

**American Library Association
Government Documents Round Table**

**REPORT
DIGITIZATION OF GOVERNMENT INFORMATION**

Submitted by:
**Ad Hoc Committee on Digitization
Of Government Information**

**Cathy Nelson Hartman
Committee Chair**

June 14, 2002

Contents:

Introduction	2
Committee Action Plan to Accomplish the Charge	3
Description of the Issues and Best Practices Report	6
Recommendations to GODORT	6
Conclusions	9
Appendix A: Charge to the Committee	A - 1
Appendix B: Committee Membership	B - 1
Appendix C: Issues and Best Practices Report.....	C - 1
Appendix D: Bibliography of Resources	D - 1
Appendix E: Digitization Projects Reported	E - 1

Introduction

In July 2000, a group of librarians¹ with a strong interest in digitizing government publications met at the ALA Conference in Chicago to discuss strategies for facilitating digital projects in libraries. The agenda for the meeting listed the following concerns for discussion: a clearinghouse for digital projects; possible resources needed to facilitate digitization of government publications; and how these issues could be addressed by documents librarians or organizations such as the ALA Government Documents Round Table (GODORT). Recognizing the extensive interest in these issues, the group decided to ask GODORT to sponsor an ad hoc committee to investigate issues related to digitization of government information and to suggest ways that GODORT might use its existing organizational structure to facilitate coordination of projects.

Ann Miller, the incoming GODORT Chair, presented a charge (See Appendix A) for an ad hoc committee at the GODORT second Steering Committee meeting at the ALA Conference, July 2000. The charge directed the Ad Hoc Committee on Digitization of Government Information to “create a report advising GODORT on the best practices and procedures in the digital library field and advise GODORT on the most effective organizational structure for support of the government information community in pursuing digital library initiatives.” Additionally, the Committee was asked to:

- Identify major issues in digital libraries.
- Examine best practices in digitization, including such issues as standards, technologies, methodologies, metadata, and authentication.
- Create a bibliography of relevant resources.
- Locate projects already completed or in progress.
- Identify key players in the digital fields.
- Seek out potential partners interested in cooperative projects to digitize government publications.
- Explore copyright implications.
- Identify potential funding sources and provide initial advice on grant acquisition.

Representatives from GODORT committees and task forces comprised the majority of the membership (See Appendix B) in the ten-member Ad Hoc committee. The GODORT Steering Committee approved the formation of the

¹The Chicago meeting was organized by Cathy Nelson Hartman and Luke Griffin. Persons present included: Ann Miller, Julia Wallace, Mary Mallory, Geoffrey Swindells, Barbara Levergood, Laura Dickson, Bill Sleeman, August Imholtz, and Bob Hook, the “Chicago 12.”

Ad Hoc Committee on Digitization of Government Information and approved the charge as drafted. Final approval for the committee membership came in November 2000, and the committee first met at the ALA Midwinter meeting in January 2001.

This document outlines the investigations of the Ad Hoc Committee and reports our findings in response to the charge from GODORT Steering. Included are:

- An explanation of the Committee's action plan to accomplish the charge,
- A description of the appended **Issues and Best Practices Report**,
- The recommendations of the Ad Hoc Committee regarding the most effective organizational structure for support of the government information community in pursuing digital library initiatives, and
- Several appendixes, including the **Issues and Best Practices Report**, an extensive bibliography of resources, text of the committee charge, and committee and work group membership lists.

Please note that the scope of this Committee's investigations regarding digital library initiatives focuses on issues directly related to the digitization of government information - federal, state, local, and international government publications. The broader world of digital library initiatives encompasses a much larger set of issues than are covered here. Additionally, the Committee believes the resources included in the report will be useful for persons involved in digital projects with skill levels from novice to expert.

Committee Action Plan to Accomplish the Charge

In preparation for the first meeting of the Ad Hoc Committee at ALA Midwinter, 2001, the committee chair prepared a proposed plan for review by the committee. The plan outlined an organizational structure consisting of four working groups to facilitate the group's work. The plan divided the topics for exploration as evenly as possible among the groups. The committee reviewed the proposal until all achieved a general understanding of the expected output of each group. The plan was accepted, but the committee recognized the unavoidability of some overlap of issues making communication between work groups important. A listserv and a Web site were created to enable communication.

Four working groups were formed, each with a work group leader from the original committee roster. Additional work group members were drawn from the committee roster and from volunteers anxious to assist with the work. Members

and volunteers were assigned to work groups based on their expressed areas of interest. All four work groups agreed to the following general guidelines and requirements:

- All would consider possible roles for GODORT in assisting and supporting libraries for digitization projects involving government publications.
- Each group would identify resources and build a bibliography of publications for use by the committee and others.
- Each group would identify any additional key issues in its area of examination that were not listed in the Work Group assigned topics.
- Each group would identify best practices or existing standards in its area of examination.
- Each group would create a list of the key players in its area of examination.

Additionally, assignments for each group outlined the specific issues to be undertaken for investigation. The assignments of the work groups included:

Work Group 1:

Work Group 1 members agreed to cover topics that were generally non-technical in nature. They investigated issues relating to:

- A. Selection of materials for digitization by:
 - Identifying projects in process.
 - Examining methods of coordinating efforts.
 - Identifying methods of sharing rare/endangered materials for digitization projects.
 - Reviewing GODORT's role in assisting such efforts and in coordinating/assisting the building of partnerships between libraries for digitizing government publications.
- B. Funding sources and costs by:
 - Identifying grant funding sources.
 - Developing methods for GODORT to sponsor partnerships for grant proposal writing.
 - Identifying resources for use by grant writers.
- C. Training and Staff development by:
 - Identifying issues.
 - Identifying resources.

Work Group 2:

Work Group 2's assignment related to the actual creation of digital copies. Made up of persons currently involved in digitization projects, their assigned tasks included the following activities:

- A. Identify best practices for digitization.
- B. Explore development of standards.
- C. Review file formats for current access and for long-term preservation (for various document types, i.e., textual, posters, photos, etc.).
- D. Create a list of vendors and explore outsourcing possibilities for digitization.
- E. Consider ADA issues related to access of digitized documents.
- F. Investigate "server-side" issues such as file management strategies and search engine software.

Work Group 3:

Work Group 3, comprised of persons with extensive experience managing digital projects, investigated three major issues related to preservation of electronic files. They examined:

- A. Storage and day-to-day maintenance of files by:
 - Determining impact/costs in hardware/software and personnel.
 - Determining current thinking on best practices for storage and maintenance.
- B. Long-term preservation of files by:
 - Reviewing changes in technology including strategies to plan for changes (digital reformatting, migration, emulation, etc.).
 - Examining issues of media degradation and file recovery.
- C. Authenticity and security issues.

Work Group 4:

This small group of highly experienced catalogers and metadata enthusiasts reviewed issues and best practices related to:

- A. Metadata
 - Reviewed metadata standards for access (Dublin Core, MARC).

- Reviewed preservation metadata elements and standards.

B. Copyright implications when digitizing government publications.

Each work group submitted a report to the Ad Hoc Committee Chair in November or December of 2001. The individual reports, as submitted, are available on the Ad Hoc Committee Web site. The combined and edited reports are attached as “Appendix C: Issues and Best Practices Report.”

Description of the Issues and Best Practices Report

The **Issues and Best Practices Report** addresses all areas of committee investigation and is attached as Appendix C. The Report follows the same organization outlined above in **Committee Action Plan to Accomplish the Charge**. Reports from the four work groups were combined and edited for style, and occasionally content was added. All Work Groups reported the issues and referred to best practices and developing standards, where they existed. The resulting document contains a wealth of resources and information for persons beginning a digitization project. The information is current as of May 2002. However, all groups found that the issues, best practices, and development of standards evolve continually as developing technology opens new avenues for innovation.

The Report includes recommendations when applicable, but, most often, the report describes “guidelines” for a successful digitization project. Best practices for a project strongly depend on the goals of the project and the type of material designated for digitization. Guidelines recognize the uniqueness of most projects and offer greater flexibility for the project manager during the creation of the project design. As Work Group 3 stated, “Perhaps the single best practice we recommend to digital project managers is to consult with local information technology resource staff and **keep abreast of developments in the broader digital arena.**”

For the convenience of the reader, each section of the Report contains a bibliography of resources. Also, a complete and separate bibliography of all resources is included as Appendix D.

Recommendations to GODORT

In the charge to the Ad Hoc Committee, GODORT Steering asked the Committee to advise GODORT on the most effective organizational structure for support of the government information community in pursuing digital library initiatives. Several recommendations grew out of the Committee investigations, but other

recommendations developed from issues brought to the Committee by librarians who personally experienced a gap in resources or assistance when planning a digital project. Outlined below are three (3) recommendations that the Committee believes will offer needed resources and assistance to librarians planning a digital imaging project for government publications. Most of the recommendations for action match the Government Information Technology Committee's (GITCO) purpose as outlined in the GODORT Policy and Procedures Manual. For this reason, it seems appropriate to assign most of the following tasks to that committee. However, recognizing the current and continuing scope of this project and GITCO's already full agenda, we also recommend that GODORT Steering create a GITCO Subcommittee, the GITCO Subcommittee on Digital Imaging (GITCO SDI). A GITCO member designated by the GODORT Chair would chair the GITCO SDI, and an additional 4 persons would be appointed by the GODORT Chair to serve on the subcommittee. Subcommittee membership does not require conference attendance and may offer flexibility with appointments. The SDI Chair would serve as a full member of GITCO and report SDI activities at GITCO meetings.

The Ad Hoc Committee recommends that GODORT:

1. Charge the Government Information Technology Committee (GITCO) to create and maintain a Web site with the following sections:
 - a. A bibliography of relevant resources section, arranged topically and based on the attached **Bibliography of Resources** (Appendix D). The Bibliography will require semi-annual updates to maintain currency with the existing topics, and new topic sections should be developed as new issues arise. The Bibliography will offer a reading list organized by topic area that can be printed and used as a handout or referenced on a Web site.
 - b. A section with detailed **models** of digital imaging projects for government publications. The section should comprise models contributed by GODORT members already involved in digital projects, and it should also link to other existing models. The models should include information about all aspects of a project, from selection, to planning for technology, costs models, preservation planning, and metadata. This section should also include links to high-quality project descriptions available on some digital projects' Web sites.
 - c. A list of volunteer mentors for digital projects, some possibly associated with the models proposed in (b). Persons beginning a first project benefit significantly from conversations with mentors.

- d. A database of digital projects that are completed, in progress, planned, or under consideration. This database would allow librarians to determine if others have already selected specific material for digital imaging, to locate partners for collaborative projects, or register a project when funded. The database should allow search and retrieval of records and data entry on a Web-based form. Work Group 1 began collecting information about projects and developed a list of fields to be included in such a database (Included in Appendix C, page C – 4). This work would serve as a basis for development of the database. The types of digital projects that should be included in the database are those that meet ALL the following requirements:
 - i. Projects undertaken by libraries (public libraries, libraries affiliated with educational institutions, or government agency libraries), government agencies, or tribal governments,
 - ii. Projects with a focus on federal, state, local, or international government publications or that use government produced data, and
 - iii. Projects that freely offer the resulting resources to the public.
 - e. A resource list of granting or funding agencies and grant writing resources based on the resources listed in the **Issues and Best Practices Report** (Appendix C) and a list of volunteer grant writing mentors. Funding for digital imaging projects is often the last piece in moving a digital imaging project forward. Funding resources and mentors for grant proposal writing would offer an important resource to the first-time grant writer.
2. Encourage and facilitate cooperative digital imaging projects that may be too extensive for any one library, such as the U.S. Congressional Serial Set or the Congressional Record. GODORT should charge GITCO to develop a plan or locate volunteers to spearhead such major projects. Communication with other groups that have similar interests, such as the Library of Congress (LC) and the U.S. Government Printing Office, is encouraged during this process. Grant proposals could be submitted to such major funding sources as IMLS, the Mellon Foundation, or possibly the National Digital Information Infrastructure Preservation Program (at LC) on behalf of multiple libraries who might participate in such a major project. GITCO should assist with solving issues for collaborative projects such as Web access to the data (all files on one server or a gateway configuration), preservation planning for all sites, or technical issues with file management or search and retrieval. All GODORT task forces and

committees should assist with publicity about collaborative projects and solicit partners for such projects.

3. Sponsor an ALA preconference and investigate sponsoring regional institutes or workshops on the topic of digital imaging for government publications. GODORT's Program Committee, in consultation with the Education Committee and GITCO should build a full-day workshop on the topic of digital imaging with a focus on government publications to be presented at the ALA Annual Conference in 2003 in Toronto. The high level of interest in digital imaging was clearly demonstrated by the impressive attendance at the 8 PM meetings of the Ad Hoc Committee at the last two ALA conferences in addition to interest shown by e-mail and phone contacts.

Regional institutes or workshops could offer opportunities to many librarians who may not be able to attend the ALA Conference in Toronto. Currently, the Library and Information Technology Association (LITA) is sponsoring highly successful regional institutes in cooperation with Amigos and other OCLC regional groups on technology topics for libraries. GODORT should investigate this LITA model and consider taking the program planned as an ALA preconference "on the road" to different regions of the U.S. and, potentially, Canada. Registration fees could cover the cost of travel for presenters or grant funding could be solicited for the project.

Conclusions

The strong interest in digitizing government publications was demonstrated repeatedly over the past two years of Ad Hoc Committee investigations. Attendance at committee meetings was excellent with an average of 35 persons present at 8 PM meetings at the 2001 ALA Midwinter and Annual Conferences. Additionally, the committee chair received numerous phone calls and electronic mail from librarians seeking information about digital imaging and about the work of the committee. The need for the resources and educational opportunities described in the recommendations above became clear to the committee, and the specific need for mentors to assist persons beginning a first digitization project was often expressed. The resources created by the committee during our investigations offer a good beginning for a permanent GODORT Web site outlining issues, best practices, and resources for digital imaging projects. Implementation of the recommendations of the committee should fill an expressed need and bring positive attention to GODORT for educating librarians about digitization and as a player in the digital library field by facilitating large digital imaging projects for government publications.

Appendix A: Charge to the Committee

Ad Hoc Committee on Digitization of Government Information

Approved at ALA GODORT Second Steering, July 9, 2000, Chicago

MEMBERSHIP:

Ten (10) members. Representatives from GITCO, REGP, Cataloging, Education, FDTF, SLDTF, IDTF, 3 at large members (including possible representation of GODORT members from GPO and ARL libraries).

THE CHARGE:

To create a report advising GODORT on the best practices and procedures in the digital library field and advise GODORT on the most effective organizational structure for support of the government information community in pursuing digital library initiatives.

Topics to be covered include:

- Identifying major issues in digital libraries.
- Identify best practices in digitization. Including principles of identifying projects, standards, technologies, methodology, cataloging and authentication issues, and create a recommended bibliography.
- Determine projects already completed or in process.
- Identify and explore key players in the digital fields, e.g., Library of Congress, Association of Research Libraries, and the Digital Library Federation, identify potential partnerships.
- Determine copyright implications.
- Determine grant sources and provide initial advice on grant acquisition.

DURATION:

Two (2) years with an option for extension.

Appendix B: Committee Membership

Cathy N. Hartman (Chair)
Head, Government Documents Dept.
UNT Libraries
P.O. Box 305190
Denton, TX 76203-5190
Phone: 940-565-3269
Fax: 940-565-2599
E-mail: chartman@library.unt.edu

Luke Griffin (GITCO Rep.)
(Committee Web Administrator)
Digital Services/Government
Docs/Ref. Librarian
Government Information
Paul V. Galvin Library
35 West 33rd St.
Chicago, IL 60616
Phone: (312) 567-3613
Fax: (312) 567-5318
E-mail: griffinL@iit.edu

John B. Phillips (REGP Rep.)
(Committee Secretary)
Head, Documents Dept.
Edmon Low Library
Oklahoma State University
Stillwater, OK 74078-0375
Phone: (405) 744-6546
Fax: (405) 744-5183
E-mail: bart@okstate.edu

Sandra K. Peterson (FDTF Rep)
(Leader, Work Group 1)
Government Documents &
Information Center
Yale University
Box 208294, 38 Mansfield St.
New Haven, CT 06520-8294
Phone: (203) 432-3212
Fax: (203) 432-3214
E-mail:
Sandra.K.Peterson@yale.edu

Suzanne L. Holcombe (Ed. Comm.
Rep) (Leader, Work Group 2)
Documents Librarian
Edmon Low Library
Oklahoma State University Library
Stillwater, OK 74078-0375
Phone: (405) 744-6546
Fax: (405) 744-5183
E-mail: sholcom@okstate.edu

Chuck Eckman (IDTF Rep)
(Leader, Work Group 3)
Principal Government Documents
Librarian & Head, Social Sciences
Resource Group
Green Library, SSRC 123A
Stanford University
Stanford, CA 94305-6004
Phone: 650-723-2982
Fax: 650-725-9348
E-mail:
ceckman@sulmail.stanford.edu

Elizabeth B. (Betsy) Richmond (Cat.
Comm. Rep.)
(Leader, Work Group 4)
Reference Librarian
McIntyre Library
University of Wisconsin - Eau Claire
Eau Claire, WI 54702-4004
Phone: (715) 836-4076
Fax: (715) 836-2949
E-mail: richmoeb@uwec.edu

Anne Watts (SLDTF Rep)
Special Projects Librarian
St. Louis Public Library
1301 Olive Street
St. Louis, MO 63103
Phone: (314) 539-0377
Fax: (314) 539-0356
E-mail: awatts@slpl.lib.mo.us

George Barnum (GPO
Representative)
Electronic Collection Manager
Library Programs Service, SLL
United States Government Printing
Office
Washington, D.C. 20401
Phone: (202) 512-1114
Fax: (202) 512-1432
E-mail: gbarnum@gpo.gov

Laura Dickson (at-large) (resigned
9/01)
U.S. Documents Librarian
Michigan State University
100 Main Library
East Lansing, MI 48824
Phone: (517) 432-8045
Fax: (517) 432-1191
E-mail: dickso15@mail.lib

Barbara Levergood (at-large)
(resigned 12/01)
Electronic Documents Librarian
Documents Section, Reference
Department, CB #3912
Davis Library
University of N. Carolina, Chapel Hill
Chapel Hill, NC 27514-8890
Phone: (919) 962-1151
Fax: (919) 962-5537
E-mail: leverg@refstaff.lib.unc.edu

Work Group Members:

Work Group I –

Sandy Peterson, Leader
Carol Watts
Barbara Levergood
John Phillips
Anne Watts
Julia Wallace
Susan Tulis
Amy West
Valerie Glenn
David Larkin

Work Group II -

Suzanne Holcombe, Leader
Luke Griffin
Elaine Winske
Donna Koeppe
Anthony E. Anderson

Work Group III -

Chuck Eckman, Leader
George Barnum
Patricia Cruse
Earl Shumaker
Amy West

Work Group IV -

Betsy Richmond, Leader
Nan Myers
Daniel Alemneh
Edward Kownslar
Kirsten Clark
Lisa Robinson

Liaison from the Library of Congress

Edward J. Malone
Head, Government Documents Section
Anglo-American Acquisitions Division
Library of Congress
101 Independence Ave., SE
Washington, D.C. 20540-4172
Phone 202-707-9470
Fax: 202-707-0380
E-mal@loc.gov

Appendix C: Issues and Best Practices Report

Edited by Cathy Nelson Hartman, Ad Hoc Committee Chair

Contributors to this report were: Sandra Peterson, Valerie Glenn, Suzanne Holcombe, Luke Griffin, Elaine Winske, Donna Koepp, Anthony E. Anderson, Chuck Eckman, Patricia Cruse, Amy West, Elizabeth Richmond, Kirsten Clark, Edward Kownslar, and Edward J. Malone (Library of Congress Representative). Additionally, Patricia Cruse contributed with the final editing and writing of sections of the draft report. Suggestions submitted during the comments period were integrated into the report. Comments were submitted by: Suzanne Holcombe, Sandy Peterson, Valerie Glenn, Elaine Winske, Kirsten Clark, Elizabeth Richmond, Suzanne Graham, Ann Branton, Amy West, Ann Miller, Sandra McAninch, Esther Crawford, Daniel Gelaw Alemneh, and Chuck Eckman.

This appendix captures the issues and best practices in the field of digitization at the time of the release of this report. Since this section of the report will not be updated, future readers should refer to the bibliography of resources (currently shown as Appendix D). The bibliography will be regularly updated by the GODORT Government Information Technology Committee.

SECTION I.

Work Group 1 – Selection of materials, training, funding, collaboration

Introduction

Work Group I was charged to review issues surrounding selection of materials for digitization, funding resources, and training and staff development. The group reviewed Internet sites and reports, interviewed selected digital project managers, surveyed the government documents community for projects planned, underway and/or completed, and incorporated information provided at a variety of American Library Association programs on digitization.

I. A. Area of Investigation: Selection

The Work Group reviewed many web sites and reports that outlined selection criteria for digital imaging. Three of the most comprehensive and/or useful sites for an overview are:

Columbia University Libraries Selection Criteria for Digital Imaging
(<http://www.columbia.edu/cu/libraries/digital/criteria.html>)

Joint RLG and NPO Preservation Conference, Guidelines for Digital Imaging.
Guidance for selecting materials for digitization, by Paul Ayris, Director of
Library Services, University College, London.
(<http://www.rlg.org/preserv/joint/ayris.html>)

Selection for Digitizing A Decision-Making Matrix.
(<http://www.clir.org/pubs/reports/hazen/matrix.html>)

Using these three documents as models, the group devised criteria particularly suited to selecting government information for digitization. They are:

Criteria for Selecting Materials for Digitization

1. Intellectual and physical nature of the source materials -
 - 1.1. Copyright assessments.
 - 1.2. Material has a preservation problem (i.e., risk of damage or loss due to high use, poor housing, or physical deterioration).
 - 1.3. Provides access to materials that cannot otherwise be easily used.
 - 1.4. Condition of originals allows them to be digitized safely.
 - 1.5. Degree to which a digital version can represent the full content of the original.
2. Number and location of current and potential users -
 - 2.1. Does the material have sufficient intrinsic value to ensure interest in a digital product?
 - 2.2. Would digitization make a contribution to local or national electronic resources
 - 2.3. Would digitization significantly enhance access, or increase use by an identifiable constituency.
3. Current and potential nature of use -
 - 3.1. Decisions must anticipate how changes in technology could enhance or make obsolete an investment in technology.
 - 3.2. If the digital resources are deemed to be of long-term value, provision must be made for archiving and maintenance of images through time and technological change.

- 3.3. An active, current audience for the materials; OR a realistic expectation of attracting new users even if current use is low.
 - 3.4. The permanent value of the data or product.
 - 3.5. Is this digitization for preservation or digitization to enhance access?
 - 3.6. Can we deliver legible images conveniently through commonly used equipment and software?
4. Format and nature of the proposed digital product and how it will be described, delivered, and archived -
 - 4.1. Does current technology yield image quality adequate to meet stated goals?
 - 4.2. Anticipation of future users with better equipment, to avoid a need to rescan in a few years.
 - 4.3. Authentication of images, if appropriate.
 5. How the proposed product relates to other digitization efforts -
 - 5.1. There is no identical or similar digital product that can reasonably meet the expressed needs.
 - 5.2. Collaboration possibilities.
 6. Projection of costs in relation to benefits -
 - 6.1. Do the benefits of digitization justify the costs?
 - 6.2. File size, associated storage needs, processing requirements, labor requirements.
 - 6.3. Long-term file management.
 - 6.4. Does the institution have the necessary expertise and resources to plan and implement the project?
 - 6.5. Staff and resources to support creation of appropriate metadata relating to document identification, technical capture information, provenance, and easy navigation within the information resource.

Bibliography of Resources About Selection:

Cornell University. *Moving Theory into Practice: Digital Imaging Tutorial, Part 2: Selection.* (<http://www.library.cornell.edu/preservation/tutorial/selection/selection-01.html>)

Outlines key concepts, addresses legal restrictions, other criteria, selection policies and additional reading

Council on Library and Information Resources. *Why Digitize?* By Abby Smith, February 1999. (<http://www.clir.org/pubs/reports/pub80-smith/pub80.html>)

Discusses what digital information is and is not.

Hazen, D., Horrell, J., & Merrill-Oldman, J. *Selecting Research Collections for Digitisation*. Amsterdam: European Commission on Preservation and Access, 1998.

Joint RLG and NPO Preservation Conference, Guidelines for Digital Imaging, 28-30 September 1998. Selection Criteria, Guidelines, Decision-Making Aids (<http://www.rlg.org/preserv/joint/selection.html>)

Links to general studies and discussions of selection criteria.

Joint RLG and NPO Preservation Conference, "Guidelines for Digital Imaging. Guidance for Selecting Materials for Digitisation," by Paul Ayris, Director of Library Services, University College, London. (<http://www.rlg.org/preserv/joint/ayris.html>)

Outlines the questions that need to be addressed in formulating guidance.

Library of Congress. Selection Criteria for Preservation Digital Reformatting (<http://lcweb.loc.gov/preserv/prd/presdig/presselection.html>)

Discusses principles on which selection of materials is based.

Handbook for Digital Projects: A Management Tool for Preservation and Access (<http://www.nedcc.org/digital/dighome.htm>)

Chapter IV focuses on Selection of Materials for Scanning, written by Diane Vogt O'Connor.

Preserving Access to Digital Information (PADI). Policies, Strategies and Guidelines (<http://www.nla.gov.au/padi/format/policy.html>)

Lists and links to international, Australasia, Europe, and North American resources which provide policies, strategies and guidelines for digital preservation.

University of California, Santa Barbara. University of California Selection Criteria for Digitization. (<http://www.library.ucsb.edu/ucpag/digselec.html>)

List of criteria recommended to guide collection development librarians and preservation librarians in selecting collections of analog materials for conversion to digital format.

Digital Library Forum. "A Framework of Guidance for Building Good Digital Collections." Draft for Comment. Institute of Museum and Library Services, 2001. Available from: <http://www.ims.gov/pubs/forumframework.htm>.

I. B. Area of Investigation: Clearinghouse / Registry of Digital Projects

One of the issues that the Group considered was the desirability of a shared registry of projects that were planned, underway or completed. A form for providing information to librarians and others about digitization projects for local, state, Federal and international government documents that are currently planned, in progress, or already completed was developed (<http://www.gliit.edu/services/ref/diggovclearinghouse.htm>). At this point, several libraries have entered data about their projects on the site. A list of the currently

registered projects is attached to this report as Appendix E. The data fields for collection of information about projects, as suggested by the Group, are:

- Institution,
- Contact Person (name),
- Phone and/or Email Address,
- Project Title,
- Project Description (agency, time period covered, specific titles, other information),
- Subject,
- SuDoc Numbers or Range (Federal projects only),
- Are you looking for partners for collaboration (yes, no, maybe)?
- Project Status:
 - Proposed, not yet funded
 - Proposed and funded
 - Not Yet Complete (Not ready for public)
 - In Process and Public
 - Complete
- Sponsoring Institution(s)
- Project URL, or Current URL
- Metadata to be used, if any
- File Format (TXT, PDF, JPEG, GIF, TIFF, other)
- Contracting work? If yes, what contractor?
- Other Comments
- Commercially available (fee based)?

Other existing sites link to digital projects, but none focus solely on government information. For example, the FAQ of the RLG DigiNews, December 15, 2000 (<http://www.rlg.org/preserv/diginews/diginews4-6.html#faq>) lists several such listings of digital projects. Some of the registries reflect “collection level” projects; others provide or plan to provide information about faithful reproductions of books and serials originally published in paper format (Registry of Digital Reproductions of Paper-based Books and Serials, proposed by the Digital Library Federation). “Faithful reproductions” are defined as copies intended to preserve the original

appearance of published materials, and must include digital images of all pages in the originals.

To facilitate collaboration and communication in the documents community, a clearinghouse of government documents digital projects seems appropriate. Such a clearinghouse database would ideally be Web based, allowing data entry and searching from a Web page. Such a database would provide librarians, who may be planning a digital project, a discovery tool to see if others are digitizing the same material or to locate other libraries interested in cooperative projects. GODORT should also continue to monitor the development of other such registries as that proposed by the Digital Library Federation.

I. C. Area of Investigation: Collaborative Projects

The following issues affect collaborative projects:

- o Awareness – how do institutions and/or individuals find out about potential collaborators?
- o Audience – are potential partners intending this for different audiences?
- o Scanning – parties in collaborative efforts must determine which will complete each part of the project? If training is necessary, can one collaborator provide the training for the other?
- o Access & storage – where will the information be stored for preservation and for access, i.e., at multiple sites and accessed via a gateway or at only one of the institutions?
- o Project funding – how will the project be funded, i.e., a cooperative grant writing project, institutional funding with both parties contributing funds, or other?
- o Communication – do all parties have a common understanding of the terminology used in a project or does language need to be defined?

GODORT seems to be uniquely positioned to serve as a forum and facilitator for collaborative projects and for exchange of information about digitization projects involving government information. Libraries in several states (Colorado, California, Oregon, Washington) are involved in regional digitization projects; some of these projects involve collaboration among libraries and museums and archives.

Bibliography of Resources About Collaboration:

Colorado Digitization Project. *Collection Policy*.

<http://coloradodigital.coalliance.org/select.html>

Defines the guiding policies of the Project, the audience, subject matter, and participants.

Stack, Jacqueline M. "Collaboration: Success for the Future." Los Alamos National Laboratory, 1999. Available from:

<http://www.osti.gov/inforum99/papers/stack.pdf>.

Dowden, Robin, et al. "*ArtsConnectEd: Collaboration in the Integration and Access to Museum Resources*." Available from:

<http://www.ims.gov/pubs/wwcp2.htm>.

Case study of collaboration between three museums. Discusses the project management and evaluation, as well as future issues. A sizable portion of the report is devoted to audience needs and communication among collaborators.

Allen, Nancy. "Collaboration through the Colorado Digitization Project." Available from: <http://www.ims.gov/pubs/wwcp5.htm>.

Discusses the digital efforts emerging from collaboration among school, public, and university libraries, museums, archives, and historical societies from all over the state of Colorado. Issues involved in this large collaboration include varying audiences, resources, communication, legal matters and preservation.

I. D. Area of Investigation: Funding

Work Group 1 identified several types of possible funding sources for projects--- Federal grants, Federal grants through state agencies, foundations, state grants, and institutional support. The Group also identified grant writing resources that are available online.

Bibliography of Resources About Funding Sources:

Colorado Digitization Project Funding Sources

<http://coloradodigital.coalliance.org/funding.html>)

Identifies public/federal sources, associations and organizations, private foundations, and information resources. Includes links to many of the sources listed below.

OCLC Preservation Organizations and Funding

<http://www.oclc.org/oclc/presres/funding.htm>

Provides links to preservation organizations, funding organizations, as well as links to preservation-related electronic publications, discussion groups, and mailing lists.

The institutions listed below have funded digitization projects before, or have stated their willingness to fund projects that use digital technology to preserve items.

Institute of Museum and Library Services (<http://www.ims.gov>)

Application forms and program guidelines may be downloaded from the IMLS web site. Brief descriptions of specific grant programs sponsored by IMLS include:

- *National Leadership Grants/Education and Training:* Funds projects that ...implement innovative approaches to education and training and enhance the availability of professional librarians with advanced skills and specializations...
Deadline: February 1.
- *National Leadership Grants/Research and Demonstration:* Model projects that conduct research and/or demonstrations to enhance library services through the effective use of new and appropriate technologies, enhance the ability of library users to make more effective use of information resources, or assist in the evaluation of library services...
Deadline: February 1
- *National Leadership Grants/Preservation or Digitization:* Projects that preserve and enhance access to unique library resources useful to the broader community; address the challenges of preserving and archiving digital media; or that lead to the development of standards, techniques, or models related to the digitization and management of digital resources. (Georgia Dept. of Archives & History was a 2000 recipient to digitize and make available Georgia's House and Senate journals and other state government records)
Deadline: February 1
- *Grants to States:* The Library Services and Technology Act (LSTA) of 1996 promotes access to information resources provided by all types of libraries. Two broad priorities for LSTA funding: 1) activities using technology for information sharing among libraries and between libraries and other community services; and 2)

programs that make library resources more accessible to urban and rural localities, and to low-income residents and others who have difficulty using library services. Funding is distributed by state library agencies; the amount of money available is based on the population of the state.

Deadline: New five-year plans due July 31, 2002

To obtain more information on the types of projects that IMLS funds, see: http://www.ims.gov/closer/cls_po.asp The Library Program Officer for IMLS is Martha Crawley, mcrawley@ims.gov.

National Endowment for the Humanities (NEH)
(<http://www.neh.gov/grants/index.html>)

The Division of Preservation and Access supports “applications for model projects that explore or resolve critical issues relating to the effective use of digital technology for preservation and access. The web site includes a “frequently asked questions” about grants page. Contact: info@neh.fed.us

National Historical Publications and Records Commission (NHPRC)
(<http://www.nara.gov/nhprc/>)

The NHPRC is the grant-making affiliate of the National Records and Archives Administration. NHPRC supports grants for the identification, preservation, and provision of public access to materials that document American history. For an indication of the type of projects funded, see <http://www.nara.gov/nhprc/projects.html>
Contact: nhprc@arch1.nara.gov

National Endowment for the Arts (NEA) (<http://www.nea.gov>)

NEA offers support for organizations in four areas---creation and presentation, planning and stabilization, heritage and preservation, and education and access. It supports projects that assist, preserve, document and present artists and forms of artistic expression that reflect our nation’s diverse cultural traditions. Funding for conservation of important work is also possible. Contact: 202-682-5400

Andrew W. Mellon Foundation (<http://www.mellon.org>)

Private foundation provides funding for projects for art conservation research and training as well as for the digitization of library and archival materials. Program Officer, Scholarly Communication: Don Waters, djw@mellon.org

Getty Grant Program (<http://www.getty.edu/grants>)

A possible source of funding is the research grant.

Other possibilities are state library agencies and foundations with regional or local focus. The *Foundation Directory*, available in many libraries, or online for a subscription fee, is another excellent source for identifying foundations that provide financial support to educational institutions for a variety of purposes.

Bibliography of Grant Writing Resources

Proposal Writing Short Course

(<http://fdncenter.org/learn/shortcourse/prop1.html>)

Step-by-step guide to writing a proposal with a short discussion of each section and what should be included.

Frequently Asked Questions

(<http://www.fdncenter.org/learn/faqs/propsample.html>)

A page from the Foundation Center that includes sections on finding funding, writing proposals, and examples of grant proposals.

Many of the funding sources' web sites also include tips for proposal writing.

GODORT's Role in Facilitating Collaborative Projects

Article II of the GODORT Bylaws states:

“The purposes of the Government Documents Round Table are: (a) to provide a forum for discussion of problems and concerns, and for the exchange of ideas by librarians working with government documents; (b) to provide a force for initiating and supporting programs to increase availability, use and bibliographic control of documents; (c) to increase communication between documents librarians and other librarians; (d) to contribute to the extension and improvement of education and training of documents librarians.”

All of the stated purposes of GODORT support an involvement that encourages the facilitation of collaborative projects among documents librarians and libraries. GODORT's long-standing role should continue into the digital environment, just as it has for the past 30 years. A clearinghouse of digital projects and other forums for communication, including *Documents to the People*, the GODORT web site, workshops, and preconferences, should be used to foster collaborative projects for digitization of government publications.

1. E. Area of Investigation: Training and Staff Development

In a report by Margaret Hedstrom and Sheon Montgomery², the lack of staff expertise was identified as a common problem both in institutions with digital preservation responsibilities and in institutions that have not yet assumed responsibility for digital materials. In the three years since this report was published, staff expertise has increased. Professional associations and coalitions, such as Solinet, Research Libraries Group, Northeast Document Conservation Center, Digital Library Federation, Multi-Media Educational Resource for Learning and Online Teaching (Merlot), Colorado Digital Project, and the American Library Association and its various divisions and round tables, have and are providing training, workshops, online guides, and tutorials.

The skills that are necessary for a successful project range from the traditional library skills (selection, creating metadata) to a new set of technical skills (scanning, web design, file management, preservation of electronic files). Additionally, and perhaps most importantly, a project manager must develop strong project management skills. A review of existing projects indicates that individual librarians have developed many of their skills through a variety of methods including online guides and tutorials, workshops, and colleagues.

The assigned issues for investigation concerning training and staff development were:

- How can GODORT contribute in this effort?
- What is the role for the Government Printing Office?

GODORT task forces and committees can contribute to the training and development of librarians' skills for digitizing collections. GODORT should conduct and/or sponsor workshops, programs and preconferences on digitizing issues and best practices or consider partnering with other ALA divisions and units to sponsor such training opportunities.

The Government Printing Office began addressing this issue through sessions at the annual Depository Library Conference. The most recent conference (October 2001) included sessions on funding (speakers from the Institute of Library and Museum Services) and project descriptions (speakers from the Documenting California project and others). GPO should be encouraged to continue these

² Hedstrom, Margaret and Sheon Montgomery (December 1998). Digital Preservation Needs and Requirements in RLG Member Institutions. Research Libraries Group. Available: (<http://www.rlg.org/preserv/digpres.htm>) [2002, May 13].

educational efforts and to use the *Administrative Notes* to communicate information about other projects involving digitization of depository materials.

Bibliography of Training Resources:

Project Management Seminars. American Management Association.
<http://www.amanet.org/seminars/cmd2/Project.htm>

AMA sponsors seminars for project managers at three levels. Courses are expensive.

Project Management Research on the Web. Maintained by Umeå School of Business and Economics, Sweden. <http://www.fek.umu.se/irnop/projweb.html>

Provides links to research, universities, project management indices, and professional associations.

Jones, Maggie and Neil Beagrie. "Preservation Management of Digital Materials: A Handbook." London: The British Library for Resource, the Council for Museums, Archives and Libraries, 2001.

Includes chapters on institutional strategies and organizational activities.

Moving Theory into Practice: Digital Imaging Tutorial, Chapter 10, Continuing Education

(<http://www.library.cornell.edu/preservation/tutorial/education/education-01.html>).

Provides links to tutorials, workshops, and guides.

Colorado Digitization Project: Workshops & Seminars

(<http://coloradodigital.coalliance.org/workshop.html>).

Listing with description of seminars and workshops on various aspects of digitization.

Illinois Digitization Institute (<http://nautilus.outreach.uiuc.edu/ldi/workshop.asp>)

Offers a series of free one-day workshops on the fundamentals of digitization.

RLG Tools for Digital Imaging (<http://www.rlg.org/preserv/RLGtools.html>)

Includes worksheets and guidelines.

Appendix E: Digitization Methods, by Stuart D. Lee, March 1999.

(<http://www.bodley.ox.ac.uk/scoping/digitization.html>)

Outlines some of the main issues surrounding digitization. Provides recommendations applicable to the strategic planning level.

Building Digital Collections: Technical Information and Background Papers.
(<http://memory.loc.gov/ammem/ftpfiles.html>)

Includes information on technical specifications, workflow and production.

HATII (Humanities Advanced Technology and Information Institute) Digitisation Summer School, Glasgow. <http://www.hatii.arts.gla.ac.uk/SumProg/> .

SECTION II.

Work Group 2 - Creation of Digital Copies

The areas of investigation assigned to Work Group 2 included:

- Identify best practices for digitization.
- Explore development of standards.
- Review file formats for current access and for long-term preservation (for various document types, i.e., textual, posters, photos, etc.).
- Create a list of vendors and explore outsourcing possibilities for digitization.
- Consider ADA issues related to access of digitized documents.
- Investigate “server-side” issues such as file management strategies and search engine software.

Introduction

The recommendations offered in this section are based on guidance offered in literature, manuals, and documentation from key digitization organizations and centers in the United States. As the leaders in the digital imaging field, their research builds the groundwork for the digitization process. Digital projects should work with existing standards and best practices to ensure quality, authority, and accessibility whether the material selected for digitization is government information or other publications. However, certain issues, such as copyright, may require less attention during project planning for digitization of government publications, but some areas may need more attention, such as coordination with government agencies or other depository libraries.

Other major initiatives within the federal community should be watched for development of standards and best practices. For example, the XML.gov project <<http://www.xml.gov/>> proposes to “facilitate the efficient and effective use of XML,” and the National Coordination Office for Information Technology Research and Development <<http://www.itrd.gov/iwg/pca/lsn.html>> is a government-sponsored organization whose mission is to formulate and promote the Federal Information Technology Research and Development Program to meet national goals.

II. A. Area of Investigation: Digitization Process

Issue:

What is the best process for creating the digital file? Scanning the document using scanning equipment or photographing it using a digital camera are the

methods most often used to create the digital file. The resulting file is then saved to a digital medium (hard drive, CD-ROM) as an image file. The file may be used for archival purposes and/or viewing on the Web. The intended use of the file and the type of information product (pages of text, photograph, poster) determine the technical specification for the digitization process, including the SPI (samples per inch) and bit depth requirements.

Best Practices:

Best practices for digitization have been explored by many of the sources listed in the bibliography. The resources noted here were selected because they describe the processes that need to be followed in order to take a project from its genesis through to completion. Each set of guidelines should be read and compared by potential digitization centers, and policies should be adopted that best meet the goals and abilities of the institution.

The Colorado Digitization Project's General Guidelines for Scanning <<http://coloradodigital.coalliance.org/scanning.html>> outlines the digitization process very comprehensively, beginning with the system components that are required (hardware and software), training, recommended standards for different types of information products, file formats, and other considerations including quality control, file naming, and costs.

Sections six and seven from the *Handbook for Digital Projects: A Management Tool for Preservation and Access* <<http://www.nedcc.org/digital/dighome.htm>>, offer "Technical Primer" and "Best Practices" sections. Section six "exposes readers to the technical terminology and concepts of the digitization process. Specifically, it provides basic technical information related to digitizing library collections, archival holdings, and other materials from cultural institutions." Section seven describes six case studies that take the reader from the theoretical views of how digitization should be conducted to the actual practice of planning, executing, and evaluating projects. Some of the sections focus primarily on the experiences of one institution, while other sections are composites of what has been learned from various situations.

The Department of Preservation and Conservation at the Cornell University Library makes an excellent tutorial available online: Moving Theory Into Practice: Digital Imaging Tutorial <<http://www.library.cornell.edu/preservation/tutorial/toc.html>>. Key sections include: Section 3, Conversion (Scanning Factors, Rich Digital Master, Benchmarking, Proposed Method, Guidelines); Section 4, Quality Control (Developing a Program, Assessing Quality); and Section 6, Technical Infrastructure (Image Creation).

The Digital Library Federation's *Guides to Quality in Visual Resource Imaging* <<http://www.rlg.org/visguides/>> contains a set of guidelines for digital imaging projects organized by a board of experts from the Digital Library Federation, the Council on Library and Information Resources, and the Research Libraries Group. Recommendations are included for such topics as: Planning an Imaging Project, Selecting a Scanner, Imaging Systems: the Range of Factors Affecting Image Quality, Measuring Quality of Digital Masters and File Formats for Digital Masters. This site is part of the greater Digital Library Federation's Digital Library Standards and Practices <<http://www.diglib.org/standards.htm>>.

See also the publication *Moving Theory into Practice: Digital Imaging for Libraries and Archives* (Kenney and Reiger) from the Research Libraries Group. Chapters three and four contain extensive information about digital benchmarking that is defined as a management tool “designed to lead to informed decision making about a range of choices.” Both chapters offer extensive information to assist with decision making about the digital imaging process.³

Bibliography of Digitization Process Resources

Archive Builders

<http://www.ArchiveBuilders.com/>

Provides an in depth understanding of document imaging as used in records management. Presentations are given on the computer industry as it relates to document imaging systems and records management for specific sites. Includes white papers, courses, handouts from presentations, etc.

Building Digital Collections: Technical Information and Background Papers (Library of Congress)

<http://memory.loc.gov/ammem/ftpfiles.html>

Documents technical activities related to the American Memory Collections produced by the National Digital Library Program (NDLP) at the Library of Congress. Includes Technical Practices, Workflow and Production, Rights and Restrictions Statements, Background Papers.

³ Kenney, Anne R. and Oya Y. Reiger (2000). *Moving Theory into Practice Digital Imaging for Libraries and Archives*, Research Libraries Group, page 24.

Colorado Digitization Center, Digital Toolbox

<http://coloradodigital.coalliance.org/toolbox.html>

Designed to guide administrators of digital projects to sites and resources on the following topics: Administrative Resources, Guidelines and Standards, Copyright and Intellectual Property, Funding, Metadata, Software, Planning, etc.

CoOL (Conservation OnLine) Digital imaging: Imaging and Imagebases

<http://palimpsest.stanford.edu/bytopic/imaging/>

Creating and Documenting Electronic Texts: A Guide to Good Practice. AHDS Guides to Good Practice.

<http://ota.ahds.ac.uk/documents/creating/>

D-Lib Forum

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

Facilitating and supporting the community developing the technology of the global digital library. See also D-Lib magazine.

Digital Imaging: Selected Bibliography. Solinet Preservation Leaflets.

<http://www.solinet.net/presvtn/leaf/imaging.htm>

Digitising History: A Guide to Creating Digital Resources from Historical Documents. AHDS Guides to Good Practice.

http://hds.essex.ac.uk/g2gp/digitising_history/index.asp

Digitizing Technologies for Preservation : a SPEC Kit / Kellerman, Lydia Suzanne; Wilson, Rebecca A.; Rounds, Laura; ed. Association of Research Libraries Office of Management Services, 1996. 173 p., map.

Guidelines for Digital Imaging (RLG and National Preservation Office)

<http://www.rlg.org/preserv/joint/confpapers.html>

Conference papers from a Joint RLG and NPO Preservation Conference: Guidelines for Digital Imaging, 28-30 September 1998. Includes Guidelines Versus Guidance for Digital Imaging: the Opportunity Before Us, Guidelines for Image Capture, etc.

Handbook for Digital Projects: A Management Tool for Preservation and Access. Maxine K. Sitts, ed. Northeast Document Conservation Center, 2000. 179 p.

<http://www.nedcc.org/digital/dighome.htm>

Introduction to Imaging: Issues in Constructing an Image Database. Besser, H. and Trant, J. 1995. [Santa Monica, CA: The Getty Art History Information Program).

Coppock, T., ed. (1999). *Making Information Available in Digital Format: Perspectives from Practitioners*. Edinburgh: The Stationery Office.

Lavender, Kenneth et. al. (2000). *Managing Preservation Technologies For the New Century*. Chicago, American Library Association. [Distributed by Teach'em], 2 sound cassettes.

Kenney, Anne R. and Rieger, Oya Y. (2000) *Moving Theory Into Practice : Digital Imaging for Libraries and Archives*. Mountain View, CA : Research Libraries Group.

Oxford Text Archive

<http://ota.ahds.ac.uk/>

Works closely with members of the Arts and Humanities academic community to collect, catalogue, and preserve high-quality electronic texts for research and teaching. See publications.

RLG DigiNews

<http://www.rlg.org/preserv/diginews/>

Selecting research collections for digitization: applying the Harvard model. Kristine R Brancolini. *Library Trends* (Urbana). v. 48, no. 4 (Spring 2000), p. 783-798.

The Seven Deadly Sins of Digitization. Online products being marketed to libraries. *Online* (Weston, Conn.). v. 23, no. 2 (Mar./Apr. '99) p. 43-6.

II. B. Area of Investigation: Development of Standards

Issues:

Choosing an image file format is one of the many technical decisions that must be made in the digitization process. The file format should be flexible and powerful to accommodate a variety of desired uses and should ensure long-term access to digitized information. The file format preferred is nonproprietary, officially endorsed by an international standards-making organization, and widely supported by computer software applications. However, de facto standards may evolve when company-developed products or specifications become widely

used. There are many de facto standards in the imaging field.⁴ Software support of a file format is absolutely critical to successful implementation.

File formats are the “containers” for the encoding of data. They must specify whether the format is binary or ASCII and how the information is organized (bit and byte order). File formats should be compared for their ability to contain detail (maximum number of pixels), tone (maximum number of bits per pixel), color (color space associated with format), and administrative metadata (number of file headers). Other technical considerations include longevity, quality, flexibility, computation efficiency, storage/transmission, and support by existing programs. Some design goals to consider are memory, accuracy, speed, device independence, robustness, extendibility, compatibility, modularity, plugability, openness, and scalability. With the variation in materials to be digitized as well as the purpose of the image, it is easy to understand how some of these design elements could be in conflict.⁵

Changes in technology have greatly increased the choice of formats available for bitmapped (raster) images. The image format is a determining factor in the choice of file format. The end use for an image should also be a consideration in that choice. There is a definite correlation between image quality and the intent of the image—be it to protect the original, represent the original, or transcend the original. And given that the end use will always result in a display to the user, it is also important to remember that current display technology is the main weak link in digitization.⁶

Anne Kenney, in her keynote address to the Joint RLG and NPO Preservation Conference on *Guidelines for Digital Imaging*, had many practical suggestions but emphasized guidance over strict recommendations in an environment where there is no stability and the ground rules change. “Guidance begins with understanding the context, provides a process for data gathering and decision-making, points to available guidelines and a means for assessing their applicability, and finally leads to the development of guidelines that are specific to

⁴ D’Amato, Donald. (2000, July) 3. Imaging Systems: the Range of Factors Affecting Image Quality in *Guides to Quality in Visual Resource Imaging* [Online]. Available: <http://www.rlg.org/visguides/visguide3.html> [2001:November 12].

⁵ Frey, Franziska. (2000, July) 5. File Formats for Digital Masters in *Guides to Quality in Visual Resource Imaging* [Online]. Available: <http://www.rlg.org/visguides/visguide5.html> [2001:November 12].

⁶ Conway, Paul. (1999). Digital Technology Made Simpler (Technical Leaflet, Section 5, Leaflet 4). In *Preservation of Library & Archival Materials: A Manual* (3d ed.) [Online]. Available: <http://www.nedcc.org/plam3/leaf54.htm> [2001, November 12].

institutional needs. Guidance should lay out a uniform process in which prejudices are stated up front, decisions are well defined and documented, and evaluation is built in all along the way.”⁷

The following table on Image File Formats and Compression Techniques lists different file types and compression techniques with brief information and a link to detailed information, usually to either the official or “unofficial” home page.

Table: Image File Formats and Compression Techniques

Format/Compression Algorithm	Explanation
CPC	Cartesian Perceptual Compression—Patented image compression method specifically designed for document image storage and transmission systems. http://www.cartesianinc.com/Tech/tech-overview.html
Flashpix	A multi-resolution, tiled file format that allows images to be stored at different resolutions for different purposes. Objects are stored in structured storage container files and the image data is stored in defined color spaces. Consistent color is achieved on both standard and color managed systems by defining color space options and providing standard ICC color management profiles. http://www.digitalimaging.org/i_flashpix.html
GIF	Graphic Image Format—widely supported image-storage format promoted by CompuServe that gained early widespread use on on-line services and the Internet. (Being replaced by PNG) http://www.dcs.ed.ac.uk/home/mxr/gfx/2d/GIF89a.txt
GIF w/LZW compression	Explains the Lempel-Ziv Welch compression algorithm, and, specifically, the implementation that GIF uses. http://www.dcs.ed.ac.uk/home/mxr/gfx/2d/GIF-comp.txt
ImagePac (.pcd)	Popular storage method for digital images. File storage format used with Kodak’s PhotoCD. http://www.kodak.com/US/en/digital/products/photoCD.shtml
ITU T.6	Facsimile coding schemes and coding control functions for Group 4 facsimile apparatus

⁷ Kenney, Anne (1998, September). Guidelines vs. Guidance for Digital Imaging: The Opportunity Before Us. In *Guidelines for Digital Imaging, Joint RLG and NPO Preservation Conference: 28-30 September 1998* [Online]. 14pp. Available: <http://www.rlg.org/preserv/joint/kenney.html> [2001, November 12]

Format/Compression Algorithm	Explanation
JBIG	<p>Joint Bi-level Image experts Group (official title: ISO/IEC JTC1 SC 29 Working Group 1) Created to produce standards for bi-level image coding. Developed IS 11544 (ITU-T T.82) for lossless compression of bi-level images (also used for coding grayscale and color images with limited number of bits per pixel). Can be seen as form of facsimile encoding similar to Group 3 or Group 4 fax, offering 20%-80% improvement in compression (about 20:1 over the original uncompressed digital bitmap image). http://www.jpeg.org/public/jbighomepage.htm</p>
JFIF	<p>JPEG File Interchange Format—file storage format for images compressed with the JPEG algorithm http://www.dcs.ed.ac.uk/home/mxr/gfx/2d/JPEG.txt</p>
JPEG	<p>Joint Photographic Experts Group (official title: ISO/IEC JTC1 SC 29 Working Group 1). Created to produce standards for continuous tone image coding. Best-known standard, IS 10918-1 (ITU-T T.81), is first in multi-part set of standards for still image compression. Generally 10:1 compression for grayscale, 20:1 for color. http://www.jpeg.org/public/jpeghomepage.htm</p>
LZW	<p>Lempel Ziv Welch—Proprietary lossless data compression and decompression technology http://www.unisys.com/unisys/lzw/</p>
MrSID	<p>Multi resolution Seamless Image Database 20:1 compression http://www.lizardtech.com/products/mrsid.html http://www.lizardtech.com/geospatial/MrSID_geospatial/whatis.htm</p>
PCX	<p>Graphics file used by PC Paintbrush http://www.dcs.ed.ac.uk/home/mxr/gfx/2d/PCX.txt</p>
PDF	<p>Portable Document Format—Open de facto standard for electronic document distribution worldwide http://www.adobe.com/prodindex/acrobat/adobepdf.html</p>
PICT	<p>Macintosh Picture—a storage format for digital images designed primarily for the Macintosh http://www.imageontage.com/Docs/PICT.html</p>
PNG	<p>Portable Network Graphics—Public domain format with lossless compression for storing bitmapped images. Approved byW3C to replace GIF for Web use. http://www.libpng.org/pub/png/</p>
SPIFF	<p>Still Picture Interchange File Format (Annex F of ITU-T Recommendation T.84 IS 10918-3)—Proposed by ISO JTC 29 Working Group 1 http://www.jpeg.org/public/spiff.pdf</p>

Format/Compression Algorithm	Explanation
TIFF	Tagged Image File Format—Public domain raster file format. Primarily designed for raster data interchange. Its main strengths are a highly flexible and platform-independent format that is supported by numerous image-processing applications. http://home.earthlink.net/~ritter/tiff/ (unofficial). TIFF Revision 6.0 http://partners.adobe.com/asn/developer/pdfs/tn/TIFF6.pdf
TIFF-IT	Tag Image File Format for Image Technology - widely used in the graphic arts industry. http://www.totalint.com/products/developer/tiffit.asp
UPF	Universal Preservation Format Goal is to bring together technology manufacturers and archivists to determine a UPF that meets the needs of both non-commercial and commercial interests http://info.wgbh.org/upf/

Since the development of file formats and standards should go hand in hand, a list of major organizations responsible for standards is included with links to their home pages:

International Standards Groups

- ISO International Organization for Standardization
(<http://www.iso.ch/iso/en/ISOOnline.openerpage>)
ISO TC42: Technical Committee Photography
ISO TC130: Technical Committee Graphic Arts
- IEC International Electrotechnical Commission (<http://www.iec.ch/>)
- ITU International Telecommunications Union
(<http://www.itu.int/home/index.html>)
- CIE Commission Internationale de l'Eclairage
(<http://www.hike.te.chiba-u.ac.jp/ikeda/CIE/index.html>)
- i³a International Imaging Industry Association (<http://www.i3a.org/>) -- formerly PIMA
- IPA International Prepress Association (<http://www.ipa.org/>)
- CEN European Committee for Standardization (<http://www.cenorm.be/>)

National Standards Groups and Associations Interested in Standards

- ANSI American National Standards Institute (<http://www.ansi.org/>)
- BSI British Standards Institute (<http://www.bsi-global.com/index.html>)
- DIN German Deutsches Institut für Normung (<http://www.din.de/>)
- NIST National Institute of Standards and Technology, US Department of Commerce (<http://www.nist.gov/>)
- NISO National Information Standards Organization (<http://www.niso.org/>)

SMPTE Society of Motion Picture and Television Engineers
(<http://www.smpte.org/>)

Other collaborative bodies

IETF Internet Engineering Task Force
(<http://www.ietf.cnri.reston.va.us/home.html>)

IEEE Institute of Electrical and Electronic Engineers (<http://www.ieee.org/>)

Best Practices:

Specific information for file format related to image types is available from several sources. The *Technical Recommendations for Digital Imaging Projects* has a “Quick Guide” which lists image type, conversion method, resolution, archival file format, screen presentation format, and print presentation format.⁸ Greenstein’s report on a DLF meeting on preservation reformatting practices gave minimum level characteristics for page images of a digital preservation master for printed texts, illustrated texts (black/white and color), and manuscripts.⁹ The Conversion section of the Moving Theory into Practice: Digital Imaging Tutorial has a table titled “Representative institutional requirements for conversion” which lists the specifications of seven institutions for printed text, pictorial materials, oversized materials and manuscripts.¹⁰ Digital Recommendations for Image Format from the Library of Congress American Memory Project also contains their requirements for digital conversion for different image types by the purpose of the image.¹¹

⁸ Image Quality Working Group of ArchivesCom, a joint Libraries/AclS committee (1997, April 2). *Technical Recommendations for Digital Imaging Projects* [Online]. Available: <http://www.columbia.edu/acis/dl/imagespec.html> [November 7, 2001].

⁹ Greenstein, Daniel (rev. 2001, July 30) *Report of a meeting of the DLF on preservation reformatting practices* [Online]. Available: <http://www.diglib.org/standards/presreformatsumpv.htm> [November 12, 2001].

¹⁰ Cornell University Department of Preservation and Conservation (2000-2001) Table: Representative institutional requirements for conversion. In *Moving Theory into Practice Digital Imaging Tutorial, Section 3 Conversion* [Online]. Available: <http://www.library.cornell.edu/preservation/tutorial/conversion/table3-1.html> [November 7, 2001]

¹¹ Fleischhauer, Carl (1998, July 13) *Digital Formats for Content Reproductions* [Online]. Available: <http://memory.loc.gov/ammem/formats.html> [2002, January 2]

Bibliography of Standards Resources

Archive Builders

<http://www.ArchiveBuilders.com/>

Provides an in depth understanding of document imaging as used in records management. Presentations are given on the computer industry as it relates to document imaging systems and records management for specific sites. Includes white papers, courses, handouts from presentations, etc.

Colorado Digitization Center, Digital Toolbox

<http://coloradodigital.coalliance.org/toolbox.html>

Designed to guide administrators of digital projects to sites and resources on the following topics: Administrative Resources, Guidelines and Standards, Copyright and Intellectual Property, Funding, Metadata, Software, Planning, etc.

Digital Library Federation, Digital Library Standards and Practices

<http://www.diglib.org/standards.htm>

Digital Library Standards.

<http://sunsite.berkeley.edu/Info/standards.html>

Includes draft standards or dreams of standards.

Digital Formats for Content Reproductions

<http://memory.loc.gov/ammem/formats.html>

Via the Library of Congress American Memory Project.

Graphics File Formats Page: 2D Bitmap Specifications (no date) [Online].

<http://www.dcs.edu.ac.uk/home/mxf/gfx/2d-hi.html> [2001, November 7]

Guides to Quality in Visual Resource Imaging (Washington, D.C.: Council on Library and Information Resources, Digital Library Federation ; Mountain View, CA: Research Libraries Group). 2000.

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

Moving Practice into Theory: Digital Imaging Tutorial

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

Tables:

1. Representative institutional requirements for conversion
<http://www.library.cornell.edu/preservation/tutorial/conversion/table3-1.html>
2. Common Image File Formats

- <http://www.library.cornell.edu/preservation/tutorial/presentation/table7-1.html>
- 3. Attributes for Common Compression Techniques
<http://www.library.cornell.edu/preservation/tutorial/presentation/table7-3.html>
- 4. Comparison of Storage Media
<http://www.library.cornell.edu/preservation/tutorial/technical/table6-2.html>
- 5. Metadata Types
<http://www.library.cornell.edu/preservation/tutorial/metadata/table5-1.html>

Report of Imaging Practitioners Meeting on 30 March 2001 to Consider How the Quality of Digital Imaging Systems and Digital Images May be Fairly Evaluated [Online]. Chapman, Stephen (2001, May 23)
<http://www.diglib.org/standards/imqualreppv.htm>

Report of a Meeting of the DLF on Preservation Reformatting Practices, [Online]. Greenstein, Daniel (rev. 2001, July 30)
<http://www.diglib.org/standards/presreformatsumpv.htm>

Technical Recommendations for Digital Imaging Projects [Online]
Image Quality Working Group of ArchivesCom, Joint Libraries/AcIS Committee (1997, April 2).
<http://www.columbia.edu/acis/dl/imagespec.html>

Universal Preservation Format Glossary, [Online]. (no date).
<http://info.wgbh.org/upf/glossary.html>

World Wide Web Consortium
<http://www.w3.org/>

Develops interoperable technologies (specifications, guidelines, software, and tools) to lead the Web to its full potential as a forum for information, commerce, communication, and collective understanding. See A-Z list of topics.

Yale Style Manual-Graphics for the Web
<http://info.med.yale.edu/caim/manual/graphics/>

II. C. Area of Investigation: *File Formats for Current Access and for Long-Term Preservation*

Issues:

The section above about development of standards contains extremely relevant discussions for this section, also. The resources listed there should be reviewed before continuing with this discussion of file formats for current access and for long-term preservation.

Multiple file formats may be necessary for a project depending on the project objectives. For example, a compressed file format suitable for a thumbnail image is not suitable as a preservation quality image requiring both file types be created to meeting the Web display requirements and the preservation needs. Another major issue is the current limitations of display technology. This goes beyond the timely loading of images and involves browser support—or lack of support—for various file formats

In part eight of the *Handbook for Digital Projects: A Management Tool for Preservation and Access* <<http://www.nedcc.org/digital/dighome.htm>>, author Janet Gertz, Columbia University Libraries states that there is a “lack of clear, nationally accepted standards and specifications” for formats for digitized materials. There are recommendations, however, several of which are included in Berkeley’s Digital Library SunSITE’s Digitizing Images and Text <<http://sunsite.berkeley.edu/Imaging/>>. Other excellent resources also exist, which are included in the Best Practices section below.

Best Practices:

Most resources distinguish between file formats for current viewing and file formats for long-term preservation (sometimes called “digital masters”). When viewing the following resources or planning a digital project be aware of that distinction and recognize that multiple file formats may be required to accomplish all of the project goals.

The Library of Congress’ *Digital Formats for Content Reproductions* <<http://memory.loc.gov/ammem/formats.html>> organizes materials by format (pictorial, textual as searchable text/images or images, maps, sound-recordings, etc.). See also the Digital Library Federation’s (Research Libraries Group) Guides to Quality in Visual Resource Imaging, *File Format for Digital Masters* <<http://www.rlg.org/visguides/visguide5.html>> (format, standards, compression) and the Colorado Digitization Project’s General Guidelines for Scanning <<http://coloradodigital.coalliance.org/scanning.html>>. This includes recommended standards for text, photographs, maps, and graphic materials.

The Department of Preservation and Conservation at the Cornell University Library provides a Table on Common Image File Formats <<http://www.library.cornell.edu/preservation/tutorial/presentation/table7-1.html>> via its Digital Imaging Tutorial <<http://www.library.cornell.edu/preservation/tutorial/presentation/presentation-02.html>>.

To see what other institutions recommend for conversion requirements, see the Table, Representative Institutional Requirements for Conversion <<http://www.library.cornell.edu/preservation/tutorial/conversion/table3-1.html>>.

This site also covers compression, “used to reduce image file size for storage, processing, and transmission.” See the discussion in this tutorial <<http://www.library.cornell.edu/preservation/tutorial/intro/intro-07.html>>. Also review the tables that summarize attributes for common compression techniques and compare file size and compression.

Columbia University makes available *Technical Recommendations for Digital Imaging Projects*. As a part of this report, the Image Quality Working Group of ArchivesCom, a joint Libraries/ACIS committee, prepared the following recommendations. The table below, taken directly from this document, offers guidelines for scanning materials in relation to their format and color.

From <http://www.columbia.edu/acis/dl/imagespec.html>

Media Type	Conversion Method	Resolution	Archive File Format	Screen Presentation Format	Print Presentation Format
Black & White Text Document	Flatbed Scanner or Digital Camera	1-bit, 600 dpi	TIFF w/CCITT Fax 4 Compression	GIF, 4-bit, 120 to 200 dpi	Acrobat (PDF), 1-bit, 300 or 600 dpi
Illustrations, Maps, Manuscripts, etc.	Flatbed Scanner or Digital Camera	8-bit grayscale or 24-bit color, 200 to 300 dpi	TIFF	Multiple JPEG, 24-bit, 512x768, 1024x1536, 2048x3072, Quality Level 50	JPEG, 24-bit, 2048x3072, Quality Level 50-100
Three-dimensional objects to be represented in two-	Digital Camera	24-bit color, 200 to 300 dpi	TIFF	Multiple JPEG, 24-bit, 512x768, 1024x1536, 2048x3072	JPEG, 24-bit, 2048x3072, Quality Level 50-100

Report to ALA GODORT – Digitization of Government Information

Media Type	Conversion Method	Resolution	Archive File Format	Screen Presentation Format	Print Presentation Format
dimensions				Quality Level 50	
35mm Black&White & Color slide or negative	PhotoCD or Slide Scanner	24-bit, 2048x3072	PhotoCD or TIFF	Multiple JPEG, 24-bit, 512x768, 1024x1536, 2048x3072, Quality Level 50	JPEG, 24-bit, 2048x3072, Quality Level 50-100
Medium to Large Format photograph, slide, negative, transparency or color microfiche	ProPhotoCD or Drum Scanner	24-bit, 4096x6144	PhotoCD or TIFF	Multiple JPEG, 24-bit, Quality Level 50	JPEG, 24-bit, 4096x6144, Quality Level 50-100
Black & White Microfilm	Microfilm Scanner	1-bit 600 dpi	TIFF w/ Fax 4	GIF, 4-bit, 120 to 200 dpi	PDF, 1-bit, 300 or 600 dpi
		8-bit, 300 dpi	TIFF	GIF, 8-bit 120 to 200 dpi	PDF, 8-bit, 300 or 600 dpi

Bibliography of File Formats Resources

Archive Builders

<http://www.ArchiveBuilders.com/>

Provides an in depth understanding of document imaging as used in records management. Presentations are given on the computer industry as it relates to document imaging systems and records management for specific sites. Includes white papers, courses, handouts from presentations, etc.

Berkeley Digital Library SunSITE

<<http://sunsite.berkeley.edu/Imaging/>

Digitizing images and text.

Colorado Digitization Center, General Guidelines for Scanning

<http://coloradodigital.coalliance.org/scanning.html>

Recommended standards for text, photographs, maps, and graphic materials.

Digital Formats for Content Reproductions (Library of Congress)

<http://memory.loc.gov/ammem/formats.html>

Organizes materials by format (pictorial, textual as searchable text/images or images, maps, sound-recordings, etc.)

Guides to Quality in Visual Resource Imaging, *File Format for Digital Masters*

Digital Library Federation's (Research Libraries Group)

<http://www.rlg.org/visguides/visguide5.html>

Handbook for Digital Projects: A Management Tool for Preservation and Access

<http://www.nedcc.org/digital/dighome.htm>

Technical Recommendations for Digital Imaging Projects. Columbia University.

<http://www.columbia.edu/acis/dl/imagespec.html>

II. D. Area of Investigation: *Explore Outsourcing Possibilities for Digitization and Create a List of Vendors*

Issues:

An organization must decide whether to digitize in-house or to contract out with a vendor. Economics, equipment, and staff are critical variables. Part Eight of the *Handbook for Digital Projects: A Management Tool for Preservation and Access* <<http://www.nedcc.org/digital/dighome.htm>> offers an introduction to issues related to digitizing in-house or contracting out to a vendor. Below is an excerpt from chapter 8, written by Janet Gertz, which clearly outlines the issues involved:

“Quality digital conversion work can be accomplished in-house or through vending out to service providers. Regardless of whether digitization is intended to serve preservation goals, it is a waste of time and money to do a poor job. A digitally converted version of a document must be fully functional. If what is digitized is illegible or so poorly indexed that end users cannot find what they need or read it when they do find it, there is a failure to provide both preservation and access.

In order for digitization to be successful, it is essential that the institution have a clear understanding of its goal for digitization and what kind of final product will serve that goal. Understanding why a project is being

undertaken will guide decision making, not only about image quality and user interfaces but also about what work should be accomplished in-house and what work may safely be vended out. It is important to:

- Involve all the relevant participants (curators, technical experts, preservation officers) in determining the project goals and making the decisions that will shape it
- Keep a careful record of what decisions are made, and why, to prevent re-inventing the wheel when problems arise
- Document fully how and why work was accomplished both in-house and by the vendor in order to aid future preservation of the digital resources themselves.

There will always be an in-house component to any digitization operation. The institution that holds the materials to be digitized must take responsibility for:

- Selecting materials to be converted
- Determining the purpose of digitization and the nature of the desired product
- Establishing necessary quality levels
- Verifying the quality of the completed work.

There are arguments in favor of working entirely in-house and arguments for employing service bureaus. The difference lies in:

- The degree of immediate control over the work
- The variety of activities that can be performed
- Efficiency
- Economics¹²

Best Practices:

The goals for a digital project may be achieved by either in-house or outsourced scanning. The decision must be based on a thorough evaluation of the project goals, costs, institutional hardware/software availability, staff skills and availability, and condition of materials to be digitized. Mentioned above, Geitz's contribution to the *Handbook for Digital Projects: A Management Tool for Preservation and Access, VIII. Vendor Relations* <<http://www.nedcc.org/digital/dighome.htm>> describes vendor relations in detail. It covers the key issues: Why Digitize In-House, Why Use Vendors, How to

¹² Sitts, Maxine K., Editor (2000) *Handbook for Digital Projects: A Management Tool for Preservation and Access, VIII. Vendor Relations*, by Janet Geitz. Available: <http://www.nedcc.org/digital/dighome.htm> [January 2, 2002].

Choose Services and Vendors, What are the Project Goals, the RFI - Request for Information, the Request for Proposal (RFP), Evaluating Responses from Vendors, the Contract, Working and Communicating with Vendors, Working with Vendors: Quality Control, and Handling Corrections.

The Colorado Digitization Project's General Guidelines for Scanning <<http://coloradodigital.coalliance.org/scanning.html>> also offers pros and cons for outsourcing digitization projects or conducting them in-house.

It is impossible to overemphasize the importance of in-house knowledge regarding the digitization process. It is just as crucial for the successful implementation of an in-house digitization project as it is for the monitoring of a project that is contracted out to a vendor.

Digital Imaging Vendors

The Colorado Digitization Project site includes a list of digital imaging vendors. It includes "links to companies that provide digital imaging technology and services. This list is provided as a service for the library, archival, and museum community. The list is updated as additional suppliers are identified.

<<http://coloradodigital.coalliance.org/vendor.html>>

Solinet (Southeastern Library Network) Preservation Leaflets include links to Digital Imaging Services. It contains a "list of companies that provide digital imaging technology or services, ... most of which have experience working with libraries and archives."

<<http://www.solinet.net/presvtn/leaf/digimage.htm>>

Solinet Preservation Leaflets also offer Reformatting Services with links to "companies that offer reformatting services for libraries and archives. These services include digital imaging, preservation microfilming, preservation photocopying and magnetic media reformatting."

<<http://www.solinet.net/presvtn/leaf/reformat.htm>>

Additional Vendors:

Axion Data Services

info@axiondata.com or <mailto:sales@axiondata.com>

Phone: 800-493-2630

Address: 1910 Cochran Road

Suite 730, Manor Oak 2

Pittsburgh, PA 15220

<http://www.axiondata.com/>

Image API
Nevin Smith, President
nsmith@imageapi.com
502 N. Adams Street
Tallahassee, FL 32301
Phone: 850-222-1400
Fax: 850-224-3367 (Fax)

ILM Corporation
E-mail: info@ilmcorp.com
Phone: 540-898-1406
Address: 216 Industrial Court
Fredericksburg, VA 22408
<http://www.ilmcorp.com/>

NBS Solutions
Phone: 651-688-0202
Address: 2919 West Service Road
Eagan, MN 55121
<http://www.nbsusa.com/>

Northern Micrographics, Inc.
2004 Kramer Street
P.O. Box 2287
La Crosse, WI 54602-2287
Phone: (608) 781-0850 ext. 222
Toll free: 800-236-0850
Fax: (608) 781-3883
http://www.normicro.com/northern_micrographics.htm

Peelle Technologies
Phone: 800-233-5006
Address: 197 East Hamilton Ave.
Campbell CA 95008
<http://www.peelletech.com>

Bibliography of Outsourcing Resources

Choosing a Vendor: Tips from the Lason Corporation. Selecting a vendor for a Web digitization project. *Texas Library Journal*. v. 75 no3 (Fall 1999) p. 107+

Colorado Digitization Center, Digital Toolbox
<http://coloradodigital.coalliance.org/toolbox.html>

Designed to guide administrators of digital projects to sites and resources on the following topics: Administrative Resources, Guidelines and Standards, Copyright and Intellectual Property, Funding, Metadata, Software, Planning, etc.

Digital Imaging Services. Solinet Preservation Leaflets.

<http://www.solinet.net/presvtn/leaf/digimage.htm>

List of companies that provide digital imaging technology or services a few resources to aid in locating and re-evaluating potential vendors.

Digital Libraries: Metadata Resources (IFLANET)

<http://www.ifa.org/ll/metadata.htm>

Metadata resources for digital libraries (from the International Federation of Library Associations and Institutions). Excellent resources.

Outsourcing Digitization. *Library Journal*. v. 124 no15 (Sept. 15 1999) p. 34+

Reformatting Services. Solinet Preservation Leaflets.

<http://www.solinet.net/presvtn/leaf/reformat.htm>

Services include digital imaging, preservation microfilming, preservation photocopying and magnetic media reformatting.

II. E. Area of Investigation: Consider ADA Issues

Issues:

As with any key material on the Web, authors of digitized information for Web display must keep in mind that the material should be accessible to all users. Several initiatives exist to expand Web information to all users and to educate those responsible for Web sites about accessibility issues. As a result of this initiative, Web designers have access to excellent resources to assist them with creating Web sites for users with disabilities.

The World Wide Web Consortium (W3C) was created in October 1994 to lead the Web to its full potential by developing common protocols (or technologies: specifications, guidelines, software, and tools) that promote its evolution and ensure its interoperability <<http://www.w3c.org/>>. One of the focuses of this key body is that of Web accessibility or universal access. It channels this effort through two main initiatives, Voice Browser Activity and the Web Accessibility Initiative (WAI). With the Voice Browser Activity, W3C works “to expand access to the Web to allow people to interact via key pads, spoken commands, listening to prerecorded speech, synthetic speech and music.” “The Web Accessibility

Initiative, in coordination with organizations around the world, pursues accessibility of the Web through five primary areas of work: technology, guidelines, tools, education and outreach, and research and development.”

The federal government has required its agencies to work with Section 508 <<http://www.section508.gov/>> of the 1998 Rehabilitation Act which “requires that Federal agencies electronic and information technology is accessible to people with disabilities. The Center for Information Technology Accommodation (CITA), in the U.S. General Services Administration's Office of Government wide Policy, has been charged with the task of educating Federal employees and building the infrastructure necessary to support Section 508 implementation. Using this web site, Federal employees and the public can access resources for understanding and implementing the requirements of Section 508.”

Several sites exist to assist Web designers with the development and testing of their work. See the Best Practices section below for a selected list of sites.

Best Practices:

When designing files for the Web, keep current with the W3C's Web Accessibility Initiative <<http://www.we.org/WAI/>>. The Resources section <<http://www.we.org/WAI/Resources/>> includes Guidelines, Checklists, and Techniques that developers should follow. These are very detailed and comprehensive.

Relevant Guidelines:

Web Content Accessibility Guidelines 1.0

Explains in detail how to make a Web site accessible for people with a variety of disabilities.

Relevant Checklists:

Checklist for Web Content Accessibility Guidelines 1.0

Prioritized list of checkpoints for making Web sites accessible; an outline of the information in the Web Content Accessibility Guidelines.

Relevant Techniques:

Techniques for Web Content Accessibility Guidelines 1.0

Detailed markup examples and explanations of how to implement the Web Content Accessibility Guidelines.

The References on the Web Accessibility section of the W3C WAI site <<http://www.w3.org/WAI/References/>> include links to all of the WAI resources plus a Mainstream Developers section that links to accessibility information for Web software developers, a Projects/Research/Resources section with links

related to Web accessibility at other organizations, and a Papers/Articles section. Of specific interest is the Mainstream Developers Accessibility Information <<http://www.w3.org/WAI/References/#mainstreamdev>>.

Also for Web designers, the Bobby Worldwide site exists to assist authors in the creation of sites that may be easily used by persons with disabilities <<http://www.cast.org/bobby/>>. “Bobby WorldWide was created by CAST. Founded in 1984 as the Center for Applied Special Technology, CAST is a not-for-profit organization whose mission is to expand opportunities for people with disabilities through innovative uses of computer technology.”

The Accessible Web Page Design site also offers a good collection of information for Web developers <<http://www.makoa.org/web-design.htm>>.

Adobe Systems, the developer of the portable document file format (.pdf), features a Web page <<http://access.adobe.com/>> that describes the capabilities of the most recent version of its Adobe Acrobat software that provides a number of new capabilities for disabled users. For Acrobat 5.0 and Acrobat Reader 5.0 this includes:

- Read Adobe PDF documents with Windows-based screen readers from vendors such as Freedom Scientific and GW Micro.
- Use keyboard navigation
- View documents in high contrast mode
- Zoom in on text and reflow to fit any size view when working with tagged (accessible) Adobe PDF files

Other software specific accessibility information available includes:

- AOL’s Accessibility Policy (links?)
- Apple Computer, Disability Solutions
- IBM Accessibility Center
- Macromedia Accessibility
- Microsoft Accessibility
- Opera Software
- Sun Microsystem’s Accessibility Program

Bibliography of ADA Resources

Accessible Web Page Design

<http://www.makoa.org/web-design.htm>

Resources that may be helpful in creating pages that are truly accessible.

Adobe Systems

<http://www.adobe.com/>

CAST Bobby

<http://www.cast.org/bobby/>

Bobby was created by CAST to help Web page authors identify and repair significant barriers to access by individuals with disabilities.

Section 508

<http://www.section508.gov/>

Section 508 requires that Federal agencies electronic and information technology is accessible to people with disabilities. The Center for Information Technology Accommodation (CITA), in the U.S. General Services Administration's Office of Government-wide Policy, has been charged with the task of educating Federal employees and building the infrastructure necessary to support Section 508 implementation.

World Wide Web Consortium

<http://www.w3.org/>

Develops interoperable technologies (specifications, guidelines, software, and tools) to lead the Web to its full potential as a forum for information, commerce, communication, and collective understanding. Two key ADA initiatives include Voice Browser Activity and the Web Accessibility Initiative.

II. E. Area of Investigation: Server-Side Issues

Issues:

With the management of any server, a method for the organization of files on the server is necessary. This is dependent upon the project, the type of files, the number of files, the users, security issues, etc. This is also dependent in part upon the metadata schemas that are used to identify files. Please refer to the Metadata discussion in this report. File management is especially important for digitization projects. There may be thousands of files of varying types and sizes that include files accessed via the Web by users and files stored for archival purposes. Issues to be considered include naming conventions, storage, and access.

Best Practices:

One set of recommendations does not exist for management of files on servers or for search engines to use on a site. Decisions concerning file management

and search engines must be made in the context of the project specifications, such as, the type of files, the number of files, the users, security issues, metadata, etc.

Naming Conventions:

The Colorado Digitization Project site discusses briefly the naming of files <<http://coloradodigital.coalition.org/scanning.html - File#File>>. “You will need to consider the nomenclature you will use to name your files before starting the project. The file name must be a unique number that uniquely identifies the image. The file name may include the name of the collection or institution as well as the image number, plus the appropriate extension (.gif, .jpg, .tif). File names should be no longer than 8 characters and should not include spaces or symbols such as ?, /, or # (etc).”

Cornell University Library’s tutorial offers information on keeping track of data (basic file system considerations)

<<http://www.library.cornell.edu/preservation/tutorial/technical/technicalC-01.html#keeping+track>>.

The authors suggest that to “some degree, the nature of the material being scanned will suggest organizing principles.” They offer some basic file system recommendations:

- “Use standard, cross-platform compatible file-naming schemes
- Use standard file extensions for different file types
- Don't overload directories with too many files
- Rely on storage management software to manage large collections across multiple physical disk drives
- Allow for collection growth.”

Storage:

Archival files may be stored on a second server, CD-ROMs, tape, or other types of storage media. Files that have taken hours of work and considerable outlay of funds to create require careful management. Loss of these files would be disastrous.

Cornell University Library’s tutorial offers an excellent discussion under Technical Infrastructure of File Management

<<http://www.library.cornell.edu/preservation/tutorial/technical/technicalC-01.html>>, which includes image databases, storage, storage types, and storage needs. It includes a Table called “Comparison of Storage Media”

<<http://www.library.cornell.edu/preservation/tutorial/technical/table6-2.html>> that contains criteria for evaluation for appropriate mass storage. They include:

- Speed (read/write, data transfer)
- Capacity

- Reliability (stability, redundancy)
- Standardization
- Cost
- Fitness to task.

To determine storage capacity requirements, the following calculation may be used:

$$\text{Total storage needed} = \# \text{ of image files} \times \text{average file size} \times 1.25$$

And storage cost may be approximated using the following formula:

$$\text{Total storage cost} = \text{total storage needed} \times \text{cost per unit of storage}$$

See also the topic of Digital Preservation (challenges, technical strategies, organizational strategies)

<<http://www.library.cornell.edu/preservation/tutorial/preservation/preservation-01.html>>

The Colorado Digitization Project site presents considerations for the storing of files <<http://coloradodigital.coalition.org/scanning.html#Storing>>, including magnetic disks, CD-ROMs and tape. The Digital Library Federation's (Research Libraries Group) Guides to Quality in Visual Resource Imaging, *File Format for Digital Masters* includes a discussion about storage concerns <<http://www.rlg.org/visguides/visguide5.html>>.

Access:

Delivery information is included in Cornell University's tutorial

<<http://www.library.cornell.edu/preservation/tutorial/technical/technicalD-01.html>> (networks, monitors, printers).

Institutions may also rely on an advanced system for file management. One example, the Digital Object Identifier (DOI®) <<http://www.doi.org/>> (International DOI Foundations) is a "system for interoperably identifying and exchanging intellectual property in the digital environment. It provides an extensible framework for managing intellectual content in any form at any level of granularity, for linking customers with content suppliers, facilitating electronic commerce, and enabling automated copyright management for all types of media."

A search engine is important for users to aid with the discovery of information on a Web site. Most commercial server software includes software that will index data stored on the server. Search engine software (such as Google, Inktomi, Verity) may also be added for site search capabilities. Some search software will index a variety of file types, such as PDF, word processing files, HTML, slide

show documents, etc.. An institution may wish to inventory all the file types on their servers and build a list of requirements for the functionality of a search engine before beginning a search for the right software to meet your needs. Some search engine software may be free but require that advertising be placed on the site or high-level local technical skills for customization of the software. Other search engine software could cost many thousands of dollars. With high-level programmer skills, a program may also be created in-house.

Bibliography of Server-Side Resources

Colorado Digitization Center

<http://coloradodigital.coalliance.org/>

Digital Object Identifier System

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

A system for identifying and exchanging intellectual property in the digital environment. It provides a framework for managing intellectual content, for linking customers with content suppliers, for facilitating electronic commerce, and enabling automated copyright management for all types of media.

Moving Theory Into Practice: Digital Imaging Tutorial

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

Excellent tutorial from the Department of Preservation and Conservation at the Cornell University Library. Includes Selection, Conversion, Quality Control, Metadata, Technical Infrastructure, Presentation, Digital Preservation, Management, Continuing Education.

Relevant Electronic Discussion Lists:

List Name: DigLibns

To subscribe: Send the message "SUB DigLibns your name" to listserv@sunsite.berkeley.edu

List Name: IMAGELIB

To subscribe: Send the message "SUB imagelib Your Full Name" to listserv@listserv.arizona.edu

SECTION III.

Work Group 3 – Preservation of Electronic Files

Introduction

Group 3 worked with three major issue areas: (1) storage and day-to-day maintenance of files; (2) long-term preservation of files; and (3) authenticity. The major issues in each area are described below. Best practices identified in each area are listed. Emphasis must be placed on the fact that the issue of best practices in the digital realm continues to be the subject of much discussion. Although there are emerging models, standards and trends in each of these three areas, best practices are continually evolving as innovation continues. Perhaps the single best practice we recommend to digital project managers is to consult with local information technology resource staff and keep abreast of developments in the broader digital arena.

III. A. Area of Investigation: The Public Interface and Storage and Day-to-Day Maintenance of Files

Major Issues

The major issues that must be addressed with regard to storage and day-to-day maintenance of digital files are the costs related to management of:

- **Hardware:** Hardware includes servers, storage media, and peripherals. When planning a digitization project, be sure to coordinate your plans with your institutional IT department to determine if hardware costs must be added to the project budget when seeking grant or other funding.
- **Software:** Includes the operating system, and for the public interface, it includes retrieval software, rendering software, and reformatting software.
- **Personnel:** This issue includes several components, chief among them are server administration, software licensing negotiation, and project management.
- **Backups:** Backups must be incorporated into the day-to-day storage and maintenance activities of any digital project. They are discussed below in the Digital Preservation section of this report.

Best Practices

Whether a project takes place within the context of a small library, or in a large organization in consultation with an information technology department, it is important to prepare estimates of project space and server needs. Server space requirements can be estimated by multiplying a representative image size by the projected number of images. Estimate the necessary server speed using a representative image size and using it in the following formula:

$$t \text{ (time in seconds)} = \text{number of megabytes in file} \div (\text{transmission speed (in MB/sec)} \times .8)$$

For a list of transmission speeds, see *Moving Theory into Practice Digital Imaging Tutorial, 6D. Technical Infrastructure*. Available at: <http://www.library.cornell.edu/preservation/tutorial/technical/technicalID-04.html>.

Estimate overall hardware costs using sections J-K of the *RLG Worksheet For Estimating Digital Reformatting Costs*. Available at: <http://www.rlg.org/preserv/RLGWorksheet.pdf>.

Creating the Public Interface:

Once digitized, the project may be marked up in a variety of languages to create metadata, which will be discussed later in this report. The term “markup” refers to computerized document preparation and is a method of adding information to the text indicating the logical components of a document, instructions for layout of the text on the page, or other information that can be interpreted by some automatic system. However, all of the relevant markup languages have the same problem with respect to viewing on the Web - they cannot be directly viewed in the major Web browsers, including Internet Explorer and Netscape. This means that the project administrator must also choose a method of converting the marked up images into a viewable format. Many programs will convert files to a viewable format. Costs of such programs vary widely, from free to very costly. However, the choice of conversion software is not limited to available funding, but also is influenced by the availability of technical staff and the goals of the project. For example, a project manager may choose an open source program because of the presence of skilled technical staff to set up the program and provide in-house support. Conversely, another institution may choose to license access to software, such as the Xpat search engine, because the continuing costs of the software are less than the costs for an equivalent number of hours of an in-house support person.

In addition, the choice of software depends on the material being marked-up and the goals of the project. For example, Ebind is structured at the page level, making it ideal for digitized books but not for digitized images. Also, project

managers may wish to integrate the digitized materials into the library catalog. Some vendors, such as Ex Libris, sell a tool that will handle both the marking up of the images, presentation of them, and integration into the catalog.

Descriptions of finished projects and their interfaces include:

Foreign Relations of the United States:

<<http://libtext.library.wisc.edu/FRUS/About.html>>

Scroll down to “Technical Environment”

American Memory Collection

<<http://memory.loc.gov/ammem/amlaw/lwdigit.html>>

Scroll down to “Creating Digital Images” and “Creating Digitized Text”

Kappler: Indian Affairs: Laws and Treaties

<<http://digital.library.okstate.edu/kappler/intro.htm>>

American Indians of the Pacific Northwest

<<http://content-dev.lib.washington.edu/aipnw/aip-sitetech.html>>

Software Costs:

Overall software costs can be estimated using sections H and I of the *RLG Worksheet For Estimating Digital Reformatting Costs*

<<http://www.rlg.org/preserv/RLGWorksheet.pdf>>. Project managers will find the *RLG Worksheet* useful when estimating costs. Since many organizations are reluctant to share information about their project costs, *the RLG Worksheet* formulas should help even the first-time project manager estimate accurately the costs of digital reformatting.

Bibliography for Storage and Day-to-Day Maintenance of Files

Collection Based Persistent Digital Archives. Reagan Moore et al. D-Lib Magazine 6:3 & 4 (March and April 2000).

<http://www.dlib.org/dlib/march00/moore/03moore-pt1.html>

<http://www.dlib.org/dlib/march00/moore/03moore-pt2.html>

Ebind <http://sunsite.berkeley.edu/Ebind/>

Kenney, Anne and Stephen Chapman. *Digital Imaging For Libraries and Archives* (Ithaca, NY: Cornell University, 1996)

Kenney, Anne R. and Oya Y Rieger. Moving Theory into Practice: Digital Imaging for Libraries and Archives (Mountain View, CA.: RLG, 2000)

Lee, Stuart D. Digital Imaging: A Practical Handbook. (New York: Neal Schuman, 2000).

List of conversion and presentation software:

<http://coloradodigital.coalliance.org/software.html>

Moving Theory into Practice

<http://www.library.cornell.edu/preservation/tutorial/technical/technicalID-01.html>

RLG Worksheet For Estimating Digital Reformatting Costs

<http://www.rlg.org/preserv/RLGWorksheet.pdf>

III. B. Area of Investigation: Long-Term Preservation of Files

Major Issues

It should be emphasized that there is considerable controversy and lack of agreement in the literature regarding a single best model, if not best practice, for an overall strategy of long-term preservation of and access to digital files. Three models of how to approach the issue of long-term preservation and access are identified by Charles Dollar in a recent essay ¹³("Electronic Archiving in the 21st Century", in *Preservation of Public Data*, 2000):

- 1) Digital Archaeology approach: a minimalist strategy that converts electronic records to new technology only when future access to them is required;
- 2) Emulation approach: a process in which one computer is used to reproduce the behavior of another computer with such fidelity that the emulation can be used in place of the original computer; and
- 3) Migration approach: concerned with maintaining dynamic electronic records that can be rendered and processed with whatever technologies the current and future marketplaces support.

A fourth approach which might be described as a demand-driven, distributed approach is based on the notion that a critical way to preserve long-term access to digital objects is to distribute them as widely as possible and employ

¹³ Dollar, Charles. "Electronic Archiving in the 21st Century: Principles, Strategy and Best Practices." *Preservation of Public Data Of Significance..A Topical Journal of the Association of Public Data Users* 2:2 (2000).

technology to ensure routine mutual verification that the object is authentic and retains its original integrity. The LOCKSS project is an example of this approach.

These approaches are not necessarily mutually exclusive. Long-term preservation and access likely depends upon a multiplicity of approaches. A significant new model for digital preservation that is on the path toward becoming an ISO standard is the "Open Archival Information System (OAIS)" model. This model provides a conceptual framework for a digital archival system and was developed initially by NASA's Consultative Committee for Space Data Systems. It was recently adopted as ISO draft standard (DIS 14621) and is described in detail at: <http://www.ccsds.org/documents/pdf/CCSDS-650.0-R-2.pdf>.

For the purposes of this report, we focused narrowly on the preservation of individual digital files, not broader collections of files or file systems.

The major issues related to the long-term preservation of discrete digital files include: (1) scale; (2) complexity; (3) fragility; and (4) disaster planning.

Scale: An increasing volume of government material is available in electronic format. As libraries digitize government documents, it will be necessary to mesh the records created locally, whether static files or aggregations of data, within the library's broader collection of digital files. Local decisions regarding format selection, metadata standards, storage and interface must take into account both local and broad-based standards in these areas.

Complexity: Preservation is difficult in an analog environment. However, digital preservation is even more complex, because an effective preservation strategy for this area must address three related elements: (1) the data itself; (2) the corresponding metadata; and (3) the rendering software or viewer technology that presents the content.

Failure to address preservation of all three threatens access for future generations of users.

Fragility: Digital media are fragile and life expectancies vary significantly with changes in environmental conditions and quality of manufacture. A 1993 General Services Administration study confirmed that archival paper and microfilm possess life expectancies far longer than emerging magnetic and optical media.

Disaster Planning: Preserving digital material is not the same as disaster planning. A library must develop a contingency plan in case of disaster that involves loss of content and access to digital material.

Best Practices

Preservation metadata should be created for all files. The metadata elements should fully document the digital object in terms of its nature, operation, and interactions with other digital objects and applications.

Digital files should be routinely checked for integrity. These procedures are referred to as “control checks”. The interval between checks should be no greater than 5 years.

A disaster recovery plan should be developed. This plan should involve a complete backup of any access system. In addition, copies of all files should be stored under proper environmental conditions at a site geographically distant from the library.

Duplicate copies should be made. Ideally there should be two: (1) a preservation master from which new working copies can be made; and (2) a security master, to guard against catastrophic events.

Files should be copied on to new media at intervals that meet the manufacturers’ recommendations for the medium to prevent the physical loss of data.

All file conversions and transfers should be carried out using applications that produce audit trails, have integrity checking and reporting features.

Application compatibility. Rendering software or viewer technology must be able to access all the formats within the department’s electronic store.

Systems compatibility. Files and rendering software must be compatible within the operating system environment.

A preservation strategy must include the contextual metadata. If a document management system is used to link metadata and files, it is important that these links are maintained through any data migration project.

When digitizing documents, retain the original paper or create an analog backup from the digital master.

Bibliography for Long-Term Preservation of Files

AHDS: Arts and Humanities Data Service (UK) “Digital Preservation: a guide to web resources”

<http://ahds.ac.uk/resource/preserve.html>

Anderson, U. 1997. Workshop on Electronic Archiving: An Evaluation of the Sesam Report (Stockholm: Swedish National Archives, Astra AB).

Bagnall, Roger S. 1995. Digital Imaging of Papyri (Washington, D.C.: The Commission on Preservation and Access).

Berkeley Digital Library SunSITE. Preservation Resources.
<http://sunsite.berkeley.edu/Preservation/>

Best Practices for Digital Archiving: An Information Life Cycle Approach. Gail Hodge. D-Lib Magazine 6:1 (June 2000).
<http://www.dlib.org/dlib/january00/01hodge.html>

The Cedars Project: CURL exemplars in digital archives (UK)
<http://www.leeds.ac.uk/cedars/>

Commission on Preservation and Access (CPA).
<http://palimpsest.stanford.edu/cpa/>

Conservation OnLine (CoOL). <http://palimpsest.stanford.edu/>

Consultative Committee for Space Data Systems. Reference Model for an Open Archival Information System (OAIS).
<http://www.ccsds.org/documents/pdf/CCSDS-650.0-R-2.pdf>

Conway, P. 1996b. Preservation in the Digital World. (Washington, D.C.: Commission on Preservation and Access)..

Day, Michael. Preservation of Electronic Information: A Bibliography.
<http://homes.ukoln.ac.uk/~lismd/preservation.html>

Eakins, J.P. and Graham, M.E., 1999. Content-based Image Retrieval: A report to the JISC Technology Applications Programme (Institute for Image Data Research, University of Northumbria at Newcastle).
<http://www.unn.ac.uk/iidr/research/cbir/report.html>

Gonzalez, P. 1998. Computerization of the Archivo General de Indias: Strategies and Results (Washington, DC: Council on Library and Information Resources).

Hendley, T. 1996. The Preservation of Digital Material (London: British Library Research and Development Department).

Kenney, A. R. and Personius, L. K. 1992. The Cornell/Xerox/Commission on Preservation and Access Joint study in digital preservation (phase 1, January 1990-December 1991: digital capture, paper facsimiles and network access) (Ithaca, NY: Commission on Preservation and Access).

Northeast Document Conservation Center
<http://www.nedcc.org/>

OCLC/RLG Working Group on Preservation Metadata. Preservation Metadata for Digital Objects: A Review of the State of the Art. January 31, 2001.
http://www.oclc.org/digitalpreservation/presmeta_wp.pdf

PANDORA: Preserving and Accessing Networked Documentary Resources of Australia
<http://pandora.nla.gov.au/>

Preservation of Public Data. Of Significance. A Topical Journal of the Association of Public Data Users 2:2(2000).

Preservation Resources Web Site (OCLC).
<http://www.oclc.org/oclc/presres/index.htm>

Preserving Access to Digital Information (PADI).
<http://www.nla.gov.au/padi/>

Puglia, S. 1999. 'Creating Permanent and Durable Information: Physical Media and Storage Standards' CRM: Cultural Resource Management 22(2), 25-27
<http://tps.cr.nps.gov/crm/archive/22-2/22-02-10.pdf>

RLG Working Group on Preservation Issues of Metadata, 1998.
<http://www.rlg.org/preserv/presmeta.html>

Research Libraries Group (RLG). (1996) Preserving digital information. Commissioned by the Commission on Preservation and Access and the Research Libraries Group. Final Report May 20, 1996.
<http://www.rlg.org/ArchTF/tfadi.index.htm>

Ross, Seamus. 2000. Changing Trains at Wigan: Digital Preservation and the Future of Scholarship. (London: National Preservation Office).

UK Public Records Office. Management, Appraisal and Preservation of Electronic Records.
<http://www.pro.gov.uk/recordsmanagement/eros/guidelines/default.htm>

Weber, H. and Dörr, M. 1997) Digitisation as a Method of Preservation (Amsterdam: European Commission on Preservation and Access).
<http://www.knaw.nl/ecpa/PUBL/weber.html>

III. C. Area of Investigation: Authenticity of Files

Major Issues

What is an “authentic” digital government publication? Much has been written on how to define authenticity. Defining the authenticity of digital objects has specific significance and implications depending on the communities and context in which the digital objects are used. Even though the concept of authenticity might be of equal importance to many communities, different criteria are used to assess the authenticity of a digital object. For example, the term authenticity has a very specific meaning to an art dealer and equally as specific a meaning to a legal scholar, but both individuals use a very different set of criteria to establish authenticity. As Bearman and Trant discuss, the very core of authenticity lays in the concept of truth. Lynch elegantly articulates the “central role of trust” in establishing authenticity. Authentication of digital information is as much a social issue as a technological issue.

However, for the sake of simplicity, authentication as used here is defined as something that is not changed from the original, that is what it purports to be, that is the same as expected based on prior reference. Terms related to authenticity are integrity, credibility, provenance, and reliability. Authenticity should not be confused with “authentication” which refers to a user’s ability and rights to access an information resource.

Best Practices

No single method of how to confirm authenticity stands out. The three prevalent models are: (1) digital signatures; (2) digital watermarks; and (3) trusted repositories.

Digital Signatures: A digital signature is represented as a string of binary digits. A digital signature is computed using a set of rules and a set of parameters such that the identity of the signatory and integrity of the data can be verified. An algorithm provides the capability to generate and verify signatures. Signature generation makes use of a private key to generate a digital signature. Signature verification makes use of a public key that corresponds to, but is not the same as, the private key. Each user possesses a private and public key pair.

The National Institute of Standards and Technology (NIST) is taking a leadership role in the development of a Federal Public Key Infrastructure (PKI) that supports

digital signatures and other public key-enabled security services. NIST is coordinating with industry and technical groups developing PKI technology to foster interoperability of PKI products and projects.

The National Archives and Records Administration (NARA) also produced a digital signature guidance document entitled "Records Management Guidance for Agencies Implementing Electronic Signature Technologies."

Digital Watermarks: Digital watermarking software programs/plugin embed information about the author into video, audio or graphics files. This information, when decoded with the appropriate software, can reveal things such as the identity of the source, author, creator, owner, distributor or authorized consumer of digital materials, video recordings or audio recordings. A digital watermark is an identification code, permanently embedded into digital data, carrying information pertaining to copyright protection and data authentication. Watermarks are not removable and unalterable. The information does not degrade with file duplication and does not perceptively disrupt the original data files

Trusted Repository: A Trusted Repository combines many of the concepts and practices found in the fields of archival administration, computer science, and librarianship. The Research Library Group (RLG) and OCLC recently drafted a report for public comment, "Attributes of a Trusted Digital Repository: Meeting the Needs of Research Resources" (<http://www.rlg.org/longterm/attributes01.pdf>). The report defines a Trusted Repository as:

"A reliable digital repository is one whose mission is to provide long-term access to managed digital resources; that accepts responsibility for the long-term maintenance of digital resources on behalf of its depositors and for the benefit of current and future users; that designs its system(s) in accordance with commonly accepted conventions and standards to ensure the ongoing management, access, and security of materials deposited within it; that establishes methodologies for system evaluation that meet community expectations of trustworthiness; that can be depended upon to carry out its long-term responsibilities to depositors and users openly and explicitly; and whose policies, practices, and performance can be audited and measured."

Bibliography for Authenticity of Files

Authenticity in a Digital Environment. Council on Library and Information Resources, May 2000. <http://www.clir.org/pubs/reports/pub92/contents.html>

Trusted Digital Repository: Attributes and Responsibilities." An RLG/OCLC Report. Available: <http://www.rlg.org/longterm/repositories.pdf>

Bearman, David, and Jennifer Trant. 1998. Authenticity of Digital Resources: Towards a Statement of Requirements in the Research Process, *D-Lib Magazine* (June) <http://www.dlib.org/dlib/june98/06bearman.html>

Carrying Authentic, Understandable and Usable Digital Records Through Time, Bikson, T.; RAND-Europe; Rothenberg, J. August 1999. http://www.si.umich.edu/CAMILEON/Camileon%20publications/final-report_4.pdf

Clifford A. Lynch, "Authenticity and Integrity in the Digital Environment: An Exploratory Analysis of the Central Role of Trust," *Authenticity in a Digital Environment* (Washington, DC: Council on Library and Information Resources, 2000), pp 32-50. <http://www.clir.org/pubs/reports/pub92/lynch.html>

FIPS (Federal Information Processing Standards Publication), Digital Signature Standard (DSS), FIPS PUB 186-2, January 2000. <http://csrc.nist.gov/publications/fips/fips186-2/fips186-2.pdf>

NARA GPEA Guidance (Records Management Guidance for Agencies Implementing Electronic Signature Technologies), Washington, DC : National Archives and Records Administration, [2000]

U.S. INTERPARES Project (International Research on Permanent Authentic Records in Electronic Systems) ; Working Bibliographies (Authenticity, Metadata, Modeling, Preservation, Archival Education, NHPRC) <http://is.gseis.ucla.edu/us-inter pares/bibgraph.htm>

Electronic Discussion Group – Preservation

PADI Forum

Specifically dedicated to the exchange of news about digital preservation. To subscribe, send the message "SUBSCRIBE padiforum-I Your Full Name" to listproc@nla.gov.au

SECTION IV.

Work Group 4 Report: Metadata & Copyright

Work Group 4 investigated issues concerning metadata and copyright.

IV. A. Area of Investigation: Metadata

Introduction:

“Metadata” is a term increasingly spoken at any gathering of information professionals and is most commonly defined as “data about data.” This simplistic definition seems inadequate as the term, “metadata,” becomes progressively more complicated as theory is applied to the specific case and as metadata schema and syntax must be selected or combined successfully. A variety of groups/communities increasingly create metadata records to describe, use, and preserve digital information objects. A metadata record, in theory, can describe **everything that is known** about texts, images, sound collections, or databases, to name just a handful of possible materials. A metadata record can range from the simple to the complex – from a simple resource description such as a MARC record, to a highly detailed description of a set of digital objects such as a database.

In spite of the ambiguities, common characteristics do prevail because metadata records are used to identify, locate, retrieve, and preserve a resource. When planning for metadata records for a digital project, determine the complexity of the metadata to be used to describe the material, then match the complexity desired with a metadata schema. How does one determine the level of complexity and the appropriate schema? There is no simple answer. Many factors contribute to these decisions, but **the intended use of the metadata and the characteristics of the item described** are the most important factors for consideration. The chart below, reproduced from an article¹⁴ by Anne S. Gilliland-Swetland, provides an excellent breakdown of the different types of metadata and their functions. Metadata types include administrative, preservation, technical, descriptive, and use. Some models fold the preservation and technical metadata into the administrative metadata.

¹⁴ Gilliland-Swetland, Anne S. *Introduction to Metadata: Setting the Stage*. Available: http://www.getty.edu/research/institute/standards/intrometadata/2_articles/index.html [May 23, 2002].

Type	Definition	Examples
Administrative	Metadata used in managing and administering information resources	<ul style="list-style-type: none"> - Acquisition information - Rights and reproduction tracking - Documentation of legal access requirements - Location information - Selection criteria for digitization - Version control and differentiation between similar information objects - Audit trails created by recordkeeping systems
Preservation	Metadata related to the preservation management of information resources	<ul style="list-style-type: none"> - Documentation of physical condition of resources - Documentation of actions taken to preserve physical and digital versions of resources, e.g., data refreshing and migration
Technical	Metadata related to how a system functions or metadata behave	<ul style="list-style-type: none"> - Digitization information, e.g., formats, compression ratios, scaling routines - Tracking of system response times - Authentication and security data, e.g., encryption keys, passwords
Descriptive	Metadata used to describe or identify information resources	<ul style="list-style-type: none"> - Finding aids - Specialized indexes - Hyperlinked relationships between resources - Annotations by users - Metadata for recordkeeping systems generated by records creators
Use	Metadata related to the level and type of use of information resources	<ul style="list-style-type: none"> - Exhibit records - Use and user tracking - Content re-use and multi-versioning information

Administrative Metadata:

Administrative metadata enables the management of an information resource. It includes information about the “who, what, when, where” of a resource – donor information, validity dates, metadata location, or ownership of rights. Administrative metadata might also note responsibility for review and update of a resource and outline a schedule for updates. Many metadata element sets contain administrative elements. Dublin Core, one of the best know element sets for descriptive metadata, is also building their Administrative Dublin Core (A-

Core) Element set.¹⁵ The A-Core set addresses the who, what, and where of *metadata assignment*. Examples of elements included in it are:

Name	Identifier
Identifier	identifier
Definition	A string or a number, which identified the metadata record
Obligation	Optional
Comment	Can be the internal number in a database

Name	Scope
Identifier	scope
Definition	Declaration of the scope of application
Obligation	Optional
Comment	Will often be declared by means of a separate form.

Name	Name
Identifier	name
Definition	The name of the entity responsible for undertaking a defined action on the content metadata
Obligation	Optional
Comment	Examples of Name include a person, an organization, or a service. Where the person has an affiliation with an organization, this information may be included. The name of a person should be provided in reverse order, that is, last name before first name, with a comma separator.

Name	Activity
Identifier	activity
Definition	The action performed on the content metadata by the responsible entity
Obligation	Optional
Comment	The actions are taken from a non-exhaustive list including: created, submitted, modified, checked, link collected, resource harvested, expired, mail sent. This list may be seen as showing the history of actions. (See also the element: Handling). Other sources may be used for the activity values such as codes from the USMARC Relator List [3]

Another example of administrative metadata elements contained in a metadata schema is the Encoded Archival Data (EAD) *Document Type Definition (DTD)*. This is a standard for encoding archival finding aids using the Standard Generalized Markup Language (SGML).¹⁶ The element called Administrative

¹⁵ Hansen, Jaytte and Lief Andresen (September 10, 2001). *Administrative Dublin Core (A-Core) Element: A proposal to be discussed in the DCMI Administrative Metadata Working Group*. Available: <http://dublincore.org/groups/admin/proposal-20010910.shtml> [May 23, 2002].

¹⁶ *Encoded Archival Description (EAD) Document Type Definition (DTD)*. Library of Congress. Available: <http://lcweb.loc.gov/ead/> [May 23, 2002].

Information <admininfo> “contains descriptive background information concerning an institution's acquisition, processing, and management of a body of archival materials. The <admininfo> element designates facts about provenance, acquisition, access and reproduction restrictions, availability of microform and digital surrogates, preferred form of citation, and other descriptive details that help readers of the finding aid know how to approach the archival materials and make use of the information they find.”

Preservation Metadata:

Often used broadly to include administrative, technical or structural, and preservation metadata, the term “*preservation metadata*” is sometimes more narrowly defined as metadata created to document the preservation processes performed on information resources. Preservation metadata records preservation actions taken (migration or emulation), notes the effects of actions taken, and ensures authenticity of an information resource over time.

An OCLC/RLG working group reviewed the state of the art for preservation metadata in a paper published in 2001.¹⁷ In addition to outlining “high-level requirements for a broadly applicable, comprehensive preservation metadata framework,” the working group reviewed several existing preservation metadata approaches. CEDARS¹⁸, the National Library of Australia¹⁹, and the NEDLIB²⁰ approaches were compared and contrasted. The working group found many similarities and concluded that work should continue toward building a consensus for a standard preservation metadata element set. Until such a consensus can be reached and a standard element set created, project managers may choose to use one of the three (CEDARS, NLA, or NEDLIB) sets examined in the white paper, or make decisions based on the use of another standard schema.

¹⁷ *Preservation Metadata for Digital Objects: A Review of the State of the Art. A White Paper by the OCLC/RLG Working Group on Preservation Metadata.* Available: <http://www.oclc.org/research/pmwg/documents.shtm> [May 23, 2002].

¹⁸ *Metadata For Digital Preservation: The Cedars Outline Specification.* Available: <http://www.leeds.ac.uk/cedars/OutlineSpec.htm> [May 21, 2002]

¹⁹ National Library of Australia, *Preservation Metadata for Digital Collections, Exposure Draft* Available: <http://www.nla.gov.au/preserve/pmeta.html> [May 21, 2002]

²⁰ Lupovici, Catherine and Julien Masanès. *Metadata for long term-preservation.* Bibliothèque nationale de France, : NEDLIB Consortium, July 2000. Available: <http://www.kb.nl/coop/nedlib/results/D4.2/D4.2.htm> [May 21, 2002]

Technical Metadata:

Technical metadata provides the description of data within an IT infrastructure. Technical metadata, which is sometimes considered part of administrative metadata, includes information about where the data is located, how to access the data, and what kinds of data types are stored. Examples of technical metadata include:

- Information on hardware.
- Information on software – version, etc.
- If the item was digitized, what formats were used.
- System response time – how long does it take to render data table?
- Security and user authentication information – passwords, who has access, etc.
- Data structure -- what is the structure of the stored data?

Descriptive Metadata:

Descriptive or *discovery* metadata allows users to discover and identify a resource. It is probably the most familiar type of metadata to the library community. Metadata such as author, title, subject, fall into this type of metadata. Libraries have been producing value-added materials (indexes, abstracts, catalog records, etc.) to create access to materials for their users for a very long time. The explosive growth of networked resources has made descriptive metadata an essential part of discovering materials in a networked environment.

The complexities of descriptive metadata assignment and maintenance should not be underestimated in a digitization project. Understanding the purpose of a particular project is key to making decisions about descriptive metadata. Also critical is an understanding of current technical capabilities.

Using metadata to enhance searching a web site, a database, or a specific digital object may be completely different than using metadata as a resource discovery tool using Internet search engines. Determining effective site searching and the capacities of Internet search engines is complicated and not standardized.

Searching within a web site

Based on descriptive metadata fields, users of a Web site typically are offered a variety of options in order to discover and use resources on the site. Examples of these options include:

- *Lists of resources* – simple list, often in html, that point to useful resources. These are often manually created. A good example of this

is the Documents Center Web site at the University of Michigan (<http://www.lib.umich.edu/govdocs/>).

- *Search* – Searching can be by keyword or controlled vocabulary. This type of descriptive metadata can either be manually created or automated. The *Catalog of U.S. Government Publications*, a search and retrieval service that provides bibliographic records of U.S. Government information products (http://www.access.gpo.gov/su_docs/locators/cgp/index.html) is a good example of this type of descriptive metadata.
- *Browsing* -- users can browse descriptive metadata which use a specific classification scheme. *Congressional Universe's* controlled subject access, from LEXIS/NEXIS, is an example of this type of descriptive metadata.

Dublin Core is one of the most widely used descriptive metadata schemas. “The Dublin Metadata Core Element Set was proposed as the minimum number of metadata elements required to facilitate the discovery of document-like objects in a networked environment such as the Internet.”²¹ The Dublin core schema can be broken down into three distinctive elements:

Elements related to the contents	Elements related to intellectual property	Elements related to physical manifestation
Title	Author or Creator	Date
Subject	Publisher	Type
Description	Contributor	Format
Source	Rights	Identifier
Language		
Relation		
Coverage		

(from Taylor, Chris. An Introduction to Metadata. <http://www.library.uq.edu.au/iad/ctmeta4.html>)

As stated above, some discovery metadata is created manually and involves an intellectual process (for example, the mapping of *GAO Reports* to LC Subject Headings). Simpler discovery metadata records can be created automatically. Generally the simpler the records are the less expensive they are to create.

Use of Metadata by Internet Search Engines

There are important resource discovery issues that will affect descriptive metadata effectiveness in searching the Internet.

²¹ Weibel, Stuart. *Metadata: the Foundations of Resource Description*. d-lib magazine, July 1995.

The current status of Internet search engines and their utilization of metadata in their search must be considered.^{22 23} Presently, most Internet search engines only make use of HTML metatags (author, title, description, keywords) if used at all. It is important to review search engine standards when making metadata decisions.²⁴

Many commercial search engines place a higher rank or priority on their “paid” subscribers, a growing trend among search engines such as GoTo, Yahoo, MSN, Excite, Ask Jeeves, About.com, Inktomi, Looksmart and Google. (Google’s sponsored links are at least so indicated in the search results.)²⁵

Some Internet search engines are beginning to index PDF files to allow searching (Google, BioWeb) and use the title-author-keyword meta tags inserted into a PDF file for discovery or display on a search engine hit list. Other search engines still do not have this capability.

It is also important to determine the capability of search engines to search the “deep web” or “invisible web” (proprietary, dynamic sites).^{26 27}

Other Descriptive Metadata Related Issues:

There are several resource discovery issues that will affect descriptive metadata effectiveness and should be monitored. These issues that may influence decisions about metadata for a project include:

²² Westera, Gillian. *Comparison of Search Engine User Interface Capabilities*. Available: <http://lisweb.curtin.edu.au/staff/gwpersonal/compare.html>. [May 21, 2002]

²³ *Search Engine Watch*. Available: <http://www.searchenginewatch.com/> [May 21, 2002]

²⁴ *Search engine standards for web site placement, ranking and positioning*. Available: <http://search-engines-web.com/> [May 12, 2002].

²⁵ Hansell, Saul. “Paid Placement Is Catching On in Web Searches.” *New York Times*, June 4, 2001. Available: <http://www.nytimes.com/2001/06/04/technology/04GOTO.html> [May 21, 2002].

²⁶ Warnick, Walter L., et al. “Searching the Deep Web Directed Query Engine Applications at the Department of Energy.” *D-Lib Magazine*. January, 2001, 7 (1). Available: <http://www.dlib.org/dlib/january01/warnick/01warnick.html>. [May 21, 2002].

²⁷ Lynch, Clifford A., “*Metadata Harvesting and the Open Archives Initiative*”, ARL Bimonthly Report 217, August 2001. Available: <http://www.arl.org/newsltr/217/mhp.html>. [May 21, 2002].

- The current status of Internet search engines and their utilization of metadata in their search.^{28 29} Presently, most Internet search engines only make use of HTML metatags (author, title, description, Keywords) if used at all. Review search engine standards when determining metadata for a site to balance effort with effective discovery when using Internet search engines.³⁰
- Many commercial search engines place a higher rank or priority on their “paid” subscribers, a growing trend among search engines such as GoTo, Yahoo, MSN, Excite, Ask Jeeves, About.com, Inktomi, Looksmart and Google. (Google’s sponsored links are at least so indicated in the search results.)³¹
- The current status of the “deep web” or “invisible web” (proprietary, dynamic sites) and their relationship to search and discovery tools.^{32 33}
- The use of metadata as a discovery tool for searching within a database to discover digital documents, after locating the database in question by a means other than metadata. Examples include the Texas Records and Information Locator (TRAIL) <http://www2.tsl.state.tx.us/servlet/TXDublinCore> and Bridges: Minnesota's Gateway to Environmental Information <http://bridges.state.mn.us>.
- The placement of metadata in a separate repository rather than being embedded within the object itself.
- Content management and publishing systems metadata tools to allow easier addition and editing of metadata entries. This is not yet a standard part of web publishing but should be an area for future

²⁸ Westera, Gillian. *Comparison of Search Engine User Interface Capabilities*. Available: <http://lisweb.curtin.edu.au/staff/gwpersonal/compare.html>. [October 23, 2001].

²⁹ Search Engine Watch. Available: <http://www.searchenginewatch.com/> [October 25, 2001].

³⁰ Search engine standards for web site placement, ranking and positioning. Available: <http://search-engines-web.com/> [May 12, 2002].

³¹ Hansell, Saul (June 4, 2001). “Paid Placement Is Catching On in Web Searches.” *New York Times*. Available: <http://www.nytimes.com/2001/06/04/technology/04GOTO.html> [January 2002].

³² Warnick, Walter L., et al. *Searching the Deep Web Directed Query Engine Applications at the Department of Energy*. D-Lib Magazine January, 2001, 7 (1). <http://www.dlib.org/dlib/january01/warnick/01warnick.html>. Accessed 4/11/01

³³ Lynch, Clifford A., *Metadata Harvesting and the Open Archives Initiative*, ARL Bimonthly Report 217, August 2001. <http://www.arl.org/newsltr/217/mhp.html>. Accessed 10/30/01

development. Because of typographical errors, outdated terms, misspellings, and other errors in metadata entry, editing and effective management of metadata is difficult. However, it is of vital importance for document discovery.³⁴

“Use” Metadata:

Use metadata tracks users of a resource and how the resource is being used. Typically these are “web access logs” that track who accesses a web-site and how. Use metadata is very sensitive. It can reveal who is accessing a system creating privacy issues for the user. Examples of use metadata include:

- At what date and time did a user access the system?
- How long was a user session?
- How did a user navigate the site?

Metadata Schemas and Encoding Syntax Schemes:

For each type of metadata described above there are many choices of metadata schemas, and the number continues to grow. Each metadata schema is generally intended for use for a specific type of resource. The type of resource and the specific goals of a project determine the metadata schema used for a project.

The list of metadata schemas below is taken from “A Review of Metadata: A Survey of Current Resource Description Formats,” by Lorcan Dempsey and Rachel Heery, et al.³⁵

Schema	Brief Description
BibTex	A program that produces the source list for a document, obtaining the information from a bibliographic database.
Categories for the Description of Works of Art (CDWA)	Developed for the communities that provide and use art information (e.g. museums and archives) and provide a structure for information used to describe works of art and images of them.

³⁴ *Metadata and Search*. Available: <http://www.searchtools.com/info/metadata.html>, [May 13, 2002].

³⁵ <http://www.ukoln.ac.uk/metadata/desire/overview/>

Report to ALA GODORT – Digitization of Government Information

Schema	Brief Description
CIMI (Computer Interchange of Museum Information)	Encompasses interchange protocols, interchange formats, and lower level network and telecommunications building blocks as well as content data standards.
Data Documentation Initiative (DDI)	An international criterion and methodology for the content, presentation, transport, and preservation of "metadata" about datasets in the social and behavioral sciences.
Dublin Core	A metadata record to describe networked electronic information.
Encoding Archival Description (EAD)	Developed for use with archives and manuscripts collections.
EELS Metadata Format	Provides an information system for quality assessed information resources on the Internet.
EEVL Metadata Format	Provides an Internet gateway to quality information resources in Engineering.
FGDC - Content Standards for Digital Geospatial Metadata	Common set of terminology and definitions for the documentation of geospatial data.
Government Information Locator Service (GILS)	Provides the general public and its own employees with a means for locating useful information generated by the many government agencies.
IAFA/whois++ Templates	Constructs a record format, which could be used by FTP archive administrators to describe the various resources available from their own archives.
LDAP Data Interchange Format (LDIF)	Primarily used for White Pages type applications, such as searching for email and postal addresses, and telephone numbers.
MARC (overview)	Standardized format for cooperating libraries to exchange and share catalogue records.
METS (Metadata Encoding and Transmission Standard)	A standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library.
USMARC	A cataloguing manual produced by the Library of Congress.
UKMARC	Published by the British Library.
UNIMARC	Guidelines resulting from meetings of the IFLA Permanent UNIMARC Committee and the requirements of the International Standard Bibliographic Description for Computer Files, ISBD (CF).

Schema	Brief Description
Summary Object Interchange Format (SOIF)	Records designed to be generated by Harvest gatherers and then used for user searches by Harvest brokers.
Text Encoding Initiative (TEI) Independent Headers	A set of generic guidelines for the representation of textual materials in electronic form, in such a way as to enable researchers in any discipline to interchange and re-use resources, independently of software, hardware, and application area.

Additional schema may be found on the International Federation of Library Associations (IFLA) Web site called "Digital Libraries: Metadata Resources" (<http://www.ifla.org/II/metadata.htm>).

Syntax is the encoding scheme used to allow the metadata for a project to be processed by a computer program. *Syntax* is not part of the metadata schema, but the two must understand each other for the data to be usable. Chris Taylor, in her *Introduction to Metadata*,³⁶ describes the following important syntax schemes:

- [HTML](#) (Hyper-Text Markup Language in Web pages, version 3.2 or 4.0)
- [SGML](#) (Standard Generalized Markup Language)
- [XML](#) (eXtensible Markup Language)
- [RDF](#) (Resource Description Framework)
- [MARC](#) (MAchine Readable Cataloging)
- [MIME](#) (Multipurpose Internet Mail Extensions)
- [Z39.50](#)
- [LDAP](#) (Lightweight Directory Application Protocol)

Taylor also states that, "Metadata may be deployed in a number of ways:

- Embedding the metadata in the Web page by the creator or their agent using META tags in the HTML coding of the page.
- As a separate HTML document linked to the resource it describes.
- In a database linked to the resource. The records may either have been directly created within the database or extracted from another source, such as Web pages."

³⁶ Taylor, Chris. *An Introduction to Metadata*, Revised 1 April 1999. Available: <http://www.library.uq.edu.au/iad/ctmeta4.html> [May 23, 2002].

Crosswalks

It is important to consider the flexibility that crosswalks between metadata standards may offer a project. Crosswalks define a mapping between metadata sets. The project manager may wish to investigate current best practices for the development of crosswalks during the project planning. Some resources include:

- MetaNet ³⁷,
- Harmony ABC model ^{38 39},
- External mapping between metadata formats by Michael Day ⁴⁰, and
- Crosswalk between MARC and Dublin Core ^{41 42}.

Changes to a Document in the Preservation Process

There are standard definitions for changes in documents and how much a document can change before it becomes a “new” item. The OAIS Reference Model discusses some of the issues of “new versions” in section 5.1.3.4 – Transformation. It states:

Digital Migrations that require some changes to the Content Information or PDI (Preservation Description Information) are referred to as Transformations. These changes will be to some of the bits in the primary Digital Object of the Content Information or PDI with corresponding changes in the associated Representation Information. In all cases the intent is to provide maximum information preservation because the resulting AIP (Archival Information Packet) is intended to be a full

³⁷ Hunter, Jane. “MetaNet – A Metadata Term Thesaurus to Enable Semantic Interoperability Between Metadata Domains,” *Journal of Digital Information*, 1:8, 2001-02-08. Available: <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Hunter/>. [May 21, 2002]

³⁸ The Harmony Project Home Page. Available: <http://www.ilrt.bris.ac.uk/discovery/harmony/>. [May 21, 2002].

³⁹ Lagoze, C., J. Hunter and D. Brickley . "An Event-Aware Model for Metadata Interoperability". *ECDL 2000*, Lisbon, September.

⁴⁰ Day, Michael. *Metadata: Mapping between Metadata Formats*. Available: <http://www.ukoln.ac.uk/metadata/interoperability/>. [May 21, 2002].

⁴¹ Library of Congress Network Development and MARC Standards Office. *MARC to Dublin Core Crosswalk*. Available: <http://www.loc.gov/marc/marc2dc.html> [May 21, 2002].

⁴² Library of Congress, Network Development and MARC Standards Office. *Dublin Core/MARC/GILS Crosswalk*. Available: <http://lcweb.loc.gov/marc/dccross.html> [May 22, 2002].

replacement for the AIP that is undergoing Transformation. The new AIP qualifies as a new **Version** of the original AIP. ⁴³

Section 5 of the OAIS Reference Model discusses this topic at length and is recommended for review during a digital project when issues arise about the point at which a file becomes a new version of a document if changes are made in the file.

General Overview of Best Practices:

Gilliland-Swetland, in her *Introduction to Metadata*,⁴⁴ outlines key questions about metadata that should be answered by a digital project manager. They involve the following areas:

- Identify the metadata schema or schemas for use in the project, depending on the specific requirements of the project undertaken. Be certain to use the most current version of the schema selected.
- Determine the metadata essentials required to achieve the project goals and the level of granularity needed for each type of metadata. Balancing costs and staffing for “developing and managing metadata to meet current needs” with “creating sufficient metadata that can be capitalized upon for future, often unanticipated uses” may present difficult decisions.

Other critical decision areas:

- Do search engines have the technical capacity to make use of assigned metadata, either at a web site or database, or across the Internet?
- Is the project cost effective? The adage “garbage in garbage out” has relevance to metadata assignment. The cost effectiveness and value to users should be the ultimate goal of any undertaking of a metadata project. Ongoing maintenance and documentation considerations may be key to the success of a metadata project.

⁴³ *Draft CCSDS Recommendations for an OAIS Reference Model (CCSDS 650.0-R-2)*. July 2001. Section 5: Preservation Perspectives. Available: <http://www.ccsds.org/rpa117/rpa117.html>. [May 21, 2002].

⁴⁴ Gilliland-Swetland, Anne S. *Introduction to Metadata: Setting the Stage*. Available: http://www.getty.edu/research/institute/standards/intrometadata/2_articles/index.html [May 23, 2002].

Relevant Electronic Discussion Group:

List Name : dc-general

List description:

The broadest of mailing lists related to the international Dublin Core effort. Unlike other lists, which relate to the tasks of specific working groups or special interest areas, this list is for discussion of all issues relevant to the development, deployment, and use of Dublin Core metadata.

To subscribe, send a message to: dc-general-request@mailbase.ac.uk

IV. B. Area of Investigation: Copyright

This section is a brief overview of copyright, based on current federal copyright provisions and inclusion of recent relevant references relating to digital copyright issues.^{45 46}

General Recommendations

- Assume that works are copyrighted except for works by the United States Government. Also be aware that some works may be labeled as protected by copyright when in fact they are not.
- If re-publishing of U.S. Government works is to occur, inquiry should be made to the agency to determine if a contractor or a U.S. Government employee created the work.
- When in doubt about using works that may infringe on copyright protections, link to a U.S. Government web site. Linking to a government site does not have the intent of infringing on copyright protection and there is reasonable belief that linking would be safe.
- Be aware of contractual restrictions or issues. In contractual arrangements, think about the relationship between Government agencies and the contractor and who is going to own the product before starting the project.
- For foreign official publications, remember that intellectual property rights are governed by national law (parallel to US state publications being governed by state law).
- For international governmental organization (IGO) publications, remember that copyright in general is governed by provisions of law in the country of the organization's headquarters.

⁴⁵ Summary of Copyright issues is at <http://www.ala.org/washoff/copyrightataglance.html>. Includes Federal, State, Digital Millennium Copyright Act, and International copyright issues.

⁴⁶ ALA Washington Office, *Copyright at a Glance*. Available: <http://www.ala.org/washoff/copyrightataglance.html>

- State and local government publications may or may not be copyrighted. Before beginning a digital project, check your state's codes and regulations and contact the state agency that published the material. For local government publication (cities or towns) and regional governmental organizations (Council of Governments, Airport Authorities, etc.), check with the publishing agency to be certain a specific publication is not copyrighted.

In general, United States government works do not have a copyright. Thus, retrieving, copying and digitizing information published by the federal government in print or on their Web sites is usually not a property rights issue. However, if that information is accessible from a non-U. S. government server, in which links or tags are added to a document retrieved from a government site, the organization providing access to that electronic document could have a copyright claim to the document on its site. While most federal government works are free from copyright, the Government Printing Office sells data to other publishers who use the data for their print and online products. These products may claim copyright on any enhancements they may incorporate into that publication, such as forewords, indexes, compilations, etc..

“WHAT WORKS ARE PROTECTED?”

“Copyright protects "original works of authorship" that are fixed in a tangible form of expression. The fixation need not be directly perceptible so long as it may be communicated with the aid of a machine or device. Copyrightable works include the following categories:

- literary works
- musical works, including any accompanying words
- dramatic works, including any accompanying music
- pantomimes and choreographic works
- pictorial, graphic, and sculptural works
- motion pictures and other audiovisual works
- sound recordings
- architectural works

“These categories should be viewed broadly. For example, computer programs and most "compilations" may be registered as "literary works"; maps and architectural plans may be registered as "pictorial, graphic, and sculptural works.”⁴⁷

⁴⁷ U. S. Copyright Office (May 7, 2002). *Copyright Basics*. Available: <http://www.loc.gov/copyright/circs/circ1.html> [May 16, 2002].

“WHAT IS NOT PROTECTED BY COPYRIGHT?”

“Several categories of material are generally not eligible for federal copyright protection. These include among others:

- Works that have **not** been fixed in a tangible form of expression (for example, choreographic works that have not been notated or recorded, or improvisational speeches or performances that have not been written or recorded)
- Titles, names, short phrases, and slogans; familiar symbols or designs; mere variations of typographic ornamentation, lettering, or coloring; mere listings of ingredients or contents
- Ideas, procedures, methods, systems, processes, concepts, principles, discoveries, or devices, as distinguished from a description, explanation, or illustration
- Works consisting **entirely** of information that is common property and containing no original authorship (for example: standard calendars, height and weight charts, tape measures and rulers, and lists or tables taken from public documents or other common sources).”⁴⁸

Considerations in Title 17 of the U. S. Code:

While Title 17, Section 105 of the U. S. Code does not provide copyright protection for works of the U. S. Government, the Government is not precluded from receiving and holding copyrights transferred by assignment, bequest or otherwise. This could affect material reprinted from other sources, such as newspapers. This could also affect copyrighted tables/information in publications, and other copyrighted materials such as NTIS copyright protected data.^{49 50}

The above general rule may not apply in foreign countries. Works of the government in most other countries are copyrighted, i.e., the laws pertaining to copyright of a country would apply in that jurisdiction for U. S. Government materials.⁵¹ When planning a scanning project for documents published by a

⁴⁸ Ibid.

⁴⁹ Jensen, Mary Brandt (1996). *Does Your Project Have a Copyright Problem? A Decision-Making Guide for Librarians*. Jefferson, North Carolina: McFarland.

⁵⁰ Hoon, Peggy, ed (1997). *Guidelines for Educational Use of Copyrighted Materials*. Pullman, Washington: Washington State University Press.

⁵¹ *Historical and revision notes*, 17 USC 105.

foreign country's government, a library should consult with that particular agency to determine if the information has a copyright. If the document is copyrighted, the library should obtain permission from that agency before digitizing and disseminating the document. For international governmental organization (IGO) publications, remember that copyright in general is governed by provisions of law in the country of the organization's headquarters

Remember also that the compilation or arrangement of information may be copyrighted.⁵²

Notice of copyright:

When a Government work is republished commercially, the copyright notice should be meaningful and not misleading. Section 403 requires that when works consist "preponderantly of one or more works of the United States Government, the copyright notice (if any) must *identify those parts of the work in which copyright is claimed*. An omission to meet this requirement would be treated as an omission of the notice, subject to the provisions of Section 405." [Emphasis added.]⁵³

Emerging issues

When copying digital works created by the U. S. Government, should one go to a federal agency site or a private organization's site? A private Web site may have value added features or may claim copyright protection related to markup language, fonts, arrangement, etc.

Creators sometimes claim copyright protection after spending money on abstracting information and presenting access through a database.

Any underlying original authorship could present problems, e.g., software coding, may preclude a public domain document.

Who owns the finished product? The "producing" agency that has repackaged it or the original creator?

How does one protect "declared" copyright?

⁵² *Historical and revision notes*, 17 USC 101.

⁵³ *Historical and revision notes*, 17 USC 403.

Best Practices

- Be aware that documents may be copyrighted. We suggest that you inquire at the publishing agency about copyright restrictions and consult an attorney if you have questions.
- Be clear about the copyright, licensing arrangements or contract restrictions made with or assumed by software providers or other non-U. S. government content providers.
- In consortial arrangements for producing digital contents, be certain that the ownership and responsibility issues are clearly stated and agreed to before the project is undertaken.

Additional Resources

Copyright Law of the United States of America from the U.S. Copyright Office.⁵⁴
Public Domain issues⁵⁵
IFLA Copyright Information⁵⁶
College and University Copyright information⁵⁷
Yale Copyright Resources⁵⁸
Digital Millennium Copyright Act⁵⁹
Library of Congress Digital Memory Project⁶⁰
UNESCO⁶¹

⁵⁴ U.S. Copyright Office (July 2001). *Copyright Law of the United States of America*. Available: <http://www.loc.gov/copyright/title17/> [May 17, 2002].

⁵⁵ Hatmaker, Scott. "Government Information: Public Domain or User Beware." *Information Outlook* 1, no.12 (December 1997): 39.

⁵⁶ International Federation of Library Associations (November 11, 2001) *Information Policy: Copyright And Intellectual Property*. Available: <http://www.ifla.org/ll/copyright.htm> [May 16, 2002].

⁵⁷ Lide, Casey. "What Colleges and Universities Need to Know about the Digital Millennium Copyright Act". *CAUSE/EFFECT Journal*, V. 22 no. 1, 1999. Available: <http://www.educause.edu/ir/library/html/cem9913.html> [May 16, 2002].

⁵⁸ *Copyright Resources Online*. Available: <http://www.library.yale.edu/%7Eokerson/copyproj.html> [May 16, 2002].

⁵⁹ U.S. Copyright Office Summary (December 1998). *The Digital Millennium Copyright Act of 1998*. Available: <http://lcweb.loc.gov/copyright/legislation/dmca.pdf> [May 17, 2002].

⁶⁰ Library of Congress, American Memory (April 14, 2000). *Building digital collections: technical information and background papers*. Available: <http://memory.loc.gov/ammem/ftpfiles.html> [May 17, 2002].

Digital Watermarking ⁶²

Digital Copyright issues ⁶³

⁶¹ Oppenheim, C. (1997). "Copyright in the electronic age." In: *UNESCO World Information Report*. Unesco, Paris. Available:
http://www.unesco.org/webworld/com_inf_reports/wirenglish/chap26.pdf [May 16, 2002].

⁶² Berghel, H. and L. O'Gorman (January 2, 1997). *Digital Watermarking*. Available:
http://www1.acm.org:82/~hlb/publications/dig_wtr/dig_watr.html [May 16, 2002].

⁶³ Balas, Janet. 1998. "Copyright in the Digital Era," *Computers in Libraries* 18 (6) June 1998.

Appendix D: Bibliography of Resources

The Bibliography of Resources first contains general resources important to all areas of digital imaging and digital preservation. This section is followed by a list of resources arranged in the same order as the discussion of issues in Appendix C: Issues and Best Practices Report.

General Resources for Digital Imaging and Preservation:

General Resources Bibliography

Kenney, Anne R., Oya Y. Rieger, and Research Libraries Group. *Moving Theory into Practice : Digital Imaging for Libraries and Archives*. Mountain View, CA: Research Libraries Group, 2000.

Sitts, Maxine K. ed. *Handbook for Digital Projects : A Management Tool for Preservation and Access*. 1st ed. Andover, Mass.: Northeast Document Conservation Center, 2000. Available: <http://www.nedcc.org/digital/dighome.htm> [May 21, 2002].

Institute of Museum and Library Services (U.S.). *A Framework of Guidance for Building Good Digital Collections*. Institute of Museum and Library Services, 2002. Available: <http://www.ims.gov/pubs/forumframework.htm> [May 14, 2002].

Digital Library Federation Web Site <http://www.diglib.org/>
The DLF site contains extensive information about developing digital collections and managing networked information. Presentations about digital projects presented at the RLG spring and fall forum may also be viewed.

Electronic Discussion Groups

Digital Libraries Research Forum (DigLib)
To subscribe, send the message "SUBSCRIBE diglib Your Full name" to listserv@infoserv.nlc-bnc.ca

Digital Librarianship (DigLIBNS)
To subscribe, send the message "SUBSCRIBE diglibns Your Full Name" to <mailto:listserv@sunsite.berkeley.edu>

Selection Issues:

Selection Issues Bibliography:

Columbia University (January 14, 2001). *Libraries Selection Criteria for Digital Imaging*. Available: <http://www.columbia.edu/cu/libraries/digital/criteria.html> [May 14, 2002].

Cornell University (2000-2001). *Moving Theory into Practice: Digital Imaging Tutorial, Part 2: Selection*. Available: <http://www.library.cornell.edu/preservation/tutorial/selection/selection-01.html> [May 14, 2002].

Outlines key concepts, addresses legal restrictions, other criteria, selection policies and additional reading

Institute of Museum and Library Services (U.S.). *A Framework of Guidance for Building Good Digital Collections*. Institute of Museum and Library Services, 2002. Available: <http://www.ims.gov/pubs/forumframework.htm> [May 14, 2002].

Hazen, Dan C., Jeffrey L. Horrell, Jan Merrill-Oldham, and Council on Library and Information Resources. *Selecting Research Collections for Digitization*. Washington, D.C.: Council on Library and Information Resources, 1998. Available: <http://www.clir.org/pubs/reports/hazen/pub74.html> [May 22, 2002].

Joint RLG and NPO Preservation Conference (1998). *Guidelines for Digital Imaging, 28-30 September 1998. Selection Criteria, Guidelines, Decision-Making Aids*. Available: <http://www.rlg.org/preserv/joint/selection.html> [May 14, 2002].

Links to general studies and discussion of selection criteria.

Joint RLG and NPO Preservation Conference (1998). *Guidelines for Digital Imaging. Guidance for Selecting Materials for Digitisation*, by Paul Ayris. Available: <http://www.rlg.org/preserv/joint/ayris.html> [May 14, 2002].

Outlines the questions that need to be addressed in formulating guidance.

Library of Congress (December 29, 1999). *Selection Criteria for Preservation Digital Reformatting*. Available: <http://lcweb.loc.gov/preserv/prd/presdig/presselection.html> [May 14, 2002].

Principles on which selection of materials is based.

Preserving Access to Digital Information (PADI) (May, 15, 2002). *Policies, Strategies and Guidelines*. Available: <http://www.nla.gov.au/padi/format/policy.html> [May 15, 2002].

Lists and links to international, Australasia, Europe, and North American resources which provide policies, strategies and guidelines for digital preservation.

Sitts, Maxine K., Editor. *Handbook for Digital Projects: A Management Tool for Preservation and Access*. Andover, Mass.: Northeast Document Conservation Center, 2000. Available:

<http://www.nedcc.org/digital/dighome.htm> [May 14, 2002].

Chapter IV focuses on Selection of Materials for Scanning, written by Diane Vogt O'Connor.

Smith, Abby, Council on Library and Information Resources, Commission on Preservation and Access, and Council on Library and Information Resources. *Digital Libraries. Why Digitize?* Washington, D.C.: Council on Library and Information Resources, 1999. Available:

<http://www.clir.org/pubs/reports/pub80-smith/pub80.html> [May 14, 2002].

Discusses what digital information is and is not.

University of California, Santa Barbara (October 8, 1998). *University of California Selection Criteria for Digitization*. Available;

<http://www.library.ucsb.edu/ucpag/digselec.html> [May 14, 2002].

List of criteria recommended to guide collection development librarians and preservation librarians in selecting collections of analog materials for conversion to digital format.

Collaboration Issues Bibliography:

Allen, Nancy. "Collaboration through the Colorado Digitization Project." *First Monday* 5, no. 6 (2000). Available: <http://www.ims.gov/pubs/wwcp5.htm> [May 14, 2002].

Discusses the digital efforts emerging from a collaboration among school, public, and university libraries, museums, archives, and historical societies from all over the state of Colorado. Issues involved in this large collaboration include varying audiences, resources, communication, legal matters and preservation.

Colorado Digitization Project (June 1999). *Collection Policy*. Available:

<http://coloradodigital.coalliance.org/select.html> [May 14, 2002].

Defines the guiding policies of the Project, the audience, subject matter, and participants.

Dowden, Robin, et al. *ArtsConnectEd: Collaboration in the Integration and Access to Museum Resources*. Available:

<http://www.ims.gov/pubs/wwcp2.htm> [May 14, 2002].

Case study of a collaboration between three museums. Discusses the project management and evaluation, as well as future issues. A sizable portion of the report is devoted to audience needs and communication among collaborators.

Stack, Jacqueline M. (1999). *Collaboration: Success for the Future*. Los Alamos National Laboratory. Available:
<http://www.osti.gov/inforum99/papers/stack.pdf> [May 14, 2002].

Funding Sources Bibliography:

Andrew W. Mellon Foundation (<http://www.mellon.org>)
Private foundation provides funding for projects for art conservation research and training as well as for the digitization of library and archival materials. Program Officer, Scholarly Communication: Don Waters, djw@mellon.org

Colorado Digitization Project Funding Sources
(<http://coloradodigital.coalliance.org/funding.html>)
Identifies public/federal sources, associations and organizations, private foundations, and information resources. Includes links to many of the sources listed below.

Getty Grant Program (<http://www.getty.edu/grants>)
A possible source of funding is the research grant, particularly for projects that "make visual arts collections, archives, research results, and other critical resources available via electronic means."

Institute of Museum and Library Services (<http://www.imls.gov>)
Application forms and program guidelines may be downloaded from the IMLS web site. Brief description of specific grant programs sponsored by IMLS include:

- *National Leadership Grants/Education and Training:* Funds projects that ...implement innovative approaches to education and training and enhance the availability of professional librarians with advanced skills and specializations...
Deadline: February 1.
- *National Leadership Grants/Research and Demonstration:* Model projects that conduct research and/or demonstrations to enhance library services through the effective use of new and appropriate technologies, enhance the ability of library users to make more effective use of information resources, or assist in the evaluation of library services...
Deadline: February 1

- *National Leadership Grants/Preservation or Digitization:* Projects that preserve and enhance access to unique library resources useful to the broader community; address the challenges of preserving and archiving digital media; or that lead to the development of standards, techniques, or models related to the digitization and management of digital resources. (Georgia Dept. of Archives & History was a 2000 recipient to digitize and make available Georgia's House and Senate journals and other state government records)
Deadline: February 1
- *Grants to States:* The Library Services and Technology Act (LSTA) of 1996 promotes access to information resources provided by all types of libraries. Two broad priorities for LSTA funding: 1) activities using technology for information sharing among libraries and between libraries and other community services; and 2) programs that make library resources more accessible to urban and rural localities, and to low-income residents and others who have difficulty using library services. Funding is distributed by state library agencies; the amount of money available is based on the population of the state.
Deadline: New five-year plans due July 31, 2002

To obtain more information on the types of projects that IMLS funds, see: http://www.ims.gov/closer/cls_po.asp The Library Program Officer for IMLS is Martha Crawley, mcrawley@ims.gov .

National Endowment for the Humanities (NEH)
(<http://www.neh.gov/grants/index.html>)

The Division of Preservation and Access supports “applications for model projects that explore or resolve critical issues relating to the effective use of digital technology for preservation and access. The web site includes a “frequently asked questions” about grants page. Contact: info@neh.fed.us

National Historical Publications and Records Commission (NHPRC)
(<http://www.nara.gov/nhprc/>)

The NHPRC is the grant-making affiliate of the National Records and Archives Administration. NHPRC supports grants for the identification, preservation, and provision of public access to materials that document American history. For an indication of the type of projects funded, see <http://www.nara.gov/nhprc/projects.html>
Contact: nhprc@arch1.nara.gov

National Endowment for the Arts (NEA) (<http://www.nea.gov>)

NEA offers support for organizations in four areas---creation and presentation, planning and stabilization, heritage and preservation, and education and access. It supports projects that assist, preserve, document and present artists and forms of artistic expression that reflect our nation's diverse cultural traditions. Funding for conservation of important work is also possible.

Contact: 202-682-5400

Other possibilities for grant funding include state library agencies and foundations with regional or local focus. The *Foundation Directory*, available in many libraries, or online for a subscription fee, is another excellent source for identifying foundations that provide financial support to educational institutions for a variety of purposes.

Grant Writing Resources Bibliography:

Frequently Asked Questions

(<http://fdncenter.org/learn/faqs>)

A page from the Foundation Center that includes sections on finding funding, writing proposals, and examples of grant proposals.

Proposal Writing Short Course

(<http://www.fdncenter.org/learn/shortcourse/prop1.html>)

Step-by-step guide to writing a proposal with a short discussion of each section and what should be included.

Many of the funding sources' Web sites also include tips for proposal writing.

Training Resources Bibliography:

Schools and Seminars:

Colorado Digitization Project: Workshops & Seminars

(<http://coloradodigital.coalliance.org/workshop.html>).

Listing with description of seminars and workshops on various aspects of digitization.

Illinois Digitization Institute (<http://nautilus.outreach.uiuc.edu/ldi/workshop.asp>)

Offers a series of free one-day workshops on the fundamentals of digitization.

Project Management Research on the Web. Maintained by Umeå School of Business and Economics, Sweden. <http://www.fek.umu.se/irnop/projweb.html>
Provides links to research universities, project management indices, and professional associations.

Project Management Seminars. American Management Association.
<http://www.amanet.org/seminars/cmd2/Project.htm>
AMA sponsors seminars for project managers at three levels. Courses are expensive.

Articles and Books:

Cornell University Library, Department of Preservation and Conservation (2001). *Moving Theory into Practice: Digital Imaging Tutorial, Chapter 10, Continuing Education*. Available:
<http://www.library.cornell.edu/preservation/tutorial/education/education-01.html>.

Provides links to tutorials, workshops, and guides.

Jones, Maggie, Neil Beagrie, Resource: The Council for Museums Archives and Libraries, and British Library. *Preservation Management of Digital Materials : A Handbook*. London: The British Library for Resource the Council for Museums Archives and Libraries, 2001.

Includes chapters on institutional strategies and organizational activities.

National Digital Library Program (NDLP), Library of Congress (April 14, 2000). *Building Digital Collections: Technical Information and Background Papers*. Available: <http://memory.loc.gov/ammem/ftpfiles.html> [May 14, 2002].

Includes information on technical specifications, workflow and production.

Research Libraries Group (May 2002). *RLG Tools for Digital Imaging*. Available: <http://www.rlg.org/preserv/RLGtools.html> [May 14, 2002].

Includes worksheets and guidelines.

Scoping the Future of Oxford's Digital Collections (March 1999). *Appendix E: Digitization Methods*, by Stuart D. Lee. Available:

<http://www.bodley.ox.ac.uk/scoping/digitization.html> [May 14, 2002].

Outlines some of the main issues surrounding digitization.
Provides recommendations applicable to the strategic planning level.

Creation of Digital Copies:

Best Practices for Digitization Bibliography:

Archive Builders

<http://www.ArchiveBuilders.com/>

Provides an in depth understanding of document imaging as used in records management. Presentations are given on the computer industry as it relates to document imaging systems and records management for specific sites. Includes white papers, courses, handouts from presentations, etc.

Besser, Howard and Jennifer Trant. *Introduction to Imaging: Issues in Constructing an Image Database*. Santa Monica, CA: The Getty Art History Information Program, 1995.

Brancolini, Kristine R. "Selecting research collections for digitization: applying the Harvard model." *Library Trends*, v. 48, no. 4 (Spring 2000), p. 783-798.

Building Digital Collections: Technical Information and Background Papers (Library of Congress)

<http://memory.loc.gov/ammem/ftpfiles.html>

Documents technical activities related to the American Memory Collections produced by the National Digital Library Program (NDLP) at the Library of Congress. Includes Technical Practices, Workflow and Production, Rights and Restrictions Statements, Background Papers.

Colorado Digitization Center, Digital Toolbox

<http://coloradodigital.coalliance.org/toolbox.html>

Designed to guide administrators of digital projects to sites and resources on the following topics: Administrative Resources, Guidelines and Standards, Copyright and Intellectual Property, Funding, Metadata, Software, Planning, etc.

Coppock, T., Editor. *Making Information Available in Digital Format: Perspectives from Practitioners*. Edinburgh: The Stationery Office, 1999.

CoOL (Conservation OnLine) Digital imaging: Imaging and Imagebases

<http://palimpsest.stanford.edu/bytopic/imaging/>

Morrison, Alan S., Michael Popham, Karen Wikander, and Oxford Text Archive. *Creating and Documenting Electronic Texts, Ahd's Guides to Good Practice*. Oxford, England Oakville, CT: Oxbow Books ; David Brown Book Co., 2000. Available: <http://ota.ahds.ac.uk/documents/creating/>

D-Lib Forum

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

Facilitating and supporting the community developing the technology of the global digital library. See also D-Lib magazine.

Digital Imaging: Selected Bibliography. Solinet Preservation Leaflets.

<http://www.solinet.net/presvtr/leaf/imaging.htm>

Townsend, Sean, Cressida Chappell, Oscar Struijvâe, Arts and Humanities Data Service, and History Data Service. *Digitising History : A Guide to Creating Digital Resources from Historical Documents, Ahds Guides to Good Practice*. Oxford: Oxbow Books, 1999. Available:

http://hds.essex.ac.uk/g2gp/digitising_history/index.asp

Guidelines for Digital Imaging (RLG and National Preservation Office)

<http://www.rlg.org/preserv/joint/confpapers.html>

Conference papers from a Joint RLG and NPO Preservation Conference: Guidelines for Digital Imaging, 28-30 September 1998. Includes Guidelines Versus Guidance for Digital Imaging: the Opportunity Before Us, Guidelines for Image Capture, etc.

Kellerman, Lydia Suzanne, Rebecca A. Wilson, Laura Rounds, Association of Research Libraries. Office of Management Services, and Association of Research Libraries. Systems and Procedures Exchange Center. *Digitizing Technologies for Preservation : A Spec Kit, Spec Kit 214*. Washington, DC: Association of Research Libraries Office of Management Services, 1996.

Kenney, Anne R., Oya Y. Rieger, and Research Libraries Group. *Moving Theory into Practice : Digital Imaging for Libraries and Archives*. Mountain View, CA: Research Libraries Group, 2000.

Lavender, Kenneth, et. al. *Managing Preservation Technologies For the New Century*. Chicago, Ill., American Library Association. [Distributed by Teach'em], 2000. 2 sound cassettes.

Oxford Text Archive

<http://ota.ahds.ac.uk/>

Works closely with members of the Arts and Humanities academic community to collect, catalogue, and preserve high-quality electronic texts for research and teaching. See publications.

RLG DigiNews

<http://www.rlg.org/preserv/diginews/>

Majka, D. R. "The Seven Deadly Sins of Digitization. Online products being marketed to libraries." *Online*, v. 23, no. 2 (Mar./Apr. '99) p. 43-6.

Sitts, Maxine K., Editor. *Handbook for Digital Projects: A Management Tool for Preservation and Access*. Northeast Document Conservation Center. 2000. Available: <http://www.nedcc.org/digital/dighome.htm> [May 17, 2002].

Development of Standards

Archive Builders

<http://www.ArchiveBuilders.com/>

Provides an in depth understanding of document imaging as used in records management. Presentations are given on the computer industry as it relates to document imaging systems and records management for specific sites. Includes white papers, courses, handouts from presentations, etc.

Chapman, Stephen (May 23, 2001). *Report of Imaging Practitioners Meeting on 30 March 2001 to Consider How the Quality of Digital Imaging Systems and Digital Images May be Fairly Evaluated*. Available: <http://www.diglib.org/standards/imqualreppv.htm> [May 17, 2002].

Colorado Digitization Center, Digital Toolbox

<http://coloradodigital.coalliance.org/toolbox.html>

Designed to guide administrators of digital projects to sites and resources on the following topics: Administrative Resources, Guidelines and Standards, Copyright and Intellectual Property, Funding, Metadata, Software, Planning, etc.

Digital Library Federation, Digital Library Standards and Practices

<http://www.diglib.org/standards.htm>

Digital Library Standards.

<http://sunsite.berkeley.edu/Info/standards.html>

Includes draft standards or dreams of standards.

Digital Formats for Content Reproductions

<http://memory.loc.gov/ammem/formats.html>

Via the Library of Congress American Memory Project.

Graphics File Formats Page: 2D Bitmap Specifications (no date) [Online].

<http://www.dcs.edu.ac.uk/home/mxf/gfx/2d-hi.html> [2001, November 7]

Greenstein, Daniel (rev. 2001, July 30). *Report of a Meeting of the DLF on Preservation Reformatting Practices*. Available: <http://www.diglib.org/standards/presreformatsumpv.htm> [May 17, 2002].

Guides to Quality in Visual Resource Imaging. Council on Library and Information Resources, Digital Library Federation, Research Libraries Group 2000. <http://www.rlg.org/visguides/>

Image Quality Working Group of ArchivesCom, Joint Libraries/AcIS Committee (1997, April 2). *Technical Recommendations for Digital Imaging Projects*. Available: <http://www.columbia.edu/acis/dl/imagespec.html> [May 17, 2002].

Moving Theory into Practice: Digital Imaging Tutorial
<http://www.library.cornell.edu/preservation/tutorial/contents.html>

Tables:

Representative institutional requirements for conversion

<http://www.library.cornell.edu/preservation/tutorial/conversion/table3-1.html>

Common Image File Formats

<http://www.library.cornell.edu/preservation/tutorial/presentation/table7-1.html>

Attributes for Common Compression Techniques

<http://www.library.cornell.edu/preservation/tutorial/presentation/table7-3.html>

Comparison of Storage Media

<http://www.library.cornell.edu/preservation/tutorial/technical/table6-2.html>

Metadata Types

<http://www.library.cornell.edu/preservation/tutorial/metadata/table5-1.html>

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

World Wide Web Consortium

<http://www.w3.org/>

Develops interoperable technologies (specifications, guidelines, software, and tools) to lead the Web to its full potential as a forum for information, commerce, communication, and collective understanding. See A-Z list of topics.

Yale Style Manual-Graphics for the Web

<http://info.med.yale.edu/caim/manual/graphics/>

File Formats for Current Access and for Long-Term Preservation

Archive Builders

<http://www.ArchiveBuilders.com/>

Provides an in depth understanding of document imaging as used in records management. Presentations are given on the computer industry as it relates to document imaging systems and records management for specific sites. Includes white papers, courses, handouts from presentations, etc.

Berkeley Digital Library SunSITE

<http://sunsite.berkeley.edu/Imaging/>

Digitizing images and text.

Colorado Digitization Center, General Guidelines for Scanning

<http://coloradodigital.coalliance.org/scanning.html>

Recommended standards for text, photographs, maps, and graphic materials.

Digital Formats for Content Reproductions (Library of Congress)

<http://memory.loc.gov/ammem/formats.html>

Organizes materials by format (pictorial, textual as searchable text/images or images, maps, sound-recordings, etc.)

Guides to Quality in Visual Resource Imaging, *File Format for Digital Masters*

Digital Library Federation's Research Libraries Group. Available:

<http://www.rlg.org/visguides/visguide5.html>.

Image Quality Working Group of ArchivesCom, a joint Libraries/AclS committee

(1997, April 2). *Technical Recommendations for Digital Imaging Projects*.

Available: <http://www.columbia.edu/acis/dl/imagespec.html> [November 7, 2001].

Sitts, Maxine K., Editor. *Handbook for Digital Projects: A Management Tool for*

Preservation and Access. Andover, Mass.: Northeast Document

Conservation Center, 2000. Available:

<http://www.nedcc.org/digital/dighome.htm> [January 2, 2002].

List of Vendors and Explore Outsourcing Possibilities for Digitization

“Choosing a Vendor: Tips from the Lason Corporation. Selecting a vendor for a Web digitization project.” *Texas Library Journal*. v. 75 no. 3 (Fall 1999) p. 107+.

Colorado Digitization Center, Digital Toolbox

<http://coloradodigital.coalliance.org/toolbox.html>

Designed to guide administrators of digital projects to sites and resources on the following topics: Administrative Resources, Guidelines and Standards, Copyright and Intellectual Property, Funding, Metadata, Software, Planning, etc.

Digital Imaging Services. Solinet Preservation Leaflets.

<http://www.solinet.net/presvtn/leaf/digimage.htm>

List of companies that provide digital imaging technology or services a few resources to aid in locating and re-evaluating potential vendors.

Digital Libraries: Metadata Resources (IFLANET)

<http://www.ifla.org/ll/metadata.htm>

Metadata resources for digital libraries (from the International Federation of Library Associations and Institutions). Excellent resources.

Tennant, Roy. "Outsourcing Digitization." *Library Journal*. v. 124 no. 15 (Sept. 15 1999) p. 34+

Reformatting Services. Solinet Preservation Leaflets.

<http://www.solinet.net/presvtn/leaf/reformat.htm>

Services include digital imaging, preservation microfilming, preservation photocopying and magnetic media reformatting.

ADA Issues Bibliography

Accessible Web Page Design

<http://www.makoa.org/web-design.htm>

Resources that may be helpful in creating pages that are truly accessible.

Adobe Systems

<http://www.adobe.com/>

CAST Bobby

<http://www.cast.org/bobby/>

Bobby was created by CAST to help Web page authors identify and repair significant barriers to access by individuals with disabilities.

Section 508

<http://www.section508.gov/>

Section 508 requires that Federal agencies electronic and information technology is accessible to people with disabilities. The Center for Information Technology Accommodation (CITA), in the U.S. General Services Administration's Office of Government-wide Policy, has been

charged with the task of educating Federal employees and building the infrastructure necessary to support Section 508 implementation.

World Wide Web Consortium

<http://www.w3.org/>

Develops interoperable technologies (specifications, guidelines, software, and tools) to lead the Web to its full potential as a forum for information, commerce, communication, and collective understanding. Two key ADA initiatives include Voice Browser Activity and the Web Accessibility Initiative.

Server-Side Issues Bibliography

Colorado Digitization Center

<http://coloradodigital.coalliance.org/>

Digital Object Identifier System

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

A system for identifying and exchanging intellectual property in the digital environment. It provides a framework for managing intellectual content, for linking customers with content suppliers, for facilitating electronic commerce, and enabling automated copyright management for all types of media.

Moving Theory Into Practice: Digital Imaging Tutorial

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

Excellent tutorial from the Department of Preservation and Conservation at the Cornell University Library. Includes Selection, Conversion, Quality Control, Metadata, Technical Infrastructure, Presentation, Digital Preservation, Management, Continuing Education.

General Resources

Association of American Publishers

<http://www.publishers.org/home/index.htm>

Covers both the general and the specific - broad issues important to all publishers as well as issues of specific concern to particular segments of the industry. The Association's "core" programs deal with matters of general interest: intellectual property; new technology and telecommunications issues of concern to publishers; First Amendment rights, censorship and libel; international freedom to publish; funding for education and libraries; postal rates and regulations; tax and trade policy. Includes conferences and publications.

Directory of Digitized Collections

<http://thoth.bl.uk/>

Digital Library Federation

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

Hazen, Dan C., Jeffrey L. Horrell, Jan Merrill-Oldham, and Council on Library and Information Resources. *Selecting Research Collections for Digitization*. Washington, D.C.: Council on Library and Information Resources, 1998.

International Federation of Library Associations and Institutions

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

Solinet Preservation Leaflets

<http://www.solinet.net/presvt n/leaf/leaflets.htm>

Systems of knowledge organization for digital libraries beyond traditional authority files. Hodge, Gail M. Washington, DC :

Key Digitization Centers

Berkeley Digital Library SunSITE

<http://sunsite.berkeley.edu/>

Builds digital collections and services while providing information and support to digital library developers worldwide. Includes Teaching and Training, Collections, Information for digital library developers, Research and Development, Tools.

Colorado Digitization Center, Digital Toolbox

<http://coloradodigital.coalliance.org/toolbox.html>

Designed to guide administrators of digital projects to sites and resources on the following topics: Administrative Resources, Guidelines and Standards, Copyright and Intellectual Property, Funding, Metadata, Software, Planning, etc.

Library of Congress, American Memory Project

<http://memory.loc.gov/ammem/>

Library of Virginia Digital Library Program

<http://www.lva.lib.va.us/dlp/>

An internationally recognized effort to preserve, digitize, and provide access to significant archival and library collections.

Northeast Document Conservation Center

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

The largest nonprofit, regional conservation center in the U.S. Its mission is to improve the preservation programs of libraries, archives, museums, and other historical and cultural organizations; to provide the highest quality services to institutions that cannot afford in-house conservation facilities or that require specialized expertise; and to provide leadership to the preservation field. Includes Handbook for Digital Projects and other publications.

Regional Alliance for Preservation

<http://www.rap-arcc.org/>

Provides comprehensive preservation information to cultural institutions and the public throughout the United States. Includes bibliographies (one on digital imaging).

University of Virginia Library Electronic Text Center

<http://etext.lib.virginia.edu/>

Builds and maintains an internet-accessible collection of SGML texts and images and a user community adept at the creation and use of these materials. The Center combines an on-line archive of tens of thousands of SGML and XML-encoded electronic texts and images with a library service that offers hardware and software suitable for the creation and analysis of text. Offers short courses, publications. Includes standards information.

Electronic Discussion List – Imaging

IMAGELIB

To subscribe send the message “SUB imagelib Your Full Name” to listserv@listserv.arizona.edu”

Preservation of Electronic Files

Storage and Day-to-Day Maintenance of Files Bibliography

Ebind <http://sunsite.berkeley.edu/Ebind/>

Kenney, Anne R., Stephen Chapman, and Cornell University. Library. Dept. of Preservation and Conservation. *Digital Imaging for Libraries and Archives*. Ithaca, N.Y.: Dept. of Preservation and Conservation Cornell University Library, 1996.

Kenney, Anne R., Oya Y. Rieger, and Research Libraries Group. *Moving Theory into Practice : Digital Imaging for Libraries and Archives*. Mountain View, CA: Research Libraries Group, 2000.

Lee, Stuart D. *Digital Imaging : A Practical Handbook*. New York: Neal-Schuman Publishers in association with Library Association Pub., 2001.

List of conversion and presentation software:

<http://coloradodigital.coalition.org/software.html>

Moore, R. , C. Baru, A. Rajasekar, B. Ludaescher, R. Marciano, M. Wan, and W. Schroeder. "Collection-Based Persistent Digital Archives: Part 1." *D-Lib Magazine* 6, no. 3 (2000). Available:
<http://www.dlib.org/dlib/march00/moore/03moore-pt1.html> [May 17,2002].

Moore, R. , C. Baru, A. Rajasekar, B. Ludaescher, R. Marciano, M. Wan, and W. Schroeder. "Collection-Based Persistent Digital Archives: Part 2." *D-Lib Magazine* 6, no. 4 (2000). Available:
<http://www.dlib.org/dlib/march00/moore/03moore-pt2.html> [May 17,2002].

Moving theory into practice : digital imaging tutorial. Cornell University Library/Dept. of Preservation and Conservation. Available:
<http://www.library.cornell.edu/preservation/tutorial/technical/technicalID-01.html> [May 17, 2002].

RLG Worksheet For Estimating Digital Reformatting Costs
<http://www.rlg.org/preserv/RLGWorksheet.pdf>

Long-Term Preservation of Files Bibliography

Andersson, U. "Workshop on Electronic Archiving, an Evaluation of the Sesam Report." *Archives and Museum Informatics* 12, no. 2 (1998): 147-50.

Bagnall, Roger S., and Commission on Preservation and Access. *Digital Imaging of Papyri : A Report to the Commission on Preservation and Access*. Washington, DC: Commission on Preservation and Access, 1995.

Berkeley Digital Library SunSITE. Preservation Resources.
<http://sunsite.berkeley.edu/Preservation/>

The Cedars Project: CURL exemplars in digital archives (UK)
<http://www.leeds.ac.uk/cedars/>

Commission on Preservation and Access (CPA).

<http://palimpsest.stanford.edu/cpa/>

Conservation OnLine (CoOL).

<http://palimpsest.stanford.edu/>

Consultative Committee for Space Data Systems. Reference Model for an Open Archival Information System (OAIS).

<http://www.ccsds.org/documents/pdf/CCSDS-650.0-R-2.pdf>

Conway, P. (1996). *Preservation in the Digital World*. (Washington, D.C.: Commission on Preservation and Access).

Day, Michael. *Preservation of Electronic Information: A Bibliography*. Available:

<http://homes.ukoln.ac.uk/~lismd/preservation.html> [May 17, 2002].

Eakins, J.P. and Graham, M.E., 1999. *Content-based Image Retrieval: A report to the JISC Technology Applications Programme* (Institute for Image Data Research, University of Northumbria at Newcastle).

<http://www.unn.ac.uk/iidr/research/cbir/report.html>

González Garcáia, Pedro, Council on Library and Information Resources. Commission on Preservation and Access, and Council on Library and Information Resources. *Digital Libraries. Computerization of the Archivo General De Indias : Strategies and Results, Economics of Information Leadership*. Washington, D.C.: Council on Library and Information Resources, 1998.

Hendley, T. "The Preservation of Digital Material." In *BLRD Report*, 210: British Library Research and Development Department, 1996.

Hodge, G.M. "Best Practices for Digital Archiving." *D-Lib Magazine* 6, no. 1 (2000). Available: <http://www.dlib.org/dlib/january00/01hodge.html> [May 17, 2002].

Kenney, Anne R., Lynne K. Personius, and Commission on Preservation and Access. *The Cornell/Xerox/Commission on Preservation and Access Joint Study in Digital Preservation : Report, Phase 1, January 1990-December 1991 : Digital Capture, Paper Facsimiles and Network Access*.

Washington, D.C.: Commission on Preservation and Access, 1992.

Northeast Document Conservation Center

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

OCLC/RLG Working Group on Preservation Metadata. *Preservation Metadata for Digital Objects: A Review of the State of the Art*. January 31, 2001.
http://www.oclc.org/digitalpreservation/presmeta_wp.pdf

PANDORA: Preserving and Accessing Networked Documentary Resources of Australia.
<http://pandora.nla.gov.au/>

“Preservation of Public Data.” *Of Significance: A Topical Journal of the Association of Public Data Users* 2:2(2000).

Preservation Resources Web Site (OCLC).
<http://www.oclc.org/oclc/presres/index.htm>

Preserving Access to Digital Information (PADI).
<http://www.nla.gov.au/padi/>

Puglia, S. 1999. “Creating Permanent and Durable Information: Physical Media and Storage Standards,” *CRM: Cultural Resource Management* 22(2), 25-27. Available: <http://tps.cr.nps.gov/crm/archive/22-2/22-02-10.pdf> [May 17, 2002].

Rieger, Oya (2000). “Projects to Programs: Developing a Digital Preservation Policy” in *Moving Theory into Practice*. Research Libraries Group.

RLG Working Group on Preservation Issues of Metadata, 1998.
<http://www.rlg.org/preserv/presmeta.html>

Task Force on Archiving of Digital Information, Commission on Preservation and Access, and Research Libraries Group. *Preserving Digital Information : Report of the Task Force on Archiving of Digital Information*. Washington, D.C.: Commission on Preservation and Access, 1996. Available: <http://www.rlg.org/ArchTF/tfadi.index.htm> [May 17, 2002].

Ross, Seamus, and British Library. National Preservation Office. *Changing Trains at Wigan : Digital Preservation and the Future of Scholarship, Npo Preservation Guidance Occasional Papers*. London: National Preservation Office The British Library, 2000.

UK Public Records Office. Management, Appraisal and Preservation of Electronic Records. Available: <http://www.pro.gov.uk/recordsmanagement/eros/guidelines/default.htm> [May 17, 2002].

Weber, Hartmut, Marianne Dèorr, Deutsche Forschungsgemeinschaft, Commission on Preservation and Access, and European Commission on

Preservation and Access. *Digitalization as a Method of Preservation? : Final Report of a Working Group of the Deutsche Forschungsgemeinschaft (German Research Association)*. Washington, D.C. Amsterdam: Commission on Preservation and Access ; European Commission on Preservation and Access, 1997.

Authenticity of Files Bibliography

Council on Library and Information Resources. *Authenticity in a Digital Environment*. Washington, D.C.: Council on Library and Information Resources, 2000. Available: <http://www.clir.org/pubs/reports/pub92/contents.html> [May 22, 2002].

Bearman, David, and Jennifer Trant. 1998. "Authenticity of Digital Resources: Towards a Statement of Requirements in the Research Process," *D-Lib Magazine* (June) <http://www.dlib.org/dlib/june98/06bearman.html>

Lynch, Clifford A., "Authenticity and Integrity in the Digital Environment: An Exploratory Analysis of the Central Role of Trust," *Authenticity in a Digital Environment* (Washington, DC: Council on Library and Information Resources, 2000), pp 32-50. <http://www.clir.org/pubs/reports/pub92/lynch.html>

FIPS (Federal Information Processing Standards Publication), Digital Signature Standard (DSS), FIPS PUB 186-2, January 2000. Available: <http://csrc.nist.gov/publications/fips/fips186-2/fips186-2.pdf>

NARA GPEA Guidance (Records Management Guidance for Agencies Implementing Electronic Signature Technologies), Washington, DC : National Archives and Records Administration, [2000].

Rothenberg, J. and T. Bikson (August 1999). *Carrying Authentic, Understandable and Usable Digital Records Through Time*. RAND-Europe. Available: http://www.si.umich.edu/CAMILEON/Camileon/publications/final-report_4.pdf [May 17, 2002].

Sarno, Luigi, InterPARES Project, and Istituto italiano di cultura (Vancouver B.C.). *Authentic Records in the Electronic Age : Proceedings from an International Symposium*. Vancouver: InterPARES Project : Istituto Italiano di Cultura Vancouver, 2000. Available: <http://is.gseis.ucla.edu/us-interpares/bibgraph.htm> [May 17, 2002].

Trusted Digital Repository: Attributes and Responsibilities." An RLG/OCLC Report. Available: <http://www.rlg.org/longterm/repositories.pdf>

Electronic Discussion Group – Preservation

PADI Forum

Specifically dedicated to the exchange of news about digital preservation.

To subscribe, send the message “SUBSCRIBE padiforum-I Your Full Name” to listproc@nla.gov.au”

Metadata & Copyright

General Resources Bibliography

Moving theory into practice : digital imaging tutorial. 5. Metadata. Cornell University Library/Dept. of Preservation and Conservation. Available: <http://www.library.cornell.edu/preservation/tutorial/> [May 23, 2002].

Hodge, Gail. *Metadata Made Simpler.* NISO Press, Bethesda, Maryland. (2001). Available: http://www.niso.org/news/Metadata_simpler.pdf. [May 24, 2002].

Gilliland-Swetland, Anne S. *Introduction to Metadata: Setting the Stage.* Available: http://www.getty.edu/research/institute/standards/intrometadata/2_articles/index.html [May 24, 2002].

Vellucci, S. L. "Metadata and Authority Control." *Library Resources & Technical Services* 44, no. 1 (2000): 33-43.

Preservation, Administrative, and Technical Metadata Bibliography

Consultative Committee for Space Data Systems. *Reference Model for an Open Archival Information System (OAIS)* <http://www.ccsds.org/documents/p2/CCSDS-650.0-R-1.pdf>. Available: [May 17, 2002].

Encoded Archival Description (EAD) Document Type Definition (DTD). Library of Congress. Available: <http://lcweb.loc.gov/ead/> [May 22, 2002].

Hansen, Jaytte and Lief Andresen (September 10, 2001). *Administrative Dublin Core (A-Core) Element: A proposal to be discussed in the DCMI Administrative Metadata Working Group.* Available:

<http://dublincore.org/groups/admin/proposal-20010910.shtml> [May 22, 2002].

Information Infrastructures for Digital Preservation: A One Day workshop. Sponsored by *The Cedars Project (JISC funded), Research Libraries Group and OCLC in association with the UK Office for Library Networking*, 6-8 December 2000, York, England. Available: <http://www.rlg.org/events/pres-2000/infopapers.html>. [May 24, 2002].

Library of Congress National Digital Library Program and Conservation Division. *Conservation Implications of Digitization projects*, (1999). Available: <http://memory.loc.gov/ammem/techdocs/conserv83199a.pdf>. [May 24, 2002].

Lupovici, C. and Julien Masanes. *Metadata for long term-preservation*, (July 2000). Available: <http://www.kb.nl/coop/nedlib/results/preservationmetadata.pdf> [May 17, 2002].

This report defines the core minimum metadata that are mandatory for preservation management purposes, in order to handle large amounts of data items in a changing technological environment.

Lynch, Clifford A. "Metadata Harvesting and the Open Archives Initiative." *ARL Bimonthly Report* 217 (2001): 1-9. Available: <http://www.arl.org/newsltr/217/mhp.html> [May 22, 2002].

Moving theory into practice : digital imaging tutorial. 5. Metadata. Cornell University Library/Dept. of Preservation and Conservation. Available: <http://www.library.cornell.edu/preservation/tutorial/metadata/metadata-01.html> [May 24, 2002].

National Library of Australia. *Preservation metadata for Digital Collections. Exposure Draft.* October 15, 1999. Available: <http://www.nla.gov.au/preserve/pmeta.html>. [May 24, 2002].

PADI (Preserving Access to Digital Information). *Preservation Metadata*, (08/25/2001). Available: <http://www.nla.gov.au/padi/topics/32.html> [May 24, 2002].

OCLC/RLG Working Group on Preservation Metadata. *Preservation Metadata for Digital Objects: A Review of the State of the Art (White Paper)*, (January 31, 2001). Available: http://www.oclc.org/digitalpreservation/presmeta_wp.pdf [May 24, 2002].

Research Libraries Group (RLG). *Working Group on Preservation Issues of Metadata, Final Report.* (May 1998). Available: <http://www.rlg.org/preserv/presmeta.html> [May 24, 2002].

Trusted Digital Repository: Attributes and Responsibilities. An RLG/OCLC Report. Research Libraries Group, (2002). Available: <http://www.rlg.org/longterm/repositories.pdf> [May 24, 2002].

Washington State Library. *Best Digital Practices: Preservation Of Digital Projects.* Available: <http://digitalwa.statelib.wa.gov/newsite/collection/preservation.htm> [May 24, 2002].

Weibel, Stuart. "Metadata: The Foundations of Resource Description." *D-Lib Magazine* (July 1995). Available: <http://www.dlib.org/dlib/July95/07contents.html> [May 24, 2002].

Descriptive Metadata Bibliography

ALCTS Taskforce on Meta Access: final report. Available: <http://www.lib.virginia.edu/alcts/about/final.html> [May 24, 2002].

Coombs, Philip. *White Paper on the Use of Numeric Tokens in Resource Descriptions* September 20, 1999. Available: <http://www.statelib.wa.gov/projects/imps/tokens.htm> [May 24, 2002].

Day, Michael. *Metadata: Mapping between Metadata Formats.* Available: <http://www.ukoln.ac.uk/metadata/interoperability/> [May 23, 2002].

Dublin Core Metadata Initiative. *Dublin Core Metadata Element Set*, (1999). Available: <http://dublincore.org/documents/1999/07/02/dces/> [May 24, 2002].

Dublin Core Metadata Initiative. *MIReG Metadata Framework - Element Set* (February 8, 2001). Available: <http://dublincore.org/groups/government/mireg-metadata-20010828.shtml> [May 24, 2002].

Dublin Core Metadata Glossary, (02/12/2001). Available: <http://library.csun.edu/mwoodley/dublincoreglossary.html> [May 24, 2002].

Harmony Project Home Page. Available: <http://www.ilt.bris.ac.uk/discovery/harmony/> [May 24, 2002].

Hillmann, Diane. *Using Dublin Core*, (July 16, 2000). Available: <http://dublincore.org/documents/usageguide/> [May 24, 2002].

- Hunter, Jane. "MetaNet – A Metadata Term Thesaurus to Enable Semantic Interoperability Between Metadata Domains," *Journal of Digital Information*, 1:8, 2001-02-08. Available: <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Hunter/> [May 17, 2002].
- Iannella, Renata. "An Idiot's Guide to the Resource Description Framework." *New Review of Information Networking* 4 (1998): 181-8. Available: <http://archive.dstc.edu.au/RDU/reports/RDF-Idiot/> [May 22, 2002].
- International Federation of Library Associations and Institutions. *Digital Libraries: Metadata Resources*. Available: <http://www.ifla.org/II/metadata.htm> [May 24, 2002].
- Lagoze, C., J. Hunter and D. Brickley (2000) "An Event-Aware Model for Metadata Interoperability". *ECDL 2000*, Lisbon, September.
- Library of Congress. *Library Of Congress Digital Repository Development: Core Metadata Elements*. Available: <http://lcweb.loc.gov/standards/metadata.html> [May 24, 2002].
- Library of Congress, Network Development and MARC Standards Office. *Dublin Core/MARC/GILS Crosswalk*. Available: <http://lcweb.loc.gov/marc/dccross.html> [May 24, 2002].
- Library of Congress, Network Development and MARC Standards Office. *LC Encoded Archival Description (EAD) Official Web Site*. Available: <http://lcweb.loc.gov/ead/> [May 24, 2002].
- Library of Congress, Network Development and MARC Standards Office. *MARC to Dublin Core Crosswalk*. Available: <http://www.loc.gov/marc/marc2dc.html> [May 24, 2002].
- Smith, Neil. "XML – the DNA of the Internet." TIPS AND TECHNIQUES column in *FreePint*, 15th March 2001 No.83. Available: <http://www.freepint.com/issues/150301.htm#tips> [May 24, 2002].
- TEI Consortium. *Text Encoding Initiative*, (2001). Available: <http://www.tei-c.org/P4X/> [May 24, 2002].
- Using Dublin Core (Readings & References): Revised Version 0.22* (March 7, 2001). Available: http://www.globility.com/~esharf/dc/dcguide_refs.html [May 24, 2002].
- Weibel, Stuart. "Metadata: The Foundations of Resource Description." *D-Lib Magazine* (July 1995). Available: <http://www.dlib.org/dlib/July95/07contents.html> [May 24, 2002].

W3C. *Resource Description Framework (RDF) Model and Syntax Specification*, REC-rdf-syntax-19990222, *W3C Recommendation*, 2 February 1999. Available: <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222> [May 24, 2002].

XML Cover Pages, By Oasis (Organization for the Advancement of Structured Information Standards). <http://www.oasis-open.org/cover/>

Related Issues and Areas of Investigation

American Library Association. Association for Library Collections and Technical Services Cataloging and Classification Section. Committee on Cataloging: Description and Access Task Force on Metadata. *Meeting Minutes*, Sunday, January 16, 2000. Available: <http://www.ala.org/alcts/organization/ccs/ccda/ff-meta5.html> [May 24, 2002].

Bridges, State of Minnesota. *Minnesota Metadata Guidelines - Dublin Core [MMG - DC]*. Available: <http://bridges.state.mn.us/metadata.html> [May 24, 2002].

Hansell, Saul. "Paid Placement Is Catching On in Web Searches." *New York Times*, June 4, 2001. Available: <http://www.nytimes.com/2001/06/04/technology/04GOTO.html> [May 23, 2002].

Knight, Jon. *MIME Implementation for the Warwick Framework* (1999). Available: <http://weeble.lut.ac.uk/MIME-WF.html> [May 24, 2002].

Lynch, Clifford A. "Metadata Harvesting and the Open Archives Initiative." *ARL Bimonthly Report* 217 (2001): 1-9. Available: <http://www.arl.org/newsltr/217/mhp.html> [May 22, 2002].

Mullen, Allen. "GILS metadata initiatives at the state level." *Government Information Quarterly* 18 (2001) 167-180.

Search Engine Watch. Available: <http://www.searchenginewatch.com/> [May 24, 2002].

TRAIL (Texas Records and Information locator). Available: <http://www2.tsl.state.tx.us/servlet/TXDublinCore> [May 24, 2002].

Warnick, W.L., R.L. Scott, L.A. Johnson, A. Lederman, K.J. Spence, and V.S. Allen. "Searching the Deep Web. Directed Query Engine Applications at

the Department of Energy." *D-Lib Magazine* 7, no. 1 (2001). Available: <http://www.dlib.org/dlib/january01/warnick/01warnick.html> [May 24, 2002].

Westera, Gillian. *Comparison of Search Engine User Interface Capabilities*, (May 2002). Available: <http://lisweb.curtin.edu.au/staff/gwpersonal/compare.html> [May 24, 2002].

Changes to a Document in the Preservation Process

Consultative Committee for Space Data Systems. *Reference Model for an Open Archival Information System (OAIS)*. Available: <http://www.ccsds.org/documents/p2/CCSDS-650.0-R-1.pdf> [May 17, 2002].

National Library of Australia. *Preservation metadata for Digital Collections. Exposure Draft*. October 15, 1999. Available: <http://www.nla.gov.au/preserve/pmeta.html#12> and <http://www.nla.gov.au/preserve/pmeta.html#23> [May 24, 2002].

Copyright Bibliography

Balas, J. "Copyright in the Digital Era." *Computers in Libraries* 8, no. 6 (1998): 38-40.

Berghel, H., and L. Ogorman. "Protecting Ownership Rights through Digital Watermarking." *Computer* 29, no. 7 (1996): 101-03. Available: http://www1.acm.org:82/~h1b/publications/dig_wtr/dig_watr.html [May 22, 2002].

Library of Congress. Copyright Office. *Copyright Law of the United States of America, and Related Laws Contained in Title 17 of the United States Code, Circular 92, April 2000*. S.l.: s.n., 2000. Available: <http://www.loc.gov/copyright/title17/> [May 22, 2002].

DTIC Guidelines for Determining Copy Rights. Available: <http://www.dtic.mil/dtic/submitting/copyright.html> [May 24, 2002].
"Because R&D managers and STI information center staff are often uncertain about what to do with documents that come to them with copyright notices, these guidelines have been written to help managers and staff in determining Government rights in copyrighted works."

Hatmaker, Scott L. "Government Information: Public Domain or User Beware." *Information-Outlook* 1, no. 12 (1997).

Historical and revision notes, Title 17, Sec.105, U. S. Code.

Hoon, Peggy. *Guidelines for Educational Use of Copyrighted Materials : Designed for Educators and Librarians in the Higher Education Setting*. Pullman, Wash.: Washington State University Press, 1997.

International Federation of Library Associations. *Information Policy: Copyright And Intellectual Property*. Available: <http://www.ifla.org/II/copyright.htm> [May 24, 2002].

Jensen, Mary Brandt. *Does Your Project Have a Copyright Problem? : A Decision-Making Guide for Librarians*. Jefferson, N.C.: McFarland & Co., 1996.

Library of Congress. *Building digital collections: technical information and background papers*. Available: <http://memory.loc.gov/ammem/ftpfiles.html> [May 24, 2002].
Site with numerous links.

Library of Congress, Copyright Office. *Copyright Basics*. Washington, D.C., Library of Congress Copyright Office, (2000). Available: <http://www.loc.gov/copyright/circs/circ1.html> [May 22, 2002].

Lide, Casey. "What Colleges and Universities Need to Know about the Digital Millennium Copyright Act," *CAUSE/EFFECT Journal*, Volume 22 Number 1, 1999. Available: <http://www.educause.edu/ir/library/html/cem9913.html> [May 24, 2002].

Okerson, Ann. *Yale Copyright Resources Online*, (2000). Available: <http://www.library.yale.edu/%7Eokerson/copyproj.html> [May 24, 2002].

Oppenheim, C. *Copyright in the electronic age*. In: *UNESCO World Information Report*. UNESCO, Paris, 1997. Available: http://www.unesco.org/webworld/com_inf_reports/wirenglish/chap26.pdf [May 24, 2002].

U.S. Copyright Office. *The Digital Millennium Copyright Act of 1998. U.S. Copyright Office Summary*. Available: <http://lcweb.loc.gov/copyright/legislation/dmca.pdf> [May 24, 2002].

Appendix E: Digitization Projects Reported

Overview

The Clearinghouse of Government Documents Digital Projects, located at <http://www.gl.iit.edu/services/ref/diggovclearinghouse.htm>, was created as a pilot project for the purposes of this study. It is intended to provide information to librarians and others about digitization projects for local, state, federal, and international government documents, which are currently planned, in progress, or already completed. The Ad Hoc Committee on Digitization of Government Information created a proposed list of data fields of information to be collected about each project and began collecting this information in a database. The database proposes to facilitate cooperation and partnerships between libraries for digitization projects.

Criteria for Inclusion

The types of digital projects that should be included in the database are those that meet ALL the following requirements:

- a. Projects undertaken by libraries (public libraries, libraries affiliated with educational institutions, or government agency libraries), government agencies, or tribal governments,
- b. Projects with a focus on federal, state, local, or international government publications or that use government produced data, and
- c. Projects that freely offer the resulting resources to the public.

The Ad Hoc Committee recommends that GITCO review these criteria periodically to determine if they should be expanded or narrowed. Additionally, colleagues must be encouraged to enter data about their digital projects in the data submission form.

Sources

In this test phase, data was added to the database from two sources. The first source was the online submission form at <http://www.gl.iit.edu/services/ref/add.htm>. The second source was the SuDocs Database listing of electronic federal documents sites hosted by the Paul V. Galvin Library at the Illinois Institute of Technology. This listing is available at <http://www.gl.iit.edu/govdocs/internet/sudocindex.htm>.

Following is a current listing of included projects as of May 24th, 2002.

Report to ALA GODORT – Digitization of Government Information

Project Title	Institution
1992 Census of Construction Industries - Geographic Area Series	Bureau of the Census, Department of Commerce
1992 Census of Construction Industries - Industry Series	Bureau of the Census, Department of Commerce
1992 Census of Retail Trade, Final Reports, Area Series	Bureau of the Census, Department of Commerce
Advisory Commission on Intergovernmental Relations	University of North Texas
Afghanistan Country Study	Illinois Institute of Technology
Agricultural Research Magazine	Agricultural Research Service, Department of Agriculture
American Housing Brief	Bureau of the Census, Department of Commerce
American Indians of the Pacific Northwest Digital Collection	University of Washington Libraries
Animal Welfare Annual Reports	Animal Care, Department of Agriculture
Annotation	National Historic Publications and Records Commission
Annual Capital Expenditures Survey	Bureau of the Census, Department of Commerce
Annual Survey of Communications Services	Bureau of the Census, Department of Commerce
Archives, Monthly Labor Review Online	Bureau of Labor Statistics, Department of Labor
Arms Control and Disarmament Agency Archive	University of Illinois at Chicago
Army Corps of Engineers Technical Publications	Headquarters, U. S. Army Corps of Engineers, Department of the Army
Census and You	Bureau of the Census, Department of Commerce
Census Briefs	Bureau of the Census, Department of Commerce
Census Catalog and Guide	Bureau of the Census, Department of Commerce
Census of Agriculture	National Agricultural Statistics Service, Department of Agriculture
Consolidated Federal Funds Report - County Areas	Bureau of the Census, Department of Commerce
County Business Patterns	Bureau of the Census, Department of Commerce
Current Construction Reports	Bureau of the Census, Department of Commerce
Current Industrial Reports	Bureau of the Census, Department of Commerce
Current Population Reports	Bureau of the Census, Department of Commerce

Report to ALA GODORT – Digitization of Government Information

Dairy Market Publications	Agricultural Marketing Service, Department of Agriculture
Digital Collections Online	University of Connecticut
Energy Information Administration Library/Archives	Energy Information Administration, Department of Energy
ERS/USDA Publications	Economic Research Service, Department of Agriculture
Everglades Digital Collection	Florida International University
Export Assistance, Food Aid, and Market Development Programs	Foreign Agricultural Service, Department of Agriculture
Fact Sheets for Arboretum Award Winners	United States National Arboretum, Department of Agriculture
Families and Living Arrangements	Bureau of the Census, Department of Commerce
Federal Expenditures by State	Bureau of the Census, Department of Commerce
Federal, State and Local Governments Summary of State and Local Government Tax Revenue	Bureau of the Census, Department of Commerce
Finances of Selected State and Local Government Employee Retirement Systems	Bureau of the Census, Department of Commerce
Fire Management Today	Forest Service, Department of Agriculture
Food Standards and Labeling Policy Book	Food Safety and Inspection Service, Department of Agriculture
Foreign Agricultural Service Publications Archive	Foreign Agricultural Service, Department of Agriculture
Foreign Relations of the United States 1900-1918	University of Wisconsin, Madison
Forest Health Protection Intermountain Region Insect and Disease Conditions	Forest Service, Department of Agriculture
Forest Health Protection Publications	Forest Service, Department of Agriculture
FPL Publications	Forest Products Laboratory, Department of Agriculture
Indian Affairs: Laws and Treaties	Oklahoma State University
International Briefs	Bureau of the Census, Department of Commerce
Internet Moving Images Archive: Movie Collection	Internet Archive
John Wesley Powell Reports	University of North Texas
List of Materials Acceptable for Use on Systems of RUS Electric Borrowers	Rural Utilities Service, Department of Agriculture
List of Materials Acceptable for Use on Telecommunications Systems of RUS Borrowers	Rural Utilities Service, Department of Agriculture
Manufacturing Profiles	Bureau of the Census, Department of Commerce

Report to ALA GODORT – Digitization of Government Information

Military Review Professional Journal (English, Spanish and Portuguese)	U. S. Army Command and General Staff College, Department of the Army
Minerals Yearbook	University of Wisconsin, Madison
Monthly Product Announcement	Bureau of the Census, Department of Commerce
Monthly Retail Trade and Food Services	Bureau of the Census, Department of Commerce
Monthly Wholesale Trade Survey	Bureau of the Census, Department of Commerce
National Agriculture Library Publications and Databases	National Agriculture Library, Department of Agriculture
National Compensation Survey	Bureau of Labor Statistics, Department of Labor
Nineteenth Century Texas Law Online: Gammel's Laws of Texas	University of North Texas
NRI Research Highlights	National Research Initiative, Department of Agriculture
Occupational Outlook Quarterly	Bureau of Labor Statistics, Department of Labor
Official U. S. Standards for Grain	Grain Inspectors, Packers and Stockyards Administration
OIG Semi-Annual Reports to Congress	Bureau of Labor Statistics, Department of Labor
Older Reports from the State Government Scanned in by the State Library	New Jersey State Library
Pacific Northwest Research Station Quarterly Lists of Recent Publications	Forest Service, Department of Agriculture
Pesticide Data Program Databases and Reports	Agricultural Marketing Service, Department of Agriculture
Preliminary Report on Manufacturers' Shipments, Inventories and Orders	Bureau of the Census, Department of Commerce
Prime Contract Awards by Region and State	Directorate for Information Operations and Reports, Department of Defense
Prime Contract Awards by Service Category and Federal Supply Classification	Directorate for Information Operations and Reports, Department of Defense
Procurement Statistics	Directorate for Information Operations and Reports, Department of Defense
Public Elementary-Secondary Education Finance Data	Bureau of the Census, Department of Commerce
Quality Standards for Fruits and Vegetables	Agricultural Marketing Service, Department of Agriculture
Quarterly Financial Report for Manufacturing, Mining and Trade Corporations	Bureau of the Census, Department of Commerce
Rural Cooperatives Magazine	Rural Business-Cooperative Service, Department of Agriculture
Service Annual Survey	Bureau of the Census, Department of Commerce

Report to ALA GODORT – Digitization of Government Information

State Basin Outlook Reports	National Water & Climate Center, Department of Agriculture
State Government Tax Collections	Bureau of the Census, Department of Commerce
Statistical Abstract of the United States	Bureau of the Census, Department of Commerce
Statistical Information Analysis Division Work Force Publications	Statistical Information Analysis Division, Department of Defense
Supplement to the American Housing Survey for selected metropolitan areas	Bureau of the Census, Department of Commerce
The Hispanic Population in the United States	Bureau of the Census, Department of Commerce
The Record	National Archives and Records Administration
Trade Leads	Foreign Agricultural Service, Department of Agriculture
Transportation Annual Survey	Bureau of the Census, Department of Commerce
U. S. Export Sales Reports	Foreign Agricultural Service, Department of Agriculture
U. S. International Trade in Goods and Services	Bureau of the Census, Department of Commerce
U. S. Trade with Puerto Rico and U. S. Possession	Bureau of the Census, Department of Commerce
U.S. Geological Survey Publications	U.S. Geological Survey, Department of the Interior
USDA Economics and Statistics System	Cornell University
USDA National Agricultural Statistics Service Publications	National Agricultural Statistics Service, Department of Agriculture
USDA/Publications	Office of Communications, Department of Agriculture
World War II Poster Collection	Northwestern University Library
Wright Air Development Center Digital Collection	Illinois Institute of Technology