IMLS   Institute of Museum and Library Services

Contact
Search
Subscribe
Site Map

Adobe Acrobat
spacer
Home

A Framework of Guidance for Building Good Digital Collections

Printer Friendly Version

November 6, 2001

The following report was prepared by members of the Digital Library Forum, a group convened by the Institute of Museum and Library Services (IMLS) to discuss issues relating to the implementation and management of networked digital libraries. Members quickly identified collection quality as a critical aspect of digital library management and interoperability, and this understanding forms the foundation of the report.

The Framework is intended as a resource for grant applicants, IMLS and other federal funding agencies, rather than as a set of requirements.

Forum members who contributed to the report include: Liz Bishoff, Colorado Digitization Alliance; Priscilla Caplan (chair), Florida Center for Library Automation; Tim Cole, University of Illinois Urbana-Champaign; Anne Craig, Illinois State Library; Daniel Greenstein, Digital Library Federation; Doug Holland, Missouri Botanical Garden; Ellen Kabat-Lensch, Eastern Iowa Community College; Tom Moritz, American Museum of Natural History; and John Saylor, Cornell University.

INTRODUCTION

This Framework is intended for two audiences: first, for people who are working in the context of projects and want to develop good digital collections; and second, for funding organizations and agencies that want to encourage the creation of good digital collections.

The use of the word *good* in this context requires some explanation. In the early days of digitization for the Web, projects could be justified as vehicles for the development of methods and technologies, as experiments in technical or organizational innovation, or simply as learning experiences. A collection could be good if it provided proof of concept, even if it disappeared at the end of the project period. As the environment matured, the focus of collection building shifted towards the more utilitarian goal of making relevant content available digitally to some community of users. The bar of goodness was accordingly raised

to include levels of usability, accessibility and fitness for use appropriate to the anticipated user group. We have now entered a third stage, where even serving information effectively to a known constituency is not sufficient. In today's digital environment, the context of content is a vast international network of digital materials and services. Objects, metadata and collections should be viewed not only within the context of the projects that created them but as building blocks that others can reuse, repackage, and build services upon. Indicators of goodness correspondingly must now also emphasize factors contributing to interoperability, reusability, persistence, verification and documentation. At the same time attention must be focused on mechanisms for respecting copyright and intellectual property law.

This document is not a guideline itself but rather a framework for identifying, organizing, and applying existing knowledge and resources that can be used as an aid in the development of local guidelines and procedures. It is built around indicators of goodness for four types of entities:

- Collections,
- Objects,
- Metadata, and
- Projects.

Note that services have been deliberately excluded as out of scope, but it is expected that if quality collections, objects and metadata are created, it will be possible for any number of higher level services to make use of these entities.

In each category, general principles relating to quality are defined and discussed, and supporting resources are identified. These resources may be standards, guidelines, best practices, explanations, discussions, clearinghouses, case studies or examples. Every effort has been made to be selective and to include only materials that are useful, current and widely accepted as authoritative. However, the value of some resources will in time be depreciated and other resources created or discovered, so it is fully expected this list will change over time. It is hoped that this framework will be flexible enough to accommodate new principles, considerations and resources, and to absorb the contributions of others.

There are no absolute rules for creating good collections, objects or metadata. Every project is unique and each has its own goals. There are almost as many ways of categorizing collections as there are collections. Projects dealing with legacy collections or with born-digital materials, for example, have different constraints than projects just embarking on new digitization. Museums, libraries, and school boards have different constituencies, priorities, institutional cultures, funding mechanisms and governance structures. The key to a successful project is not to

follow any particular path, but to think strategically and make wise choices. To use the Framework successfully, project planners should take into consideration their organizational goals, their audience, and the content available to them, and they should select the set of principles and resources that best meet their project's needs. Following sound guidelines will help guarantee that collections will not only serve known local needs but will be reusable in new and innovative contexts.

A number of excellent resources take a holistic view of digitization projects. It is recommended that projects consult these or other general guides to digitization projects.

- Northeast Document Conservation Center. *Handbook for Digital Projects: A Management Tool for Preservation & Access*. http://www.nedcc.org/digital/dighome.htm
- Anne R. Kenney and Oya Y. Rieger. *Moving Theory into Practice: Digital Imaging for Libraries and Archives*. Research Libraries Group, 2000. An online tutorial at http://www.library.cornell.edu/preservation/tutorial/ serves as an introduction to topics covered more extensively in the printed volume.

COLLECTIONS

A digital collection is more than just an assemblage of objects. In the context of this Framework, a collection can be defined as a selected and organized set of digital materials (objects) along with the metadata that describes them and at least one interface that gives access to them. As such, the whole is greater than the sum of the parts. Digital collections are generally created by organizations or groups of cooperating organizations, often as part of a project.

Principles applying to good collections

Collections principle 1: A good digital collection is created according to an explicit collection development policy that has been agreed upon and documented before digitization begins.

Of all factors, collection development is most closely tied to an organization's own goals and constituencies. Collection builders should be able to summarize the mission of their organization and articulate how a proposed collection furthers or supports that mission. Project managers should be able to identify the target audience(s) for the collection (both in the short term and in the future) and how the selected materials relate to their audience. There is an often unexamined assumption that digitization will dramatically increase the use or value of materials. If the materials exist in non-digital form, how heavily are they used? What factors specifically will influence their use or value when digitized?

Consider how the digital collection will fit in with the organization's overall collection policy, as digital collections should not stand in isolation from the original materials or from the collection as a whole.

The following documents are guidelines for selecting materials for digitization. The list does not include electronic collection development policies, which are documents drafted to guide libraries in their selection of commercially available resources.

- Joint RLG and NPO Preservation Conference, Guidelines for Digital Imaging: Guidance for selecting materials for digitisation. http://www.rlg.org/preserv/joint/ayris.html
- Moving Theory into Practice. Digital Imaging Tutorial: Selection. http://www.library.cornell.edu/preservation/tutorial/selection/selection-01.html
- Dan Hazen, Jeffrey Horrell, and Jan Merrill-Oldham. Selecting research collections for digitization (CLIR, August 1998) http://www.clir.org/pubs/abstract/pub74.html
- Towards a Learning Nation: The Digital Contribution. Recommendations proposed by the Federal Task Force on Digitization. Final Report. (December 31, 1997). Part B Issue 2: Selecting materials for digitization. http://www.nlc-bnc.ca/8/3/r3-407-e.html

A report of the DLESE Collections Committee, "How to Identify the "Best" Resources for the Reviewed Collection of the Digital Library for Earth System Education" describes a distributed selection process that could be applied to other learning resources. http://www.ldeo.columbia.edu/DLESE/collections/CGms.html

The Digital Library Federation maintains a database of digital library documents that include collection development policies of a number of DLF members. Some of these policies concern all electronic acquisitions while others focus on retrospective digitization.
http://www.hti.umich.edu/cgi/b/bib/bib-idx?c=dlf

Some examples of local collection development policies include:

- Columbia University Libraries Selection Criteria for Digital Imaging.
http://www.columbia.edu/cu/libraries/digital/criteria.html
- University of California Selection Criteria for Digitization.
http://www.library.ucsb.edu/ucpag/digselec.html
- North Carolina ECHO (Exploring Cultural Heritage Online) Guidelines, Section 2: Planning & Selection. http://www.ncecho.org/Guide/selection.htm The ECHO project includes materials from libraries, archives and museums.

There are also a number of guidelines for selecting materials for digitization specifically for preservation purposes:

- Selection Criteria for Preservation Digital Reformatting. Library of Congress Preservation Reformatting Division. http://lcweb.loc.gov/preserv/prd/presdig/ presselection.html
- Joint RLG and NPO Preservation Conference, Guidelines for Digital Imaging: Selection Guidelines for Preservation. http://www.rlg.org/preserv/joint/gertz.html

Collection builders should be aware that special constraints may exist in relation to politically and culturally sensitive materials. Even items that are unexceptional in the context of a repository can be disturbing when taken out of context. Selection guidelines with particular attention to sensitivity are included in the Northeast Documentation and Conservation Center's Handbook for Digital Projects, chapter IV: Selection of Materials for Scanning by Diane Vogt-O'Connor. http://www.nedcc.org/digital/ IV.htm.

Collections principle 2: Collections should be described so that a user can discover important characteristics of the collection, including scope, format, restrictions on access, ownership, and any information significant for determining the collection's authenticity, integrity and interpretation.

Collection description is a form of metadata (see also METADATA). Collection description serves two purposes: it helps people discover the existence of a collection (whether they are end-users seeking materials relevant to their information needs, or other collection-builders looking for similar or complementary materials), and it helps users of the collection understand what they are looking at.

To serve the first purpose, when possible, collections should be described in collection-level cataloging records contributed to a national union catalog such as the OCLC or RLIN databases. Websites and individual digital objects can be cataloged through OCLC's CORC. There are also a number of directories where collections can be registered. A few of these are listed below; for a more complete list see the inventory of directories of Web-accessible collections in the December 2000 issue of RLG DigiNews. http://www.rlg.org/preserv/diginews/diginews4-6. html#faq

- The UNESCO/IFLA Directory of Digitized Collections lists major cultural heritage collections and programs worldwide. http://thoth.bl.uk
- The University of Arizona maintains a Clearinghouse of

Image Databases. [http://www.library.arizona.edu/images /clearinghouse/clearinghouse.html](http://www.library.arizona.edu/images/clearinghouse/clearinghouse.html)

- The Association of Research Libraries maintains a database of digital initiatives that includes technical as well as collections information. [http://www.arl.org/did/index. html](http://www.arl.org/did/index.html)
- The Smithsonian Institution maintains a list of Library and Archival Exhibitions on the Web. [http://www.sil.si.edu/ SILPublications/Online-Exhibitions/online-exhibitions- frames.htm](http://www.sil.si.edu/SILPublications/Online-Exhibitions/online-exhibitions-frames.htm)

After a user has discovered a relevant collection, collection description should help him understand the nature and scope of the collection and any restrictions that apply to the use of materials within it. Incorporating a narrative description of the collection on its Web site in human readable prose is good practice. There should be a description of the materials comprising the collection, including how and why they were selected. The organization(s) responsible for building and maintaining the collection should be clearly identified, as organizational provenance is important in helping the user to evaluate the authenticity and authority of the collection. Terms and conditions of use, restrictions on access, special software required for general use, the copyright status(es) of collection materials, and contact points for questions and comments should be noted. Many project planners find a description of the methodologies, software applications, record formats, and metadata schemes used in building other collections helpful.

Good examples of collection-level terms and conditions of use are provided by JSTOR and Ad*Access Project. For examples of Web sites with extensive technical and project documentation, see Ad*Access and Historic Pittsburgh.

There do not appear to be many guidelines specifically for describing digital collections generally, as opposed to archival collections. The Collection Description project of the UK's Research Support Libraries Programme has a Web site of materials related to collection description including an RDF-based collection description schema intended to be both human and machine-readable. The set of data elements included in this schema can be used as a checklist of information a project might want to provide about its collection.

Archival collections are generally described by curators according to established principles of archival description. (See also METADATA.)

- The General International Standard Archival Description. (ISAD(G)) is a set of general rules for archival description

developed by the International Council on Archives.

- The Encoded Archival Description (EAD) is a scheme for representing archival finding aids in machine-understandable form using SGML as a markup language. http://www.loc.gov/ead/
- A standard developed for describing government agencies, collections and services. Global Information Locator Service (GILS). http://www.gils.net/

Collections principle 3: A collection should be sustainable over time. In particular, digital collections built with special funding should have a plan for their continued usability beyond the funded period.

Sustainability at the collection level is related to, but not identical with, persistence at the object level (see OBJECTS). Certainly the collection-level archiving strategy should be tied to the preservation strategy at the object level. Managers of collections containing materials of long-term importance should take steps to ensure not only that the objects within them will be preserved in usable form over time, but that collection-level access to the materials is maintained.

This implies, first and foremost, that some organizational responsibility for the ongoing maintenance of the collection is established. Collection maintenance may take different sets of skills and different commitments of resources than the original collection building. Aspects of ongoing maintenance include such functions as maintaining the currency of locations, ensuring that search systems and other access applications remain usable, logging and accumulating statistics, and providing some level of end-user support. They also include the system administration functions of upgrading server hardware and operating system software as required over time, maintaining server security, and ensuring that restoration of applications and data from backups is always possible.

Two works that focus on creating portals to third-party resources (rather than creating new digital content) focus on sustainability are:

- The *DESIRE Information Gateways Handbook*, which contains generally useful information on link checking and related maintenance activities in a section on collection management. http://www.desire.org/handbook/
- Pitschmann, Louis A. *Building Sustainable Collections of Free Third-Party Web Resources*. (Washington, DC: Digital Library Federation, Council on Library and Information Resources, June 2001) http://www.clir.org/pubs/abstract/pub98abst.html

Collections principle 4: A good collection is broadly available and avoids unnecessary impediments to use. Collections should be accessible to persons with disabilities, and usable effectively in conjunction with adaptive technologies.

At this time, the World Wide Web is the vehicle for broad availability. Collections should be accessible through the Web and should use technologies that are ubiquitous among the target user community. There is always a tradeoff between functionality and general usability; the timing of the adoption of new features such as frames and style sheets should be considered in light of how many potential users will be capable of using the technology and how many will find it a barrier. Bandwidth requirements are also a consideration, as some file formats or interfaces may not be usable by individuals on low-bandwidth connections. The minimum browser version and bandwidth requirements for use should be documented as part of the collection description.

The *webreview* site offers reference guides to style sheets and Web browsers. Their browser compatibility chart compares features supported by all versions of the major browsers. http://www.webreview.com/browsers/browsers.shtml

The report *Performance Measures for Federal Agency Websites* by Chuck McClure et. al. addresses Web site design in terms of efficiency, effectiveness, service quality, impact, usefulness and extensiveness. http://fedbbs.access.gpo.gov/library/download/MEASURES/measures.doc

Accessibility is not only good policy, it is also the law as embodied in the Americans with Disabilities Act of 1990. The International Center for Disability Resources on the Internet publishes *An Overview of Law & Policy for IT Accessibility*. http://www.icdri.org/SL508overview.html

The current de facto accessibility standard is the World Wide Web Consortium (W3C) *Web Content Accessibility Guidelines 1.0.* http://www.w3.org/TR/WAI-WEBCONTENT/

An example of how these guidelines can be applied in an institutional context is given by the Yale University Library. Their document *Services for Persons with Disabilities* has a section on Web Accessibility Guidelines which also lists other accessibility resources http://www.library.yale.edu/Administration/SQIC/spd2.html#s3..

The Bobby application will check a web page or web site for barriers to persons with disabilities. Bobby is a free service offered by CAST, the Center for Applied Special Technology. http://www.cast.org/bobby/

There are several clearinghouses that focus on Web accessibility:

- CPB/WGBH National Center for Accessible Media has a number of accessibility initiatives including projects focused on educational materials. http://ncam.wgbh.org/education/index.html
- University of Wisconsin. Trace Research and Development Center. Designing More Usable Websites. A clearinghouse of useful tools, initiatives, documentation and websites. http://trace.wisc.edu/world/web/

Collections principle 5: A good collection respects intellectual property rights. Collection managers should maintain a consistent record of rightsholders and permissions granted for all applicable materials.

Intellectual property law must be considered from several points of view in relation to any collection: what rights the owners of the original source materials retain in their materials; what rights or permissions the collection developers have to digitize content and make it available; what rights collection owners have in their digital content; and what rights or permissions the users of the digital collection have to make subsequent use of the materials. Viewed from any side, rights issues are rarely clear cut, and the rights policy related to any collection is more often a matter of risk management than one of absolute right and wrong.

There are a number of clearinghouses on law and policy related to copyright and intellectual property. The International Federation of Library Associations maintains a site with international scope at http://www.ifla.org/II/cpyright.htm. The Library of Congress Copyright Office maintains a site that combines both general and procedural information. http://www.loc.gov/copyright/

An excellent introduction to virtually all copyright-related issues is the Copyright Crash Course by Georgia Harper at the University of Texas at Austin (http://www.utsystem.edu/ogc/intellectualproperty/cprtindx.htm). There is a particularly useful section on the logistics of obtaining permission http://www.utsystem.edu/ogc/intellectualproperty/permissn.htm which takes the perspective of risk vs. benefit.

The National Initiative for a Networked Cultural Heritage (NINCH) has held a series of "Town Meetings" that combine experts' presentations with open discussion on topics such as copyright, fair use, and distance education. Reports of past meetings and the schedule of future meetings are available at http://www-ninch.cni.org/copyright/townmeetings01/2001.html

A multimedia publishing company has published primers for multimedia developers. "Intellectual Property Law Primer for MultiMedia Developers" http://www.timestream.com/stuff/neatstuff/mmlaw.html "Licensing Still Images: Some Basic Information for Multimedia Developers." http://www.timestream.com/stuff/neatstuff/license.html

Collections principle 6: A good collection provides some measurement of use. Counts should be aggregated by period and maintained over time so that comparison can be made.

Measures can include use counts ("x files retrieved"), user analysis ("this site was visited by x users from y different domains"), or "linked-to" counts ("this site is linked to by n other sites"). Since measures should be maintained over time, they take some resources to support, and the measures chosen should be designed to serve some purpose of the sponsoring project or organization. One common use is to attempt to justify resources devoted to a collection by volume of use, either generally or within a certain user population. Another use is to enlighten collection development policy. Metrics are also a tool in the evaluation of projects and collections (see PROJECTS).

There are no formal standards for measuring use of electronic content, whether remotely available commercial resources or locally provided collections. The most widely used guidelines were developed by the International Coalition of Library Consortia (ICOLC) as a guide to what measures should be reported by vendors. *Guidelines for statistical measures of usage of web-based indexed, abstracted and full text resources* (November 1998). http://www.library.yale.edu/consortia/webstats.html

The Association of Research Libraries has an initiative to develop measures for electronic resources (e-metrics) that includes both commercial resources and local digital collections. http://www.arl.org/stats/newmeas/emetrics/index.html

The National Information Standards Organization has an initiative to revise Z39.7, a standard for library statistics, to include better measures for electronic resources. Watch the NISO Web site (http://www.niso.org) for progress information about this effort. The *Report on the NISO Forum on Performance Measures and Statistics for Libraries* contains a useful "webography". http://www.niso.org/stats-rpt.html

Collections principle 7: A good collection fits into the larger context of significant related national and international digital library initiatives. For example, collections of content useful for education in science, math and/or engineering should be usable in

the NSDL.

One primary means of fitting into a larger context is paying attention to interoperability issues, particularly the ability to contribute metadata to more inclusive search engines. However, other means are also important. These include being aware of and in contact with related efforts, following widely accepted benchmarks for quality of content and of metadata, and providing adequate collection description for users to place one collection in the context of others.

Some examples of widely known national and international initiatives include:

- The Open Archives Initiative. http://www.openarchives.org/
- The National SMETE (Science Mathematics Engineering and Technology Education) Digital Library. Pathways to progress: vision and plans for developing the NSDL. http://www.smete.org/nsdl/workgroups/coordcomm/Whitepaper.doc
- The Research Libraries Group's Cultural Materials Initiative. http://www.rlg.org/culturalres/
- The Art Museum Image Consortium. http://www.amico.org/

Topical collections may fit into broader clearinghouses or cooperative portals. Project planners should search for clearinghouses in their subject area; there is an increasing number of clearinghouses, particularly in areas related to scientific or environmental information. For example:

- The Geospatial Data Clearinghouse is a collection of over 250 spatial data servers, that have digital geographic data primarily for use in Geographic Information Systems (GIS), image processing systems, and other modeling software. http://fgdclearhs.er.usgs.gov/
- The Global Biodiversity Information Facility aims for "compilation, linking, standardisation, digitisation and global dissemination of the world's biodiversity data". http://www.gbif.org/index.html

Cooperative portals are gateways to existing Web sites and other resources maintained collaboratively by a group of institutions, each taking responsibility for selecting quality resources within some subtopic of a larger subject area. Some examples include:

- Healthweb, a cooperative project of about 20 health sciences libraries for health-related resources. http://healthweb.org/

- Agnic, a portal to agricultural information being developed by the National Library of Agriculture, land grant universities and other partners. http://www.agnic.org/

OBJECTS

This Framework is concerned with two kinds of digital objects: those produced as surrogates for information objects that exist in some analog format (e.g. as books, manuscripts, museum artifacts, audio or video tapes, etc.), and those that are born digital, that is, that are produced originally in machine-readable form (scientific databases, sensory data, digital photographs, etc.). A good object that is created as a surrogate will be considered by a community to be a faithful facsimile of the artifact.

For the context of this Framework, collections (see COLLECTIONS) consist of objects. In this sense, objects are equivalent conceptually to the items that may be found amongst library holdings (books), museum collections (artifacts), and archival fonds (papers). Obviously no hard and fast line can be drawn between objects and collections. Our definition of object extends to compound objects such as the digitally reformatted book or serial publication, but not as far as a collection (which in this case would include, for example, two or more digitally reformatted book or serial publications).

When speaking of digital objects, it is often useful to distinguish between master or preservation copies and access or use copies. As their names imply, masters are typically the highest quality versions that the production technique allows while use or access copies are derivatives that are created for specific uses, distribution scenarios, or users. Thus, a master copy a of a digitally reformatted 35mm slide might be an uncompressed, 18 megabyte, TIFF file, captured in 24-bit color, at a resolution of 600 dots per inch (dpi). The access or derivative copy of this might be a 150 KB, JPEG image derived from the TIFF file, which will allow a reasonable download time for the average Web-based user. Where both master and use copies are created (in many instances, the master copy also serves as the use copy) the principals outlined below apply to the master copy, though some apply equally well to the use copy.

Among the advantages in reaching agreement about what constitutes good objects are the following:

- By agreeing to minimum level benchmarks for good objects, organizations that produce such objects can reduce the risk involved in producing and maintaining them while inspiring confidence in and encouraging their use.
- Because good objects will be considered capable of meeting known current and likely future needs,

organizations can invest in their creation secure in the knowledge that they will not be forced to re-create the objects at some future date even as production techniques improve.

- Users of good objects will develop confidence in the objects because they will have a minimum level of well-known and consistent properties, and will support a variety of known uses.
- By building consensus around the characteristics of good objects organizations that produce and support their use will be able more effectively to:

  write contracts with vendors who create such objects and to compare vendors' prices
  - commit to making good objects accessible over the longer term - good objects will be invested with an intrinsic value that makes them worth maintaining
  - level up their data creation efforts to a point where they produce objects of known quality capable of supporting a number of known uses
  - instill confidence in users who will know that good objects support their needs
  - define and narrow preservation options as may be required to migrate or emulate good objects

Principles applying to good objects

Objects principle 1. *A good digital object will be produced in a way that ensures it supports collection priorities.*

How a digital object is produced and described will determine whether, how, by whom and at what cost to whom it can be accessed and used over the longer term. Accordingly, decisions about how objects are produced and described should reflect and follow from those made about why they are being produced and for whom or what purpose. For that reason, the guidelines for selection listed in COLLECTIONS are equally relevant to the creation of good objects.

Some examples of how decisions about production and description should follow naturally from strategic collection development decisions are available in Neil Beagrie and Daniel Greenstein, "A Strategic Policy for Creating and Preserving Digital Collections (1998). http://www.ahds.ac.uk/strategic.pdf

Objects principle 2. *A good object is persistent. That is, it will be the intention of some known individual or institution that the good object will persist; that it will remain accessible over time despite changing technologies.*

Digital information is notoriously volatile. Imagine the difficulties involved ten (let alone 50 or 100!) years from now in accessing a

digital object that is created today. Even if the physical medium (e. g., CD, hard drive) that carries the object survives uncorrupted, it is unlikely that a computer will exist that is capable of reading the medium. How many computers are today are capable of handing 5.25-inch floppy disks? And even if such computers are found to exist, it isn't clear they will have the operating systems or software capable of rendering the machine-readable information into something that can be made sensible to a user with then-current software.

Two strategies are available to ensure that objects persist. The first is migration. It involves transforming objects so they can move between technical regimes as those regimes change. Migration occurs at all levels, as objects are moved:

- across media as media evolve (e.g. from diskette to CD, and from CD to optical disk or DAT tape);
- across software products as the products become outmoded (e.g. from one version of a word-processing or database package to another); and,
- across formats as formats evolve (e.g. from SGML to XML, as is the case today with so many encoded ASCII texts).

The second strategy involves emulation. This assumes that in some cases, it is better (involves less expense and/or less information loss) to emulate on contemporary systems the computer environment in which digital objects were originally created and used. Emulation strategies may be particularly appropriate for complex multimedia objects such as interactive learning modules.

Although no single production decision about format, compression, etc. will guarantee that an object will persist, some decisions are safer than others. Some formats, at least, will be easier to maintain at lower cost across changing technical regimes. A good object, then, will either have a known preservation strategy (e.g. as with SGML-encoded ASCII texts where migration through changing regimes is both known and deemed viable and cost effective) or a good chance of evolving such a strategy (e.g. where widespread commercial investment in the format- PDF - makes development of an effective preservation strategy highly likely).

A large and growing literature on digital preservation exists. Some particularly salient references include:

- PADI (Preservation Access to Digital Information), a comprehensive, well-maintained clearinghouse to all types of information resources related to digital preservation. http://www.nla.gov.au/padi/

- Reference Model for an Open Archival Information System. This reference model, currently in an ISO standards track, provides a high-level conceptual framework for thinking about persistence and preservation of digital objects. http://ssdoo.gsfc.nasa.gov/nost/isoas/overview.html
- Gregory W. Lawrence, William R. Kehoe, Oya Y. Rieger, William H. Walters, and Anne R. Kenney, Risk Management of Digital Information: A File Format Investigation (CLIR 2000). This report is based on an investigation conducted by Cornell University Library to assess the risks to digital file formats during migration. http://www.clir.org/pubs/abstract/pub93abst.html
- Jeff Rothenberg, Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation (CLIR 1999). Elaborates a proposal for emulating obsolete software/hardware systems on future, unknown systems, as a means of preserving digital information far into the future. http://www.clir.org/pubs/abstract/pub77.html
- Conservation On-Line (COOL) Preservation of Audio Materials. A clearinghouse of resources related to preserving both digital and analog audio. http://sul-server-2.stanford.edu/bytopic/audio/

Objects principle 3. A good object is digitized in a format that supports intended current and likely future use or that support the development of access copies that support those uses. Consequently, a good object is exchangeable across platforms, broadly accessible, and will either be digitized according to a recognized standard or best practice or deviate from standards and practices only for well documented reasons.

In almost every case, there is a direct correlation between the production quality of a digitized object and the readiness and flexibility with which that object may be migrated across platforms. As a result, the digitization of objects at the highest affordable quality can pay off in the long run as the objects are rendered more useful and more flexibly accessible over the longer term.

Having said that, not all objects require such investment. A spreadsheet that is used to calculate 2001 tax liabilities, or a digital image showing Michael, age 3.5 on his new bike may have substantial local and immediate value but also very limited long-term worth. The spreadsheet might be printed out and included in a personal paper archive until destroyed whenever the statute of limitations expires. The picture of young Michael may be created from a 35mm slide that is considered to be the long-term master. In both cases, there is very good reason to invest as little as possible in the creation of persistent objects. The point is that

nearly every digitization project needs to determine the value of the digitized objects themselves and to make appropriate decisions about persistence and interoperability.

Formats are presented Table 1 below. They are organized according to a typology that recognizes data types, and within data types, applications to which objects of that type may be put. The approach (derived from one that has become common in Europe) is extensible with respect both to the number of data types and applications that it recognizes.

TABLE 1. A TYPOLOGY OF FORMATS

DATA TYPE	APPLICATIONS	FORMATS	GUIDELINES and REFERENCES
Alphanumeric data	Flat files; hierarchical or relational datasets.	Comma-delimited ASCII, or portable format files recognized as de facto standards (e. g. SAS and SPSS) with enough metadata to distinguish tables, rows, columns, etc.	For social science and historical datasets, see Guide to Social Science Data Preparation and Archiving (ICPSR 2000) http://www.icpsr. umich.edu /ACCESS/ dpm.html, and Digitizing history, a guide to creating electronic resources from historic documents (HDS, 1999) http://hds. essex. ac.uk/g2gp /digitising_ history /index.aspm.
	Encoded texts for networked	SGML, XML; use documented DTD's	
	presentation and exchange of text-based information	or schema	

	Encoded texts for literary and linguistic content analysis	SGML, XML	Text Encoding Initiative (TEI) http://www. tei-c.org. Creating and documenting electronic texts (OTA, 1999) http://hds.essex. ac.uk/g2gp/ digitizing_history /index.asp and TEI text encoding in Libraries: Guidelines for Best Practice (DLF, 1999) http://www.diglib. org/standards /tei.htm
Image data (raster graphics) bitonal, grayscale and color images of pictures, documents, maps, photographs	Book or serial publication prepared as preservation digital master or access surrogate for source	Archival masters likely to be TIFF files at color depth and pixelation appropriate for application. Derivative data likely to vary depending on use	Anne R. Kenney, Oya Y. Rieger, et al Report of the Digital Preservation Policy Working Group on Establishing a Central Depository for Preserving Digital Image Collections (March 2001) at http://www.library .cornell.edu/ preservation /IMLS/image_ deposit _guidelines.pdf Library of Congress, The Preservation Digital Reformatting Program: Image Specifications (September 2001).

			The most recent consensus is available in Draft benchmark for digital reproductions of book and serial publications (DLF, 2001), at http://www.diglib. org/standards/ draftbmark.htm An example of one institution's local benchmarks: California Digital Library. Digital Object Format Standards. http:// www.cdlib. org/about/ publications/ CDLImageStd-1001.pdf
Scalable image data (vector graphics) presentations, creative graphics, computer-aided designs, clip art, line drawings, 3-D models, maps	maps, herbarium specimens	MrSid from LizardTech becoming a de facto standard although proprietary	
Audio	music audio	Archival masters likely to be IFF or AIFF. Delivery formats may be RealAudio.	A brief technical introduction to Digital Audio by the National Library of Canada http:// www.nlc-bnc. ca/9/1/p1-148-e. html. Harvard University Library Digital Initiative Audio Reformatting http://hul.harvard .edu/ldi/html/ reformatting_ audio.html.

			Currently the site has links to industry standards and will include project guidelines in the future.Sound Practice: A Report on the Best Practices for Digital Sound Meeting, 16 January 2001 at the Library of Congress http://www.rlg.org/preserv/diginews/diginews5-1.html#feature3	
	spoken word (e.g. oral histories)		See music audio above.National Gallery of the Spoken Word http://www.matrix.msu.edu/ngsw/cfm/tech_issues.cfm. Old Dominion University on-line Digital Training Workshop 2000 http://www.lib.odu.edu/services/dcenter/dtw2000/index.html	
Video				
In process				
Multimedia	GIS	GIS often combines data in multiple formats: GPS, alphanumeric data (e.g. as required to record co-ordinate data), vector and raster	GIS. A guide to good practice (ADS, 1998) http://ads.ahds.ac.uk/project/goodguides/gis/index.html	

		graphics (e.g. to represent maps)	

Objects principle 5. A good object will be named with a persistent, unique identifier that conforms to a well-documented scheme. It will not be named with reference to its absolute filename or address (e.g. as with URLs and other Internet addresses) as filenames and addresses have a tendency to change. Rather, the filename's location will be resolvable with reference to its identifier.

How an object is identified determines how (even whether) it may be found and thus made accessible over both the short and longer terms. There are at least two approaches to the provision of persistent and unique object identifiers. The first involves assigning identifiers that conform to a standard, and using applications that ensure that those names resolve to the object's filename and location.

Where application of national and international standards is beyond an institution's technical capabilities (as it is likely to be at most smaller and even medium-sized institutions), a more local approach may be considered. This involves developing and maintaining a local scheme that uniquely identifies information objects, and mechanisms for ensuring that names resolve to file locations. Where local schemes are used they should be documented and documentation should be accessible.

A third, middle way that is appropriate for Internet accessible objects is available by assigning PURLs (Persistent URLs) instead of URLs. The PURLs embedded in references to the object are resolved to true locations by a server which contains tables mapping PURLs to URLs. Although the mapping tables must be updated when an object is moved, this degree of indirection facilitates maintenance by ensuring each PURL need only be updated once in a central spot, no matter how many times it occurs in references.

The following sites contain information about standard numbers:

- International Standard Book Number system (ISBN). http://www.isbn.spk-berlin.de/
- Digital Object Identifier. http://www.doi.org
- International Standard Serial Number (ISSN). http://www.issn.org/index-eng.html
- Serial Item and Contribution Identifier. http://sunsite.berkeley.edu/SICI/

For information about the Persistent Uniform Resource Locator (PURL) see http://www.purl.org/.

For more information about Uniform Resource Names (URNs)
see http://www.ietf.org/html.charters/urn-charter.html

For information about the application of naming schemes see

- "Handle Server", Library of Congress (1998) http://lcweb2.
  loc.gov/ammem/award/docs/handle-server.html supplying
  a description of the Library's experiments using "handles,"
  one form of URN
- Harvard University Library Office for Information
  Systems, "Naming and Repository Services. An
  Introduction" http://hul.harvard.edu/ldi/resources/
  nrsdrsservice.pdf. A detailed introduction to these services
  as supplied by the Office for Information Systems and
  including a useful gentle description of the importance and
  design of naming services, good practices, etc.
- Harvard University Library Office for Information
  Systems, "Name Resolution Service (NRS) Technical
  Overview", (2000) see http://hul.harvard.edu/ldi/resources/
  nrs-overview-public.html. Supplies a technical overview
  for the Name Resolution Service (NRS) developed by the
  Library Digital Initiative at Harvard University Library.
  The NRS is a comprehensive service for creating,
  maintaining, and resolving names, which are persistent,
  location-independent identifiers for network-accessible
  resources.

*Objects principle 6. A good object can be authenticated in at least
two senses. First, a user should be able to determine the object's
origins, structure, and developmental history (version, etc.).
Second, a user should be able to determine that the object is what
it purports to be.*

Being able to authenticate an object is essential for a number of
reasons. Research is predicated on verifiable evidence. Teaching
and learning, as well as other forms of cultural engagement, also
rely on verification, although it is more frequently thought of in
terms of a user's ability to assess an information object's veracity,
accuracy, authenticity, even worth. There are some cases where
verification takes on additional significance, as for example, with
the networked representation of information that supplies
evidence about important past or current events.

Typically, information necessary for a user to determine an
object's origin, structure, and developmental history is included
with the metadata that is supplied for and about that object (see
METADATA).

Determining the veracity of a digital object is likely to rely upon
techniques that are known but whose reliability is still debated.
Techniques appropriate to digital images may include digital

signatures and water marking. Checksums and other technical routines that produce message digests are appropriate for objects in virtually all formats. They help determine by analyzing the object's structure and composition whether it has been changed in any way since some particular benchmark point.

Information may be found at

- *Authenticity in a Digital Environment* (CLIR, 2000). Report of a group of experts convened by CLIR to address the question: What is an authentic digital object? http://www.clir.org/pubs/reports/pub92/contents.html
- The importance of verifying the authenticity of an information object is well described in *The Evidence in Hand: Report of the Task Force on the Artifact in Library Collections* (2001) http://www.clir.org/activities/details/artifact-docs.html
- MD5 unofficial home page http://userpages.umbc.edu/~mabzug1/cs/md5/md5.html
- On checksum see http://www.checksum.org/
- On digital signatures see http://www.w3.org/DSig/ and information from the Electronic Privacy Information Center
- On digital watermarking see *The information hiding homepage. Steganography and digital watermarking* http://www.cl.cam.ac.uk/~fapp2/steganography/

Objects principle 7. A good object will have and be associated with metadata. All good objects will have descriptive and administrative metadata. Some will have metadata that supplies information about their external relationships to other objects (e.g. the structural metadata that determines how page images from a digitally reformatted book relate to one another in some sequence).

The Philadelphia Art Museum reports some 300,000 unique items in its collection. None of those objects would be of any use to anyone if the PMA did not also retain for each of its objects information about what it is, where it is located, when it was created, and similar information. Digital objects without metadata would be equally useless.

This principle does not prescribe what metadata will be supplied. This issue is another where fitness for purpose comes into play. Nor does it assume how metadata will be related to objects. Some objects will have metadata embedded within them (such as an encoded text with an XML header; an image with a TIFF header). With others, metadata will be stored and managed separately, as another digital object in fact.

For more information see METADATA.

METADATA

One of the most challenging aspects of the digital environment is the identification of resources available on the Web. The existence of searchable descriptive metadata increases the likelihood that collections will be discovered and used. Collection-level metadata is addressed in the COLLECTIONS section of this document. This section addresses the description of individual objects and sets of objects within collections.

Metadata is defined as "data about data" or "information about information". Anne Gilleland-Swetland, in Introduction to Metadata : Pathways to Digital Information (http://www.getty.edu/ research/institute/standards /intrometadata) states, "Perhaps a more useful 'big picture' way of thinking about metadata is as 'the sum total of what one can say about any information object at any level of aggregation.'" Gilleland-Swetland goes on to note that there are three basic kinds of metadata:

- Content, which relates to what the object contains or is about, and is intrinsic to an information object.
- Context, which indicates the who, what, why, where, and how aspects associated with the object's creation and is extrinsic to an information object.
- Structure, which relates to the formal set of associations within or among individual information objects and can be intrinsic or extrinsic.

These types of metadata are commonly known as descriptive, administrative and structural, respectively. Descriptive metadata helps users find objects, distinguish one object from another, and know something about objects they have found. Administrative metadata helps collection managers keep track of objects for such purposes as file management, rights management and preservation. Structural metadata can be thought of as the glue that binds compound objects together, relating, for example, articles, issues and volumes of serial publications, or the pages and chapters of a book.

A primary reason for digitizing collections is to increase access to the resources held by the organization. Creating broadly accessible metadata is a way to maximize access by current users and attract new user communities. Examples of metadata systems include library catalogs, archival finding aids, and museum inventory control or registrar systems. Over the years, metadata formats have been developed for a wide range of digital objects. Within this range of formats, there is a degree of consistency across all metadata schemes that supports interoperability. For

example, most if not all schemes provide for a title field, date field, and identifier field. It is important that cultural heritage institutions explore the metadata standards that are being adopted within their field, as well as across the broader cultural heritage environment, to assure the greatest likelihood of interoperability.

There is usually a direct relationship between the cost of metadata creation and the benefit to the user: describing each item is more expensive than describing collections or groups of items, using a rich and complex metadata scheme is more expensive than using a simple metadata scheme, applying standard subject vocabularies and classification schemes is more expensive than assigning a few keywords, and so on. The decisions of which metadata standard(s) to adopt, what levels of description to apply, and so on must be made within the context of the organization's purpose for digitizing the collection, the users and intended usage, approaches adopted within the community, and the desired level of access. Questions to consider include, but are not limited to:

- What type of cultural heritage institutions will be involved in the project?
- What subject discipline will be involved?
- What is the format of the original resources?
- Is there an existing metadata system used by the organization?
- Are the materials organized as a collection?
- Does information exist that supports detailed description of the object?
- Should the source object be described, or the digital version of it?

Principles applying to good metadata:

Metadata Principle 1: Good metadata should be appropriate to the materials in the collection, users of the collection, and intended, current and likely use of the digital object.

There are a variety of published metadata schemes that can be used for digital objects, Web sites, and e-resources. There will often be more than one scheme that could be applied to the materials in a given collection. The choice of scheme should reflect the level of resources the project has to devote to metadata collection, the level of expertise of the metadata creators, the expected use and users of the collection, and similar factors. Organizations should consider the granularity of description, that is, whether to create descriptive records at the collection level, at the item level, or both, in light of the desired depth and scope of access to the materials. They should also consider which schemes are commonly in use among similar organizations; using the same metadata scheme will improve interoperability among collections.

The International Federation of Library Association site Digital Libraries: Metadata Resources is a clearinghouse of metadata schemes. http://www.ifla.org/II/metadata.htm

A good general introduction to metadata issues for cultural heritage institutions is Introduction to Metadata: Pathways to Digital Information (Murtha Baca, ed.) http://www.getty.edu/research/institute/standards /intrometadata/index.html

The following are examples of the major schemes in use in cultural heritage institutions. Links to toolkits, tutorials, implementation software, and examples of projects that have adopted the standards are included in addition to links to the standards.

- Dublin Core: A simple generic element set applicable to a variety of digital object types. Dublin Core has been adapted by a number of communities to suit their own needs (such as the CIMI application profile for the museum community), and incorporated into a number of domain-specific metadata schemes.
- Dublin Core Initiative: http://dublincore.org
- Open Archives Initiative application of Dublin Core http:// www.openarchives.org
- The CIMI Guide to Best Practice for museums using Dublin Core http://www.cimi.org/public_docs/ meta_bestprac _v1_1_210400.pdf
- The GEM (Gateway to Educational Materials) application of Dublin Core http://www.geminfo.org/Workbench/ Metadata/
- IMS: A complex metadata scheme for educational resources developed by the IMS Global Learning Consortium, Inc., a group with heavy commercial participation from major hardware and software vendors. http://www.imsproject.org/metadata/
- EAD: Encoded Archival Description is set of rules for designating the intellectual and physical parts of archival finding aids so that the information can be searched, retrieved, displayed and exchanged. EAD is written in the form of a Standard Generalized Mark-up Language (SGML) Document Type Definition (DTD).
- EAD: http://lcweb.loc.gov/ead/
- The EAD Cookbook by Michael Fox http://jefferson. village.virginia.edu/ead/ cookbookhelp.html
- SAA. EAD Working Group. Encoded Archival Description Application Guidelines. SAA, 1999.
- Research Libraries Group recommended application

- guidelines for EAD. http://www.rlg.
- Online Archives of California recommended application guidelines for EAD. http://www.cdlib.org/about/publications/oacbpg-2001.pdf
- A case study of using EAD with image navigation software. http://www.library.uiuc.edu/ahx/ead/tech/CILMarch 2001.ppt
- MARC: MARC is a long established standard within the library community for exchanging cataloging information. MARC supports the Anglo-American Cataloging Rules and is maintained by the Anglo-American library community. Over the last several years, MARC has been enhanced to support descriptive elements required of electronic resources.
- MARC: http://lcweb.loc.gov/marc/
- Library of Congress. Understanding MARC Bibliographic: Machine-Readable Cataloging. (5th Edition). http://lcweb.loc.gov/marc/umb/
- MARC documentation: Extensive documentation is available at the LC site and at OCLC http://oclc.org
- Content Standard for Digital Geospatial Metadata: Developed by the Federal Geographic Data Committee, the CSDGM (FGDC-STD-001-1998) describes the content, quality, condition, and other characteristics of geospatial data.
- FGDC: http://www.fgdc.gov/metadata/metadata.html
- Factbook: http://www.fgdc.gov/publications/documents/metadata/metafact.pdf
- Tutorials: http://www.fgdc.gov/metadata/metatut.html
- Global Information Locator Service: A standard developed to describe government information resources, generally at the collection or agency level, but also usable at the item level.
- GILS: http://www.gils.net/
- U.S. National Archives and Records Administration (NARA). Guidelines for the Preparation of GILS Core Entries. http://www.ifla.org/documentslibraries/cataloging/metadata/naragils.txt
- See also NARA Bulletin 95-3. http://www.ifla.org/documentslibraries/cataloging/metadata/bull95-3.txt
- DDI Codebook: A standard for representing "codebooks" (descriptions of social science datasets) in XML. Developed by the Data Documentation Initiative (DDI), a collaborative project of the social science community.
- DDI homepage: http://www.icpsr.umich.edu/DDI/
- VRA Core Categories: The Visual Resources Association

has developed a scheme for the description of art, architecture, artifacts and other visual resources. Now in version 3, the Core Categories were designed with the awareness that there are often multiple representations of a work of art, such as the original painting and a slide of the painting used in teaching.

- VRA Core Categories, version 3 http://www.gsd.harvard.edu/~staffaw3/vra/vracore3.htm
- The VRA Data Standards Committee page http://www.gsd.harvard.edu/~staffaw3/vra/datastandardscomm.htm

Metadata principle 2: Good metadata supports interoperability.

Teaching, learning and research today operate in a distributed networked environment. Identifying resources that are distributed across the world's college and university libraries, archives, museums and historical societies is extremely difficult. Cultural heritage institutions must design their metadata systems so that they support the interoperability of these distributed systems.

Use of standard metadata schemes facilitates interoperability by allowing metadata records to be exchanged and imported into other systems that support the chosen scheme. Most standards schemes have also been mapped to other schemes. These mappings, or crosswalks, help users of one scheme to understand another, can be used in automatic translation of searches, and allow records created according to one scheme to be converted by program to another. If a locally created metadata scheme is used in preference to a standard scheme, a crosswalk to some standard scheme should be developed.

- The Getty Standards Program maintains crosswalks relevant to art, architecture and cultural heritage information on their Metadata Standards Crosswalks page http://www.getty.edu/research/institute/standards/intrometadata/3_crosswalks/index.html.
- The Library of Congress maintains crosswalks to and from MARC (http://lcweb.loc.gov/marc/marcdocz.html).
- The NSDL Standards Working Group has a metadata resources page largely devoted to crosswalks. http://128.253.121.110/NSDLmetaWG/IntroPage.html

One way to increase interoperability is to support the metadata format and harvesting protocol of the Open Archives Initiative. Systems that support OAI can expose their metadata to harvesters, allowing their metadata to be included in large databases and used by external search services. http://www.openarchives.org/

Another way to increase interoperability is to support protocols for cross-system searching. Under this model, the metadata remains in the source repository, but the local search system accepts queries from remote search systems. The best know protocol for cross-system search is the international standard Z39.50 http://lcweb.loc.gov/z3950/agency/.

Metadata principle 3. *Good metadata uses standard controlled vocabularies to reflect the what, where, when and who of the content.*

Content should be expressed in a standard form selected from standard lists. Examples of controlled vocabularies, include standard subject heading lists (e.g. Library of Congress Subject Headings), thesauri (e.g. the Art & Architecture Thesaurus) and taxonomic lists (e.g. TRITON, Taxonomy Resource and Index to Organism Names). Locally defined vocabularies, where appropriate, can be utilized. Classification systems (e.g. Dewey Decimal Classification) can also be used to provide subject access. Vocabularies should be consistently applied and the application documented.

Controlled vocabularies, thesauri and classification systems available in [sic] the WWW lists several dozen web-accessible controlled vocabularies by subject area. http://www.lub.lu.se/metadata/subject-help.html.

The High Level Thesaurus Project (HILT) is a clearinghouse of information about controlled vocabularies, including related resources, projects, and an alphabetical list of thesauri. http://hilt.cdlr.strath.ac.uk/Sources/index.html

The Getty Vocabulary Program builds, maintains, and disseminates several thesauri for the visual arts and architecture:

> Art & Architecture Thesaurus® (AAT) http://www.getty.edu/research/tools/vocabulary/aat/
> - Union List of Artist Names® (ULAN) http://www.getty.edu/research/tools/vocabulary/ulan/
> - Getty Thesaurus of Geographic Names™ (TGN) http://www.getty.edu/research/tools/vocabulary/tgn/

Some other controlled vocabularies:

- Library of Congress Subject Heading List-Available through OCLC, RLG and other cataloging services and on CD ROM from the Library of Congress.
- Medical Subject Heading List: http://www.nlm.nih.gov/meshhom.html

- Thesauri of Graphic Materials I: http://lcweb.loc.gov/rr/print/tgm1/
- Thesauri of Graphic Materials II: http://lcweb.loc.gov/rr/print/tgm2/
- Thesaurus of Graphic Names: http://www.gii.getty.edu/vocabulary.tgn.html

Metadata principle 4. Good metadata includes a clear statement on the conditions and terms of use for the digital object.

Terms and conditions of use include legal rights (e.g. fair use), permissions and limitations. The user should be informed how to obtain permission for restricted uses, and how to cite the material for allowed uses. Special technical requirements, such as the required viewer or reader should also be noted

If this information is the same for all the materials in a collection, documenting it in collection-level metadata is adequate (see COLLECTIONS). Otherwise metadata records for individual objects should contain information pertaining to the particular object. Many metadata schemes have designated places to put this information; if they do not, some locally-defined element should be used.

Metadata principle 5: Good metadata records are objects themselves and therefore should have the qualities of good objects, including archivability, persistence, unique identification, etc. Good metadata should be authoritative and verifiable.

Metadata carries information that vouches for the provenance, integrity and authority of an object. Metadata's own authority must be established. Clues to the authority of a metadata record include the identification of the institution that created it and what standards of completeness and quality were used in its creation. The institution should provide sufficient information to allow the user to assess the veracity of the metadata, including how it was created (automated vs. manually created), what standards/schemes were used, and what vocabularies were used.

The problem of non-authentic and inaccurate metadata is real and serious. Many Internet search engines deliberately avoid using metadata embedded in HTML pages because of pervasive problems with spoofing (one organization supplying misleading metadata for a resource belonging to another organization) and spamming (artificially repeating keywords to boost a page's ranking). The same techniques used to verify the integrity and authenticity of digital documents (e.g. digital signatures) can also be applied to metadata (see OBJECTS).

6. *Good metadata supports the long-term management of objects in collections.*

Administrative metadata is information intended to facilitate the management of resources. It can include data such as when and how an object was created, who is responsible for controlling access to or archiving the content, what control or processing activities have been performed in relation to it, and what restrictions on access or use apply. Technical metadata, such as capture information, physical format, file size, checksum, sampling frequencies, etc., may be necessary to ensure the continued usability of an object, or to reconstruct a damaged object. Preservation metadata is a subset of administrative metadata aimed specifically at supporting the long-term retention of digital objects. It may include detailed technical metadata as well as information related to the rights management, management history, and change history of the object.

The Dublin Core Metadata Initiative proposed but never finalized a simple set of administrative data elements. Despite the unfinished and unapproved nature of the work, some implementers have found it useful. http://metadata.net/admin/draft-iannella-admin-01.txt

Two of the most widely reviewed preservation metadata element sets are the National Library of Australia's Preservation Metadata for Digital Collections (http://www.nla.gov.au/preserve/pmeta.html) and the RLG PRESERV element set (http://www.rlg.org/preserv/presmeta.html).

The PADI (Preserving Access to Digital Information) clearinghouse at (http://www.nla.gov.au/padi) has a long annotated listing of resources related to preservation metadata at http://www.nla.gov.au/padi/topics/32.html.

The Digital Imaging Group's DIG35 Specification: Metadata for Digital Images. Version 1.0, August 30, 2000 (http://www.digitalimaging.org/) specifies technical metadata for images created by digital cameras. A draft NISO standard under development, Data Dictionary for Technical Metadata for Digital Still Images (http://www.niso.org/DataDict.html) focuses on images created by scanning.

Structural metadata relates the pieces of a compound object together. If a book consists of several page images, it is clearly not enough to preserve the physical image files; information concerning the order of files (page numbering) and how they relate to the logical structure of the book (table of contents) is also required. Most schemes for recording structural metadata are local to a given institution or application. There is, however, an emerging standard that provides a framework for encoding descriptive, administrative, and structural metadata called the

Metadata Encoding and Transmission Standard (METS) http://www.loc.gov/standards/mets/.

PROJECTS

Projects are initiatives of finite duration, designed to accomplish a specific goal. Often a grant application contains the project plan, which is begun when the grant is awarded and ends when grant funding runs out. With good luck and good planning, this is coterminous with the accomplishment of the objectives of the project. However, it is important to distinguish between the project, which is transient, and the collection, which in most cases should persist. If the intent is for the collection to be maintained after the end of the project period, plans must be made for incorporating collection maintenance into the normal operating procedures of the responsible institution.

Projects to build digital collections often involve a cross-disciplinary subset of one institution's staff, but may also involve representatives from multiple institutions. Different people will contribute different skills and perspectives. However, it is important that there be one individual who is responsible for coordinating the work of all project participants and maintaining the project plan and timeline. The project manager may report to a higher manager, to a board of directors, or to an advisory board. However, the project manager should have the authority to delegate work, make decisions, and take remedial actions within the parameters set by the higher agency.

Projects principle 1: A good project has a substantial design component.

Design includes all aspects of project planning, from processing workflow to the ultimate look and feel of the collection website. A realistic assessment of the functional requirements of users needs to be a key element in design. Some early projects are notorious for devoting major resources to sophisticated display functionality when their users mostly wanted printed documents.

The Washington State Library Digital Best Practices site has a section on Project Management with a focus on market research as a tool for both design and promotion. http://digitalwa.statelib.wa.gov/newsite/projectmgmt/index.htm. The Colorado Digitization Alliance provides an example of the identification of market segments and their varying needs. http://coloradodigital.coalliance.org/users.html

RLG/DLF Guides to Quality in Visual Resource Imaging: 1. Planning an Imaging Project. http://www.rlg.org/visguides/visguide1.html

Northeast Document Conservation Center. *Handbook for Digital Projects: A Management Tool for Preservation & Access. III: Considerations for Project Management.* http://www.nedcc.org/digital/dighome.htm

Projects principle 2*: A good project has an evaluation plan.*

The IMLS encourages outcomes-based evaluation for their funded projects, and points to supporting resources. http://www.imls.gov/grants/current/crnt_obe.htm. A generic Basic Guide to Outcomes-Based Evaluation for Nonprofit Organizations with Very Limited Resources is available at http://www.mapnp.org/library/evaluatn/outcomes.htm.

The University of Texas has received an IMLS grant to develop tools and guidelines that libraries, museums and other information agencies can use to evaluate and improve the utility of their websites. http://imls.lib.utexas.edu/

Projects principle 3*: A good project produces a project report.*

The primary goal of any project should be to accomplish its stated objectives within the time and budget allowed. However, the knowledge gained in implementing a digital collection should not be lost to other organizations. Although most funding agencies require some sort of report at the end of the project period, these are not always generally available. A project report providing a detailed description and honest assessment of work accomplished should be produced and remain accessible on the Web indefinitely.

Some examples of useful, comprehensive project reports:

- Library of Congress. Manuscript Digitization Demonstration Project. Final Report. October 1998. http://lcweb2.loc.gov/ammem/pictel/
- Anne R. Kenney Digital to Microfilm Conversion: A Demonstration Project 1994-1996. Final Report to the National Endowment for the Humanities PS-20781-94. http://www.library.cornell.edu/preservation/com/comfin.html
- Preserving and Digitizing Plant Images: Linking Plant Images and Databases for Public Access. November 2000. Final report from the Missouri Botanical Garden to the IMLS. http://ridgwaydb.mobot.org/mobot/imls/final.asp

Questions, comments, or problems? Contact IMLS via email imlsinfo@imls.gov
or by phone (202) 606-8536.

Questions, comments, or problems? Contact IMLS via email imlsinfo@imls.gov
or by phone (202) 606-8536.