

# **Preserving Legal Materials in Digital Formats**

Prepared for the Legal Information Preservation Alliance

By

Judith Cobb, OCLC Online Computer Library Center, Inc.  
Joan Allen-Hart, San Diego County Public Law Library

Submitted: 4 February 2005

# Table of Contents

Executive Summary .....	3
Section 1: Background and purpose.....	5
Section 1.1: The importance of legal information and the need for preservation efforts	5
Section 1.2: The traditional preservation problem of paper legal materials .....	7
Section 1.3: Problems and challenges presented by digital materials .....	9
Section 2: Identification of Content and Risks Associated with Its Preservation.....	14
Section 2.1: Types of legal materials and their provenance .....	14
Section 2.2: Digitization trends and how they affect preservation .....	16
Section 3: Digital Preservation Overview and Current Projects.....	19
Section 3.1: Overview .....	19
Section 3.2: Government projects .....	20
Section 3.3: Commercial projects .....	22
Section 3.4: Non-profit projects.....	24
Section 3.5: What can be learned from these projects?.....	26
Section 4: Requirements for a Successful Digital Preservation Program .....	27
Section 4.1: Program and policy requirements .....	27
Section 4.2: Selecting materials for digital preservation.....	28
Section 4.3: Operational requirements .....	30
Section 5: Roles and Responsibilities .....	33
Section 5.1: Coordinated collaboration .....	33
Section 6: Recommendations.....	35
Appendix A: Digital Preservation Resources .....	39
Glossaries.....	39
Recommended Resources .....	39
Additional Resources.....	40
Policies.....	41
Digital Preservation Projects.....	42
Appendix B: Digital Imaging Resources.....	44
General Resources .....	44
Metadata.....	44
Besser Principles .....	45
Bibliography .....	46

# Executive Summary

Legal information is at the core of a democracy. From the U.S. Constitution to local legislation and court decisions, legal information is fundamental to the principles of open government to which citizens have free access. Legal information includes primary records and information created by the government as well as secondary writings and compilations that organize, explain and evaluate those primary records. Today, legal resources come in a variety of formats – from leather bound books to microfilm to computer systems. No matter how the information is recorded, all are vital resources for our democratic system and need to be preserved in order to ensure ongoing and sustainable access.

Librarians, archivists and other information professionals have long been the caretakers of legal information, assuring these resources are maintained in such a way that the valuable information will not be lost while providing access to them. Law libraries, a main caretaker of legal materials in the U.S., are particularly concerned with their primary constituencies of law faculties, law students, practitioners, judges, and government agencies<sup>1</sup> while maintaining a strong belief “in the fundamental principle that the availability of government and legal information to all people is a necessary requirement for a just and democratic society and a valuable public good created at taxpayer expense.”<sup>2</sup>

Legal information in all formats is vital. Yet legal information recorded in digital formats is at a much higher risk of being lost than information in analog formats. Digital materials, created via e-government initiatives, by publishers who now publish to the web, by local courts who utilize court management systems, and by libraries who digitize analog resources so that the digital surrogate can be made web accessible, all create challenges for preservation.

Digital materials are at risk due to several factors including storage media obsolescence, software obsolescence, organizational and cultural challenges, and the need for sustainable access. Although both public and private sector organizations are working to manage these risks and facilitate the preservation of authentic digital materials, there are no simple solutions. As a result, legal information is at risk of being lost, particularly legal information that is not redundantly captured and maintained and for which there is no possibility of an analog surrogate.

The challenges inherent in digital preservation will require a cross community, coordinated and collaborative effort. For legal materials, the Legal Information Preservation Alliance needs to act as the primary advocate and organizer for these efforts. LIPA will need to work with many different types of organizations – public and private –

---

<sup>1</sup> AALL Preservation Policy. available at [http://www.allnet.org/about/policy\\_preservation.asp](http://www.allnet.org/about/policy_preservation.asp) accessed 14 December 2004

<sup>2</sup> AALL Supports The E-Government Act of 2001, AALL WAO News Release, May 1, 2001, available at <http://www.ll.georgetown.edu/aallwash/pr05012001.html> accessed 14 December 2004

that have different areas of interest and expertise, including records and information management, technologists, archivists, librarians, and policymakers at all levels of government to advocate a fundamental change in how legal materials are created, managed, maintained, preserved, and accessed.

# Section 1: Background and purpose

## Section 1.1: The importance of legal information and the need for preservation efforts

Legal information is at the core of a democracy. From the U.S. Constitution to local legislation and court decisions, legal information is fundamental to the principles of open government to which citizens have free access. Legal information includes primary records and information created by government and secondary writings and compilations that organize, explain and evaluate those primary records. Today, legal resources come in a variety of formats – from leather bound books to microfilm to computer systems. No matter how the information is recorded, all are vital resources for our democratic system and need to be preserved in order to ensure ongoing and sustainable access.

Librarians, archivists and other information professionals have long been the caretakers of legal information, assuring these resources are maintained in such a way that the valuable information will not be lost while providing access to them. These materials are preserved so all those who need them can use them. Law libraries, a main caretaker of legal materials in the U.S., are particularly concerned with their primary constituencies of law faculties, law students, practitioners, judges, and government agencies.<sup>3</sup> However, the American Association of Law Libraries (AALL) puts forth a strong belief “in the fundamental principle that the availability of government and legal information to all people is a necessary requirement for a just and democratic society and a valuable public good created at taxpayer expense.”<sup>4</sup>

Legal materials have a broad audience and serve a fundamental purpose that must be remembered:

*“The American public, as taxpayers, has a right to access government information regardless of format, to locate it easily and conveniently, and to know that valuable government information accessible today in electronic formats will be preserved and remain available for continuous, permanent public access.”<sup>5</sup>*

This statement speaks to the very purpose of this white paper. Legal information is vital to many types of users and to a democratic society. It must be preserved. The potential loss of legal information, particularly primary legal materials and their supporting documents, threaten the very foundations of our democracy. This threat has become increasingly apparent as official entities have transitioned to providing government

---

<sup>3</sup> AALL Preservation Policy. available at [http://www.allnet.org/about/policy\\_preservation.asp](http://www.allnet.org/about/policy_preservation.asp) accessed 14 December 2004

<sup>4</sup> AALL Supports The E-Government Act of 2001, AALL WAO News Release, May 1, 2001, available at <http://www.ll.georgetown.edu/aallwash/pr05012001.html> accessed 14 December 2004

<sup>5</sup> AAAL Government Relations Committee and Washington Affairs Office, State-By-State Report on Permanent Public Access to Electronic Government Information, Page 2, June 2003, available at <http://www.ll.georgetown.edu/aallwash/> accessed on 14 December 2004

services and information through electronic means and via the Internet. During the 1990's, the federal government and most state governments passed a variety of E-Government acts, which mandated the provision of online services to their citizens.

The transition to providing electronic information made a vast amount of primary documents, which had previously only been available to citizens through the libraries and archives, instantly and widely available. Statutes, regulations, administrative agency forms, policy and procedure manuals, as well as court decisions and related materials (court rules and forms, for example) are now available on government websites. The positive effects of the move to digital workflows and access are obvious:

*“The free flow of government information via the Internet facilitates transparency and accountability in government. It increases accessibility of government at all levels and encourages citizen participation in our democracy.”*<sup>6</sup>

While the beneficial aspects of providing government information electronically were immediately obvious, after a few years, the less positive aspects of this transition became apparent, as many government websites were not preserved. Information that was available one day often “disappears” the next, depending on the changing political winds or information technology policy.

In addition to e-government initiatives, the non-profit and commercial sectors are expending significant resources to digitize and provide electronic access to materials via the web. And, publishers of legal information are publishing directly to the web. Thus, the universe of legal information is vast, easier to access, and widely available to researchers. Yet, as we will see throughout this paper, digital materials in any format, are more volatile than their analog counterparts and measures to collect, preserve and provide ongoing access to them are in their infancy. The problems with digital preservation are widely acknowledged in the information professions, yet even after years of research, the solutions are less than clear.

In response to and recognition of this growing problem, the AALL 2000-2005 Strategic Plan mandated the development of a "national plan for the preservation of legal materials in all formats." They identified several components of a national plan, which include: 1) mission and purpose, 2) content, 3) infrastructure, 4) standards and best practices, 5) partnerships and collaborations, and 6) sustainability.<sup>7</sup>

In March 2003, “A group of more than 40 law librarians and other experts gathered at the Georgetown Law Center for the conference, ‘Preserving Legal Information for the 21<sup>st</sup> Century,’ to take the first steps toward developing a national agenda for preserving legal

---

<sup>6</sup> Ibid

<sup>7</sup> Patricia K. Turpening, “AALL Takes First Steps to Develop a National Preservation Plan” Abbey Newsletter Volume 26, Number 6, Oct 2003 available at <http://palimpsest.stanford.edu/byorg/abbey/an/an26/an26-6/an26-605.html> accessed on 14 December 2004.

information in all formats.”<sup>8</sup> Robert Oakley, the convener of this group, the Legal Information Preservation Alliance (LIPA), stated, “Such information is the foundation of our society, and we are its custodians. Preservation of the information deserves nothing less than our full commitment.”<sup>9</sup>

LIPA’s mission is as follows:

*“The mission of the Legal Information Preservation Alliance [LIPA] is to provide the leadership, the necessary organizational framework, and the professional commitment necessary to preserve vital paper and electronic legal information by defining objectives, developing and/or adopting appropriate standards and models, creating networks, and fostering financial and political support for long term stability.”<sup>10</sup>*

The importance of both preservation of and access to legal materials is well established. Today, legal materials take the form of both analog and digital formats. A comprehensive preservation program should be established for both. Although the preservation of digital formats is the focus of this paper, we will briefly look at the preservation problems for analog legal materials in order to establish the context for digital preservation efforts.

## **Section 1.2: The traditional preservation problem of paper legal materials**

As we plan for the future of digital preservation, we need to make sure we understand and learn from our past preservation experience. Providing an overview of all the preservation issues, projects and programs that have addressed traditional preservation is beyond the scope of this document. Instead, we will look at two particular problems that provide insight into the digital preservation effort – brittle books and audiovisual materials.

### **Brittle Books**

The brittle book problem is more accurately stated as the brittle paper problem – as not only books are affected by it, but many paper based materials. Brittle books, as they are often referred to, are books in which the paper is actually disintegrating and crumbling. The main cause of the problem is a high level of acid found in paper manufactured between 1840 and 1950 – this high acid level created the chemical reactions in the paper which caused it to become brittle and in some cases, actually disintegrate when handled.

---

<sup>8</sup> Margaret Maes Axtmann, “Preservation – The Time is Right,” AALL Spectrum Volume 7 Number 9, June 2003

<sup>9</sup> Robert Oakley, quoted in Paul D. Healey, “Georgetown Conference Outlines Preservation Agenda,” AALL Spectrum, Volume 7 Number 9 June 2003

<sup>10</sup> Legal Information Preservation Alliance, Conference Report: Preserving Legal Information for the 21<sup>st</sup> Century: Toward a National Agenda, March 6-8, 2003. May 2003 available at [http://www.aallnet.org/committee/lipa/LIPA\\_Conference\\_Report.pdf](http://www.aallnet.org/committee/lipa/LIPA_Conference_Report.pdf) accessed on 14 December 2004

This was the first large-scale preservation crisis that libraries had faced. Until this crisis, preservation efforts were made on an item-by-item basis. It was clear that a mass preservation effort was necessary, and that steps needed to be taken to ensure that the culprit, acidic paper, was controlled in library collections.

By the 1960s, it was possible to manufacture alkaline (non acidic) paper cost effectively. This development was followed by the development of an ANSI standard for permanent paper. These advances can be demonstrated in the current AALL preservation policy, which supports development and revision of standards related to traditional preservation, and indicates that AALL will communicate the needs of members to publishers and information vendors regarding the need for “permanent, durable” materials.

Responding to the millions of books that were deteriorating, however, was a more difficult problem. This was indeed a preservation crisis, a true threat to the destruction of vital information in all disciplines – information that quite possibly could not be re-created. Further, studies undertaken to quantify the problem and forecast what actions could be taken in an appropriate time frame showed that collaboration was vital to ensure efforts were not duplicated and save as much as possible.

Librarians were faced with the knowledge that they could not save everything. They also faced a situation that required a national program implemented by local efforts. The crisis required a coordinated and non-duplicative effort to ensure that as much as possible was rescued. It also required technical solutions – mass de-acidification and reformatting. Mass de-acidification did not strengthen or restore the existing damage; it simply removed the acid from the paper. Reformatting sometimes resulted in the loss of the original media, but the preservation of the content.

The challenge with digital materials is analogous. The communities involved must create a national strategy that encourages appropriate selection of digital materials for preservation, non-duplicative efforts, and develops appropriate technological solutions. Further, we must acknowledge that we will not be able to rescue everything, and that some resources have been and will continue to be lost. And finally, we must consider the technical options for preservation. Although these solutions are not clear, it does seem likely that the two general models will remain the same in the digital environment -- stabilizing digital materials and maintaining them in their current format, or reformatting them to a more stable format.

Recorded information is as fragile as the medium on which it is stored; it will be preserved only with intervention. There are two things we know definitively: not all information will survive and we cannot predict accurately what information will be in demand in the future.<sup>11</sup>

---

<sup>11</sup> Abby Smith, *The Future of the Past: Preservation in American Research Libraries*. CLIR: Washington D.C. April 1999 available at <http://www.clir.org/pubs/reports/pub82/pub82text.html> accessed on 14 December 2004



We also know that information stored on media that requires equipment to access the content is significantly more fragile than information stored on “eye readable” media. Recording media that carries non-textual information and requires an intermediary machine has not fared well over time. 80% of motion pictures made before 1940 have perished because no one saved them or because of physical deterioration beyond recovery.<sup>12</sup> We need time to understand how to preserve these kinds of machine-readable resources, yet in this instance, we ran out of time. We need organized collecting policies to ensure that we have the materials that need to be preserved. We need funding and expertise to take preservation actions. All of these elements were lacking in the recording media realm, and we have lost irreplaceable materials. However, the community needs to learn from this mistake. It must not be repeated.

### **Section 1.3: Problems and challenges presented by digital materials**

By looking at the responses of the community to the brittle books and recording media problems, we begin to see that digital materials present us with very similar problems. However, the volume of digital materials and the much higher rate of technological change exacerbates the problems and creates a much higher risk that materials are being lost.

#### **Background and Definitions**

Digital materials can be classified into several groups. These include:

**Data** – “Representation of facts, concepts, or instructions in a formal manner suitable for communication, interpretation or processing by manual or automated means. Basic units of information. The representation forms of information. Examples of data include a sequence of bits, a table of numbers, the characters on a page.<sup>13</sup> Digital materials are made of data.

**Digitally born** – “Digital materials which are not intended to have an analog equivalent, either as the originating source or as a result of conversion to analog form.” Born digital materials are differentiated from other materials such as 1) digital materials which have been created as a result of converting analog originals; and 2) digital materials, which may have originated from a digital source but have been printed to paper, e.g. some electronic records.<sup>14</sup>

**Digitization** -- The process of creating digital files by scanning or otherwise converting analog materials. The resulting digital copy, or digital surrogate,

---

<sup>12</sup> Ibid

<sup>13</sup> NEDLIB Glossary available at <http://www.kb.nl/coop/nedlib/glossary.pdf> accessed on 14 December 2004

<sup>14</sup> Maggie Jones and Neil Beagrie, Preservation Management of Digital Materials: A Handbook available at <http://www.dpconline.org/text/intro/definitions.html> accessed on 14 December 2004

would then be classed as digital material and then subject to the same broad challenges involved in preserving access to it, as "born digital" materials.<sup>15</sup>

The current trend toward reformatting analog materials to digital (digitization) will result in many digital resources that may need to be preserved. When we discuss digital preservation, the same issues apply whether materials are born digital or digitally reformatted. The main difference is that in many cases, there will be an analog back-up for digitally reformatted materials, whereas there may not be for born digital materials. Thus, there is significantly less risk of loss of content (although certainly loss of access is a concern) with digitally reformatted materials. The preservation of digital images will be handled best if images are created according to existing standards and best practices. It is beyond the scope of this report to consider these standards, however we do recommend that best practices for the digitization of legal materials are developed and adopted throughout the community in order to facilitate the creation of high quality images that are best suited to digital preservation if that is warranted. These best practices should include scanning and digitization specifications, metadata creation and management best practices, guidance for selecting materials to be digitized, and planning for digital preservation, if that is deemed necessary for digital surrogates. As with digital preservation, a programmatic and coordinated approach to digitization will result in the most satisfactory results for the community.

**Electronic information** -- Digital data that has been assembled to enable someone to make a decision, to distinguish between alternatives, or to narrow a range of possibilities.<sup>16</sup>

**Electronic Records** -- Records created digitally in the day-to-day business of the organization and assigned formal status by the organization. They may include word processing documents, emails, databases, or web pages.<sup>17</sup>

Digital preservation refers to “the managed activities necessary for ensuring both the long-term maintenance of a bit stream and continued accessibility of its contents.”<sup>18</sup> Digital preservation should also ensure that materials remain authentic – the material is what it purports to be. In the case of electronic records, it refers to the trustworthiness of the electronic record as a record. In the case of "born digital" and digitized materials, it refers to the fact that whatever is being cited is the same as it was when it was first created unless the accompanying metadata indicates any changes. Confidence in the

---

<sup>15</sup> Ibid

<sup>16</sup> NEDLIB Glossary available at <http://www.kb.nl/coop/nedlib/glossary.pdf> accessed on 14 December 2004

<sup>17</sup> Maggie Jones and Neil Beagrie, Preservation Management of Digital Materials: A Handbook available at <http://www.dpconline.org/text/intro/definitions.html> accessed on 14 December 2004

<sup>18</sup> OCLC and RLG, Trusted Digital Repositories Report available at <http://www.rlg.org/longterm/repositories.pdf> accessed on 14 December 2004

authenticity of digital materials over time is particularly crucial owing to the ease with which alterations can be made.<sup>19</sup>

### **The Challenges of Digital Materials**

Digital materials, created via e-government initiatives, by publishers who now publish to the web, by local courts who utilize court management systems, and by libraries and archives that digitize analog resources so that the digital surrogate can be made web accessible, all create challenges for preservation and sustainable access. The challenges involve three primary areas: 1) lack of resources, 2) the need for digital preservation to become integrated into existing organizational cultures, and 3) technological needs and questions. None of these challenges is insurmountable.

Digital materials are at risk due to several factors. These include:

**Storage Media Obsolescence:** Storage media for digital materials changes fairly quickly. Over the last 10 years we have seen several generations of storage, including 3 ½ inch diskettes, 5 ¼ inch diskettes, CD, DVD, magnetic tape, and memory sticks. Storing digital materials on media requires an ongoing commitment to moving the data from one storage medium to another. This is called refreshing the data, and it can be costly and time consuming for large quantities of data.

**Software Obsolescence:** Like storage hardware, software programs change often as we upgrade to new versions of programs. File formats change, and new software programs may not be compatible with older files. Proprietary formats are characterized by full documentation not being available consistently; license and patent rules; license agreements that are subject to change; and restrictions for use and modifications may apply. Open or non-proprietary formats have no license fees, no patent owners, full documentation available, and are open for self-made modifications. (The distinction between proprietary and non-proprietary/open can also be made for computer systems.) Open formats and systems are preferable for preservation purposes, but are generally used less frequently.

**Organizational and Cultural Challenges:** Essentially, digital materials are at risk because of the high rate of technological change. However, contributing to this risk is a general lack of management of digital materials by creators and other caretakers. If one looks at the literature in the records management, archival, and information management fields, it is clear that the creators and caretakers of these materials are not managing these resources like they once managed paper-based resources. Materials are placed on the web for publication, and removed and deleted. Computer systems generate electronic records without incorporating recordkeeping and records management. Publishers make content available via the web, but cannot assure that their materials will be available in the long term.

---

<sup>19</sup> Maggie Jones and Neil Beagrie, Preservation Management of Digital Materials: A Handbook available at <http://www.dpconline.org/text/intro/definitions.html> accessed on 14 December 2004

These substantive, organizational problems must be addressed. Digital preservation is not solely a technical problem. The technical problems are made significantly worse by a lack of acknowledgement that information is a resource which must be managed, funded and maintained appropriately.

**Access:** With the advent of digital technology, and the World Wide Web in particular, access has become the driver for information. What is being forgotten is that there will be no access through time without preservation.

*“Access is assumed to mean continued, ongoing usability of a digital resource, retaining all qualities of authenticity, accuracy and functionality deemed to be essential for the purposes the digital material was created and/or acquired for.”<sup>20</sup>*

Even in the library and archival professions, most organizations are digitizing materials for access, while not addressing issues of preservation. Certainly access and preservation must work together in order for us to ensure the longevity of our vital legal information. There are a myriad of examples of digitization to facilitate access. In California, cases from 1850 to the present, became accessible online to the public at no charge. LexisNexis, the publisher of the official California reporters, provides a link from the court website, [www.courtinfo.ca.gov](http://www.courtinfo.ca.gov), to its website. Online researchers can search and retrieve California case law by parties' names, citation, key word or natural language at no charge.<sup>21</sup> Previously, Findlaw had made California cases from the 1930s to the present available to users who had registered through its free “My Findlaw” registrations. Although access to these materials is greatly increased, no digital preservation program is in place.

Further, particularly for legal materials, we must acknowledge that unmediated access to materials is not always the best choice for our users. Legal materials are notoriously complex and users may require the services of a reference specialist in order to meet their information needs.

*“A patron who is attempting to do extensive case law research, even with a fairly clear set of facts, should probably be referred to the closest law library open to the public. This is because searching for case law on the Internet can be frustrating, as many of the reporter series containing appellate decisions are copyrighted publications which are not available for free on-line. There are some exceptions, however, such as decisions of the United States Supreme Court which can be found on a number of web sites such as Cornell’s Legal Information Institute ([www.law.cornell.edu](http://www.law.cornell.edu)). In addition, many of the court web sites are publishing their own recent opinions, but they do not keep them on-line indefinitely. The California Courts web site, for example, only retains the opinions of the state Supreme Court and the District Courts of Appeals for 100*

---

<sup>20</sup> Maggie Jones and Neil Beagrie, Preservation Management of Digital Materials: A Handbook available at <http://www.dpconline.org/text/intro/definitions.html> accessed on 14 December 2004

<sup>21</sup> This appears to be a fairly unique arrangement between a state court and a commercial vendor at this time.

*days ([www.courtinfo.ca.gov](http://www.courtinfo.ca.gov)). In addition, many of the court web sites do not offer search engines which allow for boolean searching. Attempting to locate relevant case law on-line (unless you have access to LexisNexis or Westlaw) can be like searching for the proverbial needle in the haystack.”<sup>22</sup>*

Digital preservation must take the access requirements and needs of the users into consideration. Just as access without preservation is not an adequate solution, preservation without considering user needs and access issues is not complete. Access and user considerations are one of the least researched areas of digital preservation and much more work is required in order to ensure that preservation and access are balanced appropriately.

---

<sup>22</sup> Locating the Law: A Handbook for Non-Law Librarians. Southern California Association of Law Libraries available at <http://www.aallnet.org/chapter/scall/locating.htm> accessed on 14 December 2004

## Section 2: Identification of Content and Risks Associated with Its Preservation

American law comes from a variety of sources including the federal government, fifty state governments, the District of Columbia, and local government and courts. Each of these sources is structured similarly, with legislative, executive, and judicial components that each “make law” and create legal information. Thus legal information comes from a variety of sources, and researchers must understand the types of information created by each branch. Legal information is also found in the form of published manifestations of the law – books, electronic databases, microforms, optical disks, and other media all contain legal information.<sup>23</sup>

### Section 2.1: Types of legal materials and their provenance

As we stated in the previous section, legal information is complex. There are 51 different legal systems in operation in the United States - the federal system and each of the state systems. In a specific case, federal law alone may apply, state law only may be relevant, or there may be a mixture of state and federal issues.

Generally, the federal courts are responsible for applying and interpreting federal statutes and the U.S. Constitution. The state courts apply and interpret the state’s constitution and state statutes. There are also county and municipal ordinances as well as state and/or federal administrative regulations and rulings that may apply to a legal issue.<sup>24</sup>

#### Primary Sources

The courts and government entities that create these legal materials create primary sources of legal information. Some primary sources may be published and some may not. Published primary sources include “official sources” which are published directly by or authorized for publication by the creating entity. They also may include “unofficial sources” which are published by commercial vendors and not authorized by the creating entity.

Primary sources are created at all levels of government and include:

**Statutes and Codes:** The current laws of a jurisdiction that are organized by subjects (codes). The statutory codes are published in both official and non-official versions in many jurisdictions. Unofficial codes are published by commercial publishers and are usually *annotated* (editorial enhancements which include references to other primary materials, such as court decisions which have applied or interpreted the statute, as well secondary sources which explain the application of the statute.)

---

<sup>23</sup> Roy M. Mersky and Donald J. Dunn, *Fundamentals of Legal Research*, 8th ed. 2002. Steven M. Barkan, author. “An Introduction to Legal Research,” Chapter 1, p.1.

<sup>24</sup> *Ibid*

*Federal:* At the federal level, statutes and codes include the U.S. Constitution, U.S. Congress legislative information, bills, session laws, statutes, codes. They also include all documents produced during the legislative process, such as committee reports, and transcripts of committee hearings.

*State:* At the state level, statutes and codes include state constitutions, state legislative information such as bills, session laws, statutes, and codes. These also include all documents produced during the legislative process, such as committee reports and transcripts of committee hearings.

*Local:* Locally, statutes and codes may take the form of local charters, ordinances, municipal codes, and administrative regulations.

**Court decisions:** The written opinion of a judge or of a panel of judges is also referred to as a case. A case usually includes a brief description of the factual background of the situation and a short history of the legal procedures that brought the dispute before the court now writing the opinion. The cases that are published, or reported, are usually the opinions of judges at the appellate court level. Judicial decisions are published only if they change or clarify a rule of law. Judicial law is published in both official and unofficial versions. The difference between the official and unofficial version lies in the editorial enhancements. The text of an opinion is the same in both versions. Special topic headings and numbers may also be assigned by the publisher to the various points of law discussed in the opinion. These topic headings and numbers can then be used to find similar cases in the case finding tools called digests.

Legal researchers today expect to access all or most of the documents prepared during the litigation process. Federal judicial policy recognizes this expectation as legitimate in carrying out its case management process:

*“Case files are maintained by the clerk of court as the official record of litigation in the federal courts. As a general rule the public case file consists of all pleadings, orders, notices, exhibits, and transcripts filed with the clerk of court. It is standard practice that case files are open for inspection and copying during normal business hours. There is also a general presumption that court files are available to anyone upon request. Courts do not make access determinations based on the status of the requestor. The federal judiciary also offers various electronic public access (EPA) services that allow the public to gain quick access to official court information and records from outside the courthouse.”<sup>25</sup>*

*Federal:* At the federal level, these materials include decisions of the U.S. Supreme Court and of the lower federal courts (circuit courts of appeals and district courts), and the case documents filed, such as pleadings, motions, briefs,

---

<sup>25</sup> *“Privacy and Access to Electronic Case Files in the Federal Courts,”* Office of Judges Programs of the Administrative Office of the United States Courts, December 15, 1999, available at <http://www.privacy.uscourts.gov> accessed on 14 December 2004

which are increasingly available online through PACER (Public Access to Court Electronic Records), administrated through the Administrative Office of the Courts, see <http://pacer.psc.uscourts.gov>.

*State:* Court documents at the state level include the states' supreme court decisions as well as documents filed with the courts during the appellate process.

*Local:* At the local level these materials are made up of local court decisions and records.

**Regulations:** Administrative law is created when a state or federal agency issues regulations, or when an administrative official decides a dispute in the subject area for which the agency is responsible. Administrative decisions, or cases, are usually published in special loose-leaf services whose coverage is limited to a particular subject, such as antitrust or labor law. These are generally issued by commercial publishers. However, many federal and state agencies are now putting their current decisions, procedures and rules on their web sites.

*Federal:* Federal executive branch documents include the regulations and rulings of federal administrative agencies, and the orders of the executive.

*State:* These materials include the regulations and rulings of state administrative agencies, other public documents and records, and orders of the states' governors.

*Local:* At the local level, these may include regulations and rulings of local administrative agencies.

**Government documents:** Published government information, a subset of which includes legal materials.

Government documents are published at all levels of government. Increasingly documents are published via the web and consequently are not acquired by a depository program for preservation and access.

### **Secondary Sources of Legal Information**

Secondary sources usually explain and/or describe the law, discuss a legal problem, or set out a model piece of legislation. Examples include treatises, legal encyclopedias, practice guides, law reviews and other legal periodical articles, and the reports of law reform commissions or other organizations such as the American Bar Association.

## **Section 2.2: Digitization trends and how they affect preservation**

Now that we have defined what legal information is and where it comes from, we will look at some online resources for legal information. Although digitization does not necessary equal preservation, digitization does require a significant investment and may



imply that the material being digitized is considered important enough to warrant that expenditure of resources. Some examples of the digitization of legal materials include:

- HeinOnline-- The database includes full-text access to the complete contents of indexed legal journals. Multiple search capabilities, including an official citation search option, link to exact page images of journal issues. HeinOnline has four online library collections: an image-based Law Journal Library collection, which contains nearly 370 of the top academic legal periodicals, the Federal Register Library, which is also image-based and contains pre-1981 coverage of the United States Federal Register, the Treaties and Agreements Library and the Supreme Court Library.
- The ongoing Maryland State Archives<sup>26</sup> publication series, *Archives of Maryland Online*, currently provides access to over 471,000 historical documents that form the constitutional, legal, legislative, judicial, and administrative basis of Maryland's government. Legal materials included in this project consist of Codes, Compilations of Laws, Rules and Regulations, Constitutional Records, judicial records, legislative records, probate records, and session laws.
- CRS/West is engaged in a collaborative project in California to digitize the appellate briefs which are filed with the state's courts of appeals. CRS receives the paper copies of the briefs from one of the brief depository law libraries, they are microfilmed and digitized and made available on the CRS/West website. Issue of concern: previously library users could use the paper format at no charge in a depository library. Now there is a charge to download and access the briefs online.
- Law Library Microform Consortium Digitization Project (LLMC-Digital) is specifically access oriented. As of Dec. 2003, LLMC's current back file of filmed materials, as described on their web site, numbered approximately 50-million page images. It is anticipated that additional data capture over the next decade will be roughly 50-million page images. All of this material will gradually be made available via *LLMC-Digital*, which by 2013 is expected to be serving some 100-million page images.<sup>27</sup>
- The Government Printing Office is considering the digitization of government documents in order to increase their accessibility. Many government documents are vital to the legal community, and this effort to increase access will be important to watch as it progresses.
- Another initiative to track is the digitization work being tested by the commercial source Google. Google is partnering with several libraries, including Harvard, in order to digitize and make available library materials via their web based search

---

<sup>26</sup> Maryland State Archives Web Site available at <http://www.mdarchives.state.md.us> accessed on 20 December 2004.

<sup>27</sup> *LLMC Offers Digital Access While Remaining Committed to Film for Preservation and Local-Backup Purposes* available at [http://www.llmc.com/general\\_description.htm](http://www.llmc.com/general_description.htm) accessed on 14 December 2004

service. The project is not directly addressing digital preservation, but like all digitization initiatives, ongoing access will require preservation.

The increasing amount of digitization being done today will raise an interesting question in the future. Do we spend time and money preserving digital images that have an analog back up; or do we use our resources for preservation of born digital information that has no analog back up? Because users of legal information require ongoing access to both current and non-current information, the plethora of digitized resources will need to be sustained for the long term in order to meet access needs. In other communities, digitized resources might not be candidates for digital preservation, particularly as their access wanes; however, legal resources require ongoing access. Further, in the digital environment, we may find that access actually drives preservation. This is antithetical to the analog environment, where increased access inhibits preservation, yet it may be the case that an increasing and ongoing need for digital access to materials will actually be the driver for the preservation of that material. We will look at this question more in Section 4 in an effort to assess how much risk legal information is exposed to.

## Section 3: Digital Preservation Overview and Current Projects

Determining how to preserve digital materials continues to be a challenge for the larger library and archival community. Although the development of standards and strategies is underway, all are in their infancy and require significant testing, improvement and proof of concept before they are truly reliable. Moreover, digital preservation has yet to truly be incorporated into most library and archival organizations in terms of staffing, funding, education, and infrastructure.

The goal of digital preservation is to sustain the digital resource through time without losing significant informational and contextual properties. Preservation is the foremost concern, but access and usability must also be ensured. Preservation may be achieved by converting all digital resources to a “canonical format” which is considered the preservation format – such as TIFF, PDF, or XML. Or, it may be achieved by performing preservation actions on the original files or by emulating the original environment in which the digital objects were accessible. No matter what the technical preservation strategy (or strategies) utilized, it is desirable that the long-term preservation and use of digital materials will allow us to use digital data in new ways: to consolidate reports, to see long-term trends, to mine for metadata that is inherent within a document. Digital information is more user-friendly in our current environment and allows us to meet the needs of our users more efficiently than ever before.

*“In the broadest sense, technology extends our abilities to change the world: to cut, shape, or put together materials; to move things from one place to another; to reach farther with our hands, voices, and senses. We use technology to try to change the world to suit us better.... But the results of changing the world are often complicated and unpredictable. They can include unexpected benefits, unexpected costs, and unexpected risks—any of which may fall on different social groups at different times. Anticipating the effects of technology is therefore as important as advancing its capabilities.”<sup>28</sup>*

### Section 3.1: Overview

We have already looked at the challenges of digital preservation. This section will discuss the current state of affairs within the digital preservation communities, particularly in the U.S. It is important to remember that digital preservation work is in its infancy and there is much work to be done. We will focus on implementation oriented projects, most are specifically oriented to digital preservation, yet some are access oriented. These access oriented programs may be potential candidates for digital preservation work in the future.

---

<sup>28</sup> American Association for the Advancement of Science. *Science for All Americans Online: Chapter 3 The Nature of Technology* available at <http://www.project2061.org/tools/sfaol/chap3.htm> accessed on 14 December 2004

Before discussing implementation, we must be aware of a core document – the Open Archives Information System Reference Model (OAIS).

The OAIS is not a programmatic model, but a reference model that outlines the required functions of an archive. It was developed within the scientific community, but has great utility within the library and archival communities. It outlines the core functions of an archive in some detail, and has been used to build digital archive implementations. These functions include acquisition, management, administration, access, and preservation planning. The OAIS is an ISO standard and compliance with it is viewed as vital to a functioning trusted digital repository. See Appendix A for a list of references for more information on the OAIS.

Digital preservation requires three core components: resources, policy infrastructure, and technical infrastructure. Resources are necessary to ensure that a digital preservation program is stable and long-lasting. A policy infrastructure will include integration of digital preservation into existing policy and planning documentation; for example, integrating digital preservation into strategic planning, collection policies, access policies, security policies, disaster prevention and recovery policies and plans, information technology planning, acquisitions documentation and workflows, preservation planning, and legal considerations, particularly rights management and privacy issues. Technically, digital preservation requires hardware, software, staffing, and the implementation or development of new tools to facilitate both access and preservation to digital resources in a sustainable manner.

### **Section 3.2: Government projects**

**National Archives and Records Administration/San Diego Supercomputer Center:** The work of the National Archives and Records Administration (NARA) and the San Diego Supercomputer Center (SDSC) has included XML self-describing documents; storage resource broker technology; and data grid technology. Since legal information is generally text-based information, the use of XML to create self-describing documents should be considered as an option for long-term preservation and access. Additionally, the SDSC currently has data grid technology available and in use at several organizations. The technology appears to be able to support large amounts of data with a relatively small investment.

**National Archives and Records Administration, Electronic Records Archive:** NARA has also recently released a request for proposal and requirements documents for their Electronic Records Archive (ERA) project. The ERA development will bear watching, as it will be the first large-scale implementation of a digital archive for government information in the U.S. At the federal level, the NARA ERA may have a significant impact upon the preservation of federal legal information and, if implemented for legal information, should provide long-term sustainability for those resources. NARA has awarded Lockheed Martin and Harris Corporation one-year design contracts. Both

companies will submit their designs to NARA and one will be awarded the contract to build the Electronic Records Archive solution.

**Library of Congress, National Digital Information Infrastructure and Preservation**

**Program:** “In December 2000, Congress appropriated \$100 million (rescinded to \$99.8 million) for a national digital-strategy effort, to be led by the Library of Congress. The Library will work closely with federal partners to assess considerations for shared responsibilities. Federal legislation calls for the Library to work jointly with several federal agencies and the private sector. The program will seek to provide a national focus on important policy, standards and technical components necessary to preserve digital content. Investments in modeling and testing various options and technical solutions will take place over several years, resulting in recommendations to the U.S. Congress about the most viable and sustainable options for long-term preservation.”<sup>29</sup> The work that takes place in the various projects funded through NDIIPP will need to be monitored.

**Government Printing Office:** The Government Printing Office (GPO) is pursuing a number of pilot projects and partnerships designed to maximize the results of their efforts to update and modernize the Federal Depository Library Program (FDLP). FDLP is fundamentally a partnership program based on shared responsibility that is coordinated and led by the Superintendent of Documents.

A key concept inherent within the GPO model is that of distinguishing between “*permanent public access*” and “*permanent preservation*” of the “record” copy. The federal information management infrastructure is segmented between GPO and the National Archives and Records Administration (NARA). The GPO is responsible for printing, distribution, and access to publications, and NARA is responsible for the selection and preservation of official agency records, including the subset of records that includes publications. Under the current model, NARA will be the caretaker of digital information that needs to be preserved. However, GPO does have initiatives underway that promote preservation, including projects to capture and preserve federal web sites and the development of an analog collection of last resort – analog copies of digital resources that are held as a failsafe.

This distinction between preservation and access is one the LIPA may need to address and will be discussed further in later sections of this document.

**Washington State:** The Washington State Digital Archives is the nation's first archive dedicated specifically to the preservation of electronic records from both State and Local agencies that have permanent legal, fiscal or historical value.<sup>30</sup> The archive’s storage system is expandable to 800 terabytes and was developed by Microsoft and EDS.<sup>31</sup>

---

<sup>29</sup> NDIIPP web site About the Program available at <http://www.digitalpreservation.gov/index.php?nav=1> accessed on 14 December 2004

<sup>30</sup> Washington State Digital Archives Grand Opening, Washington Secretary of State Web site. Available at <http://www.digitalarchives.wa.gov/Content.aspx?txt=news> accessed on 14 December 2004

<sup>31</sup> Washington Opens Digital Archive. Forbes.com October 4, 2004 available at <http://www.forbes.com/business/feeds/ap/2004/10/04/ap1575094.html> accessed on 14 December 2004

Washington is the only state with an operational digital archive. It appears that legal information is not specifically a priority for the program. As these state programs develop, an advocate for legal materials should be present.

**Public Access to Court Electronic Records (PACER):** Although not a true preservation project, the PACER system<sup>32</sup> is now accessible to the public via the Internet at a very nominal search fee. PACER provides links to all the federal appellate, district and bankruptcy courts. This system originated in 1996 when the Administrative Office of the U.S. Courts began development of the Case Management/Electronic Case Filing (CM/ECF) system which allows courts to maintain electronic case files and offer electronic filing of all documents in a federal case. The goal is to make all case information immediately available electronically through the Internet. Essentially PACER is an electronic recordkeeping system and should be a high priority for digital preservation.<sup>33</sup>

### **Section 3.3: Commercial projects**

**IBM Universal Virtual Computer:** The UVC is an intermediate language of relatively simple design that promotes sustainability. Over time, the UVC may have to be migrated, and thus is probably not the ultimate digital preservation solution – although it's not evident that there is an ultimate solution. It is being implemented by the National Library of the Netherlands which plans to use it as their main preservation method. It is not known if the implementation will adequately allow for the Library to change to another vendor and this reliance on a specific vendor is questionable.<sup>34</sup>

**Adobe:** Adobe has been working for several years, in conjunction with information professionals, on the creation and implementation of pdf-a. Using its Extensible Metadata Platform (XMP), pdf-a is envisioned to be a standard that is designed for long term preservation. It is a file format that incorporates appropriate metadata and should promote long term sustainability. A file format is not a solution, but it is a component of a total solution that could be considered.<sup>35</sup> Since pdf files are so highly used for publications, the development of this preservation-friendly file format is a significant step forward.

---

<sup>32</sup> See <http://pacer.psc.uscourts.gov> accessed on 14 December 2004

<sup>33</sup> As previously stated, legal researchers now expect to have access to both current and historical case file information. Court records retention acts which have existed to protect public record information in the court do not really address the long-term preservation needs for historical research, thus leading to the destruction of case files after the required statutory period. Appellate briefs filed in U.S. Supreme Court cases and the federal appellate courts have been digitized and are available on the web sites of the two major legal vendors, LexisNexis and Westlaw, but these are not true preservation projects. Retention of state court files varies widely from one state to another.

<sup>34</sup> IBM/KB Long Term Preservation Study. Erpanet Topic of the Month, March 2004 available at <http://www.erpanet.org/topic/digitalrepositories/show.php?id=1073637964&t=3> accessed on 14 December 2004

<sup>35</sup> William G. LeFurgy, PDF/A: Developing a File Format for Long-Term Preservation available at [http://www.rlg.org/preserv/diginews/v7\\_n6\\_feature1.html](http://www.rlg.org/preserv/diginews/v7_n6_feature1.html) accessed on 14 December 2004

**HeinOnline:** In 2000, the William S Hein & Co., a respected legal publisher for more than 80 years, launched its online product, HeinOnline. HeinOnline is built on Dienst Technology, which provides image-based data, fully searchable, allowing the online researcher to view all pages exactly as they originally appeared in the hardcopy volume. According to its November 2004 title directory it now offers more than 13.6 million pages of legal literature. There are plans to soon release a Legal Classics Library and a Presidential Documents Library. The goal is to ultimately make at least 80 million pages of legal materials available in electronic format to the research community. Although not a digital preservation project, this project could provide opportunities for future collaborative work.<sup>36</sup>

---

<sup>36</sup> <http://heinonline.org> accessed on 14 December 2004

### Section 3.4: Non-profit projects

**Lots of Copies Keeps Stuff Safe:** LOCKSS is a Stanford-based project, originally designed to create caches, or copies, of digitally published academic journals on local servers and to automatically add new content. The LOCKSS staff has also worked on the viability of implementing LOCKSS for the Federal Depository Library Program. Their 2003 report states, “It is unclear whether the LOCKSS journal model – in which a single plug-in is associated with an individual journal publisher platform – could directly be applied to Federal agency content.”<sup>37</sup> LOCKSS announced in late November 2004 that they have invented a way to serve old web content to users via newer browsers that is transparent to the user. This is LOCKSS first preservation action.

Although not specific LOCKSS projects, this concept of multiple copies reducing the risk of information loss is being utilized by other organizations. Rutgers and Emory University libraries are sharing legal information while also providing redundant storage for this joint information. Cornell hosts mirror sites for the web sites of the International Labour Organization and the International Court of Justice. The mirror sites are updated every six months and are freely available to the public. This concept of redundancy is an important one that will have an impact on assessing the risk level of legal materials later in Section 4.

**DSpace:** DSpace was created to house research and publications for faculty and researchers at the Massachusetts Institute of Technology.<sup>38</sup> It is open source software built upon free tools such as Linux, Apache Web server and the Tomcat Servlet engine, and the PostgreSQL relational database system. DSpace offers the advantages of digital distribution and support for long-term preservation for a variety of formats, including text, audio, video, images, datasets, and more. It offers the opportunity to provide access to electronic information from multiple sources through one Web interface. Kansas ran a brief pilot test of DSpace, titled KSpace, to gauge its utility in a government documents environment to capture at risk publications; however, large-scale implementation is currently being sought and no reports of the outcome of the pilot test are available.<sup>39</sup>

*“With the ability to collect and manage the vast amount of grey literature, the DSpace system will provide a framework to enable the academic community to challenge the way research is communicated and distributed. Even after acknowledging that many questions remain unanswered and that DSpace is currently more of a platform for development rather than a solution to the problem of long-term preservation of research materials, it is evident that DSpace will play a valuable role in helping academic libraries and archives develop*

---

<sup>37</sup> “LOCKSS for Government Documents: A Needs Assessment for the Federal Depository Library Community” available at <http://lockss-docs.stanford.edu/NAfinal.pdf> accessed 14 December 2004

<sup>38</sup> More information about DSpace is available at <http://www.dspace.org>

<sup>39</sup> National Association of Government Archives and Records Administrators, “Kansas State Publications Archival Collection,” *Crossroads*, available at [http://www.nagara.org/crossroads/2003\\_2.html](http://www.nagara.org/crossroads/2003_2.html) accessed 14 December 2004.



*necessary strategies and technologies. Long-term management plans are not yet in place at MIT (beyond guaranteed preservation of bit streams), but with the obvious demand for this system, and the development of the DSpace federation of institutions in large academic research institutions in the United States, Canada, and the United Kingdom, DSpace should become a valuable tool and resource.*<sup>40</sup>

**Internet Archive.** The Internet Archive's mission is to prevent "born-digital" materials from disappearing into the past. For this reason, it is building a digital library of Internet sites and other cultural artifacts in digital formats. This material is obtained both through donations of material and capture of web sites. The Internet Archive provides free and open access to its collections to researchers, historians, and scholars.

The biggest of the Internet Archive's collections is the on-line Web. The Internet Archive is the most comprehensive attempt to capture and preserve the web environment. They are acquiring vast amounts of data, working on preservation related copyright issues and demonstrate an awareness of digital preservation issues. The unanswered questions and problems related to digital preservation have not hindered them from very active acquisition.<sup>41</sup>

In the library and archival communities, the Internet Archive is sometimes overlooked as a possible source of content and mechanism for digital preservation. The Internet Archive has worked cooperatively with many organizations, and seems open to such collaborations.

**Flexible Extensible Digital Object and Repository Architecture:** The Mellon FEDORA Project, conducted by the University of Virginia Library and Cornell University and funded by the Andrew W. Mellon Foundation, created a digital object repository system, based on the Flexible Extensible Digital Object and Repository Architecture and including access services through web-based technologies. The system will manage large digital collections by providing ingest, access and interoperability functionality. Although not being preservation-focused, the project's findings on digital collections management can be of interest to the preservation community.<sup>42</sup>

**OCLC Digital Archive:** The Online Computer Library Center has developed an OAI compliant digital archive. Users can place objects in the archive via two mechanisms: 1) a web harvester that allows web content to be described, have preservation metadata applied, and copies the selected files to the archive; and 2) a batch loading process for

---

<sup>40</sup> DSpace. Erpanet Topic of the Month, March 2004 available at <http://www.erpanet.org/topic/digitalrepositories/show.php?id=1045041083&t=3> accessed on 14 December 2004

<sup>41</sup> The Internet Archive. Erpanet Topic of the Month, March 2004 available at <http://www.erpanet.org/topic/digitalrepositories/show.php?id=1037347440&t=3> accessed on 14 December 2004

<sup>42</sup> The Mellon FEDORA (Flexible Extensible Digital Object and Repository Architecture) Project. Erpanet Topic of the Month, March 2004 available at <http://www.erpanet.org/topic/digitalrepositories/show.php?id=1037267118&t=3> accessed on 14 December 2004

digital image masters. Currently the archive ensures that no loss will occur to the files that are stored in it, although preservation actions are being planned for.

**Law Library Microform Consortium (LLMC) and LLMC-Digital:** Based in Hawaii, its mission as a library cooperative was to preserve primary law and legal literature in microform format. During its first 27 years, LLMC filmed more than 7500 titles, and by December 2003, its backfile of filmed images numbered approximately 50 million page images. LLMC-Digital was officially launched in September 2003, with the stated goal of digitizing the complete LLMC backfile, as well as all future filmed titles, in order to make the information available online to subscriber libraries. In addition, LLMC has “explicitly and indefinitely pledged to the retention of an analog format, currently Silver-Halide film, as its storage medium of choice for the purposes of long-term preservation. LLMC-Digital’s data delivery model is comprised of 3 digital formats: TIFF images are the primary format, while back up formats are OCR and pdf text.”<sup>43</sup>

### **Section 3.5: What can be learned from these projects?**

These projects, and the many others that are working on digital preservation and access related issues, each have their own unique viewpoint and methodologies. And working with many of them will be advantageous to the preservation of legal materials. However, equally important are the broader conclusions that can be drawn from them.

**Digital preservation depends on collaboration and partnerships.** Every single project listed here is dependent on more than one organization for its development and success. Collaboration occurs to provide funding, testing, ideas, content, technical development, and data redundancy. And organizations from both the public and private sector are involved.

**There is no single solution.** There is not now, and most likely will never be, a single solution for the preservation of digital materials. In fact, given the diversity of solutions, using many of them, even redundantly, may provide the best solution for the foreseeable future.

**The solutions are not always clear, but we can collect and manage now.** These projects do not provide specific and infallible solutions for digital preservation. But it is clear that the tools exist to begin collecting and managing digital materials while planning for their preservation. We must collect now, before the content is lost.

**Digitization does not ensure digital preservation.** Digitization projects need to conform to standards and best practices to ensure that the digitized resources are high quality and preservation-ready. See Appendix B for a list of projects, standards and best practices that should be taken into consideration.

---

<sup>43</sup> <http://www.llmc.com> and <http://www.LLMC-Digital.org> accessed on 14 December 2004

## Section 4: Requirements for a Successful Digital Preservation Program

The technical challenges of digital preservation are complex and immense. Yet digital preservation requires more than technical expertise. Organizations must be prepared with a policy infrastructure and workflows and staffing that support digital preservation. The technical answers may not be clear, but organizations can create and implement these non-technical components of a digital preservation program today. Ultimately, the success of a digital preservation program for legal materials will be measured by the selection of materials that were deemed worthy of preservation and the ability, in 100 years, of a user to locate and use authentic legal materials from 2004.

### Section 4.1: Program and policy requirements

A coordinated effort is needed to develop an environment in which information is seen and managed as an asset by its producers – in this case, the information producers are most often the government. New laws, policies and responsibilities need to be developed to support this effort. This notion is taking shape and being articulated at state and federal levels – but all its advocates need to come together, including archivists, librarians, records managers, technologists, and policymakers. Until this occurs, digital preservation work will be very difficult and will most likely take the form of triage operations designed to rescue materials that are about to be lost or pockets of activity instead of national coordinated effort.

For the caretakers of legal information, the first step toward successful implementation is to advocate such changes while working on the architecture of a digital preservation program in parallel. Specifically a digital preservation program must develop a preservation policy and strategy that answers the following questions:

- What do you want to preserve? (selection criteria)
- For how long?
- How do you want to access and use the material?
- What resources (knowledge, skills, money) are available?
- In what juridical and organizational framework do you want to operate?<sup>44</sup>

erpaNet has a “Digital Preservation Policy” tool available online that is designed to facilitate the development of a preservation policy.<sup>45</sup> This preservation policy should take into account that there is and will continue to be a variety of efforts on digital

---

<sup>44</sup> Selecting Technologies Tool, Electronic Resource Preservation and Access Network. September 2003 available at [http://www.erpanet.org/www/products/tools/ERPANETSelect\\_Techno.pdf](http://www.erpanet.org/www/products/tools/ERPANETSelect_Techno.pdf) accessed on 14 December 2004

<sup>45</sup> Digital Preservation Policy Tool, Electronic Resource Preservation and Access Network. September 2003 available at <http://www.erpanet.org/www/products/tools/ERPANETPolicyTool.pdf> accessed on 14 December 2004

preservation. Every system will be designed to meet the goals of its sponsoring organization. This heterogeneity actually enhances digital preservation efforts, as it guards against a failure of all the systems. From a policy perspective, however, the policy must be broad enough to accommodate several systems and should not advocate the one size fits all solutions.<sup>46</sup>

## **Section 4.2: Selecting materials for digital preservation**

Determining which materials in digital formats are truly at the greatest risk of being lost must be considered using a combination of traditional selection criteria and a risk analysis of the nature of the digital resources.

Before discussing this further, however, we will acknowledge that any digital surrogates that have an analog counterpart are considered at low risk for loss as long as the analog counterparts are stored and preserved properly. In most library and archives, storage and preservation are adequate or more than adequate; thus, even if the digital surrogate is lost, the analog counterpart will most likely be preserved. This does not preclude the need for standards-based digitizing and best practice back up and management procedures for digitized resources. It simply acknowledges that the existence of the analog source is an adequate preservation strategy.

For materials with no analog back up, the risk is much greater that the content of the digital resource will be lost or corrupted over time. All of these materials are at risk due to the systemic problems mentioned above and the unknown element of preservation actions that will need to be taken in the future. These two factors make determining states of satisfactory and at risk difficult to assess. Indeed, even defining what makes a current preservation strategy “satisfactory” is difficult to determine at this time in digital preservation research and development. This determination ultimately becomes a risk management activity.

### **Managing Risk**

#### **Lowest Risk**

Digital surrogates with an analog counterpart that is properly curated. Long-term sustainability of the content is very likely, however, long term digital access may not be guaranteed.

#### **Low Risk**

Digital content that is curated properly and maintained in digital repositories with preservation policies and strategies specifically stated. Long-term sustainability of the content is likely.

---

<sup>46</sup>Update to the NDIIPP Architecture available at [http://www.digitalpreservation.gov/repor/NDIIPP\\_v02.pdf](http://www.digitalpreservation.gov/repor/NDIIPP_v02.pdf) accessed on 14 December 2004

### Medium Risk

Digital content held by several independent organizations, possibly including the creator, caretakers, and publishers. The redundant storage of this type of materials makes it less likely that the content will be lost in the near term. Although this material may not be stored in true digital repositories, the redundant storage reduces risk of loss significantly. The fact that the content is held redundantly also implies that it is vital for access; this ongoing demand for access to the materials may lead to the ongoing sustainability of the content in digital form. Short term sustainability of the content is very likely; medium term sustainability is likely; long term sustainability is not assured.

### High Risk

Digital content that is not held in redundant systems or a digital repository is at the highest risk of loss. Short term sustainability of the content is likely, but not assured. Medium and long term sustainability is not assured.

### Highest Risk

Digital content that is not held in redundant systems or a digital repository and that cannot be reformatted to eye readable formats. Materials that can be reformatted to eye-readable formats, even if they currently exist in a high risk environment, have a better chance of surviving via reformatting than do these materials which are completely machine dependent. Short term sustainability of the content is likely, but not assured. Medium and long term sustainability is not assured.

### **Where does legal information fall within these levels of risk?**

Placing the variety of types of legal information into these risk categories is helpful, but we must also consider the nature of the legal information itself. The categories listed above determine risk from the perspective of digital preservation; but not from the perspective of the nature of the information. Although all legal information is important, it is fair to say that some is more important than others. Losing decisions of the Supreme Court or the U.S Code would be disastrous; losing a local ordinance or some government publications would be bad, but most likely not an information disaster.

Further analysis is needed in this area; however, it seems fair to speculate that most of the highest use and most valuable legal information fall in the low and medium risk categories. Additionally, the nature of legal material eases the digital preservation needs somewhat. Legal materials in digital formats are generally either held as text or as digital images; two formats that should be somewhat easier to sustain than more complex digital objects and systems. This is a fairly good situation to be in; although far from ideal.

### Lowest Risk Legal Materials

Digitized resources with analog counterparts are at the least risk for loss of content. Many libraries are embarking on digitization projects and programs for these types of materials. With the analog counterpart in place, these materials are at very little risk of loss.

### Low Risk Legal Materials

Legal materials that are managed in digital repositories and even content management systems are at relatively low short and medium term risk. Legal materials held in OAIS compliant digital repositories are at even less risk.

### Medium Risk Legal Materials

Significant amounts of legal materials fall into this category. Although these materials are not currently being managed for preservation purposes, it seems likely that having several organizations managing the content and providing access to it will result in the information being sustainable in the medium term. It is also likely that a significant amount of very vital legal information falls in this category – statutes, high court decisions, etc. Thus, although the materials are “safe for now” it should be strongly stated that risk for these vital materials should be lowered in the near future.

### High and Highest Risk Legal Materials

These materials most likely reside at the creating entity in their own internal systems and on their web sites. Certainly, the risk should be lowered for these materials. However, the importance of the information will need to be determined. It is likely that these materials will be the most difficult and costly to preserve; and it is not clear if the return on the investment is adequate.

In the analog world, organizations were able to curate information resources that were very valuable, valuable, and somewhat valuable. They could do this because there was confidence that the materials, even if they were seldom used, would stand the test of time in the stacks and could be reformatted if they began to deteriorate. The digital environment does not give us this luxury. Digital preservation will require an ongoing commitment of resources and action to ensure that the objects remain accessible and authentic. Benign neglect does not apply. The community will be faced with difficult selection decisions that are uncomfortable and seem almost impossible. However, the reality is that we have limited resources, and we must save what is the most valuable and do our best with the rest.

## **Section 4.3: Operational requirements**

A preservation policy is just one piece of the organizational infrastructure that needs to be created for digital preservation. Earlier we looked at the national response to the brittle books crisis and noted that we must learn from these types of experiences and apply that learning to our digital preservation efforts. The brittle books program provides us with four central questions that need to be answered and implemented to ensure a successful and coordinated national program:

- How do we document the information to be saved, on what medium, and using what standards?
- How do we record the fact that a title has been preserved?

- What should be preserved?
- Who should be responsible for accomplishing the preservation?<sup>47</sup>

These are not easy questions. Yet answering them will provide the basis for a national program that works to ensure the preservation of our nation's legal heritage. The questions raise several issues that need to be addressed.

### **Architecture**

Given the distributed provenance of legal materials, a distributed architecture for ensuring the preservation of those materials is the most viable architecture for a digital preservation system. In this instance, distributed implies that several preservation systems exist and they are geographically distributed and use a variety of technical approaches. Ideally, some basic level of interoperability would exist among the systems, although currently this may not be achievable.

### **Technical requirements**

Distributed systems and caretaker organizations imply a variety of system implementations. However, OAIS compliance for all systems should be implemented if at all possible to ensure that the essential functions for each system are in place. Some level of metadata standardization, or the ability to crosswalk to a common standard, is preferable. The Metadata Encoding Transmission Standard<sup>48</sup> may provide a common standard that can support a variety of implementation types and interoperability among them.

Practically speaking, digital preservation currently requires ongoing data storage and maintenance, systematic back-ups, verification of file formats, creation and storage of metadata, algorithms that ensure data has not been corrupted (fixity checking), procedures to maintain relationships among the files that make up complex digital objects (objects that are made up of many files), disaster preparedness and recovery planning, ongoing preservation planning, network connectivity to ensure data can be transferred appropriately and securely, system administration and maintenance, security measures to ensure access to the digital repository is not compromised, and staffing, tools, and equipment to support these functions. The complexity of implementing the technical components of digital preservation depends on the context, scope and complexity of the digital repository.

### **Staffing**

Again, distributed systems imply distributed and varied staffing. The NARA Electronic Records Archive puts forth the following functions that will be necessary to support a digital preservation program: organizational change management, communications, configuration management, program management, quality management, research, risk

---

<sup>47</sup> Abby Smith, *The Future of the Past: Preservation in American Research Libraries*. CLIR: Washington D.C. April 1999 available at

<http://www.clir.org/pubs/reports/pub82/pub82text.html> accessed on 14 December 2004

<sup>48</sup> See <http://www.loc.gov/mets>

management, and systems engineering.<sup>49</sup> Curatorial functions will also need to be taken into account, such as selection, description, and access.

### **Access**

As mentioned above, the federal model currently places an access requirement on the GPO, while placing the preservation requirement on NARA. In the realm of legal materials, a model will be necessary to ensure that materials are both preserved and accessible. If publishers agree to provide access, does the preservation system need to ensure access as well? Should the preservation system only provide access when other access mechanisms are no longer available? Or should the preservation system take responsibility for access? In a distributed environment, these decisions may have to be made on a case-by-case basis for each system, and in some cases for particular types of materials.

### **Financing and Sustainability**

All of these programmatic components have one commonality. They require funding. A commitment to digital preservation will require ongoing funding and organizational commitment to it. In fact, digital holdings are threatened more by budget problems than by technological problems – there must be long term financial planning from the beginning.<sup>50</sup> Ultimately, this may be the most challenging element to creating the programmatic infrastructure for a preservation program for legal materials.

---

<sup>49</sup> NARA ERA web site available at [http://www.archives.gov/electronic\\_records\\_archives/program\\_management.html](http://www.archives.gov/electronic_records_archives/program_management.html) accessed on 14 December 2004

<sup>50</sup> Selecting Technologies Tool, Electronic Resource Preservation and Access Network. September 2003 available at [http://www.erpanet.org/www/products/tools/ERPANETSelect\\_Techno.pdf](http://www.erpanet.org/www/products/tools/ERPANETSelect_Techno.pdf) accessed on 14 December 2004



## Section 5: Roles and Responsibilities

### Section 5.1: Coordinated collaboration

Where does responsibility for preservation lie? The AALL preservation policy recognizes the need for collaboration, cooperation and communication among groups creating national preservation programs, since preservation is not the sole concern of law libraries.<sup>51</sup>

For primary materials, preservation responsibility lies with the creators and legally mandated caretakers of legal materials – government archives, libraries and other agencies. The preservation of primary legal materials is really a subset of the management and preservation of government records. However, the policies and programs are not always in place to ensure that this will occur at the federal, state and local levels. For secondary materials, preservation responsibility is not clearly defined and may currently be at the whim of the publisher.

Practically speaking, responsibility may lie with those who are concerned about and willing to take responsibility. This situation again reinforces the need for the distributed model mentioned above in which responsibility for digital preservation is shared among a variety of government agencies, commercial entities, and non-profit entities. The challenge is to implement this model in a coordinated manner. Each of these three entities has strengths and weaknesses, but working together should ensure the strongest possible effort.

Government agencies are predominantly the statutorily mandated preservation providers. They are also often under-funded and many, particularly at the state and local levels, have little or no capacity for a digital preservation program. These agencies are also responsible for the preservation of many types of government information and may not have a particular focus on legal materials. In a distributed architecture, government entities bring the strength of the legal mandate to the broader initiative.

Commercial entities often see the benefit of collaboration due to the changing nature of business models. There are two arenas in which the commercial sector may be helpful. First, commercial organizations could develop solutions for digital preservation. There seems to be little evidence of a market for digital preservation even though there appears to be a market for access to legal information. However, NARA argues that the creation of their ERA will provide an ongoing business for the chosen company. If NARA's assumption is correct, the market for digital preservation may be somewhat hidden. And further, the larger question for commercial entities is whether or not the market has the ability to pay for the tools and services they offer.

---

<sup>51</sup> AALL Preservation Policy available at [http://www.aallnet.org/about/policy\\_preservation.asp](http://www.aallnet.org/about/policy_preservation.asp) accessed on 14 December 2004

Commercial publishing entities could also have a role in digital preservation if they can be convinced to create and maintain digital preservation systems. If they do so, contingency plans will need to be in place if the publisher goes out of business. In the proposed distributed model, commercial entities can provide leadership in business models, planning, research, and development.

Non-profit entities encompass a variety of organizations, including organizations, libraries, and archives. Often non-profit organizations have funding difficulties and because they are not motivated by mandate or profit, their motives for getting into digital preservation must be service oriented. This is the strength of the non-profit – creative collaborations and a commitment to service.

Again, we must be guided by reality. There are a handful of organizations interested in pursuing digital preservation work, and there are others who should be interested, but aren't. There are also many organizations that are responsible for the preservation of legal materials, in any format. What the role of these organizations is has yet to be determined, but the ongoing and over-arching reality is that a coordinated and collaborative effort is paramount to ensuring legal materials are preserved.

## Section 6: Recommendations

LIPA is the only organization specifically concerned about the preservation of digital legal materials. Others are concerned about digital preservation generally or about other types of materials. So, in order to move forward, LIPA's role will need to be that of leadership and advocacy. These recommendations are high level and a significant amount of work will be necessary to drill down and understand the intricacies necessary for a coordinated national program. It should be stressed, however, that these recommendations are not technical in nature, but organizational. The technical challenges of digital preservation cannot adequately be confronted until the organizational and planning groundwork has been laid.

We offer the following recommendations:

### **Planning**

#### **1. LIPA needs to define what a successful coordinated national digital preservation program for legal information will be.**

Whether the goals be modest or overarching, without defining success LIPA will be unable to evaluate whether or not they have accomplished what they set out to do. Its activities and related activities should be evaluated regularly to assure the initiative is moving ahead appropriately.

#### **2. LIPA needs to develop a program vision and act as the coordinating agency of a national digital preservation program for legal materials.**

We have outlined a very high level vision here – one consisting of a distributed architecture with flexible implementation models.

#### **3. LIPA needs to develop a high level and long term action plan for achieving this vision.**

Such a plan should incorporate the following recommendations:

- a. LIPA needs to advocate for an infrastructure within government that promotes and supports good management of digital information and that reflects the age of e-government.**

Identifying and facilitating appropriate legislative and policy changes at the federal, state and local level which ensure the preservation of legal information. This effort may include areas such as records and information management programs, copyright, advocating funding, and creating and/or reviewing technical best practices. The need for funding to be built into this infrastructure is vital.

- b. LIPA needs to create selection criteria for identifying legal materials that are the highest priority for preservation based on their informational content and technical environment.**

A coordinated effort then needs to take place to implement the criteria and collect digital legal materials. Materials at high risk and high informational value content need to be the first priority for action. Materials with medium risk and high informational content should be the second priority.

- c. LIPA needs to foster education and training opportunities for libraries and other organizations that need to plan for and implement digital preservation systems.**

Digital preservation is new to many in the library and archival fields. LIPA should facilitate educational opportunities at AALL meetings and stand alone meetings. Additionally, LIPA should make every effort to keep its members informed of regional and local educational opportunities that they can take advantage of.

- d. LIPA needs to identify the technical options available for a coordinated effort, and create standards and best practices that are appropriate for such a distributed model.**

Since system implementation will vary, such standards and best practices will most likely be functional, not prescriptive and high level, not detailed. Some questions to be answered: is OAIS compliance required? Do all the systems need to be interoperable? Is there a minimum amount of metadata that needs to be maintained by each system? What happens if a system fails (due to organizational issues, technical problems, or natural or man-made disasters), or a preservation action fails, and data is lost? The technical infrastructure will need to be thoroughly and thoughtfully considered.

### **Coordination and Cooperation**

- 4. LIPA needs to specifically identify partners in this endeavor and cultivate those relationships**

- 5. LIPA needs the full support of AALL to pursue a national digital preservation initiative for legal materials.**

LIPA should be formally recognized as an AALL interest group. Further, AALL should revise its current preservation policy so as not to rule out digital formats for preservation. AALL should create a digital preservation policy that acknowledges that reality of digital preservation and sets out its priorities for digital preservation.

**6. LIPA needs to become actively involved with related projects that are ongoing at the federal and state level as an advocate of collaborative, coordinated action on behalf of legal materials.**

Partners should include government, public and private sector representatives. Given the current ongoing programs and projects in this area, potential partners include LIPA members that have existing digital preservation projects, NARA, digitization projects such as LLMC, PACER, GPO, Washington State Archives, Cornell University, SDSC, the Internet Archive, etc. And, since primary legal materials are ultimately government records, partnerships with government agencies should be substantial. LIPA should consider outreach to organizations such as the National Association of Government Archivists and Records Administrators. Internationally, the Legal Information Institute is a potential partner as it pursues work on capturing and sustaining versions of legal information.

**Funding**

**7. LIPA and its partners need to advocate for and establish a funding base and funding opportunities for digital preservation systems.**

As stated above, funding should be programmatic, not project based. Preservation, whether digital or analog, implies a long-term commitment to and responsibility for these legal materials that cannot be supported by one-time grant monies. An ongoing, sustainable source of funding must be obtained to ensure the preservation of legal materials, and all materials in digital formats.

Grant funding will be instrumental in pursuing start-up projects today and research and test bed projects in the future. Currently, there are new and developing opportunities for grant funding of digital preservation projects. LIPA needs to facilitate grant writing for projects and systems; act as an advocate for funding for the infrastructure necessary to promote digital preservation and access at the federal, state and local levels; and should investigate business models to fund the preservation of digital materials.

## **Staffing**

**8. LIPA needs a minimum of one full time paid staff to perform outreach to other interested parties, act as an advocate to uninterested parties, and provide leadership.**

Staff needs to research digital preservation projects and issues, create case studies, attend meetings, interface with vendors and AALL, coordinate projects, write grant proposals, and advocate preservation of legal materials. Such a position requires a person with technical knowledge as well as skills in information management, outreach and grant writing. One position is not fully adequate to this work, but it is a minimum and could be supplemented with an organized volunteer effort.

# Appendix A: Digital Preservation Resources

(This list of resources has been adapted from the Cornell University Digital Preservation Online Tutorial.)

## Glossaries

Attributes of a Trusted Digital Repository, Appendix B  
<http://www.rlg.org/longterm/attributes01.pdf>

Colorado Digitization Program, Digitization Glossary  
[http://www.cdpheritage.org/resource/introduction/rsrc\\_glossary.html](http://www.cdpheritage.org/resource/introduction/rsrc_glossary.html)

Definitions and Concepts from Preservation Management of Digital Materials: A Handbook by Maggie Jones and Neil Beagrie.  
<http://www.dpconline.org/text/intro/definitions.html>

National Library of Australia Information Paper Digital Services Project Appendix G  
<http://www.nla.gov.au/download/dsp/appendices.pdf>

NEDLIB Glossary  
<http://www.nla.gov.au/download/dsp/appendices.pdf>

Universal Preservation Format Glossary  
<http://info.wgbh.org/upf/glossary.html>

Working Definitions of Commonly Used Terms for the purposes of the Cedars Project.  
<http://www.leeds.ac.uk/cedars/documents/PSW01.htm>

## Recommended Resources

Consultative Committee for Space Data Systems, Reference Model for an Open Archival Information System (OAIS), Blue Book, January 2002.  
<http://ssdoo.gsfc.nasa.gov/nost/isoas/overview.html> (overview)  
<http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf>  
(standard)

Trusted Digital Repositories: Attributes and Responsibilities, An RLG-OCLC Report. May 2002.  
<http://www.rlg.org/longterm/repositories.pdf>

Cornell Digital Preservation Tutorial.  
<http://www.library.cornell.edu/iris/tutorial/dpm/>

Jones, Maggie, and Neil Beagrie. Preservation Management of Digital Materials: A Handbook. 2001.

<http://www.dpconline.org/graphics/handbook/index.html>

Preserving Access to Digital Information (PADI), National Library of Australia.

<http://www.nla.gov.au/padi/>

Preservation Metadata and the OAIS Information Model, A Metadata Framework to Support the Preservation of Digital Objects: A Report by the OCLC-RLG Working Group on Preservation Metadata, June 2002.

[http://www.oclc.org/research/projects/pmwg/pm\\_framework.pdf](http://www.oclc.org/research/projects/pmwg/pm_framework.pdf)

National Digital Information Infrastructure and Preservation Program (NDIIPP), Library of Congress. Reports and Papers.

<http://www.digitalpreservation.gov/>

Guidelines for the Preservation of Digital Heritage. Webb, Colin. 2003.

[http://portal.unesco.org/ci/ev.php?URL\\_ID=8967&URL\\_DO=DO\\_TOPIC&URL\\_SECTION%3E=201&reload=1049879672](http://portal.unesco.org/ci/ev.php?URL_ID=8967&URL_DO=DO_TOPIC&URL_SECTION%3E=201&reload=1049879672)

Preserving Digital Information: Report of the Task Force on Archiving of Digital Information. The Commission on Preservation and Access and RLG. 1996.

<http://www.rlg.org/ArchTF/>

Rothenberg, Jeff. Ensuring the Longevity of Digital Information. Scientific American.

<http://www.clir.org/pubs/archives/ensuring.pdf>

Dollar, Charles. Authentic Electronic Records: Strategies for Long-Term Access.

Cohasset Associates. January 1999.

<http://www.mybestdocs.com/CSum1.html>

### **Additional Resources**

Byrnes, Margaret. "Assigning permanence levels to NLM's electronic publications"

<http://www.rlg.org/events/pres-2000/byrnes.html>

Cornell University Electronic Student Records Systems Project Report. 2000.

<http://rnc.library.cornell.edu/online/studentRecords/>

Council on Library and Information Resources, digital preservation publications.

<http://www.clir.org/pubs/>

Data Dictionary - Technical Metadata for Digital Still Images NISO Z39.87-2002 AIIM 20-2002.

[http://www.niso.org/standards/resources/Z39\\_87\\_trial\\_use.pdf](http://www.niso.org/standards/resources/Z39_87_trial_use.pdf)



Digital Preservation Needs and Requirements in RLG Member Institutions. RLG. 1998.  
<http://www.rlg.org/preserv/digpres.html>

Invest to Save: Report and Recommendations of the NSF-DELOS Working Group on Digital Archiving and Preservation, Network of Excellence for Digital Libraries (DELOS), 2003.  
<http://delos-noe.iei.pi.cnr.it/activities/internationalforum/Joint-WGs/joint-wgs.html>

It's About Time: Research Challenges in Digital Archiving and Long-Term Preservation, Final Report, Workshop on Research Challenges in Digital Archiving and Long-Term Preservation, April 12-13, 2002. August 2003.  
<http://www.clir.org/pubs/film/future/discussion.html>

Metadata Standards Framework - Preservation Metadata. National Library of New Zealand, November 2002.  
[http://www.natlib.govt.nz/files/4initiatives\\_metaschema.pdf](http://www.natlib.govt.nz/files/4initiatives_metaschema.pdf)

National Historical Publications and Records Commission, Electronic Records Projects.  
[http://www.archives.gov/grants/electronic\\_records/electronic\\_records.html](http://www.archives.gov/grants/electronic_records/electronic_records.html)

National Library of Australia, "Guidelines: Managing Web Resources for persistent access".  
<http://www.nla.gov.au/guidelines/2000/persistence.html>

Preservation Metadata for Digital Objects: A Review of the State of the Art. White paper by the OCLC/RLG Working Group on Preservation Metadata, March 2001.  
[http://www.oclc.org/research/projects/pmwg/presmeta\\_wp.pdf](http://www.oclc.org/research/projects/pmwg/presmeta_wp.pdf)

Producer-Archive Interface Methodology Abstract Standard (CCSDS), June 2002.  
[http://bill.cacr.caltech.edu/cfdocs/usvo-pubs/files/CCSDS\\_651\\_W2.pdf](http://bill.cacr.caltech.edu/cfdocs/usvo-pubs/files/CCSDS_651_W2.pdf)

Selection and Collaboration in Digital Preservation: An RLG-JISC Symposium, 24-25 March 2003.  
<http://www.rlg.org/events/2003rlgjisc/>

Submission Information Package (SIP) Specification, Version 1.0 Draft. Harvard E-Journal Archive. December 19, 2001.  
<http://www.diglib.org/preserve/harvardsip10.pdf>

Waibel, Günter. Like Russian Dolls: Nesting Standards for Digital Preservation.  
<http://www.rlg.org/preserv/diginews/feature2>

## **Policies**

Harvard Digital Repository Services Policy Guide

<http://hul.harvard.edu/ois/systems/drs/policyguide.html>

National Library of Australia, Digital Preservation Policy  
<http://www.nla.gov/policy/digpres.html>

National Archives of Australia, An Approach to the Preservation of Digital Records  
[http://www.naa.gov.au/recordkeeping/er/digital\\_preservation/summary.html](http://www.naa.gov.au/recordkeeping/er/digital_preservation/summary.html)

National Preservation Office, Building Blocks for a Preservation Policy  
<http://www.bl.uk/services/preservation/np08.pdf>

Neil Beagrie: A Continuing Access and Digital Preservation Strategy for the Joint Information Systems Committee (JISC) 2002-2005. October 2002.  
<http://www.jisc.ac.uk/dner/preservation/dpstrategy2002b.rtf>

Neil Beagrie and Daniel Greenstein: A strategic policy framework for creating and preserving digital collections. Version 5.0 (2001):  
<http://ahds.ac.uk/strategic.htm>

Columbia University Libraries Policy for Preservation of Digital Resources. July 2000.  
<http://www.columbia.edu/cu/libraries/services/preservation/dlpolicy.html>

InterPARES Strategy Task Force Report: An Intellectual Framework for Policies, Strategies, and Standards. In: The Long-Term Preservation of Authentic Electronic Records: Findings of the InterPARES Project.  
<http://www.interpares.org/book/index.htm>

PRO (Public Record Office), the National Archives: Corporate policy on electronic records – Standards for the management of Government records; September 2000.  
<http://www.pro.gov.uk/recordsmanagement/eros/RMCorpPol.pdf>

PRO: Management, Appraisal and Preservation of Electronic Records – Volume 2: Procedures, Preservation of electronic records.  
<http://www.pro.gov.uk/recordsmanagement/eros/guidelines/default.htm>

### **Digital Preservation Projects**

Arts and Humanities Data Service (AHDS) Preservation Management of Digital Materials  
<http://www.ahds.ac.uk/about/publications/index.htm>

Creative Archiving at Michigan and Leeds: Emulating the Old on the New (CAMiLEON)  
<http://www.si.umich.edu/CAMiLEON/>

CURL Exemplars in Digital Archives (CEDARS)

<http://www.leeds.ac.uk/cedars/>

Harvard University Digital Library Initiative

<http://hul.harvard.edu/ldi/>

InterPARES Project (International Research on Permanent Authentic Records in Electronic Systems)

<http://www.interpares.org>

Kulturaw Heritage Project

<http://www.kb.se/kw3/ENG/Description.htm>

Lots of Copies Keep Stuff Safe (LOCKSS)

<http://lockss.stanford.edu>

PANDORA (Preserving and Accessing Networked Documentary Resources of Australia)

<http://pandora.nla.gov.au/index.html>

PRISM (Preservation, Reliability, Interoperability, Security, Metadata)

<http://www.library.cornell.edu/iris/research/prism/index.html>

National Archives and Records Administration Electronic Records Archives

[http://www.archives.gov/electronic\\_records\\_archives/index.html/](http://www.archives.gov/electronic_records_archives/index.html/)

NEDLIB (Networked European Deposit Library)

<http://www.kb.nl/coop/nedlib/>

OAIS (ISO Reference Model for an Open Archival Information System)

<http://ssdoo.gsfc.nasa.gov/nost/isoas/us/overview.html>

OCLC Digital Archive System Guides: Digital Archive Metadata Elements

<http://www.oclc.org/support/documentation/digitalarchive/default.htm>

San Diego Supercomputer Center and NPACI Data-Intensive Computing (DICE)

<http://www.npaci.edu/DICE/>

Stanford Archival Digital Library Architecture: Archival Digital Libraries Repositories

<http://www-db.stanford.edu/>

# Appendix B: Digital Imaging Resources

## General Resources

Cornell Digital Imaging Tutorial

<http://www.library.cornell.edu/preservation/tutorial/contents.html>

Western States Digital Imaging Best Practices

[http://www.cdpheritage.org/resource/scanning/documents/WSDIBP\\_v1.pdf](http://www.cdpheritage.org/resource/scanning/documents/WSDIBP_v1.pdf)

## Metadata

Western States Dublin Core Metadata Best Practices

<http://www.cdpheritage.org/resource/metadata/wsdcmbp/index.html>

Dublin Core

<http://www.dublincore.org>

MODS

<http://www.loc.gov/standards/mods/>

VRA

<http://www.vraweb.org/>

NISO Technical Metadata

[http://www.niso.org/standards/resources/Z39\\_87\\_trial\\_use.pdf](http://www.niso.org/standards/resources/Z39_87_trial_use.pdf)

MIX

<http://www.loc.gov/standards/mix/>

METS

<http://www.loc.gov/standards/mets/>

## **Besser Principles**

- Scan at the highest resolution appropriate to the informational content of the originals
- Scan at an appropriate level of quality to avoid rescanning and re-handling of the originals in the future--scan once
- Create and store a master image file that can be used to produce derivative image files and serve a variety of current and future user needs
- Use system components that are non-proprietary
- Use image file formats and compression techniques that conform to industry standards
- Create backup copies of all files on a stable medium
- Create meaningful metadata for image files or collections
- Store media in an appropriate environment
- Monitor and recopy data as necessary
- Outline a migration strategy for transferring data across generations of technology
- Anticipate and plan for future technological developments

# Bibliography

## AALL publications and documents:

*AALL Preservation Policy*, adopted by the AALL Executive Board, July 1994, Revised July 10, 1998, available online at:

[http://www.aallnet.org/about/policy\\_preservation.asp](http://www.aallnet.org/about/policy_preservation.asp)

AALL Washington Affairs Office, Issue Brief, *The Next Generation Electronic Government Information Access Act of 1999*, July 1999, available online at:

<http://www.ll.georgetown.edu/aallwash/Isbrnext.html>

AALL Washington Affairs Office, Issue Brief, *Law Library of Congress*, July 1999, available online at:

<http://www.ll.georgetown.edu/aallwash/Isbrlaw.html>

AALL Washington Affairs Office, Issue Brief, *Equal and Equitable Public Access to the Rule of Law*, June 1999, available online at:

<http://www.ll.georgetown.edu/aallwash/ib069901.html>

AALL Washington Affairs Office, Issue Brief, *Public Access to Government Information*, July 1999, available online at:

<http://www.ll.georgetown.edu/aallwash/Isbrpubl.html>

AALL Washington Affairs Office, Working Document, *Enhanced Library Access and Dissemination of Federal Government Information: A Framework for Future Discussion*, June 1995, available online at:

<http://www.ll.georgetown.edu/aallwash/rep06251995.html>

Legal Information Preservation Alliance, *Conference Report: Preserving Legal Information for the 21<sup>st</sup> Century: Toward a National Agenda, March 6-8, 2003*. May 2003.

[http://www.aallnet.org/committee/lipa/LIPA\\_Conference\\_Report.pdf](http://www.aallnet.org/committee/lipa/LIPA_Conference_Report.pdf) accessed on 21 October 2004.

*Permanent Public Access of State Government Information: A State-by-State Compilation and Resource Guide*. A project of the American Association of Law Libraries Government Relations Committee, funded by a grant from Aspen Publishing Company, 2002/03, available online at:

<http://www.ll.georgetown.edu/aallwash/PPAreport.htm>.

### **Legal research and legal periodical articles:**

Administrative Office of the United States Courts, Office of Judges Programs. “*Privacy and Access to Electronic Case Files in the Federal Courts*”, December 15, 1999, available online at <http://www.privacy.uscourts.gov>

Adam, Nabil R., et al. *The Global Legal Information Network (GLIN)*, 46 *American University Law Review* 477 (1996)

Axtmann, Margaret Maes. *Preservation - The Time is Right*, *AALL Spectrum*, June 2003 p.22

Cohen, Morris L. *An Historical Overview of American Law Publishing (Order from Chaos: Contexts for Global Legal Information)*, 31 *International Journal of Legal Information* 179 (2003)

Crocker, C. Anne. *The Official Version: Authenticating, Preserving and Citing Legal Information in Digital Form*, 26 *International Journal of Legal Information* 23 (1998)

Danner, Richard A. *Contemporary and Future Directions in American Legal Research: Responding to the Threat of the Available (Order from Chaos: Contexts for Global Legal Information)*, 31 *International Journal of Legal Information* 168 (2003)

Eisenberg, Theodore and Schlanger, Margo. *The Reliability of the Administrative Office of the U.S. Courts Database: An Initial Empirical Analysis*, 78 *Notre Dame Law Review* 1455 (2003)

Ford, Don. *Preservation of Electronic Government Documents in the Post-9/11 World*, *AALL Spectrum* Sept. 2003 p.26

Healey, Paul D. *Georgetown Conference Outlines Preservation Agenda*, *AALL Spectrum*, June 2003 p.18

Hyde, Janice S. *Building the Global Legal Information Network (GLIN)*, 19 *Legal Reference Services Quarterly* 157 (2001)

Joergensen, John P. *Are Non-Profit Internet Publishers the Future of Legal Information (The Political Economy of Legal Information: The New Landscape)*, 17 *Legal Reference Services Quarterly* 33 (1999)

LaVigne, Mark. *Electronic Government: A Vision of a Future that is Already Here*, 52 *Syracuse Law Review* 1243 (2002)

Marke, Julius J. *Use of the Global Legal Information Network*, *New York Law Journal*, Sept. 16, 1997, p.5

Merksy, Roy M. and Dunn, Donald J. *Fundamentals of Legal Research*, 8<sup>th</sup> ed. 2002.

Miller, Diana Frazier. *Here Today, Gone Tomorrow? The Archivability of Electronic Records*, 26 International Journal of Legal Information 39 (1998)

Office of Judges Programs of the Administrative Office of the United States Courts, "Privacy and Access to Electronic Case Files in the Federal Courts," December 15, 1999.

<http://www.privacy.uscourts.gov> (accessed on 14 December 2004)

Southern California Association of Law Libraries. *Locating the Law: A Handbook for Non-Law Librarians*.

<http://www.aallnet.org/chapter/scall/locating.htm> (accessed on 14 December 2004)

Rothenberg, Jeff. *Ensuring the Longevity of Digital Information*, 26 International Journal of Legal Information 1 (1998)

Wei Luo, et al. *Government Documents Around the World: Access Policies and Strategies of Disclosure*, AALL Spectrum, Oct. 2002 p. 14

### **Traditional and digital preservation publications:**

Alemneh, Daniel Gelaw, et al. "A Metadata Approach to Preservation of Digital Resources: The University of North Texas Libraries' Experience." *First Monday* 7:8 (August 2002).

[http://firstmonday.org/issues/issue7\\_8/alemneh/index.html](http://firstmonday.org/issues/issue7_8/alemneh/index.html) (accessed 9 July 2004)

American Association for the Advancement of Science. *Science for All Americans Online: Chapter 3 The Nature of Technology*.

<http://www.project2061.org/tools/sfaaol/chap3.htm> (accessed 14 December 2004)

Axtmann, Margaret Maes. "Preservation – The Time is Right," AALL Spectrum Volume 7 Number 9, June 2003.

Barata, Kimberley. "Archives In The Digital Age: A Study For Resource." London, UK: Missenden Consulting for the Museum Library and Archives Council, November 2002.

<http://www.mla.gov.uk/documents/id435rep.doc> (accessed 29 July 2004)

Barnum, George D. and Steven Kerchoff. "The Federal Depository Library Program Electronic Collection: Preserving a Tradition of Access to United States Government Information." *Proceedings of the Information Infrastructures for Digital Preservation*



*and Preservation 2000: An International Conference on the Preservation and Long Term Accessibility of Digital Materials.* York, England, 6-8 December 2000.

<http://www.rlg.org/events/pres-2000/barnum.html> (accessed 29 July 2004)

Biehl, Kathy. *The Internet Archives: Preserving the History of Web Pages*, November 1, 2001.

<http://www.llrx.com/columns/webcritic12.htm>

*CAMiLeon: Creative Archiving at Michigan & Leeds: Emulating the Old on the New.* Ann Arbor, MI: University of Michigan, 2003.

<http://www.si.umich.edu/CAMILEON/> (accessed on 14 December 2004)

Categorization of Government Information (CGI) Working Group, U.S. Federal Interagency Committee on Government Information. *Defining What Government Information Is To Be Categorized.* Washington, DC.

<http://www.gpoaccess.gov/cgiwg/pdf/cgiwggroup.pdf> (accessed 29 July 2004)

*Collection Care - Digital Preservation.* London, U.K.: British Library.

<http://www.bl.uk/about/collectioncare/digpresintro.html> (accessed 29 July 2004)

*Collection Of Last Resort : Discussion Draft, April 6, 2004.* Washington, DC: U.S. Government Printing Office, April 2004.

[http://www.access.gpo.gov/su\\_docs/fdlp/pubs/clar.pdf](http://www.access.gpo.gov/su_docs/fdlp/pubs/clar.pdf) (accessed 29 July 2004)

Cornell University. *Digital Preservation Management; Implementing Short Term Strategies for Long Term Management .*

<http://www.library.cornell.edu/iris/tutorial/dpm/> (accessed on 14 December 2004)

Council on Library and Information Resources. *Authenticity in a Digital Environment.*

<http://www.clir.org/pubs/abstract/pub92abst.html> (accessed 29 July 2004)

Council on Library and Information Resources and Library of Congress. *Building a National Strategy for Digital Preservation: Issues for Digital Media Archiving*, April 2002.

<http://www.clir.org/pubs/reports/pub106/contents.html> (accessed on 14 December 2004)

Electronic Resource Preservation and Access Network. *Digital Preservation Policy Tool.* September 2003.

<http://www.erpanet.org/www/products/tools/ERPANETPolicyTool.pdf> (accessed 14 December 2004)

Electronic Resource Preservation and Access Network. *DSpace.* Erpanet Topic of the Month, March 2004.

<http://www.erpanet.org/topic/digitalrepositories/show.php?id=1045041083&t=3>  
(accessed 14 December 2004)

Electronic Resource Preservation and Access Network. *The Mellon FEDORA (Flexible Extensible Digital Object and Repository Architecture) Project*. Erpanet Topic of the Month, March 2004.

<http://www.erpanet.org/topic/digitalrepositories/show.php?id=1037267118&t=3>  
(accessed 14 December 2004)

Electronic Resource Preservation and Access Network. *IBM/KB Long Term Preservation Study*. Erpanet Topic of the Month, March 2004.

<http://www.erpanet.org/topic/digitalrepositories/show.php?id=1073637964&t=3>  
(accessed 14 December 2004)

Electronic Resource Preservation and Access Network. *The Internet Archive*. Erpanet Topic of the Month, March 2004.

<http://www.erpanet.org/topic/digitalrepositories/show.php?id=1037347440&t=3>  
(accessed 14 December 2004)

Electronic Resource Preservation and Access Network. *Selecting Technologies Tool*. September 2003.

[http://www.erpanet.org/www/products/tools/ERPANETSelect\\_Techno.pdf](http://www.erpanet.org/www/products/tools/ERPANETSelect_Techno.pdf) (accessed 14 December 2004)

*The Fedora™ Project: An Open-Source Digital Repository Management System*. Charlottesville, VA: University of Virginia, 2004.

<http://www.fedora.info/> (accessed 29 July 2004)

Feldman, Susan and Chris Sherman. "The High Cost of Not Finding Information: An IDC White Paper." Framington, MA: IDC, July 2001.

[http://monkey.biz/Content/Default/Support/Resources/IDC\\_TheHighCostOfNotFindingInformation\\_1510.pdf](http://monkey.biz/Content/Default/Support/Resources/IDC_TheHighCostOfNotFindingInformation_1510.pdf)

Germain, Claire M. *Digital Legal Information: Ensuring Access to the "Official" Word of the Law Cornell Law School*. Cornell Law School Working Papers Series. Working Paper 2.

<http://lsr.nellco.org/cornell/clsops/papers/2> (accessed on 21 October 2004.)

Gwinn, Nancy E. *A National Preservation Program for Agricultural Literature*. Smithsonian Institution Libraries, May 16, 1993.

<http://www.nal.usda.gov/preservation/AgLit/natpresprog.shtml> (accessed on 22 October 2004)

Healey, Paul D. "Georgetown Conference Outlines Preservation Agenda," AALL Spectrum, Volume 7 Number 9 June 2003.

*InterPares: The International Research on Permanent Authentic Records in Electronic Systems*. Vancouver, BC: InterPARES Project, School of Library, Archival and

Information Studies. The University of British Columbia, 2004.  
<http://www.interpares.org/> (accessed 29 July 2004)

Jones, Maggie and Neil Beagrie. *Preservation Management of Digital Materials: A Handbook*.  
<http://www.dpconline.org/text/intro/definitions.html> (accessed on 14 December 2004)

*LC21: A Digital Strategy for the Library of Congress*. National Research Council. Washington, DC: National Academy Press, 2000.  
<http://books.nap.edu/html/lc21/> (accessed 29 July 2004)

Lavoie, Brian F. "Of Mice and Memory: Economically Sustainable Preservation for the Twenty-first Century." In *Access in the Future Tense*. Washington, D.C.: Council on Library and Information Resources, 2004.  
<http://www.clir.org/pubs/abstract/pub126abst.html> (accessed on 14 December 2004)

Lavoie, Brian F. *The Open Archival information System Reference Model: Introductory Guide*. "A joint report of the Digital Preservation Coalition (DPC) and OCLC, published electronically as a DPC Technology Report, 2004.  
[http://www.dpconline.org/docs/lavoie\\_OAIS.pdf](http://www.dpconline.org/docs/lavoie_OAIS.pdf) (accessed online 14 December 2004)

Lavoie, Brian F. *The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making*, 2003.  
<http://www.oclc.org/research/projects/digipres/incentives-dp.pdf>

Law Library Microform Consortium. *LLMC Offers Digital Access While Remaining Committed to Film for Preservation and Local-Backup Purposes*.  
[http://www.llmc.com/general\\_description.htm](http://www.llmc.com/general_description.htm) (accessed on 14 December 2004)

LeFurgy, William G. *PDF/A: Developing a File Format for Long-Term Preservation*.  
[http://www.rlg.org/preserv/diginews/v7\\_n6\\_feature1.html](http://www.rlg.org/preserv/diginews/v7_n6_feature1.html) (accessed 14 December 2004)

"*LOCKSS for Government Documents: A Needs Assessment for the Federal Depository Library Community*."  
<http://lockss-docs.stanford.edu/NAfinal.pdf> (accessed 29 July 2004)

Moehle, Frank. *The Role of File Formats in Digital Preservation: Opportunities and Threats*. Powerpoint as presented at *ERPANET File Formats for Preservation*, Vienna. Monday 10th - Tuesday 11th May 2004 at Austrian National Library, Vienna, Austria.  
[http://www.erpanet.org/events/2004/vienna/presentations/erpaTrainingVienna\\_Moehle.pdf](http://www.erpanet.org/events/2004/vienna/presentations/erpaTrainingVienna_Moehle.pdf) (accessed on 14 December 2004)

National Association of Government Archives and Records Administrators, "*Kansas State Publications Archival Collection*," Crossroads Newsletter.  
[http://www.nagara.org/crossroads/2003\\_2.html](http://www.nagara.org/crossroads/2003_2.html) (accessed 14 December 2004)

National Information Standards Organization. *Data Dictionary – Technical Metadata for Digital Still Images*.

[http://www.niso.org/committees/committee\\_au.html](http://www.niso.org/committees/committee_au.html) (accessed 29 July 2004)

National Information Standards Organization. *The Dublin Core Metadata Element Set*.

<http://www.niso.org/standards/resources/Z39-85.pdf> (accessed 29 July 2004)

National Science Foundation and Library of Congress. *It's About Time: Research Challenges in Digital Archiving and Long-Term Preservation*. August 2003.

[http://www.digitalpreservation.gov/repor/NSF\\_LC\\_Final\\_Report.pdf](http://www.digitalpreservation.gov/repor/NSF_LC_Final_Report.pdf) (accessed on 14 December 2004)

NEDLIB Glossary.

<http://www.kb.nl/coop/nedlib/glossary.pdf> (accessed on 14 December 2004)

*Report on GPO's Transition to a More Electronic FDLP*. US Depository Library Council Meeting. Newport, Rhode Island, Spring 2000.

[http://www.access.gpo.gov/su\\_docs/fdlp/council/etreport.html](http://www.access.gpo.gov/su_docs/fdlp/council/etreport.html) (accessed 29 July 2004)

*Report On The Assessment Of Electronic Government Information Products*. United States National Commission On Libraries And Information Science Washington, DC: U.S. Government Printing Office, March 30, 1999.

[http://www.access.gpo.gov/su\\_docs/nclisassessment/report.html](http://www.access.gpo.gov/su_docs/nclisassessment/report.html) (accessed 29 July 2004)

RLG and OCLC. *Trusted Digital Repositories: Attributes and Responsibilities*. May 2002.

<http://www.rlg.org/longterm/repositories.pdf> (accessed on 22 October 2004)

Smith, Abby. *The Future of the Past: Preservation in American Research Libraries*. CLIR: Washington D.C. April 1999.

<http://www.clir.org/pubs/reports/pub82/pub82text.html> (accessed on 21 October 2004)

Sproull, Robert F. & Jon Eisenberg, eds. *“Building an Electronic Records Archive at the National Archives and Records Administration: Recommendations for Initial Development.”* Computer Science and Telecommunications Board, National Academies, 2003.

<http://www.nap.edu/books/0309089476/html/> (accessed 29 July 2004)

Turpening, Patricia K. *AALL Takes First Steps to Develop a National Preservation Plan* Abbey Newsletter Volume 26, Number 6, Oct 2003

<http://palimpsest.stanford.edu/byorg/abbey/an/an26/an26-6/an26-605.html> (accessed on 21 October 2004)

U.S. National Commission on Libraries and Information Science. *A comprehensive assessment of public information dissemination, final report.* Washington, DC: U.S. Government Printing Office, 2001.

<http://www.nclis.gov/govt/assess/assess.voll.pdf> (accessed on 14 December 2004)

Washington Secretary of State. *Washington State Digital Archives Grand Opening.*

<http://www.digitalarchives.wa.gov/Content.aspx?txt=news> (accessed 14 December 2004)

*Washington Opens Digital Archive.* Forbes.com October 4, 2004.

<http://www.forbes.com/business/feeds/ap/2004/10/04/ap1575094.html> (accessed 14 December 2004)

#### **Other Web Sites Accessed:**

DSpace <http://www.dspace.org>

Global Legal Information Network (GLIN) <http://www.loc.gov/law/glin>

HeinOnline <http://heinonline.org>

Law Library of Congress <http://www.loc.gov/law/public/law.html>

Law Library Microform Consortium (LLMC) and LLMC-Digital

<http://www.llmc.com> and <http://www.LLMC-Digital.org>

Library of Congress "A Century of Lawmaking for a New Nation: U.S. Documents and Debates 1774-1875" <http://lcweb2.loc.gov/ammem/amlaw/>

National Archives and Records Administration, Electronic Records Archive

<http://www.archives.gov>

National Digital Information Infrastructure and Preservation Program (NDIIPP)

<http://www.digitalpreservation.gov>

Public Access to Court Electronic Records (PACER) <http://pacer.psc.uscourts.gov>

Thomas <http://thomas.loc.gov>