Preserving
Our Digital Heritage:

The National Digital Information Infrastructure and Preservation Program 2010 Report

A Collaborative Initiative of The Library of Congress

Digital PRESERVATION

National Digital Information Infrastructure & Preservation Program

First issued, January 2011

Contents


JAMES MADISON
1751–1836

I. Executive Summary

A. What Is at Stake

Digital information technologies are the foundation of our nation's knowledge capital. They are key to economic prosperity and crucial for maintaining the United States' global competitiveness. They deepen citizens' engagement with democracy and enrich their daily lives by enabling new forms of communication and creativity. Now, more easily than ever, citizens can have access to the information they need to govern themselves and engage in lifelong learning. But the great promise of new information technologies also brings unprecedented challenges because digital information is inherently fragile. How does our nation ensure that the knowledge and wisdom endowed to us by generations of Americans, continuously collected and preserved since the founding of the Library of Congress (the Library) in 1800, will continue to grow?

What is at stake is the loss of data representing billions of dollars of investment in new information technology, new scientific discoveries, and new information upon which our economic prosperity and national security depend. Also at stake is the transmission of ideas, knowledge, and the American people's legacy of creativity to future generations. Preserving digital content is as important today as preserving the records of the Founding Generation was in 1800.

In 2000, grasping the seriousness of this situation, Congress charged the Library to create the National Digital Information Infrastructure and Preservation Program (NDIIPP) to develop a strategy to meet the challenge of digital preservation. A network of institutions committed to preserving the nation's digital heritage is now poised to carry forth this strategy. This report summarizes the Program's accomplishments to date and outlines its next steps.

B. NDIIPP Legislation

In 2000, Congress authorized $100 million to be directed to the Library of Congress for "a major undertaking to develop standards and a nationwide collecting strategy to build a national repository of digital materials" (P.L. 106-554). The digital era presents clear challenges for such an undertaking: the escalating scale of data creation, the globalization of information exchange, and the immaturity of standards and best practices, among many.

In response to this congressional charge, the Library undertook a process of consultation with a variety of stakeholders from the public and private sectors that resulted in a long-term plan to ensure that content of value to the nation will be available for present and future users. On the basis of these consultations, in 2003, the Library proposed and

A popular Government, without popular information, or the means of acquiring it, is but a Prologue to a Farce or a Tragedy; or, perhaps both. Knowledge will forever govern ignorance: And a people who mean to be their own Governors, must arm themselves with the power which knowledge gives.

James Madison, 1822

Statue of James Madison by Walter K. Hancock. Photograph by Carol M. Highsmith.

Saving digital records is vital to continuity of state governments

Challenge: *State government legislative digital information, including bills, acts, mandated reports and house and senate journals, are increasingly at-risk due to technological obsolescence.*

Solution: *The Minnesota Historical Society, in partnership with the nine other state governments, is developing model practices for preserving legislative information resources. The project has implemented a trustworthy information management system and is testing the capacity of different states to adopt the system for preserving state records for current and future law and policy makers.*

Congress approved a plan calling for a distributed, networked stewardship capacity to be developed and maintained under the leadership of the Library of Congress. A multiphased implementation of the plan has effectively leveraged the strengths of existing organizations and led to more communities joining this national effort for sustainable solutions to digital preservation.

The Program's plan grew from the recognition that to create a successful and sustainable digital preservation infrastructure, NDIIPP should focus on four major goals:

1. **Stewardship network:** Develop a growing national preservation network.
2. **National digital collection:** Develop a content collection plan that will seed a national collection and preserve important at-risk content.
3. **Technical infrastructure:** Build a shared technical platform for networked preservation.
4. **Public policy**: Develop recommendations to address copyright issues and to create a legal and regulatory environment that both encourages incentives and eliminates disincentives to preservation.

To date, the Library has recruited more than 185 digital preservation partners in more than 44 states and 25 nations to execute a multiphased plan to collect and preserve a broad spectrum of high-value digital content, with special attention to the needs of the public policy, education and research, and cultural heritage communities.

C. Key Outcomes and Findings

NDIIPP organized its initiatives and investments around these four strategic goals identified in the plan. The key outcomes and findings associated with each goal are summarized below. More detail and full reports from each initiative are available on the Program's website (www.digitalpreservation.gov).

1. Stewardship Network

Building distributed, networked capacity for digital preservation and long-term stewardship is a complex undertaking. It demands that action be taken before all the critical factors involved are fully understood. Supporting the growth of that networked capacity is equally complex and requires sustained, dedicated coordination. Through building a network, NDIIPP engages a diverse set of preservation partners. The Library serves as the central node in this network of networks, articulating and coordinating roles and responsibilities. Through regular meetings, strategy sessions, and outreach projects, partners share outcomes and lessons learned from their local preservation programs.

Key Findings

- Preservation is a societal good undertaken by committed organizations. These organizations are motivated by their own interests and incentives, and as a group they also act on behalf of the public interest.
- Each participating institution brings to the network its own resources, interests, and organizational culture. Under the auspices of a neutral convener and honest broker, natural networks emerge over time through participation in shared activities, problem solving, and meetings.
- As natural networks form, the larger network becomes more complex, but also stronger and better able to withstand stresses and strains.
- The Library has proven to be a reliable and trustworthy broker.

2. National Digital Collection

Geospatial data, web-based content, digital text and images, broadcast television, and audiovisual content are examples of materials having significant value to current and future users. Collecting and preserving such materials is highly complex. Because of its volume and dynamic nature, digital content requires coordinated action by diverse stakeholders with targeted approaches and skills. NDIIPP partners and the Library of Congress have coordinated, selected, collected, and preserved content in areas that provide the greatest value to the public policy, education and research, and cultural heritage communities. The work to date informs an approach to the future establishment of a national digital collection.

Key Findings

- Digital content is ephemeral. To ensure that it will be available in the future, it must be collected and preserved now.
- The scale of content production will continue to outpace the ability of any single institution to collect and preserve a national digital collection, requiring an alliance of several organizations committed to stewardship.
- The value of content grows when it can be shared across content domains, and a distributed approach to collecting has proven an effective strategy to achieve breadth and depth of content.

3. Technical Infrastructure

In the first year of planning for NDIIPP, a group of technology experts proposed a three-layer architecture model for digital preservation. The bottom layer stores and maintains the data. The middle layer provides services for content description and management, and is the layer associated with libraries and archives. The top layer, the access layer, provides services to view and use the content. NDIIPP partners have tested and validated a distributed approach to preservation, with clearly

Public safety and environmental policy planning rely on digital mapping information

Challenge: *Geospatial mapping information is critical for states to provide effective services to their citizens, but preserving older data to enable "change over time" analysis is difficult.*

Solution: *The Geospatial Multistate Archive and Preservation Partnership is expanding the capacity of state governments to provide long-term access to geospatial data. By pairing state government archives staff with geospatial experts, GeoMapp is addressing statewide issues in formulating best practice guidelines for geospatial data stewards in all 50 states.*

Archiving digital primary source materials sharpens critical and creative thinking for K-12 students

Challenge: *Young people often cannot discern valuable and reliable digital information available on the web.*

Solution: *Over a dozen primary and secondary classes participated in a year-long class activity to learn how web sites have lasting value as historical documents. They identified and described dozens of collections ranging across topics such as local history and news, business and commerce, international events, and sports and entertainment. Their teachers noted the high value of the project for the development of the students' critical analysis skills within an authentic learning experience.*

articulated roles among partners. Tools, services, and domain-specific standards have been developed, shared, and implemented throughout the preservation network.

Key Findings

- The diversity, complexity, and volume of digital objects and formats require that stewardship of digital content be shared across many organizations.
- Technical work across NDIIPP partnerships has affirmed the value of open development of tools.
- Maintaining redundant copies, distributed geographically, organizationally, and across diverse systems, is a better means of ensuring long-term security than keeping only one copy or holding copies in a single environment.
- Developing several approaches to the same problem is preferable to relying on a single approach, which risks data corruption or irretrievable loss should that approach fail.

4. Public Policy

Building an infrastructure for the stewardship network requires fostering a public policy environment that is conducive to preservation. Putting in place legal and incentive structures is as important to preservation success as collecting the content itself. Experts in copyright law and in the creation and preservation of digital content made recommendations to update U.S. copyright law to enable and encourage digital preservation. NDIIPP also co-sponsored a comparative analysis of international copyright laws that affect preservation activities globally.

Key Findings

- Both the copyright and regulatory environments need to be updated for the digital era; currently they discourage preservation best practices or even make them illegal.
- Privately owned digital content often has high historical and cultural value for the public and such value should be protected. When owners have no further use for their content, they should be given incentives to entrust it to a stewardship organization.
- There are few, if any, economic incentives for preservation. Because preservation and stewardship of a national collection serve the public interest, creating economic incentives is a priority.

D. Moving Forward

The Library's strategy for expanding and sustaining a national digital collection of high-value content for the public policy, education and research, and cultural heritage sectors is to build a *distributed stewardship network for a national collection*. Through collaboration, the network will achieve economies of scale, lower costs to network participants, and bring under stewardship a large and diverse body of

valuable content. Building on the foundation created by NDIIPP, the Library and its partners are moving into the next phase: establishing the infrastructure necessary to provide sustainable, long-term access to a national digital collection.

1. Chartering the National Digital Stewardship Alliance

To foster and sustain the stewardship network, the Library is formalizing the National Digital Stewardship Alliance. Each Alliance member commits to collection, provision, or curation of content for the Alliance; or to provision of services for the Alliance, including storage, infrastructure, tools, software or hardware, or cataloging.

The functions of the National Digital Stewardship Alliance are sixfold:

1. Build and sustain a national digital collection—a rich and diverse record of the national experience and information vital to the nation.
2. Support collaboration among Alliance member organizations and affiliates.
3. Identify, develop, and maintain services, standards, best practices, and sustainable business models of direct benefit to the Alliance.
4. Facilitate the development of a networked technical infrastructure, tools, and storage capacity for the Alliance.
5. Ensure the advancement of digital preservation science and technologies by encouraging research and development.
6. Foster an environment that supports long-term access to digital content through education, advocacy, and encouragement of sound public information policies.

2. Developing a Framework for a National Digital Collection

NDIIPP works with more than 185 partners from 44 states and 25 countries to identify, collect, and preserve at-risk digital content. The knowledge gained through the early Program activities has informed NDIIPP's approach to developing a national digital collection that serves not only Congress and government agencies but also the American people and an international audience. Working collaboratively, Alliance partners will continue to build and sustain a national digital collection that covers a broad scope of education and research, cultural heritage, and public policy subject areas. From 2010 to 2013, collection-building activities will focus on:

- Government, politics, and law.
- Maps and geography.
- Gews, media, and journalism.

3. Strengthening and Enabling Public-Private Partnerships

The Library will explore the costs and benefits of creating an independent entity that could provide a flexible means by which

Preserving research data maximizes the federal investment

Challenge: *Despite the recognition by the National Science Foundation and the National Institutes of Health that data sharing maximizes the impact of research dollars, a 2008 inventory of over 1,600 federally-funded social science research projects revealed that at least 25% of the data had been lost and was not available to researchers.*

Solution: *The Data Preservation Alliance for the Social Sciences (Data-PASS) has rescued and is preserving over 800 data collections that were identified as being at risk. The continuing collaboration of ICPSR, Odum Institute, Murray Archive and Roper Center developed an infrastructure for the distributed preservation and access of social science data sets.*

Innovation in a global knowledge economy is driven by digital information

Challenge: *New ventures are an essential pillar of the nation's innovation ecosystem, yet born-digital business records of new ventures, especially failed ones, are rarely preserved.*

Solution: *The Business Plan Archive—established at the Robert H. Smith School of Business at the University of Maryland at College Park—hosts business plans and related digital business records from more than 3,000 new ventures created during the Dot Com boom of the late 1990s. Cost savings are realized by designing the repository to hold the records of a population of firms, rather than building firm-by-firm collections.*

to fund innovation and develop shared tools and services for the Alliance, especially its federal members, by enabling robust public-private partnerships. Any tools and services developed would be designed to meet the infrastructure needs of preservation and would be deployed throughout the Alliance, allowing partners with complicated service-contracting requirements or limited resources access to cutting-edge technology and research. The core functions of the entity would include:

- sponsoring technical research and development;
- serving as a broker for core infrastructure services such as storage; and
- raising funds from private sources and brokering public-private partnerships.

4. Fostering a Public Policy Environment Conducive to Digital Preservation

The Library undertook a broad-gauged review of the public policy environment for digital preservation and access. The primary finding of this review is that there are too few incentives, and many disincentives, to preserve digital content in the public interest.

The Library will explore three major areas that can be addressed through federal policy.

1. Work with the U. S. Copyright Office and Congress to pursue Section 108 Study Group Report recommendations for updating copyright law for digital preservation.[1] Key recommendations of the study group are as follows:
   - Make museums eligible for section 108 exceptions.
   - Increase the number of digital copies that libraries, archives, and museums are allowed to make for preservation.
   - Permit libraries, archives, and museums to make preservation copies of at-risk works prior to damage or loss.
   - Allow libraries, archives, and museums to capture and preserve publicly available online content.
2. Convene a panel of national experts to explore the creation of tax-related incentives for digital preservation. These could include tax credits for a portion of the costs of preservation or tax incentives that would encourage individuals and corporations to donate digital cultural assets.
3. Create a pilot project in which the Library of Congress may explore with copyright owners the digital display and/or dissemination of certain works that are in its collection and protected by copyright (for example, text, audio, visual or audio visual works) under terms to be mutually agreed upon.

[1] The Section 108 Study Group Report. An independent report sponsored by the United States Copyright Office and the National Digital Information Infrastructure and Preservation Program of the Library of Congress. March 2008. Available at http://www.section108.gov.

4. Form a study group to investigate ways of reducing barriers to preserving historically significant business, corporate, and privileged records that are held privately but if preserved would provide a significant public good. A possible solution is to create closed archives, formed by congressional charter or authorization, for business and confidential records.

E. Securing Knowledge for the Future

Much has changed, been learned, and been achieved since NDIIPP legislation was passed in 2000. The next phase of NDIIPP will ensure growth of the National Digital Stewardship Alliance, with a goal of establishing partnerships in all 50 states. In addition, a structure will be put in place to support the development and deployment of tools and services across the Alliance, so that all partners have access to cutting-edge technologies. Building on collections already preserved by the NDIIPP partners and the Library's collection areas, the Alliance will establish a national digital collection that will ensure long-term access to digital resources of high value to Congress and the American people. NDIIPP will also work toward creating a public policy environment that supports best practices and provides incentives to preserve for the public good.

Much is at stake if we do not act now. The nation's educational system, economic security, energy infrastructure, and the continuing creativity and innovation that assure the people's well-being all depend on a secure knowledge base. What is at stake is no less than the ability to show our children and grandchildren where we have come from, to help them understand how our democracy grows, and to empower them with the knowledge and wisdom to make the difficult choices that the Founders well understood would confront us as a free people.

State libraries and archives leverage shared expertise and systems in times of fiscal constraints to support their missions

Challenge: Applying existing digital preservation tools, services and processes is difficult for state archives and libraries when they do not have a shared collaborative approach.

Solution: Seven state governments are adapting technology for preservation through the Persistent Digital Archives and Library System project. The project developed software components and applications to automate digital preservation tasks to demonstrate an inexpensive "digital stacks" that can preserve the authenticity and integrity of the collections. In parallel, the Washington State Digital Archives led nine other states to develop a centralized regional repository for state and local digital information. By implementing a cost-effective interstate technological archiving system, the project demonstrated a scalable approach to preserving and making available at-risk digital government information.

II. Digital Information Needs of the Nation

Thomas Jefferson wrote the words on the right shortly before Congress lost its library when the U.S. Capitol was set afire in the War of 1812. Upon hearing of the destruction, Jefferson offered to sell his personal collection of books, charts, and primary records—the largest library in the New World—to restock the congressional library, nearly doubling its size. Congress gratefully accepted Jefferson's offer.

Jefferson's legacy has grown into the largest and most comprehensive collection of knowledge in the world—a collection that serves the people's government, documents our history and creativity, and seeds the innovation of the nation. Today, nearly two centuries after the devastating blaze that all but destroyed the original Library of Congress collection, we stand at a similar point of inflection in the history of the nation's library. New information technologies have spurred an exponential growth in knowledge and have generated vital new information that Congress consults in the course of business.

Digital information technologies can be a boon to the country's economy. They deepen citizen engagement with democracy and enrich our lives by enabling new forms of communication and creativity. Now, more easily than ever before, citizens can have access to the information they need to govern themselves and engage in lifelong learning. But the great promise of new information technologies also brings unprecedented challenges because digital information is inherently fragile. How does the nation ensure that the knowledge and wisdom endowed to us by generations of Americans, continuously collected and preserved since the founding of the Library in 1800, will continue to grow?

What is at stake today is not just the loss of data representing billions of dollars of investment in new information technology, new scientific discoveries, and new information upon which our economic prosperity and national security depend. What is at stake is the transmission of ideas, knowledge, and the American people's legacy of creativity to future generations. Preserving digital content is as important today as preserving the records of the Founding Generation was in Jefferson's lifetime. Grasping the seriousness of this situation, Congress charged the Library in 2000 to create the National Digital Information Infrastructure and Preservation Program (NDIIPP) and to assign it the responsibility of developing a strategy to meet the challenge of collecting and preserving high-value digital content.

The Program has completed its planning, development, and early implementation phases. It is now poised to build on this

He who receives an idea from me, receives instruction himself without lessening mine; as he who lights his taper at mine, receives light without darkening me.

Thomas Jefferson, 1813

The Thomas Jefferson Library, the Jefferson Building, Library of Congress. Photography by David Sharpe.

foundational work to continue constructing the nation's digital infrastructure to serve the information needs of the public policy, education and research, and cultural heritage communities. This report summarizes the Program's accomplishments to date and plans for next steps.

A. The Preservation Challenge for the Nation

Imagine it is 2059.

education

A high school class is learning about citizen participation in representative government, and one group of students is doing a research project on how the primary and caucus system worked in 2008. Searching online, they find multiple references to citizens organizing get-out-the-vote campaigns for their candidate of choice. Knowing that such efforts were spread virally, they search online for e-mail listservs, blogs, and social software sites but find no information is available. All they find are the candidates' official sites. No citizen network sites have been preserved, even though such sites were ubiquitous on the public web at the time.

economic security

An economist notices unusual volatility in one sector of the bond market. She looks for when similar instability was last noted and finds it was between 2008 and 2010, when a number of investment banks failed. She looks for the records of the major banks of the time—many of them defunct—to investigate the internal decision-making processes that led to the series of fatal decisions. She learns that all such records are private: the banks either destroyed them when they closed or the records are under seal. She has no way to understand what happened at that time, and no way to learn lessons from history.

national energy infrastructure

Geologists are investigating renewable wave-energy sources in the Gulf of Mexico and need to predict the effects of hurricanes on the Mississippi Delta and the inland levee systems. They find that there are decades-long gaps in the aerial surveys of the delta in some important parishes and municipalities, making it impossible to develop reliable analytical models. Having seen this problem in numerous projects, they know that they have no choice but to abandon this project.

creativity

A filmmaker learns about an independently produced film by an influential filmmaker from the first decade of the twenty-first century. When he tries to find the film to view, he learns that it was never preserved because the original filmmaker did not have the time, expertise, or resources to do it on his own. Early studio films by the same filmmaker are known to exist, but the researcher is not able to find them either. The studio was

eventually bought by a holding company that decided the films were not profitable and the costs of preserving the old formats were too high. The films were never transferred to a library for preservation.

cultural heritage

A biographer of a famous writer and poet laureate from the turn of the twenty-first century is looking for blogs that wrote about her. He finds a site that claims to have archives of contemporary poetry blogs, but when he searches the source, he finds that none of the information can be read in current software.

Key Terms

Content sectors are the information producers, providers, and users who cluster around content with a common profile, such as geospatial, audio, or web content. Content sectors usually have shared interests in developing standards and best practices for the creation and description of that content, and often collaborate to develop tools and services that are of special utility for that content. The content sectors of particular interest to NDIIPP are geospatial, text and image, audiovisual, broadcast television, and web-based content. NDIIPP gives priority to three user communities with particular interest in these content sectors: public policy, research and education, and cultural heritage.

Distributed digital preservation occurs when functions such as storing and maintaining data, description and management of digital collections, and access services are performed in coordinated ways across several institutions. Distributed digital preservation has several notable advantages over preservation undertaken by a single institution, wherein one organization is responsible for every aspect of content storage, description, collection management, and access services. Distributed digital preservation is able to achieve economies of scale, for example. It is premised on the leveraging of expertise across a variety of technical and content domain experts. As a result, in such a system there is no single point of failure. That said, to be successful this approach requires explicit commitments by participating institutions with respect to roles and responsibilities.

Stewardship is the full range of activities that support the collection, preservation, description, interpretation, display, and dissemination of cultural, artistic, intellectual, and research resources.

Stewardship organizations are those with a primary mission of ensuring the stewardship of the resources under their care over the course of two or more generations of users. Such organizations are usually, though not exclusively, nonprofit organizations, though they may offer services that can recover some of the expenditures of providing stewardship. Examples of stewardship organizations include research libraries, archives, museums, historical societies, data archives, and research institutions.

what is at stake

These are some examples of what may lie in store if we do not act now to build a national capacity for long-term preservation and stewardship of digital content. As the scenarios illustrate, digital content can be at risk for many reasons—including the rapid pace of development of hardware and software, outdated provisions for preservation in the copyright law, and organizational failure to maintain important information assets and records. These problems affect everyone, from corporations to individuals, from local to national organizations. Moreover, they are global in scope.

What Is At-Risk Content?

Much digital content is at risk of loss because there is little robust and secure infrastructure for its collection, management, and preservation. In building a national preservation strategy, our nation must address four categories of risk.

Technological Risks

- *Hardware and software*, both proprietary and open source, can be a challenge to maintain and keep current.
- *Content formats* can be complex and fragile. They are often not well documented and frequently become obsolete.
- *Lifecycle management risks* such as data migration, file degradation ("bit rot"), or unauthorized use can make content unusable.

Legal and Policy Risks

- *Copyright* laws are unclear about libraries' rights to create and keep preservation copies.
- *Privacy claims* can prohibit collection and documentation of content.
- *Sarbanes-Oxley regulations* can induce content owners to destroy historically valuable documents.
- *The law does not recognize public value in preserving digital content.* There are few policy incentives for concerned parties to preserve content in the public interest.

Content Risks

- *The volume or complexity* of content makes it difficult to collect comprehensively.
- *Insufficient description* of content makes it challenging to discover or retrieve it for use.

Organizational Risks

- *Insufficient resources* to maintain information can lead to content loss.
- *Lines of authority and responsibility* for maintaining digital content are often not aligned with the demands of such content.
- *Insufficient skilled personnel* can prevent even routine best practices from being implemented.

As the Government Accountability Office (GAO) reports, our energy, transportation, and water infrastructures are in serious need of updating and repair.[2] But to be productive, these crucial sectors themselves require a robust, secure, and reliable information infrastructure to manage the risks and capitalize on the efficiencies of digital information in today's complex knowledge economy. More important, the sectors must manage critical data for decades to come.

information infrastructure needed

The fundamental risk to the future of our nation's knowledge capital and our success in a global economy is the lack of a distributed, networked, cost-efficient content stewardship infrastructure designed to perform in a dynamic environment.

Once a book is printed, it can rest on the shelf for more than a century and remain readable. In

[2] The Upcoming Transition: GAO's Efforts to Assist the 111th Congress and the Next Administration. Statement of Gene L. Dodaro, Acting Comptroller General of the United States. Testimony before the Subcommittee on Government Management, Organization, and Procurement, Committee on Oversight and Government Reform, U.S. House of Representatives. September 24, 2008. Available at http://www.gao.gov/new.items/d081174t.pdf.

contrast, digital content requires active management throughout its entire period of use. Each step in the lifecycle of digital content, from creation and distribution to selection, description, and preservation, requires choices—choices that will determine its long-term availability. The choice of widely used, well-documented formats by the creator, the deposit of content in a stewardship organization, and the appropriately resourced management of content by skilled personnel are critical for enabling access over time. Good lifecycle management not only ensures that content will remain available over time but also achieves economies of scale and obviates the need for expensive recovery efforts for endangered data.

B. Congressional Charge to the Library

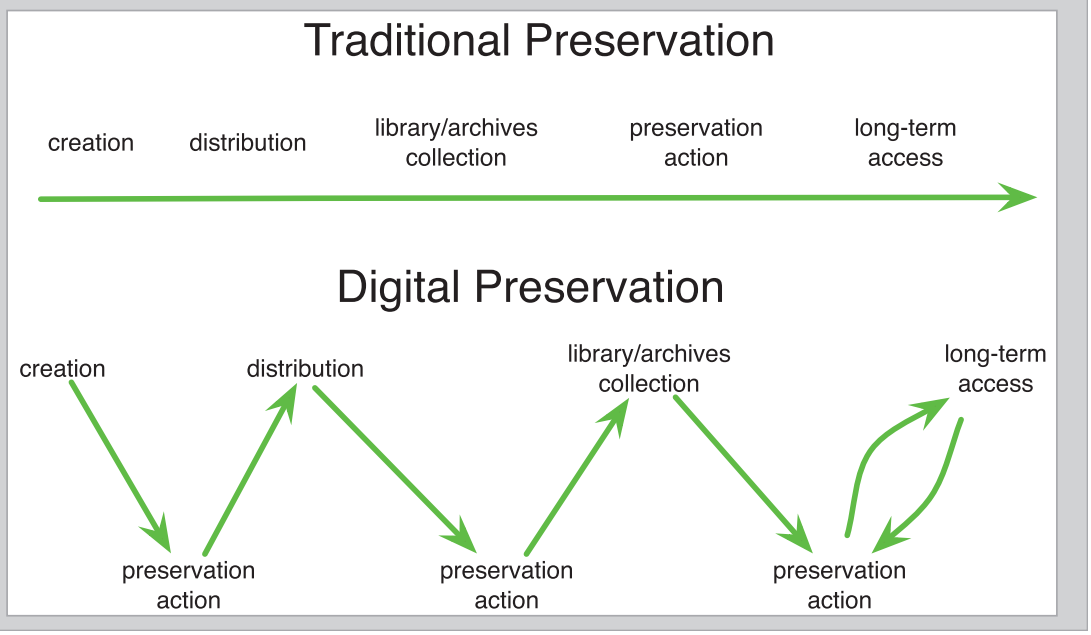
In 2000, Congress responded to the urgent challenge of digital preservation by authorizing 100 million to be directed to the Library of Congress for "a major undertaking to develop standards and a nationwide collecting strategy to build a national repository of digital materials" (P.L. 106-554). The digital era presents clear challenges for such an undertaking—the escalating scale of data creation, the globalization of information exchange, and the immaturity of standards and best practices, among many.

digital content requires active management

Other nations, notably those of the European Union, understand

Fig. 1. Traditional Preservation Versus Digital Preservation

Digital content requires active management throughout its entire period of use.



this challenge and are taking action. They have declared long-term digital stewardship a top priority in their drive to compete globally in science, technology, commerce and trade, and to foster social well-being.[3] NDIIPP is helping build our nation's capacity to maintain international leadership in these areas.

mandate to meet the challenge

In addition to asking the Library to tackle the long-term challenge of building distributed stewardship capacity, Congress charged the Library to "collect or preserve essential digital information which otherwise would be uncollectible . . ." The Library has worked with major content creators and stewardship organizations to identify digital content of high value to the nation. Together, they have secured this content for use by members of Congress and the Library's users. A special objective of the program is to capture especially vulnerable materials that will vanish without aggressive action.

When NDIIPP began in 2000, few people understood the fragility of digital content. Even fewer predicted the explosive growth of web-based information, how the web would enable unprecedented public engagement with national political events, and the types of innovation and creativity that digital technology would foster. Only a handful of people understood how important it was to begin securing digital

NDIIPP Legislation

P.L. 106-554

The congressional charge to NDIIPP calls for short-term action and a long-range plan.

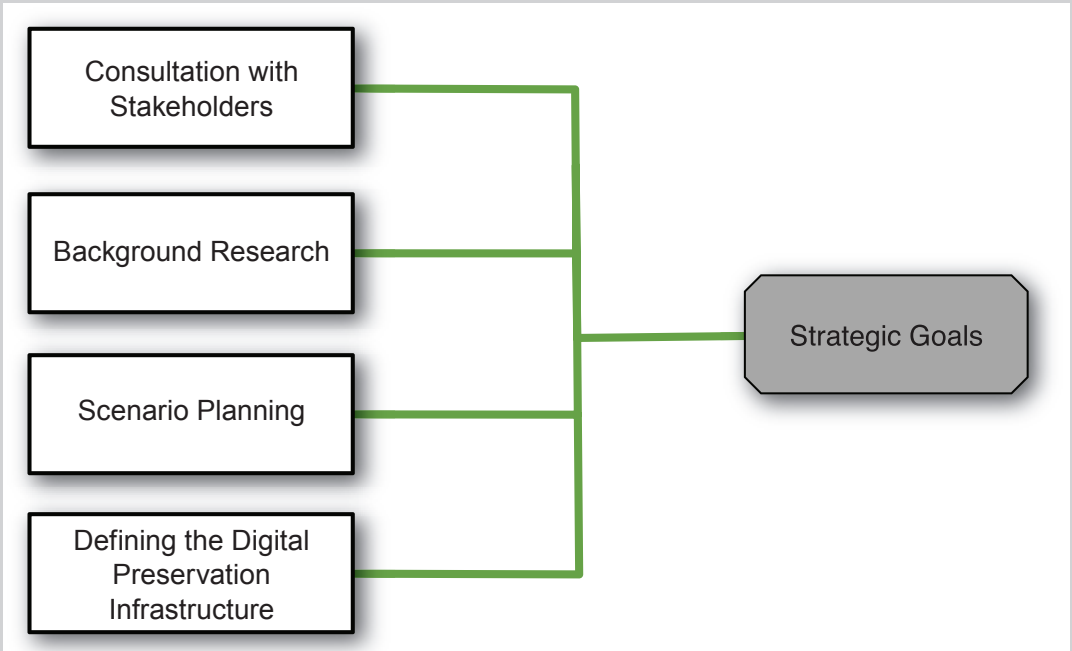
This program is a major undertaking to develop standards and a nationwide collecting strategy to build a national repository of digital materials. . . .

The Library is directed to develop a phased implementation plan for this program jointly with Federal entities with expertise in telecommunications technology and electronic commerce policy and with participation of other Federal and non-Federal entities. . . .

The overall plan should set forth a strategy for the Library of Congress, in collaboration with other Federal and non-Federal entities, to identify a national network of libraries and other organizations with responsibilities for collecting digital materials that will provide access to and maintain those materials. In addition to developing this strategy, the plan shall set forth, in concert with the Copyright Office, the policies, protocols, and strategies for the long-term preservation of such materials, including the technological infrastructure required at the Library of Congress.

[3] European Commission. Commission Recommendation on the digitization and online accessibility of cultural material and digital preservation. August 25, 2006. Available at http://ec.europa.eu/information_society/newsroom/cf/itemlongdetail.cfm?item_id=2782.

Fig. 2. NDIIPP Planning Process
The NDIIPP planning process comprised four key activities.



information for present and future generations.

C. National Plan for Preservation

To raise awareness about digital preservation and to gauge the readiness of content sectors to respond, in 2001–2002 the Library undertook a consultative planning process that involved major stakeholders, including the creative industries, libraries and archives, institutions of higher education, state and local entities, and legal and technical experts. The Library also conducted extensive research into the preservation problem; worked with leading technologists to define the main requirements for technical infrastructure for preservation; and did scenario planning to test working assumptions and anticipate disruptions in technology, the economy, and national security.

why a network?

On the basis of its findings, the Library proposed, and Congress approved, a plan calling for a distributed, networked stewardship capacity to be developed and maintained under the leadership of the Library.

Given the scope of the challenges, the Library took early actions to mobilize important

stakeholders as a network. Without a broad-based, distributed approach, the goals would seem "too big, too intractable," as the Library heard many times. Concerned stakeholders, coming from many different content and technical communities, felt isolated. They believed that collective action was the only way to achieve scale, but in 2000 they had little in common beyond a need to preserve their own content. Acting as honest broker and neutral convener, the Library brought them together to begin the work. An early and critical step in executing the program's strategy for preservation was to build and nurture a networked community.

National Digital Strategy Advisory Board

To guide the program's development, the Library formed a National Digital Strategy Advisory Board comprising members of key federal agencies, information organizations, libraries, archives, and content creators. Because the network works within a global context, board members include individuals from national libraries other than the Library of Congress. A list of Advisory Board members appears in Appendix A. The board advises Library of Congress managers on the development and operation of NDIIPP and sponsors the Federal Digitization Guideline Working Groups, whose mission is to develop best practices and standards for preservation across major federal agencies.

D. Implementing the Plan

The Library invested in testing approaches, developing standards, and adopting best practices through collaborative networks incubated by the program. By enabling stakeholders to take early action and learn by doing, NDIIPP has leveraged the efforts of content communities already undertaking preservation.

take early action, learn by doing

NDIIPP is developing a distributed technical infrastructure to support digital preservation; fostering collaboration between federal and state sectors; addressing

Digital Preservation Environment 2000

In 2000, digital content was growing, but few understood the need to preserve it.

- Growth of use and content on the web
- Growth of digital library services
- Growth of digitization
- Limited awareness of digital preservation
- Few technical solutions
- Limited technical expertise among preservation professionals
- Lack of cooperation and trust among content stakeholders
- No provisions for digital preservation in copyright law
- Young and unproven models for economic sustainability

information policies that impede digital preservation; and leading a national effort to educate the public and promulgate best practices for digital preservation.

adapt and build on achievements

Since the approval of the NDIIPP plan in 2002, here has been widespread adoption of radically new information technologies and social behaviors—changes that are most visible on the web. This period has also been characterized by deep disruptions in the national and global economies. During this period of change, the digital preservation network has grown rapidly. As the volume of valuable content grows, the program has distributed responsibility for content stewardship through the community by recruiting partners from increasingly diverse sectors. NDIIPP leverages the strengths of its partners by catalyzing new activity. Through regular assessment and consultation, it helps partners develop the program iteratively, learning from their experiences and building on their achievements. Above all, NDIIPP offers its partners a strategic approach to digital preservation, addressing the important components of the preservation network through a balance of short- and long-term actions and investments.

To date, the Library has recruited more than 185 digital preservation partners in more than 44 states and 25 nations to

Fig. 3. Strategic Goals

Through a network of committed preservation partners, NDIIPP has built and sustained a national digital collection, developed a shared technical infrastructure, and promoted public policy that encourages distributed digital preservation.

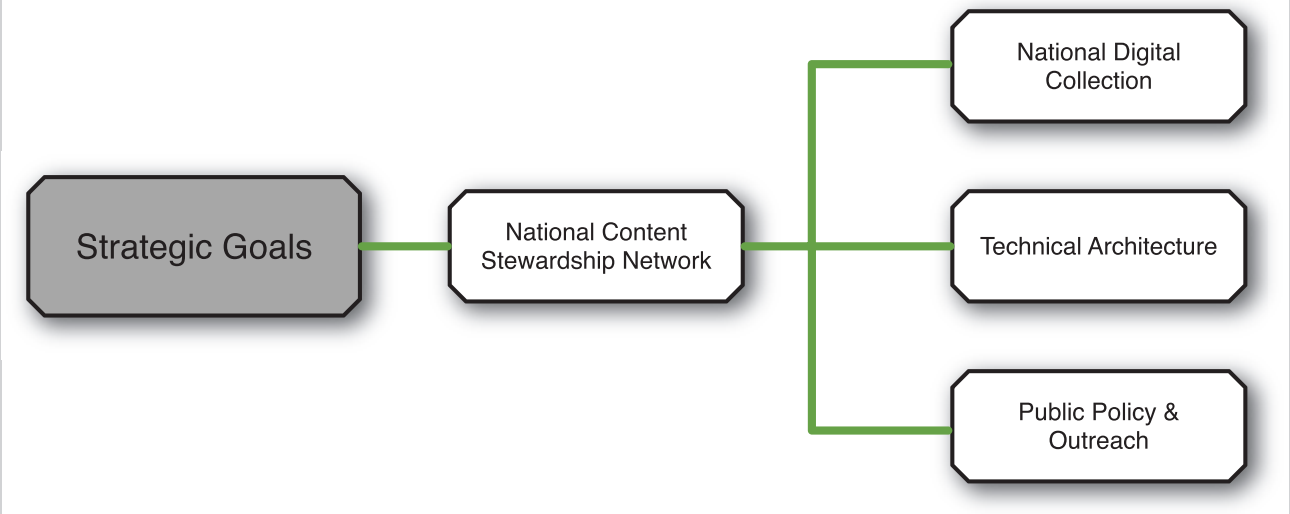
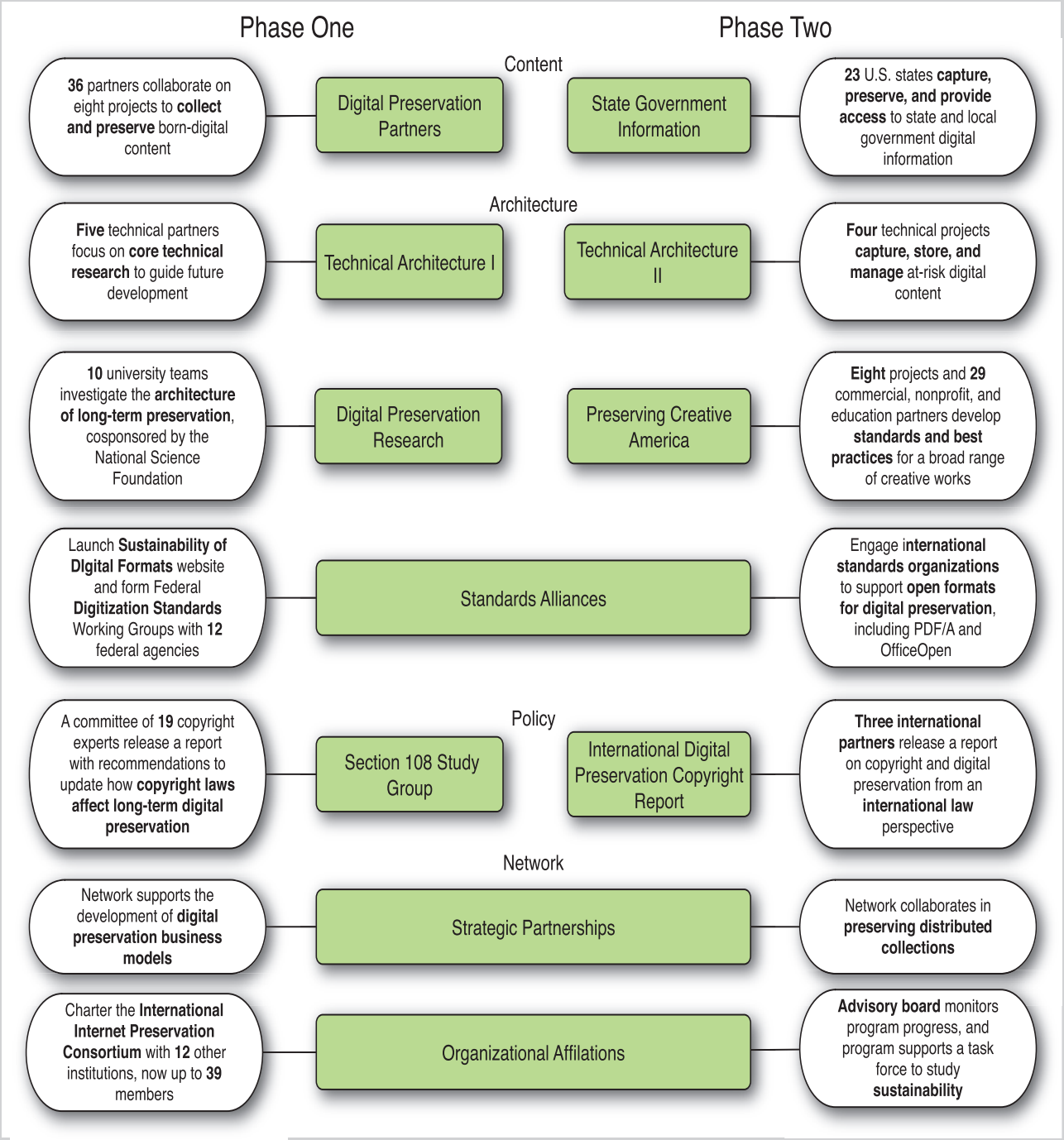


Fig. 4. Strategic Phased Initiatives

The program grew iteratively with partners crossing boundaries to leverage expertise, share resources, and build capacities across the network.



execute a multiphased plan to collect and preserve a broad range of high-value digital content, with special attention to the public policy, education and research, and cultural heritage communities. A list of partners is provided in Appendix B.

1. Phase One Initiatives

Phase one initiatives funded early actions to collect content, model and test preservation approaches, advance preservation science, and address rights issues presenting obstacles to sustainable digital content.

Several major events that occurred in the early years of the program underscored the importance—as well as the inherent fragility—of digital content. These included elections that had significant presences on the web and tragic events such as the 9/11 attacks and global natural disasters, in which the immediate need for information related to recovery highlighted the nation's growing dependence on electronic communication and information. In view of these challenges, the program took immediate actions to collect at-risk content. In addition to collecting websites, the Library collected and preserved social science data sets used by policy makers, political scientists, and historians to analyze social patterns and change. Geospatial data documenting land use and environmental modification were also given priority. The program made investments to ensure that this historically significant content was secured for the future.

capturing at-risk content

2. Phase Two Initiatives

Phase two initiatives funded continued growth of the network through new partnerships between content partners and technical partners to develop targeted tools and services, partnerships among state libraries and archives, and partnerships with commercial content creators and professional associations to develop and promote standards and best practices.

Section III describes key outcomes and findings of NDIIPP. The findings document the critical need to formalize the existing preservation network of partners so that they may, under the Library's leadership, continue collaborating to preserve and provide long-term access to digital content.


W E
JUNE. 16. 1891.

III. Building the National Digital Preservation Network

NDIIPP organized its initiatives and investments around four strategic goals:

1. **Stewardship Network:** Develop a growing national preservation network.
2. **National Digital Collection:** Develop a content-collection plan that will seed the national collection and preserve important at-risk content.
3. **Technical Infrastructure:** Build a shared technical platform for networked preservation.
4. **Public Policy**: Develop recommendations to address copyright issues and to create a legal and regulatory environment that both encourages incentives and eliminates disincentives to preservation.

The outcomes and findings of each goal are summarized below. Full reports from each initiative are available at www.digitalpreservation.gov.

A. Stewardship Network

Building distributed, networked capacity for digital preservation and long-term stewardship is a complex undertaking—one that demands action before all of the critical factors involved are fully understood. Growing and supporting that networked capacity, an equally complex task, requires sustained, dedicated coordination and support.

1. Key Outcomes

The key outcomes in this area are:

- Development of a stewardship network through recruitment of preservation partners.
- Articulation of partners' roles and responsibilities in the network, including the role of the Library of Congress as the central node in the network of networks.
- Development of a sustaining entity to govern a network of networks committing to the stewardship of digital content.

development of a stewardship network

Preservation partners are the stakeholders—creators and producers, owners, collectors, service providers, coordinating bodies, users, and others—who collaborate to preserve and provide access to digital content. They include libraries, public television networks, foreign news broadcasters, commercial content creators, state libraries and archives, industry groups and nonprofit organizations, and national libraries abroad. Each preservation partner brings its own network of associates and partners to participate in the larger constellation of organizations, creating, in effect, a network of networks. An important network in the program is the Federal Agencies Digitization Guidelines Initiative[4] with representatives from 15 federal agencies. This initiative has two working

NDIIPP Mission:

To ensure access over time to a rich body of digital content through the establishment of a national network of partners committed to selecting, collecting and preserving at-risk digital information.

[4] http://www.digitizationguidelines.gov.

Construction of the Jefferson Building, Library of Congress, with the U.S. Capitol in the background, June 16, 1891.

groups collaborating on the development of standards for digitization and preservation of digital content at the national level.

articulation of partners' roles and responsibilities

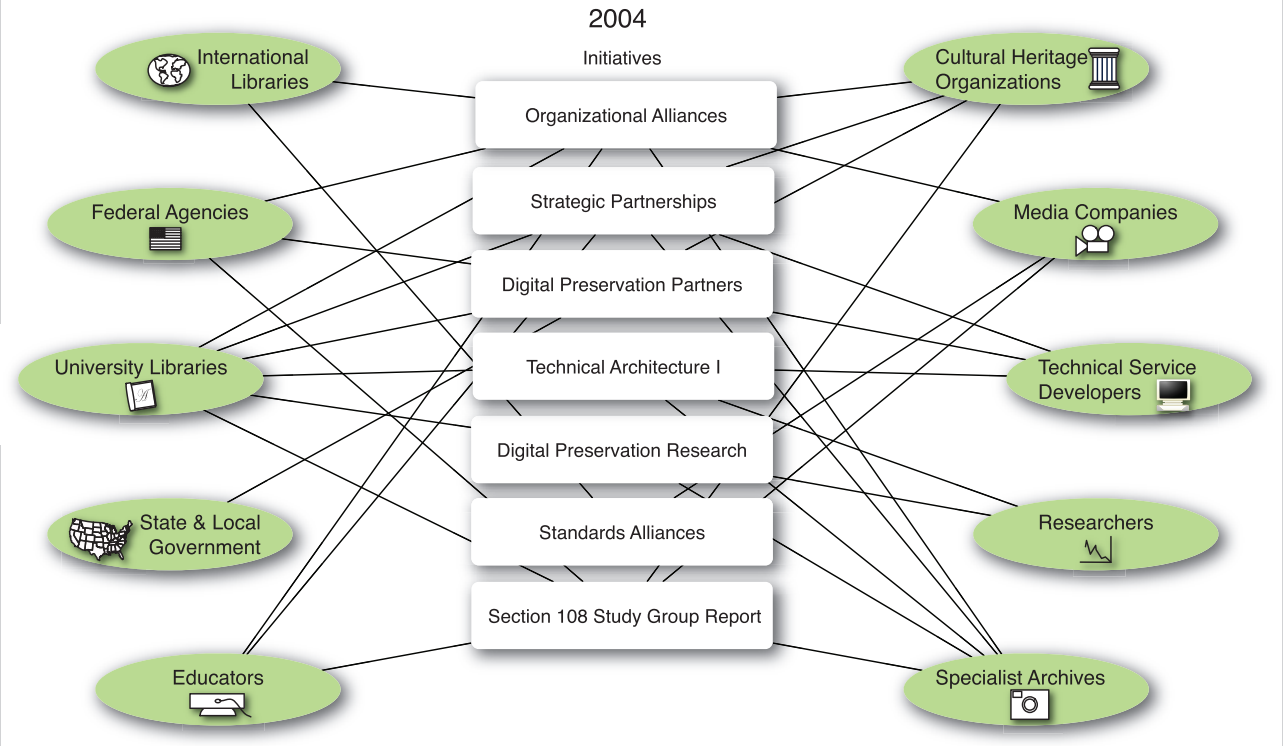
The complexity and diversity of the emerging network is a major strength. It has an inherent ability to leverage the expertise and experience of a broad range of participating institutions. Participation is driven by partners who recognize the benefits of becoming part of a larger whole. In that sense, the network acts like a web: it gets stronger as it gets larger and more complex. Affinity groups and working partnerships arise spontaneously as natural networks among content partners, technology experts, and service providers form, broaden, and deepen. Working within the trust-building environment of the network, member organizations share the responsibilities of preservation across a spectrum of groups and individuals with whom they might not otherwise interact.

central node in the network of networks

The network, as a peer-to-peer organization, requires leadership

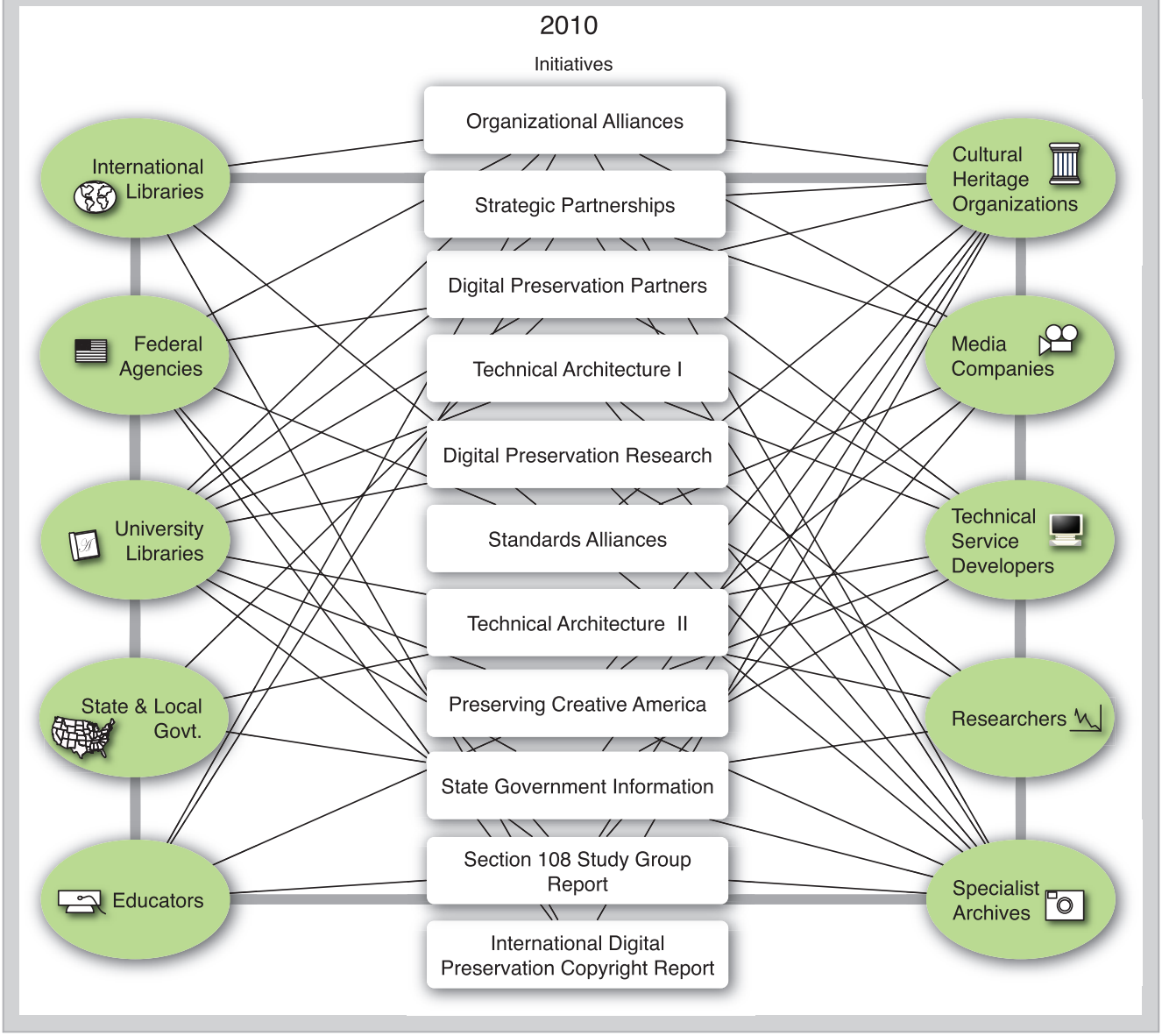
Fig. 5. Digital Preservation Communities, 2004

Between 2004 and 2010, natural networks grew as content communities collaborated on shared activities and problem solving.



to set and guide its agenda and to develop and sustain important relationships and collaborations. The Library of Congress led this effort by articulating a compelling vision for the network, by recruiting partners across a broad spectrum of content communities, and by performing a neutral convening function for all stakeholders. Because of its history of trusted relations with a wide range of content communities in the public policy, education and research, and creative sectors, the Library has been able to catalyze actions among content and technology stakeholders, pro-

Fig. 5. Digital Preservation Communities, 2010



vide incentives to spur and sustain participation, and represent the achievements and promote the interests of U.S. preservation initiatives at national and international forums.

Network partners have called for a transparent and accountable entity—the Library—to continue building trust among participating institutions by convening partners, leading program development, coordinating content collection, and ensuring self-governance among the preservation partners. Recommendations for advancing this network appear in Section IV.

2. Key Findings

- Preservation is a societal good undertaken by committed organizations. These organizations are motivated by their own interests and incentives, but as a group they also act on behalf of the public interest.
- Each participating institution brings to the network its own resources, interests, and organizational culture. Under the auspices of a neutral convener and honest broker, natural networks emerge through shared activities, problem solving, and meetings.
- As natural networks form, the larger network grows more complex, but also becomes stronger and better able to withstand stresses and strains.
- The Library has proven to be a reliable and trustworthy broker.

Federal Agencies Digitization Guidelines Initiative

Federal working groups for still images and audiovisual materials develop best practices for digital content creation, preservation, and access.

Goals of federal working groups:

- Identify and establish common digitization standards, methods, practices, and guidelines for the digitization of documents, printed matter, pictorial collections, and audiovisual content.
- Promote the creation of sustainable content.
- Encourage collaborative digitization practices and projects among federal agencies and institutions.
- Provide the public with a product of uniform quality.
- Set a common benchmark for digitization service providers.
- Enhance the exchange of research results and developments

Participating members

- Defense Visual Information Directorate
- Government Printing Office
- Institute of Museum and Library Services
- Library of Congress
- National Aeronautics and Space Administration
- National Agricultural Library
- National Archives and Records Administration
- National Gallery of Art
- National Institute of Science and Technology
- National Library of Medicine
- National Technical Information Service
- National Transportation Library
- Smithsonian Institution
- U.S. Geological Survey
- Voice of America

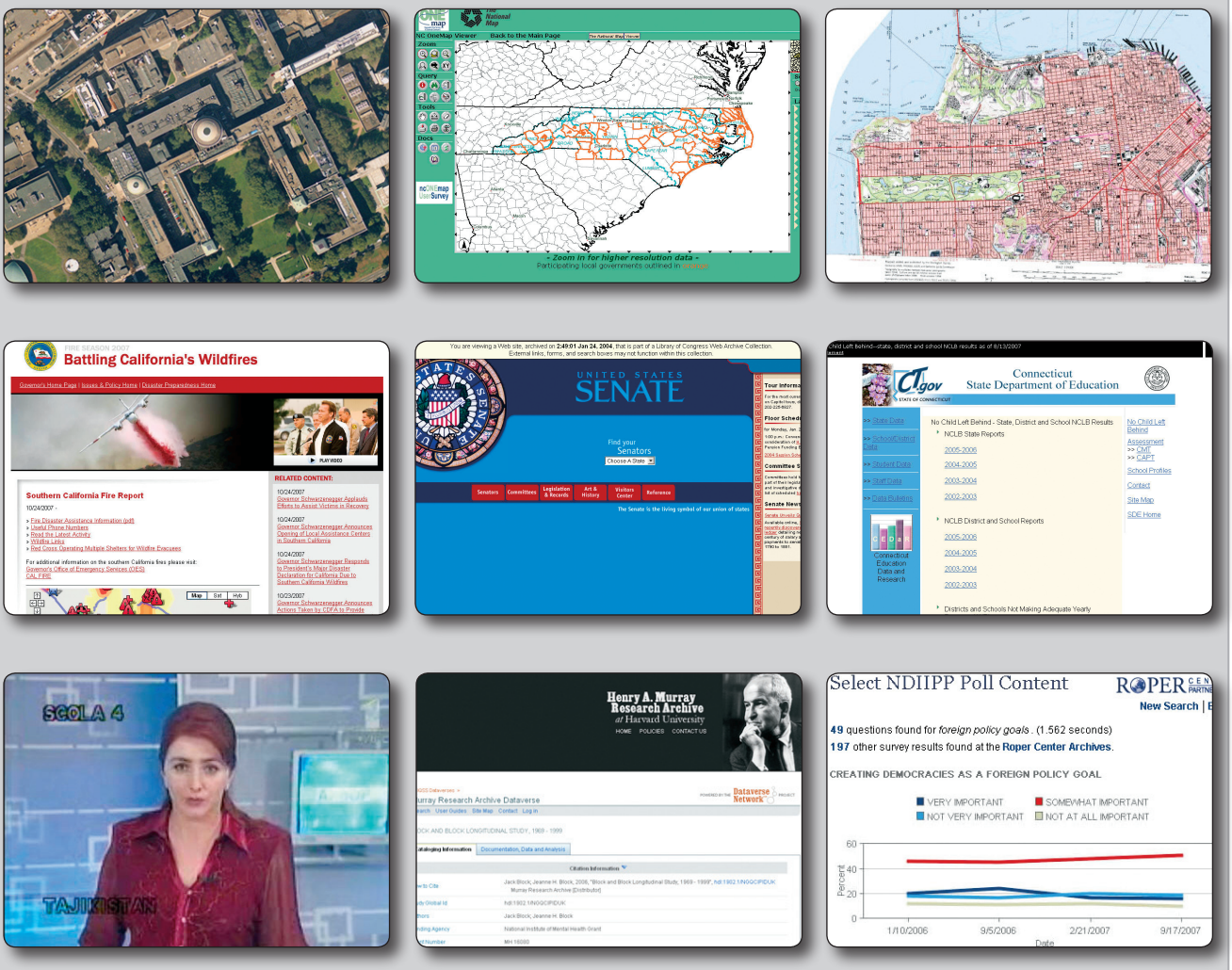
Fig. 6. Content Collected by NDIIPP Partners

Both historical and contemporary **geospatial data** are used for critical services, including environmental monitoring, transportation planning, political redistricting, and homeland security.

Events drive the preservation of **web-based content**. Collections include information on topics such as California wildfires, U.S. national elections, and the Iraq War; legal blogs; online-only government publications; and records of defunct government agencies and commissions.

Digital text and images are at the core of digital library collections. From historic American culture to statistical data sets to contemporary opinion polls, these collections have preserved important and rich content.

Audiovisual materials capture our dynamic history like no other medium. Digital technologies allow for broad and immediate distribution of news, culture, and ideas from a variety of view points. Preserving these resources will give future historians a deeper understanding of our time.



Credits left to right: North Carolina Geospatial Digital Archives, Chatham County Aerial Imagery; NC One Map Viewer; National Geospatial Digital Archive, California Spatial Information Library; California Digital Library, California Wildfires Collection; Library of Congress Web Archives, 2004 Election Collection; Connecticut State Library, CT State Department of Education Web Archives; SCOLA, Television News Collection from Tajikistan; Harvard University, Henry A. Murray Research Archive; Roper Center for Public Opinion, Roper Center Archives.

B. National Digital Collection

Digital content is dynamic, and its preservation demands approaches and skills that differ from those used for traditional preservation. Significant collections of digital content are being captured and preserved for future users by NDIIPP partners. The work to date informs an approach to the future establishment of a national digital collection. The scope of these collections is illustrated in figure 6; descriptions of collections preserved through NDIIPP are listed in Appendix D.

1. Key Outcomes

The key outcomes in this area are:

- Identification of collection priorities to inform a national digital collection.
- Identification and engagement of organizations committed to preserving digital content for the nation.
- Preservation of at-risk digital materials of high importance for the research, scholarship, and cultural heritage communities.

collection priorities

In harmony with the Library of Congress's historical collecting responsibilities, areas of highest priority for a national digital collection are those of greatest value to the public policy, education and research, and cultural heritage communities. Especially important in the digital realm are geospatial data, web-based content, digital text and images, and audiovisual content.

collecting now for future generations

Building a rich collection of content valuable to the nation requires a network of partners brought together by a shared vision of a national digital collection with value for present and future generations. As that collection grows, preservation partners will be guided by collection strategies that address the need to collect content *now* to ensure access in the *future*. Because digital content is "born ephemeral," collectors cannot wait for the passage of time to determine its value for the future. While preservation partners work to standardize and promote "born-archival" digital content formats to lessen the threat of built-in obsolescence, timely action to secure content today is the only strategy for preserving much content of value.

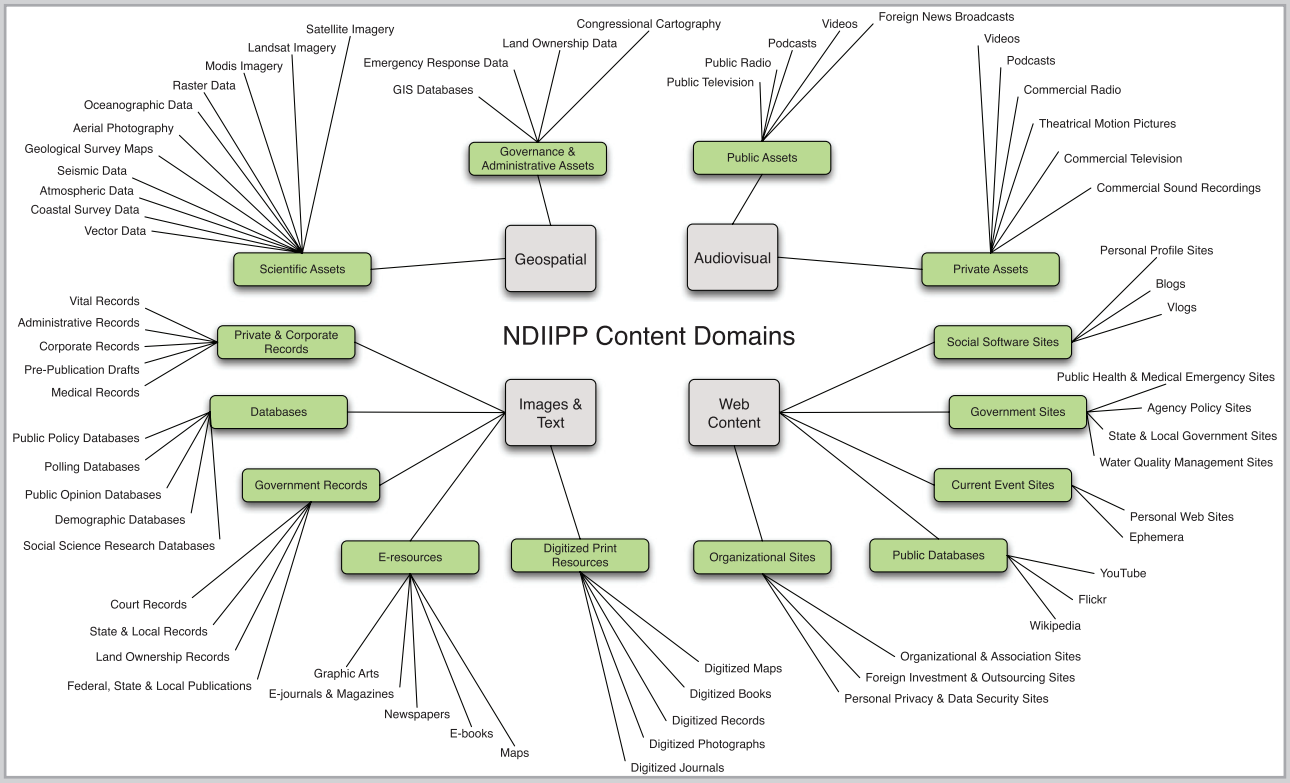
The partners share short-term strategies to capture at-risk content, as well as long-term strategies with multigenerational time horizons. In addition, because digital content may be more valuable when it is used in combination with other data, distributing collecting responsibilities to ensure broad coverage of key areas has emerged as a priority. For example, time-series aerial images of Hurricane Katrina were combined in real time with locally generated web-based

resources and analyzed using geospatial data to help emergency workers render assistance. Foreign broadcast news programs, political commentary websites, satellite imagery of global hot spots, and historical data sets that reveal long-term migrations of people can be vitally important to understanding global political developments from multiple perspectives. The collection framework for the network takes an integrated view of content and leverages the expertise of many disparate content communities to create a whole greater than the sum of its parts. The program also works closely with libraries and stewardship organizations abroad to ensure appropriate collection coordination.

The long-term goal of the preservation network is to ensure access to high-quality content over time, so that none of the imaginary scenarios sketched at the beginning of this report comes to pass. No single institution is able to collect and serve the universe of digital content that is of long-term value. Building and sustaining the national collection is an ongoing, distributed activity.

Fig. 7. NDIIPP Content Domains

The diversity of domains that create content vital for public policy, education and research, and cultural heritage communities indicates the breadth the network will achieve over time.



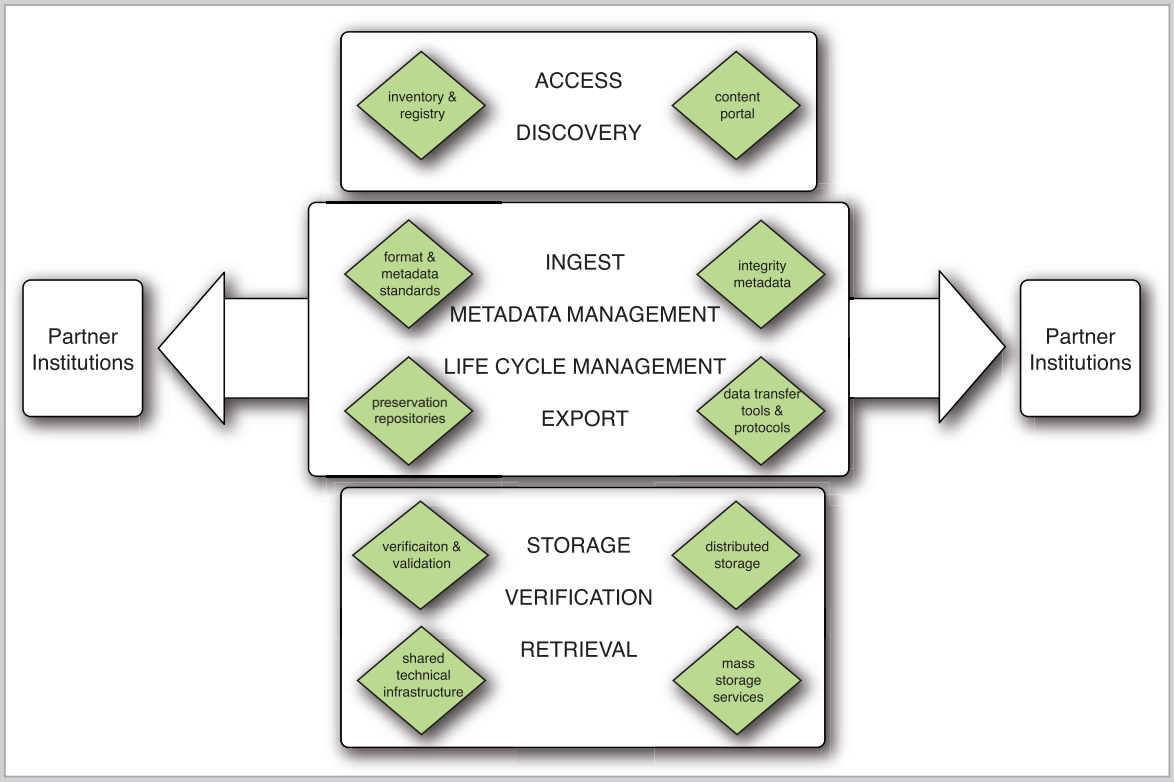
Network partners have explored divisions of collecting responsibilities. In the short term, the network has focused on a set of collecting priorities—geospatial, audiovisual, images and text, and web content. As indicated in figure 7, encompassing more content over time will require recruiting content partners from additional domains. It will also require developing new strategies for managing high-volume, ephemeral data and, ultimately, restructuring copyright and other policy-driven incentives for contributing content to the national collection. Proposed approaches for expanding the national collection by including more collecting institutions in the network are described in Section IV.

2. Key Findings

- Digital content is ephemeral. To ensure that it will be available in the future it must be collected and preserved now.
- The scale of content production will continue to outpace the ability of any single institution to collect and preserve a national digital collection,

Fig. 8. Technical Architecture for Digital Preservation

The three-layer architectural framework tested and modeled by digital preservation partners in a variety of projects has proven to be flexible and extensible across a diverse community of partners.



requiring a number of organizations committed to stewardship.
- The value of content grows when it can be shared across content domains, and the distributed approach to collecting has proven an effective strategy to achieve breadth and depth of content.

C. Technical Infrastructure

The technical infrastructure comprises both the architecture that enables content preservation and data exchange, and the tools and services to support partner participation in the architecture. The ability of the infrastructure to respond to the growing needs of preservation partners validates the distributed approach to preservation.

1. Key Outcomes

Key outcomes for the technical architecture are as follows:
- Testing of a three-layer preservation architecture and validation of the distributed approach to preservation, with clearly articulated roles among partners.
- Development and testing of tools and services to support distributed preservation (see Appendix C).
- Development of format and workflow standards specific to content domains of digital text and images, geospatial information, websites, and audiovisual materials.

three-layer preservation architecture

In the first year of planning for NDIIPP, a group of technology experts proposed a three-layer architecture for digital preservation (see figure 8). The lowest layer stores and maintains the data. The middle layer provides services for content description and management; it is the stewardship layer that is associated with libraries and archives. The top layer, the access layer, provides services to view and use the content.

supporting distributed preservation

These three layers provide a framework for distributed roles to support preservation across time and technological change. Diverse organizations and systems can be configured to leverage the best capabilities in each layer. This model accommodates the culture of access on the Internet that encourages innovative use of information in the top layer and the deliberate action of stewardship organizations that are concerned with the longevity of digital data in the middle layer. Commercial and nonprofit data centers provide expertise and services for storage and management at the bottom layer.

development of domain-specific standards

One of NDIIPP's earliest activities was to test this architectural framework by simulating the changes to which digital content would be subject over time.[5] The change of technology, including systems and formats, is usually the first

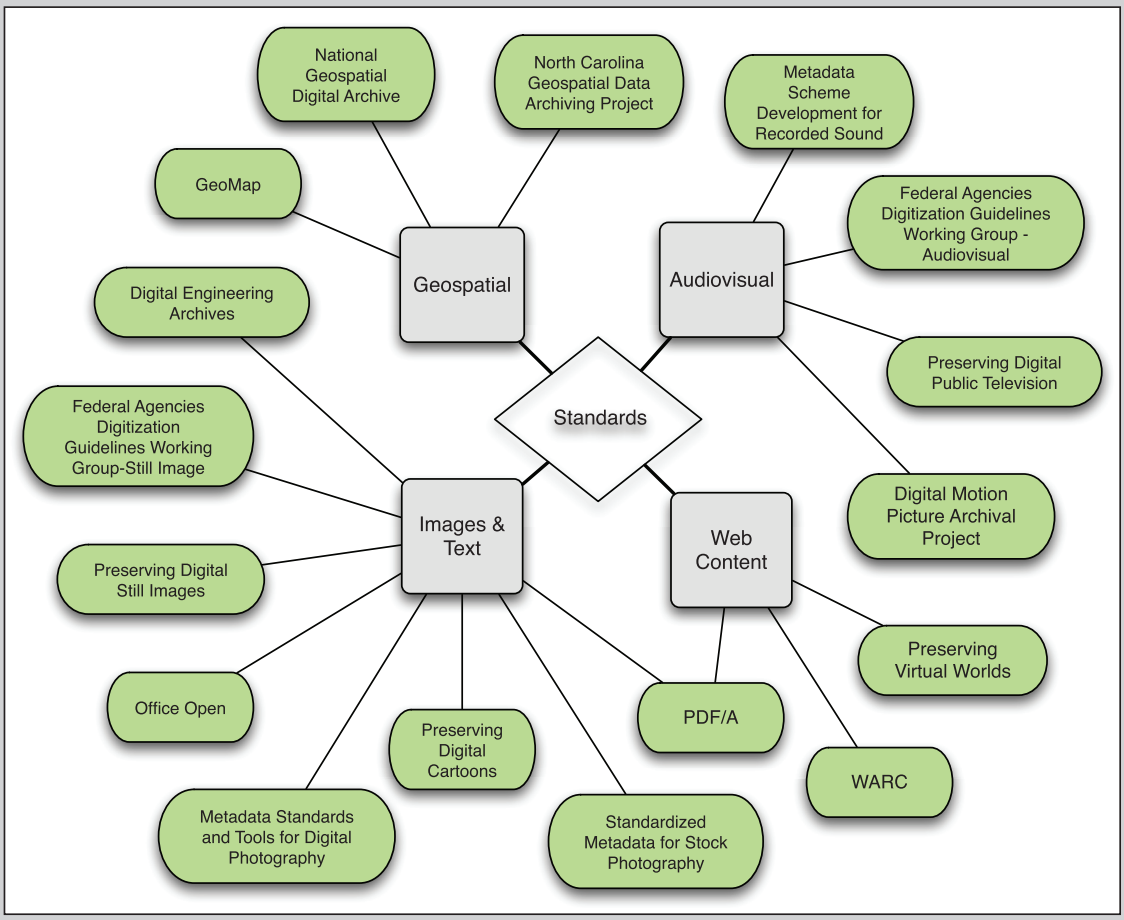
[5] http://www.digitalpreservation.gov/partners/aiht/high/ndiipp_aiht_final_report.pdf.

concern. However, changing organizational roles and responsibilities for stewardship over time should also be anticipated. An entity may cease to be capable of maintaining digital content, due to loss of resources or redirection of its mission and goals.

During the course of the program, each digital preservation project challenged assumptions and functions of this architectural model. The most important result was validation of the model in which the lower storage layer is generalized for all types of data, while the management layer in the middle is customized for the content type. The top access layer has proven to be the most volatile with the emergence of user-created content. Appendix C lists tools and services developed and tested for this architectural model.

Fig. 9. NDIIPP Standard-Making Activities

NDIIPP has supported standards activities across all content domains. The use of sustainable formats and adherence to technical standards are critical for reliable and cost-effective preservation.



The program has invested in standards work because content must be created in sustainable formats to ensure cost-effective preservation over time. See figure 9. Working closely with major content communities, the Library has taken the lead in making such work available through the Sustainability of Digital Formats website.[6] The continuation of this work across all content domains is critical to the long-term usability of content as hardware and software change.

2. Key Findings

- The diversity, complexity, and volume of digital objects and formats require that stewardship of digital content be shared across many organizations.
- Technical work across NDIIPP partnerships has affirmed the value of open development of tools.
- Maintaining redundant copies, distributed geographically, organizationally, and across diverse systems, is a better means of ensuring long-term security than is keeping only one copy or holding a copy in a single environment.
- Developing several approaches to the same problem is preferable to relying on a single approach, which risks data corruption or irretrievable loss should a single approach fail.

D. Public Policy

Building an infrastructure for the stewardship network requires a public policy environment that is conducive to preservation. Legal and incentive structures are as important to preservation success as is collecting the content itself.

1. Key Outcomes

Key outcomes in the policy arena are as follows:

- Recommendations for revising Section 108 of the U.S. copyright law to enable and encourage preservation of digital content.
- A comparative analysis of international copyright laws that affect preservation activities globally.
- A review of national and international policies that enable and encourage private entities to preserve content with public value.

U.S. copyright law

NDIIPP and the U.S. Copyright Office sponsored a cross-sectoral study group to conduct a detailed analysis of Section 108. Current 108 exceptions are optimized for print-on-paper copyright requirements; they do not match the technical realities of managing digital content and inadvertently often make digital best practices illegal. The study group delivered its recommendations to the Librarian of Congress and Register of Copyrights in a report dated March 31, 2008.

[6] http://www.digitalpreservation.gov/formats/.

The Register reviewed the study group's report and began seeking public comment in early 2009.

international copyright law

Recognizing that digital content crosses national boundaries, the Library examined the impact of international intellectual property laws and policies on digital preservation. In partnership with the United Kingdom, the Netherlands and Australia, the program sponsored the International Study on the Impact of Copyright Law on Digital Preservation.[7]

creating incentives and diminishing public policy impediments

NDIIPP convened working groups of partners to help develop policy recommendations for legal and regulatory changes that would recognize the broad public interest in long-term access to digital content. The working groups' discussions were based directly on the impediments to digital preservation that partners had encountered—impediments that could be diminished or eliminated through the types of incentives that federal and local governments provide for historic preservation of buildings, natural environments, and other significant heritage of the American people. The groups further explored these public policy matters by conducting a global review of the policy incentives that federal, local, and international governments use to encourage stewardship of national heritage. These issues are described in Appendix E.

Some privately held digital collections have great public interest for future generations, either because they document important parts of our history (such as business records) or because they document American creativity and constitute part of our cultural heritage (such as the manuscripts of public figures and writers, the outtakes of singers, and the unused footage of filmmakers). It is important to ensure that the owners and creators of such content can enjoy rights-protected use of this material. It is equally important that the historical and cultural heritage value of these materials be preserved for the public interest and made accessible to future generations.

sustainable economic models

Digital content is also at risk of loss because the economic models for sustaining content over time are immature and lack robust incentives to support the societal goal of long-term preservation. To address this gap, the program collaborated with the National Science Foundation and other organizations in a far-reaching study of the economic models

[7] *International Study on the Impact of Copyright Law on Digital Preservation*. A joint report of the Library of Congress National Digital Information Infrastructure and Preservation Program, the Joint Information Systems Committee, the Open Access to Knowledge (OAK) Law Project, and the SURFfoundation. July 2008. Available at http://www.digitalpreservation.gov/library/resources/pubs/docs/digital_preservation_final_report2008.pdf.

required to sustain digital content for long periods of time.[8] Economists and preservation experts of national renown developed models of sustainability for all communities that have a stake in long-term access to digital content. *Sustainable Economics for Digital Planet Ensuring Long Term Access to Digital Information* was published in February of 2010. [9]

2. Key Findings

- Both the copyright and regulatory environments need to be updated for the digital era; currently they discourage preservation best practices, or even make them illegal.
- Privately owned digital content often has high historical and cultural value for the public, and such value should be protected. When owners have no further use for their content, they should be given incentives to entrust it to a stewardship organization.
- There are few, if any, economic incentives for preservation. Because preservation and stewardship of the national collection serve the public interest, creating economic incentives is a priority.

Shared infrastructure increases the capacity for organizations to commit to digital preservation

Challenge: *The cost and expertise required for the installation of preservation storage and management systems prohibit organizations from preserving their digital content.*

Solution: *The MetaArchive project adapted a low-cost, fault tolerant auditing and monitoring software called LOCKSS to establish and run a distributed network for preservation storage of cultural heritage content. The system relies on cloud infrastructure and lower-cost commodity storage as an economical solution. Members in the network share their storage capacity with others in the network to provide for multiple copies that are audited and checked regularly by the software to ensure against data corruption.*

[8] *Sustaining the Digital Investment: Issues and Challenges of Economically Sustainable Digital Preservation*. Interim Report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access. December 2008. Available at http://brtf.sdsc.edu/biblio/BRTF_Interim_Report.pdf.

[9] Available at http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf.

IV. Securing Knowledge for the Future

The Library can now leverage the work of the NDIIPP program through its existing network with external and internal stakeholders to create leadership around digital preservation. According to an IBM study of NDIIPP in 2008, there are two principal mechanisms for the Libray to build on its success to further the goals of digital preservation. The first is leadership. The Library can lead, influence and provide additional momentum, resources, partners, and direction to the projects. Governance is another mechanism the Library can use to influence the structure and functioning of the partners to enhance their success.

This section describes actions under way to advance NDIIPP to the next phase; a self-sustaining network bound by common interests, synergies, and commitments. These actions have been developed from the key findings of the Program's initiatives as well as in-depth working sessions with the preservation partners, the Advisory Board, and other stakeholders. The next stage of progress entails:

- Chartering and developing the National Digital Stewardship Alliance.
- Expanding a national digital collection.
- Strengthening and enabling public-private partnerships.
- Fostering a public policy environment conducive to digital preservation.

A. Chartering the National Digital Stewardship Alliance

In response to the congressional charge to "develop standards and a nationwide collecting strategy" for the digital era, the Library of Congress developed and tested a distributed preservation network. The network has been effective in leveraging the strengths of a diverse set of partners, and has proven resilient in the face of technological volatility, economic downturn, and explosive growth of digital creation.

To sustain and foster the stewardship network, the Library is formalizing the National Digital Stewardship Alliance. Through a charter agreement, each Alliance member commits to work together to protect the investment made in digital resources across the Alliance. Members will serve in a variety of roles and working groups, including those focused on content, standards and best practices, infrastructure, innovation, and outreach.

1. Purpose of the Alliance

The purpose of the National Digital Stewardship Alliance is to create and sustain a national network of trusted partners to collect, preserve, and ensure long-term access to a national collection of digital content of value to Congress and the American people. The Alliance's functions are sixfold:

1. Build and sustain a national digital collection, a rich and diverse record of the national

Let us save what remains: not by vaults and locks which fence them from the public eye and use in consigning them to the waste of time, but by such a multiplication of copies, as shall place them beyond the reach of accident.

Thomas Jefferson

formalizing collaborative partnerships

Main Reading Room, Thomas Jefferson Building, Library of Congress
Photo by Carol M Highsmith

experience and information vital to the nation.
2. Support collaboration among Alliance member organizations and affiliates.
3. Identify, develop, and maintain services, standards, best practices, and sustainable business models of direct benefit to the Alliance.
4. Facilitate the development of a networked technical infrastructure, tools, and storage capacity for the Alliance.
5. Ensure the advancement of digital preservation science and technologies by encouraging research and development.
6. Foster an environment that supports long-term access to digital content through education, advocacy, and encouragement of sound public information policies.

formalizing collaborative partnerships

2. Role of the Library of Congress in the Alliance

The Library of Congress is uniquely positioned to serve as convener, promoter, and trusted voice for the Alliance. Trust among Alliance members—from rights holders to preservation organizations—is essential. The Library's centuries-old history of stewardship and the widespread recognition that it acts on behalf of the public interest will continue to open doors for the Alliance to recruit members and forge partnerships among diverse communities. Members of the Alliance are committed to serving as digital stewards of America's national collection of historically-significant digital content. The Library serves as the Executive Secretariat for the Alliance.

B. Developing a Framework for a National Digital Collection

Information that is essential for the study and development of public policy—government and nonprofit websites, news distributions, legal journals, maps and charts—is increasingly all digital. This content is at risk for loss for several reasons, but primarily because channels for its dissemination and content forms are digital and inchoate. Finding a means to capture and preserve such content is the Alliance's top priority.

NDIIPP has worked with more than 185 partners from 44 states and 25 countries to identify, collect, and preserve at-risk digital content. The experience of collecting and preserving this content from early program activities informs the Alliance's approach to expanding a national digital collection. The Alliance will work in collaborative partnerships to build and sustain a national digital collection intended to cover a broad scope of education and research, cultural heritage, and public policy subject areas. Partners have articulated a framework for a national digital collection that corresponds with the collections that libraries, and particularly the Library of Congress, have built over the past two centuries. While the genres and formats of this content will change over time, there

is an essential continuity between the breadth and depth of valuable analog collections and future born-digital collections.

Public and private organizations and individuals are the sources for the content. Libraries and archives produce digital library collections even as they are working to preserve them. Federal, state, and local government agencies create and publish public information. Commercial and non-profit publishers of news and commentary, and commercial and independent multimedia producers are important sources for many materials that will become part of a national digital collection.

Convening and consulting with recognized experts and stakeholders have been successful practices for NDIIPP. Experts, both at the Library and at other nationally recognized organizations, will be asked to help identify and collect the most authoritative and important content. Congressional Research Service analysts and staff will validate the relevance and importance of the selections for the development of public policy support for Congress. House and Senate librarians and archivists will be consulted in identifying government documents that are at risk of loss. Other expertise can be drawn from academic centers and specialized research centers.

Through the framework, the Alliance will build a collection of digital content that serves not only Congress and government agencies, but also the American people and an international audience. Content partners in the Alliance will collaborate on articulating collecting priorities; this will ensure appropriate levels of redundancy among collecting institutions, and will enable local and regional stewardship organizations to focus on collecting for their respective users.

expert & stakeholder selection

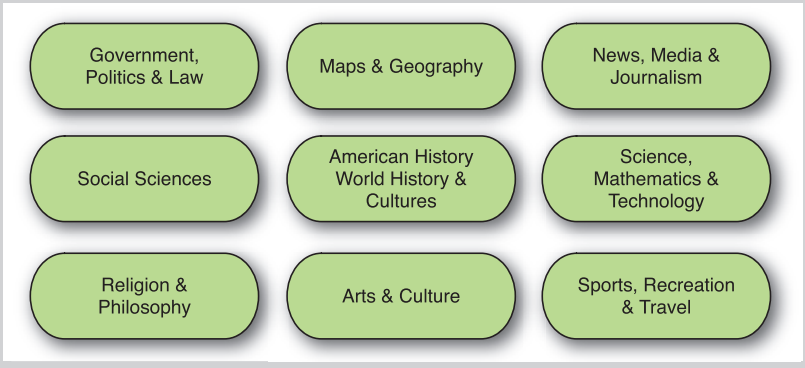
For a full discussion of the national digital collecting strategy, see Appendix G.

C. Strengthening and Enabling Public-Private Partnerships

The Library will seek to establish an independent entity that could provide a flexible means by which to fund innovation and develop shared tools and services for the Alliance, especially its federal members, by enabling robust public-private partnerships. Any tools and services developed would be

Fig. 10: National Digital Collection Framework

The major content areas of the national digital collection framework build on the strengths of the nation's analog collections and will serve the needs of a democratic society.



designed to meet the infrastructure needs of preservation and would be deployed throughout the Alliance, allowing partners with complicated service-contracting requirements or limited resources access to cutting-edge technology and research. The core functions of the entity would include:

- Sponsoring technical research and development.
- Serving as a broker for core infrastructure services such as storage.
- Raising funds from private sources and brokering public-private partnerships.

seek sustainable partnerships

An independent, non-profit entity such as a 501(c)(3) would have the ability to invest in new models of preservation and new technologies, and the flexibility to employ the necessary human resources throughout the country. It would provide the most nimble and cost-efficient mechanism for supporting digital preservation needs. An example for consideration in the establishment of a federally-chartered, non-profit entity that can receive private and public funds is the In-Q-Tel model used by the Federal intelligence community.[10] Preservation in the digital age, like intelligence gathering in the digital age, must not only keep pace with, but also anticipate technological change. Current contracting requirements and lengthy procurement processes create a vulnerability to technological obsolescence and inefficient use of public funds. These risks could be minimized or limited through public-private partnerships.

1. Partnerships for Innovation

The Library is fostering digital technology innovation through strategic partnerships with the private sector. These partnerships range from work with creative industries—to ensure the development and use of standards in the creation of information and entertainment assets—to a series of targeted partnerships with key firms to address emerging technical challenges. Thus, the Library joins innovation of the private sector with a commitment to the public trust and equitable access intrinsic to the public sector.

The Library will also form partnerships with educational institutions which will recruit the brightest young talent in higher education—at the undergraduate and graduate level—to work on cutting-edge technological challenges facing libraries. Thus, the Library takes advantage of the freshest perspectives on problem solving and feeds the pipeline of talent that libraries, archives, and all stewardship organizations need in the 21st century.

D. Fostering a Public Policy Environment Conducive to Digital Preservation

The overall goal of NDIIPP is to ensure that today's digital heritage

[10] *In-Q-Tel is an organization that serves the intelligence community to advance research and development. See https://www.cia.gov/library/publications/additional-publications/in-q-tel/index.html.*

is accessible to future generations. To this end, NDIIPP tackled critical technical challenges, collaboratively modeled and tested organizational approaches for a national strategy, and secured high-value content for a national digital collection. Beyond these three areas of engagement, however, is a fourth that shapes all of these while at the same time transcending them—the public policy environment in which the creation, use, and preservation of digital content occur.

At the beginning of the Republic, the Founding Fathers created legal structures such as copyright to encourage American creativity. At the same time, they created institutions such as the Library of Congress to ensure future generations have access to our nation's creative and intellectual output. Creating a public policy environment that is conducive to stewardship of our heritage ensured the growth of and continued access to knowledge. Recognizing the importance of a balanced public policy environment to the productivity and cultural richness of our nation, the Library undertook a broad review of the current legal structure for digital preservation and access.

The primary finding of the review is that there are few incentives, and too many disincentives, to preserve digital content in the public interest. This misalignment of incentives is a recent development. National and local governments have created a suite of complementary policy incentives that operate effectively in the analog realm. Many of these can and should be adapted for the digital environment. Policies should provide incentives and mitigate disincentives to act on behalf of the public good. This is the only way to ensure fair and equitable access to our intellectual, scientific, and creative output. Without such policies, stewardship organizations will encounter insuperable obstacles to the work they can undertake on behalf of the public interest. Without such policies, we are at risk of creating a digital dark age.

It is important to craft federal policies that work in a complementary fashion rather than at cross-purposes. Moreover, the federal government must set the pace for this activity. Although local and state incentives are also required, they will not emerge without federal leadership. Appendix E presents a full discussion of proposals to create a public policy environment conducive to the long-term access to digital materials.

create preservation incentives

In the near term, the Library will explore four major areas that can be addressed through federal policy: (1) updating copyright law for digital preservation; (2) creating tax-related incentives for digital preservation; (3) pilot a digital access project with copyright owners at the Library of Congress; and (4) reducing barriers to preservation of historically significant business, corporate, and privileged records.

1. Updating Copyright Law for Digital Preservation

Current copyright law and the lack of exceptions that clearly apply to digital preservation are insuperable obstacles to collecting and securing digital content for the future. Because preserving digital content requires creating multiple copies, copyright law becomes inadvertently implicated in even the most basic tasks in preserving digital content. The Section 108 Study Group examined what exceptions would be necessary and desirable to meet the needs of the digital era. In light of its findings, the Study Group's principal recommendations were as follows:

1. Include museums, which perform many of the same functions as libraries and archives, within Section 108 eligibility.
2. Create a new exception to permit qualified libraries, archives, and museums to make preservation copies of at-risk published works prior to any damage or loss. Access to these "preservation-only" copies will be limited.
3. Create a new exception to permit libraries, archives, and museums to capture and preserve publicly-available online content and make it accessible for research and scholarship. Permit rights holders to opt out of this provision.
4. Permit libraries, archives, and museums to make a limited number of copies, as reasonably necessary, to create and maintain a replacement or preservation copy. This alteration to the current three-copy limit would, among other things, enable all such entities to more securely preserve digital materials in their collections.

2. Creating Tax-Related Incentives for Digital Preservation

Digital content that has significant cultural value to society may have little value to its owner, and in such cases the owner may have no incentive to invest in its preservation. Such content might include sound recordings and films that have exhausted their commercial potential, geospatial data that records significant environmental conditions from the past that are of little value to a land developer, or corporate websites with outdated information.

Tax incentives have proven effective in encouraging the preservation of cultural assets by helping owners realize greater value from their cultural properties. But existing tax laws do not provide incentives for the preservation of historical digital materials. Tax credits that would cover a portion of the costs of preservation or enable individuals and corporations to donate digital cultural assets may provide enough of a subsidy to ensure long-term access to that content. The Library will investigate how best to develop tax-related incentives analogous to those that exist for the preservation of real property. This investigation may focus on issues such as:

1. An incentive modeled on tax

credits available for the rehabilitation and preservation of real property. The U.S. tax code allows taxpayers to claim deductions for certain expenditures made to rehabilitate qualified historic buildings. An analogous tax credit for digital preservation would allow taxpaying entities to deduct a portion of the costs of qualified digital preservation activities. These activities might include the development of technologies, services, or infrastructure necessary to advance digital preservation, or the commitment to preserve particular materials. Qualifying conditions and limitations could prevent the use of the credit or deduction for purposes other than "qualified digital preservation." A further requirement might be that the content to be preserved be certified as at-risk or otherwise as preservation-worthy by the Library of Congress.

2. An incentive modeled on deductions for charitable contributions of certified historic structures or historically important land areas. A similar tax deduction could be provided for contributions of digital collections to qualified cultural heritage institutions. This deduction would provide incentives for private individuals or entities to donate potentially-valuable digital assets to qualified cultural memory institutions and government entities for conservation or preservation purposes.
3. An incentive to contribute digital content for preservation through existing deductions for charitable contributions of tangible and intangible property. Amend the non-cash charitable deductions provision of the tax code to make deductions for donations of "qualified digital preservation works" to cultural heritage institutions applicable where the content has archival value but no established current market value. Because Congress's perception of widespread abuses of the charitable deduction has led it to restrict the availability and value of non-cash charitable deductions, any proposed charitable deduction amendments intended to encourage the preservation of digital materials should be carefully crafted to address such potential abuses.

3. Pilot Digital Project

Create a pilot project in which the Library of Congress may explore with copyright owners the digital display and/or dissemination of certain works that are in its collection and protected by copyright (for example, text, audio, visual or audio visual works) under terms to be mutually agreed upon.

4. Reducing Barriers to Preservation of Historically Significant Business, Corporate, and Privileged Records

The confidential or privileged nature of business records and other sensitive materials presents

additional challenges to preservation. These issues are described in detail by David Kirsch in his recent article, "The Record of Business and the Future of Business History: Establishing a Public Interest in Private Business Records."[11]

American business history—an important element of this country's broader history—has been built on the records that businesses have retained, often inadvertently, and that made their way into the hands of archives and historians decades later when their historic value had become apparent is at great risk especially in a digital environment. Generally, corporations—particularly public corporations with fiduciary duties to their stockholders—place no value on the historical benefits of their records, especially the public benefit derived from generalized findings.

The Library will study remedies with particular attention to the creation of specialized archives for business and confidential records by congressional charter or authorization. These are just some of the changes in public policy that could help ensure that the rich cultural heritage and full historical record of the American people will be accessible to future generations.

E. Investment Strategy for 2010–2020

The investment by the Library through NDIIPP continues to focus on forming strategic partnerships, content stewardship and practice, developing distributed infrastructure and shared tools and services, continuing policy study, and expanding outreach. The findings of the first ten years of work through the Program have demonstrated that these areas encompass the critical categories of effort.

Continuing into the next decade, priorities for investments will be based on strategic objectives to:

Respond to Information Challenges through Innovation and Action.

- Develop and promote effective standards and practices for selecting, organizing, preserving, and serving digital content; sponsor and maintain tools for curation and preservation; and providing services for the long-term preservation of digital content.
- Identifying, selecting, and stewarding a distributed, national digital collection to be preserved and made available to current and future generations.

Catalyze Collaboration for Digital Stewardship

- Bring together diverse sectors—government agencies, educational institutions, non-profit

[11]Kirsch, David A. 2009. The Record of Business and the Future of Business History: Establishing a Public Interest in Private Business Records. *Library Trends* 57(3): 352-370.

organizations, and business entities—to preserve a national collection of significant digital content.

- Support approaches for content stewardship and services that have the broadest application across communities.

Increase National Capacity for Stewardship of Digital Content

- Provide robust communication and outreach regarding all aspects of digital preservation and access; provide a central clearinghouse of information useful for all stakeholders.
- Develop expertise for digital preservation through the education and training of working professionals and students.
- Study and create awareness of public policy that contributes to long-term access to digital information as a national heritage resource.
- Encourage flexible, cost-effective services that make stewardship of digital content accessible to organizations large and small.

See Appendix H: Strategic Objectives 2000 - 2020 for more details.

F. Moving Forward

Much has changed, been learned, and been achieved since Congress passed NDIIPP legislation in 2000. The next phase of the national digital preservation plan will ensure growth of the National Digital Stewardship Alliance, with a goal of establishing partnerships in all 50 states. A structure will also be put in place that provides for the efficient development and deployment of tools and services across the Alliance so that all partners, regardless of size, have access to cutting-edge technologies. Building on collections already preserved and the Library's collection areas, the Alliance will establish a national digital collection that will ensure long-term access to digital resources of high value to Congress and the American people. The Alliance will also work toward creating a public policy environment that supports best practices and provides the incentive to preserve for the public good.

As the opening scenarios in this report point out, much is at stake if we do not act now: the nation's educational system, economic security, energy infrastructure, and the continuing creativity and innovation that assure the people's well-being all depend on a secure knowledge base. What is at stake is no less than the ability to show our children and grandchildren who we are and where we come from, to help them understand how our democracy grows, and to empower them with the knowledge and wisdom to make the difficult choices that the Founders well understood would confront us as a free people.

Appendixes

Paired sculptures of Minerva of War and Minerva of Peace.
Photo by Carol M Highsmith

Appendix A

National Digital Strategy Advisory Board

Jim Barksdale	Barksdale Management Corp.
Lynne Brindley	The British Library
Mary Chute	Institute of Museum and Library Services
Nancy Eaton	The Pennsylvania State University
Stephen M. Griffin	National Science Foundation Division of Information and Intelligent Systems
Margaret Hedstrom	School of Information, University of Michigan
Ellen Herbst	Department of Commerce
Betsy Humphreys	National Library of Medicine
Larry Irving	Irving Information Group
Glenn R. Jones	Jones International, Ltd.
Brewster Kahle	The Internet Archive
Donald A. B. Lindberg	National Library of Medicine
Clifford A. Lynch	Coalition for Networked Information
Carol Mandel	New York University
Victor McCrary	Applied Physics Laboratory, Johns Hopkins University
Charles E. Phelps	Office of the Provost, University of Rochester
Richard S. Rudick	John Wiley & Sons, Inc.
Robert C. Tapela	Government Printing Office
Ken Thibodeau	National Archives and Records Administration
Donald J. Waters	The Andrew W. Mellon Foundation
Allen Weinstein	National Archives and Records Administration
William A. Wulf	National Academy of Engineering
Peter Young	National Agricultural Library

Appendix B

Projects and Partners

Location	Partner Institutions	Project Name
Alabama	Alabama Department of Archives and History	Persistent Digital Archives and Library System (PeDALS)
	Auburn University Libraries	MetaArchive
Alaska	Alaska State Archives	Multi-State Preservation Consortium Utilizing the Washington State Digital Archives Framework
Arizona	Arizona State Library, Archives, and Public Records	ECHO DEPository (Exploring Collaborations to Harness Objects with a Digital Environment for Preservation)
		Persistent Digital Archives and Library System (PeDALS)
	University of Arizona, Eller College of Management	Investigating Data Provenance in the Context of New Product Design and Development
Arkansas	Arkansas State Library	A Model Technological and Social Architecture for the Preservation of State Government Digital Information
California	Academy of Motion Picture Arts and Sciences, Science & Technology Council	Digital Motion Picture Archive Framework Project
	California State Library	A Model Technological and Social Architecture for the Preservation of State Government Digital Information
	Internet Archive (IA)	ArchiveIT
		International Internet Preservation Consortium (IIPC)
		Web Capture
	J. Paul Getty Trust	Section 108 Study Group
	Legislative Counsel of California	A Model Technological and Social Architecture for the Preservation of State Government Digital Information
	Linden Labs	Preserving Virtual Worlds
	Morrison & Foerster, LLP	Birth of the Dot Com Era
	National Academy of Recording Arts and Sciences Producers and Engineers Wing	Federal Agencies Digitization Guidelines Initiative (Still Image Working Group)
		Metadata Schema Development for Recorded Sound
	Ropers Majeski Kohn & Bentley PC	Birth of the Dot Com Era
	Stanford University Libraries and Academic Information Resources	Stanford Agreement (LOCKSS/CLOCKSS)
		Archive Ingest and Handling Test
		National Geospatial Digital Archive (NGDA)
	Stanford University, Stanford Humanities Lab	Preserving Virtual Worlds

Appendix B

Location	Partner Institutions	Project Name
	Universal Mastering Studios	Section 108 Study Group
	University of California at Santa Barbara Libraries (UCSB)	National Geospatial Digital Archive (NGDA)
	University of California, California Digital Library	A Model Technological and Social Architecture for the Preservation of State Government Digital Information
		JHOVE2: A Next Generation Architecture fo Format Aware Digital Preservation Processing
		Web at Risk: A Distributed Approach to Preserving Our Nation's Political Cultural Heritage
	University of California, Los Angeles, Film & Television Archives	Preserving Digital Independent Film
	University of California, San Diego, San Diego Supercomputer Center	Chronopolis
		Digital Preservation Lifecycle Management: Building a Demonstration Prototype for the Preservation of Large Scale Multimedia Collections
		Distributed Storage and Preservation Services
		Multi-Institution Testbed for Scalable Digital Archiving
	Scripps Institute of Oceanography	Multi-Institution Testbed for Scalable Digital Archiving
Colorado	Colorado State Library	Multi-State Preservation Consortium Utilizing the Washington State Digital Archives Framework
	National Conference of State Legislatures (NCSL)	A Model Technological and Social Architecture for the Preservation of State Government Digital Information
Connecticut	Connecticut State Library	ECHO DEPository (Exploring Collaborations to Harness Objects with a Digital Environment for Preservation)
	University of Connecticut, Roper Center for Public Opinion Research	Data Center Data Preservation Alliance for the Social Sciences (Data-PASS)
District of Columbia	American Library Association	Section 108 Study Group
	Biodiversity Heritage Library	DuraCloud
	Business Software Alliance	Section 108 Study Group
	District of Columbia Office of Public Records	Geospatial Multistate Archive and Preservation Project (GeoMAPP)
	District of Columbia Office of the Chief Technology Officer	Geospatial Multistate Archive and Preservation Project (GeoMAPP)
	Folger Shakespeare Library	MetaArchive
	Georgetown University Law Center	Section 108 Study Group
	National Archives and Records Administration	Federal Agencies Digitization Guidelines Initiative (Still Image Working Group)
	National Gallery of Art	Federal Agencies Digitization Guidelines Initiative (Still Image Working Group)
	National Transportation Library	Federal Agencies Digitization Guidelines Initiative (Still Image Working Group)
	Recording Industry Association of America	Metadata Schema Development for Recorded Sound

Appendix B

Location	Partner Institutions	Project Name
	Smithsonian Institution	Federal Agencies Digitization Guidelines Initiative (Audio-Visual Working Group)
	U.S. Government Printing Office	Federal Agencies Digitization Guidelines Initiative (Still Image Working Group)
		Federal Agencies Digitization Guidelines Initiative (Audio-Visual Working Group)
		International Internet Preservation Consortium (IIPC)
	Voice of America	Federal Agencies Digitization Guidelines Initiative (Audio-Visual Working Group)
	Walt Disney Company	Section 108 Study Group
Florida	Florida State University Libraries	MetaArchive
	State Library and Archives of Florida	Persistent Digital Archives and Library System (PeDALS)
Georgia	Emory University, Robert W. Woodruff Library	MetaArchive
	Georgia Archives	Geospatial Multistate Archive and Preservation Project (GeoMAPP)
	University of Georgia, Carl Vinson Institute of Government, Office of Information Technology Outreach Services Division	Geospatial Multistate Archive and Preservation Project (GeoMAPP)
	Georgia Institute of Technology Library and Information Center	MetaArchive
	MetaArchive Cooperative/ Educopia	MetaArchive
Idaho	Idaho Commission for Libraries	Multi-State Preservation Consortium Utilizing the Washington State Digital Archives Framework
	Idaho State Historical Society	Multi-State Preservation Consortium Utilizing the Washington State Digital Archives Framework
Illinois	Illinois State Library	A Model Technological and Social Architecture for the Preservation of State Government Digital Information
		ECHO DEPository (Exploring Collaborations to Harness Objects with a Digital Environment for Preservation)
	University of Illinois at Urbana-Champaign, Graduate School of Library and Information Science, and UI Library (UIUC)	ECHO DEPository (Exploring Collaborations to Harness Objects with a Digital Environment for Preservation)
		Preserving Virtual Worlds
Indiana	Indiana State Archives and Library	Multi-State Preservation Consortium Utilizing the Washington State Digital Archives Framework
Iowa	SCOLA	Preserving International Television
Kansas	Kansas Legislative Computer Services	A Model Technological and Social Architecture for the Preservation of State Government Digital Information

Appendix B

Location	Partner Institutions	Project Name
	Kansas State Historical Society	A Model Technological and Social Architecture for the Preservation of State Government Digital Information
Kentucky	Kentucky Commonwealth Office of Technology, Division of Geographic Information	Geospatial Multistate Archive and Preservation Project (GeoMAPP)
	Kentucky Department for Libraries and Archives	Geospatial Multistate Archive and Preservation Project (GeoMAPP)
	Kentucky State University	Geospatial Multistate Archive and Preservation Project (GeoMAPP)
	University of Louisville Libraries	MetaArchive
Louisiana	Louisiana State Archives	Multi-State Preservation Consortium Utilizing the Washington State Digital Archives Framework
Maine	Maine Office of GIS	Geospatial Multistate Archive and Preservation Project (GeoMAPP)
	Maine State Archives	Geospatial Multistate Archive and Preservation Project (GeoMAPP)
Maryland	Johns Hopkins University	Securely Managing the Lifetime of Versions in Digital Archives
		Archive Ingest and Handling Test
	Maryland State Archives	Geospatial Multistate Archive and Preservation Project (GeoMAPP)
	Maryland Department of Natural Resources	Geospatial Multistate Archive and Preservation Project (GeoMAPP)
	National Agricultural Library	Federal Agencies Digitization Guidelines Initiative (Audio-Visual Working Group)
	National Archives and Records Administration, Electronic and Special Media Records Service Division	Data Center Data Preservation Alliance for the Social Sciences (Data-PASS)
		Federal Agencies Digitization Guidelines Initiative (Audio-Visual Working Group)
	National Institute of Standards and Technologies	Federal Agencies Digitization Guidelines Initiative (Audio-Visual Working Group)
	National Library of Medicine	Federal Agencies Digitization Guidelines Initiative (Audio-Visual Working Group)
		Federal Agencies Digitization Guidelines Initiative (Still Image Working Group)
		Section 108 Study Group
	University of Maryland Institute for Advanced Computer Studies	Robust Technologies for Automated Ingestion and Long-term Preservation of Digital Information
	University of Maryland Institute for Advanced Computer Studies	Chronopolis
	University of Maryland Institute for Technology in the Humanitites, Electronic Literature Organization	Preserving Virtual Worlds
	University of Maryland Robert H. Smith School of Business	Birth of the Dot Com Era

Appendix B

Location	Partner Institutions	Project Name
Massachusetts	Boston College	MetaArchive
	DuraSpace	DuraCloud
	Educational Broadcasting Corporation (EBC) (13/WNET NY)	Preserving Digital Public Television
	Harvard University Library	Archive Ingest and Handling Test
		Global Digital Format Registry
	Harvard University, Institute for Quantitative Social Science, Harvard-MIT	Data Center Data Preservation Alliance for the Social Sciences (Data-PASS)
	Tufts University, Peruses Project	ECHO DEPository (Exploring Collaborations to Harness Objects with a Digital Environment for Preservation)
	WGBH Educational Foundation	Preserving Digital Public Television
	Woods Oceanographic Institution	Multi-Institution Testbed for Scalable Digital Archiving
Michigan	Michigan State University Library, Vincent Voice Library	ECHO DEPository (Exploring Collaborations to Harness Objects with a Digital Environment for Preservation)
	University of Michigan, Inter-University Consoritum for Political and Social Research (ICPSR)	Data Center Data Preservation Alliance for the Social Sciences (Data-PASS)
	University of Michigan, School of Information	Incentives for Data Provenance in the Content of New Product Design and Development
Minnesota	Minnesota Department of Administration, Geospatial Information Office	Geospatial Multistate Archive and Preservation Project (GeoMAPP)
	Minnesota Historical Society	A Model Technological and Social Architecture for the Preservation of State Government Digital Information; Geospatial Multistate Archive and Preservation Project (GeoMAPP)
	Minnesota Legislative Reference Library	A Model Technological and Social Architecture for the Preservation of State Government Digital Information
	Minnesota Office of the Revisor of Statutes	A Model Technological and Social Architecture for the Preservation of State Government Digital Information
	Thomson Reuters/West	A Model Technological and Social Architecture for the Preservation of State Government Digital Information
Mississippi	Mississippi Department of Archives and History	A Model Technological and Social Architecture for the Preservation of State Government Digital Information
Missouri	Universal Press Syndicate	Preserving Digital Cartoons
Montana	Montana Historical Society	Multi-State Preservation Consortium Utilizing the Washington State Digital Archives Framework; Geospatial Multistate Archive and Preservation Project (GeoMAPP)
Nebraska	Nebraska State Historical Society	A Model Technological and Social Architecture for the Preservation of State Government Digital Information
Nevada	Nevada State Library and Archives	Multi-State Preservation Consortium Utilizing the Washington State Digital Archives Framework

Appendix B

Location	Partner Institutions	Project Name
New Jersey	Portico	Preservation of E-journals
	Stock Artists Alliance	Metadata Standards for Stock Photography
New Mexico	Los Alamos National Library Research Library	Memento: Adding Time to the Web
		American Physical Society eJournal Transfer and Ingest
	New Mexico Commission of Public Records State Records Center and Archives	Persistent Digital Archives and Library System (PeDALS)
New York	ARTstor	Preserving Digital Still Images
	Association of American University Presses	Section 108 Study Group
	Audio Engineering Society	Metadata Schema Development for Recorded Sound
	Columbia University Libraries	Section 108 Study Group
	Columbia University Law School	Section 108 Study Group
	Cornell University Library	Section 108 Study Group
	Cowan, DeBaets, Abrahams & Sheppard, LLP	Section 108 Study Group
	Gallivan, Gallivan & O'Melia	Birth of the Dot Com Era
	JSTOR	Section 108 Study Group
	New York Public Library	DuraCloud
	New York State Archives	Persistent Digital Archives and Library System (PeDALS); Geospatial Multistate Archive and Preservation Project (GeoMAPP)
	New York State Office of CyberSecurity & Critical Infrastructure Coordination	Geospatial Multistate Archive and Preservation Project (GeoMAPP)
	New York University	Preserving Digital Public Television
		Web at Risk: A Distributed Approach to Preserving Our Nation's Political Cultural Heritage
	Penguin Group (USA)	Section 108 Study Group
	Rochester Institute of Technology, Game Programming, Department of Information Technology Council	Preserving Virtual Worlds
	The Andrew W. Mellon Foundation	Section 108 Study Group
	Time, Inc.	Section 108 Study Group
North Carolina	North Carolina Center for Geographic Information and Analysis	Geospatial Multistate Archive and Preservation Project (GeoMAPP)
	North Carolina State Archives	Geospatial Multistate Archive and Preservation Project (GeoMAPP)
	North Carolina State Library	ECHO DEPository (Exploring Collaborations to Harness Objects with a Digital Environment for Preservation)
	North Carolina State University Libraries (NCSU)	Geospatial Multistate Archive and Preservation Project (GeoMAPP)
		North Carolina Geospatial Data Archiving Project (NCGDAP)

Appendix B

Location	Partner Institutions	Project Name
	The Geographic Information Coordinating Council (GICC)	North Carolina Geospatial Data Archiving Project (NCGDAP)
	University of North Carolina, Howard W. Odum Institute for Research in Social Science	Data Center Data Preservation Alliance for the Social Sciences (Data-PASS)
	University of North Carolina at Chapel Hill, School of Information and Library Science	A Model Technological and Social Architecture for the Preservation of State Government Digital Information
		Preserving Video Objects and Context: A Demonstration Project
	University of North Carolina School of Law	Section 108 Study Group
	NC OneMap	North Carolina Geospatial Data Archiving Project (NCGDAP)
North Dakota	North Dakota Legislative Council Library	A Model Technological and Social Architecture for the Preservation of State Government Digital Information
	State Historical Society of North Dakota	A Model Technological and Social Architecture for the Preservation of State Government Digital Information
Ohio	OCLC (Online Computer Library Center)	ECHO DEPository (Exploring Collaborations to Harness Objects with a Digital Environment for Preservation)
Oregon	Oregon State Archives	Multi-State Preservation Consortium Utilizing the Washington State Digital Archives Framework
	Oregon State Library	Multi-State Preservation Consortium Utilizing the Washington State Digital Archives Framework
Pennsylvania	American Society of Media Photographers	Workflow, Archiving, and Metadata Standards for Digital Photography
	Drexel University, Geometric and Intelligent Computing Laboratory	Digital Engineering Archives
	Pennsylvania State University	MetaArchive
South Carolina	Clemson University	MetaArchive
	South Carolina Department of Archives and History	Persistent Digital Archives and Library System (PeDALS)
	University of South Carolina	MetaArchive
Tennessee	BMS/Chace	Metadata Schema Development for Recorded Sound
	Tennessee State Libraries and Archives	A Model Technological and Social Architecture for the Preservation of State Government Digital Information
	University of Tennessee at Knoxville Computer Science Department	National Geospatial Digital Archive (NGDA)
		Planning a Globally Accessible Archive of MODIS Data
	Vanderbilt University	National Geospatial Digital Archive (NGDA)
Texas	Rice University	MetaArchive
	Texas Natural Resources Information Systems	Geospatial Multistate Archive and Preservation Project (GeoMAPP)
	Texas State Library and Archives Commission	Geospatial Multistate Archive and Preservation Project (GeoMAPP)

Appendix B

Location	Partner Institutions	Project Name
	University of North Texas Libraries	International Internet Preservation Consortium (IIPC)
		Web at Risk: A Distributed Approach to Preserving Our Nation's Political Cultural Heritage
		MetaArchive
Utah	Sundance Institute	Preserving Digital Independent Film
	Utah Automated Geographic Reference Center	Geospatial Multistate Archive and Preservation Project (GeoMAPP)
	Utah Division of Archives	Geospatial Multistate Archive and Preservation Project (GeoMAPP)
Vermont	Vermont State Archives	A Model Technological and Social Architecture for the Preservation of State Government Digital Information
Virginia	Defense Visual Information (DVI) Directorate	Federal Agencies Digitization Guidelines Initiative (Audio-Visual Working Group)
	George Mason University, Center for History and New Media	Birth of the Dot Com Era
	National Science Foundation Office for Cyberinfrastructure	Blue Ribbion Task Force on Sustainable Digital Preservation and Access
	National Technical Information Service	Federal Agencies Digitization Guidelines Initiative (Still Image Working Group)
	Old Dominion University, Department of Computer Science	Archive Ingest and Handling Test
		Memento: Adding Time to the Web
		Tools for a Preservation-Ready Web: Shared infrasturucre preservation models
	Public Broadcasting Service	Preserving Digital Public Television
	U.S. Geological Survey	Federal Agencies Digitization Guidelines Initiative (Audio-Visual Working Group)
		Federal Agencies Digitization Guidelines Initiative (Still Image Working Group)
	Virginia Polytechnic Institute and State University Libraries (VA Tech)	MetaArchive
Washington	Washington State Archives	Multi-State Preservation Consortium Utilizing the Washington State Digital Archives Framework
	Washington State Library	Multi-State Preservation Consortium Utilizing the Washington State Digital Archives Framework
Wisconsin	Wisconsin Historical Society	Persistent Digital Archives and Library System (PeDALS)
	Wisconsin State Library	ECHO DEPository (Exploring Collaborations to Harness Objects with a Digital Environment for Preservation)
	University of Wisconsin-Madison, Arthur H. Robinson Map Library	Geospatial Multistate Archive and Preservation Project (GeoMAPP)
Wyoming	University of Wyoming, American Heritage Center	Geospatial Multistate Archive and Preservation Project (GeoMAPP)
	University of Wyoming, Geographic Information Science Center	Geospatial Multistate Archive and Preservation Project (GeoMAPP)

Appendix B

Location	Partner Institutions	Project Name
Oceania - Australia	National Library of Australia	International Internet Preservation Consortium (IIPC)
	Open Access to Knowledge (OAK) Law Project	International Study on the Impact of Copyright Law on Digital Preservation
South America - Brazil	PUC Rio de Janeiro	MetaArchive
North America - Canada	Bibliothèque et Archives Nationales du Québec (BANQ)	International Internet Preservation Consortium (IIPC)
	Library and Archives Canada	International Internet Preservation Consortium (IIPC)
Asia - China	National Library of China	International Internet Preservation Consortium (IIPC)
Europe - Croatia	National and University Library of Croatia	International Internet Preservation Consortium (IIPC)
Europe - Czech Republic	Národní knihovna České republiky (National Library of the Czech Republic)	International Internet Preservation Consortium (IIPC)
Europe - Finland	Kansalliskirjasto (Helsinki University Library, The National Library of Finland)	International Internet Preservation Consortium (IIPC)
Europe - France	Bibliotheque Nationale De France (National Library of France)	International Internet Preservation Consortium (IIPC)
	Ina (Institut National de l'Audiovisuel)	International Internet Preservation Consortium (IIPC)
Europe - Germany	Deutsche Nationalbibliothek (German National Library)	International Internet Preservation Consortium (IIPC)
Europe - Iceland	Landsbokasafn Islands - Haskolabokasafn (National and University Library of Iceland)	International Internet Preservation Consortium (IIPC)
Asia - Israel	Jewish National and University Library	International Internet Preservation Consortium (IIPC)
Europe - Italy	Biblioteca Nazionale Centrale di Firenze (National Library of Italy, Florence)	International Internet Preservation Consortium (IIPC)
Asia - Japan	National Diet Library, Japan	International Internet Preservation Consortium (IIPC)
Oceania - New Zealand	National Library of New Zealand	International Internet Preservation Consortium (IIPC)
Europe - Norway	Nasjonalbiblioteket (The National Library of Norway)	International Internet Preservation Consortium (IIPC)

Appendix B

Location	Partner Institutions	Project Name
Europe - Poland	National Library of Poland	International Internet Preservation Consortium (IIPC)
Europe - Scotland	National Library of Scotland	International Internet Preservation Consortium (IIPC)
Asia - Singapore	National Library Board, Singapore	International Internet Preservation Consortium (IIPC)
Europe - Slovenia	National and University Library (Slovenia)	International Internet Preservation Consortium (IIPC)
Asia - South Korea	National Library of Korea	International Internet Preservation Consortium (IIPC)
Europe - Spain	Biblioteca de Catalunya	International Internet Preservation Consortium (IIPC)
Europe - Sweden	Kungl. biblioteket (National Library of Sweden)	International Internet Preservation Consortium (IIPC)
Europe - Switzerland	Swiss National Library	International Internet Preservation Consortium (IIPC)
Europe - The Netherlands	European Archive Foundation	International Internet Preservation Consortium (IIPC)
	Koninklijke Bibliotheek (National Library of the Netherlands)	International Internet Preservation Consortium (IIPC)
	Netarchive.dk	International Internet Preservation Consortium (IIPC)
	SURFfoundation	International Study on the Impact of Copyright Law on Digital Preservation
	Virtual Knowledge Studio – Royal Netherlands Academy for Arts and Sciences	International Internet Preservation Consortium (IIPC)
Europe - United Kingdom	British Library	International Internet Preservation Consortium (IIPC)
	Hanzo Archives Limited	International Internet Preservation Consortium (IIPC)
	Joint Information Systems Committee (JISC)	International Study on the Impact of Copyright Law on Digital Preservation
	National Archives of England, Wales, and the United Kingdom	International Internet Preservation Consortium (IIPC)
	University of Hull (UK)	MetaArchive

Appendix C

NDIIPP Partner Tools and Services Inventory

This is a list of tools and services designed, developed, or used by NDIIPP partners during their projects. The Library encourages NDIIPP partners and others in the preservation community to share in, and take advantage of, these resources.

Archive-It

A subscription service from the Internet Archive that allows institutions to build and preserve collections of born-digital content. Through a web application, Archive-It partners can harvest, catalog, manage, and browse their archived collections. Collections are hosted at the Internet Archive data center and are accessible to the public with full-text search. More than 65 stewardship institutions around the world partner with Internet Archive to archive the web using Archive-It.

- **Developer**: Internet Archive
- **NDIIPP project:** Internet Archive
- **Written in**: Java
- **OS and run-time environment**: Web-based
- **Application**: http://www.archive-it.org/
- **Documentation**: http://webteam.archive.org/confluence/display/ARIH/Welcome
- **License**: Fee-based

Audit Control Environment (ACE)

A prototype tool that validates the integrity of digital files through mathematical techniques. Its purpose is to ensure the authenticity of digital objects in long-term archives. ACE consists of a third-party Integrity Management Service (IMS), which issues integrity tokens for digital objects, and a local archive Audit Manager (AM), which periodically validates the repository. Consistency is guaranteed through the use of the archive-independent IMS to validate integrity tokens and with the publication of witness values to prove the correctness of the system.

- **Developer:** University of Maryland
- **Written in:** Java
- **OS and run-time environment:** Web-based and platform-independent. Requires Java 1.4 or greater.
- **Application:** V 1.4, http://adapt.umiacs.umd.edu/ace
- **Documentation:** https://wiki.umiacs.umd.edu/adapt/index.php/Ace
- **License:** To be decided

BagIt

A specification for the packaging of digital content for transfer. Content is packaged (the bag) along with a small amount of machine-readable text (the tag) to help automate the content's receipt, storage, and retrieval. There is no software to install. A bag consists of a base directory

containing the tag and a subdirectory that holds the content files. The tag is a simple text-file manifest, like a packing slip, that consists of two elements: (1) an inventory of the content files in the bag; and (2) a checksum for each file.

A slightly more sophisticated bag lists URLs instead of simple directory paths. A script then consults the tag, detects the URLs, and retrieves the files over the Internet, 10 or more at a time. This simultaneous multiple transfer reduces overall data-transfer time. In another optional file, users can supply metadata that describe the bag.
SEE ALSO: *BagIt Library* and *BagIt Transfer Utilities*

- **Developer**: University of California, California Digital Library; Library of Congress
- **NDIIPP project:** Web-at-Risk
- **Written in**: N/A
- **OS and run-time environment**: N/A
- **Application**: N/A
- **Documentation**: Bagit specification http://www.digitalpreservation.gov/library/resources/tools/docs/bagitspec.pdf
- **License**: N/A

BagIt Library (BIL)

A Java software library that supports the creation, manipulation, and validation of bags.
SEE ALSO: *BagIt* and *BagIt Transfer Utilities*

- **Developer**: Library of Congress
- **Written in**: Java
- **OS and run-time environment**: N/A
- **Application**: http://sourceforge.net/projects/loc-xferutils/
- **Documentation**: http://sourceforge.net/projects/loc-xferutils/
- **License**: Public domain

BagIt Transfer Utilities

A collection of tools developed by the Library of Congress and its partners in NDIIPP for the purpose of validation and transfer of bags.

The Parallel Retriever optimizes the retrieval of bags through parallelization, and produces a bag when given a file manifest and a "fetch.txt" file. VerifyIt verifies a bag manifest using parallel md5 processes. The Bag Validator validates a bag against the BagIt specification, as well as checking for files in the manifest that are missing from the disk, files on the disk that are not listed in the manifest, and duplicate entries in manifest.
SEE ALSO: *BagIt* and *BagIt Library*

- **Developer**: Library of Congress
- **Written in**: Python and UNIX shell
- **OS and run-time environment**: UNIX
- **Application**: http://sourceforge.net/projects/loc-xferutils/
- **Documentation**: http://sourceforge.net/projects/loc-xferutils/
- **License**: BSD

Conspectus Database for LOCKSS Private Network

LOCKSS software provides inexpensive digital preservation through replication of data storage in multiple locations; the Conspectus Database provides a central "catalog" with records that describe the data in each location. This is used by the MetaArchive project, whose group members are jointly developing

—a prioritized survey of at-risk digital content held at the partner sites;
—a harvested body of the most critical content to be preserved at the partner sites; and
—a distributed preservation network infrastructure based on the LOCKSS software.

The conspectus database is web-based, searchable, and browsable, and it requires only a login ID and password.

- **Developer**: Emory University
- **NDIIPP project:** MetaArchive
- **Written in**: N/A; web-based
- **OS and run-time environment**: N/A; web-based
- **Application**: http://www.metaarchive.org/conspectus/
- **Documentation**: http://www.metaarchive.org/conspectus/
- **License**: N/A

Note: The MetaArchive software engineer is currently working with the LOCKSS team to develop new tools to accomplish curation and monitoring tasks for Private LOCKSS Networks in Ruby on Rails. The software will likely be released under an open-source license in the near future.

ContextMiner

A framework to collect, analyze, and present contextual information along with the data. It is based on the idea that while describing or archiving an object, contextual information helps make sense of that object or makes it possible to preserve it better.

- **Developer**: University of North Carolina at Chapel Hill, School of Information and Library Science
- **NDIIPP project:** Vidarch
- **Written in**: N/A; web-based
- **OS and run-time environment**: N/A; web-based
- **Application**: http://www.contextminer.org/index.php
- **Documentation**: http://www.contextminer.org/index.php
- **License**: N/A

Dataverse Network

The Dataverse Network software is an open-source, digital library system for management, dissemination, exchange, and citation of virtual collections (dataverses) of quantitative data. Dataverses can be used or administered through web-based clients that communicate with a host Dataverse Network.

A Dataverse Network, usually running at a major institution, requires installation of application software. Individual dataverses are self-contained *virtual* data archives, served out by a Dataverse Network and appearing on the websites of their owners (e.g., individuals, departments, projects, or publications). Dataverses are branded in the style of the owning entity but

are easy to set up, require no local software installations, and offer the services of a modern archive controlled by the dataverse owner. Data are displayed in a hierarchy; descriptive information (Data Documentation Initiative [DDI]) can be searched.

Depending on the policies of the dataverse owner, end-users may be able not only to download files but also to extract subsets and to perform statistical analysis online. Dataverses and Dataverse Networks can federate with each other and with other systems through open protocols (OAI-PMH and Z39.50).

- **Developer**: Institute for Quantitative Social Science, Harvard University
- **NDIIPP project:** Data-PASS
- **Written in**: Java Platform, Enterprise Edition (Java EE) 5, including Enterprise Java Beans (EJB) 3 and Java Server Faces
- **OS and run-time environment**: Individual dataverses are managed and used through web-based graphic user interface. Software installations are required only to create an entire Dataverse Network. Software runs on top of the Glassfish Application Server. Harvard uses PostgreSQL for database software. The data analysis component uses R and Zelig for statistical computing.
- **Documentation**: http://thedata.org/
- **Application**: http://dvn.iq.harvard.edu/
- **License**: Gnu Affero General Public License, version 3: http://gplv3.fsf.org/comments/agplv3-draft-1.html (a version of GPLv3: http://gplv3.fsf.org/)

Digital Archive

Provides a secure storage environment to manage and monitor the health of master files and digital originals. It also provides a managed storage environment for digital master files that fits in with the workflows for acquiring digital content.

For users of CONTENTdm(R) (either hosted or direct), the Digital Archive is an optional capability integrated with the various workflows for building collections. Master files are secured for ingest to the archive using the CONTENTdm Acquisition Station, the Connexion digital import capability, and the web harvesting service. For users of other content management systems, the Digital Archive provides a low-overhead mechanism for safely storing master files.

- **Developer**: OCLC
- **Written in**: Java
- **OS and run-time environments**: Linux, MySQL, Apache, Tomcat
- **Application**: http://oclc.org/digitalarchive
- **Documentation**: http://www.oclc.org/digitalarchive/support/default.htm
- **License**: Fee-based

DiscoverInfo

A tool to explore a collection of documents. It enables the user to

—*Search*: Runs a full-text search in the collection. DiscoverInfo indexes text, HTML, XML, and PDF documents.

—*Browse*: Builds term clouds based on the term occurrences in the collection as well as across the documents. Users can browse through the clickable term clouds to discover documents.

—*Discover*: Retrieves relevant information from the indexed collection; also evaluates the novelty of information in documents with respect to other documents in the collection.

- **Developer**: University of North Carolina at Chapel Hill, School of Information and Library Science
- **NDIIPP project:** Vidarch
- **Written in**: N/A; web-based
- **OS and run-time environment**: N/A; web-based.
- **Application**: http://idl.ils.unc.edu/~chirag/DIToolkit/
- **Documentation**: http://idl.ils.unc.edu/~chirag/DiscoverInfo/index.html
- **License**: N/A

EchoDep Hub and Spoke Framework Tool Suite

A hosted service and open technology developed that makes it easy for organizations and end users to use cloud services. It offers cloud storage across multiple commercial and non commercial providers and compute services to unlock the digital content stored in the cloud. It provides services that enable digital preservation, data access, transformation and data sharing. Customers are offered "elastic capacity" coupled with a "pay as you go" approach.

- **Developer: Duraspace, a collaboration of the Fedora Commons and the DSpace Foundation**
- **NDIIPP Project: Duracloud**
- **Written in: N/A. Web based.**
- **OS and run-time environment: N/A. Web based.**
- **Application: Release 0.5**
- **Documentation: https://wiki.duraspace.org/display/duracloud05/DuraCloud+Release+0.5**
- **License: Apache License, Version 2.0**
- **Last tool update: August, 2010**

EchoDep Hub and Spoke Framework Tool Suite

With a set of simple tools, Hub and Spoke provides a method for exchanging digital files and metadata among different types of digital management systems built on different platforms. It provides basic interoperability among repositories via a common METS-based profile, a standard programming API, and a series of scripts that use the API and METS profile for creating SIPs and DIPs that can be used across different repositories. Key architectural components are:

—the *METS profile*, which remains mostly neutral regarding content files and structure but defines a minimum level of descriptive (MODS) and administrative (PREMIS) metadata, with an emphasis on preserving technical data and provenance.

—the *REST-based Lightweight Repository Create, Retrieve, Update, and Delete Service (LRCRUDS)*, which maps URIs to local identifiers and uses HTTP methods (PUT, GET, POST, and DELETE} to handle packages submitted or disseminated from a repository. Packages are shipped as Zip archives containing a header, METS file, and content files in a format suitable for repository import.

—the *Hub*, which converts from and to the METS profile and manages generation and validation of technical and provenance metadata. At present, the Hub is a package-staging area; the goal is to develop it into a digital repository capable of disseminating packages and handling submissions from other repositories.

- **Developer**: University of Illinois, Urbana-Champaign
- **NDIIPP project:** ECHO DEPository: Exploring Collaborations to Harness Objects with a Digital Environment for Preservation
- **Written in**: An interpreted language (Java, Perl)

- **OS and run-time environment**: OS-independent
- **METS profile**: http://www.loc.gov/standards/mets/profiles/00000015.html
- **LRCRUDS**: http://dli.grainger.uiuc.edu/echodep/HnS/LRCRUDS.htm
- **Application**: http://sourceforge.net/projects/echodep/
- **Documentation**: http://dli.grainger.uiuc.edu/echodep/hands/
- **License**: University of Illinois/NCSA Open Source License, http://www.opensource.org/licenses/UoI-NCSA.php

Federated Archive Cyberinfrastucture Testbed (FACIT)

A technology testbed that explores the use of geographically distributed storage in a networked environment. It builds on logistical networking technology (see http://loci.cs.utk.edu/) using the Internet Backplane Protocol (IBP) (see http://loci.cs.utk.edu/ibp/) to provide a generic interface for managing distributed storage resources. Each FACIT archive will use L-Store (see http://www.lstore.org) to manage data storage in both its private infrastructure and in the shared storage pool that the federation makes available.

Using L-Store, and leveraging IBP, FACIT archives will automatically mirror each other's content to provide fault-tolerance and increased accessibility. For its wide area storage infrastructure, FACIT archives will participate in the larger Research and Education Data Depot Network (REDDnet) storage network (see http://www.reddnet.org/). Since REDDnet is based on IBP and supports L-Store, FACIT archives will have seamless access to this larger, shared pool of storage.

- **Developer**: University of California, Santa Barbara; Vanderbilt University; University of Tennessee at Knoxville
- **NDIIPP project:** National Geospatial Digital Archive
- **Written in**: L-Store is written in Java; IBP is written in C.
- **OS and run-time requirements**: Linux, Unix
- **Application**: Command line interface; GUI in development
- **Documentation**: http://www.ngda.org/FACIT.php
- **License**: Berkeley BSD

GIS Archiving Toolset

The Toolset prepares vector and raster datasets for archive ingest. Basic pre-ingest functions include limited format validation, fidelity management, virus scanning, data set characterization, metadata creation and remediation, and SIP organization.

- **Developer**: North Carolina State University
- **NDIIPP project**: North Carolina Geospatial Data Archiving Project
- **Written in**: Python
- **OS and run-time requirements**: The Toolset was written to run cross-platform, but has been tested only in Linux. Core requirements are met by Python. Extended functionality requires calls to external applications including ClamAV, NOID, 4Suite XML, Unix File, and JHOVE.
- **Application**: Tool is not shared.
- **Documentation**: Tool is not shared.
- **License**: Tool is not shared.

Heritrix

A flexible, extensible, robust, and scalable web crawler capable of fetching, archiving, and analyzing Internet-accessible content.

- **Developer**: Internet Archive
- **NDIIPP project:** Internet Archive
- **Written in**: Java
- **OS and run-time requirements**: Written in Java. Must have Java Runtime Environment (JRE, http://www.java.com/en/download/index.jsp) and at least Java version 5.0 installed. Default heap size is 256MB RAM. Heritrix not tested, packaged, or supported on platforms other than Linux at this time.
- **Application**: http://crawler.archive.org
- **Documentation**: http://crawler.archive.org/articles/user_manual and http://webteam.archive.org/confluence/display/Heritrix/Home
- **License**: GNU Lesser General Public License 2.1 (http://crawler.archive.org/license.html); migrating to Apache License 2.0 in future

integrated Rule-Oriented Data Systems (iRODS)

A data grid that allows the end-user control over storage management policies and procedures through definition of business rules tailored to the characteristics of the files being managed. It provides an abstraction for data management processes and policies in the same way as the Storage Resource Broker provided abstractions for data objects, collections, resources, users, and metadata, but is flexible and customizable.

This is accomplished by coding the processes as microservices that are controlled by explicit rules. Management policies are mapped to sets of rules, and management processes are mapped to sets of microservices. Assessment criteria are mapped to queries on the persistent state information generated by execution of each microservice. A distributed rule engine is installed at each storage location to ensure enforcement of policies independently of the choice of access mechanism.

iRODS architecture features include:

—peer-to-peer data grid servers, based on a client/server model and distributed storage resources;

—a database system for maintaining the attributes and states of data and operations; and

—a rule system for enforcing and executing adaptive rules.

- **Developer**: San Diego Supercomputer Center
- **Written in**: iRODS servers written in C. iRODS clients are written in the appropriate language; Java I/O library, PHP web browser, Python web browser.
- **OS and run-time environment**: Linux, Solaris, Macintosh, and AIX. The iCAT Platforms page at http://irods.sdsc.edu/index.php/iCAT_Platforms lists the supported operating systems and configurations for ICAT-enabled servers. Currently either a PostgreSQL or Oracle database may be used for managing state information.
- **Application**: http://irods.sdsc.edu/index.php/Downloads
- **Documentation**: http://irods.sdsc.edu/index.php/Documentation
- **License**: BSD open source (http://irods.sdsc.edu/index.php/License)

JSTOR/Harvard Object Validation Environment

An extensible system designed to provide automated and efficient identification and validation of the format of digital files with minimal human intervention. JHOVE can:

- identify the format to which a digital object conforms;
- determine the compliance of an object to its format's specification, both in terms of syntax (well-formedness) and semantics (validity); and
- characterize an object in terms of its format-specific significant properties.

JHOVE defines a Java API and also provides a stand-alone application that runs in either command line or GUI mode. JHOVE supports the following formats: AIFF, ASCII, GIF, HTML, JPEG, JPEG 2000, PDF, TIFF, UTF-8, WAVE and XML.

- **Developer**: Harvard University
- **Written in**: Java 1.4
- **OS and run-time environment**: JHOVE should be usable on any UNIX, Windows, or OS X platform with an appropriate J2SE installation. It should run on any operating system that supports Java 1.4 and has a directory-based file system.
- **Application**: http://sourceforge.net/projects/jhove/
- **Documentation**: http://hul.harvard.edu/jhove/documentation.html
- **License**: GNU Lesser General Public License (LGPL) (http://www.gnu.org/licenses/lgpl.html)

LOCKSS

LOCKSS provides inexpensive digital preservation through replication of data storage in multiple locations. Copies of the same content in multiple LOCKSS replicas are automatically compared to each other, and can be repaired by the comparisons automatically.

- **Developer:** Stanford University
- **Written in:** Java
- **OS and run-time environment:** All POSIX (Linux/BSD/UNIX-like OS), Linux. Most LOCKSS installations use a CD that bundles the LOCKSS daemon with an operating system based on OpenBSD. The LOCKSS team also supports running the daemon on RPM-based Linux distributions and on Solaris. The LOCKSS daemon can run in any environment with a Java VM 1.5 or above and a Unix-like file system. The hosting PC needs at least 1 GB of memory, a CD drive, and at least 250 GB of storage. The current CD distribution supports parallel (PATA) and serial (SATA) ATA and SCSI drives. On Linux and Solaris the daemon can use the full set of storage options.
- **Application**: http://sourceforge.net/projects/lockss/
- **Documentation**: http://www.lockss.org/lockss/Installing_LOCKSS
- **License**: BSD, http://www.lockss.org/lockss/Software_License

L-Store (Logistical Storage)

L-Store is low-level system software that leverages the basic powerful protocols of the Internet to move and manage large chunks of data through digital networks, much as the Internet moves and manages e-mails and other traffic. It is built on the Internet Backplane Protocol (IBP) (see http://loci.cs.utk.edu/ibp/). The L-Store client provides a storage framework for distributed, scalable, and secure access to data. It is to be used on the Research and Education Data Depot Network (REDDnet) infrastructure (see http://www.reddnet.org/). L-Store is

designed to provide:
—high scalability in both raw storage and associated file;
—a decentralized management system;
—security;
—fault-tolerant metadata support;
—user-controlled replication and striping of data on file and directory level;
—scalable performance in both raw data movement and metadata queries;
—a virtual file system interface in both a web and command line form; and
—support for the concept of geographical locations for data migration to facilitate quicker access.

- **Developer:** Vanderbilt University
- **Written in:** Java
- **OS and run-time requirements**: Java 1.6 or better
- **Application**: http://www.lstore.org/pwiki/pmwiki.php?n=Docs.CLI-ClientIntro
- **Documentation**: http://www.lstore.org/pwiki/pmwiki.php?n=Docs.CLI-ClientIntro
- **License**: BSD: http://www.opensource.org/licenses/bsd-license.php

Logistical Distribution Network (LoDN)

An experimental content distribution tool, LoDN allows users to store content on the REDDnet and to manage or retrieve that stored content without installing anything or learning to use complicated software. LoDN comprises three elements: (1) upload; (2) download clients (powered by Java Web Start) for storing and retrieving data; and (3) a web interface for managing stored data and browsing public content.

LoDN uses the Logistical Networking infrastructure provided by the Internet Backplane Protocol (IBP) (see http://loci.cs.utk.edu/ibp/) deployed on REDDnet (http://www.reddnet.org) to store file content on IBP storage "depots." Content publishers can use LoDN's web interface to manage stored data. Content distributors can make LoDN data files available by including an active LoDN link on a web page, in an e-mail, or through the LoDN content directory. Users access a file by clicking a LoDN link, thereby starting the LoDN Download Client and then using the download client to retrieve the file content directly from IBP storage.

- **Developer**: University of Tennessee at Knoxville
- **Written in**: Web-based; uses Java Webstart
- **OS and run-time requirements**: Any Java capable, version 1.4.2 or better
- **Application**: https://ln.eecs.utk.edu/lodn/
- **Documentation**: https://ln.eecs.utk.edu/lodn/
- **License**: BSD: http://www.opensource.org/licenses/bsd-license.php

National Geospatial Digital Archive (NGDA) Tools: Main Page

NGDA Tools provide a suite of tools for graphical search and display of geospatial and map digital data; see http://www.ngda.org/research.php.

NGDA/Alexandria Digital Library: ADL Middleware Server

A distributed, peer-to-peer software component that provides mediated access to digital library collections.

- **Developer**: University of California, Santa Barbara
- **NDIIPP projec**t: National Geospatial Digital Archive
- **Written in**: Java and Python
- **OS and run-time environment**: Can be run as a web application inside a servlet container, as an RMI server, or both; has been tested and run under Tomcat in Windows, *nix, and MacOSX. Build requirements include Java and the Apache Ant build tool. Python modules are run inside of Java through an interpreter, so Python is not a requirement.
- **Application**: Not directly accessible to the public. Users can send queries through a user interface, http://clients.alexandria.ucsb.edu/globetrotter/.
- **Documentation**: http://www.alexandria.ucsb.edu/~gjanee/middleware/
- **License**: Open source for noncommercial use with attribution; see source code for details. UCSB uses CVS (Concurrent Versioning System) to store the most up-to-date versions of its code. Anyone interested in downloading the source code for this tool should contact programmers@library.ucsb.edu.

NGDA/Alexandria Digital Library: Globetrotter

A Google Maps-based web client for the Alexandria Digital Library middleware. Globetrotter enables a user to perform spatial searches on spatial data. Users can tune their searches by adjusting a number of different constraints.

- **Developer**: University of California, Santa Barbara
- **NDIIPP project**: National Geospatial Digital Archive
- **Written in**: XHTML, JavaScript, XSLT, and the Velocity Templating language
- **OS and run-time environment**: Runs under Tomcat, tested only on *nix. Build requirements include Java (1.5 or 1.6) and the Apache Ant build tool.
- **Application**: http://clients.alexandria.ucsb.edu/globetrotter/
- **Documentation**: http://clients.alexandria.ucsb.edu/globetrotter/
- **License**: Open source for noncommercial use with attribution; see source code for details. UCSB uses CVS (Concurrent Versioning System) to store the most up-to-date versions of its code. Anyone interested in downloading the source code for this tool should contact programmers@library.ucsb.edu.

NGDA: ArchiveView

A service that provides a consistent, stylized view of objects in the NGDA Archive. XSLT style sheets can be added or customized to change the available views of an object.

- **Developer**: University of California, Santa Barbara
- **NDIIPP project:** National Geospatial Digital Archive
- **Written in**: Java
- **OS and run-time environment**: Powered by a servlet written in Java and requires a servlet container. Tested under Tomcat 5 in Windows and *nix. Build requirements include Java (1.5 or 1.6) and the Apache Ant build tool.
- **Application**: http://www.ngda.org/ArchiveView/
- **Documentation**: Currently documented only within the code

- **License**: Open source for noncommercial applications with attribution. UCSB uses CVS (Concurrent Versioning System) to store the most up-to-date versions of its code. Anyone interested in downloading the source code for this tool should contact programmers@library.ucsb.edu.

NGDA: Bulk Ingest Tool

Used for preparing large collections of data for addition to the archive. After the user has created a template and a configuration file, the Ingest Tool is able to collect files and other data and tie them to an Archive Object identifier. This information is later used to create objects within the archive itself.

- **Developer**: University of California, Santa Barbara
- **NDIIPP project**: National Geospatial Digital Archive
- **Written in**: Java
- **OS and run-time environment**: Uses a MySQL database for persistent data storage. Users must have a user account with write access to a MySQL database. Tested in Windows and *nix. Requires Java (1.5 or 1.6). Current build runs in NetBeans, but code is not dependent on NetBeans as a platform.
- **Application**: An offline tool
- **Documentation**: http://www.ngda.org/research.php
- **License**: Open source for noncommercial applications with attribution. UCSB uses CVS (Concurrent Versioning System) to store the most up-to-date versions of its code. Anyone interested in downloading the source code for this tool should contact programmers@library.ucsb.edu.

NGDA: Format Registry

A wiki-based expert community website for collaborative description of geospatial formats.

- **Developer**: University of California, Santa Barbara
- **NDIIPP project**: National Geospatial Digital Archive
- **Written in**: Built out of the Mediawiki software: http://www.mediawiki.org. Written in PHP.
- **OS and run-time environment**: Tested and released on *nix with Apache 2, PHP 5, and MySQL 5.
- **Application**: http://ngda.library.ucsb.edu/format
- **Documentation**: http://ngda.library.ucsb.edu/format/index.php/Help:The_Process (Community Participation Rules); http://ngda.library.ucsb.edu/format/index.php/FormatRegistry:FlatSpace (FlatSpace Extension)
- **License**: Open source for noncommercial use with attribution. UCSB uses CVS (Concurrent Versioning System) to store the most up-to-date versions of its code. Anyone interested in downloading the source code for this tool should contact programmers@library.ucsb.edu.

NGDA: NGDA Server

Software responsible for the creation of Archive Objects within the archive. Accepts requests with attached data, and properly formats and places those data within the archive.

- **Developer**: University of California, Santa Barbara
- **NDIIPP project**: National Geospatial Digital Archive

- **Written in**: Java, using the Spring Framework, an open-source web application framework that works with any servlet container
- **OS and run-time environment**: Runs in a servlet container. Tested and run under Tomcat 5 in Windows and *nix. Build requirements include Java (1.5 or 1.6) and the Apache Ant build tool.
- **Application**: Not available for public use
- **Documentation**: http://www.ngda.org/research.php
- **License**: Open source for noncommercial applications with attribution. UCSB uses CVS (Concurrent Versioning System) to store the most up-to-date versions of its code. Anyone interested in downloading the source code for this tool should contact programmers@library.ucsb.edu.

NGDA: Workflow Tool

A GUI-based tool for taking items prepared by the Bulk Ingest Tool and inserting them into the archive.

- **Developer**: University of California, Santa Barbara
- **NDIIPP project**: National Geospatial Digital Archive
- **Written in**: Java
- **OS and run-time environment**: Should work on any OS that supports Java. Tested under Windows XP and Ubuntu with Java 1.5 and 1.6.
- **Application**: An offline tool
- **Documentation**: To be posted
- **License**: Open source for noncommercial applications with attribution. UCSB uses CVS (Concurrent Versioning System) to store the most up-to-date versions of its code. Anyone interested in downloading the source code for this tool should contact programmers@library.ucsb.edu.

NutchWAX

Software for indexing ARC files (archived web sites gathered using Heritrix) for full-text search. NutchWAX is based on the open-source web-search software, Nutch.

- **Developer:** Internet Archive
- **NDIIPP project:** Internet Archive
- **Written in:** Java
- **OS and run-time environment:** Platform-independent Java, though tested and primarily used only on Linux machines.
- **Application:** http://archive-access.sourceforge.net/projects/nutchwax/
- **Documentation:** http://archive-access.sourceforge.net/projects/nutchwax/apidocs/overview-summary.html
- **License:** GNU Lesser General Public License 2.1; Nutch itself is under Apache License 2.0. Goal is to merge all NutchWAX functionality into Nutch.

Producer-Archive Workflow Network (PAWN)

A workflow system designed for individuals who have small collections of digital files that need to be processed into preservation systems for management and future access. PAWN is not a long-term archiving or content-management system; rather it is a flexible environment that can map the requirements of different producers into various archival states. It can be used

to provide bulk ingestion from distributed producers into an archive. A gateway to archive a storage resource broker (SRB) environment is already in place; gateways are planned to other commonly used digital management systems (Fedora and dSpace). Components include:

—*Client*: Ingests data, manages users and records organization, and triggers transfer into an archive;

—*Management server*: Tracks accounts, records schedules, records sets, packages lists, and provides security for multiple domains;

—*Scheduler*: Allocates space on a receiving server for the transfer; controls security and configuration for receiving servers; and

—*Receiving server*: Receives data from clients into a package, allows modification of data depending on user credentials, and transfers data to a back-end archive at the direction of an approved user.

- **Developer:** University of Maryland
- **Written in:** Java
- **OS and run-time environment:** Web-based application. Requires Java 1.5 (Java 5) or higher, an account on a PAWN manager, and a keystone to secure traffic through PAWN.
- **Application:** http://adaptwiki.umiacs.umd.edu/twiki/bin/view/Main/PawnDemoClient Client software, keystore and demonstration accounts
- **Documentation:** http://narawiki.umiacs.umd.edu/twiki/bin/view/Main/PAWN

Recollection

Developed by Zepheira in partnership with the Library of Congress, Recollection is a platform which is used to enhance discovery and visualization of NDIIPP collections, making them easier to find, access, and share. NDIIPP Partners can upload their collections to the Recollection platform and create a custom web interface including maps, timelines, tables, and pie chart views. These views can then be immediately published to the Web to share with others. Recollection views can be embedded on other web sites, so partners can share these new visualizations from any site as desired.

- **Developer:** Library of Congress, Zepheira
- **NDIIPP Project: n/a**
- **Written in: Akara, Python, MIT SIMILE, Buildout, Django, Pinax, Freemix Core**
- **OS and run-time environment: n/a - web-based**
- **Application: http://recollection.zepheira.com/**
- **Documentation: http://recollection.zepheira.com/about/userguide/**
- **License: n/a**
- **Last tool update: 16 July 2010**

Replication Monitor and Verification

The Replication Monitor is designed to monitor copies of data in a federated SRB installation. The monitor periodically checks a master site for new data and ensures that copies are created at designated sites. Each replica site operates independently of other sites, ensuring that replication will occur even if the entire data grid is in a degraded state. Extensive logging of any action on data in a collection is provided. In addition, it provides a web interface for quickly reporting the current state of a distributed collection and its copies. Extensions of the current tools to other distributed environments are planned.

- **Developer**: University of Maryland

- **Written in**: Java
- **OS and run-time environment**: Web-based application. Installation requires Tomcat 6.0+, mysql 4.1+ and Java 1.6+.
- **Application**: https://wiki.umiacs.umd.edu/adapt/index.php/Replication:Replication_Monitor_2.0
- **Documentation**: Same as for application
- **License**: Source code TDB; binaries available for download without restriction

Storage Resource Broker (SRB)

SRB is a software tool that allows end-users to organize their digital files in a way meaningful to them, without having to be knowledgeable about the underlying storage technologies. Data may be stored in file systems, tape archives, object-relational databases, and object ring buffers. State information is maintained for each registered entity, enabling uniform access support.

A Metadata Catalog (MCAT) supports retrieval based on queries on attributes instead of physical names or locations. The logical name used to identify a file does not change as the file is moved to other storage systems. The access controls on the file do not change as the file is moved, and the metadata associated with the file remain attached to the file or directory.

- **Developer**: San Diego Supercomputer Center
- **Written in:** SRB servers written in C. The SRB clients are written in the appropriate language: Perl load library, Python load library, Java I/O library, C library calls.
- **OS and run-time environment:** SRB has been ported to UNIX platforms including Linux, Mac OS X, AIX (ex. SP-2 machines), Solaris, SunOS, SGI Irix, and Windows. If you are setting up an MCAT-enabled SRB, you will require an Oracle, DB2, Sybase, mySQL, or PostgreSQL database. The SRB software system itself requires only about 200 MB of storage. For MCAT-enabled servers, the DBMS will require additional space; on Linux, for example, the SRB with PostgreSQL and ODBC take about 700 MB. Any Linux system with a 1.5 GHz CPU should have good performance. Memory size of 1/2 GB or 1 GB is sufficient. For a heavy-load instance of SRB, it is best to use a commercial DBMS like Oracle. PostgreSQL works well for initial testing and light-to-moderate data loads.
- **Application:** http://www.sdsc.edu/srb/index.php/Downloads
- **Documentation:** http://www.sdsc.edu/srb/index.php/Documentation
- **License:** Freely available only to academic organizations and government agencies through a source code distribution. http://www.sdsc.edu/srb/index.php/Client_License

TubeKit

A toolkit for creating YouTube crawlers, TubeKit allows the user to build a tool that can crawl YouTube based on a set of seed queries and collect up to 17 different attributes. TubeKit assists in all the phases of the process, from database creation to browsing and searching interfaces that provide access to the collected data.

- **Developer**: University of North Carolina at Chapel Hill, School of Information and Library Science
- **NDIIPP project:** Vidarch
- **Written in**: PHP; web-based
- **OS and run-time environment**: N/A; web-based

- **Application**: http://www.tubekit.org/download.php
- **Documentation**: http://www.tubekit.org/index.php
- **License**: N/A

Wayback Machine

A powerful search and discovery tool for use with collections of website "snapshots" collected through web harvesting, usually with Heritrix (ARC or WARC files).

- **Developer:** Internet Archive
- **NDIIPP project:** Internet Archive
- **Written in:** Java
- **OS and run-time environment:** Platform independent. Wayback has been successfully tested on Tomcat, an Apache.org Java-based web server.
- **Application:** http://archive-access.sourceforge.net/projects/wayback/
- **Documentation:** http://archive-access.sourceforge.net/projects/wayback/administrator_manual.html
- **License:** GNU Lesser General Public License 2.1 (http://archive-access.sourceforge.net/projects/wayback/license.html); migrating to Apache License 2.0 in future

Web Archives Workbench

A suite of web capture tools based on principles of managing archived content in aggregates rather than as individual objects. The suite comprises:

—*Discovery Tool*, which helps identify potentially relevant websites by crawling relevant "seed" entry points to generate a list of domains to which they link;

—*Properties Tool*, which enables users to maintain information about content creators, associate them with the websites they are responsible for, and enter high-level metadata;

—*Analysis Tool*, which permits user to look at the structure of the website to see what kind of content is represented by the file directory; and

—*Harvest Tool*, which allows user to monitor crawl status, review and modify harvest settings, and package harvests for transfer to a repository. Also offers a separate Quick Harvest feature that schedules one-time harvests of content. Harvest packages are encoded in METS with Dublin Core metadata embedded.

- **Developer**: OCLC
- **NDIIPP project:** ECHO DEPository: Exploring Collaborations to Harness Objects with a Digital Environment for Preservation
- **Written in**: Java, JavaScript, JSP
- **OS and run-time environment**: Linux
- **Application**: Download from SourceForge, http://sourceforge.net/projects/webarchivwkbnch
- **Documentation**: Available on SourceForge
- **License**: Available on SourceForge

Web Archiving Service (WAS)

WAS is a web-based curatorial tool that enables libraries and archivists to capture, curate, analyze, and preserve web-based government and political information. The WAS allows users to set parameters of web crawls, capture sites, provide metadata for archived sites, and build collections of archived websites.

- **Developer:** California Digital Library
- **NDIIPP project:** Web-at-Risk
- **Written in:** Java, Ruby on Rails
- **OS and run-time environment**: Web Page: Javascript must be enabled in the user's browser. User must be able to install browser bookmarklets to use the "add sites while browsing" feature. Log in and password required. Back End: Infrastructure consists of Solaris 10 and Linux machines. The heaviest infrastructure demands are processing power for crawling, processing power for indexing, and storage. Other tools used are Heritrix, NutchWAX, Open Source Wayback Machine, MySQL and Storage Resource Broker.
- **Application:** http://webarchives.cdlib.org
- **Documentation:** http://was.cdlib.org
- **License:** N/A

Web Harvester

A service that enables users to harvest content from the web, review it, and add the harvested items to their CONTENTdm® collections during the Connexion cataloging process. By integrating digital collection development and capture with standard cataloging workflows, the Web Harvester provides an additional option for expanding participation in growing and maintaining digital collections.

Harvested items added to CONTENTdm Digital Collection Management Software using the Web Harvester are discoverable from the CONTENTdm web interface, as well as WorldCat.org, WorldCat Local, and OCLC FirstSearch. Each harvested item added to CONTENTdm using the Web Harvester is associated with its WorldCat record via a persistent URL based on the OCLC number of the WorldCat record. With an additional subscription to the OCLC Digital Archive, master files will be automatically placed in the Archive's secure, managed storage system.

- **Developer**: OCLC
- **Written in**: Java
- **OS and run-time environments**: Linux, MySQL, Apache, Tomcat, Heritrix
- **Application**: http://oclc.org/webharvester
- **Documentation**: http://www.oclc.org/webharvester/support/default.htm
- **License**: Fee-based

Appendix D

NDIIPP Partner Collections

A key aim of the National Digital Information Infrastructure and Preservation Program (NDIIPP) is to identify digital content that is at risk of loss and bring that content into the care of stewardship organizations. The body of at-risk digital content is growing at an estimated rate of tenfold annually. More than 1,000 collections have been the target of preservation action by NDIIPP-sponsored projects. More than half of the current collections are concerned with government, politics, and law, and provide rich research opportunities and support for public policy development.

The collections have been grouped into the following National Digital Collection framework categories:

- Arts and culture
- Government, politics, and law
- Maps and geography
- News, media, and journalism
- Religion and philosophy
- Social sciences
- Science, mathematics, and technology
- World history and cultures

ARTS AND CULTURE

MetaArchive of Southern Digital Culture

Partner institution: Auburn University
Type: Text and/or Image
Project: MetaArchive Cooperative
Description: Digital materials of relevance to the study of Southern cultures and histories. Auburn University collections preserved are: Alabama Cooperative Extension Service (ACES) Photographs 1920s–1960s; Auburn University Numbered Photographs Collection; Auburn University Sesquicentennial Lecture Series; Glomerata: Auburn University Yearbooks, 1897–; Auburn Football Programs Photos; Caroline Dean Wildflower Collection; Eugene B. Sledge Collection.

Partner institution: Emory University
Type: Text and/or Image, Audio/Video
Project: MetaArchive Cooperative
Description: Digital materials of relevance to the study of Southern cultures and histories. Emory University collections preserved are: the e-journal Southern Spaces with Masters; Sam Nunn Constituent Mail System Files; Special Collections and Archives Digital Image Master Files;

Southern Changes; The Civil War in America from the Illustrated London News; Special Collections Sound Recordings; University Archives Sound Recordings; Conspectus Database; Music of Social Change; the Famous Guy.

Partner institution: Florida State University (FSU)
Type: Text and/or Image
Project: MetaArchive Cooperative

Description: Digital materials of relevance to the study of Southern cultures and histories. Florida State University collections preserved are: Digitized Juvenile Literature; FSU "Flying High" Circus Collection Photographs; FSU Biological Scientist, Dr. A.K.S.K. Prasad Diatomscapes I and II Collections Photographs; FSU Claude Pepper Library Historical Photos Collection Photographs; FSU Cuneiform Tablets Collection Photographs; FSU Department of Oceanography Technical Reports; FSU Dissertations 1952–2002; FSU ETDs 2003–Present; FSU Heritage Protocol Collection; FSU Historic Photograph Collection; FSU Historic Theses Collection; FSU History and Heritage Rare Books; FSU Napoleonic Era Collection; FSU Provost Dr. Lawrence Abele Collection; FSU Provost Dr. Lawrence Abele Collection Photographs; FSU Special Collections; FSU Undergraduate Honors in the Major Theses 2004–2008.

Partner institution: Georgia Institute of Technology
Type: Text and/or Image, Audio/Video
Project: MetaArchive Cooperative

Description: Digital materials of relevance to the study of Southern cultures and histories. Georgia Institute of Technology collections preserved are: A Photographic Atlas of Selected Regions of the Milky Way; George Griffin Photograph Collection; Georgia Tech Advertisements; "Splendid Growth": Architectural Drawings of the A. French Textile Building; The Buildings of Georgia Tech from 1888–1908; Photographs of the Historic American Buildings Survey Georgia; An Illustration and Measuration of Solid Geometry; Deceased Faculty Biographies; Georgia Tech Photograph Collection; Georgia Tech Publications; SMARTech; Aardvark.

Partner institution: University of Louisville
Type: Text and/or Image, Geospatial
Project: MetaArchive Cooperative

Description: Digital materials of relevance to the study of Southern cultures and histories. University of Louisville collections preserved are: Bernheim Foundation interviews; Kentucky Quilt Project image masters; Jean Thomas Collection; African American Oral History Collection; Kate Matthews Collection; Kentucky Maps; Newton Owen Postcard Collection; The Herald-Post Collection; Macauley's Theater Collection; Arthur Younger Ford (1861–1926) Photograph Albums.

Partner institution: Virginia Polytechnic University
Type: Text and/or Image, Audio/Video
Project: MetaArchive Cooperative

Description: Digital materials of relevance to the study of Southern cultures and histories. Virginia Tech collections included are: First International Students at Virginia Tech; "My Precious Loulie ...": Love Letters of the Civil War; The History of Blacksburg, Virginia; American Indian History at Virginia Tech; Ann Hertzler, Professor Emerita; Annual Reports of the Digital Library and Archives and the University Libraries; BiblioTech; Black History at Virginia Tech; Black Women at Virginia Tech; Blacksburg's Bicentennial 1798–1998; Book for Receipts–1731; Bugle Archive; Campus Unrest at Virginia Tech–1970; Center for Ulster Migrations, Cultures, and Societies; Christiansburg Church Minutes Book; Collegiate Times Archive; CONTENTdm Test Collections at Virginia Tech; Cooking Recipes; ETDs@ VT; Fenwick Collection; Ferry Hill Ledger; George Marvin Scrapbook; German Soldier Scrapbook; Graduate School Reporter; H. E. Valentine Scrapbook; Historical Virginia Tech; History of Architecture Catalogue for Hypertext; International Archive of Women in Architecture Biographical Database; International Archive of Women in Architecture Image Base Metadata; John Hilton Papers; John McLaren McBryde (1891–1907); Joseph Dupuy Eggleston (1913–1919); Library Friends; Mapping the Blues Genes, Early Blues Music: 1900–1930; Margaret Morris–The April 16th Project; Mountain Slavery; Nancy Figgat Recipe Book; Nutrition Central; Papers of Judge Wm. M. Harris; Paul Brandon Barringer (1907–1913); Pre-World War II Thanksgiving at VPI; Principalship Project: Oral History of the Public School Principalship; Raymond Dessy (in Faculty Archives); Receipts and Home Remedies circa 1869; Research Publications Virginia Agricultural Experiment Station; *Roanoke Times*; Solitude–The Future Home of the Appalachian Center; Song of the Mountain; South Atlantic Humanities Center; South Atlantic Humanities Center–Current; Special Collections at the Virtual Library of Virginia; Spectrum; Technology for all Americans; Thomas Marshall Hahn (1962–1974); Tin Horn; VPI in the Spanish American War; Virginia Agricultural & Mechanical College Catalogue; Virginia Libraries; Virginia Tech Conductor; Virginian Pilot; VPI & SU Historical Data Book Centennial Edition; VPI Cadet Uniforms Image File; VT Corps Female Cadet Scrapbook; VT Staff Employees of the Week; Walter S. Newman (1947–1962); WDBJ7; White Sulphur Springs Ledger; Women, Work and Family in the Antebellum Mountain South; Yellow Sulphur Springs Ledger.

GOVERNMENT, POLITICS, AND LAW

AFL-CIO/Change to Win Web Archive

Partner institution: Institute of Industrial Relations Library, University of California, Berkeley

Type: Websites

Project: Web-at-Risk

Description: Web collections of materials from labor organizations that capture transitions in social, workforce, and work-and-family trends of the U.S. workforce.

Alaska State Government Electronic Documents

Partner institution: Alaska State Archives
Type: Text and/or Image, Audio/Video, Websites
Project: Multi-State Preservation Consortium Utilizing the Washington State Digital Archives Framework
Description: Vital records, land ownership and use documentation, court records, and web-based state and local government reports and other electronic government documents.

Arizona Government Digital Collections

Partner institution: Arizona State Library, Archives, and Public Records
Type: Text and/or Image, Websites
Project: Persistent Digital Archives and Library System (PeDALS) Project
Description: Digital content from all branches of Arizona state government, including marriage certificates, civil case files, well reports, agency and legislative web pages, proceedings of the Arizona House and Senate, state agency publications, and e-mails of the state's governor.

Arizona State Agencies Web Publications

Partner institution: Arizona State Library
Type: Websites
Project: Persistent Digital Archives and Library System (PeDALS) Project
Description: Website collections of the Territorial and State Agency Publications Depository Program. Arizona history, law, and genealogy are some key subject areas covered.

California Political Blogs and Interest Group Websites

Partner institution: Institute of Governmental Studies Library, University of California, Berkeley
Type: Websites
Project: Web-at-Risk
Description: California politics as represented in political blogs and interest group websites.

Colorado State Government Electronic Documents

Partner institution: Colorado State Library
Type: Text and/or Image
Project: Multi-State Preservation Consortium Utilizing the Washington State Digital Archives Framework
Description: Colorado historical census information; birth, death, and divorce records; bar admissions; land patents and land records; lot and block index; Gilpin court filing and county bankruptcy; tax lists; school records; wills; and voter registration.

CyberCemetery

Partner institution: University of North Texas
Type: Websites
Project: Web-at-Risk
Description: Government websites that have ceased operation (usually

websites of defunct government agencies and commissions that have issued a final report).

Federal Budget Materials

Partner institution: University of California, Berkeley
Type: Websites
Project: Web-at-Risk
Description: Websites that cover economics, governmental studies, political science, public policy, and sociology.

Florida State Government Digital Collections

Partner institution: Florida State Archives
Type: Text and/or Image, Websites, Audio/Visual
Project: Persistent Digital Archives and Library System (PeDALS) Project
Description: Audio recordings of Florida House of Representative debates, Florida Senate debates, and committee sessions. Photographs of Governor Bush, constituent correspondence, the Governor's weekly newsletters, analysis files from the Office of Policy and Budget, and press releases and statements. Rules of Civil Procedures, Rules of Judicial Administration, and Rules of Criminal Procedure.

Idaho State Government Electronic Documents

Partner institution: Idaho State Library
Type: Text and/or Image
Project: Multi-State Preservation Consortium Utilizing the Washington State Digital Archives Framework
Description: Alturas County mining claims; Idaho Department of Correction records; Idaho Bureau of Homeland Security; Idaho naturalization records; Idaho State Guard; Nez Perce County District Court; Old Age Pensions; and Washington County District Court civil cases.

Indiana State Government Electronic Documents

Partner institution: Indiana State Archives
Type: Text and/or Image
Project: Multi-State Preservation Consortium Utilizing the Washington State Digital Archives Framework
Description: State of Indiana Social Security Death Index collection.

Local Government and Local Area Flood Control Collection

Partner institution: University of California, Davis
Type: Websites
Project: Web-at-Risk
Description: Local government websites of the City of Davis, Yolo County, Sacramento city and county government with emphasis on flood control levees and dams in the Sacramento area. General websites relating to water resources, water control, and water agencies in Northern California and from regional divisions of federal agencies.

Louisiana State Government Electronic Documents

Partner institution: Louisiana State Archives
Type: Text and/or Image
Project: Multi-State Preservation Consortium Utilizing the Washington State Digital Archives Framework
Description: State of Louisiana Social Security Death Index collection.

Montana State Government Electronic Documents

Partner institution: Montana Historical Society
Type: Text and/or Image, Audio/Video
Project: Multi-State Preservation Consortium Utilizing the Washington State Digital Archives Framework
Description: Audio recordings of the Montana Joint Legislative Administration–Interim Committee meeting recordings and a Social Security Death Index collection.

Monterey Bay Area Local and Regional Government Websites

Partner institution: University of California, Santa Cruz
Type: Websites
Project: Web-at-Risk
Description: Local and regional government websites for the Monterey Bay area, including the counties of Santa Cruz, Monterey, San Benito, San Mateo, and Santa Clara. Also included are websites for federal regional agencies with jurisdiction in these counties, and for special districts, city governments, and California regional agencies such as pollution control districts, in addition to nongovernmental non-profit organizations with missions that align themselves with local government agencies, for example, the Land Trust of Santa Cruz County.

Multi-Media Legislative Data

Partner institutions: Minnesota Historical Society, the Minnesota Office of the Revisor of Statutes, and the Minnesota Legislative Reference Library
Type: Audio/Video
Project: A Model Technological and Social Architecture for the Preservation of State Government Digital Information project
Description: Multimedia legislative content, including legislative audio files from the state legislatures of Tennessee, Kansas, and Minnesota.

New Mexico State Archives Electronic Documents

Partner institutions: New Mexico State Archives, New Mexico State Library
Type: Text and/or Image, Audio/Video, Websites
Project: Multi-State Preservation Consortium Utilizing the Washington State Digital Archives Framework
Description: Vital records, land ownership and use documentation, court records, and web-based state and local government reports and other electronic government documents.

New York State Government Digital Collections

Partner institution: New York State Archives
Type: Text and/or Image, Websites
Project: Persistent Digital Archives and Library System (PeDALS) Project
Description: State government reports, policies, laws and legislative histories, rules and regulations, and administrative and judicial hearings and decisions.

Orange County Government Information Web Collection

Partner institution: University of California, Irvine
Type: Websites
Project: Web-at-Risk
Description: Government materials include online material issued by Orange County, city governments, special districts, and active regional agencies with emphasis on budgets, general plans, and key government department publications that supply development, social, political, and economic data about the area. In addition, nongovernment materials issued by "special interest groups" operating within the county that produce relevant reports are included.

Oregon State Government Electronic Documents

Partner institutions: Oregon State Archives, Oregon State Library,
Type: Text and/or Image, Websites
Project: Multi-State Preservation Consortium Utilizing the Washington State Digital Archives Framework
Description: Records from the Oregon Board of Medical Examiners; Carbon Allocation Task Force; Clandestine Drug Lab Cleanup Program; Computing and Network Infrastructure Consolidation Governing Board; Joint Finance and Technology Subcommittee; Construction Claims Task Force; a variety of records from the Bureau of Labor and Industries; Center for Health Statistics; Coastal Management Program; Commission on Children and Families; Forest Resources Institute; Emergency Management; Geospatial Enterprise Office; Health Services Commission; State Boards of Forestry and Nursing; Task Force and Land Use Planning; Transportation Commission; Oregon Vital Records; Oregon Watershed Enhancement Board; the Secretary of State and many other collections.

Riverside California Inland Empire Web Archive

Partner institution: University of California, Riverside Libraries
Type: Websites
Project: Web-at-Risk
Description: Websites that record Inland Empire growth and development along with the challenges associated with that growth. Focuses of the archive are: city/county/regional planning documents; land use; transportation; water supply, consumption, and quality; and air quality and pollution.

San Diego Area Local Government Information Archive

Partner institution: University of California, San Diego (Web-at-Risk project)
Type: Websites
Project: Web-at-Risk

Description: Materials from government agencies in San Diego County (e.g., City of San Diego, County of San Diego) and semiofficial (e.g., League of Women Voters of San Diego) information related to topics of local interest. Web publications represent a variety of documents, including administrative materials, reports, statistics, data, and maps.

South Carolina State Government Digital Collections

Partner institution: South Carolina (SC) Department of Archives and History
Type: Text and/or Image
Project: Persistent Digital Archives and Library System (PeDALS) Project

Description: State Senate journals, State House of Representatives journals, SC Code of Laws and Regulations, published opinions from the State Supreme Court, orders from the Public Service Commission, voter registrations from the State Election Commission, death certificate indexes from the Department of Health and Environmental Control, admission/discharge records from Department of Corrections, incorporation documents from the Secretary of State, trademark records from the Secretary of State, e-mail from the Governor's Office, Order of the Palmetto database from the Governor's Office, and meeting minutes of professional licensing boards from the Department of Labor, Licensing, and Regulation Special Subjects.

State Library Web Collection

Partner institutions: Various state libraries (AZ, CT, IL, NC, WI)
Type: Websites
Project: ECHO DEPository: Exploring Collaborations to Harness Objects with a Digital Environment for Preservation

Description: Electronic publications of state governments and state agencies. This content contains the historic record of government. This record is now beginning to exist only in electronic form, which state libraries and archives are mandated to collect.

State of California Legislative Data

Partner institutions: California Digital Library, California State Library, Legislative Counsel of California
Type: Text and/or Image
Project: A Model Technological and Social Architecture for the Preservation of State Government Digital Information project

Description: Legislative content, including bills, acts, mandated reports, and house and senate journals. California content will be coordinated by the Legislative Counsel of California. Final edited session laws, final edited statutes, and administrative rules with all associated text, tables, indexes, and graphic figures will be included.

State of California Legislative Data and Reports

Partner institutions: California Digital Library, California State Library, Legislative Counsel of California

Type: Websites

Project: A Model Technological and Social Architecture for the Preservation of State Government Digital Information project

Description: Web resources that relate to the content of a California legislative bill, including reports, schedules, meeting minutes, and/or journals that complement the content of bill-drafting systems and elaborate on the records of the legislature.

State of Kansas Legislative Data

Partner institutions: Kansas State Historical Society, Kansas Legislative Computer Services

Type: Websites

Project: A Model Technological and Social Architecture for the Preservation of State Government Digital Information project

Description: Web resources that relate to the content of a Kansas legislative bill, including reports, schedules, meeting minutes, and/or journals that complement the content of bill-drafting systems and elaborate on the records of the legislature.

State of Kansas Legislative Data

Partner institutions: Kansas State Historical Society, Kansas Legislative Computer Services

Type: Text and/or Image

Project: A Model Technological and Social Architecture for the Preservation of State Government Digital Information project

Description: Legislative content, including bills, acts, mandated reports, and house and senate journals. Final edited session laws, final edited statutes, and administrative rules with all associated text, tables, indexes, and graphic figures will be included.

State of Minnesota Legislative Data

Partner institutions: Minnesota Historical Society, the Minnesota Office of the Revisor of Statutes, and the Minnesota Legislative Reference Library

Type: Text and/or Image

Project: A Model Technological and Social Architecture for the Preservation of State Government Digital Information Project

Description: Legislative content, including bills, acts, mandated reports, and house and senate journals supplied by the Minnesota Office of the Revisor of Statutes. Final edited session laws, final edited statutes, and administrative rules with all associated text, tables, indexes, and graphic figures will be included.

State of Minnesota Legislative Data and Reports

Partner institutions: Minnesota Historical Society, the Minnesota Office of the Revisor of Statutes, and the Minnesota Legislative Reference Library

Type: Websites

Project: A Model Technological and Social Architecture for the Preservation of State Government Digital Information project

Description: Web resources that relate to the content of a Minnesota legislative bill, including reports, schedules, meeting minutes, and/or journals that complement the content of bill-drafting systems and elaborate on the records of the legislature.

Tamiment Library at New York University

Partner institution: New York University
Type: Websites
Project: Web-at-Risk

Description: Political communication of left activists and labor organizations that are engaged in the struggle for progressive social and political change, including social movements, labor unions, antiglobalization groups, radical arts and culture, and left-media organizations.

UCLA Nongovernmental and Local Government Information

Partner institution: UCLA Library
Type: Websites
Project: Web-at-Risk

Description: Los Angeles municipal, county, and regional government information including nongovernmental materials (e.g., nonprofits, policy institutes, consultant firms, citizen groups) that are integral resources at the local level.

UCLA Online Campaign Literature Archive

Partner institution: University of California, Los Angeles
Type: Websites
Project: Web-at-Risk

Description: Campaign websites created for elections affecting the Los Angeles area, including those for local, state, and federal offices and ballot measures.

University of California, Santa Barbara, Ventura, and San Luis Obispo Counties Local Planning Documents and Water Archive

Partner institution: University of California, Santa Barbara
Type: Websites
Project: Web-at-Risk

Description: Government resources for Santa Barbara, San Luis Obispo, and Ventura counties. In particular, reports and documents from local planning agencies, as well as sites from U.S. and California government agencies, relating to water issues for these counties.

Washington State Government Electronic Documents

Partner institution: Washington State Library
Type: Text and/or Image, Websites
Project: Multi-State Preservation Consortium Utilizing the Washington State Digital Archives Framework

Description: Washington historic census, justice, and

county history information; electronic publications from Department of Agriculture, Department of Fish and Wildlife, Department of Early Learning, and many other key state departments; records from the Biodiversity Council; Community Trade and Economic Development; Department of Corrections records; birth, death, and divorce records; records from the Economic and Revenue Forecast Council; the Education Research and Data Center; historic photos of the House of Representatives; county land records; military records; audio and video recordings; legislative proceedings; city council minutes; and many others.

Wisconsin State Government Digital Collections

Partner institution: Wisconsin Historical Society
Type: Text and/or Image, Geospatial
Project: Persistent Digital Archives and Library System (PeDALS) Project
Description: Correspondence from the Secretary of Natural Resources, correspondence from the Department of Workforce Development, employer record system employment data, Department of Workforce Development Public information news releases, minutes from the Department of Health and Family Services, selected speeches and public comments of Governor Tommy G. Thompson (1993–1996), selected speeches and public comments of Governor Scott McCallum (2001–2002), railroad maps, the index to building permit records, and bill-drafting files.

MAPS AND GEOGRAPHY

Aerial Photographs Collection

Partner institution: University of Illinois at Urbana-Champaign Map and Geography Library
Type: Text and/or Image
Project: ECHO DEPository: Exploring Collaborations to Harness Objects with a Digital Environment for Preservation
Description: A digital subset of the Library's aerial photograph collection: 239 photographs, scanned at 720dpi, from 1939–1954 of Will County. Also includes Champaign, Cook, Fulton, Mason, and Peoria counties. Commissioned by the United States Department of Agriculture in the 1930s–1950s, this collection provides a unique physical history of Illinois lands, showing urban and rural development in Illinois beginning in the 1930s.

Digital Maps in the Electronic Records Archives

Partner institution: Kentucky Department for Libraries and Archives
Type: Geospatial
Project: NC GeoMAPP project
Description: A wide variety of digital maps from the state of Kentucky. Many are digitized historic maps and represent intellectual content and meaning beyond that found in the underlying data sets. The maps are a result of a number of processes including data layer selection and ordering, symbolization, classification, output of data models, and annotation.

Digital Orthophoto Quadrangle (DOQ) Collection

Partner institution: US Geological Survey/ Illinois State Geological Survey via National Center for Supercomputing Applications

Type: Geospatial

Project: ECHO DEPository: Exploring Collaborations to Harness Objects with a Digital Environment for Preservation

Description: Digital orthoimagery of Lake County, IL, which may be used for many large-scale mapping projects such as urban planning for smart growth, environmental assessment, crop science, agriculture, geology, soils science, watershed management, geography, landscape architecture, civil engineering, biology, and pollution prevention. This collection forms part of the national orthoimagery coverage of the United States.

KY GeoNet

Partner institution: Kentucky Commonwealth Office of Technology, Division of Geographic Information

Type: Geospatial

Project: NC GeoMAPP project

Description: A variety of data sets and static map products and images will be made available through the Geospatial Data Clearinghouse for the Commonwealth of Kentucky. The content includes periodic snapshots (2008–present) of the KyVector database, the master repository of vector-based data in Kentucky, and is based on all vector data that are available on the Kentucky Geography Network. These data describe government or administrative boundaries, infrastructure (e.g., streets, water/sewer, industry sites), and/or the physical environment (e.g., streams and elevation). Images appropriate to the period covered in the raster data sets are maintained in an archive. The database contains all the aerial images, topographic maps, digital elevation models, hillshade, SPOT satellite imagery, trucolor imagery, land cover imagery, slope, and other critical raster GIS base layers in the Kentucky system.

National Digital Geospatial Archive

Partner institutions: University of California and Stanford University

Type: Geospatial

Project: National Digital Geospatial Archive project

Description: The collections include early maps from the David Rumsey Map Collection/Cartography Associates; geospatial information for California from the California Spatial Information Library; geospatial information for the United States as a whole from Novacell and the U.S. Geological Survey; and original scanned aerial photography from the Alexandria Digital Library. The archive also includes the Stanford

[1] Feature data are vector data resources that model features on the earth's surface as points, lines, or polygons and may be output as maps or displayed and analyzed. County government data resources are typically developed in association with tax assessment and emergency response functions and so usually prioritize data layers such as land parcels, street centerlines, and jurisdictional boundaries.

[2] Digital orthophotography produces essentially photographic maps. County government orthophoto data are generally of much higher resolution than the state/federal data.

[3] A wide variety of digital maps are being acquired. Local government data are typically more detailed, more current, and more accurate than state or federal data. This type of data is currently being acquired as part of the Homeland Security Information Program.

Geological Survey Collection containing thousands of maps and hundreds of notebooks spanning the 100-year life of the survey.

North Carolina (NC) County Geospatial Data

Partner institutions: North Carolina State University Libraries and North Carolina Center for Geographic Information & Analysis
Type: Geospatial
Project: NC Geospatial Data Archiving Project
Description: Preserved GIS systems have been developed in 99 of 100 NC counties. Data have been acquired from over 70 counties. The data set includes feature data,[1] digital orthophotography,[2] and digital maps.[3]

North Carolina (NC) Municipal Geospatial Data

Partner institutions: North Carolina State University Libraries and North Carolina Center for Geographic Information & Analysis
Type: Geospatial
Project: NC Geospatial Data Archiving Project
Description: Feature data and digital maps from many municipalities in North Carolina.

North Carolina (NC) State Agency Data

Partner institutions: North Carolina State University Libraries and North Carolina Center for Geographic Information & Analysis
Type: Geospatial
Project: NC Geospatial Data Archiving Project
Description: Feature data and digital maps from more than 20 state agencies in North Carolina, including the Department of Transportation, the Department of Environment and Natural Resources, the North Carolina Flood Mapping Program, the Department of Health and Human Services, and the Department of Public Instruction.

Selected Items from the Utah State Geographic Information Database

Partner institution: Utah Automated Geographic Reference Center
Type: Geospatial
Project: NC GeoMAPP project
Description: A variety of data sets and static map products and images from the state of Utah. These include local data sets such as Salt Lake and Davis County parcels, zones, and municipality data; orthoimagery from Salt Lake County; centralized data sets, including framework, biota, inland waters, flooding, structures, transportation, and environmental data; project files, including those used to supply evidence during the debate on Utah's Enhanced Drug Penalty Zone Law; and digitized maps, including the Salt Lake County 7.5-minute quadrangles.

NEWS, MEDIA, AND JOURNALISM

Foreign News Broadcasts

Partner institution: SCOLA (Satellite Communications for Learning)
Type: Audio/video
Project: Preserving Foreign Television News Broadcasts
Description: 8,000 high-interest hours of foreign news broadcasts per year from 30 countries, including Qatar (Al-Jazeera) Iran, Pakistan, Russia, Egypt, South Africa, and the Philippines.

Frontline: Selections from Public Television National Productions

Partner institutions: Thirteen/WNET-TV, New York; WGBH-TV, Boston, Public Broadcasting Service
Type: Audio/video
Project: Preserving Digital Public Television
Description: Signature documentary and current affairs programs on the national public television prime time lineup. Selections from the series *Frontline* include in-depth investigative reports on political issues. This content is at risk because there is no clear designated entity that has responsibility for program preservation, so no resources or procedures are in place to preserve programs after broadcast unless the producer takes steps independently.

Vincent Voice Library Collection

Partner institution: Michigan State University Libraries
Type: Audio/video
Project: ECHO DEPository: Exploring Collaborations to Harness Objects with a Digital Environment for Preservation
Description: Primary source sound material, found mainly in speech, interview, lecture, and performance formats. Content donated to the ECHO DEPository project at the University of Illinois at Urbana-Champaign is in the public domain, and includes material from the Vincent Voice 20th Century U.S. Presidents gallery, the Geographers gallery, and the Michigan Writers gallery.

WILL AM-FM-TV Collection

Partner institutions: WILL AM-FM-TV, University of Illinois at Urbana-Champaign College of Communications
Type: Audio/video
Project: ECHO DEPository: Exploring Collaborations to Harness Objects with a Digital Environment for Preservation
Description: *Focus-580* radio program, containing interviews with newsmakers and experts on international affairs and daily life; includes topics in news, public affairs, higher education, and culture. The collection reflects the output of local programming on the public radio and television station at the University of Illinois during a time of rapid political, cultural, and technological change. It contains many items documenting scholarly work at the university, and analyses of local, regional, national, and international affairs.

RELIGION AND PHILOSOPHY

***Religion and Ethics Newsweekly*: Selections from Public Television National Productions**

Partner institutions: Thirteen/WNET-TV, New York; WGBH-TV, Boston, Public Broadcasting Service

Type: Audio/video

Project: Preserving Digital Public Television

Description: Signature documentary and current affairs programs on the national public television prime time lineup. Selections are from the series *Religion & Ethics Newsweekly*, a current affairs magazine program reporting on news stories related to religion. This content is at risk because no clearly designated entity has responsibility for program preservation, so no resources or procedures are in place to preserve programs after broadcast unless the producer takes steps independently.

SOCIAL SCIENCES

Data Preservation Alliance for the Social Sciences

Partner institution: Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan

Type: Text and/or Image

Project: Data-PASS

Description: Federally funded social science research, especially data collections funded since the 1970s by the National Science Foundation and the National Institutes of Health. Data-PASS has preserved more than 800 significant, at-risk data collections built on the partnering archives' experience and expertise.

Data Preservation Alliance for the Social Sciences

Partner institution: Roper Center for Public Opinion Research at the University of Connecticut

Type: Text and/or Image

Project: Data-PASS

Description: Public opinion survey data conducted on behalf of the U.S. Information Agency Office of Research from 1952 through 1999, pioneering National Opinion Research Center surveys from the 1950s and 1960s, and opinion surveys conducted by private research organizations such as the Public Agenda Foundation and AARP.

Data Preservation Alliance for the Social Sciences

Partner institution: Howard W. Odum Institute at the University of North Carolina-Chapel Hill

Type: Text and/or Image

Project: Data-PASS

Description: Harris Poll data and data collections from the National Network of State Polls, a confederation of organizations that conduct state-level public opinion surveys.

Data Preservation Alliance for the Social Sciences

Partner institutions: Henry A. Murray Research Archive, Harvard-MIT Data Center, members of the Institute for Quantitative Social Science at Harvard University

Type: Text and/or Image

Project: Data-PASS

Description: Numeric data, studies on women, studies with diverse samples, and longitudinal studies, such as the landmark Longitudinal Study of Personality Development.

Data Preservation Alliance for the Social Sciences

Partner institutions: The custodial electronic records division of the National Archives and Records Administration; Roper Center for Public Opinion Research at the University of Connecticut

Type: Text and/or Image

Project: Data-PASS

Description: USIA surveys and related documentation.

SCIENCE, MATHEMATICS, AND TECHNOLOGY

DeLIver Engineering Journal Database Collection

Partner institutions: University of Illinois at Urbana-Champaign and participating publishers (American Institute of Physics; American Physical Society; American Society of Civil Engineers; Institution of Electrical Engineers; Corporation for National Research Initiatives; Naval Research Laboratory; NTT J-Stage; Association for Computing Machinery; American Society for Materials; Elsevier Science; Institute of Electrical and Electronics Engineers Computer Society)

Type: Text and/or Image

Project: ECHO Depository project

Description: Articles from more than 50 journals, 1995–present, published by partners listed above.

Digital Archive of the Birth of the Dot Com Era

Partner institutions: University of Maryland, Robert H. Smith School of Business; Center for History and New Media, George Mason University; Gallivan, Gallivan & O'Melia LLC; Morrison & Foerster, LLP; Ropers Majeski Kohn & Bentley PC

Type: Text and/or Image, Audio/Video

Project: Birth of the Dot Com Era

Description: At-risk digital materials from the American business culture during the early years of the commercialization of the Internet, from 1994 to 2001, including business, marketing, and technical plans; venture presentations; and other business documents from more than 2,000 failed and successful Internet startups.

Nature: Selections from Public Television National Productions

Partner institutions: Thirteen/WNET-TV, New York; WGBH-TV, Boston, Public Broadcasting Service
Type: Audio/video
Project: Preserving Digital Public Television
Description: Signature documentary and current affairs programs on the national public television prime time lineup. Selections from the series *Nature* include topics of natural history, animals, and the environment.

WORLD HISTORY AND CULTURES

International Governmental Organizations and Developing Countries

Partner institution: University of California, Berkeley
Type: Websites
Project: Web-at-Risk Project
Description: Websites from the areas of (1) international economic development, (2) demographics and population growth, (3) poverty and developing countries, and (4) HIV/AIDS.

Islamic and Middle Eastern Political Web

Partner institution: Stanford University
Type: Websites
Project: Web-at-Risk Project
Description: Websites of political parties and dissident groups from Islamic and Middle Eastern countries. The collection of all relevant web materials in all formats in support of teaching and research in Islamic and Middle Eastern Studies at Stanford University.

Appendix E

Proposals for the Creation of a Public Policy Environment Conducive to Digital Preservation

By Mary E. Rasenberger[1]

1. BACKGROUND: PUBLIC POLICY AND INCENTIVES FOR DIGITAL PRESERVATION

> *"What is at stake is the transmission of ideas, knowledge and the American people's legacy of creativity to future generations."*[2]

Since the early days of our Republic, legislators have enacted policies and created institutions and laws to support the preservation of our cultural and intellectual heritage. Like education and infrastructure, preservation is an area in which the public's stake is high, yet the financial benefits to the private sector are too low to rely on private sector investment alone. Such public goods generally require the support of public programs and policies. As described in this paper, Congress already has crafted many creative tools, institutions, laws, regulations, and precedents over the history of the Republic to support the public goals of preserving our cultural and intellectual resources—but these are in urgent need of adaptation for the digital environment.

Congress has created institutions, including the Library of Congress, the National Archives, and the Smithsonian Institution, as well as a number of programs and organizations responsible for ensuring that future generations have access to our nation's creative and intellectual output.[3] Congress has also enacted various laws to support preservation of cultural heritage, including requirements to deposit published creative materials with the Copyright Office in order to create a national record of such materials with the Library of Congress. Just as important are the incentives for private investment in preservation provided through tax law, for instance, by allowing cultural memory institutions to obtain not-for-profit tax status and by providing tax credits and deductions for historic preservation and donations to cultural institutions. These actions have created a public policy environment that is conducive to stewardship of our cultural and intellectual heritage.

Congress's investment in preservation of our cultural and intellectual heritage has not been made through federal dollars alone. Its creation of policies and organizations that promote private sector participation has played a significant role. To date, these policies have protected our heritage, regardless of the embodiment or technical format, adjusting to the introduction of

---

[1] Many thanks to Morrison & Foerster for their assistance in researching the tax and foreign law, and especially to Anthony Ramirez for his help with the foreign law, even interrupting his bar study to help update research.

[2] Library of Congress, *National Digital Information Infrastructure and Preservation Program 2010 Report* (hereinafter "NDIIPP Report"), *available at* http://www.digitalpreservation.gov, at [14].

[3] Congress has also created a number of federally chartered organizations and agencies to support preservation efforts, including, for example, the Institute of Museum and Library Services, the National Film Preservation Foundation, Preserve America, the National Trust for Historic Preservation, and the Advisory Council on Historic Preservation.

new technologies over time. But digital preservation poses challenges that are different in kind from those of any previous technological format, and few policies have yet to be set in place to address the preservation of digital materials. True, preservation policies have often lagged behind the introduction of new technologies, but the ephemerality of digital technologies makes any such lag far more likely to equate to loss than was true in the past.

The astounding growth of digital information has outpaced the evolution of institutions and policies to support its preservation. Increasingly, our cultural and intellectual assets are being created and distributed in digital format. They include news and blogs, videos, multiplayer games, music, film, and books, as well as corporate and business documents, e-mails, and scientific and legal records. If properly managed, these assets will become the raw material from which future knowledge will be built and from which historians will reconstruct the complete story of our unique and changing times. Yet the solutions for long-term preservation of these materials are still very much in the development stages. Meeting the new challenges presented by digital preservation[4] will require public policy support to encourage the development of the necessary infrastructure, technologies, and preservation practices.

Digital preservation is complex and expensive: it requires a technical infrastructure and expertise that may not be readily available in traditional cultural memory institutions. It also requires the ability to conduct long-term planning and careful, secure data management and manipulation. Moreover, the costs of digital preservation are front loaded: decisions concerning what to preserve need to be made early in the life of the work. Digital materials need to be actively managed, not just left on shelves. Owing to the inherent instability of many digital media formats and the frequent obsolescence of formats and equipment to render digital files readable, active steps to preserve materials, including, for instance, migrating material to preservable formats and adding standardized metadata, may need to be taken quite early in the life of a digital work. At the same time, an exponential growth in the volume of digital material being created today makes it impossible for any one institution to collect and preserve more than a small fraction of the total.

By contrast, stewards of comparable paper and other analog materials do not bear the costs of preservation until far down the road in the work's lifecycle. Up-front costs—those incurred when a work is first published or released—are generally limited to acquisition. An analog copy of content, such as a book, manuscript, photo, or even film, can remain on the shelf for many decades with little or no signs of age. Preservation becomes a concern only when deterioration becomes visible to the naked eye, at which point measures can be taken to prevent further deterioration or to make a copy. Further, by the time an analog work has started to deteriorate, it has in all likelihood withstood—or not withstood—the test of time. In the analog realm, resources can be effectively allocated to those works that have acknowledged value under a historical lens.

Traditionally, the responsibility for preserving our cultural artifacts and knowledge has rested with public institutions, including libraries, archives, historical societies, and museums. The

---

[4] The challenges associated with preservation of digital content are very different from those associated with the preservation of traditional analog materials. For a discussion of some of the differences between preserving analog and digital works, *see The Section 108 Study Group Report* (March 2008) (hereinafter Section 108 Report) at 43-46, *available at* http://www.section108.gov/docs/Sec108StudyGroupReport.pdf; *see also* NDIIPP Report, *supra* note 1, at [15-16].

[5] NDIIPP is congressionally mandated to build a network of committed partners to collect, preserve, and ensure long-term access to digital materials for the benefit of Congress and the nation. *See* http://www.digitalpreservation.gov.

costs and resources required for digital preservation layer an additional responsibility on cultural heritage institutions, a responsibility that most libraries and archives cannot support under current funding. Nor can the public sector by itself take on the burden of shepherding our digital culture. Congress recognized this in creating the National Digital Information Infrastructure Preservation Program (NDIIPP)[5]—namely, that an effective long-term national strategy for digital preservation necessarily will entail a collective effort among a wide-ranging set of public and private entities.

The private sector, including individual collectors, has always played a key role in the collection, preservation, and provision of access to the nation's cultural and intellectual heritage.[6] Given the vast amounts of digital information now being created and disseminated, as well as the costs associated with preservation, private initiatives to collect, preserve, and/or donate digital collections will be especially crucial in the coming years to the building and preservation of a national collection of digital information and creativity.

At the same time, it is unrealistic to expect the private sector to make the investments necessary to steward the nation's digital culture without government support. Currently, there are insufficient economic incentives to expend the resources required for effective digital preservation. The financial rewards of a robust preservation system for digital assets are generally far from immediate, and the high up-front costs may be difficult to justify in a tight economy, particularly in sectors driven by the need for near-term profit.[7]

2. PROGRAM ACTIVITIES AND PROPOSALS

Recognizing the importance of a sound public policy environment for the productivity and cultural richness of our nation, NDIIPP partners and staff undertook a preliminary review of the types of public policy measures that would create an environment conducive to digital preservation. NDIIPP convened representatives from its partners and other interested parties to assist in the development of policy recommendations for legal and regulatory changes that would recognize the broad public interest in the preservation of, and long-term access to, digital content. This work was informed by partners' experiences as they encountered obstacles in preserving historically significant content.

The primary finding of this review is that there currently are too few incentives, and too many disincentives, for the private sector to engage in the preservation of digital content of public interest. Moreover, direct federal financial investment, while crucial to the incubation of a national infrastructure for digital preservation, is not sustainable; it does not provide a solution

---

[6] Examples from the early days of our nation include Jefferson's sale of his vast personal collection to Congress to replace the library lost in a fire in 1812 and the establishment of the Mount Vernon Ladies' Association in 1854 to save the home of George Washington, attributed with starting a national movement for private national historic preservation projects. *See* Mount Vernon, *Historic Preservation in America*, http://www.mountvernon.org/visit/plan/index.cfm/pid/525/; *see also* Murtagh, William. *Keeping Time: The Theory and History of Preservation in America.* New York, NY: Main Street Press, 1988.

[7] For instance, although the motion picture industry is rapidly moving from film to digital media, there are no industry standards or accepted solutions for digital motion picture preservation. The Academy of Motion Picture Arts and Sciences report *The Digital Dilemma* describes the challenges that digital preservation poses and the need for a preservation plan as good as the 100-year plan developed for analog film. A principal concern is cost and the fact that motion picture owners motivated by financial concerns will be reluctant to make the large investments necessary to ensure long-term access to digital film assets. *See* The Science and Technology Council of the Academy of Motion Picture Arts and Sciences, *The Digital Dilemma: Strategic Issues in Archiving and Accessing Digital Motion Picture Materials* (2007), *available at* http://www.oscars.org/science-technology/council/projects/digitaldilemma/login.php.

[8] Roy Rosenzweig. (2003). Scarcity or Abundance? Preserving the Past in a Digital Era. American Historical Review, 735 (June).

even for the near term. Accordingly, there is an urgent need for public policies that will provide incentives to encourage private sector participation in the huge tasks of collecting, cataloging, preserving, and making available our digital heritage of tomorrow.

Models for workable, effective incentives already exist in the domains of real property and analog cultural materials. They include economic incentives, such as tax breaks, and potential legal and other nonmonetary benefits, such as special copyright exceptions. Many of these can and should be adapted for the digital environment.[8] Such incentives will be key to driving the development of the infrastructures, technologies, and practices necessary for creating and preserving a national digital collection.

NDIIPP analyzed two major areas of the law that pose practical obstacles to digital preservation: (1) copyright law; and (2) discovery rules that enable business records to be subpoenaed from any entity (even a closed archive) and that are partially responsible for driving practices that dictate systematic destruction of digital business records. NDIIPP investigated lowering barriers created by these obstacles, in addition to creating affirmative incentives.

Without incentives, whether carrots or sticks, it may take the loss of vast amounts of digital information and culture before we, collectively as a society, private and public sectors combined, are willing to invest the resources necessary to creating effective digital preservation programs. At that point it will be too late, and the "digital dark ages" of the early twenty-first century will be a reality.[9]

The activities conducted by NDIIPP partners and the resulting proposals are described below. The proposals fall into four categories: (1) amendments to the copyright law; (2) tax and related economic incentives; (3) creation of a specialized business and confidential records archive; and (4) preservation requirements for federally funded research.

2.1 Amendments to Copyright Law

Early in its work, NDIIPP found that copyright law, namely, the lack of clearly applicable exceptions, was a primary obstacle to preservation of digital content. Effective preservation of digital works entails making lots of copies. Digital preservation anticipates failure by copying the data early and often—before it is lost or becomes inaccessible as hardware or software formats evolve—and by housing copies in multiple locations. Digital content that is actively managed for preservation purposes must be acquired, copied into a robust preservation repository, appropriately tagged with metadata, kept secure, monitored for integrity, consistently refreshed, backed up on a regular basis, and reformatted and migrated (emulated) to new media as prior media become obsolete. Each of these steps involves the making of copies and, unless an exception to copyright law applies, copyright law is implicated with each copy.[10]

Section 108 of the Copyright Act provides certain exceptions to the exclusive rights of copyright that allow libraries and archives to conduct preservation-related activities.[11] These exceptions were written with analog materials in mind and thus do not contemplate the need for

[9] Rosenzweig, *supra* note 8.

[10] *See* Section 108 Report *supra* note 4 at 43-46.

[11] *Id.* at 17-21.

[12] *Id.* at 81-84.

the multiple copies necessitated by digital preservation. Obtaining permission from the rights holder is one strategy to eliminate any risk of infringement, but it is impractical to attempt to obtain permission for entire collections of works with many different owners, who may or may not be locatable.[12] As a result, much digital preservation is currently being conducted under theories of fair use, but there is no assurance for libraries or archives or others conducting digital preservation that these activities are in fact fair use. The absence of copyright exceptions, coupled with the already-high costs of preservation, discourages many entities from engaging in digital preservation for any materials other than public domain content.

2.1.1 Section 108 Study Group Recommendations for Copyright Law Amendments

Recognizing that solving certain copyright issues was crucial to advancing long-term preservation solutions for digital content, NDIIPP, in conjunction with the Copyright Office, convened the Section 108 Study Group (study group), an independent committee named for the relevant section of the copyright law and comprising experts in the library, archives, and museum communities; scholarly communities; related not-for-profits; various rights holder communities; and other relevant professional disciplines. Group members were specifically chosen to represent a wide variety of perspectives, including those of all the major interest groups. NDIIPP and the Copyright Office asked the study group to consider how copyright exceptions for libraries and archives should be revised to respond to the challenges and opportunities presented by digital technologies.[13]

The Section 108 Study Group's Principal Recommendations Related to Digital Preservation

- Include museums, which perform many of the same functions as libraries and archives, within section 108 eligibility.
- Create a new exception to section 108 to permit qualified libraries, archives, and museums to proactively preserve at-risk publicly disseminated works prior to any damage or loss. Access to these "preservation-only" copies will be limited.
- Create a new exception to section 108 to permit libraries, archives, and museums to capture and preserve publicly available online content and make such content accessible for research and scholarship. Rights holders would be able to opt out of this provision.
- Permit libraries, archives, and museums to make a limited number of copies, as reasonably necessary, to create and maintain a replacement or preservation copy. This alteration to the current three-copy limit would, among other things, enable these entities to more securely preserve digital materials in their collections.
- Permit libraries, archives, and museums to use outside contractors to assist them in activities authorized under section 108.

[13] The study group adopted the following mission statement:

> The purpose of the Section 108 Study Group is to conduct a reexamination of the exceptions and limitations applicable to libraries and archives under the Copyright Act, specifically in light of digital technologies. The group will study how section 108 of the Copyright Act may need to be amended to address the relevant issues and concerns of libraries and archives, as well as creators and other copyright holders. The group will provide findings and recommendations on how to revise the copyright law in order to ensure an appropriate balance among the interests of creators and other copyright holders, libraries, and archives in a manner that best serves the national interest.

Id. at 1 (emphasis in original).

[14] Section 108 Report *supra* note 4.

[15] A joint report of the Library of Congress National Digital Information Infrastructure and Preservation Program, the Joint Information Systems Committee, the Open Access to Knowledge (OAK) Law Project, and the SURFfoundation. *International Study on the Impact of Copyright Law on Digital Preservation*. July 2008. *Available at* http://www.digitalpreservation.gov/library/resources/pubs/docs/digital_preservation_final_report2008.pdf.

In its report,[14] submitted to the Librarian of Congress in March 2008, the study group proposed several amendments to the copyright law that would enable libraries, archives, and museums to preserve important at-risk digital materials without harming the rights of the copyright owner. The report included several key recommendations related to preservation, highlighted below.

Proposal

- NDIIPP endorses the study group's recommendations related to preservation. In addition, NDIIPP recommends permitting the Library of Congress to delegate the mandatory deposit function to other qualified entities for specialized collections.

The study group's recommendations related to digital preservation were endorsed by the International Study on the Impact of Copyright Law on Digital Preservation, cosponsored by NDIIPP. Recognizing that digital content crosses national boundaries and that solutions can be found by working with other countries facing similar challenges related to digital preservation, NDIIPP conducted a multicountry examination of the impact of international intellectual property laws and policies on digital preservation in partnership with the United Kingdom and the national libraries of the Netherlands and Australia.[15]

The study group's proposals for two new exceptions, one for the preservation of publicly available content generally and the other for publicly available online content, are briefly described below. The details of each proposal, including recommended conditions, are set forth in the Report of the Section 108 Study Group (the Section 108 Report).[16]

2.1.2 Preservation of Publicly Disseminated Materials: New Exception for "Qualified" Cultural Institutions

The study group's principal recommendation for the preservation of publicly disseminated materials is to create a new exception for the preservation of published and other publicly disseminated works. The exception would:

- Permit a qualified digital preservation library, archives, or museum "to make a limited number of copies as reasonably necessary" to create and maintain a preservation copy of any at-risk publicly disseminated work in its collections.[17]

---

[16] Section 108 Report, *Supra* note 4 at 43-90.

[17] *Id*. at 69-79.

[18] *See infra* section 2.1, at 7.

[19] *See* Section 108 Report, *supra*. at 69-70. The recommended criteria to determine if a particular entity is "qualified" for the proposed exception are that it:

a. Maintains preservation copies in a secure, managed, and monitored environment utilizing recognized best practices. The following general principles for "best practices" should be observed for digital preservation (and for analog preservation to the extent applicable):
   i) a robust storage system with backup and recovery services;
   ii) a standard means of verifying the integrity of incoming and outgoing files, and for continuing integrity checks;
   iii) the ability to assess and record the format, provenance, intellectual property rights, and other significant properties of the information to be preserved;
   iv) unique and persistent naming of information objects so that they can be easily identified and located;
   v) a standard security apparatus to control authorized access to the preservation copies; and
   vi) the ability to store digital files in formats that can be easily transferred and used should the library or archives of record need to change.
b. Provides an open, transparent means of auditing archival practices;
c. Possesses the ability to fund the cost of long-term preservation;
d. Possesses a demonstrable commitment to the preservation mission; and
e. Provides a succession plan for preservation copies in the event the qualified library or archives ceases to exist or can no longer adequately manage its collections.

Id.

This exception would enable libraries, archives, and museums to make copies of published and other public digital works already in their collections for the purpose of preserving them. Many published works are at risk of loss if copies are not made before harm occurs. This is particularly true with respect to works in digital form, which can deteriorate very quickly to the point at which they cannot be used or restored. Preservation must begin early in the work's life and requires making multiple copies over a work's life, as described above.[18] Section 108 as currently drafted, however, does not provide for the making of preservation copies of published works—only of unpublished works. The proposed new exception would allow libraries and archives qualified for digital preservation to undertake preemptive preservation of "at-risk" publicly disseminated works in their collections—which includes works in their collections that have been publicly disseminated but are considered unpublished under the copyright law.

To ensure that the exception is used only for preservation purposes, the study group recommended that only those libraries, museums, and archives that are capable of conducting effective digital preservation and maintaining adequate security over digital copies be permitted to take advantage of it. The study group saw no public benefit to allowing entities that do not have such capacities to make copies and label them as preservation copies. The study group identified criteria for determining whether a library, archives, or museum is qualified, including whether it conducts effective, managed, robust preservation, utilizes then-current best practices, and employs adequate security to prevent the proliferation of unsecure copies.[19]

Because best practices are developing and likely will continue to change over time, the study group did not believe they could be set out in any particularity in the law. Instead, the study group suggested that best practices might be defined with reference to trusted sources of best practices, or through regulations that could be updated from time to time. The Section 108 Report describes several possible approaches for determining whether an entity complies with then-current best practices. One of those proposals would dovetail particularly well with certain other recommendations described in this paper and the *National Digital Information and Infrastructure Preservation Program 2010 Report*; namely, the Library of Congress, through the National Digital Stewardship Alliance (the Alliance), would assume the role of certifying best practices from time to time itself, through an advisory group, or by reference to practices adopted by others. The same approach could be used both to qualify entities as full members of the Alliance and to qualify entities to take advantage of the digital preservation exception in the copyright law. In other words, compliance with then-current best practices, as designated by the Library of Congress, would be a condition of full membership in the Alliance, and full membership would automatically qualify an entity to take advantage of the preservation exception. Non-members also might be eligible for qualification for the exception, subject to the same standards.

The study group's recommendation includes other conditions and qualifications that are described in more detail in the Section 108 Report. The proposals for qualification for digital preservation are highlighted because of their particular relevance to and synergies with the NDIIPP report's suggested structure for the Alliance, as well as the proposed tax and other incentives described below.

Proposals Specific to Exception for Preservation of Publicly Disseminated Works

- Adopt the proposed exception for digital works to allow qualifying institutions to

[20] *Id.* at 80-87.

preserve publicly disseminated works pursuant to best practices without risk of copyright infringement. This would help ensure that robust and effective preservation practices are employed, and that more digital material is preserved for the long term.
- Provide the Library of Congress with the authority to (or to delegate an advisory committee of experts or third-party organization(s) to) identify, update, and maintain best practices for digital preservation and to determine whether libraries, archives, and museums comply with such best practices and thus are eligible for the digital preservation exception.

2.1.3 Adopt a New Web Archiving Exception

The Internet exceeds the reach of any prior distribution channel by an exponential measure. Vast amounts of information and creative content are being published on the web, and some of it has significant historical or cultural value. But much of the content published online is ephemeral; it is not archived in any systematic way. Once taken down, it is potentially lost for all time.

The Section 108 Study Group proposed a new web archiving exception to address this problem. The exception would apply to all libraries, archives, and museums and would:
- Permit libraries, archives, and museums to capture and reproduce publicly available online content for purposes of preserving the content and providing scholarly access.[20]

Unlike other section 108 exceptions, this provision would enable libraries, archives, and museums to collect and preserve at-risk digital materials. As a general rule, other section 108 exceptions permit making copies only of materials already in these organizations' collections. An exception permitting libraries, archives, and museums to capture and archive online content would eliminate any risk of infringement, and the ability to provide access for research and scholarly use would help justify costs and thereby provide incentives for such entities to collect and preserve important online content that might otherwise be lost. Because the content is already freely available online, rights holders generally should not object to the collection, preservation, and research use of such content, and if they do, they may opt out. Further, allowing *all* libraries, archives, and museums to take advantage of the exception increases the likelihood that collectively our cultural heritage institutions will collect and preserve a meaningful portion of online content of potentially significant historical value.

Proposal
- Adopt the study group's proposal for a new exception to copyright that would permit libraries, archives, and museums to capture and reproduce publicly available online content for preservation purposes and to make those copies accessible for research and scholarship.

2.1.4 Authorize the Library of Congress to Delegate the Copyright Office's Mandatory Deposit Function

Another major obstacle to digital preservation is the sheer quantity of material to be collected and preserved. While the study group's proposed Internet archiving exception, described immediately above, partially addresses this problem with respect to online content, it solves only some of the problems of creating a coherent, comprehensive national collection of digital works. An organized system for collecting online materials is needed if the nation is to build

[21] 17 U.S.C. § 407.

a national digital collection that in any way mirrors our traditional collections in breadth and scope.

The Library of Congress has an unparalleled collection of paper and other hardcopy forms of content, including books, monographs, serials, music, newspapers, and other printed materials; maps; recorded sound (including CDs); and moving and still images. It has built this collection largely through the deposit provisions of the copyright law. Section 407 of the Copyright Act requires that owners of the publication right of any work published in the United States (subject to exemptions promulgated by the Register of Copyrights) deposit two copies of the work with the Copyright Office for the disposition of the Library of Congress.[21] Copies of unpublished works registered with the Copyright Office are also deposited for the use of the Library of Congress. Copyright deposit has enabled the Library of Congress to amass the largest, most comprehensive collection in the world.

The current copyright deposit system is unlikely to be sufficient to meet the demands of a comprehensive national collection in the digital age. The vast amount of digital material being produced today calls into question the Library of Congress's continued ability to serve as a "universal collection." No one entity has the capacity to collect all works published online, even with the benefit of mandatory deposit. Hence, one of NDIIPP's key findings is the necessity of a distributed approach for collecting and preserving digital information.

A proposed strategy for building a comprehensive national digital collection is to allow the Library of Congress to designate specially qualified institutions as agents for mandatory deposit for specific types of content. The Library might exercise this prerogative for areas in which it does not have sufficient resources to carry out collection and preservation activities. It might also use the prerogative in cases where it does not wish to divert resources to acquire the necessary expertise for the creation or management, whether technical or curatorial, of the collection. For instance, particular institutions might be charged with collecting and preserving specific types of Internet content in which they have proven expertise, such as websites related to state politics or the environment; and other institutions might be the designated national repositories for certain specialized content, say, social science data, scientific data, local geospatial data, or business records. Each such designated agent would serve as a specialized center for the content type. The Library of Congress would continue to collect and preserve the types of content it already collects and perhaps others, with some support, as appropriate, from designated agents.

Distributing the national collection through the delegation of the mandatory deposit function would create powerful incentives for others to engage in systematic digital preservation activities. This would enable the Library of Congress to focus its resources on those areas that are strategically most important to it, as well as on those it has traditionally devoted itself to, without having to spread its resources too thin, or alternatively, having to simply ignore the myriad new forms of content. The authority to delegate mandatory deposit would allow the Library of Congress to outsource this function on a case-by-case basis to other entities that have proven expertise in collecting and preserving particular content types.

Of course, any entity wishing to serve as an agent for mandatory deposit would have to meet

[22] Other national libraries have shown that a distributed approach can work. For instance, in the United Kingdom, there are six legal deposit libraries across England, Scotland, Wales, and Ireland.

specific requirements, including compliance with best practices. These would include, for example, a noncommercial mission, guaranteed 24/7 full access at the Library of Congress for its patrons as well as Congress and staff, robust backup and preservation strategies, compliance with Library of Congress policies, and a fail-safe succession plan for the designated content in case the entity ceased to exist or wished to deaccession the materials. The Library of Congress would be required to oversee, approve, and monitor all activities employing the mandatory deposit privilege. Any collection acquired through this agency would be held for the benefit of the public. The content so collected would be deemed part of the Library of Congress's collection, albeit housed and stewarded by a third party.

There are precedents for delegation of national collections, such as the Government Printing Office's (GPO) Federal Depository Library System and the National Library Services for the Blind and Physically Handicapped's national network of cooperating libraries. Moreover, the Library of Congress already informally delegates certain collections through arrangements with the National Agricultural Library and the National Library of Medicine. Because of its experience with NDIIPP, the Library of Congress is well positioned to lead this kind of distributed national collection.[22]

Authorizing the Library of Congress to delegate the mandatory deposit function with respect to particular collections would allow institutions with the requisite expertise to engage in far broader preservation initiatives than otherwise would be the case. It would give these entities the ability to demand copies of published works within the specifically circumscribed area of its content mandate. And it would permit the Library of Congress, in partnership with others, to develop and execute a comprehensive national digital collection plan that would approximate the scope of the Library of Congress's collection of traditional formats. Such a plan would ensure that the United States continues to maintain as near a universal collection as possible of the country's cultural and intellectual output in the digital age.

In addition, certain amendments need to be made to the mandatory deposit provisions of section 407 of the Copyright Act to enable the Library of Congress to collect, preserve, and provide access to materials published online without running afoul of the copyright law. This includes allowing the Library of Congress to make a limited number of copies of works deposited under section 407 as reasonably necessary to preserve those works and to make copies available to users on the premises of the Library of Congress, with appropriate protections for right holders. The Library of Congress also should have the ability to demand deposits in formats other than those in which a work is published. Deposits currently are limited to a "best edition" of the work, which the statute identifies as a published version. Publishers of online content often publish content in formats, such as html, other than those in which it produces or archives the content, and those other formats may be far more suitable to the Library's needs, especially for preservation.

The Library of Congress's agents for mandatory deposit also should be able to avail themselves of these copyright exceptions in fulfilling their duties as agents.

---

[23] The United States Internal Revenue Code of 1986, as amended, 26 U.S.C. § 1 *et seq.* (the "Code").

[24] State and local laws also provide tax incentives for historic preservation, and a number of government and not-for-profit grants are available. For a synopsis of resources available for preservation of historic buildings, *see* the National Trust for Historic Preservation website at http://www.preservationnation.org/resources/find-funding/nonprofit-public-funding.html.

Proposals

- Enact legislation that would permit the Library of Congress to delegate the U.S. Copyright Office's mandatory legal deposit function to qualified entities within the Alliance. This would address the vast quantities of material and varied expertise needed to effectively collect and preserve the vast amount of important digital content being produced.
- Amend section 407 of the Copyright Act to allow the Library of Congress and those to whom it delegates its mandatory deposit authority to (1) make a limited number of copies of works deposited under section 407 as reasonably necessary to preserve those works and to make them available to users on the premises of the Library of Congress, and (2) require deposit in a format suitable for preservation or other needs of the Library of Congress, whether or not a published version.

The Study Group also examined but did not offer specific conclusions to the question of libraries and access to content, specifically looking at the intersection between copyright law and the ability of libraries to provide access to content, whether through digital display, dissemination or public performance. This is a complex issue that involves various fact patters, including the purpose of the use and the nature of the content. For example, is the content published or unpublished? Is it out of print or being offered by the rights holder on the commercial market? Is the Library engaged in "fair use" or another applicable exception or limitation under the law? Is there a licensing opportunity that might serve the rights holder and the library's mission?

Proposal

- Create a pilot project in which the Library of Congress may explore with copyright owners the digital display and/or dissemination of certain works that are in its collection and protected by copyright (for example, text, audio, visual or audio visual works) under terms to be mutually agreed upon.

---

[25] 26 U.S.C. §§ 38, 47.

[26] 26 U.S.C. § 170(a).

[27] 26 U.S.C. § 38, 47(a), 50(d).

[28] 26 U.S.C. § 47(c)(2). The credit percentage is 20 percent with respect to buildings that meet the criteria for "certified historic structures" and 10 percent for buildings not meeting such criteria but that were originally placed in service prior to 1936. 26 U.S.C. §§ 47(a), 47(c)(1). With respect to buildings other than certified historic structures, certain additional criteria apply in terms of the percentage of the original external walls and internal structural framework that must be retained in the rehabilitation.

[29] "Certified historic structure" means any building (and its structural components) that (i) is listed in the National Register or (ii) is located in a "registered historic district" and is certified by the Secretary of the Interior to the Secretary of the Treasury as being of historic significance to the district. 26 U.S.C. § 47(c)(3)(A). "Registered historic district" means (i) any district listed in the National Register and (ii) any district (I) that is designated under a statute of the appropriate state or local government, if such statute is certified by the Secretary of the Interior to the Secretary of the Treasury as containing criteria that will substantially achieve the purpose of preserving and rehabilitating buildings of historic significance to the district and (II) that is certified by the Secretary of the Interior to the Secretary of the Treasury as meeting substantially all of the requirements for the listing of districts in the National Register. 26 U.S.C. § 47(c)(3)(B).

[30] 26 U.S.C. § 47(c). "Certified rehabilitation" means any rehabilitation of a certified historic structure that the Secretary of the Interior has certified to the Secretary of the Treasury as being consistent with the historic character of such property or the district in which such property is located. 26 U.S.C. § 47(c)(2)(B)(iv).

[31] Section 108 Report *supra* note 4, at 69-79.

2.2 Tax-Related Incentives

Creative or intellectual content that may have significant cultural or historic value to society may have little value to its owner, and so the owner may have insufficient incentives to make the investments necessary to preserve it. In the case of digital assets, one such resource might be a sound recording or film that no longer has a sufficient market to justify the costs of preservation; another might be scientific data used for a study since completed. A third example is the e-mail correspondence of a significant individual.

Government policies, including tax incentives, could provide encouragement to preserve cultural assets by helping owners of cultural properties recognize greater value in them. Tax credits that cover a portion of the costs of preservation or enable individuals and corporations to donate digital cultural assets may provide enough of a subsidy to make the difference between whether or not those works are preserved.

NDIIPP explored some of the ways in which the laws of the United States and other countries encourage preservation of culturally significant real and personal property—through tax or other indirect economic incentives. NDIIPP partners were interested in finding appropriate models for tax benefits that would encourage the preservation of digital content. The United States Tax Code[23] currently provides little in the way of incentives to preserve digital materials of potential historical or cultural value, but provides benefits that encourage the preservation of real property, as well as the donation of culturally significant property to preservation institutions.[24] Laws relating to tax credits and deductions for the historic preservation of real property provide especially useful models for tax incentives that are potentially applicable to digital cultural assets.

2.2.1 Tax Credits Related to the Preservation of Real Property

First, NDIIPP researched the types of federal tax deductions and credits allowable for the preservation of real property. The Code provides tax incentives for the preservation of historic real properties in the form of (1) a tax credit (the "rehabilitation credit") with respect to the substantial rehabilitation of "qualified rehabilitated buildings"[25] and (2) a deduction for contributions of real property to qualified charities and governmental units.[26]

2.2.1.1 Rehabilitation Credit

A rehabilitation credit is available under the Code to reduce, within certain limits, taxes that would otherwise be owed by a taxpayer who has an adequate ownership (or leasehold) interest in a "qualified rehabilitated building."[27] The credit creates incentives to rehabilitate historic buildings and put them back into productive use by allowing the owner to take as a tax credit a percentage of "qualified rehabilitation expenditures,"—capital expenditures for real property and additions and improvements thereto.[28] The credit vests over a five-year period, such that if the ownership (or leasehold) interest is not maintained during that time, the unvested portion

---

[32] 26 U.S.C. § 170(a).

[33] There may be a reduction in the deduction to reflect rehabilitation tax credits received by the same taxpayer. 26 U.S.C. § 170(f)(14).

[34] 26 U.S.C. § 170(h).

[35] *Id.* §170(h)(C). "Conservation purposes" include, among other things, the preservation of a "historically important land area" or of a "certified historic structure." 26 U.S.C. §§ 170(h)(4)(A)(iv)"Certified historic structure" includes (i) any building, structure, or land area that is listed in the National Register or (ii) any building that is located in a "registered historic district" and is certified by the Secretary of the Interior to the Secretary of the Treasury as being of historic significance to the district. 26 U.S.C. § 170(h)(4)(C).

[36] Gifts to private foundations are subject to more-restrictive deduction limitations. *See* 26 U.S.C. § 170(c).

is recaptured in the taxpayer's tax bill for the year of recapture. Generally, in order to qualify for the 20 percent rehabilitation credit, the rehabilitation must be of a "certified historic structure"[29] and must be a "certified rehabilitation."[30] The various conditions and qualifications are meant to ensure against abuse.

Proposals

- Institute a tax credit similar to that available for the preservation of real property that would allow taxpaying entities to deduct a portion of the costs of qualified digital preservation activities. This might include the development of technologies, services, or infrastructure necessary to advance digital preservation, as well as the commitment to preserve particular materials.
- Qualifying conditions and limitations similar to those provided for the rehabilitation credit would prevent the use of a digital preservation credit or deduction other than for "qualified digital preservation" purposes. The deduction could be limited to entities—or services or technologies—that meet best-practices standards for digital preservation. A further requirement might be that the content to be preserved be certified as at risk or otherwise as preservation-worthy. The Library of Congress could provide the same "certification" role as the Secretary of the Interior does in certifying historic structures and districts. The Section 108 Study Group's proposals for qualification for the proposed preservation exception for publicly disseminated works provides some guidance on the form such certification might take.[31]

2.2.1.2 Deduction for Charitable Contributions of Certified Historic Structures or Historically Important Land Areas

The Code allows taxpayers to claim a deduction for contributions of historic real property, namely, "certified historic structures" or historically important land, to qualified charities and governmental units.[32] In the case of a corporate donor, the aggregate deductions for charitable contributions for any taxable year cannot exceed 10 percent of the taxpayer's taxable income. In no case can the amount of the deduction exceed the appraised fair market value of the property, and in certain cases it may be limited to a lesser amount. There are a number of conditions for qualification, limitations, and restrictions to prevent abuse.

---

[37] 26 U.S.C. § 170(c)(1).

[38] Ottawa Silica Co. v. United States, 699 F.2d 1124 (Fed. Cir. 1983).

[39] *See* Transamerica Corp. v. United States, 15 Cl. Ct. 420 (1988), *aff'd on other grounds*, 902 F.2d 1540 (Fed. Cir. 1990) (no deduction for donation to Library of Congress of film negatives; donor retained all intangible rights and access to the property for commercial use). Under copyright law, the ownership of copyright or any of the exclusive rights under the copyright is distinct from ownership of the material object in which the work is embodied. Unfortunately, this is not true for purposes of the charitable deduction. In the case of a copyright, the copyright and the underlying "art," i.e., the tangible personal property for which the law gives a copyright, cannot be separated under the partial interest rule for income tax purposes. Treas. Reg. § 170A-7(b)(1) & Ex.I. Congress liberalized this rule in 1981 to allow them to be separated for gift and estate tax purposes, thus permitting a charitable deduction of a gift of art, for example, when the copyright interest is retained by the donor. However, Congress left the partial interest rule intact for income tax purposes with respect to the transfer of copyrights separate from the underlying art. Thus, the partial interest rule denies a deduction for a gift of a material object while retaining the copyright, or conversely, a gift of the copyright while retaining the object.

[40] "Gifts or bequests or devises to or for the benefit of the Library of Congress, including those to the board, and the income therefrom, shall be exempt from all Federal taxes, including all taxes levied by the District of Columbia." 2 U.S.C. § 161. Case law suggests that the actual availability of a deduction for such gifts will still turn on donative intent, whether the transfer actually qualifies as a completed gift, and valuations. *See*, *e.g.*, *Transamerica Corp.*, 15 Cl. Ct. 420.

[41] 26 U.S.C. § 170(e)(1)(A).

[42] *Transamerica Corp.*, 15 Cl. Ct. at 475 (no deduction for donation to Library of Congress of film negatives; no fair market value); Strasser v. Comm'r, T.C. Memo. 1985-579 (1986) (Library of Congress employee testified that donated manuscripts have research value but no commercial value; deduction denied) *aff'd*, Strasser v. C.I.R., 838 F.2d 1203 (2d Cir. 1987).

[43] 26 U.S.C. § 170(e)(1).

[44] Treas. Reg. § 1.170A-4(b)(1)

[45] 26 U.S.C. § 1221(a)(3); Treas. Reg. § 1.1221-1(c)(1).

Special rules apply permitting a charitable deduction for the value of "qualified conservation contributions."[33] "Qualified conservation contributions" include contributions to governmental units, public charities, or their supporting organizations of a "qualified real property interest,"[34] if the contribution is made "exclusively for conservation purposes."[35]

Proposals

- A tax deduction for contributions of digital collections to qualified cultural heritage institutions could be structured in a manner similar to that provided for certified historic structures and historically important land. Such a deduction would provide incentives for private individuals or entities to donate potentially valuable digital assets to qualified cultural memory institutions and government entities for conservation or preservation purposes.
- Similar types of qualifying conditions and limitations as those pertaining to the historic real property deduction would prevent the use of the deduction for purposes other than "qualified digital preservation" and would ensure that the deduction is taken only for at-risk or other preservation-worthy material. These would complement and augment the deductions already available for charitable contributions of tangible property, described below.

2.2.2 Deductions for Charitable Contributions of Tangible and Intangible Property

Income tax deductions are currently available under a separate provision of the Code for charitable contributions of tangible and intangible property, including intellectual property. Congressional action over the years to curb abuse has significantly reduced the tax benefits, however, making such contributions less attractive to donors or altogether inapplicable in the case of materials that have no or little current market value, even where they have significant potential archival value.

The charitable contribution rules are quite complex. The availability and amount of an allowed deduction depend on the type of property, the donor, the type of donee, and any related restrictions on the transfer. A qualifying charitable contribution can be for educational, research, public, or other exempt purposes, including for preservation purposes.

Deductible contributions may be made to governments or to U.S. charities or private foundations that are exempt from federal income taxation under Code Section 501(c)(3).[36] The contribution must be a bona fide gift; if there is any expectation of receiving a financial benefit or something substantial in return for the gift, then the deduction is limited to the value of the property that exceeds the benefit received by the donor in return.

Charitable gifts of property to the U.S. government or a state or local government qualify as

[46] 26 U.S.C. § 1221(a)(3); Treas. Reg. § 1.1221-1(c)(2). This has become an issue in several cases involving gifts of historical papers or corporate archives. The IRS and courts have held that as self-created documents and compilations, they are ordinary income property in which the donor corporation has no basis. Chronicle Publishing Co. v. Comm'r, 97 T.C. 445 (1991) (zero basis in newspaper clipping library; not deductible); Morrison v. Comm'r, 71 T.C. 683, *aff'd*, 611 F.2d 98 (5th Cir. 1980); TAM 200119005 (zero basis in film library; not deductible).

[47] 26 U.S.C. § 170(e)(1)(A).

[48] Rev. Rul. 82-9; 1982-1 C.B. 39 (no deduction allowed for oil- and gas-drilling documents donated to educational institution; basis is zero because costs of developing the items were previously deducted as drilling and development costs).

[49] There are two exceptions to this general provision for long-term capital gain assets. If the donee is a private nonoperating foundation, certain rules apply to reduce the charitable deduction. The deduction may also be reduced for gifts of tangible personal property that are not going to be used for a purpose related to the donee's exempt or charitable purposes. 26 U.S.C. § 170(e)(1)(B).

[50] *See, e.g.*, Rimmer v. Comm'r, T.C. Memo 1995-215.

charitable contributions as long as the gift is "made for exclusively public purposes."[37] As a practical matter, in cases involving governmental donees, the Internal Revenue Service has not examined closely the actual use of the contributed property by the donee, but rather has focused on whether the gift was motivated by a sense of disinterested generosity and public purpose, or was made in anticipation of a quid pro quo or special treatment from the government.[38] If the donor retains any interest in or power over the property, then it is not a completed gift and no deduction is available.[39]

A practical consideration is the authority of the donee agency to accept such a gift. This authority might be found in the federal statutes creating and empowering the agency, in its regulations, or in an executive order. Numerous statutes authorize gifts to federal agencies and often declare that such gifts are deductible, although such statutory authority is not a necessary precondition to a charitable deduction if the gift is made for an exclusively public purpose and accepted as such. Gifts to the Library of Congress are expressly exempt from all federal taxes by virtue of a special provision of the law.[40]

The Code provides that, for purposes of calculating the amount of the deduction, the fair market value is used as a starting point.[41] This presents a problem for tangible or intangible property that no longer has any commercial value but may have significant historical or archival value. The courts have held that "values which are not amenable to separate measurement in a commercial market, such as values for archival and historical uses and scholarly research, do not have a 'fair market value' for the purpose of a charitable deduction."[42] For such property, then, it appears that no charitable deduction may be available.

For property that does have a fair market value, the deduction will be reduced if the property is ordinary income property.[43] Ordinary income property[44] includes, inter alia:

- property held for sale in the ordinary course of business (i.e., inventory property);
- a copyright, literary, musical, or artistic composition, theatrical production, radio

---

[51] UNESCO's Cultural Heritage Laws Database, *available at* http://www.unesco.org/culture/natlaws/, was the source of many of the translations used to conduct research. Much of the research was also done through the noted secondary sources due to the difficulty of researching foreign law systems and language barriers. It is recommended that any study group formed to investigate possible incentives for digital preservation follow up with the applicable cultural heritage institutions of the other countries to better understand the implementation of the provisions and ensure up-to-date information is recommended for provisions of particular interest.

[52] *See generally* UNESCO Cultural Heritage Laws Database, *available at* http://www.unesco.org/culture/natlaws/; *see* Susan Shearing, *One Step Forward? Recent Developments in Australian State and Territory Indigenous Cultural Heritage Laws*, 3 Macquarie J. Int'l & Compl. Envtl. L. 35, 46-51 (2006). *See* Aboriginal Cultural Heritage Act 2003, §§ 23-33 (Queensland, Australia) (*Available at* http://www.legislation.qld.gov.au/LEGISLTN/ACTS/2003/03AC079.pdf). In Queensland, Australia, the duty may be automatically satisfied if the owner follows "best practices" promulgated by the government. These best practices are called the "Duty of Care Guidelines" and are *available at* http://www.nrw.qld.gov.au/cultural_heritage/legislation/duty_of_care.html. Many countries impose similar duties not to harm designate cultural properties, including for example, Spain, Greece, Spain, and Mexico. *See* Law 16/1985 dated 25 June, on the Spanish Historical Heritage (Official State Bulletin of 29 June 1985) (Spain); Federal Law on monuments and archaeological, artistic, and historical zones (translation) (1986) (Mexico); Law 3028 on the protection of antiquities and cultural heritage in general (Greece); Code of the Cultural and Landscape Heritage (translation)(2004) (Italy); *see also* Andrea Boggio, *From Protections to Protection: Rethinking Italian Cultural Heritage Policy*, 24 Colum.-VLA J.L. & Arts 269 (2000); *see generally* Halina Nieč *Legislative Models of Protection of Cultural Property*, 27 Hastings L.J. 1089 (1975).

[53] In the United Kingdom, in certain cases, the owner may sell the title to the property to the government but retain possession, where an agreement is reached on a preservation plan, security, and public access. *See* Valerie M. Fogleman, *A Capital Tax System to Preserve America's Heritage: A Proposal Based on the British National Heritage Capital Tax System*, 23 Vand. J. Transnat'l L. 1, 14-43 (1990); *see* France's *Code du Patrimoin*e (2004).

[54] *See* Chester H. Liebs, *Listing of Tangible Cultural Properties: Expanded Recognition for Historic Buildings in Japan*, 7 Pac. Rim. L. & Pol'y J. 679, 692-93 (1998). In France, these subsidies are partially administered by a private, nonprofit organization, La Fondation du Patrimoine, which considers and approves potential properties. *See* http://www.fondation-patrimoine.net/en/adhesion/porteurs-projet/prive.php.

program, newspaper cartoon strip, or any other property *created by the donor*, and eligible for copyright protection;[45] and

- letters, memoranda, or similar property, such as corporate documents or collections of papers, **in the hands of the person who prepared them or the person or corporation for whom they were prepared.** Such items might include a draft of a speech, a manuscript, a research paper, an oral recording, a transcript of an oral recording or interview, a personal or business diary, a log or journal, a corporate archive, office correspondence, a financial record, a drawing, a photograph, or a dispatch.[46]

The deduction for a contribution of property that is considered "ordinary income" property is limited to the donor's adjusted tax basis in the property,[47] which generally means the amount the donor has spent in acquiring or developing the property and that has not previously been deducted for tax purposes. This ordinary income property could well include digital materials of potential long-term value to a cultural heritage institution, including, for example, sound recordings donated by a record company or films donated by the motion picture company holding the copyright. As a practical matter, since most business expenses incurred in developing these forms of property are expensed or amortized, meaning a business expense tax deduction has already been taken, the company's basis in such property is zero, and no charitable deduction will be available.[48]

If the types of property listed above were not created by or for the donor, and if they qualify as long-term capital assets, then a fair market deduction generally is available.[49] This might apply, for example, to a compilation of archival materials gathered by a third-party collector who did not create the materials.[50]

Amending these non-cash charitable deductions provisions of the Code so that they are more readily available for donations of "qualified digital preservation works" to cultural heritage institutions would encourage organizations and individuals who have created or acquired collections of important digital materials to care for and donate such materials to cultural heritage institutions for the benefit of the public.

Proposal

- Amend the non-cash charitable deductions provisions of the Code so that they are more readily available for donations of "qualified digital preservation works" to cultural heritage institutions. For instance, special rules could be created for donations of

---

[55] *See* C. Franklin Sayre, *Cultural Property Laws in India and Japan*, 33 UCLA L. Rev. 851, 872 (1985).

[56] *See* Boggio, *supra* note 53.

[57] *See* Jason C. Roberts, *The Protection of Indigenous Populations' Cultural Property in Peru, Mexico and the United States*, 4 Tul. J. Comp. & Int'l L. 327, 341 (1996).

[58] *See* Liebs, *supra* note 55, at 692-93; Fogleman, *supra* note 54, at 14-43 (1990); *see generally* Nieč*supra* note 53.

[59] *See* Historic Buildings and Ancient Monuments Act 1953 (United Kingdom), Art. 3A; Liebs, *supra* note 55, at 692-93 (1998).

[60] *See, e.g.*, Inheritance Tax Act 1984, §§ 30-35A (United Kingdom), *available at* http://www.legislation.gov.uk/RevisedStatutes/Acts/ukpga/1984/cukpga_19840051_en_1; *see* Fogleman, *supra* note 54, at 14-43.

[61] *See, e.g.*, Inheritance Tax Act 1984, §§ 30-35A (United Kingdom) at §§ 230-31.

[62] *See, e.g.*, Law on Cultural Heritage, Law # 28/2001/QH10, Art. 15 (Socialist Republic of Vietnam).

[63] The Central Archives for Finnish Business Records was established in Finland to preserve the business records of all Finnish companies. *See* http://www.elka.fi/.

[64] United Nations Educational, Scientific and Cultural Organization (UNESCO) *Convention for the Safeguarding of the Intangible Cultural Heritage*, adopted at its 32nd session, Paris, October 17, 2003; *See also* UNESCO *Convention Concerning the Protection of the World Cultural and Natural Heritage*, adopted by the General Conference at its 17th session, Paris, November 16, 1972.

"qualified digital preservation works"—i.e., digital content and other materials that are at risk of loss, meet certain quality standards, and have potential historical value to "qualified digital preservation entities"—entities qualified to conduct digital preservation in the manner described above. An alternative method to fair market valuation, such as the value of the material to the qualified preservation entity, could be explored to address the fact that much of the qualifying material would not have a market value. Such proposed amendments should be carefully crafted to address potential abuses of the charitable deduction.

2.2.3 Preservation Incentives Around the World

NDIIPP conducted preliminary research into some of the laws and policies of other countries that encourage preservation of cultural heritage assets.[51] Nations around the world use various types of legal measures to provide incentives, in the form of both carrots and sticks, for individuals and entities to preserve cultural property. In most of these countries the cultural property is treated as a national asset, whether in private hands or the hands of the State, and joint public-private responsibility is imposed on preservation of the asset. Historic buildings, lands, and artifacts are covered by the various policies. In some countries the laws may be broad enough to apply to digital cultural properties, but in general the laws are directed toward the preservation of older artifacts, artwork, or structures.

Some countries provide incentives to preserve the property within the country's borders or to provide public access to the cultural heritage property. In other cases, the government has the right to purchase the cultural heritage property, or might subsidize private entities or individuals to preserve and maintain it, directly or through tax credits, provided certain conditions, such as proper care or public access, are met. A number of countries provide tax relief for donations of cultural heritage property to the government similar to the provisions in the Code described above. Some, such as Britain, permit tax deductions for stewarding culturally valuable tangible property within the country, under conditions that ensure the stewardship is conducted on behalf of the public interest. These deductions are somewhat akin to the rehabilitation credits that the Code provides for real property. Other countries impose "duties of care" to encourage preservation of culturally significant property and artifacts. The following is a list of such types of incentives and disincentives to encourage the preservation of cultural heritage materials that have been found in the laws of other countries:

- Duty of care to avoid harmful actions imposed on certain cultural properties or removal from country, even if privately owned, and fines for violations.[52]
- Government has right to purchase cultural property, or owner has right to sell it to the government.[53]
- Government provides subsidies to owners of monuments and cultural property for maintenance of the property or preserves qualified registered property at its expense.[54]
- Government will purchase at market price an antiquity that is in danger of being destroyed, removed, or injured.[55]
- Private entities and persons may sell cultural artifacts only if not contrary to the interest in public collections.[56]
- Requirements to list cultural properties in national inventories, with benefits and/or penalties.[57]

---

[65] David A. Kirsch. (2009). The Record of Business and the Future of Business History: Establishing a Public Interest in Private Business Records, Library Trends, 57(3): 352-70 (included in a special edition of Library Trends "The Library of Congress National Digital Information Infrastructure and Preservation Program" (Patricia Cruse & Beth Sandore eds.)).

- Government organization may outsource management of a cultural asset to a private entity.
- Exemptions from and reductions in inheritance and other taxes.[58]
- Low-interest loans.[59]
- Tax exemptions for transfers, including by inheritance, of cultural properties, conditioned on:
  —Owner must provide public access to the property.
  —A maintenance plan for the property, developed in consultation with the government.[60]
- Ability to offer cultural properties to the government in lieu of outstanding taxes. In some cases, the property may stay in place on permanent loan to the former owner pursuant to an agreed plan for preservation, security, and public access.[61]
- If the owner of an item of cultural heritage does not have the means or ability to protect or promote it, it should be sent to a state museum or authority.[62]
- Establishment of national archives with cooperation of commercial sector to preserve business records.[63]

NDIIPP also reviewed the UNESCO Convention for the Safeguarding of the Intangible Cultural Heritage.[64] The convention provides for a number of ways in which the treaty members should encourage the safeguarding of cultural heritage properties at a national level, including taking necessary measures to ensure the safeguarding of the intangible cultural heritage present in its territory; drawing up inventories of the intangible cultural heritage present in its territory; fostering scientific, technical, and artistic studies, as well as research methodologies, with a view to effective safeguarding of the intangible cultural heritage, in particular the intangible cultural heritage in danger; and adopting certain legal, technical, administrative, and financial measures. None of the provisions of the treaty specifically targets and encourages preservation of digital cultural assets, but it nonetheless serves a recent model for international efforts to preserve cultural heritage.

Proposals

Examples of foreign laws provided fodder for new ideas and further support for some of the ideas already developed by NDIIPP partners—namely, for incentives to encourage private sector participation in the preservation of significant digital cultural materials. Some of the ideas found support in that foreign laws include:

- Tax deductions, subsidies, or loans for those who agree to follow best practices for digital preservation and agree to provide some measure of public access, as appropriate.
- Creation of a national inventory of preserved digital content.
- Funding to create a robust electronic deposit system and repository for the Library of

---

[66] University of Maryland, *Closed Archive Methodology* (Aug. 3, 2006), *available at* http://www.brobeckclosedarchive.org/court_docs.html.

[67] The *NIH Public Access Policy*, *available at* http://publicaccess.nih.gov/ implemented Division G, Title II, Section 218 of P.L. 110-161 (Consolidated Appropriations Act, 2008), made permanent by Division F Section 217 of P.L. 111-8 (Omnibus Appropriations Act, 2009). The law states: "SEC. 217. The Director of the National Institutes of Health ("NIH") shall require in the current fiscal year and thereafter that all investigators funded by the NIH submit or have submitted for them to the National Library of Medicine's PubMed Central an electronic version of their final, peer-reviewed manuscripts upon acceptance for publication, to be made publicly available no later than 12 months after the official date of publication: Provided, That the NIH shall implement the public access policy in a manner consistent with copyright law."

[68] *Public Research in the Public Interest Act of 2006*, S. 4040 (2006) *available at* http://thomas.loc.gov/cgi-bin/query/z?c109:S.4040, *but see* H.R. 801, 111th Cong. (2009), *available at* http://thomas.loc.gov/cgi-bin/query/z?c111:H.R.801: A bill, the *Fair Copyright in Research Works Act*, was introduced by Representative Conyers on February 3, 2009. The law would prohibit any federal agency from requiring a license or transfer of copyright as a condition of a funding agreement. This bill is intended to expressly prohibit policies such as NIH's Public Access Policy or those that would be required by S. 4040.

Congress to obtain digital content through the mandatory deposit provisions of the Copyright Act and legislation to allow it to fully utilize mandatory deposit for online and other digital materials.
- Legislation to allow the Library of Congress to delegate the collection and preservation of certain parts of the national digital collection to private institutions that utilize best practices pursuant to agreed plans for preservation, security, and public access.
- Creation of a national business records archive.

There are many ways in which the laws of a country can support the maintenance and preservation of cultural heritage in private or public hands; the ideas described in this paper are just the tip of the iceberg.

2.3 Creation of a Specialized Archives for Business Records and Other Sensitive or Confidential Records

Business records and confidential, privileged, and other sensitive materials carry additional burdens for preservation. These issues are described in detail in "The Record of Business, and the Future of Business History: Establishing a Public Interest in Private Business Records."[65]

The preservation of digital business records present a separate set of challenges. American

Summary of Policies Necessary to Encourage Digital Preservation Efforts

- Adopt the Section 108 Study Group's recommended amendments to section 108 of the Copyright Act.
- Enact legislation that would (1) permit the U.S. Copyright Office and Library of Congress to delegate the mandatory legal deposit function to qualified entities within the Alliance and (2) allow the Library of Congress and those to whom it delegates its mandatory deposit authority to (a) make a limited number of copies of deposited materials as reasonably necessary to preserve those works and to make them available to users on the premises of the Library of Congress and (b) require deposit in a format suitable for preservation or other needs of the Library of Congress, whether or not a published version.
- Study the possible adoption of:
  —provisions in the U.S. tax code, similar to those available for historic preservation, allowing qualified preservation entities to deduct certain costs of qualified digital preservation;
  —amendments to the U.S. tax code charitable contribution provisions that would make donations of "qualified digital preservation works" to cultural heritage institutions more readily available;
  —authorizing legislation for the creation of a business and confidential records archive that would be immune to discovery, where access would be limited to on-site use by legitimate researchers, and that would allow publication only of aggregate data; and
  —provisions for preservation of research data and results as a condition of federal research grant money.
- Most important, foster congressional recognition of this new, unique, and urgent problem and urge Congress to commit the needed resources to it.

business history, key to an understanding of our history generally, has been fed by the records that businesses have retained, often inadvertently, and that made their way into the hands of archives and historians decades later, when their historic value was apparent. This historic value often is a public value only. Corporations, particularly public corporations whose fiduciary duties are to their stockholders, tend to place little or no value on the historical benefits of their records or on their use to create generalized findings for business history.

U.S. businesses have a strong incentive to destroy their records—the fear of litigation. Technology has made it possible for litigants to search multitudes of documents for a smoking gun. Even if the risk is small that a document creating or proving liability exists, public corporations generally are not willing to take that risk. The litigiousness of the American business world and broad discovery rules in U.S. courts have led to the broad adoption of what are known as "document retention policies"—policies that set out rules for retaining business documents only as long as required by business considerations or by law (e.g., the Sarbanes–Oxley retention rules). Digital technologies have permitted the policies to be automated, implemented, and enforced broadly and effectively across organizations. These streamlined, so-called retention policies are eradicating the record of U.S. business and are paving the way for a future that cannot learn from its past.

Confidential, privileged, and other sensitive materials, such as personal data collected by the Census Bureau, social scientists or health care providers, and law firms, may contain important historical information. Yet these data are of necessity kept locked away under applicable policies, laws, or regulations, and may not be available for research or preservation, while at the same time the owner of the data has little incentive to preserve it. At some point in the future the data may lose their sensitivity and also take on historic value. The data can also provide research value when analyzed in aggregate form in ways that do not violate privacy or other confidentiality codes.

NDIIPP partners investigated ways in which these types of materials may be preserved in spite of the obstacles imposed by their very nature. The program recommends the exploration of the creation of specialized archives for business and confidential records with protocols modeled on those developed by one of the NDIIPP partners for an archive of the digital records of a defunct law firm.[66]

Proposal

Study the possible creation of specialized archives for business and confidential records.

2.4 Preservation Requirements for Federally Funded Research Data and Results

By statute, Congress directed the National Institutes of Health (NIH) to adopt the NIH Public Access Policy, a policy whereby all peer-reviewed journal manuscripts that arise from NIH-funded research must be submitted to the digital archive PubMed Central. The purpose of the policy is to ensure that the public has access to the published results of NIH-funded research in order "to help advance science and improve human health."[67] A bill introduced in the 109th Congress, but never passed, provided for all federal agencies to develop public access policies for research funded by such agency.[68]

Similar legislation could require that the results of federally funded research be deposited in

a certified preservation repository, as described in the following proposal. Such legislation would ensure that access to such learning be preserved for future generations.

Proposal

Enact a statute similar to the NIH Public Access Policy requiring all researchers who receive federal funding to deposit the data and results of the federally funded research in a qualified digital preservation repository. It would be important to identify such repository(ies) by name, as the NIH legislation does; to define the standards for a "qualified digital preservation repository"; or to identify an agency such as the Library of Congress that would have responsibility for certifying such repositories.

3. CONCLUSION: CATALYZE AN ENVIRONMENT CONDUCIVE TO DIGITAL PRESERVATION

A recurring theme throughout NDIIPP's work is that, without public policy support of preservation for historically and culturally important digital materials, much of that content could be lost. There are a number of ways—some of which are identified in this paper—that federal policies could support and encourage preservation of our growing digital heritage without a large financial investment. While state and local policies are important, the creation of federal policies is a crucial first step.

Further study is necessary to determine the practicality, potential effectiveness, and affordability of extending these policies and programs into the realm of digital preservation.

Proposal

- Study the adoption of policies and possible legislative or regulatory reforms that would provide incentives and eliminate disincentives to preserve digital works. Investigate other incentives that might catalyze an environment conducive to digital preservation.
- Consult with experts in the wide-ranging fields from which input would be required, including digital preservation, standards development, tax law, economics, and the technologies required for digital preservation, as well as with representatives from various types of producers and distributors of digital content. The group would be charged with studying the proposals described in this paper and considering other possible policies and laws that would create incentives and diminish disincentives for preserving the vast body of digital knowledge and culture. Separate groups with specialized expertise could be charged with investigating particular sets of incentives. For instance, tax experts would be required in a group that looked at potential tax incentives, while copyright experts would be included in a group reviewing the recommended copyright exceptions, and archivists of confidential data would be essential for creation of the confidential records archives.

Appendix F

Partnership Networks: Recommendations for the Library of Congress NDIIPP[1]

September 30, 2008

1 PROBLEM STATEMENT AND OVERVIEW

An increasing volume of material is "born digital" or is being stored in a digital format. This is due to a number of factors, including the ease with which digital information can be created, the ease of sharing digital information via electronic mechanisms such as the web and e-mail, and the fact that a lot of automatically created content (e.g., geospatial information) originates in digital form. However, digital information can be modified or destroyed just as easily as it can be created. This fact has raised concerns that without adequate attention to digital preservation, we could someday face a "digital dark age."

In December 2000, Congress asked the Library of Congress to address this threat by leading a program on digital stewardship and preservation. This program is known as the National Digital Information Infrastructure and Preservation Program (NDIIPP). The program today consists of a number of initiatives addressing key areas in digital preservation, including content-specific preservation, technical architecture, network creation, and standards. The network of partners involved in NDIIPP has grown to include more than 130 partners, including libraries, archives, universities, research centers, not-for-profit organizations, and for-profit entities that span both a national and international scope. More information about NDIIPP is available at http://www.digitalpreservation.gov.

One of the key goals of NDIIPP is to create a sustainable ecosystem of preservation stewards. It is the desire of the Library of Congress that this ecosystem continue to function well with the Library at the helm but in the absence of continued funding. In late 2007, the Library had initial conversations with the Governance Science Research Group at the IBM Watson Research Center regarding questions of how to create long-term sustainable structures in NDIIPP. The consulting group officially began an engagement with the Library to begin to address these questions in August 2008. This report presents a summary of our findings based on conversations with the Library, conversations with partners in the NDIIPP, and research we performed around some of the key issues in creating a network of stewards to preserve digital content. (See Appendix for a list of the partners who were interviewed). In particular, we address the following key questions in the report:

- What are the characteristics of the networks that the library has already funded which make them likely to succeed in becoming self-sustaining (and why)? Given these characteristics, how can the Library influence other distributed stewardship networks and repositories so that the Library is not the only source for preservation?
- For areas that are not self-sustainable, how can the Library strategize the allocation of funds to maximize investments in the most critical areas?
- What role should the Library play with respect to the partners?
- Has the Library defined the problem of digital preservation sufficiently well?

[1] This report was prepared by consultants from the Governance Science Research Group at the IBM Watson Research Center.

This document addresses these questions in three sections. The first two sections present findings and observations based on our discussions with NDIIPP partners. First, we discuss the range of leadership activities that the partners believe it would be useful for the Library to own. Next, we present an overview of useful governance models that the Library may want to consider as it evolves its role in NDIIPP.

2 FINDINGS AND RECOMMENDATIONS

In the six years since NDIIPP began, the Library has had considerable success overseeing the emergence of a distributed network of partners around key content areas including geospatial, digital television, Web content, digital images, digital sound recordings, and datasets (Anderson 2008). The partners we interviewed confirmed the value of NDIIPP for their own project consortia citing especially their positive interactions with the Library, the Library's role of trusted intermediary bringing partners together with each other and with people outside NDIIPP, how much was learned from the partners including those from different disciplines, and their ability to accelerate efforts in digital preservation that otherwise might not have happened or happened at a much slower pace.

Based on our analysis of the interviews, we see two principal mechanisms for the Library to build on its success to further the goals of digital preservation:

Leadership. Leadership encompasses the ways in which the Library can exert influence to gain additional momentum, resources, and partners, and provide direction and coherence to the projects.

Governance. Governance encompasses the ways in which the Library can influence the structure and functioning of the partners to enhance their success.

We start by looking at how the Library can leverage its world-class reputation and existing network with external as well as internal stakeholders to create leadership around digital preservation.

2.1 LEADERSHIP

2.1.1 SET/GUIDE AN AGENDA AROUND DIGITAL PRESERVATION

NDIIPP has played a key role in creating a public agenda around digital preservation. It has been visible through its website and participation in conferences and other meetings. From the perspective of a technology adoption lifecycle, the Library and the NDIIPP partners have been innovators in their use of technologies and methods for digital preservation. Part of the challenge going forward is to leverage the current successes to bring additional participants to the program. By virtue of its international reputation and its position as a government agency, the Library has an opportunity to be an important voice in expanding the agenda around digital preservation to bring it into the mainstream. This is important not just for the general goals of digital preservation but for the partners as well. By making the agenda more visible to a broader population, it helps the NDIIPP partners gain increased visibility and credibility, which in turn helps them gain funds and resources for their projects. It also helps position the work on NDIIPP in the context of a larger national and international effort.

There is a limit to the amount of material that can ever be feasibly preserved. Therefore it is important for the Library to set guidelines and priorities to allow wise decisions to be made about what should be preserved and what may be safely ignored. These are not decisions to be taken

lightly, so the Library will want to partner with other groups to outline its agenda. The Library is viewed by the partners we interviewed as being the right group to set the vision and priorities for digital preservation in general and collections in particular—what collection areas are most important, what pieces are most at risk. Guidance from the Library would ensure that more attention is paid to critical areas than might otherwise happen and prevent the scarce resources of librarians and archivists from being spread too thin. The Library can help set priorities and help facilitate the organization of content domains dedicated to particular types of collections. This also helps to build a shared commitment to the preservation of particular types of data or content.

2.1.1.1 RECOMMENDATIONS

- **Promote broad concerns of digital preservation**. On those occasions when legislative acts are needed to address significant issues that are common to all or most of the partners (e.g., issues related to copyright), the Library should facilitate the communication with Congress.
- **Create a visible agenda for digital preservation**. The Library can foster scholarly discussion about preservation priorities with NDIIPP partners and others by forming and enabling a task force. One method for communicating the Library's priorities might be an "endangered species" list for collections that are at immediate risk of being lost.
- **Facilitate adoption of digital preservation by new groups**. The Library may also develop a re-usable set of mechanisms and best practices that will help communities build collections in particularly critical areas.

2.1.2 BRAND DIGITAL PRESERVATION

Digital preservation is still a relatively new concept and initiative. One way of increasing awareness of it, as well as enhancing the legitimacy of the process, is to "brand" material. This would provide two benefits. First, it would distinguish material developed within the context of NDIIPP as trustworthy. Second, branding would increase the visibility of NDIIPP as a whole.

2.1.2.1 RECOMMENDATIONS

- **Make current digital preservation efforts more visible**. Publicize the importance of digital preservation and the Library's leadership role by providing a "seal of approval" for content that has been selected and approved by the Library. The actual form this seal would take is not specified here but could be something like "powered by the Library of Congress" or "enabled by NDIIPP." Examples from the Department of Energy and National Science Foundation might be an appropriate starting point.

2.1.3 PROMOTE KEY RELATIONSHIPS AND COLLABORATIONS

Digital preservation is a distributed problem that no single institution or discipline can solve. The Library has acknowledged this by recognizing the importance of content, technology, and research, and by using funding to bring diverse constituents together. It has also had success in getting projects to share what they have learned by hosting regular partner meetings. We heard repeatedly that these meetings were valuable both for the networking opportunities as well as for getting feedback, direction, and validation on project work. The Library has also been very active in supporting open source projects that enable groups to reuse and leverage technology advances of others. The Library's support of the LOCKSS technology has been especially successful in this regard and has led to its adoption by a number of groups, notably the MetaArchive Project, which used LOCKSS as a catalyst for informal and more formalized community formation.

The continuing focus on digital preservation is going to be challenged by a reduction in targeted funding. Librarians and archivists, who form the majority of lead partners, are already well aware of the need for digital preservation and the processes by which it happens and will continue to seek ways to continue their work. However, there is a need to educate a broader population, especially content providers, about the need for and importance of digital preservation.

Required capabilities for effective digital preservation include management of standards, architecture, tools, processes, procedures, policies, and roles. Partners also need to bring together people with different knowledge and skills including:

- People and institutions who provide the content as well as the rationale for how and why that content will be used. For instance, in the case of a project focused on geospatial data we heard that state and local government needs a historical record of the data for schools, tax assessments, and other public works projects.
- Computer scientists who understand how to develop new technologies and/or adapt existing ones to achieve all the steps of digital preservation.
- Librarians and archivists who have curatorial expertise, knowledge of document formats, and how the preserved artifacts are managed and used.

Bringing these perspectives together can be challenging because each group often has a different agenda and speaks a different "language." (One key to success is to establish some kind of common ground. We will say more about this topic in the next section.) Effective knowledge sharing, especially for diverse, distributed groups, depends on high levels of social capital that are built on a foundation of trust and reciprocity.

2.1.3.1 RECOMMENDATIONS

- **Partner meetings**. Continue sponsoring partner meetings since these are universally viewed as valuable opportunities to share project information, network with other partners, and gain knowledge of a broader but relevant range of topics and issues.
- **Build social capital**. By virtue of its leadership role in NDIIPP, the Library's relationships with the partners are very visible and set expectations for others to follow. By making its trust in the partners visible, the Library models the kind of behavior that others can emulate, promoting the necessary social capital to keep the network vibrant.
- **Broker relationships**. The Library can broker relationships between content area communities and the technical communities supporting them.
- **Trusted intermediary**. The Library is recognized by the partners for its role as "trusted intermediary," especially in terms of brokering relationships between universities that might be competing with each other for grants and professors. The Library can use its reputation to foster communities of practice around specialized topic areas.
- **Clearinghouse for funding**. The Library could help provide funding to organizations or support their search for funding by serving as a clearinghouse for funding sources and approaches.
- **Alternate funding models**. There are several alternate funding models. For example, funds could be solicited from content consumers for appropriate collections. Another possibility is a membership model in which members pay dues to support the content area community. MetaArchive is an example of this kind of model.
- **Outreach programs**. The Library can start outreach programs with a broad spectrum of content providers, including those in the private sector, to establish awareness of the importance of digital preservation. This can drive demand for digital preservation services.

- **Education**. The ecology of preservation requires more education of researchers and others about the need for preservation of digital content.

2.2 GOVERNANCE: MODELS, ROLES, AND RESPONSIBILITIES

The Library faces a number of difficult challenges as it prepares to pull back funding from the partner organizations, including distributed collaboration, organizations with varying interest and incentives, a complex legal/regulatory environment, limited partner motivation to participate in the face of dwindling financial incentives from the Library, and the heterogeneous nature of its communities. As part of our analysis, we evaluated existing governance models of successful, stable, reasonably long-running organizations that have many of the same key characteristics as NDIIPP. These characteristics include geographically distributed members (some previous collaborators and some well-known competitors); little or no financial rewards available to serve as motivation for network participants; and heterogeneous communities involving academic, research, and commercial members. Each member had different goals for the network, different contributions to bring to the network, and different benefits they wanted to accrue from their participation. We also sought organizations that faced some of the same issues that the Library of Congress faces with NDIIPP, including contracts, intellectual property, protecting the network from the sudden departure of a member, and ensuring that for-profit organizations would be welcomed and could profit within the structure of the network. Our goal was to identify existing governance models that have already been demonstrated to be viable, and to explore whether and how they could be adapted for use for NDIIPP.

2.2.1 LESSONS FOR NDIIPP FROM THE ECLIPSE AND APACHE FOUNDATIONS

Two organizations in particular struck us as similar to the NDIIPP along these dimensions: the Apache Software Foundation (http://www.apache.org) and the Eclipse Foundation (http://www.eclipse.org). Open-source software development as a whole shares much in common with the NDIIPP network, though the domains and participants are quite different. These two organizations have reached prominence and evolved well-defined governance structures after starting out as much smaller, less well-organized groups. Eclipse and Apache grew out of fairly different communities of developers, so it was intriguing to note that these two organizations converged on a similar, and successful, governance structure.

The following lessons learned from the Eclipse and Apache foundations are particularly noteworthy for NDIIPP:

- **The foundations provide support for the community of projects.** Similarly, NDIIPP is itself intended to be a supportive community of projects (content preservation efforts), perhaps eventually steered by some entity analogous to the Apache and Eclipse foundations.
- **Projects are characterized by a collaborative, consensus-based development process.** Just as open-source software development seeks to address project concerns in a collaborative, agile way, we believe successful NDIIPP partners will be characterized by collaborative and community-driven governance of their projects.
- **An open and pragmatic software license.** Both Apache and Eclipse had to strike a balance between open availability of their software and commercial participation in order to succeed. Doing so allowed an ecosystem of projects to flourish, protected from the burden of independent legal negotiations and the creation of licenses. Although not the same, the questions NDIIPP faces concerning copyrighted content appear to be analogous. Addressing copyright concerns in a centralized way can similarly save the

NDIIPP partners from this difficult task.

- **A desire to create high-quality software that leads the way in its field.** This is a key tie that binds the members of all of the most successful open-source communities—the desire to produce something that is high quality, to see the results be widely used and expanded, and to be recognized as important contributors to the field. We have seen a corresponding desire to do high-quality digital preservation among several of the current NDIIPP partners.
- **Not simply a group of projects sharing a server, but rather a community of developers and users.** The importance of community building and community support cannot be overemphasized as a key success factor for both these organizations and for other successful open-source projects.

Successful governance structures reflect and promote the goals of an organization, and given the similarities between the NDIIPP network's goals and those of the Eclipse and Apache foundations, it is not hard to imagine that some of the governance models that Eclipse and Apache use might also be appropriate to NDIIPP.

In the remainder of this section, we present a distilled version of the Eclipse and Apache foundation governance structures, mapped as we see them to the NDIIPP domain. The next section expands on this to identify specific roles and responsibilities that we see as important for an NDIIPP network, including both Library-only roles and roles that can (and hopefully will) be assumed by others in the network. We will generally assume some form of merit-based approach to assigning roles—anyone may participate in the network (subject to whatever restrictions the Library of Congress and/or NDIIPP choose to impose), but people must participate in the community for some period of time and demonstrate both their commitment and their abilities before they are assigned significant roles—both to ensure that the best available people fill each role and to help mitigate the risk that someone might take on a role and then leave abruptly, causing significant disruption in the network. We assume—and recommend—the merit-based approach because it has demonstrated itself to be both reliable and well accepted by members of the open-source community, and it has also been used successfully in many other not-for-profit organizations for a long time.[1]

2.2.2 ADAPTING A GOVERNANCE MODEL FOR NDIIPP

As part of our research, we spent some time harvesting critical features of the Eclipse and Apache governance models—features that we saw as critical to their success—and evaluated them with respect to the requirements of, and goals for, NDIIPP. This section presents a similar, though somewhat different, governance framework that maps to the requirements and goals we gathered from the Library and its current NDIIPP partners for the NDIIPP network. A key goal of this effort was to identify a structure that might enable NDIIPP to leverage members' strengths and motivations for participating in the network to ensure that they fill the right roles and obtain the benefits they expect in exchange for their participation. Ensuring that the needs of the participants are satisfied—as well as the needs of the network—has been of paramount importance to the success of the open-source communities we have studied, and we anticipate that this will also be the case for an NDIIPP network that functions without Library of Congress funding.

[1] The merit-based approach has provided some interesting and unexpected benefits to some members of the open-source community. In particular, although they do not receive financial remuneration for their work on open-source projects, developers who rise to levels of influence and trust have been able, in some cases, to use their acknowledged success to attain better paid positions for their "day jobs." They have also built up significant social networks, which they have been able to leverage when they wanted or needed new positions.

2.2.2.1 Framework

As we interviewed some of the current NDIIPP partners, two different dimensions of concern for organizing the NDIIPP network—i.e., two different but equally relevant ways of dividing the members of the NDIIPP community—emerged (see Table 1). One dimension is the *content domain*, such as "geospatial" or "Southern culture." The shared focus on a particular content domain appeared to be a critical factor in the success of the NDIIPP sub networks—that is, partners who shared a commitment to preserving content in a given domain tended to bind more strongly, and to be more successful in the Library's view, than those who partnered for other reasons.

The second dimension of concern is *preservation issues*. Table 1 is intended to be suggestive, not exhaustive, but a few key preservation issues we heard about included archival strategy (the approach taken to choosing, cataloging, and organizing digital content), consumer applications (applications that enable end users to query, navigate, or otherwise manipulate preserved content), and technologies (notably, those that enable the physical storage, distribution, and secure access to content).

Table 1. Sample NDIIPP Framework Dimensions

Sample Preservation Issues	Geospatial	Chinese	Southern Culture
Archival strategy			
Consumer applications			
Technologies			

Depending on one's focus or goal, one could profitably identify and promote different, successful communities based on either horizontal or vertical "slices" through Table 1. Thus, for example, the MetaArchive subcommunity of NDIIPP is already dedicated to the preservation of Southern culture (a vertical slice). There is also a horizontal slice corresponding to general preservation technologies (something like LOCKSS) that apply to multiple content domains. We have noted, however, that it appears to be critical for participants in horizontal slices to participate in vertically sliced communities to be successful; in other words, the content domains are of paramount importance in addressing preservation issues. In cases where groups have sought to address a horizontal slice without specifically understanding and addressing the needs of one or more vertical communities, we have seen the production of unfocused results that were less successful in the Library's evaluation. Conversely, where vertical communities have not worked closely with appropriate experts in each horizontal area—such as technology—those communities have ended up ignoring critical issues, reinventing knowledge, or failing to produce consumable results. In short, the vertical and horizontal communities need each other to be successful and some care should be taken to explicitly coordinate their communication and alignment. We suggest introducing the following roles in the NDIIPP network:

- **Facilitate "vertical integration."** As noted above, only a subset of the preservation issues were addressed by a particular community; limitations were notable on the technologies front. Ensuring that some person or organization addresses each of the preservation issues for a particular content domain community—and hence, achieving vertical integration across all the horizontal slices—appears to be important.
- **Promote "horizontal integration."** Although we do not know at this time whether it will make sense for all NDIIPP partners who are addressing a particular preservation issue (horizontal slice) to form active communities within the network, it is clear that at

least some interaction horizontally will be important. In particular, we note that standardization efforts are likely to arise across horizontal slices, as different technologies, applications, and archive strategies are demonstrated to be "best of breed," and as the digital preservation domain matures enough to begin articulating standards beyond the several that have already emerged. There will undoubtedly be multiple standardization-related efforts, instituted across multiple standards organizations, and this will certainly require management and catalysis.

We have also identified several "cross-cutting concerns" that do not properly fit into either the content domain or preservation issues dimensions, yet touch both. We use this term to refer to issues that affect multiple areas of the framework—e.g., multiple content domains, or some or all of the technologies. These include:

- Copyright issues;
- Standardization efforts, which may vary significantly across content domains, technologies, applications, etc.;
- Community-wide issues, such as organization of community-wide conferences, shared infrastructure support (such as message board hosting), addressing pervasive legal issues (e.g., creation of license agreements), coordination efforts to align community goals and work; and
- Community continuity and direction issues, such as helping to identify and promote benefits for participants and identifying and managing risks to the network.

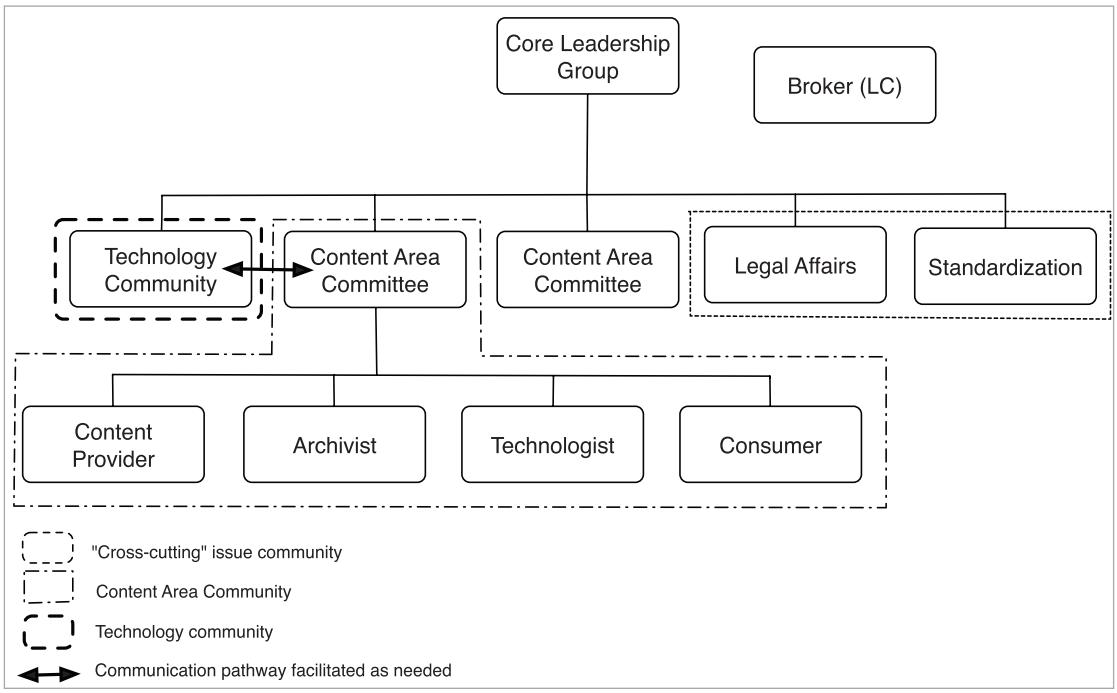
Each of the communities that addresses (subsets of) these cross-cutting concerns must work closely with the horizontal and vertical communities affected by the particular concerns to ensure that the particular needs and issues in those communities are addressed. For example, the Library itself has taken on the cross-cutting copyright issue that affects many of the NDIIPP partners. In so doing, it has had to look carefully at the particular copyright issues arising in each content area (vertical slice) and preservation issue community (horizontal slice) to ensure that subsequent legislation addresses the problems appropriately.

2.2.2.2 COMMUNITIES, ROLES, RESPONSIBILITIES, AND COORDINATION

Given the frameworks described in the previous section, we illustrate some key features of a potential governance structure for NDIIPP in figure 12.

A **core leadership group**, including the Library, might be tasked with defining directions for, and governing, the NDIIPP network as a whole. By sharing responsibility for the overall governance of NDIIPP, the Library frees up some of the time it previously focused on the partners, which it can now use to advance the leadership agenda outlined in the previous section. The Eclipse Foundation's bylaws identify different kinds of directors, and a similar kind of structure might work for the NDIIPP network. *Strategic Developer Member* organizations (ones that commit a certain number of developers from their organization to doing Eclipse-related work full-time, plus who pay significant annual dues) and *Strategic Consumer Member* organizations (which make significant resource commitments to the Eclipse Foundation) are each allowed to seat one director from their organization. Other director positions come from the Eclipse community and/or are elected. The role of core member is a significant one, as the core members help shape future directions and make policy decisions for the network as a whole; hence, it should be given only to those who have demonstrated a significant commitment to the NDIIPP network and have also demonstrated the vision and influence required for this role.

Fig. 11. A Potential NDIIPP Governance Structure



One of the cross-cutting issues is addressed by a specific community within both the Eclipse Foundation and the Apache Software Foundation: namely, a legal department. Both foundations have encountered significant legal hurdles in achieving their goals, in a manner not dissimilar to what several NDIIPP members reported. These hurdles include protecting those who adopt or extend the Eclipse or Apache technologies from lawsuits over intellectual property issues (a particularly critical issue for commercial organizations that participate in the communities and ecosystems). Hence, both foundations have legal counsel whose duties include such activities as creating "safe" licenses under which all foundation software is required to be offered; creating any required legal contracts between the foundation and members; and addressing provenance issues (to ensure that developers donate only code that they themselves have written as a protection against copyright/patent/other intellectual property infringements). The Eclipse Foundation employs its own legal department; we were not able to determine what the Apache Foundation does. NDIIPP does not yet appear to have any such organization to aid the members of the network with common legal issues, but providing one may lessen the shared legal burden, improve the odds of a successful network, and provide a solid benefit for participating in the network.

Each **content management committee** might be responsible for one content domain, including its metadata, tools, services, and access. These committees, along with the set of people who preserve content in that domain, the technologists who provide the hardware and software that enables the preservation, the service providers who may ultimately offer value-add services on top of the preserved content, and all who participate in the content domain (e.g., by reporting problems, offering insights, participating in discussions), form one of the vertically sliced content domains shown in Table 1. The content management committee corresponds to concepts such as PMC (project management committee) that are found in the open-source governance structures. In both the Eclipse and Apache communities, these content areas are

required to operate openly and transparently, and according to the best practices of the open source community. Metrics are collected on an ongoing basis, and areas that are found not to be operating openly or transparently, or that have very low or no activity, are subject to termination or other actions. Although metrics are difficult to employ early in a network's lifetime, NDIIPP has existed for long enough that it may be time to consider collecting information about the network to help identify what is working well and what is not while it is still early enough to correct the problems and to identify and promote best practices that emerge.

The approach we have outlined here is based on ideas of networked governance aligned with those espoused by Milward and Provan (2006). They propose that the management of tasks in public networks depend on management of accountability, legitimacy, conflict, design (governance structure), and commitment. In addition, we suggest that loosely coupled networks such as NDIIPP also depend on two important brokerage roles: people who bridge between groups or functions, and people who help bond a group around a particular idea or content area. In both cases, the brokers help create a more robust, and hence more sustainable, network that also has an opportunity to grow. We have not yet had a chance to identify examples of these people within the current NDIIPP structure but we hope to do so in the next phase of work.

We believe—both from prior experience in the open-source and other communities, and from what we heard from some of the NDIIPP partners—that the channels of communication across different subcommunities of the NDIIPP network are absolutely critical for the network to grow beyond its current bounds; indeed, these communications already appear urgent for the continued well-being of the present network. For example, we learned from the NDIIPP partners that there is a dearth of mechanisms for delegating problems. One key role in any network is that of broker—one or more persons who identify subcommunities that should be communicating with one another to achieve some shared goal. Of course, as the Library has pointed out, when issues require legislation or other heavyweight solutions, they may take considerable time to resolve. This is a complex area that requires more research.

Library of Congress broker roles: In sections 2.1.1–2.1.3 we discussed a set of leadership actions the Library could take and their associated roles. For completeness we recap these roles below:

- **Trusted intermediary**. Since trust among peer organizations appears to be low at times, the Library may be the only organization trusted by all the communities. In this capacity, the Library may facilitate critical partnerships and mediate critical issues.
- **Channel to Congress**. On those occasions when only legislative acts can address significant issues facing the network as a whole (such as the pervasive copyright law problems that all the partners reported), the Library is clearly the only member of NDIIPP with the ability and mission to seek legislative redress (e.g., the Section 108 Study Group).
- **The vision**. The Library is viewed by some members of NDIIPP (and undoubtedly elsewhere in the preservation community) as uniquely positioned to set the vision and priorities for collection. In this capacity, the Library would articulate key goals that the preservation partners would pursue and help ensure that progress was being made toward reaching them.
- **The brand**. It apparently means something significant for a physical document to be labeled as a part of a Library collection. For this to continue to be the case in a digital future, the Library will need to continue branding digital collections as well.

3 CONCLUSION

In this document, we have presented an overview of the Library of Congress NDIIPP initiative and a set of high-level recommendations for the Library to address the long-term sustainability of a network of preservation stewards. The recommendations cover such activities as program branding, preservation agenda management, collaboration, and governance.

The Library has a strong and successful start toward preserving at-risk digital content, and we hope that our recommendations are useful in the development of a robust, sustainable network of preservation stewards.

REFERENCES

Anderson, M. (2008). Evolving a Network of Networks: The Experience of Partnerships in the National Digital Information Infrastructure and Preservation Program. *The International Journal of Digital Curation*, 1(3): 4-14. Retrieved September 30, 2008, from http://www.digitalpreservation.gov/library/resources/pubs/docs/anderson_ijdc_2008.pdf_

Milward, H. B., & Provan, K. G. (2006). A *Manager's Guide to Choosing and Using Collaborative Networks*. Washington, DC: IBM Center for the Business of Government.

Raymond, E. S. (1999). *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. Sebastopol, CA: O'Reilly Media.

APPENDIX: NDIIPP PARTNER ORGANIZATIONS INTERVIEWED FOR THIS REPORT

We conducted one-hour phone interviews with the principal investigators of six current NDIIPP projects, recommended by the Library. The interviews were conducted between August 22 and September 9, 2008.

Affiliation	Project
University of North Carolina	North Carolina Geospatial Data Archiving Project (NCGDAP)
Emory University Library	MetaArchive; MetaArchive Cooperative
North Carolina Center for Geographic Information and Analysis	Geospatial Multistate Archive and Preservation Project (GeoMAPP); NCGDAP
University of Illinois at Urbana-Champaign	ECHO DEPository Project
California Digital Library	Web-at-Risk
Stanford University	LOCKSS

Appendix G

National Digital Collecting Strategy

The need for up-to-date, reliable, and comprehensive information is greater than it has ever been. Vital decisions about the environment and energy, education and research, health and economic recovery, and disaster preparedness and security sectors are being made in a complex information landscape. Decision makers must be able to rely on digital stewardship organizations to select, capture, and preserve important resources for analysis now and over the long term.

Stewardship organizations must adapt to new modes of publishing and authorship, massive volumes of material, and ever-rising user expectations for immediate, integrated access. A collecting strategy in the second decade of the 21st century must be undertaken with strong partnerships with other stewardship organizations and digital content stakeholders. Without preservation in the present, there will be no access to digital information in the future.

The National Digital Information Infrastructure and Preservation Program (NDIIPP) has worked for a decade to build a network of trusted stewardship organizations. NDIIPP partners are experts in selecting, collecting, preserving, and providing access to a wide variety of digital materials, from website archives to social science research data. The technical infrastructure developed by NDIIPP and its partners is flexible, and able to respond to constant changes—new formats, new uses, and new users. The NDIIPP network, now formalized as the National Digital Stewardship Alliance (the Alliance), is a group of institutions committed to working together in service of the national digital collection.

Many traditional forms of materials for Library collections—namely academic journals, newspapers, motion pictures, research data, and government documents—are changing rapidly from paper and film to digital forms. Content distribution is moving from a limited group of publishers to many individuals. Anyone can create, re-create, and rapidly share digital content, creating a dynamic and changing landscape. In addition, content that is produced and distributed digitally often have audiovisual or other interactive elements embedded into a work. New media and more traditional publishers are all evolving to take advantage of new digital delivery systems and formats to better serve their users.

Examples include the federal government, which hosts over 24,000 websites publishing information for citizens that used to appear only in print. News outlets such as the *Christian Science Monitor* have ended their print operations and are publishing solely on the Internet. Authoritative legal commentary and discourse that previously appeared in professional journals are now available only through blogs on the Internet. Magazines like *Wired* and *Time* are producing applications and multimedia content for mobile devices.

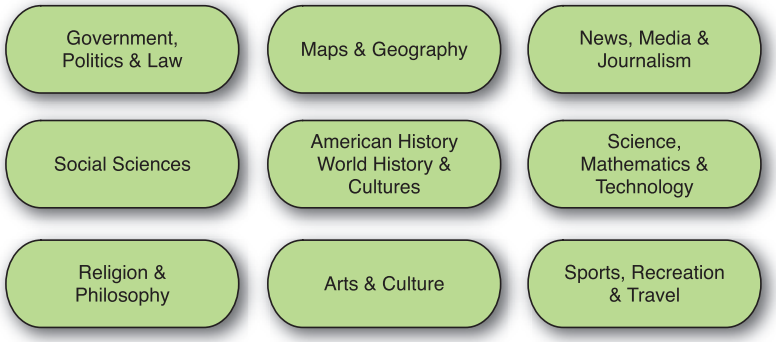
Developing flexible and realistic collection strategies in this environment demands fresh approaches to identifying and collecting content with long-term value. Developing such strategies begins with a framework identifying the content that will be in demand in the future. Working within this framework, the NSDA can leverage complementary expertise, commitment, and resources to anticipate future demand.

Based on these principles, below is a description of the national digital collection framework. It outlines the a ten-year implementation plan, with special priority given to public policy resources. It also explains the major collecting needs and opportunities that the Library and its Alliance partners will encounter in this new dynamic information environment. It summarizes the results of the Library's first steps in building a national collection of born-digital content: convening expert communities of practice—content creators, distributors, and users—to identify collecting priorities for the near term. Finally, it identifies the key benchmarks of success and failure as we look to the future when the government, education, research, and culture of our nation will all be produced, distributed, accessed, and stored digitally.

National Digital Collection Framework

The Library of Congress, in partnership with the Alliance, has developed the national digital collection framework to guide the growth and maintenance of a national digital collection in the coming decades. This collection framework builds on centuries of the Library collecting, preserving, and providing access to high-value content to the nation and its government, and moves this tradition firmly into the digital age. The framework is rooted in an ongoing evaluation of the access needs of the nation, and targets those areas of particular interest to public policy, education and research, and cultural heritage—the traditional preservation and access strengths of the Library.

Flexibility and the capacity to adapt to changes in the information landscape is crucial to meeting the needs of tomorrow's users. The Library is able to respond to dynamic changes in the digital environment because of the broad reach of its distributed network of collecting organizations: over 185 partners from 44 states and 25 countries, and growing. The high-level scope of the national digital collection framework includes public policy and cultural heritage subject areas. The framework builds on the knowledge gained from NDIIPP to guide the development of a national digital collection in the coming decades.



The National Digital Collection Framework includes a broad scope of educational, cultural heritage, and public policy subject areas; priority areas for 2010–2013 are those most relevant for public policy.

A national digital collection starts with each stewardship organization setting local priorities that build on local strengths. Sustainable digital collections must begin with:

- identifying the new ecosystem of creation, dissemination, and use;
- understanding the new dynamics of content creation and use; and
- developing digital collecting strategies that are firmly aligned with these new dynamics and support sustainability.

Collection priorities: Content relevant for public policy

Primary and secondary sources that libraries collect take all conceivable forms, both established and emerging: geospatial, pictorial, audiovisual, and textual. Content targeted for acquisition and preservation comes from numerous sources: public, corporate, and individual. Federal, state, and local government agencies create and publish public information. Commercial and non-profit publishers of news and commentary, and commercial and independent multi-media producers are key sources for many materials that will become part of a national collection.

Priority subject areas for 2010–2013 are those most relevant for public policy.

- Government and legal information includes federal, state, and local government agency documents, records and publications, videos and podcasts, digital photographs, websites, and blogs produced by government agencies and nonprofit organizations.

- News and journalism information includes television and radio broadcasts, websites, blogs, videos, and podcasts produced by public and commercial producers, as well as citizens.

- Geospatial information is available as digital maps and geospatial data sets produced by government, non-profit, and commercial sources and collected from satellites.

In addition, highly relevant sources for all collecting areas include social science data sets, surveys and interviews, electronic literature, journals and news publications, web-based government and industry sources, and relevant foreign sources. The digital collection also includes materials digitized from analog library holdings, such as newspapers, manuscripts, books, maps, photographs, film, and sound recordings.

Anticipating future demand for digital content

What is different about collecting in this new environment? In some ways, much remains the same: the end goal of collecting is to provide reliable access to relevant information in a timely fashion to policy makers, researchers, journalists, scientists, teachers, and students of all ages. Many of the acquisition decisions that the Library and others made decades ago are still relevant. Many of the topics of greatest import for public policy content users remain the same. But in the digital environment, identifying and securing these kinds of high-value resources is challenging because of the volume of content created, the multiple sources from which the content comes, the poorly documented and often ephemeral nature of the content, and the many novel uses of content made possible by the power of computing.

Just as we could not have predicted Web 2.0 in 2000, when NDIIPP started, or anticipated the economic disruptions of 2007–2010, we cannot predict with certainty what new technologies, computing devices, or formats will appear. We should anticipate that the changes in content creation and use will have dramatic and unforeseeable shifts, and that stewardship organizations must constantly assess the information landscape and adapt to change. How can a national collection strategy adequately manage such uncertainty?

As the *Blue Ribbon Task Force on Sustainable Digital Preservation and Access* determined after

studying the nature of future demand for digital information, the best method for anticipating future uses is to look at current uses. For example, we see the use of demographic information in conjunction with maps to assess population trends in many instances, from the local to the national and international. This tells us that demographic information—always crucial for policy issues—is more useful when there are good mapping interfaces available. When policy makers analyze and compare several different options for a piece of regulatory legislation, for example, they want the best demographic data and current base maps and an array of geographic information services to query the data, and ways to model different possible policy outcomes, and ways to display the results in easily-legible maps for comparison.

In addition to developing such current use cases that will recur, it is important to track emerging trends, such as data mining across many sources to find patterns. For example, when looking at healthcare policy options, the Congressional Research Service (CRS) will use demographic and public health sources—geospatial mapping technologies to look at regional distributions, consult the legislative history, read various legal sources (including blogs by legal experts), study evidence-based reports from policy groups, and so forth. Sources become more valuable the more they can be used with others. Computing technologies can mine existing data for information hidden due to the sheer volume. As the Task Force urged, especially in cases of uncertainty, it is vital to act now to preserve (at low levels of curation when possible) in order to leave open the option of selecting for re-use and retention later.

Convening Stakeholders and Communities of Expertise

The Library and its partners rely upon a network of domain experts to keep abreast of current and innovative uses and emerging trends in content creation and distribution. These experts come from a range of knowledge domains, in both the public and private sectors, and include both those who create and use content, librarians and archivists who collect and serve content, and experts in various technologies.

In the summer and fall of 2010, the Library hosted high-level discussions about content selection in three subject areas germane to public policy:
1. government, politics, and law;
2. maps and geography; and
3. news, media, and journalism.

The high-level questions that each convening session addressed included:

- Selection—What will users need in the future? What are the expected time horizons of use? Who creates the content and who owns or has control over it?
- Collection—How can we identify and collect sources with the greatest efficiency? How will the data be tagged or cataloged for easy retrieval?
- Preservation—Who will collect and preserve the content?

The convening sessions focused on describing the ecosystem of content production and use, and of shared responsibility for preserving among producers and users, as well as traditional libraries and archives. Participants heard testimony from a variety of expert communities. They compared relative priorities among different types of information. In the end, they made recommendations about targeted collecting efforts based on needs and opportunities. They evaluated ways to align top-down and bottom-up approaches for identifying high-value con-

tent, leveraging community-based expertise and commitment, and ensuring that the information meets local requirements for vetted and validated content. The group also evaluated ideas for providing incentives for individuals and groups to do self-archiving.

While the communities of practice in the three collecting areas have distinct profiles in many ways, it significant that these stakeholders share many of the same challenges and see very similar opportunities. In addition to these stakeholder meetings, in 2008 the Library conducted in-depth interviews with experts in Web archiving, geospatial information, broadcast news, and e-journals to identify key opportunities and challenges in collecting and preserving content from these sources.

Digital Content for Public Policy: Needs

The Library of Congress has a long history of serving key bodies that create, regulate, and advise on public policy: Congress and other branches of government, independent policy groups, universities and research institutions, among others. The need for reliable, authenticated, transparent, and persistent information to inform and monitor public policy is greater than ever. And the public's right to know about its government remains paramount. In effect, the collecting priorities in this area have not changed in substance in the digital realm; high-quality information must be available to serve Congress and other bodies in a timely and reliable way.

What public policy users need

Policy makers and those who serve them, such as the Congressional Research Service and Law Library of Congress, need information about national and international affairs. They need to:

- integrate analog and digital sources seamlessly;
- have facts relevant for an issue and its context and history;
- track the history of an issue over a long period of time;
- use raw data for analysis and the analysis itself;
- have persistent access to reliable and authentic content; and
- have sources in multiple formats and languages.

Digital content will be visualized, mapped, used, packaged, and re-packaged. For the purposes of accountability, the whole process of research, analysis, and presentation needs to be well documented and transparent.

None of these are fundamentally new requirements, and in many ways digital delivery can make these processes faster, more efficient, and better sourced. That said, there are major disruptions in the production and dissemination of public policy sources which create entirely new challenges. Many of the news producers—from network to cable and print to radio—are themselves going digital and making heavy investments in technology to deliver content faster and in new ways. But these enterprises are simultaneously struggling to find the right business model in a radically altered journalism environment. New workflows, digital initiatives, and publishing streams are launched in a constant cycle but the preservation of new digital content is often neglected. Documentary sources that have been crucial in making policy can disappear over time, breaking the chain of accountability and crippling future policy decisions that need to understand the historical context of past events. Government publications are also at-risk; they are now distributed exclusively online, and can disappear with each monthly update.

Digital Content for Public Policy: Opportunities

These constitute fundamental changes that will have permanent effects on the information landscape. There are four salient trends currently shaping the creation and use of high-value content:

1. the rapidly growing use of geospatial sources in all knowledge domains;
2. the high impact of social media on both content creation and use;
3. the erosion of boundaries between the public and private spheres online; and
4. the ready availability of local and regional content and perspectives online.

These trends represent extraordinary opportunities to improve service to users by getting greater value out of existing content, leveraging networks of expertise, and building cost efficiencies into digital collecting strategies.

Growing use of geospatial sources

A new element in this environment is the ubiquity of geospatial imagery and data that can provide an unprecedented context for analysis and presentation. Events happen in time and space, and are shaped by them. Maps and mapping tools allow a sharper and more accurate picture of the context in which events unfold over time. Geospatial data are crucial for studies of energy and environment, global health issues, foreign policy and global conflicts, endangered species, and agriculture and aquaculture worldwide.

Congress and state legislatures have a particular need for accuracy in spatial data with respect to public policy legislation. Agencies and activities such as the U.S. Census Bureau also need precise, granular data, at the same time they need to control for privacy issues. Every agency—from the Federal Aviation Administration to the Environmental Protection Agency, the Fish and Wildlife Service to the National Oceanographic and Atmospheric Administration—have high-volume demands for both historical and contemporary data. With respect to both human and natural resources, tracking flows over time and space are very significant.

The challenges for archiving geospatial data and technologies for access are significant. Geospatial data formats are among the most technically complex. Base-layer maps, and the software that makes them useful, are often proprietary. There is a need to engage with the private owners of this data and give them incentives to preserve in the public interest. In government agencies, preservation mandates require better processes and compliance, and an evaluation of existing selection criteria. Building preservation processes into content creation would not only lower the cost of preservation and ensure the integrity of the data, but would also greatly reduce future costs for access to these data.

Libraries also need to provide models for smaller organizations willing, but currently unable, to preserve. Built on a collaborative model, shared responsibility in the NDIIPP-sponsored GeoMAPP group (Geospatial Multistate Archive and Preservation Project) shows how self-organizing groups can leverage economies of scale among many under-resourced agencies in states and regions.

High impact of social media

We have witnessed a remarkable phenomenon on the Web since the growth of user-generated content that started with Web 2.0. Users are now creators as well as consumers, enabled by

social-networking formats such as blogging and tweeting, and wiki formats that allow collective production of content.

As more government, news, and policy sources go online, there is an embrace of social media to allow users to comment, download, and modify, and themselves contribute content back. The line between the public and private spheres, once so well demarcated, has eroded. As more citizen commentators and policy activists create online conversations, the volume of content to consult, collect, and preserve grows exponentially. Validating that content, assessing its value, and trying to secure it for long-term access is challenging. As one web archivist has noted, "Because so much of Web content is user-generated, there is much less control and standardization and adherence to guidelines, which makes it particularly challenging."

The need to preserve the output of creators such as "citizen journalists" is crucial. As the 2009 Iranian protests show, news can be a video taken on a mobile phone, or a tweet shared with the world. Even traditional corporate news sources are putting more and more content online and soliciting input from the public. For example, "Post a comment" and "Send us your photos" links on newspaper sites are becoming more common. What broadcasters air is only a small portion of what they make available online. These "outtakes" are very heavily used and often get repurposed for documentaries and other productions.

In addition, the commentary that is so widespread online has become an influential—if problematic—source of public opinion and persuasion. It is equally important to capture the mood of public opinion, including those websites that influence people's political persuasions.

Public and private boundaries eroding

The blurring of distinction between public and private spheres is the direct result of people using the Web to post information about themselves to their peer groups, and in the case of social networking sites like Facebook, to potential peers and "friends" whom one has not even met. Because the Web is also the medium for posting official information circulating in the public sphere, there is a curious—and most likely unintended—blurring of boundaries between the two. This erosion of boundaries between public and private on theeb is accompanied by a blurring of the distinction between news and entertainment; a trend that began in the 1970s, but has accelerated on the Web.

This blurring of public and private, news and entertainment, poses interesting challenges for collecting. "Is it private data even though it is published on a public website?" is one question preservationists are often asked. Web archivists note that "The general trend is to assume it is fair use to collect the stuff that is publicly available on the Web, but that redistribution of that material requires more care, and involves asking permission or providing notification of intent." Taking care to evaluate the ethics, as well as, legality of collecting privately-created Web content that is publicly available will be an important community commitment for the Library and its NDSA partners.

Local and regional content and perspectives

With the rise of social media, there is greater and greater access to local and regional content and perspectives. This means communities that have been marginalized and underrepresented by mainstream media are now part of the national conversation. We can now track microtrends in political opinion and action, determine their geographic distribution, demographic

make-up, time frames, and so forth. In the geospatial realm, it means that policy makers and regulators have access to crucial local and regional information sources about land use, natural disasters, watersheds and how they are changing, and the effect of climate change on local species, among other things.

Such small-scale analysis and comparison allows for better policy and compliance assessments, as well as more targeted and more effective legislation. It also allows analysts to see which developments are truly local, and which are universal. As one expert remarked, "National elections, and state and other local elections, tend to be a very big area of focus when you are talking about Web archiving. It's interesting to see that it is one area that crosses geographic boundaries, because the web has become a strong influence on how people choose to vote."

Digital Content for Public Policy: Challenges

A handful of challenges, common among all sources, were identified by participants in convening sessions and fact-finding surveys. These common problems indicate areas of special attention and opportunity for the collecting priorities of the Library and NDSA.

Identification and Selection

The sheer volume of content from which to choose materials that are of potential long-term value means many tried-and-true selection procedures used with print materials are no longer effective. One television project manager noted that "It used to be when people were shooting on film, the ratio of non-published to published material was about 5-to-1. Now, however, with video someone told me they were shooting 100-to-1."

Besides the challenges of selecting from among voluminous materials, there are also challenges posed by content that is essentially dynamicly limitless. Which versions of a news site, which is updated every 15 minutes, should be preserved? How frequently should geo-databases be sampled for permanent retention? Which parts of expert legal blogs should be preserved—the text alone? The text with some or all content? The full links embedded in the blog as well?

Digital content is usually quite fluid, easy to update, modify, and customize for delivery to targeted audiences. However policy makers and regulators need data with full transparency and accountability in all stages of decision-making, from research to analysis and presentation. Permanence and stable linking is essential for reliability in decision making. An expert commented, "Libraries look for what they consider authoritative sources, and the nature of authority has shifted drastically from the authority of the mediated published document to the authority of the crowd, of the community. It comes together from the many, rather than from the one. This is a real dilemma for institutions that come from a traditional library mindset."

Collecting

A major obstacle encountered in collecting is the way that content is locked into silos—tied to a proprietary software, for example, or kept behind gates. Ownership of content can be opaque. It is inefficient to hand-select sources and to seek permission to look at content and decide about its long-term value in each and every instance. This is an area where the creators and distributors can really help by making their content more available to authorized collecting bodies, such as the Library and its trusted collecting partners.

According to U. S. copyright law there is almost no way to say that even a single website is

entirely in the public domain, even if the organization is a government organization. They sometimes integrate materials and imagery that are privately owned… Also, there are barriers that public institutions can put up to prevent capture of data, e.g., the White House can put rules on their servers to prevent collection of the data, even though it is public domain data.

Most vexing of all are the barriers that government sites routinely throw up to bar their content from being collected. National security considerations aside, there are large numbers of government sources that put `robots.txt` exclusions on their sites that prevent conscientious preservation organizations from collecting valuable content. A more open Government 2.0 approach—the mandatory use of open, well-documented formats, where possible—would lower barriers to efficient preservation and access.

Drivers to preserve

Creators and producers play critical roles in making information that is valuable, reliable, well-documented, easy to find, and easy to preserve. A national digital collection cannot be created and maintained without their participation. Format and bibliographic standards, such as acid-free paper and ISBN numbers for print publications, were cooperatively developed by collecting institutions and publishers. Digital formats are lacking such widely recognized standards and best practices so collectors must again engage creators, producers, and distributors to create awareness of the benefits of preservation and their roles.

For government, preservation benefits are usually mission-related; for commercial entities, they are business-driven; for Web 2.0 and community-created content there are personal and local heritage benefits; and in academia, the motivating preservation factors are the benefits to research and learning. In all of these cases, but especially in community-created content, the intellectual property, format, and distribution standards and best practices are quickly evolving, making systematic selection, collection and preservation highly challenging while the content is becoming more and more valuable. It will be important for the successful implementation of the National Digital Collection Framework to provide incentives for people and organizations that produce content of value for public policy to take the preliminary step of saving it, and making it available for long-term preservation.

LOOKING AHEAD

Strategy for Collection 2012–2020

The work of building a national digital collection of use to Congress and the American people has begun with focusing on public policy content. Over the next ten years, the Library and its partners have identified other areas to target for collection.

2012 –2014
Coordinated collection of digital materials related to Science, Mathematics and Technology, and the Social Sciences is already underway at many partner institutions. Due to the increased focus on the data collected during publicly-funded research and the various open-access initiatives in academic research, the systematic selection, collection, and preservation of these burgeoning materials will be a key outcome of the national digital collection.

2015–2017
Today's artifacts of digital culture and communication will be key resources in American His-

tory, and World History and Cultures collections. Manuscript collections will begin to be filled with electronic drafts and documentation of creation. International collections will be brought together in a scaled, global digital collection. The collection Framework will outline areas of cooperation, and leverage local strengths, to provide future users with a comprehensive and accessible view of the U.S. and the world's history and culture that is recorded digitally.

2018–2020
The areas of Arts and Culture, Religion and Philosophy, and Sports, Recreation and Travel are often the subjects that have the most wide appeal and use. However, the materials in these subject areas also have the most irretraceable intellectual property rights. Without changes in copyright laws that would allow the preservation of these materials by trusted institutions, the Library and its partners cannot systematically preserve them. Therefore, these materials will be targeted for preservation at the end of this decade.

Coordination and Collaboration

Given the dynamism of the digital content environment, the need for a flexible and responsive preservation infrastructure to be in place and ready to act is paramount. Due to the range of content and sources, a coordinated, collaborative approach is required. The crucial role of the convener and coordinator, played by the Library, has been effective in bringing together committed parties to seize opportunities and secure time-sensitive materials. There is much work to be done to move from these initial steps to a sustained effort.

Collecting anything close to a universal collection of knowledge cannot be done by a single institution. To succeed, stewardship organizations need to collaborate around sustainable infrastructure and shared tools, standards, and coordinated collecting initiatives. A coordinating group is essential to ensure that users have full coverage of important content without creating expensive and unnecessary duplication. The Library of Congress and its partners need to collaborate. Most crucially, they need to partner with content creators to ensure that valuable information created today can and will be preserved. Although the collecting approaches must be new, the key benchmarks of success remain the same; that the citizens of tomorrow have ready access to high-quality, reliable, and authentic information about the actions taken today by their governments and fellow citizens.

Appendix H

Strategic Objectives 2000–2020

Objective: Catalyze Collaboration for Digital Stewardship

Area of Action	A Decade of Action 2000–2009	Work in Progress 2010–2013	The Decade Ahead 2013–2020
Strategic Partnerships	• Brought together over 185 stewardship partners in more than 44 states and 25 nations to preserve at-risk digital content. • Defined roles and responsibilities within the network to leverage local strengths and priorities for sustainable partnerships. • Developed and promoted digital preservation best practices and standards throughout the partner network.	• Chartering the National Digital Stewardship Alliance with organizations from across the government, academic, non-profit and commercial sectors. • Establishing working groups to advance a national digital preservation agenda around content, standards and practices, infrastructure, innovation, and outreach.	• Establish and sustain digital preservation programs in all 50 states at the university, state library and archives, and community levels. • Enable the exchange of information and content between private or commercial entities and public stewardship organizations through well-defined relationships and mechanisms. • Establish a network of federal agencies and private-sector funders in support for ongoing preservation action.

Objective: Respond to Information Challenges through Innovation and Action

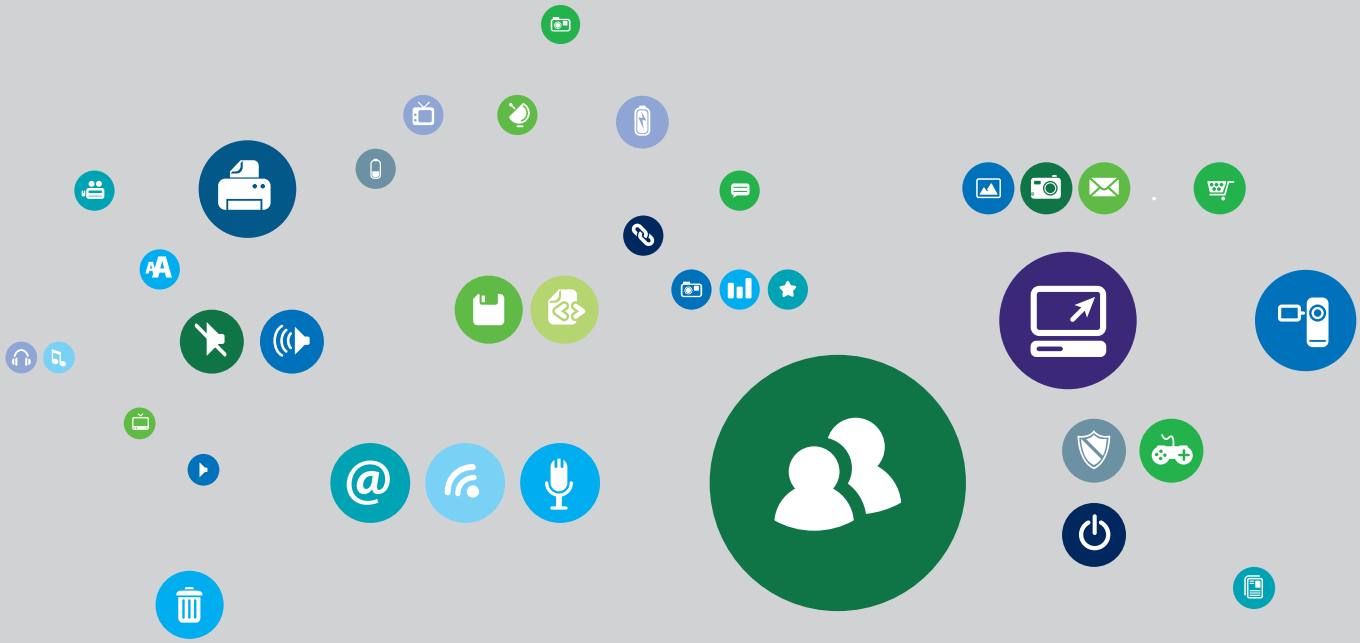
Area of Action	A Decade of Action 2000–2009	Work in Progress 2010–2013	The Decade Ahead 2013–2020
Content Stewardship and Practice	• Selected and preserved at-risk digital content in over 1400 digital collections primarily open for public or researcher access. • Fostered committed organizations to collect, preserve, and provide access to digital content on a programmatic basis, especially in the area of public policy. • Characterized the nature of digital content domains, and the challenges and opportunities for the future of cultural heritage collections. • Led digital standards and best practices for formats, workflows and protocols.	• Establishing a shared National Digital Collection Framework to enable the cooperative and coordinated collection of content with value for present and future generations. • Convening panels of experts to advise the alliance about existing and emerging digital forms of content and their characteristics that impact preservation and access. • Initiating a phased strategy to collect digital materials with a strong focus on government, politics and law, maps and geography, and news, media and journalism. • Launching a web portal and associated access tools to provide integrated and enhanced access to preserved collections.	• Assemble a comprehensive national digital collection across a breadth and depth of subject areas—even those with challenging rights issues—including world history and cultures, American history, arts and culture, religion and philosophy, and sports, recreation, and travel. • Preserve the national digital collection through multiple trusted organizations. • Make appropriate access available to collections and metadata on desktops around the world. • Enable users to mine data and combine collections, and answer new research questions through new visualization and mapping tools.

Objective: Increase National Capacity for Stewardship of Digital Content

Area of Action	A Decade of Action 2000–2009	Work in Progress 2010–2013	The Decade Ahead 2013–2020
Distributed Infrastructure and Shared Tools and Services	• Modeled a layered and distributed architecture for digital preservation. • Fostered the testing and release of tools and services to support distributed preservation. • Built a storage and transfer infrastructure for over 200TB of preserved content at the Library. • Developed more than 25 shared technical tools and services available for public download. • Fostered the development of mechanisms, standards, and tools for cost-effective storage and secure management of collections distributed across the partnerships. • Engaged the storage industry to establish avenues for collaboration with the digital preservation community.	• Supporting and extending the development of open source tools and services by the entire digital preservation community. • Investigating computer forensics and its application for collecting institutions. • Establishing large-scale and distributed storage architectures to deal with large-scale archives. • Enabling smaller organizations to preserve digital materials through distributed partnerships. • Promoting the development of cloud services that can support the growing needs for storage and computing. • Encouraging communities with highly-specialized needs (e.g., geospatial, data sets, observational data) to develop specialized storage networks or access services that can serve the entire community.	• Provide management, preservation, and access through certified digital repositories services across the Alliance. • Closely collaborate with storage and technology industry partners on infrastructure requirements for the digital preservation community. • Work through a diverse and robust open source community for the development of new content management and access tools as new technologies emerge. • Support ongoing innovation to adapt to changing technologies and media.

Objective: Increase National Capacity for Stewardship of Digital Content (cont.)

Area of Action	A Decade of Action 2000–2009	Work in Progress 2010–2013	The Decade Ahead 2013–2020
Policy Study	• Supported a study and made recommendations to the Register of Copyrights to revise the U.S. copyright law to enable and encourage the preservation of digital content. • Issued an international report with three other countries on the impact of copyright law on digital preservation. • Co-sponsored and published a study by the NSF Blue Ribbon Task Force on Economic Sustainability of Digital Preservation and Access.	• Sponsoring the study of privacy issues in the digital age and their relationship to the preservation of our cultural heritage. • Chartering a working group to make recommendations on creating incentives to archive private assets for the public good. • Continuing to identify policy issues that impact access to preserved content.	• Support preservation action under an updated U.S. copyright law. • Encourage incentives for the preservation of digital materials. • Support policies for archives to preserve private assets for the public good.
Outreach	• Launched a monthly newsletter about digital preservation that in the first year of publication tripled its audience reach from 8,000 to 24,000. • Launched a digital preservation video series. • Initiated a K-12 Web Archiving Program with 15 primary and secondary school teachers and students. • Promoted student awareness of the risk of loss of digital cultural heritage content through a workshop and video.	• Promoting general public awareness about digital preservation through a Personal Archiving section on digitalpreservation.gov and an open public event with the Library Preservation Directorate called Personal Archiving Day. • Supporting robust communication and outreach regarding all aspects of digital preservation and access. • Providing a central clearinghouse of information useful for all stakeholders. • Engaging in collaborative partnerships to deliver educational opportunities for the broader preservation community. • Launching a fellowship in advanced digital technologies to develop new relationships with top university graduate programs.	• Create a broad and deep awareness about digital preservation issues and solutions in the digital preservation community and the public. • Enable cultural heritage institutions to operate in an environment that values preservation of digital content and enables the beneficial use of America's digital heritage. • Provide digital preservation education opportunities in every state and online. • Establish certification for digital preservation practitioners.



www.digitalpreservation.gov