

EXECUTIVE SUMMARY

The exponential growth in the creation and dissemination of digital objects by authors, corporations, academicians, governments, and even librarians, archivists and museum curators, has emphasized the speed and ease of short-term dissemination with little regard for the long-term preservation of digital information. Digital information is inherently more fragile than traditional technologies such as paper or microfilm. It is more easily corrupted or altered, without recognition. Digital storage media have shorter life spans and require access technologies that are changing at an ever increasing pace. Because of these technological advances, the time frame in which we consider archiving becomes much shorter. Groups or individuals who did not previously consider themselves to be archivists are now being drawn into the role, either because of the infrastructure and intellectual property issues involved or because user expectations demand it.

This has raised the awareness of the issues surrounding digital archiving and preservation among information managers, librarians, publishers, and archivists. ICSTI, being a community which represents many of these information industries, has been involved in this issue for several years. Based on the most recent efforts by the ICSTI Electronic Publications Archive Working Group, this study was undertaken to provide information on the state-of-the-art and practice in digital electronic archiving.

Purpose, Scope and Methodology

In this project, “digital electronic archiving (DEA)” is defined as the long-term storage, preservation and access to information that was “born digital” (created and disseminated primarily in electronic form) or for which the digital version is considered to be the primary archive. [This does not include the digitization of material from another medium (such as digitization of paper or microfiche) unless the digital becomes primary.] Based on the analysis during this project, there is no common agreement on the definition of long-term preservation; the time frame is long enough to be concerned about changes in technology and changes in the user community. Depending on the particular technologies and subject disciplines involved, this time span may vary from 2-10 years.

The purpose of this study is to identify the state-of-the-art and practice related to DEA policies, models, and best practices, with an emphasis on the most “cutting edge” approaches. The study emphasizes those areas of most concern and interest to ICSTI members and those research areas previously identified by ICSTI as necessary to move the digital archiving discussion forward. Primary attention is given to operational and prototype projects involving scientific and technical information. The study is international in scope. It includes a variety of data types applicable to scientific and technical information, including data, text, images, audio, video and multimedia, and a variety of object types, such as electronic journals and monographs, satellite imagery, biological sequence data, and patents.

The study methodology involved an initial survey of the ICSTI and CENDI members (see Appendix A-1 for a full copy), as well as a literature review and contacts with experts, to identify the most “cutting edge” projects. The highlighted projects cover six countries (U.S. (9), UK (2), Canada (1), Australia (1), Sweden (1) and Finland (1)). Four organizations are considered to be international in scope, because their funding sources and scope are not bound to a particular country. The projects come from a number of sectors including government scientific and technical programs, national archives, national libraries, publishers, and research institutes. Information about other projects is included where applicable.

After the initial questionnaire, follow-on discussion questions (see Appendix A-2) were developed and aimed at identifying emerging models for the relationship between the various entities in the information chain (users, intermediaries, primary publishers, secondary publishers, online vendors, and others) as they relate to archiving; the metadata information that is being gathered; how the archive will be maintained and accessed; an estimate of the costs to be incurred for start-up and maintenance; and outstanding issues and possible best practices. While technologies for storage and retrieval may be mentioned in the report, technology is of secondary interest to the understanding of policy and practice.

General State of the Art/Practice

The issue of archiving digital objects brings together several normally diverse communities -- archivists, records managers, librarians, data center managers, and data producers. There is so much activity among various groups that it is difficult to encapsulate the general state of DEA. However, there are a few general models that can be highlighted as emerging. The models have genesis in one of the diverse communities, but may have applicability to others.

It is noteworthy that many of the major projects in digital archiving are of a cultural or historic nature. While the emphasis in this study has been on scientific and technical projects, the humanities-related projects have provided the basis for much of the current thinking in this area. They have been used peripherally in this study for what they offer to science and technology, or for the scientific and technical information components that many have.

Identified Organizational Models

The highlighted projects were analyzed for commonalities that would identify organizational models for DEA. The approach taken is an organizational one, loosely based on the previous work sponsored by the Arts and Humanities Data Service (AHDS) (Beagrie and Greenstein, 1999). Four major organizational models -- Data Centers, Institutional Archives, Third Party Repositories, and Legal Depositories -- were identified. An additional conceptual model for interoperable archives is also described. These models are based on differences in the information flow, the management of the life cycle functions of the archive (creation, management, preservation, and access), responsibility and ownership of the data, and the economic model.

The most mature archival model is that of the data center. Three subcategories of data centers were identified based on the degree of homogeneity and centralization. Centralized data centers, such as the National Digital Archive for Datasets (UK), have numerous contributors, but a central repository and administration. This model allows for easier integration of the data and more consistent adherence to standards. However, there may be little backup for the central repository, particularly if funding is cut. It is also difficult to include new data producers with varying data models, standards and primary audiences. Federated data centers, like the NASA Distributed Active Archive Centers (DAACs) operate in a distributed, but closely guarded environment with common standards and practices, and a single user interface. There is redundancy in the federation's ability to respond to user needs. With looser standards, more partners may be involved more easily. Cooperative data centers do not currently exist, but there is a prototype under development between the San Diego Supercomputer Center, the U.S. National Center for Environmental Analysis and Synthesis, and the Long Term Ecological Research Network. The aim is minimal metadata and system standards, acknowledging the diversity of data types, models and structures in ecological science.

On the whole, the data centers are also the simplest organizational model. The intellectual property rights are generally clear, because the owner is the funding agency. The economic model allows free access by the funding agency and, since many of these are government sponsored or internationally developed data banks, the public also has free access. Additional charges may be levied for extraordinary services or for access for commercial purposes. However, it is unclear how well the practices of these data centers, which have large volumes of relatively simple data, would migrate to other communities and object types.

Institutional archives are generally a department or branch of an institution that collects and preserves the intellectual capital for that institution. These institutions can include publishers, data producers, societies, cultural organizations, government agencies, academic institutions and industries of various types. Institutional archives generally have some level of ownership of the information. Often access is limited to members of the organization, to subscribers, or to partners in a particular project or venture. Many corporations and institutions archive only what is required by regulation, fearing legal ramifications if certain information is retained. However, there are organizations such as pharmaceutical, chemical and petroleum companies, where internal scientific and technical information is critical to the perpetuation of research and development. Institutional archives may also increase as the knowledge management technologies connected to intranets reach a wider market.

Third party repositories are the third model. They tend to derive from the journal publishing industry, rather than government data centers or institutional records needs. They can be divided into two types: Publication Service Providers and Repository Management Agents.

Publication Service Providers serve other roles in the information chain. In addition to their primary role as vendors, electronic publishers, or jobbers, they may also provide digital archiving as a service to their clients, which are primarily learned societies and publishers. This is the most

complex organizational model for archiving, because there are numerous roles being played by the participants. Often the economic model for the archiving is not clear, because it is bundled with the other services that the publication service is providing. Examples of Publication Service Providers who also provide archiving services include ingenta, Ltd. and HighWire Press.

Repository Management Agents are an emerging model in DEA. These organizations act as trusted third party repositories, but do not serve any other function in the value-added chain. They provide a safety net by continuing to provide access to the digital object should the publisher or producer of the object determine that it can no longer archive the material or if it goes out of business. Examples include JSTOR and OCLC's Electronic Journals Project. Both projects have substantial numbers of journals available. The majority of JSTOR's current titles are in the humanities and social sciences. However, they have recently begun a project on a Science Cluster, which will include AAAS's *Science* and the publications of the National Academies of Science (US). In both cases, the charges are borne by the user or library. JSTOR's pricing model is based on a yearly subscription to the JSTOR service, with rates differing by size of institution. OCLC's model is based on the library's subscription to the electronic journal directly through the publisher, through a jobber, or in some cases through OCLC. The agreement requires linked access to the publisher's archive or deposit of a digital copy with OCLC. OCLC is currently working on the long-term business and pricing model for this service.

The fourth model is that of the Legal Depository. There are generally two types of legal depositories: national depositories and national libraries. The national depository (or archive) has tended to document the business of government, which includes administrative documents. The national libraries are generally charged with maintaining the culture, history and intellectual output of the country by collecting what is published within that country. Both national libraries and national depositories have sought to handle digital material. As part of digital government initiatives, archives such as the UK Public Records Office and the U.S. National Archive and Records Administration have extensive electronic projects. In the UK, the PRO has separated the responsibility for archiving digital datasets from the archiving of digital office records.

Some national libraries have sought to extend their mandate to digital information. In many cases, they are doing this without the benefit of legislation. The PANDORA Project of the National Library of Australia has the most extensive guidelines for the selection of Web-based "Australiana". The National Library of Canada's Electronic Collection incorporates electronic books and journals published in Canada in its regular workflow, based on the results of the Electronic Publications Preservation Project pilot study. The National Library of Sweden is using robots to harvest all relevant domain names and Web servers, archiving the content without review. Projects are also underway at the National Library of Finland. The Networked European Depository Libraries (NEDLIB) project is funded by the European Union to investigate the procedures, standards and infrastructure needed to support a multinational library network for digital archiving.

Though not an operational model, the interoperable archive model described in the recently

drafted *Reference Model for an Open Archival Information System (OAIS)* (Consultative Committee for Space Data Systems, 1999) provides insight into the future of a hierarchy of archival organizations and heterogeneous archives, and is worth evaluation in this context. This reference model provides terms of reference, conceptual data models, and functional models for open archives that can interoperate. The models are based on packets of information, including the data object itself, descriptive metadata, representational information which helps to interpret the bits in the data object (e.g., the ASCII table), and specific information needed for preserving the object. Based on the exchange of these packets, and the standardization and crosswalks among the metadata formats used to present the information, objects can move from one archive to another, and archives can be searched simultaneously. Many experts, including the CEDARS project in the UK, are investigating whether this data-centered model could be generalized across other data types.

Life Cycle Managers and Their Roles

The results of the study were also analyzed for the changes in the roles of the traditional players in the information dissemination chain. The roles analyzed include creator (author), publisher, secondary publisher, library and consortia, funding source, and user.

The analysis found that creators and users are not very involved in the digital archiving process. However, this is changing as organizations are requiring metadata creation with digital objects, and as software is developed to make the creation of such metadata (and even its automatic extraction) easier.

Publishers are involved in digital archiving in a number of ways. The most vocal are the learned society publishers who consider this to be part of the mission for their discipline or organization. However, the economics and long-term viability of such preservation (as the content of the system grows) is unclear.

Few secondary publishers have expressed an interest in digital archiving according to an informal study conducted by the National Federation of Abstracting and Information Services. However, many of these services have a long history of migrating and maintaining archival collections of bibliographic records in a discipline. Third party repositories, particularly OCLC, and national libraries (the National Library of Australia) have designed systems to take advantage of the bibliographic records as the catalog record that provides access to the full archival object.

Libraries, particularly consortia, have been instrumental in raising digital electronic archiving issues. As they seek to provide access to electronic journals, which no longer provide a consistent physical copy that can be owned and preserved, libraries have developed guidelines for license agreements which include statements regarding digital electronic archiving. Licenses generally provide for a trusted third party or the library itself to receive and archive an electronic copy immediately or when it is no longer available from the publisher.

Funding is a key driver in the evolution of archive models. Funding is provided by government organizations, national and international science initiatives, private foundations, research institutes, and museums. Funding organizations in many quarters have espoused the need for archiving digital information. Unfortunately, in many cases, particularly at the government level, there have been mandates without supportive funding. In many cases, guidelines have been developed, but they are not detailed enough to provide real guidance on issues of long-term preservation, media migration, and planning for the related costs in program and project budgets.

Best Practices

The evaluation of the research results was organized by again looking at the best practices by the information life cycle for archiving material across the various models. The life cycle functions are creation, acquisition/collection development, cataloging and identification, storage, preservation and access.

Practices used when a digital object is created ultimately impact the ease with which the object can be digitally archived and preserved. The preservation and archiving process is made more efficient when attention is paid to issues of consistency, format, and standardization in the very beginning of the information life cycle. Institutions are beginning to require a more limited number of formats for some objects created under their auspices.

All groups involved acknowledge that creation of good metadata at the source of data creation is where the long-term archiving and preservation must start. As standards groups and vendors incorporate Extensible Mark-up Language (XML) and RDF (Resource Description Framework) architectures in their word processing and database products, creating metadata when the digital object is created will be more efficient and more rapidly adopted. However, work remains to identify the specific metadata elements needed for long-term preservation, particularly for non-textual data types like images, video and multimedia. Others in the information creation chain for formal materials, e.g., publishers, funding sources, learned societies, etc. can play a large part in promoting such attention on the part of creators and the development of relevant preservation standards.

Cataloging and identification issues are often interrelated with decisions about what to archive and how long it will likely be retained. The metadata to be collected, and the degree to which a standard will be used, depends on the type of organization doing the archiving, the resources available, the type of material to be used, and the requirements of funding organizations. The most common formats are MARC and Dublin Core. Only the traditional publishers appear to be using the Digital Object Identifier. Other stakeholders have developed their own identification schemes.

The national libraries are taking the lead in the development of guidelines related to the acquisition and collection of digital objects in archives. The PANDORA project has extensive guidelines for a variety of Web-based (primarily textual) material, including ephemera. Issues

addressed in the guidelines include determining what should be archived, determining the extent or the boundaries of the digital work, and archiving related links.

Storage issues center around hardware/software migration. New releases of software can be expected every 2-3 years. Migration to new media and hardware occur less frequently, but can be expected at least once every 10 years. The general response from those queried about these issues is that they have no firm plans for migration, but will plan to stay up to date with current technologies by migrating the content to each new technology. The issues of cost have the biggest concern here, and there is now a sense of having to deal with it as best we can as the technologies change. All the respondents followed industry best practices related to refreshing the media, back-up, recovery and remote storage for disaster recovery.

Preservation is the aspect of archival management that preserves the content as well as the look and feel of the digital object. In cases where the archiving is taking place while changes or updates may still be occurring to the object, such as with datasets or electronic journals, attention is being given to refreshing the site contents. The National Library of Australia allocates a gathering schedule to each “publication” in its automatic harvesting program. Obviously, the burden of refreshing the contents increases as the number of sources stored in the archive increases.

Most organizations lack formal retention policies, because they are relatively new to digital information and storage costs continue to decrease at a faster rate than the increase in the size of most archives. The most common answer is that the organization will archive “everything” for “all time”. Other than legal depositories, there is little recognition of the need for more definitive policies in the future based on the value of the information to potential users, the resources available on the part of the archiving organization, and the desires of the funding agency. Those who recognize the need for such policies also acknowledge that we do not have a crystal ball, and, therefore, it is difficult to determine precisely what will be of value in the future. When the burden gets too great, particularly for commercial institutions, it may be necessary for public institutions to intervene and provide a backup archiving service for objects that are no longer of sufficient commercial value to warrant inclusion in the commercial organization’s archive.

Preservation has also involved the decision of whether to transform the incoming information into a new, more standardized format, or to retain the native format. While the answer to this depends to some extent on the user community being served by the archive, and the degree to which the transformed format matches the native format, there appears to be a tendency to transform to the newest related format, for example from the current version of TIFF format to the next. However, in some cases where legal responsibilities intercede, the original is always retained, along with the transformed format for access.

Regardless of the decision about transformation versus native format, preserving the “look and feel” of the object remains an issue. If the digital information is transformed, the question is how much does this impact the “look and feel”? If the information is retained in native format, how

will the “look and feel” be provided when the technology changes in the future. Migration is the most common answer to this issue, realizing that the “look and feel” may not always be retained. An alternative is an emulation strategy. Emulation involves reconstructing the behavior of the hardware and software in the future environment in order to recreate the “look and feel” of the original digital object in its old environment. This will involve cooperation on the part of hardware and software vendors to provide access (or perhaps restricted registries) to proprietary information about the hardware and software. However, to-date there have been no large-scale pilot projects that would indicate that the emulation approach is practical or scalable.

Finally, the life cycle of archived material requires access or the ability to reuse the information. Currently, all projects reviewed have or are planning Web-based interfaces to their archives. Additional interfaces are available for certain specialized information, such as the datasets available from the data centers. However, digital archivists are looking beyond the Web to another as yet unknown interface, and they consider the interface to be another technology that can change rapidly.

Depending on the intellectual property and licensing issues, the access to the objects may be restricted. Archives that store copyrighted materials, proprietary information, or restricted government information must also deal with security and authentication issues. Processes being investigated or put into place may include digital signatures and certificates, in addition to the more traditional IP address and user name/password log on procedures. The ability to download and reuse the information also differs depending on the archive, the license agreements with the rights holders, the type of user and his relationship to the archive or rights holder, and the amount and type of material being downloaded. Because of the ease with which digital material can be altered, either knowingly or unknowingly, mechanisms such as watermarks or encryption are viewed as key tools in the process of digital preservation.

Best practices are also beginning to emerge for different format and object types. Image archives are particularly concerned with the type of metadata information needed for preservation and access to these images, including changes in resolution and compression techniques. The Research Library Group, the Digital Library Federation, and the U.S. National Information Standards Organization, partnered with a variety of European organizations, are involved in developing such guidelines and metadata elements which will be available for review in the next few months.

All the issues related to the various data types, and more, are bundled into the issues surrounding the archiving of multimedia works. Since efficient archiving, access, reuse and preservation differ based on data type, multimedia, which combines various data types, cannot be dealt with by a single approach. In addition to the archiving of a series of objects that make up the multimedia object, it is important to be able to bring the collective multimedia object back together again. Projects in this area are underway within the US Department of Defense and the US National Library of Medicine. A standard file format for multimedia is being developed by Microsoft.

Costs/Resources

Although cost is recognized as a basic driver in DEA, it was also the most difficult aspect on which to gather information. In some cases, a lack of response was because of the proprietary nature of this information. However, in most cases, the respondents indicated that they just didn't know how much the archive was costing or would cost in the future. For publishers and producers, the cost of archiving is still tied up in the cost of manufacture. This is also true of publications services where the archiving is considered an added benefit to the publishers who are served. Until several large archives have gone through at least one or two migrations or emulation developments, it will not be possible to separate the cost for the archives from the cost of doing business.

Anecdotal information is available from several national library or institutional projects that are archiving Web sites, electronic journals and other digital publications from the Internet. However, the information is generally presented in terms of the number of full or part-time staff being devoted to the effort at this time, with no indication of hardware/software or other infrastructure costs now or in the future.

In addition to questions of start-up and ongoing operation, there is a serious issue of the long term financial commitment to archives. Increasing recognition by scientific authors and funding sources is key to the success and sustainability of an archive. Several experts interviewed suggested that an endowment model might be needed. This would set aside a portion of the payment for the use (whether storage or access) of the archive for its perpetual care.

Conclusions

Based on the analysis of the organizational models, the changing roles of traditional stakeholders, and best practices in digital life cycle management, general conclusions can be made in the areas of most interest to ICSTI. These include policies, organizational models, and economic models.

The policy issues of major concern seem to be the intellectual property issues, and with them the related security and authentication concerns. To greater or lesser degrees, all stakeholders in the archiving and preservation chain are concerned about intellectual property. For many of the data centers, the issue is put in public versus commercial use terms, and is reflected in the types of access and services provided and the charges placed on them. For publishers and producers, intellectual property concerns are reflected in the kinds of business arrangements used to promote their archives. Intellectual property concerns have led some organizations to consider institutional archives, where the information remains under their control. Others, lacking the resources to do this, but still concerned about their intellectual assets, are contracting with publication services or trusted third-party repositories. Part of these contracts requires security and authentication on the part of the archive, as well as specific procedures for granting and continuing access. Libraries, consortia and users are increasingly attuned to intellectual property issues, and their concerns for fair use in a digital environment are often reflected in the license agreements that are signed.

Five organizational models for digital archiving have been identified. Aggregation on the part of repositories, publication services and legal archives is likely to continue as stakeholders struggle with how to make the information accessible with common interfaces, in the midst of cost and intellectual property concerns. Based on the numbers and types of organizations involved, the need to integrate across format and object types in the sciences, increased emphasis on multimedia, and ever-changing technologies, the organizational model for archives in the foreseeable future appears to be a loose network of archives covering special disciplines, geographic areas, or object types. Using network technologies and interoperable standards, the future model will likely be a network of disparate but interoperable archives. Individual communities are likely to develop standards and common practices. Interoperability in a heterogeneous environment is likely to be required. The Open Archive Information System (OAIS) reference model, described earlier, appears poised to promote this interoperability beyond the realm of data-centered archives.

Similarly, it is likely that there will be a variety of economic models for digital archiving. This will impact not only the way the archives are managed and who manages them, but the value (and the cost) involved in retaining older materials. Some archives will be commercially viable, others will not. Some will need to charge for services, while others will not. When archives are governmentally appropriated, there is increasing recognition of a long term maintenance commitment, but there does not seem to yet be sufficient definitive action and funding to support this recognition.

With a large number of models and increased interest in the future of digital information, many stakeholders are getting into the archiving business. There are many organizations that appear to consider this a reasonable avenue for business growth. With the large infrastructure and varying skills needed to perform digital archiving satisfactorily, we may be seeing the rise of a new industry. Smaller publishers in particular may continue to look for avenues by which they can contribute to one or more archives, without undertaking the infrastructure development themselves.

Multiple economic and organization models are likely to persist in the DEA environment. As the report of the ICSTI Electronic Publications Archive Working Group suggested, a hierarchy of archiving organizations may be needed to overcome the economic and intellectual property issues that continue to abound in the digital environment.

It appears the discipline specific, as well as national and global archives, will be built incrementally on the basis of pilot projects that lead the way and evolve into a complex network of content infrastructure. The issue has been recognized and the bandwagon is growing. In summarizing best practice areas, we see building blocks for future developments. The trick will be the coordination of these archives to reduce the expense of unnecessary redundancy, to tie the system together in an integrated fashion for the user, to ensure long-term funding for these archives, and to mechanisms to protect the rights of both copyright holders and users.

Recommended Next Steps

Based on the survey and analysis conducted during this project, the following actions are recommended for consideration.

ICSTI

1. Many models are evolving and taking hold. Each stakeholder will be affected and the activities should be monitored for more specific and ongoing relevance to ICSTI member groups:

! Hold discussions on impacts of the various models (both organizational and economic) for classes of ICSTI members. Monitor projects selected by members to be models for their part of the industry, and provide opportunities for interaction between these projects and appropriate communities within ICSTI.

Projects that include the specific stakeholder group or the portion of the information life cycle function in which a particular organization is interested should be monitored with specific reports back to ICSTI members interested in these particular areas. In addition to project monitoring, opportunities should be provided for interaction between the project managers of the selected projects and ICSTI members. The next annual meeting, or a special meeting cosponsored with ICSU, UNESCO or some other organization, would provide a forum for the discussion of these specific projects. It might also be valuable to hold the session concurrent with a major meeting where these projects might already be represented.

! Interpret the draft Open Archive Information System (OAIS) Reference Model for the ICSTI Communities

Since heterogeneity and a complex network seem to be evolving, the OAIS Reference Model is one worth further group exploration. It stands as a possible framework for data interchange needed across the various functions of an archive (regardless of the players involved), and across archives. However, the current reference model is still very data-centered. ICSTI should convene a small group or groups of stakeholders to interpret the reference model for the different communities -- primary publishers, secondary publishers, and libraries. During this process it should be possible to determine if the reference model has utility for a variety of stakeholders and a variety of data types. The CEDARS project in the UK has expressed an interest in working together with ICSTI on this review. This follow-on project should be done in the context of the ISO review of the draft reference model and should consider interoperability, standards, common practices and economic models that will have to coexist. The benefit to ISO and the Consultative Committee on Space Data Systems is that they will obtain a review by an expert community, outside the data community. The benefit to ICSTI is that it may find a model that can be used across its members and to inform the community at large.

! Develop a Digital Electronic Archive Registry Emphasizing Digital Publications

The Electronic Archive Registry, recommended by the ICSTI Electronic Publications Archive Working Group, may act as a transitional mechanism between the current distributed, unintegrated archiving projects for electronic publications and the fully networked environment envisioned by the OAIS. The Working Group envisioned this registry as a finding aid for the location of where, by whom, in what format, and what parts of a publication are electronically archived. The data elements required for such a registry and the procedures whereby the registry is created, maintained and accessed must be developed. The Working Group suggested that the registry could be added to the ISSN system. The concept should also consider the work of other groups such as the Digital Object Identifier (DOI) Foundation and the national libraries/bibliographies.

! Monitor and report on the key projects related to the cost and organizational issues of digital archiving

This review has identified that there are still significant unanswered cost and economic questions related to long-term digital archiving. Some of these questions are related to the speed of technological change, while others are institutional. However, there are several significant projects under way that have been briefly identified in this report. They should continue to be monitored and progress on them reported to the ICSTI community. Recommendations for projects to be monitored include NEDLIB, the objective of which is the networking of depository libraries and the development of digital depository format standards for publishers; CEDARS, which is looking at the networking of UK archives; and Cornell University's Digital Library 2-Initiative which will address cost and organizational issues. Relationships should be established with these projects in order to learn about their progress and be able to report on the outcomes to the ICSTI listserv.

2. As appropriate, work at individual organization levels to promote digital archiving practices:

! Recommend to ICSTI organizations that digital standards for metadata and object identification that are under consideration be reviewed with a particular eye to their ability to support long term preservation and access.

In particular, work to ensure that the concept of archives and preservation is developed and used within existing and forming standards for metadata and identifier.

! Provide testbed material for projects when possible.

A significant way for ICSTI members to become involved and to learn more about the challenges and best practices in this area is to provide material for digital archiving testbeds. This is already being done by Elsevier, Kluwer and Springer in the NEDLIB project. There may be similar

opportunities with other projects, including CEDARS and the Cornell University DLI-2 projects.

! Promote multilateral projects, to promote the development of best industry practices in digital archiving

Promote round-table sessions at a follow-on ICSTI meeting that would bring together ICSTI members working on similar issues related to digital archiving so that resources, lessons learned, and pilot projects could be shared. Of particular importance would be discussions and pilot projects related to business models for digital archiving and intellectual property issues (particularly between national libraries and publishers).

Both ICSTI and CENDI

1. Make ICSTI/CENDI's interest in this area known so the organizations stay involved with the forefront of activities and continue to keep the debate visible with customers, suppliers, and funding sources.

! Present a paper at the World Science Conference

As suggested by the ICSTI Executive Board and planned in the proposal, the results of this study will be presented by Dr. David Russon at the World Science Conference in July 1999.

! Develop a Statement of Concern regarding digital electronic archiving

As many survey participants mentioned, the current projects in digital electronic archiving are often being done without adequate commitment and funding. There is concern that funding will not be sustained, and is not consistent with mandates to collect and preserve electronic information. As suggested by the ICSTI Working Group, ICSTI and CENDI should produce a Statement of Concern, either jointly or consecutively, that raises the issues of electronic archiving and continued preservation and access to these archives with stakeholders, policy makers and funding sources. Many of the stakeholder groups are represented by members of ICSTI and CENDI, and therefore, it should be in a unique position to "work through" this difficult task. As the ICSTI Digital Electronic Working Group indicated in its report, the statement should not only identify the need for and benefits to be gained by electronic archiving and continuing access, but it should identify guidelines for what constitutes an electronic archive and sufficient access. It should emphasize the need to support verbal commitments to digital archiving with proper programming and funding. The Statement of Concern should also identify further activities in which ICSTI and others can participate to ensure that the statement is put into action.

! Publish an article on the results of the ICSTI/CENDI study

While the report to the World Science Conference will provide some level of visibility for the efforts of ICSTI and CENDI as well as for the next steps necessary to move digital archiving

forward, this will not reach all stakeholder audiences. It is suggested that an article be prepared from the study and published in a relevant journal. The investigators have already been approached by the editor of the *Journal of Electronic Publishing* for such an article.

! Develop a topical area on either the open ICSTI or CENDI Web site that highlights digital electronic archiving. (This could also be done as a joint effort.)

The topic of archiving was highlighted in the report from the June 1997 meeting and in a subsequent issue of the *ICSTI Forum*. Those documents, a summary of this report and other possible information gleaned from ICSTI members should be included as a special theme on the Web site. (There are many good sites that already address this issue, and there is no need to replicate them. However, links from a specific ICSTI or CENDI page to these other sites may be of value to ICSTI and CENDI members and others interested in this subject.) CENDI could consider highlighting this area as a special adjunct to the broader STI Manager part of its Web site.

This survey has emphasized that DEA issues require collaboration and coordination among a variety of stakeholders. There are numerous projects underway at many levels. The ICSTI and CENDI members can benefit from staying informed of ongoing activities. They also have experience and practical needs that can help to inform and move the state of DEA implementation forward.