Digital preservation initiatives in the United States:
a summary

DEANNA B. MARCUM

A B S T R A C T

General presentation on how digital preservation issues are being faced in the United States of America. Special reference to the National Digital Information Infrastructure and Preservation Program, that is found at the Library of Congress, and aims to implement a national strategy for the long-term preservation of digital content.

R E S U M O

Breve apresentação da forma como os EUA têm equacionado os problemas da preservação digital. Especial referência ao Programa National Digital Information Infrastructure and Preservation, liderado pela Biblioteca do Congresso, o qual visa a implementação de uma estratégia nacional para a preservação a longo prazo de conteúdos digitais.

In the United States, preservation is traditionally a distributed activity. Each library or archives assumes responsibility for its own holdings and for providing services to its users. The national collection is not simply the holdings of the Library of Congress; rather, it is the sum of all holdings of the academic and public libraries in the country. This tradition has raised some interesting challenges as we begin to consider the preservation of digital information.

The activities related to digital preservation now underway in the United States fall into four categories:
• Developing common understandings among stakeholders;
• Practical preservation projects;
• Experimental preservation activity;
• Preservation research.

The groups that are believed to have a strong interest in (or stake) in preservation of digital materials include the creators or authors, the publishers and the archival repositories. All have different motivations, but each has an interest in preserving digital information for long-term access and use.

What all of these groups have in common is a need to select a single or set of standard formats. All have an interest in agreeing on terms or conditions for the distribution of archived information. For materials to be accessible over time, all must agree on what constitutes a trusted repository. For many of the stakeholders, there is increasing interest in the Open Archival Information System as the standard reference model.

Practical preservation projects now underway provide a great deal of guidance as to what works and what does not. The National Archives and Records Administration has mounted a massive electronic records project that will help all organizations understand what is involved in maintaining these records. Commercial publishers are beginning electronic journal archiving projects to satisfy customer demand. A number of non-profit organizations are reformatting print-based collections and building archival repositories for the digital surrogates. Finally, service organizations are building collaborative digital archives to serve the library community.

Important experimental activity is also underway. The Andrew W. Mellon Foundation funded seven academic libraries to work with publishers or creators to create digital repositories for electronic journals. Standford University

has developed software known as LOCKSS (Lots of Copies Keeps Stuff Safe) that allows institutions to manage the caches from Web searches for electronic journal articles. And the IBM Research Center at Almaden is experimenting with a universal virtual machine for digital preservation.

Preservation research is going on in a number of institutions, all aimed at answering one or more of these questions:
• What are the attributes of preservable digital information?
• What are the attributes of trusted archival repositories?
• What are the appropriate standards related to digital archives?
• What are the best methods of automatic copying and distributing digital information?

The most comprehensive program related to digital preservation in the United States is found at the Library of Congress. In December 2000, Congress passed legislation establishing the National Digital Information Infrastructure and Preservation Program (NDIIPP) in the Library of Congress (LC) (PL 106-554). The legislation calls for LC to lead a national planning effort for the long-term preservation of digital content and to work collaboratively with representatives of other federal, research, library, and business organizations. The legislation allocates $100 million for the program, to be released in stages: $5 million to be immediately authorized, $20 million to be made available after Congressional approval of an NDIIPP plan, and the final $75 million to be contingent upon raising $75 million in matching funds.

It is a challenging mandate, requiring LC to mediate among many technical, organizational and management concerns: What does it mean to have a national strategy for digital collections, for example? What are the roles and responsibilities? Does LC partner with all libraries? All research libraries? Other cultural institutions? Other federal agencies with information management missions – of which there are many? What are we preserving for whom and for how long?

The challenge of digitizing rare and unique materials is immense and important. The National Digital Library Program LC accumulated substantial experience in the operational details of building digital collections that represent both robust systems for today and archives for tomorrow. But conversion of works largely in the public domain pales in comparison with the challenge of dealing

with *born digital* – those items that have been created in digital form, which can range from Web sites to databases of scientific data, to streaming content from one of the record labels. These are rife with hardware and software issues as well as intellectual property concerns and there may be no analog version in reserve should the archiving system fail.

Some people believe that the only collection policy for digital preservation is ***everything***, and they bewail the thought that Web pages are routinely removed or somehow lost. Others, librarians and archivists among them, who are familiar with collection development and management, acknowledge that selection is necessary. Even if it is theoretically possible to save *everything*, it may be very expensive to make that material valuable to potential users. Collection development is but one example of the possible choices and trade-offs to be considered. Planning is about evaluating alternatives and finding the middle ground among extremes.

In early 2001, LC began its planning process under the leadership of Laura E. Campbell, the Associate Librarian for Strategic Initiatives. Its overriding goals are to encourage shared responsibility and to seek national solutions for:
• The continued collection, selection, and organization of the most historically significant cultural materials, regardless of evolving digital formats;
• The long-term storage, preservation, and survivability of those digital materials; and
• Ensuring access to the electronic historical record of the American people under conditions that respect the legitimate interests of rights holders.

A broad-based advisory board, consisting of representatives from other federal agencies, research libraries, private foundations, and industry, was organized. Based on a series of formal and informal discussions, LC made an initial set of scoping decisions.

The first phase of work would be focused on *born digital* information and would be organized around formats in which LC's collections are strong or where the digital materials are aligned with LC's traditional mission: Web sites, electronic journals, electronic books, digitally recorded sound, digital television, and digital moving images (e. g., *film*).

The planning process would involve four dimensions. These would function roughly in parallel but would also allow for feedback and cross-fertilization:

>

stakeholder meetings; collaborative research; conceptual framework; and scenario planning.

An important milestone in the planning process will be submission of a formal plan to Congress later this year, which will include recommendations for future investments. As of this writing in December 2002, the stakeholder meetings have been held; work has begun on an approach to developing a collaborative research agenda; a conceptual framework has been drafted, and three scenario workshops have been held, and a plan has been submitted to Congress for approval.

The next steps will include experimental projects with a variety of partner institutions, and the results will be posted on the Library of Congress' Web site.