

# Columbia University Libraries

## Report to the Digital Library Federation August 15, 2003

### TABLE OF CONTENTS

#### [I. Collections, services, and systems](#)

#### [II. Projects and programs](#)

---

## I. Collections, Services, and Systems

### A. Collections

#### *Current Material*

1. **Licensed Electronic Collections.** In March 2003, The Libraries completed a year-long project to upgrade its 1997-vintage LibraryWeb publishing system for current, licensed electronic resources. Our new, more powerful process now involves extraction of records for ejournal and electronic reference tools directly from our online catalog, loading into the enterprise SQL system (currently IBM's dB2 product) that acts as our [Master Metadata File](#), and the generation of real-time lookup and browse functionality. This enhanced system allows us to create interfaces and retrieval methods that are optimized for electronic resources in a way that cannot yet be done in our online catalog; it also provides a bibliographic knowledge base that will be used for other types of value added services, such as database advisers, research guides, and quick bibliographies. ([Columbia Eresources working page](#))
2. **Columbia Educational Resources Online** is a collection of approximately 100 "E-seminars", three to five hour online classes, that bring together a professor's instruction with supplemental teaching tools in History, International Affairs, Public Health, and Science. Teaching tools include primary texts, interactive maps and timelines, simulations and animations. E-seminars are often taught by the most distinguished faculty of the University and incorporate videos of lectures delivered on campus. The collection is available through license. ([ceri.columbia.edu](http://ceri.columbia.edu))
3. **Columbia American History Online (CAHO)** is a subscription-based collection of resources designed for students and teachers at the high-school and college levels. CAHO integrates curricular materials, such as document-based questions, simulations, and teaching activities with e-seminars derived from four semester-length Columbia courses: Intellectual and Cultural History of the United States, 1890-1945 (Casey Blake), America Since 1945 (Alan Brinkley), Slavery and Emancipation (Eric Foner), The History of the City of New York (Kenneth Jackson) ([caho.columbia.edu](http://caho.columbia.edu))

#### *Historical & Specialized Collections*

1. **Advanced Papyrological Information System (APIS).** The APIS project, for which Columbia is currently the technology host, completed its Phase III in March 2003 and was funded by NEH for an additional two-year grant beginning in July. At the end of Phase II, there were approximately 19,000 metadata records, some 3400 of which have translations, and 16,000 unique digital images. Phase III brought two new papyrus collections into APIS (U. of Toronto and Washington University). It also included a complete rewrite of the APIS search and retrieval application and a migration of the database to a new, more current metadata platform using Columbia's [Master Metadata File](#) version 2. ([APIS Public Home Page](#) -- [Columbia APIS working page](#))
2. **Bunraku Puppet Theater Collection.** The Barbara C. Adachi Bunraku Collection was donated by Ms. Adachi to Columbia University's C. V. Starr East Asian Library in 2000. This extensive collection

documents the significant post-World War II revival of popular interest in bunraku, a type of traditional Japanese puppet theater. The collection spans the 1960s through the 1990s and consists of more than 12,500 slides and nearly 7,000 black-and-white photographs of rehearsals, performances, and workshops, as well as theater programs in Japanese and English, texts of the plays performed, and audio and video recordings of interviews with masters of the modern Japanese puppet theater. A selection of the images from this collection have been digitized and made available while additional funding is being sought for a more comprehensive project.

([Bunraku Collection Public Home Page](#) -- [Columbia Bunraku Project working page](#))

3. **Digital Anthropological Resources for Teaching (DART).** DART is a partnership with the London School of Economics, funded jointly by the NSF (US) and the JISC (UK). Four postdoctoral Fellows, two in each partner Anthropology department, are developing digital collections to support teaching at the undergraduate level. The project is exploring the differences between the institutions both in the study and teaching of Anthropology and in our technology approaches. We will identify and hopefully address scalability issues in the existing Columbia technical infrastructure for acquisition, cataloging, and automated tools for site generation and conceptual navigation.
4. **Digital Scriptorium.** The Digital Scriptorium project, funded chiefly by grants from the National Endowment for the Humanities, completed its most recent grant period in 2003. The database includes images and cataloging for the holdings of Berkeley and Columbia as well as additional holdings from the Huntington Library, the University of Texas, Austin, the New York Public Library along with those of a number of smaller collections. As of July 2003, the Digital Scriptorium includes some 4,000 catalog records and 15,000 digitized images. Columbia has submitted a grant application to NEH to allow for expansion of the database's content with additional key U.S. research library holdings and migration to an updated and enhanced technology and metadata environment.  
([Digital Scriptorium Public Home Page](#) -- [Columbia Digital Scriptorium working page](#))
5. **Greene & Greene Virtual Archive.** The "Greene & Greene Virtual Archive" — a joint project of The Gamble House/USC, Berkeley and Columbia and funded by the Getty Foundation — was completed in June 2003. Its objective was the creation of a scholarly web site presenting architectural drawings and photographs by Charles and Henry Greene, American Arts and Crafts movement architects working in the early 20th century. Cataloging (in the form of archival finding aids) and images were contributed by Columbia, Berkeley, and the Huntington Library; Berkeley is providing technical support for the metadata compilation and web presentation.  
([Greene & Greene Virtual Archive Website](#) -- [Columbia Greene & Greene WebSite](#))
6. **John Jay Papers.** The two-year "Papers of John Jay" project, funded by a \$150,00 grant from NEH and supplementary contributions from the Florence Gould Foundation and the Columbia Libraries' Friends fund, was completed in April 2003. It provides an online index to all known documents written by or to John Jay, 1745-1829, one of America's 'founding fathers,' a distinguished statesmen and a graduate of Columbia, then King's College. The project includes the creation of a searchable database of descriptive metadata — including abstracts — at the item level for all of the approximately 13,400 unique letters, memos, diaries, etc. in the collection. Some 100,00 page images will be scanned as 8-bit grayscale TIFFs at 300 dpi and linked to the metadata records. Additional enhancements to database content and functionality are underway and will be completed Fall 2003.  
([Jay Papers Public Home Page](#) -- [Columbia Jay Papers Project working page](#))
7. **Medieval & Renaissance Manuscripts.** Columbia's Medieval Renaissance Collection is the first virtual collection to be mounted in the Columbia's local implementation of the [Luna Insight®](#) system. It will contain cataloging and selected images for all the ca. 1400 medieval and renaissance manuscripts held by the numerous libraries of Columbia University. Approximately 4,000 images will be included initially. These manuscripts represent both documentary (or archival) sources and those that are or began their lives as books. Descriptions are schematic; chosen images range from one to 123 per manuscript, with an average of six images per codex. Under the name, [Digital Scriptorium](#), but without the "Insight" interface, the same Columbia descriptions and images, as well as those of a number of other institutions are

available at:

(<http://sunsite.berkeley.edu/scriptorium/form.html> -- [Columbia Medieval & Renaissance Manuscripts Project working Page](#))

8. **Joseph Urban Theatre Collection.** The Joseph Urban Stage Design Models and Documents stabilization and access project, funded by a \$207,289 from NEH will preserve 240 three-dimensional stage models created by Joseph Urban for New York theaters between 1914-1933, including productions for the Ziegfeld Follies, the Metropolitan Opera, and a variety of Broadway theaters. The project will also create and link digital images of related stage design documents and drawings to the existing online finding aid. ([Urban Theatre Collection Working Page](#))

## B. Services

1. **National Science Digital Library (NSDL).** Columbia participates in the NSF-funded NSDL Core Integration team. Our work focuses on developing sustainability models and a scalable access management system for the library. For the latter, we have been active developers of the Internet-2 Shibboleth system, have worked with NSDL collections and commercial publishers on their pilots of the software, and have integrated Shibboleth into the central NSDL portal (uPortal framework).
2. **Online Cost and Use Evaluation.** This project, funded by the Andrew W. Mellon foundation, is exploring how digital resources are affecting the overall scholarly communications process, in terms of cost throughout the life cycle of the publication process; how the use of these resources is affecting qualitatively and quantitatively the research and teaching patterns of scholars and students; and the financial models that will allow for sustainability of these products over the long term.
3. **BorrowDirect.** The "*Borrow Direct*" rapid book request and delivery service was expanded beyond its original three members (Columbia, Penn and Yale) in the fall of 2002. The service now provides access to the circulating collections of seven libraries: Brown, Columbia, Cornell, Dartmouth, Penn, Princeton and Yale. The system runs on Dynix's URSA software, with a central server maintained at Yale and project coordination at Penn. Users search a virtual catalog that checks shelf status and filters out non-circulating items. Target time from online request to notification for pick up is four working days or less. The service is increasingly popular: between October 2002 and June 2003 more than 55,000 items were borrowed/lent by members, with an average fill rate of 84.5%. Approximately half of Columbia's interlibrary borrowing is now filled by Borrow Direct, with the remainder being handled through ILL Manager and "traditional" mediated ILL. ([BorrowDirect @ Columbia Home Page](#))

### 4. Online Reference

'*Ask Us Now*' Columbia Libraries' on-line reference service began operation in May 2002. Since that time, reference librarians from Columbia, Barnard, and Teachers College libraries have staffed the service, responding to over 1,000 questions from affiliated students, faculty, and staff from the three institutions.

*Ask Us Now* is available to anyone with a valid Columbia e-mail ID and password, and is available via any World Wide Web connection, on or off campus. The service hours are 1-5 p.m., Monday through Friday. Beginning February 24, *Ask Us Now* switched to software from Live Assistance (<http://www.liveassistance.com/>). This software extended the availability of the service to users of Macintosh and Linux operating systems and to a greater variety of Internet browsers, and enhanced the speed of communications.

([Ask Us Now FAQ](#))

5. **Institutional Repository / DSpace Planning.** Columbia Libraries, the Electronic Publishing Initiative at Columbia (EPIC) & Academic Information Systems have been active in institutional repository work. Over the last year, we have been participating in the initial development of the DSpace federation. Technical work includes the integration of local access management, workflow, and acquisition systems into the DSpace software. More generally, we are investigating policy and process options for a Columbia Institutional repository, without regard to particular software. This includes projects that have mediated,

unmediated, and 'curated' methods for contribution; electronic publishing projects; course management projects (course reserves); a campus wide image repository; and a long-term digital archiving plan.

## C. Systems

- A. **Semantic Indexing.** Several discovery and navigation applications are now based on ISO 13250 topic-maps. Columbia has developed tools to edit these data structures and to generate several kinds of web navigation and indexes from them.
- B. **Workflow Tools** have been developed to manage deposit and track digital asset life cycle, progressively attaching metadata while enforcing approval processes, from rights clearance, to descriptive metadata (editorial), to technical and structural metadata (production and deployment).
- C. **Metadata framework.** This system stores and converts metadata, allowing libraries of digital resources to be reused in a variety of ways. It can enforce conformance to custom or standard metadata schemas and can represent multiple hierarchies and structural relationships.
- D. **OAI-PMH server.** An OAI server is currently serving Earth Science metadata as part of our participation in the National Science Digital Library. Over the next year we plan to add records from several other collections to this server.
- E. **Endeavor Voyager Implementation.** Columbia Libraries implemented Endeavor's Voyager system on July 7, 2003. Columbia worked with the vendor to successfully integrate Voyager's authorization system with the standard campus authentication system (LDAP). The Voyager system replaces the Libraries' previous NOTIS-based system, implemented in 1989.

## II. Projects & Programs

- A. **Libraries Digital Program Division.** Columbia Libraries and Information Services have established a new Digital Program Division to advance digital technology tools and resources at the Libraries. The Division, launched in September 2002, consists of seven staff members, bringing together programmers and specialists already working on digital projects elsewhere in the Libraries' organization and creating three additional positions. Stephen P. Davis, previously director of Columbia's Library Systems Office, is the Director of the new division.

In the near term the Libraries Program will focus on three main areas: developing new tools to improve access to, and enhance the use of published electronic resources required for University teaching and scholarship; creating innovative scholarly tools and new digital presentations from Columbia Libraries' archival and special collections and collaborating with other research libraries and institutions where appropriate; and implementing a plan for the long-term archiving and preservation of Columbia's digital content by developing institutional and other types of digital repositories. A key goal of the program is to acquire or develop a manageable, scalable and robust software systems platform to deliver digital library resources to the University and scholarly community.

([Press Release](#) -- [LDPD Home Page](#))

### B. New Media Center Joins Libraries

In January 2003, the Center for New Media Teaching and Learning (CCNMTL) joined the University's Information Services Division, which also includes the University Libraries as well as the Electronic Publishing Initiative at Columbia (EPIC), the Center for Research on Information Access (CRIA), and Academic Information Systems (AcIS). The move of CCNMTL to the Information Services Division signals the University's commitment to the Center as an important educational and research unit at Columbia. ([Press Release, 1/31/03](#) -- [CCNMTL Home Page](#))

- C. **Computational Linguistics for Metadata Building Project (CLIMB).** In April 2002, The Center for Research on Information Access (CRIA) at the Columbia University Libraries received a \$542,000 grant

from the Andrew W. Mellon Foundation for the "Computational Linguistics for Metadata Building Project" (CLIMB). This project brings together the most recent developments in natural language processing and applies them to the problems of automatically extracting metadata from text. A two-year research project, CLIMB proposes to develop innovative uses of computational linguistic techniques for the identification and extraction of descriptive metadata with the purpose of improving access to image collections. The strategy proposed has the potential to provide rich, subject-oriented indexing for large image collections that would otherwise be prohibitively expensive to describe and index using manual techniques. A further advantage of the approach is that the descriptive metadata generated may be derived from authoritative scholarship in a way not normally feasible in standard cataloging practice.

([CLIMB Website](#))

#### **D. New Geographic Information Systems (GIS) Initiative**

A full-time GIS librarian position has been created in the Electronic Data Service (EDS) in order to expand the services offered in support of instruction and research that involve geo-spatial data. Geographic Information Systems (GIS) software and the supporting manuals have long been available in the EDS computer lab. Last year, through the EDS web site, ESRI online tutorials in the use of ArcGIS and ArcView software were made available to the Columbia community. Currently spatial files are being added to the EDS data collection and the EDS computer lab has expanded.