

**National Digital Information Infrastructure and Preservation Program
(NDIIPP)
Convening Sessions, November 5-6, 7-8, 15-16, 2001
*Summary Report***

140 individuals representing a broad range of stakeholder communities, primarily among content creators, distributors and users, were invited to participate in one of three one- and one-half day sessions in Washington, D.C. Given travel restrictions and other inhibitions arising from the current political climate, about 70 individuals attended one of the three sessions for all or part of the program. Attendees represented media and entertainment (film, television, music); scholarly, textbook, commercial and newspaper publishing; research libraries; heritage preservation organizations; universities; private foundations and independent authors and artists. Participants also included representatives of other concerned federal agencies: National Library of Medicine, National Agricultural Library, National Archives, Smithsonian, National Science Foundation, Department of Commerce, U.S. Copyright Office, and other units of the Library of Congress. Many attendees had experience across multiple functions and industries, enriching the conversations through their diverse personal experiences and informed perspectives.

The organization of each of the three sessions was identical. Attendees were invited to dinner, hosted by the Librarian, in the Jefferson Building at the Library of Congress on the first day, where various digital projects and rare materials were showcased. This provided an opportunity for participants to get to know one another as well as familiarize themselves with some of the Library's activities. Members of the Library's staff were also invited to these dinners. The second day began at 8:00 am and concluded by 4:00 pm. Following introductions and an update on NDIIPP (for those who had been unable to attend dinner the night before), participants were invited to volunteer topics they wished to see addressed in the course of the day, thus building the agenda from the ground up. Discussion followed, moderated by Peter Schwartz of the Global Business Network, who was assisted by Richard P. O'Neill. The sessions were videotaped and subsequently transcribed.

In the last hour, attendees were encouraged to offer visions of the future, guided by two questions: "What in the year 2101 will you wish people in the year 2001 had done with regard to archiving of digital information?" And, "What are you glad that they did?" The session ended with participants offering specific suggestions concerning priority issues and next steps. These are summarized in the last section of this report. The remainder of this document is focused on reporting topics and themes that reoccurred across all three days' discussions.

Participants voiced broad consensus behind the initiative, willingness to seek collaborative solutions and a sense of urgency. There was also consensus behind some form of distributed or decentralized solution, involving multiple entities and possibly affording varying degrees of access according to groups of users. Moreover, there was widespread agreement that decentralization required a degree of centralization. Said one computer science professor, "It's not a yes-or-

no answer to distribute it or centralize it.” Decisions associated with centralization, sharing, community consensus, or codes of conduct might concern technical standards, archival management standards and metadata.

Frustrations were shared over the multiplicity of formats, rapid technological changes, hardware and software obsolescence, both too many and too few technical standards, the formal and informal standards setting processes and problems associated with playback (which are particularly acute for film, television and sound as well as early computer files). Representatives of film, television and music industries agreed that there are substantial technical differences among the formats, despite commonalities, and there is a question as to whether it is better to subdivide and focus on the issues that arise within formats or to start by trying to solve the commonalities. Moreover, it was clear that the new technologies are resulting in shifts in institutional roles and functions that are not clearly understood. From the perspective of library community, one senior university official noted, the challenge is going from a model based on loaning information that the institution owns, to a model based on providing access to information that the institution does not own wherein “access” can be calibrated.

Archiving: Whose responsibility?

Information in digital form, whether “born digital” or “reborn digital,” as many composite, multimedia works in film, music and television are, pose a challenge for libraries and archives. As one major figure in archival preservation observed at the first session, “Until now, libraries have been society's place for preserving information, and it seems to me the reason we're here today is that the kinds of things that are now becoming part of the documentary record are not the kinds of things that typically go into libraries.” Moreover, libraries have historically served a civil society function. The challenge, she continued, “is to figure out how will we serve society in the future by thinking carefully about the documentary record, whatever form that takes, and a shared responsibility for its preservation.” Commercial interests are accustomed to coping with technological changes, another participant observed, “that's the nature of their business.” The question for the library community is, “How do all those changes [affect current roles and functions] and how would you see your roles in a different system, if you built the system [from] a bottom-up standpoint where you don't know, because the role is going to evolve under your feet?”

The new information technologies fundamentally alter many existing relationships and expose ambiguities that may have formerly existed but that had been overcome in practice. According to one industry executive, “I think it's very important to distinguish between publisher and creator and recognize that publishers are one way that creators get their works to the public but not the only way.” A newspaper publisher urged the group not to think of the process as a linear one of creation, distribution and use wherein publishers were distributors. Rather, publishers contribute to the creative process in editorial judgment and selection: “If you don't keep the concept of editing and packaging it in the center of this, you do run the chance of running amok of asking what in

the world is it we're trying to save and why are we trying to save it and where does the value come from.”

Indeed, even definition of what constitutes “publication” is currently ambiguous as are notions of what is appropriate for archivists and librarians to collect, including material in corporate archives, material that the creators themselves may not wish to have made public, and information that is believed to have been deleted but, because of the nature of the medium, may, in fact, have been silently retained. (One example given concerned contracts and material potentially extant in corporate archives. In the case of a contract, clauses may be struck by mutual agreement of the parties, yet are silently retained by the software and therefore unintentionally accessible to future readers.) No longer can archiving be seen as a “just in time” capture or an activity that takes place after distribution. Rather, archiving must occur at the time of creation (e.g., specifying metadata), potentially putting an additional burden in the creators. But said another participant, “how do you incentivize? Usually a young, struggling artist doesn't have the time and doesn't have the resources to preserve how he created his work. He might actually keep a copy of the manuscript of the music he wrote, but you know, he's a struggling artist. He's starving. He doesn't have the space or the room or the time or the money.”

Finally, representatives from industries as diverse as scholarly publishing and the commercial music labels agreed that there was a *business* case to be made for collaborative long term archiving of digital content and that the challenge was to work out an architecture and set of policies that balanced legitimate economic and cultural interests. Said a representative from a media conglomerate, “We don't preserve them [works] because we wanted to be archivists. We preserve them because we think there's some follow-on value to turning out those films again, and as they start to turn vinegar, there are decisions made.” Owners of lucrative properties seemed willing to propose compromises, e.g., relaxing some restrictions after an agreed-upon period of time, based on the economic value of the property; releasing segments but restricting access to the entirety; releasing partial versions for limited use, rather than the commercial version. Notable was the “can-do” attitude of many, mainly but not solely in the entertainment industries, who are working the problems internally and through industry committees and appear to welcome the opportunity to exchange ideas and to come up with solutions that align business and cultural needs: One desirable outcome, said one representative of a music label, would be to take a few of the ideas offered and formulate them as “what if” models that basically I can take to my senior management and put something on the table for discussion with them. That has not happened because I've never been given anything to even set up a meeting.”

Representatives of traditional libraries and archives and representatives from the music and print publishing industries found that they shared a fundamental conceptual distinction: the master copy (also called the “crown jewels” or a “content bank”) and the derivative work, which might also be represented in different formats depending on how the company expected to distribute the work to a given set of users/listeners. This led to broad agreement on the

relevance of distinguishing between the archive and the services which might be built thereon. This strategy requires “active management” of the underlying archive as well as attention to issues of authorization and authentication and ways that access might be afforded while protecting the “asset.” Some emphasized the importance of managing the archive (format and software obsolescence, storage media deterioration, signal degradation, playback) while others emphasized issues related to intellectual property rights management.

Protecting intellectual rights and other sources of tension

Managing intellectual property rights was one of several issues over which there was substantial disagreement. Other sources of tension arose over collection policies, roles of participating institutions, standards processes and economic support. Discussions in these topics ranged from fairly amicable to somewhat intense, particularly with regard to the content of the Digital Millennium Copyright Act and its implications for libraries and archives about which there is substantial ambiguity, partial information and a legacy of mistrust among libraries and publishers. Moreover, there was substantial frustration with identifying rights holders for individual works.

Differences in corporate cultures were apparent in discussions of legal and economic issues. Very consistently, librarians expressed concern over actions, many of them core to maintaining their collections (e.g., making a copy for purposes of preservation), that they believed might be illegal – or even criminal – under existing law. Maintained one university librarian, “It seems like we need to have some resolution of some of these issues so we can get ahead with some of these things.” On the other hand, those with commercial interests saw the issues in terms of protecting the economic value of an asset. Said one industry representative, who acknowledged that some of the contradictions that librarians saw in the DMCA had not previously occurred to her, “it’s [provisions concerning copying] intended to be -- allow for the preservation, but the whole prevention item is one of not allowing somebody to illegally make money off of somebody else’s intellectual property.”

Coping with new technologies and new products and services is compelling changes among libraries and archives as well as within the industries represented among the participants. One senior university official noted that libraries and archives were not accustomed to negotiating “unbundled” rights but that unbundling was likely to be an effective strategy in the coming environment. In rather marked contrast to the concerns about intellectual property rights among many in the library and publishing communities, one of the technical representatives from the entertainment industries suggested that the approach should be to design the system first and then let the legal specialists work out the implications. Moreover, with respect to encryption and other measures employed to protect digital works, one representative of the entertainment industry commented, “It doesn’t have to be Fort Knox.” A representative from a major non-profit multimedia publisher echoed the sentiment. When asked if someone could hack the encryption protecting their new product, he answered, yes, but “you’d really have to know a lot, and by the

time you do that, you might as well take a printed version of the [name of the print version] and put it on a high-resolution copier and go, whoa, this is really cool.”

The challenge, a participant pointed out, was not protecting rights from legitimate users but rather to guard against users who took advantage of flexibility of the medium to behave improperly or illegally. One proposed solution to the question of protecting rights to digital material was to allow use of archival material on site, thus affording a measure of control through location. However, even if the issues associated with the DMCA were fully resolved there would remain other legal issues, such liabilities that potentially arise in commercial archives (for example, the unreleased documentation and reporting associated with a newspaper) that might be sought as part of a collection policy but that would pose a significant risk for the corporate owner and rights associated with individuals’ privacy (e.g., use of Humphrey Bogart’s image) and confidentiality (e.g., medical records).

Topics in which the discussions diverged broadly but less heatedly include mission, roles, standards setting and collection policies. Some participants advocated models based on the notion of a universal library and linked the Library of Congress to a perhaps idealized notion of the role of public libraries. Yet academic and research librarians quickly noted that university libraries recognize multiple clienteles and do not necessarily serve all of them equally. Representatives from the National Archives noted that the Archives saves about two percent of the material submitted to it and that there can be substantial disagreement about which two percent. Historically, the focus there and in the Library of Congress and elsewhere has been on famous figures. Yet, it was pointed out, much can be learned about the workings of government by looking at transactional records, and second-guessing the future’s interests in historical matters is challenging as many scholars around the table pointed out in the second and third sessions. One historian described the challenges of doing histories of federal agencies and program, where preservation of relevant material is haphazard, and another scholar cited various World War II era records that subsequently proved to have unanticipated value for documenting looting and hence reparations.

A number of voices noted that technical standards both enable development but potentially impede it, particularly if the standard is either premature or overly restrictive. Still, another participant commented, in effect, the standards process is and has been messy and imperfect and it is important to “just get on with it.” One participant encouraged planners to think in terms of the Internet standards process, which has enabled innovation and is organized around notions of rough consensus, running code and two independent implementations. The role of the Library of Congress might then be to oversee the standards setting process or to function in some sense as the honest broker.

In addition to functions that the Library might play relative to corporate interests, the question of roles, including relationships with other federal libraries, university and research collections and public libraries, was linked to

collections policies, which were, in turn, linked to different profiles of possible users. Thus, the notion of the collector of “last resort” implied different responsibilities from a vision of the collection that might be built around copyright deposition. Still another proposal called for allocating funds to enable selected libraries with particularly strong collections in a given area to digitize those collections as part of a larger system of inter-working libraries. But do not, a representative from one of the major federal libraries implored, sacrifice the heterogeneity in collection policies that has created a robust information base.

With respect to users, however, one futurist cautioned that the values represented by current scholars and archivists are not widely shared among teenagers and young adults. The Library was encouraged to conduct a focus group with 18-22 year olds as part of its ongoing planning activities and more generally to articulate a vision, embodying American values behind which support might be mobilized. Others echoed the need for attention to the values that any future system might embody, including traditional respect for values embodied in the First Amendment.

Who should pay?

Related to the question of developing the collection were discussions of “who should pay?” Related to the question of “who should pay” are issues related to valuing donated content and use of possible tax incentives, a strategy that has been employed for cultural preservation projects. There is substantial ambiguity surrounding economic issues and modeling costs remains a research topic. Moore’s Law notwithstanding, a representative of a major research university library lamented, the cost of storage remains a line item in the annual budget. And costs are not going down.

One model is to attach requirements for support to acceptance of private collections; another is public funding, which is implicit in the popular notion of a public library and the perception that the taxpayer has already paid for library and archival services. The representative of the National Archives acknowledged the immensity of the digital archiving effort and described efforts of the agency to leverage work done by others – with enlightening results. As a result of his experience with the San Diego Super Computing Center, “They've totally turned around my view of media,” he said. “They migrate their entire storage system every two years to save money. . . But you know, I used to worry about how long the media lasts. Now what I worry about -- will it last long enough to be cost effective?” The Archives is also looking into work done by the Defense Advanced Research Projects Agency (DARPA), the storage resource broker. “The principle of the architecture for this preservation system is you want your preservation system to be independent of the infrastructure you're using at any given time, independent in that you want to be able to replace any component with minimal impact on the rest of the system and replace any component with negligible impact on what you're preserving.”

Technology

Technology informed almost every aspect of the discussion. Yet no one seemed to believe that purely technological solutions would be sufficient. As previously noted, representatives from the entertainment and broadcasting industries were well aware of the technological differences among their respective media – as well as the common concerns – evidenced by the existing technical working groups on digital archiving within some of these industries. More generally, though, several observers pointed to a technical and professional culture within computer scientists that appears indifferent to the problem of preservation. One academic with very strong ties to the entertainment industry observed, “You know, as someone who worries about content creation and what that means from the relationship between how things exist physically and how they were actually intended to be created, we find that there is very little recognition by people coming purely from the technology side, often, of how the standard for storage alters the actual content of the material and makes it into something different.”

Differences in cultures and motivations notwithstanding, there were grounds for collaboration around notions of “digital asset” protection and management, as these systems are known in corporate circles. As previously observed, one proposed architecture that provoked constructive discussion posted an archive “layer” and a digital library services “layer” on top of it, which can allow for highly heterogeneous services, appropriate to local conditions, while preserving a consistent archive. The model is also congruent with industry-based distinctions between digital asset protection (the archive) and digital asset management (the digital library services). The Library was urged to continue the process of engagement. One representative of the film industry commented, “We have tremendous resources from people who are involved on the technology side and we would like to be a resource for the Library of Congress as this goes forward.”

Preservation and access

The digital medium offers many technical and organizational challenges, of which two of the fundamental are volume and the ability to manage volume in ways beyond storing it in increasingly efficient storage media – itself an area of both research and concern. Thus, a computer scientist said, “The most important thing [in this ability to capture huge volumes of data] is -- for it to have value -- is to be able to search it, retrieve from it, and summarize from it, because there's so much, you can't--it doesn't do you any good to get, you know, 10 million results when you make a query, or even 1,000 results.”

No one advocated a totally dark archive. Print publishers, for example, consistently talked about the importance of understanding what constituted a “preservable” electronic or digital document. In part, this arises from technical concerns. As one publisher of scholarly journals pointed out, technical glitches and bugs in a digital collection will not surface unless users “beat on it.” From a business perspective, the economic value of an entertainment asset is reduced if

people cannot use and enjoy it and re-use of older footage has proved valuable, sometimes unexpectedly. Finally, scientists, scholars and creative artists want their legacy preserved and appreciated. A balance between protection and use seems essential. How that balance should be achieved, where the boundaries should be drawn and what the priorities should be remain issues on the table.

That said, representatives of the music industries, seconded by film and television, consistently expressed greatest concern for material “born” or “reborn” digital that have a life-span of five to ten years. “But we need to define how we get away from the born digital to stop the death and how can we bounce with each other as a depository or some method to basically enhance what you want and basically keep alive what we want” These materials represent our “dowry”, to use the phrase employed by a representative of public television. Said another representative of public broadcasting: “In public broadcasting, we want to use our assets to engage people in a civil society and to engage them in their community. So that's our goal. They're actionable. They're about access. They're about use rather than mere preservation.”

Priorities and next steps

Conveners recommended the following:

1. Continue the discussion among relevant players in the concerned industries and through participation in existing industry working groups. There was clear support for this among the entertainment industries (music, film, television, publishing) as well as among related groups, such as the Annenberg Center (University of Southern California), with interests in these media. This approach includes leveraging other research efforts, such as those sponsored by DARPA and the National Science Foundation (NSF). “The archives market is not sufficient to do this alone, we have to ride the coat-tails of others”
2. “Stop the loss.” Take advantage of key individuals in entertainment industries to get started with protecting assets and then develop a process to encourage preservation.
3. Create a “safe harbor” where experiments might be conducted and where concerns over potential violations of DMCA might be allayed. Related to the notion of experiments was the advice that pilot projects address both the most and least tractable archiving problems and that such experiments be used to build broad public and congressional support. One idea was to build a core collection based on voluntary deposit and/or material harvested (“snarfed down”) on which experimental models might be built and in which issues might be solved as they arose and cross licensing/contractual and cost recovery and funding arrangements might be attempted. Then, the challenge would be to see whether and how the model might be scalable.
4. Examine conditions under which restricted or perpetual access might be appropriate with respect to intellectual property rights.

5. Develop standards, tools and frameworks for a decentralized technical architecture and begin putting content into that framework. This should include identifying what is urgent and what is “archivable in principle.” This might also include identifying types of information that are most stable and that are not stable -- supplementary material to scientific journals, chat rooms, etc. One source of content might be the Web, but in any event, it might be prudent to narrow the focus. These standards should reflect U.S. values (e.g., broad access to information).
6. Streamline and expedite procedures for identifying rights holders. This might include establishing procedures whereby a “default” option might exist in which permission was granted if a good faith effort could not identify the relevant rights holder, for example.
7. Develop a recommended set of guidelines for donations of digital material. Provide a length of time during which collections, policies, etc. are re-evaluated. Identify a set of standards reflective of: (1) trust, what represents a good job (area in which CLIR has made contributions) and (2) quality of service for its customers.
8. Provide a set of tools that would enable use, and migration.
9. Partition the problem – identify a set of technologies that will act as bookshelves, an archival structure that will be stable for at least two decades, or at least remains constant for some period of time. In addition, develop guidelines for standards for media, formats and metadata.
10. Develop a process for ubiquitous, continuous capture, that less intrusive but not indexed; then develop tools to search, retrieve and prioritize. Accessibility should cross boundaries of media, ownership, physically distinct content.
11. Clarify legal requirements with regard to copyright, privacy, confidentiality and so on. Pick the strongest friction points and try to deal with them, otherwise they will fester as exemplified by the discussions over the DMCA.
12. Consider the international dimensions. This would include identifying standards and best practices in other national libraries (Australia, Netherlands, British Library) as well as building system that are compliant to international systems.
13. Articulate a series of high level design principles, a vision that everyone can understand. This should include a well-defined scope that identifies critical institutions, relationships, generic types and products.
14. And finally, with thanks to Nike, Inc., “Just do it.”

*Submitted to:
The Council on Library and Information Resources
and the Library of Congress*

*by:
Amy Friedlander, Ph.D.
Center for Information Strategy and Policy
Science Applications International Corporation
McLean, Virginia*

December 2001