World Library and Information Congress: 70th IFLA General Conference and Council

22-27 August 2004
Buenos Aires, Argentina

Programme: http://www.ifla.org/IV/ifla70/prog04.htm

The U.S. National Preservation Strategy: Challenges and Opportunities
(La Estrategia Nacional de Preservación de EE.UU.: Desafíos y Oportunidades)

Laura E. Campbell

Associate Library for Strategic Initiatives / Bibliotecaria Asociada para Iniciativas Estratégicas
Library of Congress / Biblioteca del Congreso
Washington, DC USA

Abstract

As the volume of digital material escalates, creative expression in science, technology, arts, and humanities and social sciences is increasingly embodied in this fragile, ephemeral, and dynamic medium. Its loss will have significant and irreversible effects upon the cultural literacy of future generations. As a result, the U.S. Congress has charged the Library of Congress to lead a national effort to forge an infrastructure to identify, acquire, manage, and preserve important works in digital form through the National Information Infrastructure and Preservation Program (NDIIPP), legislation passed in December 2000 (PL-106-554). This paper describes our work to date and offers ideas about ways that government can foster partnerships to achieve this important national goal.

Sumario
Simultáneamente con el aumento en el volumen de materiales digitales, también aumenta la expresión creativa en ciencia, tecnología, artes, humanidades y ciencias sociales representada en este frágil, efímero y dinámico medio. Su pérdida tendría efectos significativos e irreversibles en la educación cultural de generaciones futuras. Como resultado, el Congreso de los EE.UU. ha encargado a La Biblioteca del Congreso el liderazgo de un proyecto nacional para generar una infraestructura para identificar, adquirir, dirigir, y preservar trabajos relevantes en formato digital a través del Programa Nacional para Infraestructura de Información y Preservación

---

As the volume of digital material escalates, creative expression in science, technology, arts, and humanities and social sciences is increasingly embodied in this fragile, ephemeral, and dynamic medium. Its loss will have significant and irreversible effects upon the cultural literacy of future generations. Therefore, the U.S. Congress has charged the Library of Congress to lead a national effort to forge an infrastructure to identify, acquire, manage, and preserve important works in digital form through the National Information Infrastructure and Preservation Program (NDIIPP). The legislation, passed in December 2000 (PL-106-554), allocates approximately $100 million for the program, to be released in stages: $5 million to be immediately authorized to support planning activities, $20 million to be made available after Congressional approval of a plan for the effort, and the final $75 million to be contingent upon raising $75 million in matching funds. The plan, *Preserving Our Digital Heritage*, was submitted to Congress in the fall of 2002. And in December of that year, the plan was formally accepted, enabling us to move to the next phase of work based on about $20 million in public funds and an estimated $15 million to be raised from private, non-federal sources.

Since that time, we have pursued a multidimensional agenda, engaging partners and stakeholder communities around the world in different activities. We held a competition to support building a network of partnerships to identify and capture at-risk digital content. Our work on refining a proposed technical architecture continues with the goal of identifying protocols necessary to the preservation of data during ingest and transfer. The Archive Ingest and Handling project, as it has become known, has four partners who are testing seven preservation technologies. The test data set is a 12 gigabit archive of materials collected from the public by George Mason University during a period following September 11, 2001, which GMU donated to the Library and agreed to have used a test archive. In cooperation with NSF and other federal agencies, the Library also plans to fund a program of basic research; a call for proposals is expected later this year.

Finally, we are cooperating with other national libraries to support an international consortium to preserve Web content. Our partners include: the Bibliothèque nationale de France, the British Library, and the national libraries of Australia, Canada, Denmark, Finland, Iceland, Italy, Norway, and Sweden as well as the Internet Archive. Key objectives include:

- collaborative working, within each country's legislative framework, to identify, develop and facilitate implementation of solutions for selecting, collecting, preserving and providing access to Internet content;

- facilitating international coverage of internet content archive collections within national legal frameworks and in accordance with individual national collection development policies;

- international advocacy for initiatives that encourage the collection, preservation and access to internet content.

To achieve these objectives, the Consortium will:

- provide a forum for sharing knowledge about Internet content archiving both within the Consortium and beyond;

- develop and recommend standards;

- develop interoperable tools and techniques to acquire, archive and provide access to web sites;

- raise awareness of Internet preservation issues and initiatives through conferences, workshops, training events, publications, etc.

Our partnership agreements with major research university libraries and their partners will eventually engage some 25 organizations among the U.S. higher education, state libraries, not-for-profit and for-profit sectors. We expect to learn a great deal about how to structure and administer partnerships while we deepen our understanding about collecting digital information. The U.S. National Archives is also partnering with one of the projects, but because of the legislative constraints on our funding and the matching requirements, the Archives is a "silent" – but active – partner. Other U.S. state and federal agencies are also involved. We are extremely pleased with the geographic scope of the project as well as with the diversity of formats represented: geospatial (in several formats), digital television, text, Web sites, e-mail and databases of social science information. Among the topics that we hope to address are privacy and confidentiality of data, intellectual property, migration of legacy systems, maintenance of open source software tools, and a variety of technical issues related to identification and collection of at-risk information.

At a general level, the technical issues are fairly well understood. Storage media and encoded signals degrade; hardware and software obsolescence means that playback is a problem; metadata schema seem to proliferate and are not necessarily consistent from one format to the next, given the different requirements of, say, cinema versus text. The simple act of copying data from one storage medium to another (known as "refreshing the data") potentially introduces errors, and although some of those errors may be immaterial to the content of the work, others may not, and both categories of error raise questions about notions of integrity and authenticity of the works that are stored. Digital preservationists have debated the relative merits of migration, emulation, and encapsulation, and for now, there seems to be consensus around the notion that different data will require different preservation strategies, depending on frequency of use and the importance of preserving "look and feel".[1]   Intellectual

---

[1] This point is made with some care in the excellent review article by the team at the National Institute of Standards (NIST); Kyong-Ho Lee, Oliver Slattery, Richard Lu, Xiao Tang, and Victor McCrary, "The State of the Art and Practice in Digital Preservation," *Journal of Research of the National Institute of Standards and Technology* 107 (1) (January-February 2002): 93-106.

property rights are complex and their management has historically varied among formats and media so broadcast radio and scholarly publishing, for example, both function under the same overall rubric of copyright law but industry structure and practice vary.

Finally, despite enormous technical achievements in processor speed and storage capacity, preservation is simply more than the ability to cram data on high density discs and then play them back. The challenge we all acknowledge is to make that stored information accessible – not just accessible in the literal sense that we can display it, hard as that is, but also accessible in the sense that future users will be able to find what they need and to use it in a way that respects the integrity of the information and the legitimate interests of stakeholders. And that means trying to anticipate the needs of the 22nd century user. Or at least doing no harm in the present while allowing technological progress to take care of the needs of the user one hundred years off.

More recently, investigators have started to ask questions at a higher level: How do the components fit together? How does the system meet the needs of the institution and its present and future users? That is, what does a technical architecture of preservation look like?

We convened a small group of technical experts in February 2001 who outlined a set of design principles and sketched an architecture that we have continued to vet as we have discussed it with various groups. We refined the original proposal as a result of another round of meetings in 2002, but the principles remain the same. Collectively they support the values of transparency, collaboration, incremental development, stability, flexibility, heterogeneity, and innovation and suggest a flexible approach that allows multiple parties to work on different pieces while ensuring overall coherence.

The technical papers are posted to our web site, www.digitalpreservation.gov, and I hope those of you who are building and managing digital archives will spend some time with the material we are developing. But I would like to take a moment and cull out a few points made by our consultant Clay Shirky in *Update to the NDIIPP Architecture, Version 0.2* that speak directly to the relationship between organization and technology:

1   We recognize that great energy is going into work digital preservation among a many of institutions. As a result, the architecture recognizes the need for interfaces among institutions to exchange individual objects and whole collections and for institutions to be able to perform different roles simultaneously.

2   There are any number of systems for digital preservation ready to test and a strong desire for federating systems or otherwise creating ways for those systems to interoperate – hence our archiving ingest and testing project.

3   Quick convergence on a single system or set of interoperable systems is unlikely. Every system rightly is designed to meet the goals of its home or sponsoring institution and institutional goals necessarily differ. Digital preservation will be the healthier for this approach because heterogeneity guards against system wide

failure. But interoperability of based on the same tools and formats or even interoperability based on the same conceptual model is unattainable. Simple interoperability based on the nation, "everyone uses the same tools and formats" and the deeper interoperability of "everyone uses the same conceptual model" are both unattainable, now and for the foreseeable future.

4   As a consequence, the NDIIPP architecture is set up to support institutions as they cooperate with one another but who may have differing technological systems in place.

This is hard. But we believe that it is possible and necessary. The documents we and others put forward will facilitate conversations among NDIIPP and its partners and within the wider world of digital preservation projects that might not be directly or formally linked to our initiative. We also have no doubt that the real-world experience we and others garner will serve to enrich and refine the work.

Over the decades, libraries and archives have evolved practices, standards and cooperative relationships that amount to an intangible infrastructure to support collecting and protecting the intellectual output of our patrons. In concert with such organizations as the International Federation of Library Associations (IFLA), the Library of Congress and our counterparts at home and abroad have expanded our respective national missions so that we employ internationally consistent standards and collect the record of human creativity globally. Our grandchildren's children will thank us.