Digital Preservation Archiving Workshops -- Focusing the Agenda

and the Coalition for Networked Information

Following are excerpts from a proposal submitted to NEH by Peter Graham on behalf of the proposed sponsoring organizations (ARL, CLIR and CNI), and the Rutgers University Libraries. The full Proposal text is available as a PDF file. ARL, CNI and CLIR have committed themselves to be Sponsors of this proposal and will form the Advisory Committee to work with the Project Director (PG) on substantive and organizational issues. The proposal was submitted July 1, 1997; NEH word can not come until April 1, 1998. Meanwhile, CLIR has offered a planning grant for the interim period. A first organizing meeting of the Sponsors and PD was held in Washington, DC on September 17, 1997. Tentative plans are for the first workshop to be held in June, 1998.

PROPOSAL SUMMARY: The proposal requests funding for five two-day workshops on specific problem topics of digital archiving, with participants both invited and self-selected through competition. Each workshop will devote itself to the particular topic with the aim of noting what progress has been made so far, specifying what the major problems in the topic are, and what would be fruitful means of investigating the problems and solving them. Reporters will produce summary papers outlining conclusions for both print and World Wide Web publication in prompt, convenient formats. The aim is to produce research and action agendas to guide already-active digital library creators in furthering effective digital preservation and archiving. The Sponsors for the Project and the workshops will be the Association of Research Libraries, the Coalition for Networked Information, and the Commission on Library and Information Resources.

Participants will include knowledgeable individual practitioners from organizations active either in planning or in creation of digital repositories. Participants will also include self- (or other-) nominated individuals. Workshop attendance should comprise approximately 30-50 people. This will be large enough to assure

multiplicity of views and ability to work in smaller breakout groups, but not so large as to prevent considerable informal exchange. Participants should expect to fund their own expenses in attending the workshop (thus this is a matching fund proposal with respect to most participants). The Project, using grant funds, will fund a small number who could not otherwise come.

At each workshop the goals will be:

- To identify what research or development on the topic has been done, if any, and where clear consensus may already have been reached.
- To find a consensus on the most pressing research and action agendas for potential practitioners;
- Identify and prioritize, if possible, the next action steps that should be taken in the topic area.
- Where possible, identify which organizations or institutions are the best candidates for proceeding, including pressing for volunteer organizations at each workshop. The workshops will each last for approximately two days (probably requiring two overnight stays for most attendees). The workshops will include invited speakers, breakout sessions and plenary sessions.

Summary statements reflecting each workshop outcome will be prepared immediately after each workshop. These statements and documents will be made available on the network through a Web site maintained by either the Director or one of the Advisory Committee participants; the editing and presentation values will be at minimal levels to expedite their appearance. At the conclusion of the Project a more formal version will be published both in print form and on the Web and made widely available.

PROPOSED WORKSHOPS:
(See the link to the full Proposal, above, for fuller discussion.) Following are proposed workshop topics, with related questions for which further direction is needed. The five workshops will be selected from among the six described below, after wide consultation. The topics of the workshops may change as further discussion with the active community is held, and later topics may change as workshops are held. Each workshop should include a review of present activities in the topic area, as described above. The planning for each workshop will no doubt result in further questions and subtopics:

1. Archival contents and redundancy ("collection development"): What is the necessary agenda?
What needs to be archived? What doesn't? What distinguishes what research libraries archive from what other organizations archive (e.g. ICPSR, NARA)? To what extent are existing collection development practices valid in the archiving environment? What needs to change? What do libraries need from parent institutions in this regard? What is possible? What role will universities play? Will

universities agree to serve as archives? What are the repository redundancy needs? Issues include storage costs and network bandwidth, disaster issues, geographic concerns, loci of authoritative sites, etc. What are repository agreements likely to be based on? Consortial arrangements, political convenience (e.g. statewide arrangements), common vendor relations, subject area agreements, etc. What are the pros and cons of these possibilities?

2. Technological Obsolescence and Migration techniques: What is the necessary agenda?

Just in case or just in time? (Migrate all electronic data through all technologies or wait for demand to determine what is migrated?) What are practical technology emulation possibilities? Migrate the information object only or migrate its tool for use? Link them together or provide only implicit linking information? What are the major issues in migration?

3. Current and immediately future hardware technologies: What is the necessary agenda?

What are effective current storage media? To what extent should future refreshing/migration be taken into account in choosing current storage media? Is vendor cooperation possible? What are the prospects for future hardware technologies as archiving media (HD-Rom, ceramics, etc.)? What is the interplay between hardware permanence and software permanence? What are the strategic consequences? Online, nearline and offline: when to use what.

4. Authenticity (integrity; intellectual preservation): What is the necessary agenda?

What are potential tools and/or mechanisms for assuring integrity? What is the experience so far? What are the user burdens? How well do they work? What infrastructure do they require? What costs? What needs to be assured, and what need not be? What principles can be established in this area so that constant individual decisions are not necessary? What does integrity mean? Can content assurance be provided though presentation modes change? Does that matter? What does integrity mean in a migrating environment? What is the fundamental "content" to be preserved? What can afford to be lost? Are there differences in document type that affect the integrity mechanisms? What are the relations between integrity mechanisms and version control? Do they matter?

5. Repository certification (NOTE: this workshop best held after the others): What is the necessary agenda?

What should be the standards for certification as a repository? Should institutions self-certify? If not, who should, or how should they be? Should products be part of the certification process, or process? Is research library repository certification a stand-alone issue, or part of a larger issue (are research libraries one component of an archiving repository community)? Should certification standards be different for different communities?

6. Intellectual Property: What is the necessary agenda?

What are the intellectual property constraints placed on preservation distinct to the preservation activity ? What are the potentials for libraries gaining some form of intellectual trusteeship over information, whether in partnership with producers or in the event of their failure to act ("fail-safe" triggers)? What intellectual property legislation, or international agreements (e.g. through WIPO), or commerce conventions, are needed to facilitate preservation of the human record?