

*Harvard University*  
*Report to the Digital Library Federation*  
*October, 2005*

**TABLE OF CONTENTS**

- i. [Collections, services, and systems](#)
- ii. [Projects and programs](#)
- iii. [Specific digital library challenges](#)
- iv. [Digital library publications, policies, working papers, and other documents](#)

---

**I. Collections, services, and systems**

**A. Collections**

**Ben Shahn at Harvard**

Ben Shahn at Harvard is a searchable database of images and information relating to works by the American artist Ben Shahn (1899 - 1969) in the collections of the Harvard University Art Museums and the Harvard University Library.

<http://www.artmuseums.harvard.edu/shahn/>

**Biomedical Image Library (BIL)**

A central catalog and collection of biomedical images produced in support of basic biomedical research. Biologists, medical scientists, and clinicians will be able to use the Biomedical Image Library to distribute their work to the community or to identify and retrieve data for novel analysis. Educators and students will find a ready collection of images to support learning. BIL provides access to data such as stacks of serial sections that cannot be published through traditional means. <http://nrs.harvard.edu/urn-3:hul.eresource:bioimlib>

**The Boston Transit Collection at the Fogg Art Museum**

Approximately 700 glass plate negatives and prints documenting the construction of the city's subway and elevated railway systems between 1895 and the beginning of World War II. The Boston Transit Collection is a component of the Carpenter Center Photograph Collection on deposit at the Fogg Art Museum since 2002.

<http://www.artmuseums.harvard.edu/btc/pages/introduction.html>

## **Botanical and cultural images of Eastern Asia from the Arboretum archives - Ernest Henry Wilson's photographs**

Thousands of botanical and cultural images of Eastern Asia by E. H. Wilson from the Arnold Arboretum Horticultural Library.

[http://www.arboretum.harvard.edu/programs/eastern\\_asia/wilson.html](http://www.arboretum.harvard.edu/programs/eastern_asia/wilson.html)

## **Bracton Online**

A digital presentation of *Bracton: De Legibus Et Consuetudinibus Angliæ* (Bracton on the Laws and Customs of England), the first comprehensive attempt to rationally articulate English law. The 13th century document is commonly attributed to the English judge and scholar Henry of Bratton. Here the Latin and an English translation can be searched and viewed individually or side-by-side.

<http://hls.law.harvard.edu/bracton/>

## **Degas at Harvard**

A searchable web site, offered as an introduction to the holdings of works of art by Hilaire-Germain-Edgar Degas in the collections of Harvard University. Although most of the Degas works at Harvard are in the Fogg Art Museum, Harvard University Art Museums, there are also works held by The Houghton Library (<http://hcl.harvard.edu/houghton/>) and The Dumbarton Oaks Research Library and Collection, Washington, D.C. (<http://www.doaks.org/>).

<http://www.artmuseums.harvard.edu/degas/pages/frameset.html>

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

Thousands of pages of scanned images of rare and unique musical scores drawing on Harvard's extensive collections of first and early editions of Bach family composers, Mozart, and multiple versions of 19th-century opera from the Eda Kuhn Loeb Music Library, Harvard College Library.

<http://hcl.harvard.edu/libraries/loebmusic/collections/digital.html>

## **Dürer's Passions**

Highlights and excerpts from a Fall 2000 exhibition on display at the Busch-Reisinger Museum, Harvard University Art Museums.

<http://www.artmuseums.harvard.edu/exhibitions/featured/passion/index.html>

## **A Grand Legacy: Arts of the Ottoman Empire**

Highlights and excerpts from a 1999-2000 exhibition on display at the Arthur M Sackler Museum, Harvard University Art Museums.

<http://www.artmuseums.harvard.edu/exhibitions/featured/GL/intro.html>

### **Harvard Daguerreotypes**

Images of more than 3,500 daguerreotypes, the first publicly-announced photographic process, from photograph collections throughout the University used in research and instruction.

<http://preserve.harvard.edu/exhibits/daguerreotypes.html>

### **Harvard/Radcliffe Online Historical Reference Shelf (H/R OHRS)**

Electronic access to frequently consulted sources on the history of Harvard and Radcliffe including annual reports, narrative histories, writings, statistics, founding documents, Massachusetts legislation concerning Harvard, Harvard music songs sung at football games and other ceremonial occasions, serial publications, and media coverage.

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

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

More than 5,000 photographs from the Harvard-Yenching Library, Harvard College Library taken by Hedda Hammer Morrison (1908-1991) during her residence in Beijing from 1933 to 1946. Her photographs document lifestyles, trades, handicrafts, landscapes, religious practices, and architectural structures that in many cases have all but disappeared from modern China.

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

### **Investigating the Renaissance**

An interactive demonstration by conservators and scientists in the Straus Center for Conservation, located in the Fogg Art Museum showing the ways in which computer technology can be harnessed to add to our knowledge about Renaissance paintings and how they were made. Computer-assisted imaging can reveal aspects of the process of making art not visible to the unaided eye. It also reveals the alterations of intervening centuries, alterations that were intended to repair the ravages of time and use, and to adjust images to reflect changing aesthetic preferences.

<http://www.artmuseums.harvard.edu/Renaissance/help.html>

### **Legal Portraits Online**

Over 4,000 portrait images of lawyers, jurists, political figures, and legal thinkers dating from the Middle Ages to the late twentieth century from the Harvard Law School's Legal Portrait Collection. The collection of these prints, drawings, and photographs depict legal figures prominent in the Common Law as well as those associated with the Canon and Civil Law traditions.

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

### **The Mercator Globes at Harvard Map Collection**

A presentation of images of the Mercator Globes at the Harvard Map Collection, Harvard College Library with zooming and navigation. Mercator was a prolific publisher of maps and atlases, but he is only known to have produced one version of a globe pair: a terrestrial globe in 1541 and a matching celestial globe in 1551. Surviving examples of the Mercator globes are rare, and the pair at the Harvard Map Collection are the only known matched pair in America.

<http://hcl.harvard.edu/mercatorglobes/>

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

40,304 digital images of Maya archaeological photographs selected from the Carnegie Institute of Washington Collection in the Photographic Archives of the Peabody Museum of Archaeology and Ethnology. Many of the buildings, monuments, and artifacts that are recorded in the photographs no longer exist, are badly damaged or are so difficult to access that they are unavailable to researchers. Discovery and delivery of the digital images has improved access to the photographs for government researchers working on accurate restoration and reconstruction of the sites, linguists needing undamaged scripts, archaeologists, historians, publishers, and producers. View the images in VIA:

<http://via.harvard.edu>

### **Nineteenth-Century American Trade Cards**

More than 1,000 images of 19th c. advertising trade cards selected from the Historical Collections at Baker Library, Harvard Business School. As one of the most popular forms of advertising in the nineteenth century, and as an indicator of consumer habits, social values, and marketing techniques, trade cards are of interest to scholars of business history, American studies, graphic design and printing history, and social and cultural history.

[http://www.library.hbs.edu/hc/19th\\_century\\_tcard/](http://www.library.hbs.edu/hc/19th_century_tcard/)

### **The Nuremberg Trials Project**

The Nuremberg Trials Project provides access to digitized documents from the Harvard Law School Library relating to the trial of military and political leaders of Nazi Germany before the International Military Tribunal (IMT) and to the trials of other accused war criminals before the US Nuremberg Military Tribunals (NMT).

[http://nuremberg.law.harvard.edu/php/docs\\_swi.php?DI=1&text=nur\\_13tr](http://nuremberg.law.harvard.edu/php/docs_swi.php?DI=1&text=nur_13tr)

### **Piet Mondrian: The Transatlantic Paintings**

The term "transatlantic paintings" refers to a group of 17 works that Mondrian started (and in some cases finished) in Europe between 1935 and 1940, and finished (or refinished) in New York after his arrival there in the fall of 1940. Many of Mondrian's

works crossed the ocean during his lifetime, but the seventeen transatlantic paintings are the only ones he worked on in both Europe and America.

<http://www.artmuseums.harvard.edu/mondrian/index.html>

### **Sargent at Harvard**

Sargent at Harvard is a searchable database of images and information relating to the American artist John Singer Sargent (1856-1925) in the collections of the Harvard University Art Museums (Fogg Art Museum) and the Harvard University Portrait Collection.

<http://www.artmuseums.harvard.edu/sargent/>

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

Selected audio recordings and text files of oral literature made by Professor Milman Parry of the Department of the Classics at Harvard University in 1933-35 in Yugoslavia and the epic texts collected by Professor Albert B. Lord in 1950-51; and an exhibition of photographs. The Milman Parry Collection of Oral Literature is the largest single repository of South Slavic heroic songs in the world.

<http://www.chs.harvard.edu/mpc/>

### **Sunk in Lucre's Sordid Charms - South Sea Bubble Resources in the Kress Collection at Baker Library**

Digital images and full text from pamphlets, books, broadsides, prints, and ephemera focusing on the South Sea Bubble stock market crisis in the early part of the eighteenth century. The resources are found in the Kress Collection and additional materials at Baker Library, Harvard Business School.

<http://www.library.hbs.edu/hc/ssb/index.html>

### **Western China and Tibet: Hotspot of Diversity**

A selection of Harvard's historic and contemporary ethnographic and natural history collections related to western China and Tibet including material from the collections of the Arnold Arboretum, the Harvard Map Collection, the Botany Libraries, the Museum of Comparative Zoology, the Harvard-Yenching Institute and the Harvard University Herbaria. Beginning in 1924 with the Arnold Arboretum's Expedition to northwestern China and northeastern Tibet led by Joseph F. Rock, the historic collections include plant and bird specimens, as well as photographs of the region's landscape, architecture and people. The Herbaria have been collecting contemporary biological specimens from the same region. By relating the historic and contemporary material from various repositories, the collection provides students and scholars with access to information about the area's natural and ecological resources, as well as the social and cultural history of the region.

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

## **Women Working, 1800-1930**

Thousands of digitized historical, manuscript, and image resources selected from Harvard's library and museum collections. Women Working, 1800-1930 explores women's roles in the US economy between 1800 and the Great Depression and includes documentation of working conditions, conditions in the home, costs of living, recreation, health and hygiene, conduct of life, policies and regulations governing the workplace, and social issues.

<http://ocp.hul.harvard.edu/ww/>

## **B. Services**

### **Reserves List Tool**

As one of the first collaborative efforts between them, the Harvard University Library, the Provost's Office and the Instructional Computing group for the Faculty of Arts and Sciences have jointly developed a reserves list tool for use by instructors, library staff and students. The Reserves List Tool allows faculty members to submit reserves requests to the library via a new module in the course "toolbox" available for the creation and maintenance of course web sites. The information is transmitted to the library, where reserves library staff fulfill the requests and update the citation descriptions. Students can then view the list with availability information, including direct digital links to e-resources such as journal articles, on their course web site.

## **C. Systems**

### *Catalogs and Discovery Systems*

#### **E-Research @ Harvard Libraries**

On June 30, 2005, the Harvard Libraries implemented MetaLib software from Ex Libris to launch a set of new tools for access to e-resources and e-journals and introduce federated searching and personalization features.

<http://e-research.lib.harvard.edu>

#### **OASIS Reimplementation**

In January 2005, OIS launched version 2.0 of OASIS (Online Archival Search Information System), the online union catalog of finding aids for archival collections at Harvard University. Searching more than 2,400 finding aids from 18 Harvard archives and libraries with a single query, OASIS facilitates the discovery of a wide range of primary research materials, including letters, literary manuscripts, business records, musical scores, diaries, photographs, drawings, printed material, and realia. Version 2 is the first major redesign of the OASIS system since its debut in July 1998. Highlights of the new release include a completely new user interface consistent with the look and feel of other OIS applications (such as VIA and the Page Delivery Service); an underlying XML database that supports indexed and wild card searching; documents represented in

the EAD ("Encoded Archival Description") standard format; the ability to browse alphabetically and by repository; the ability to download or print finding aids; the ability to refine searches by searching within a results set, or combine searches to broaden the results; a "Search History" that lets users see what search strategies they have tried throughout a session and re-display the results; and the ability of curators to redact portions of a finding aid to protect confidential information. OASIS finding aids are also exposed to the wider community for searching by external search engines such as Google and RLG.

<http://oasis.harvard.edu>

### *Delivery and Search Systems*

#### **Reimplementation of Page Delivery Service (PDS)**

Page Delivery Service (PDS) was reimplemented during FY2005. The PDS provides a navigational environment for digital surrogates of monographs, serials, manuscripts, musical scores, and other page-oriented materials. The FTS provides services for indexing and searching textual content in the PDS as well as content delivered through other systems. As part of the PDS reimplementation, the internal structure of page-turned digital objects was changed to use METS (Metadata Encoding & Transmission Standard), an XML schema widely adopted as a standard within the digital library community. Over 8,000 page-turned objects pre-existing in the DRS were automatically converted to METS form. The new PDS also features a newly designed user interface that provides simplified page navigation, a graphic representation of the logical structure of the entire page-turned object, and maximizes the screen area available for the page images.

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

#### **Reimplementation of Full-text Search Service (FTS)**

Full-text Search Service (FTS) was reimplemented during FY2005. The reimplemented system is based on Lucene, an open-source text search engine that provides significantly-improved system response time compared to the previous version of the FTS, an important consideration as the FTS now indexes over 615,000 pages of text. Lucene is also Unicode-aware, which will provide support for non-English text in the future.

<http://hul.harvard.edu/ois/systems/#fts>

### *Digital Infrastructure Tools*

#### **JHOVE (JSTOR/Harvard Object Validation Environment)**

JHOVE, the JSTOR/Harvard Object Validation Environment (pronounced 'jove') is a software tool that was locally designed and developed for format-specific identification, validation, and characterization of digital objects. The ability to identify, validate, and characterize digital objects properly is a fundamental requirement for effective long-term preservation. By fully automating what had previously been a primarily manual process, JHOVE significantly enhances the timeliness and sophistication with which preservation



institutions are able to deal the ever increasing amounts of digital data requiring preservation handling. JHOVE provides support for the most prevalent digital formats routinely used for representing audio (AIFF, WAVE), image (GIF, JPEG, JPEG 2000, TIFF), text (ASCII, UTF-8), and document (PDF, XML) content, but is designed as an extensible framework to facilitate the integration of additional formats over time. OIS is integrating JHOVE into the Digital Repository Service (DRS) workflows for deposit, archival storage, and preservation planning. Beyond Harvard, JHOVE has gained widespread international acceptance and use by most major library and archival institutions with significant digital library and preservation programs.

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

---

## II. Projects and programs

### A. Projects

#### New Project Announcements

##### **Google Project**

In December, 2004 the University announced that it had entered into a pilot agreement with Google to investigate the possibility of digitizing much or all of the bound-volume collection of Harvard's libraries. The digitized data would be used by Google to provide searching across the full content of books (thus providing a new way for readers to search for materials in the Harvard collection). In addition, a copy of the data would be returned to Harvard for future use. The pilot project involves digitizing a sample of library books to allow both the University and Google to explore such issues as workflow, possible damage to books caused by the digitization project, the quality of the data created through Google's unique digitization process, the integration of Google's systems and services with those of the Harvard libraries, and costs. In addition, Harvard wanted to have time to gather reactions from the many communities (Harvard students and faculty, alumni, other libraries and universities, publishers and authors) that were expected to be interested in the project.

[http://hul.harvard.edu/news/2004\\_1214\\_faq.html](http://hul.harvard.edu/news/2004_1214_faq.html)

#### Update on Existing Projects

##### **Archive Ingest and Handling Test (AIHT)**

Harvard University participated in the Archive Ingest and Handling Test (AIHT) organized by the Library of Congress as part of its National Digital Information Infrastructure Preservation Program (NDIIPP) initiative. The intent of the test was to assess the feasibility of large-scale transfer of digital resources between institutions utilizing quite different



technical infrastructures. The test corpus was a collection of over 57,000 files (13 TB) in more than 100 data formats, provided with no accompanying technical metadata. The technical metadata required for deposit into the Digital Repository Service (DRS) was automatically extracted from the files themselves. The test successfully proved that the four project participants could import and export the data amongst themselves, despite the heterogeneity of their repository systems, variously using RDBMS, XML, grid storage, and MPEG-21 technologies. This project phase did uncover a few deficiencies in the DRS in dealing with arbitrary, web-harvested content, rather than with the highly-curated materials typically deposited. These issues will be addressed in future DRS system enhancements. A further stage of the AIHT project involved the systematic migration of objects in the collection. The investigation included the automated transformation of GIF, JPEG, and TIFF images (over 15,000 files) to the JPEG 2000 format. The systems and workflows devised for this transformative process will be incorporated in the DRS in the future, providing collection managers with the option to request retrospective conversion of existing visual resources.

### **GDFR (Global Digital Format Registry)**

The Harvard University Library (HUL) has received a grant from the Andrew W. Mellon Foundation for the development of a registry of authoritative representation information about digital formats. Detailed representation information, which defines the syntactic and semantic rules by which content is encoded in digital form, is fundamental to the preservation of digital resources. The two-year project will result in a new Global Digital Format Registry (GDFR), which will become a key international infrastructure component for the digital preservation programs of libraries, archives and other institutions with the responsibility for keeping digital resources viable over time.

The wide diversity and rapid pace of adoption and abandonment of digital formats present an ongoing problem for long-term preservation efforts. Preservation programs must document the format of the objects they are preserving. Without precise knowledge of format, a digital object is merely a collection of undifferentiated bits. Creating a shared registry will save an enormous amount of duplicative effort in acquiring and recording such documentation. It also allows the community to share expertise in formats, so that each institution does not require deep local expertise in every format of data it is preserving.

GDFR will be established as a distributed service in which participating research libraries, archives, and other organizations with preservation responsibilities can contribute, as well as use, format-typing information <http://hul.harvard.edu/gdfr/>

## **B. Programs**

### **The Library Digital Initiative (LDI)**

Harvard University launched the LDI in July 1998 to develop the University's capacity to manage digital information by creating a robust technical infrastructure for the acquisition, organization, delivery, and archiving of digital library materials; by providing a team of specialists to advise librarians and others in the University community on key issues in the digital environment; by providing librarians and staff with experience in digital library projects; and by enriching the Harvard University Library system with a significant set of digital resources. Now in its eighth year, LDI is making it easier for Harvard's libraries to maintain their collections and services in the digital era, without each library having to individually acquire the expertise and systems needed to support digital resources. The development of the collections, systems and services documented in this report were funded by LDI.

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

### **The Digital Acquisitions Program**

Initiated as part of LDI, the Digital Acquisitions Program 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. Program staff are also involved in assisting libraries with collection decisions involving print resources, such as canceling unneeded duplicate print journal subscriptions in order to control acquisitions. During FY 2005, approximately 709 new resources - including 645 e-journals and 64 databases - were licensed and made available to the Harvard community through the Harvard Libraries web site. User sessions on commercial resources increased 22% this year from 4,269,955 to 5,193,132.

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

### **LDI Internal Challenge Grant Program**

Managers and staff throughout Harvard's libraries, archives, museums and special collections have participated in LDI through the Internal Challenge Grant Program. They have assisted LDI by prioritizing, testing and demonstrating new systems and services while contributing valuable online content for research and education. Projects have had a range of goals including basic digital conversion of a single collection; the creation of a virtual collection by digitizing related material from multiple repositories; and the development of new delivery systems for natively digital material. 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. To date, the LDI grant program has funded 42 projects through which more than 200 Harvard staff members have gained experience in working with digital projects. seven

projects were completed in FY 2005. Four projects employed LDI's Management Assistance and Planning programs (LDI MAP), a cost-recovery service that has provided customized, hands-on assistance to managers of LDI grant-funded projects. New grants in FY06 will focus on archiving web sites.

[http://hul.harvard.edu/ldi/html/funded\\_projects.html](http://hul.harvard.edu/ldi/html/funded_projects.html)

### **Open Collections Program (OCP)**

Through Harvard's Open Collections Program, resources from Harvard's libraries are made available online to benefit students and teachers around the world. With the generous support of the William and Flora Hewlett Foundation and the Lisbet Rausing Charitable Fund, the Open Collections Program (OCP) is developing efficient, replicable methods for the creation of comprehensive, subject-based digital resources drawn from the holdings of Harvard library system. The goal is to create a new model for digital collections that will benefit the Harvard community and the general public alike. In FY2005, OCP completed their pilot collection, Women Working, 1800-1930 and began work on their second collection, Emigration and Immigration, 1789-1930 which will focus on immigration to the United States.

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

## **III. Specific Digital Library Challenges**

### **IV. Digital library publications, policies, working papers, and other documents**

- Abrams, Stephen and Chapman, Stephen et. al. "Harvard's Perspective on the Archive Ingest and Handling Test" D-Lib Magazine, December 2005; Volume 11 Number 12. ISSN: 1082-9873  
<http://www.dlib.org/dlib/december05/abrams/12abrams.html>
- Abrams, Stephen. "Establishing a Global Digital Format Registry," Library Trends; 2005 Volume 54, issue 1, 125-43.
- Abrams, Stephen. "The Role of Format in Digital Preservation," VINE; 2004 Volume 34, Issue 2, 49-55. ISSN: 0305-5728
- Chapman, Stephen. "Microfilm: A Preservation Technology for the 21st Century?," IS&T's 2005 Archiving Conference: Final Program and Proceedings, Society for Imaging Science and Technology; 2005, 228-32.

- Chapman, Stephen, and Merrill-Oldham, Jan. "Reply to ARL's Recognizing Digitization as a Preservation Reformatting Method," *Microform & Imaging Review*; Autumn 2004 Volume 33, No. 4.
- Mandell, Lee and Sue Kriegsman. "Digital Repository Planning and Policy" *IS&T's 2005 Archiving Conference* Washington, D.C.; April 27, 2005; p. 5-8; ISBN: 0-89208-255-0.  
<https://www.imaging.org/store/epub.cfm?abstrid=32216>
- Wendler, Robin. "The Eye of the Beholder: Challenges of Image Description and Access at Harvard." *Metadata in Practice*. Chicago, Ill: American Library Association, 2004. p 51-69. ISBN: 0-8389-0882-9
- 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 Harvard. The site includes information about preservation and imaging services for both traditional and digital materials.  
<http://preserve.harvard.edu>
  - Online exhibitions and searchable collections from the Harvard University Art Museums are available at:  
<http://www.artmuseums.harvard.edu/home/index.html>