

THE LIBRARY OF CONGRESS AND THE NATIONAL DIGITAL LIBRARY IN THE U.S.A.

**Winston Tabb, *Associate Librarian for Library Services*
*Library of Congress (U.S.A.)***

ABSTRACT

National libraries have multiple roles depending on the history and tradition of their nation, the country's information infrastructure and policies, and the responsibilities of other libraries within the country. But national libraries also have much in common, as indicated in UNESCO's definition of national libraries as: Libraries which, irrespective of their title, are responsible for acquiring and conserving copies of all significant publications published in the country and functioning as a deposit' library, whether by law or under other arrangements. Among the most important responsibilities of a national library is acquiring (often through copyright deposit or *deposito legal*), preserving and making accessible the comprehensive creative record of its citizens. This traditional role continues to be of primary importance, but must constantly be re-examined and adapted as new forms of creativity and communication emerge, while resources remain steady at best.

The Library of Congress conducts a number of programs to assure that it can continue to meet its traditional national library responsibilities with digital media. Its National Digital Library Program is a partnership between the public and private sectors. The Library of Congress also has established cooperative arrangements with other research libraries, publishers, and corporations to ensure that digital resources are acquired and made enduringly accessible.

As the *de facto* national library of the United States, the Library of Congress has a leadership role in the U.S. national digitization project to preserve America's cultural and historical heritage for future generations through digital means and to collect new materials being produced in digital formats in order to ensure that such current digital materials are available to the Library's main constituencies: the Congress of the United States, the scholarly research communities of the U.S. and other countries, and America's 25,000 other libraries and their users. For the Library of Congress, digitization projects are a natural extension and expansion of its traditional responsibilities to acquire, preserve, and make accessible a universal collection of knowledge and creativity for present and future generations--responsibilities the Library embraces in its mission statement and strategic plan. The Library is carrying out its digitization projects in a climate in which staffing resources and federal government

funding remain steady at best and may decline in future years; this awareness has encouraged the Library to seek private-public funding partnerships and cooperative enterprises with other U.S. libraries in order to ensure support for the national digitization project.

The Library of Congress is enjoying success and wide acclaim for its retrospective digitization project, American Memory, an extensive, multiyear project to convert a critical mass of Americana from its collections and those of other selected institutions to digital formats. Prospectively, the Library is leading the way with CORDS, its Copyright Office Electronic Registration, Recordation, and Deposit System, which obtains original digital materials through the Library's traditional function as the United States Copyright agency, and the recently launched Electronic Resources Project to obtain and provide access to online materials including databases and electronic serials. Because in today's environment of steady resources and burgeoning quantities of converted and original digital materials no single library can hope to keep up on its own, the Library of Congress participates in several partnerships, such as the LC/Ameritech Digital Library Competition, the Digital Library Federation, and the Digital Libraries Initiative Phase II. For the same reason, the Library has expanded the traditional national library function of issuing standards--e.g. for cataloging--to include adopting and promoting standards for use in the digital library.

American Memory

The Library of Congress envisions the national digital library' as a distributed virtual library of converted, digitized collections and digital originals to which many American institutions contribute. The Library of Congress' contribution of converted, digitized materials is American Memory, the Web site of retrospectively digitized collections created by the Library's National Digital Library Program. There are currently 25 multimedia historical collections on American Memory, accessible via the Library of Congress Home Page or at the specific URL: <http://lcweb2.loc.gov/ammem/>

The goal of the LC National Digital Library Program is to make a critical mass of Americana maximally and freely accessible to all Americans and citizens of the world by the year 2000, the bicentennial of the founding of the Library of Congress. By the end of 1997, more than 400,000 digital files were available, making American Memory one of the largest digital library sites on the World Wide Web. The 44 members of the National Digital Library Program production staff are actively working on the digitization of another three million images. American Memory is already extremely successful. It receives an average of 10.5 million visits per month and has garnered many honors in the United States from both popular and scholarly critics:

Time magazine called it one of the best Web sites of 1996;' The New York Times listed it as an Internet Hit' in 1997; The newspaper USA Today named it a Hot Site;' It was highlighted in the July 20, 1997 issue of Parade magazine, a supplement to many Sunday newspapers in the U.S.; PC Magazine has rated included it in its Top 100 List

five times since July 1996; Britannica selected it as one of the forty best sites on the World Wide Web; Magellan Internet Guide gave it a four-star review; Net Guide Best of the Web gave it five stars; The History Channel On-line listed it as a recommended Web site for history; The Scout Report, an eminent reviewer of cutting-edge Web content, has recognized American Memory as well as other Library of Congress Internet content.

Such acclaim--all these accolades were bestowed within the span of a single fiscal year--shows the enormous American interest in Internet historical resources, a hunger for historical materials, particularly primary documents and images, that is simply insatiable. The Library has been able to launch American Memory relatively quickly thanks to the bipartisan support of both houses of Congress and the Executive Branch of the U.S. federal government and the generosity of America's executive and entrepreneurial leadership. Such success raises questions in regard to selection and audience, technology, and funding.

Selection and audience are intertwined. Library staff recognized very early, in the planning for the five-year pilot which commenced in 1990, that the national digital library invited, indeed required, the nation's great research libraries to expand their user communities in dramatic and unlooked-for ways. Libraries which traditionally defined themselves in terms of a well-defined user community consisting of resident scholars, their students, and a few guest researchers on site now realize that their digital collections can be used to support study and research by any interested user who has Internet access. Thus, while the national digital library cannot define its user community with the same ease that is possible for the traditional library, it also has the opportunity to involve people in all walks of life and all locations in study and research. The Library of Congress therefore chose to begin its National Digital Library Program by digitizing primary materials about the United States' history and culture, a topic of very broad interest to all age groups. By selecting materials mainly from the nineteenth and early twentieth centuries, the Program has avoided issues of copyright infringement while bringing primary sources to the attention of faculty, students, and the general public as never before. The Library of Congress traditionally limited its services to users older than eighteen, and this is still the rule for on-site use of the traditional collections; but recognizing that American Memory expands the Library's user base to include younger students, the NDLP has made a focused educational outreach effort. The pilot American Memory Fellows Program, funded by the Kellogg Foundation, invited selected teams of school media specialists and teachers of kindergarten through twelfth grade to spend a week at the Library in the summer of 1997 to learn about digital primary sources and their use in school curricula. In addition, private-sector funding will endow the Papamarkou Chair in Education at the Library of Congress, with the intention that the incumbent will help integrate American Memory materials into curricula for elementary and secondary schools (i.e. for students ages five through eighteen).

Technology poses ongoing threats and opportunities for the national digital library. Like other research libraries, the Library of Congress had a substantial investment in the Gopher technology, the hierarchical menu-driven interface to the Internet which has been overshadowed by the World Wide Web. Planning for American Memory must ensure that the vast digital collections can be accessed by whatever software is dominant in the future. The Library has needed to upgrade its online processing power and storage capacity to support the National Digital Library Program, and further upgrades will undoubtedly be needed as demand for access to the Library's digital collections continues to grow. Funding is undoubtedly the most critical question in determining the prospects of the national digital library, whether we are referring to its Library of Congress component or to contributions distributed throughout other institutions.

The materials selected for digitization are, by dint of being rare or old, highly vulnerable to damage in the scanning process for digitization. Therefore they require particularly careful handling and storage which calls for state of the art equipment and highly skilled staff. Given these requirements, the scanning process is unavoidably labor-intensive, and the scale of the project would call for millions of dollars even if the materials in question were not fragile or rare. In order to meet its ambitious goal of making several million images of Americana available on American Memory by the year 2000, the Library therefore proposed a budget for the NDLP of \$60 million, with \$15 million to be appropriated by the U.S. Congress and \$45 million to be raised from the private sector. To date, Congress has released \$9 million, or \$3 million a year since 1995. Private sector funding began with two extraordinarily generous founding gifts of \$5 million each from Mr. John W. Kluge and the David and Lucile Packard Foundation. Corporate and other private-sector gifts to date now exceed \$28 million, including \$6.5 million in new pledges during fiscal year 1997. During 1997 Mr. Alexander Papamarkou pledged \$2 million to fund the Library's first endowed chair, the Harissios Papamarkou Chair in Education. The Library now has a total of \$48 million in public- and private-sector pledges for support of American Memory.

The LC/Ameritech National Digital Library Competition combines a public-private funding partnership with a cooperative program between the Library of Congress and other American institutions. The international corporation Ameritech has pledged \$2 million to fund a competition for three years to enable public, research, and academic libraries, museums, historical societies, and archival institutions (excluding those in the U.S. Federal government) to create retrospective digital collections of their primary materials in American history. The applications are judged by scholars, educators, and other specialists external to the Library of Congress. The winners' digitized collections will be accessible via the American Memory Web page and will be searchable in the same way as the Library of Congress collections found on the American Memory Web site. This cooperative competition thus amplifies the American Memory collections while providing support for smaller institutions which could not digitize and maintain digital repositories with their own resources. I hope everyone in the audience will have a chance to view at least a few of the American Memory collections, all accessible via the Library of Congress Home Page (URL: <http://lcweb.loc.gov>) or directly via the

American Memory URL (<http://lcweb2.loc.gov/ammem/>). Many collections are multimedia, and the site as a whole includes documents, digitized manuscripts, films, photographs, maps, and sound recordings. One fascinating offering in popular culture is The American Variety Stage: Vaudeville and Popular Entertainment, 1870-1920, a multiformat collection that includes much of the Library's renowned material on the famed magician and escape artist' Harry Houdini. Another multiformat collection is California Gold: Northern California Folk Music, which includes digitized versions of sound recordings, photographs, drawings, and manuscripts. It is a companion collection to California As I Saw It: First-Person Narratives of California's Early Years, 1849-1900. Collections which are chiefly photographic include America's First Look into the Camera: Daguerreotype Portraits and Views, 1839-1864 and Around the World in the 1890s: Photographs from the World's Transportation Commission, 1894-1896.

American Memory collections focusing on documents to support historical research include: African American Perspectives: Pamphlets from the Daniel A.P. Murray Collections, 1818-1907; A Century of Lawmaking for a New Nation: U.S. Congressional Documents and Debates, 1774-1873; and Documents from the Continental Congress and the Constitutional Convention, 1774-1789.

Finally, the Washington/Jefferson Papers Project will make the thought and actions of the United States' first and third presidents available to a worldwide audience. Made possible through a gift of one million dollars from Reuters America, Inc., and the Reuters Foundation, the project will digitize more than 65,000 items totaling about 176,000 pages by or relating to George Washington, tagged to permit keyword searching. This part of the project is being carried out in eight phases, all to be complete by the end of 1999. The first release is now available on the American Memory Web page, including some 8,000 pages of Washington's correspondence, letter books, commonplace books, diaries, financial accounts, and military records and reports. Preparation of the Jefferson Papers database has been completed and a start has been made on preparing the microfilm reels for conversion to digital form. In addition to the digitized collections of American Memory, the retrospective portion of the Library of Congress's National Digital Library Program includes more than fifteen online exhibitions, digitized versions of three-dimensional exhibitions which have been displayed at the Library in the past several years. Here the Library moves beyond its focus on the United States' own historical and cultural legacy, mounting digitized exhibitions such as Dresden (items loaned by the Saxon State Library), Creating French Culture (treasures from the Biblioth#que nationale de France), Scrolls from the Dead Sea, Rome Reborn (items loaned by the Vatican Library), and Revelations from the Russian Archives (documents from the former Soviet Presidential Archives and the archives of the former KGB and Central Committee of the Communist Party.)

GLIN

The Global Legal Information Network, GLIN, is a cooperative venture of the Law Library of Congress and its international partners to exchange primary legal materials in the vernacular via the Internet. GLIN partner countries submit legal information published in their official gazettes, along with abstracts and subject headings they have assigned to the materials, to the Law Library of Congress for entry in the GLIN database. GLIN partner countries may retrieve images of the actual legal documents; nonmembers may search the database to retrieve citations only. Information about GLIN and its searchable database are accessible via the Library of Congress World Wide Web home page at URL: <http://lcweb.loc.gov>

GLIN currently has 12 contributing member' nations: Albania, Argentina, Brazil, Kuwait, Lithuania, Mexico, Romania, South Korea, Tunisia, Ukraine, the United States, and Uruguay. In addition, Hungary, Mauritania, Poland, and Sweden have sent personnel for training to contribute materials and are considered trained member' countries. Albania, Sweden, Tunisia, and Uruguay joined GLIN in 1997, and Kuwait became a contributing member in March 1998. The World Bank provided start-up funds to support Tunisia's participation, and the Inter-American Development Bank is considering ways to support an expansion to countries of Central America, the Andean countries, and the English-speaking countries of the Caribbean region. Since some GLIN member countries lack sufficient bandwidth in their traditional telecommunications infrastructures to transmit GLIN data, the Law Library is working with the U.S. National Aeronautics and Space Administration (NASA), Intelsat, and Comsat to test satellite capabilities for transmission of legal data.

Prospective Digital Projects: CORDS and the Electronic Resources Project

If it is to fulfill the duty of a national library in collecting, preserving, and making the digital library accessible, the Library of Congress cannot limit its role to retrospective conversion of existing materials. The production of digital originals--journals and monographs which are issued only in digital form, with no counterpart in print or hard copy--will only increase in future years. The Library, with its obligation to sustain and preserve a universal record of knowledge for the use of Congress, the American people, and future generations, must ensure that it acquires digital originals with the same comprehensiveness that it pursues the acquisition of materials in traditional formats. The Library is now beginning to consider the challenges that digital originals pose for large-scale acquisition, collection development, preservation, and service to the Library's publics. Two major efforts should be noted: CORDS and the Library of Congress Electronic Resources Project.

CORDS is the Copyright Office Electronic Registration, Recordation, and Deposit System, the automated system which will permit electronic U.S. copyright registration and deposit via the Internet. Research, development, and testing of the system occupied three years before the first electronic deposit on February 27, 1996: an unpublished

doctoral thesis from Carnegie-Mellon University in Pittsburgh, Pennsylvania. Because of the wide range of copyright applicants, testing of CORDS will continue in a series of tests from different groups of copyright owners. Phase One, testing of unpublished theses and technical reports, is still underway. Before a successful full-scale implementation of CORDS is achieved, the Library of Congress and the copyright community will need to address issues surrounding:

Digital signatures Access management Secure storage and retrieval of content
Pioneering system architecture Volume scalability (ability to scale the system up to handle quantities of material) Platform scalability (ability to make the system usable on any platform) Multiple payment mechanisms Protection of multiple data types

Full implementation of CORDS also assumes that there will be a major expansion of available Internet bandwidth and an increase in speed of transmission to permit electronic submission of hundreds of thousands of large files each year. Encouraging progress has been made in the past two years, however, particularly in the areas of access management and cooperative planning with publishers and other members of the copyright community in the United States. When fully implemented, CORDS will be the Library of Congress's first source of supply for copyrighted digital original works--a logical extension of the Library's traditional role in obtaining works for its collections via Copyright deposit.

The Library of Congress Electronic Resources Project resulted from the recognition, late in 1996, that the challenge of collecting and serving digital resources to our various publics is so complex that a coordinated, Library-wide approach was necessary to ensure our success. Over the past ten years or so, many groups within the Library have worked on meeting the various challenges posed by digital originals, but narrowly focused efforts could not successfully meet the challenge of electronic resources at the Library of Congress. Therefore, in March 1997, I issued a contract to expedite the Library of Congress Electronic Resources Project. The contractor worked with major stakeholders in the Library who formed the Electronic Resources Project Management Steering Committee to develop recommendations for dealing with acquisitions, selection, and collection development; access to electronic resources via Copyright deposit; and reference and access. (Other Library groups are responsible for considering questions of bibliographic control, preservation, and archiving and storage of digital materials.) The ERP Management Steering Committee submitted its report and recommendations in December, 1997, and Library management and stakeholders are considering their implications now. In general, the Report asks Library management to recognize that:

A senior level manager should be appointed to be responsible for full-time coordination of the Library of Congress's electronic resources efforts.

The Library must recognize that dealing with the challenges posed by electronic resources will require an effort stretching over some years.

The Library needs to develop more formal working relations with other national libraries and institutions that are beginning to cope with the challenges of electronic resources.

Staff will need specialized training to handle, acquire, and serve electronic resources to the public.

The Library will need to be aware of the ongoing complex legal questions surrounding copyright, access, and secure storage of digital resources.

Appropriate personnel have drafted a Collections Policy Statement on Electronic Resources, analogous to the usual collections policy statements issued for various formats of material and topical areas to ensure that the Library collects materials appropriately for its role as a national library. The collections policy statement is subject to comment from the Copyright Office and the Library's recommending officers and other appropriate staff. The ERP Management Steering Committee identified several potentially problematic areas:

The Library's best edition statement,' which defines the best edition' which copyright depositors are obligated to supply to the Library, must be updated to reflect the conditions posed by digital works.

Current requirements for mandatory Copyright deposits must be updated to reflect the realities of digital works.

The CORDS Working Group on Mandatory Deposit needs to include a broader range of Library of Congress stakeholders.

The Library should plan for an automated interface between CORDS and its new integrated library system, which is due to be implemented in fall 1999.

The Library's Copyright Office, public service divisions, and automation support units must ensure that the Library has an adequate digital repository for secure and accessible storage of online works, particularly those selected from CORDS.

Closer coordination is needed between Library of Congress units which produce electronic resources and units which use and serve them to the Library's publics.

The Library must negotiate license agreements with digital vendors that are as permissive as possible.

The Library needs a strategy for migrating electronic resources from format to format and medium to medium as technology evolves, to ensure that digital resources will be accessible via hardware and software likely to be commonplace in the future.

Digital Library Federation

Recognizing that cooperation among research libraries is essential if we are to grapple successfully with the challenges of the digital library, the Library of Congress was one of the founding members of the National Digital Library Federation, a group of fifteen research libraries and archives that have agreed to cooperate on defining activities leading to broadly accessible digitized materials relating to the building and dynamics of U.S. heritage and culture.' The term 'National' was dropped from the Federation's name in September 1997 to reflect the broadening of the Federation's membership and purpose to that of an international enterprise. The Library of Congress hosts and maintains the Digital Library Federation World Wide Web Home Page at URL: <http://lcweb.loc.gov/ndlf>

Digital Libraries Initiative, Phase II

The Library of Congress will participate with several other United States federal government agencies--the National Science Foundation, Defense Advanced Research Projects Agency, National Library of Medicine, National Aeronautics and Space Administration, and National Endowment for the Humanities--in the Digital Libraries Initiative, Phase II, which was announced just this February. The Library will allow researchers and institutions which receive funding from the Digital Libraries Initiative to use content from the American Memory collections for testing new technologies and tools for building, managing, and accessing collections of digital content, permitting validation of such new technologies in a user-centered environment. Thus a collection of retrospectively digitized historical source materials, which was made possible by technologies that were new a few years ago, will serve as a testbed for validation of the next generation of technology, which can in turn promote the generation of more digital resources, both original and converted.

