate license agreements for their local collections. Increasingly, LDI DAP is involved in assisting libraries with collection decisions involving print resources, such as canceling unneeded duplicate print journal subscriptions in order to control acquisitions costs.

Harvard-wide digital acquisitions continued at a steady pace of growth in FY 2003. Approximately 1,210 new resources—including 1,150 electronic journals and 60 databases—were licensed and made available to the Harvard community through the Harvard Libraries web site. Combined expenditures for electronic resources shared among the Harvard faculties rose to \$1.6 million in FY 2003, a 6 percent increase over the previous year. Statistics on use, including those that have been collected by electronic resource vendors, are now available through Harvard IP addresses on the Digital Acquisitions web site.

<http://hul.harvard.edu/digacq>

Harvard's online collections are expanding in breadth, depth, and linguistic diversity. Significant retrospective collections added in FY 2003 include the *New York Times*, the *Wall Street Journal* Historical Edition, and the *Nation* Archive. New "born digital" works added in FY 2003 include Proteome BioKnowledge Library, the ACLS History E-Book Project electronic micrographs, and new digital compilations of pre-existing source material—such as Black Drama and North American Women's Letters and Diaries—gathered together for the first time. The libraries have also acquired important collections of non-Latin language materials, including the China Academic Journals database (3,000 scholarly and scientific journals published in China) and Russian-language newspapers, journals, and government information from the Russian republics and the Baltic region.

One of the main functions of LDI DAP is to foster coordination among the Harvard libraries in order to achieve cost savings in journal and other acquisitions. One avenue for accomplishing this is Harvard's participation in the Northeast Research Libraries Consortium (NERL), a federation of leading research libraries formed in 1996 to contain the costs of electronic purchases and to influence market developments through joint licensing agreements.

Many of Harvard's most costly and widely used resources, including the ISI Web of Science citation databases and electronic journals from numerous major publishers, are acquired through NERL. In the spring of 2003, NERL issued its "Principles for Electronic Journal Licenses" to guide negotiations with publishers in a time of fiscal uncertainty. Cost containment and the ability to build suitable collections through individual title selection as well as package offerings are among the key objectives that Harvard is actively pursuing both through NERL and independently.

Acquiring and providing access to electronic resources also poses significant new challenges in information management and workflow. Traditional library systems that are designed to support the acquisition and processing of print materials have not yet caught up with the requirements of electronic information. The DLF Electronic Resource Management Initiative is defining standards and best practices for electronic resource management to provide guidance to system vendors and libraries alike. Ivy Anderson, Harvard's digital acquisitions program librarian, serves on the steering committee of this project.

Local planning efforts for an electronic resource management (ERM) module for the Harvard libraries complemented the work of the DLF. In collaboration with MIT and Ex Libris, the company that developed Aleph, LDI staff defined functional requirements and specifications for a commercial ERM system based on existing DLF documents. This system would interact with Aleph, SFX, and MetaLib. Participation in the Harvard/MIT work was extended to include members from the North American Aleph Users Group (NAAUG) and the International Consortium of Aleph Users (ICAU).

### **Harvard Open Collections Program**

In November 2002, the Harvard Open Collections Program was launched as an 18-month pilot project with funding from the William and Flora Hewlett Foundation. The goal of the Open Collections Program (OCP) is to increase the availability and use of Harvard's rich and historically significant collections for teaching, learning, and research. Using the digital library infrastructure developed by LDI, selected resources in broad topic areas will be digitized and made available to the larger academic community through union catalogs and the web. The Open Collections pilot will focus on women and work in the United States in the late 19th and early 20th centuries. The original source material for the project will include monographs, manuscripts, and visual resources drawn from many of Harvard's libraries, museums, and other collections. The resulting digital resources will be stored in DRS, given persistent identifiers in NRS, and described in the appropriate catalogs (monographs in HOLLIS, manuscripts in OASIS, visual material in VIA). In addition, a subject-specific web site will be created to provide a contextual environment for discovery and exploration of these resources.

## Internal Challenge Grant Program

Managers and staff throughout Harvard’s libraries, archives, museums, and special collections have made enormous contributions to LDI through the Internal Challenge Grant Program. While adding valuable content for research and education, these grant projects have advanced LDI by prioritizing development work and by testing and demonstrating new systems and services. Project goals have ranged from straightforward digitizing of images or text collections to creating virtual collections that combine related modern and historic material from different repositories across the University to developing new systems for the cataloging and delivery of specialized data. Many projects have focused on providing access to previously inaccessible collections and making them available online for use by students and scholars at Harvard and around the world.

Since 1998, the LDI grant program has funded 30 projects through which nearly 200 Harvard staff members have gained experience in working with digital projects. In FY 2003, four projects were completed and 12 were newly funded. Four projects utilized LDI’s Management Assistance and Planning program (LDI MAP), a cost-recovery service that has provided customized, hands-on assistance to managers of LDI grant-funded projects.

<http://hul.harvard.edu/ldi/html/grants.html>

### Completed in FY 2003

#### Biomedical Image Library (BIL)

Countway Library of Medicine (Harvard Medical School *et al*)

*in collaboration with the Biomedical Imaging Laboratory at the Harvard School of Public Health*

A central catalog and collection of biomedical images produced in support of basic biomedical research.

<http://nrs.harvard.edu/urn-3:hul.eresource:bioimlib>

#### Maya Archaeological Photographs from the Carnegie Institute of Washington Collection

Peabody Museum of Archaeology and Ethnology (Faculty of Arts and Sciences)

Online access in VIA, including descriptive cataloging records, to approximately 10,000 Mayan archaeological photographs—selected from the Photographic Archives of Harvard’s Peabody Museum—of buildings, monuments, and artifacts, many of which are no longer in existence.

<http://nrs.harvard.edu/urn-3:hul.eresource:viaxxxxx>

*To view the holdings, enter the search term “Maya” in the first box, select “Anywhere” in the drop-down window, limit to holdings of the Peabody Museum, and restrict the search to records that have images.*

## **South Central China and Tibet: Hotspot of Diversity**

Arnold Arboretum Library of Harvard University (Faculty of Arts and Sciences)

A digital collection of specimens, correspondence, maps, and images related to modern and historic botanical expeditions to South China and Tibet, including those of explorer Joseph Rock in the 1920s.

<http://arboretum.harvard.edu/library/tibet/expeditions.html>

## **Loeb Design Library Electronic Finding Aid Project**

Frances Loeb Library (Harvard Design School)

A reconfiguration of the library's database to enable the export of EAD-formatted finding to OASIS. Select the link for Loeb Design Library at <http://oasis.harvard.edu> to view the nine EAD finding aids available online in OASIS as a result of this project.

## **Proposals Funded in FY 2003**

### **Digitization of the Slide Library**

Fine Arts Library (Harvard College Library/FAS)

To digitize 100,000 teaching slides and to make them available through VIA. The creation of cataloging records for these images is also being funded through LDI.

### **Enabling Access to Historical Images of the Harvard Medical School**

Countway Library of Medicine (Harvard Medical School *et al*)

To create a finding aid for a collection of historical images of the Harvard Medical School with links to 1,600 images.

### **Legal Portrait Collection**

Harvard Law School Library

To digitize, catalog in OLIVIA, and make available through VIA 4,000 portraits of lawyers, jurists, and legal thinkers of the Middle Ages to the late 20th century that are included in the Law Library's Special Collections.

<http://www.law.harvard.edu/library/special/collections/portraits>

### **The Pickens Collection on China's Muslims**

Harvard-Yenching Library (Harvard College Library/FAS)

To digitize a collection of images documenting Muslims in China. 1,000 photographs in three albums and 50 broadsides will be digitized, cataloged in OLIVIA, and made accessible in VIA. Images of album pages will be made available through links in a finding aid to the collection.

### **Russian Theatrical Designs in the Harvard Theatre Collection**

Houghton Library (Harvard College Library/FAS)

To digitize, catalog in OLIVIA, and make available through VIA nearly 600 images of Russian stage designs.

### **Jacques Burkhardt and the Thayer Expedition to Brazil (1865–1866)**

Ernst Mayr Library of the Museum of Comparative Zoology (FAS)

To provide online access to the Jacques Burkhardt collection of watercolors and pencil drawings (approximately 1,000 items), as well as correspondence, field notes, diaries, sketches, photographs, monographs, specimens, and specimen records from the 1865 Thayer Expedition to Brazil with Louis Agassiz.

### **Digital Scores from the Collections of the Eda Kuhn Loeb Music Library**

Eda Kuhn Loeb Music Library (Harvard College Library/FAS)

To create scanned images of 30 rare, unique, and/or fragile musical scores, including first and early editions of Bach family composers, Mozart, and multiple versions of 19th-century operas. The project demonstrates the concept of using multiple-variant print and manuscript versions of a single musical work for research into historical performance practice.

<http://hcl.harvard.edu/loebmusic/online-ir-digitalscores.html>

### **The Nature of Eastern Asia: Botanical and Cultural Images from the Arnold Arboretum Archives**

Arnold Arboretum Library (Central Administration of Harvard University)

To create online finding aids in OASIS and to digitize 4,521 images for nine collections from the photographic archives. These images of eastern Asia, depicting the area's social and cultural history, landscapes, artifacts, people, and natural and ecological resources, will be made accessible through VIA.

### **Maya Archaeological Photographs from the Carnegie Institute of Washington Collection—Phase II**

The Photographic Archives of the Peabody Museum of Archaeology and Ethnology

*with sponsorship from Tozzer Library (Harvard College Library/FAS)*

To provide access in VIA to the remaining 30,000 images in the Carnegie Institute of Washington collection of Maya Archaeological Photographs.

### **Architectural Views of the World, 1870–1920: Digitization of Lantern Slides from the Fine Arts Library Collection**

Fine Arts Library (Harvard College Library/FAS)

To digitize 15,000 of the department's 95,000 lantern slides and to convert the cataloging data for the images. The selected lantern slides document architectural views of North America (including Harvard University), Europe, and Northern Africa. The images and data will be made available online through VIA.

### **Pre-1601 English Law Collection Access Project**

Harvard Law School Library

To create HOLLIS records for approximately 2,000 English law titles printed before 1601 that are held in the Law Library's Special Collections Department.

### **Creation of Descriptive Metadata for Images Used in Teaching a Sequence of Required Architectural History Courses**

Frances Loeb Library (Harvard Design School)

To create catalog records in OLIVIA for approximately 2,400 images used in teaching a sequence of required architectural history courses at the Harvard Design School.

## **Ongoing Grant Projects**

### **Asian Art Images**

Harvard University Art Museums (HUAM) and Fine Arts Library (Harvard College Library/FAS)

A collaborative project to provide online access to 3,600 Asian art images.

### **Music from the Archive: A New Model of Access to Rare and Unique Sound Recordings**

Eda Kuhn Loeb Music Library (Harvard College Library/FAS)

To provide online access to the finding aids, images, and music from three collections in the Music Library:

- The Laura Boulton Collection of Byzantine and Orthodox Musics  
<http://oasis.harvard.edu/html/mus00001frames.html>
- The Joseph Jeffers Dodge Duke Ellington Recordings Collection  
<http://oasis.harvard.edu/html/mus00006frames.html>
- The Rubin Collection of Indian Classical Music  
<http://oasis.harvard.edu/html/mus00008frames.html>

### **Harvard College Library Finding Aids Conversion Pilot Project** Houghton Library (Harvard College Library/FAS)

To develop a model for full-scale conversion to EAD of finding aids at Harvard. Thousands of pages of finding aids will be converted and submitted to OASIS as part of this project.

<http://oasis.harvard.edu/hou.html>

### **Schlesinger Library Encoded Archival Description Evaluation and Retrospective Conversion Project**

Schlesinger Library (Radcliffe Institute for Advanced Study)

To investigate, evaluate, and select an EAD markup methodology to convert finding aids to browser-usable formats, and to contribute a significant number of finding aids to OASIS.

<http://oasis.harvard.edu/sch.html>

### **Web-Based Course Material Archiving Project: Study Phase**

Frances Loeb Library (Harvard Design School)

To study the issues confronting the documentation and archiving of digital course materials in various formats.

<http://hul.harvard.edu/ois/projects/gsdarchiving>

### **Incunabula and Solomon M. Hyams Collections Access Project**

Countway Library of Medicine (Harvard Medical School *et al*)

To catalog some 800 incunabula titles and approximately 4,000 other items from the Hyams Collection of Judaica, including pamphlets, incunabula, manuscripts, and monographs.

### **A Project to Digitize, Process, and Save Widener's Latin American Pamphlets**

Widener Library (Harvard College Library/FAS)

To catalog, microfilm, digitize from the film, and preserve 3,000 Latin American pamphlets in Widener beginning with 19th-century pamphlets from Argentina, Bolivia, Cuba, Mexico, and Peru. A representative pamphlet is available at <http://nrs.harvard.edu/urn-3:FHCL:66737>.

## **Retrospective Conversion of the Slide Library Card Catalog**

Fine Arts Library (Harvard College Library/FAS)

To convert catalog cards for nearly 340,000 core teaching slides, to import the data into OLIVIA, and to make the records available through VIA.

## **Baker Library Trade Catalogs**

Baker Library (Harvard Business School)

To create catalog records for 4,500 trade catalogs from the Historical Collections Department.

## **The Singer Continues the Song: Text and Music from the Milman Parry Collection**

The Milman Parry Collection of Oral Literature

*with sponsorship from Harvard's Center for Hellenic Studies Library (FAS)*

To provide networked access to a selection of sound recordings and text images contained in the collection, the largest single repository of South Slavic heroic song in the world.

<http://nrs.harvard.edu/urn-3:hul.eresource:milparco>

## **Projects Launched and in Use**

LDI grant projects are contributing several hundred thousand cataloging records and more than a half million digital objects to Harvard's library system. The following completed projects now have online resources available for use by students and scholars at Harvard and around the world.

### **The Harvard–Radcliffe Online Historical Reference Shelf (HROHRS)**

Harvard University Archives (Harvard University Library) and Radcliffe Archives (Radcliffe Institute for Advanced Study)

Electronic access to frequently consulted sources on the history of Harvard and Radcliffe, including annual reports, narrative histories, and founding documents.

<http://nrs.harvard.edu/urn-3:hul.eresource:hronhif>

### **19th-Century American Trade Cards**

Baker Library (Harvard Business School)

Descriptions and digital images in VIA of 1,000 advertising trade cards selected from the Historical Collections at the Baker Library. As an indicator of consumer habits, social values, and marketing techniques, trade cards are of interest to scholars of American social, cultural, and business history.

<http://www.library.hbs.edu/hc/exhibits/tcard>

### **The Hedda Morrison Photographs of China**

Harvard-Yenching Library (Harvard College Library/FAS)

Descriptions and digital images in VIA of 4,800 photographs made by German photographer Hedda Morrison.

Taken between 1933 and 1946, the photographs document the architecture, streetscapes, clothing, religious practices, and crafts that in many cases have all but disappeared from modern China.

<http://hcl.harvard.edu/harvard-yenching/morrison>

### **Harvard Geospatial Library**

Social Science Program (Harvard College Library/FAS)

A catalog and repository for geospatial data held by Harvard University.

<http://nrs.harvard.edu/urn-3:hul.eresource:hgeodesy>

## **Systems and Services: Digitizing Laboratories**

### **Fine Arts Library Digital Imaging Lab (FAL DIL)**

The Digital Imaging Lab (DIL) is part of the Slides and Digital Images Department in HCL's Fine Arts Library. The lab provides digital images of slides for use with the Instructional Computing Group's (ICG) digital carousel tool and study images of slides for VIA. The lab serves the faculty and students of the Department of the History of Art and Architecture, as well as faculty from throughout FAS, including the Harvard Extension School and the Institute for Learning in Retirement.

In FY 2003, FAL DIL deposited 20,697 images into Harvard's DRS, including images scanned in house as well as some image files from vendors. In addition, the DIL does scanning for FAL's LDI grant projects and additional programs, including a forthcoming exhibition and publication on the history of the Fogg Art Museum.

<http://hcl.harvard.edu/finearts/sdi.html>

### **Harvard College Library Digital Imaging Group (HCL DIG)**

HCL DIG, a division of HCL's Preservation & Imaging Department, produces high-quality digital reproductions of library and archival materials, and offers image processing, metadata creation, and DRS deposit services on behalf of HCL and other repositories throughout the University.

In FY 2003, DIG created and deposited 38,825 digital objects into DRS, including 17,950 master archival images with their associated derivatives and 1,228 XML-formatted structural metadata files. During the year, HCL DIG provided reformatting services for seven LDI-funded grant projects, as well as scanning and processing of 18,389 page images for HCL's Reserves Program.

<http://preserve.harvard.edu/dig>

### **Harvard University Art Museums (HUAM) Digital Imaging and Visual Resources (DIVR)**

HUAM DIVR creates high-quality digital images of art objects and ephemera in the collections of the Harvard University Art Museums through direct digital capture and conversion of film surrogates. HUAM DIVR handles internal requests from curators, registrars, and staff in exhibitions, publications, and public relations, as well as external requests for scholarly, nonprofit, and commercial use in research and publications.

During FY 2003, HUAM DIVR created over 70,000 images and deposited 94,719 digital objects to DRS. The deposits comprise two terabytes of space and correspond roughly to 21,864 unique images with their associated derivatives.

### Peabody Museum of Archaeology and Ethnology

In FY 2003, the Peabody Museum of Archaeology and Ethnology developed the capacity to make batch deposits to DRS. As part of their LDI grant project, the museum contracted for the creation of digital images for photographs from 35mm copy-positive reel film and deposited to DRS three digital versions of each image—an archival master, a reference image, and a thumbnail. During the year, the Museum deposited a total of 31,254 image files to DRS, representing approximately 10,000 photographs. The images and their associated catalog records are available to the public through VIA, and additional information is made available for staff use through the museum's collection management database.

### Harvard College Library Audio Preservation Services (HCL APS)

Audio Preservation Services, part of the Eda Kuhn Loeb Music Library of the Harvard College Library, was established for the purpose of preserving, reformatting, and reproducing library and archival audio materials. In addition to preservation services, APS offers a full range of audio remastering services, including audio processing, de-noising, metadata creation, and DRS deposit services.

During FY 2003, digital preservation masters and delivery versions were created from analog source recordings, including 100 open-reel tapes and twelve 78-rpm records for an LDI grant-funded project, Music from the Archive: A New Model of Access to Rare and Unique Sound Recordings. The studio will begin depositing audio files and associated metadata to DRS in the winter of 2004. <http://hcl.harvard.edu/loebmusic/about-audiopreservation.html>

## Catalogs and Discovery

LDI systems provide online access to a rich set of library resources that includes digital content through a number of general and specialized catalogs and discovery tools. As the focus of LDI shifts to user needs, these systems and services are continually enhanced.

### HOLLIS (Harvard Online Library Information System)

Harvard's HOLLIS Catalog is a database containing over 9 million records for books, journals, electronic resources, manuscripts, government documents, maps, microforms, music scores, sound recordings, visual materials, and data files owned by the University and its libraries. The union catalog is updated continually as material is ordered, received, and cataloged.

In FY 2003, LDI staff:

- completed the loading and indexing of 515,300 CJK (Chinese, Japanese, and Korean) records, which enables HOLLIS users to search these records in their native scripts. For Chinese materials, records that were formerly in Wade-Giles Romanization were also converted to pinyin.
- implemented Z39.50 access to the HOLLIS Catalog, enabling authorized Harvard users with valid IDs and PINs to conduct HOLLIS searches using a Z39.50 client such as EndNote in addition to a web browser.
- improved the integrated library system, or ILS, staff functions by adding a desktop reporting module that allows users to report on data extracted from HOLLIS, including acquisitions and financial data, circulation history, reserve courses and bibliographic data, and selected fields from the MARC bibliographic and holdings data.
- beta-tested the newest release of the Aleph software, version 16, with full implementation in HOLLIS expected by January 2004.

<http://holliscatalog.harvard.edu>

### Harvard Libraries Web Site

The "Harvard Libraries" site is a comprehensive web interface that presents a single, organized view of web-accessible resources available to the Harvard community. The site also serves as an electronic gateway to Harvard's union catalogs and to comprehensive information about Harvard's libraries.

Substantial enhancements were made to the top page of the web site in late spring of 2003.

<http://lib.harvard.edu>

### E-reserves

The E-reserves program provides students with online access to course reserve reading materials. Through the new HOLLIS Catalog, users have integrated access both to E-reserves and to information about print reserves.

### VIA (Visual Information Access)

VIA is Harvard's web-based union catalog of visual resources in art, architecture, and material culture. VIA records include descriptive information about slides, photographs, drawings, paintings, objects, and other artifacts held by the university's libraries, museums, and archives. In FY 2003, detailed functional specifications and technical analysis for a new system architecture were completed. The new system with improved functionality will be implemented in FY 2004.

<http://nrs.harvard.edu/urn-3:hul.eresource:viaxxxxx>

### OLIVIA

OLIVIA is a visual resources cataloging system for the creation of descriptive metadata that will be exported to VIA for public access. In FY 2003, more than 40 catalogers worked in OLIVIA. Through the year, LDI effected a number of small system enhancements to increase cataloging efficiency, including a merging and de-duplicating function for duplicate catalog records and the capability to link OLIVIA records to restricted images stored in the DRS.

<http://hul.harvard.edu/ois/systems/olivia>

### OASIS (Online Archival Search Information System)

OASIS is Harvard's online catalog of electronic finding aids, which provide detailed information about the University's archival and manuscript collections. OASIS contributors are increasingly providing links within electronic finding aids to digital content such as correspondence, audio recordings, photographs, and other images. During FY 2003, LDI completed a set of detailed functional specifications and a technical analysis for a new OASIS system, which will be implemented in FY 2004.

<http://nrs.harvard.edu/urn-3:hul.eresource:oasisxxx>

### Harvard Geospatial Library (HGL)

HGL is both a discovery tool and a data-mining environment for geospatial data sets. Unique to the digital library world, HGL provides researchers with detailed information about geospatial data and with the tools necessary to capture and deliver subsets of the data into their research environments. In FY 2003, LDI upgraded the program by adding a feature that allows researchers in

other systems to pass information into HGL—combining their data with HGL's to create customized maps. Other work on HGL in FY 2003 included metadata and cartographic searching improvements, cataloging and data-loading efficiencies, and infrastructure enhancements.

<http://nrs.harvard.edu/urn-3:hul.eresource:hgeodesy>

### Templated Database Service (TED)

A centrally supported web-based database, TED can be customized for collections or catalogs that do not fit within the scope of existing library catalogs at Harvard. TED offers collection managers an opportunity to create specialized databases that will be supported and upgraded along with other centralized systems such as VIA and OASIS.

<http://hul.harvard.edu/ois/systems/ted>

### Full-Text Search Service (FTS)

FTS is a discovery tool that provides researchers with the ability to search full text associated with scanned images.

## Delivery Services

LDI offers a number of format-specific delivery services developed to enable the delivery of digital objects stored in DRS to web browsers. These services include:

- Image Delivery Service (IDS) for delivery of still image files.
- Page Delivery Service (PDS) for delivery of scanned page images within the context of logical navigation—in other words, PDS mimics the page-turning functionality of a book.
- Streaming Delivery Service (SDS) delivers streamed media to web browsers. Note: SDS currently delivers audio files, but it is capable of delivering video as well.
- Asynchronous Delivery Service (ADS) allows curators and researchers to request large objects or sets of objects from DRS for downloading upon e-mail notification. Note: This new service is primarily used to deliver large image files from the Biomedical Image Library for printing or for creating three-dimensional models.

In FY 2003, significant efforts went into analyzing additional functionality for improvements to IDS that will be implemented next year; the user interface to PDS was redesigned; and SDS and ADS were developed as new services.

## **Storage and Management Systems**

### **Digital Repository Service (DRS)**

DRS is an integrated set of services to manage, maintain, preserve, and deliver Harvard's digital materials. In FY 2003, LDI upgraded the system to support audio files and established processes and procedures for auditing all copies of each digital object stored. As a repository, DRS is not visible to researchers and most curators.

<http://hul.harvard.edu/ois/systems/drs>

### **Name Resolution Service (NRS)**

Harvard's NRS assigns persistent identifiers to digital objects. Persistent identifiers provide curators and researchers with confidence that the URL they cite will always work.

<http://hul.harvard.edu/ois/systems/nrs>

### **Access Management Service (AMS)**

AMS provides secured access to Harvard's licensed or copyrighted materials. Using the University Personal Identification Number (PIN) and Directory Services, AMS protects the electronic assets of the University from unlawful access and also restricts access to the Harvard community as required by curators. In FY 2003, AMS was upgraded to work with the newest version of the University's Directory Service.

### Public Forums

LDI keeps the Harvard library community informed about infrastructure development, digital library projects, and related activities through articles and announcements in *Harvard University Library Notes*, presentations throughout the University, and a number of other public forums. Public forums in FY 2003 included brown-bag luncheons and open meetings on the Templated Database System (TED), the SFX dynamic resource-linking technology, the planned implementation of Harvard's new research portal using MetaLib software, and LDI grant-funded projects.

Additional digital library information, documentation, and publications are linked from the following web sites:

- The Library Digital Initiative (LDI) site focuses on information about the initiative, including technical developments, advisory services, and the grant program.  
<http://hul.harvard.edu/ldi>
- The Office for Information Systems site contains information about available Harvard University Library systems and services, including resources for the staff at Harvard's libraries, museums, and archives, and for information technology offices using LDI systems and services.  
<http://hul.harvard.edu/ois>
- The "Library Preservation at Harvard" site is a collaborative effort of the Weissman Preservation Center in the Harvard University Library and the Preservation & Imaging Department in the Harvard College Library. The site includes information about preservation and imaging services for both traditional and digital materials.  
<http://preserve.harvard.edu>

### LDI Steering Committee

Hugh Wilburn, *chair*, Librarian, Frances Loeb Library, Harvard Design School  
Nancy M. Cline, Roy E. Larsen Librarian of Harvard College  
Dale P. Flecker, Associate Director of the Harvard University Library for Planning and Systems  
Barbara S. Graham, Associate Director of the Harvard University Library  
for Administration and Programs  
Harry S. Martin III, Henry N. Ess III Librarian and Professor of Law, Harvard Law School  
Judith Messerle, Librarian, Countway Library of Medicine, Harvard Medical School  
Thomas Michalak, Executive Director, Baker Library, Harvard Business School  
Wendy Gogel, *committee staff*, Digital Library Projects Liaison, Harvard University Library

### Grant Review Committee

Donald Pfister, *chair*, Asa Gray Professor of Systematic Botany and  
Curator of the Farlow Library and Herbarium, Faculty of Arts and Sciences  
Robert Buckwalter, Associate Librarian, Harvard Law School Library  
Dale P. Flecker, Associate Director of the Harvard University Library for Planning and Systems  
Margaret Hale, Harvard College Library Digital Projects Librarian  
Judith Messerle, Librarian, Countway Library of Medicine, Harvard Medical School  
Wendy Gogel, *committee staff*, Digital Library Projects Liaison, Harvard University Library

### LDI Review Board

Sarah Thomas, *chair*, University Librarian, Cornell University  
Nancy Cline, Roy E. Larsen Librarian of Harvard College  
John F. Keane, Sr., Chairman, Keane, Inc.  
Clifford A. Lynch, Coalition of Networked Information  
Stuart Shieber, James O. Welch, Jr., and Virginia B. Welch Professor of Computer Science (FAS)  
Megan Sniffin-Marinoff, Librarian and Deputy Director, Arthur and Elizabeth Schlesinger Library  
on the History of Women in America, Radcliffe Institute for Advanced Study  
Karen Trainer, University Librarian, Princeton University  
Elizabeth Hess, *committee staff*, Senior Specialist for Academic Computing, Office of the Provost

July 2003

### Library Digital Initiative

Dale P. Flecker  
Harvard University Library  
Office for Information Systems  
1280 Massachusetts Avenue, Suite 404  
Cambridge, MA 02138

T 617.495.3724

F 617.495.0491

<http://hul.harvard.edu/ldi>