Building Preservation Partnerships: The Library of Congress National Digital Information Infrastructure and Preservation Program*

By William Lefurgy
Digital Initiatives Project Manager
Library of Congress Office of Strategic Initiatives

The Need for Partnership

While it has been evident for some time that management and preservation of digital information is challenging, until recently there has been little in the way of a coordinated approach to meet the challenge. The reasons for this are familiar: tools and best practices for preservation are developmental; resources available to address the issue are limited; and digital content itself continues to evolve. Absent as well has been a mechanism that links into a collaborative partnership all the various institutions and other entities that manage digital assets. But as more and more significant details about our society are recorded in bits, the need for moving beyond these limits grows.

Millennia of dependency on preserving knowledge and cultural expression are starkly threatened in a digital environment. Analog objects can survive with minimal care for centuries, but no electronic format can hope to persist more than a short while without careful (and perhaps expensive) intervention. There will be no digital equivalent of the Lascaux cave paintings, Mayan stone scripts, Dead Sea scrolls, or other kinds of rediscovered ancient knowledge. For that matter, there may not even be the digital equivalent of Emily Dickinson's poetry, which languished for only a few years in original form before its posthumous publication. Today's digital record of creativity and knowledge is at risk of wholesale loss tomorrow from obsolete software applications and file formats, degraded tape and other recording media, and other hazards wrought by rapid information technology advances. There will be little opportunity to recover anything that is untended.

Tending to digital information is, however, a complex undertaking. Digital objects have come into prominence only within the very recent past and there is little collective experience to draw upon about how best to create, manage, and preserve them. There are huge—and growing—quantities of content available at any given moment. At the same time, much of this content is constantly changing or disappearing in favor of something newer. Thorny copyright, privacy, and other rights-related issues loom over all aspects of the digital life cycle. And while entities ranging from universities to corporations to government agencies are rapidly accumulating important digital content, there is no precedent for these stakeholders working in concert to preserve significant digital information.

In 2000, Congress recognized that the nation needed an exceptional effort to prevent the loss of our digital heritage. Legislation enacted the National Digital Information Infrastructure and Preservation Program (NDIIPP) and directed the Library of Congress to determine the shape of the effort and set forth a strategy for its

implementation. Public Law 106-554 provided up to $100 million was authorized to support NDIIPP, with $75 million contingent on a dollar for dollar match from non-federal sources. Congress understood that the Library, with a core mission to make information available and useful, and to sustain and preserve a universal collection of knowledge and creativity regardless of format for current and future generations of Congress and the American people, was uniquely qualified for this assignment.

After spending nearly two years meeting with diverse stakeholder communities across the nation and studying critical aspects of the challenge, the Library issued a comprehensive plan for tackling the digital preservation problem. The plan, *[Preserving our Digital Heritage (PDF)](#)* outlined an approach to build a national network of entities committed to digital preservation and that are linked through a shared technical framework. This strategy also recognized the need for identifying best practices and supporting advanced research into tools, repositories, and overall models for digital preservation. Underlying this approach was a strong committment to partnership: given the scope and size of the digital preservation challenge, no single institution—not even the Library itself—could realistically hope to meet the challenge alone. Instead, the most effective way forward lay in harnessing the collective interest, talent, and resources of individual institutions. Collaboration is key to making partnerships work, and NDIIPP rests on a firm commitment to sharing information and building on the insights of others. The Library's role is to provide leadership in building the partnership network and also in spurring awareness of and cooperation with preservation issues among content creators, distributors, stewards and users.

Launching the First Set of Partnerships

The Library issued a Program Announcement in 2003 for proposals to start building the partnership network. Proposals could seek awards of between $500,000 and $3 million for up to three years; applicants were also required to provide matching resource contributions. The call specified that proposals provide for three outcomes. (This information is reprinted from the [Program Announcement [PDF]](#).)

1. **Partnership models for allocating collecting roles and responsibilities across collaborating institutions**
   This includes defining roles and responsibilities among and between the partners and the Library, and developing and testing cooperative collecting agreements among libraries, archives, and other institutions in the public and private sectors.

2. **Collections of at-risk digital content**
   Proposed digital collections may encompass a variety of cultural heritage materials. Among the collections of high interest to the Library at this time are those with holdings of historical and cultural materials or information from around the globe that document key social and political developments necessary to understand contemporary events of high importance to national legislators and policy-makers. Such subject areas might include American Law, Domestic Social Policy, Foreign Affairs, Defense and Trade, Government and Finance, and Science and Industry. Collections with holdings in languages other than English may be included within the scope of the project.

Formats of interest include textual, numeric, visual, audio, and geospatial, among others. Content collected under this program announcement must be accessible and transferable to the Library upon its request.

3. **Strategies and best practices for identifying, capturing, and retaining content**
   These may include but are not limited to:
   - o Definition and selection of at-risk content of long-term value - including strategies for making these definition and selection decisions (for example, historical significance, user surveys, interests of scholars, faculty and researchers, relative institutional strengths);
   - o Identification, development and testing curatorial best practices for defining and selecting complex and dynamic objects, such as Web-based objects, broadcast and streaming media, GIS materials, interactive objects;
   - o Identification and testing curatorial best practices for selecting non-English-language materials; and
   - o Identification and testing of methods and/or practices for collecting digital content (such strategies may include capturing content from the Web or other sources or receiving content directly from publishers or other creators and providers).

Each of the eight projects consists of a lead institution and at least one additional partner. A senior Library staff member serves as a Program Officer and chief liaison for each project. The Library hosted an opening kick-off meeting for all the partners in January 2005. Subsequent meetings will occur twice a year over the three-year period of performance for the initiative. Partners are also invited to participate in four so-called affinity groups, which represent significant topics that cut across all the interests of the partners. The affinity groups focus on intellectual property rights, content collection and selection, technical infrastructure, and the economics of sustaining digital preservation over the long term. Each group will identify priorities for action over the near term and will undertake a variety of activities, the results of which will be shared among the other partners. Library staff are facilitating the work of each group.

The projects represent a diverse cross section of institutions and content. What unites the projects is a dual effort to identify, get, and sustain significant material while also collaborating with the Library and the other partners to advance digital preservation methods and best practices. Each of the eight projects is outlined below. (This information is reprinted from the Library's press release for the awards. Additional information has been drawn from additional sources; Web Uniform Resource Locator addresses for these sources are noted).

Winning Proposals

California Digital Library at the University of California
Lead institution: California Digital Library at the University of California.
Partners: New York University; University of North Texas, The Libraries; and the Texas Center for Digital Knowledge.

Collaborators: San Diego Supercomputer Center; Stanford University Computer Science Department and Sun Microsystems, Inc.; New York University's Tamiment Library; Stanford University Library's Social Sciences Resource Center; Arizona State Library and Archive; and the UC libraries, including the UCLA Online Campaign Literature Archive and UC Berkeley's Institute for Government Studies Library and Institute of Industrial Relations Library.

Subject: The award is for a project to develop Web archiving tools that will be used by libraries to capture, curate and preserve collections of Web-based government and political information. This literature is a critical element of our nation's heritage and is increasingly found exclusively online, putting it at greater risk of being lost. The collections will focus on local political activities and movements, such as the California gubernatorial recall election of 2003. The issue of digital preservation has become more important in recent years, especially for government information. More than 65 percent of all government publications are now posted directly online without a print counterpart. With the half-life of government Web pages at four months, much of this information is at risk of being permanently lost. The grant will support development of infrastructure and tools that libraries and other organizations will need to build collections of selected web-based materials.

California Digital Library at the University of California announcement (MS Word document)

University of California at Santa Barbara

Lead institution: University of California at Santa Barbara.
Partner: Stanford University.
Subject: The University Libraries of UCSB and Stanford are leading the formation of the National Geospatial Digital Archive (NGDA), a collecting network for the archiving of geospatial images and data. Geospatial information has played an important role in the history of the United States. From the first colonial maps to the satellite imagery of the 21st century, cartographic information has helped define and frame our view of the United States. Project objectives include:

- Create a new national federated network committed to archiving geospatial imagery and data.
- Investigate the proper and optimal roles of such a federated archive, with consideration of distant (dark) backup and migration, directly serving content to users, vs. referring requestors back to the originators of the data for copies or assistance, active or passive quality/integrity monitoring, application of metadata, federated searching, dissemination of metadata, etc.
- Collect and archive major segments of at-risk digital geospatial data and images.
- Develop best practices for the presentation of archived digital geospatial data.
- Develop partner communication mechanisms for the project and then ongoing.
- Develop a series of policy agreements governing retention, rights management, obligations of partners, interoperability of systems, and exchange of digital objects.

National Geospatial Digital Archive announcement

Educational Broadcasting Corporation (Thirteen/WNET New York)

Lead institution: Educational Broadcasting Corporation (Thirteen/WNET New York).
Partners: WGBH Educational Foundation; Public Broadcasting Service (PBS); New York University (NYU).
Subject: Partners in this project will collaborate to establish procedures, structures and national standards necessary to preserve public television programs produced in digital formats. Thirteen and WGBH are the two largest producers of public television content in the United States. Through PBS, their productions are made available to audiences from coast-to-coast. Together, these three entities produce and distribute the majority of public television in the United States. NYU is home to one of America's most distinguished research libraries and the university recently established a graduate-level program in moving image preservation, which includes the exploration of digital technologies. The four partners will focus on such influential series as "Nature," "American Masters," "NOVA" and "Frontline," which are increasingly being produced only in digital formats, including the new high-definition standard (HDTV). Issues associated with the preservation of important corollary content, such as websites that accompany broadcasts, will also be examined.
New York University announcement

Emory University

Lead institution: Emory University.
Partners: The University of Louisville Libraries; Virginia Polytechnic Institute and State University Libraries; Florida State University Libraries; Auburn University Libraries; Georgia Institute of Technology Library and Information Center.
Subject: This project will develop a MetaArchive of Southern Digital Culture by creating a distributed digital preservation network for critical and at-risk content relative to Southern culture and history. The partners will select and preserve institutional digital archives and other institutionally relevant born-digital materials such as electronic theses and dissertations, as well as ephemeral works such as online exhibitions and cultural history Web site displays. This body of digital content includes a wide variety of subjects complementary to Library of Congress collections such as the Civil War, civil rights movement, slave narratives, Southern music, handicrafts and church history. The partner institutions of this project envision a three-year process to develop a cooperative for the preservation of at-risk digital content with a particular content focus: the culture and history of the American South. The project group members will jointly develop:

- a prioritized conspectus of at-risk digital content in this subject domain held at the partner sites;
- a body of content from the partner institutions, selected as most critically in need of preservation, harvested into a "dark archive";
- a cooperative agreement for ongoing collaboration
- a distributed preservation network infrastructure based on the LOCKSS software.

The proposed work plan for this project builds on relationships and workflows developed during previous projects of the MetaScholar Initiative and other collaborating consortia.
[MetaArchive announcement](#)

University of Illinois at Urbana-Champaign Library, Graduate School of Library and Information Science

Lead institution: University of Illinois at Urbana-Champaign Library, Graduate School of Library and Information Science.
Partners: OCLC Online Computer Library Center, Tufts University Perseus Project, Michigan State University Library, and an alliance of state library agencies from Arizona, Connecticut, Illinois, North Carolina, and Wisconsin.
Subject: This project will develop scalable software tools to facilitate selection and preservation of digital materials. In addition, it will configure and test digital repository architectures to evaluate functionality with regard to content, user and uses, interoperability, implementation of standards, and technical requirements. This undertaking will work with sound and video recordings, historical aerial photography, web-based government publications from the partner states, and primary and secondary historical materials made available by the Perseus Project. The project also provides an opportunity for information professionals with traditional library backgrounds and those with digital library expertise to work together to address these challenges. Illinois also will explore ways for libraries and repositories to share and preserve digital information existing in a wide variety of formats including Web-based government publications, historical documents and photos, sound and video recordings, Web sites and other varied digital resources that will be of historical interest to future generations.

University of Maryland Robert H. Smith School of Business

Lead institution: University of Maryland Robert H. Smith School of Business.
Partners: Center for History and New Media at George Mason University; Gallivan, Gallivan and O'Melia LLC; Snyder, Miller, Orton Lawyers LLP; and the Internet Archive.
Subject: This project will preserve at-risk digital materials from the American business culture during the early years of the commercialization of the Internet -the "Birth of the Dot Com Era," specifically 1994-2001. The materials, collected through Web portals at www.businessplanarchive.org and www.dotcomarchive.org and through direct contact with former participants in the Dot Com Era, will be of incalculable historical value to Americans eager to make sense of this remarkable period of venture creation. Content associated with this project includes business plans, marketing plans, technical plans, venture presentations, and other business documents from more than 2,000 failed and successful Internet start-ups.
[University of Maryland Robert H. Smith School of Business announcement](#)

University of Michigan Inter-university Consortium for Political and Social Research

Lead institution: University of Michigan Inter-university Consortium for Political and Social Research.

Partners: The Roper Center for Public Opinion Research at the University of Connecticut; the Howard W. Odum Institute for Research in Social Science at the University of North Carolina-Chapel Hill; the Henry A. Murray Research Archive and the Harvard-MIT Data Center (both members of the Institute for Quantitative Social Science at Harvard University); and the Electronic and Special Media Records Service Division of the National Archives and Records Administration.

Subject: These institutions will create a partnership to identify, acquire and preserve data used in the study of social science to ensure that future generations of Americans have access to this vital digital material that will allow them to understand their nation, its social organization and its policies and politics. Surveys have done more than predict the outcomes of elections or tell us when presidents gain or lose popularity. They inform us about aging, health and health care, race relations, women's rights, employment, and family life—the full story of the social and cultural tapestry that makes up our nation. They provide the data necessary for sound, empirically based policy-making. Yet a huge quantity of this data is missing or at-risk. Examples of data that will be preserved by this project include opinion polls, voting records, large-scale surveys on family growth and income, and focused studies on effects of events such as factory closings or the need to care for aging parents. Together the partners will build a shared catalog, adopt a common standard for describing survey data and develop strategies for ensuring that the data remains available for analysis.

[University of Michigan Inter-university Consortium for Political and Social Research announcement](#)

North Carolina State University Libraries

Lead institution: North Carolina State University Libraries.
Partner: North Carolina Center for Geographic Information & Analysis.
Subject: The project will collect and preserve digital geospatial data resources, including digitized maps, from state and local government agencies in North Carolina. Geospatial data are created by a wide range of state and local agencies for use in applications such as tax assessment, transportation planning, hazard analysis, health planning, political redistricting, homeland security and utilities management. The geospatial resources targeted by the NCSU Libraries' project include digitized maps, geographic information systems (GIS) data sets, and remote sensing data resources such as digital aerial photography. A wide range of state and local agencies create these forms of data for use in tax assessment, transportation planning, hazard analysis, health planning, political redistricting, homeland security and utilities management. State and local agencies frequently offer more detailed and up-to-date geospatial data than federal agencies. However, these entities are by definition decentralized, and their dissemination practices focus almost exclusively on providing access to the most current data available, rather than any older versions. Although this project will focus solely on North Carolina, it is expected to serve as a demonstration project for other states.

[North Carolina State University Libraries](#)

Seeking to Add States and Territories to the NDIIPP Network

The Library is presently seeking to expand the network of preservation partners beyond those noted above through an exploratory initiative with all U.S. states and territories for

preserving significant state and local government information in digital form. State libraries and archives typically have broad responsibility for preserving and providing public access to state and local government information of enduring value, and are important components of a national preservation network. (The information below is reprinted from the Library's [announcement](#) for the initiative.)

State and local governments are creating vast amounts of information solely in digital form, including land data, school records, official publications and court records. Much of this material is of permanent legal, legislative, or cultural value, yet is at risk because of fragile media, technological obsolescence, or other hazards. State libraries, archives, and other state and local institutions face complex barriers in developing an effective strategy to meet this challenge.

During 2005, the Library will sponsor collaborative workshops to help states identify their needs and priorities for digital preservation. A toolkit will be made available to each state to enable this analysis. The Library will support attendance by representatives from state libraries, archives, and other stakeholder organizations at workshops to first refine and then to implement the toolkit. This activity is intended to help states assess their preservation capabilities and target content for collection. Findings and results from participating states will be assembled into a final report, which will provide direction for the Library in building out its preservation network with the states and will also impart important conceptual and practical knowledge to all communities with an interest in digital preservation.

The Center for Technology in Government, a leading digital government research center at the University at Albany, State University of New York, will assist the Library in this work. The Institute of Museum and Library Services (IMLS), the primary source of federal support for the nation's libraries, will also be a partner. Along with experience in supporting collaborative projects to manage, preserve and provide digital access to collections, IMLS has significant expertise administering state based library service programs that encourage planning and evaluation.

As NDIIPP continues to move forward, the Library anticipates continuing to add partners to the national preservation network. Over time, the intent is for partners to define and undertake specific roles and responsibilities in connection with this participation. The Library will continue to play a leadership role in facilitating network activities and in advancing digital preservation knowledge and practice.