THE STATE OF PRESERVATION METADATA PRACTICES IN NORTH CAROLINA REPOSITORIES

by
Claire Eager

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

August, 2003

Approved by:

_____

Advisor

Claire Eager. The State of Preservation Metadata Practices in North Carolina Repositories. A Master's paper for the M.S. in I.S. degree. August, 2003. 77 pages. Advisor: Helen Tibbo

Grand metadata schemas and detailed technology applications may well be ideal for the effective preservation of digital images, but they may not be practical for the vast majority of cultural heritage repositories. If resource-challenged organizations do not receive assistance in scaling these efforts to meet their available resources, hundreds of repositories may waste valuable assets on ineffective measures, or they may do nothing at all. The aim of this study is to determine where North Carolina repositories are in their efforts to digitize collections; how they are incorporating preservation metadata, if at all; and to solicit their assessment of the Metadata for the Administration and Preservation of Digital Images (MAPDI) schema created by the North Carolina Exploring Cultural Heritage Online Preservation Metadata Working Group. The accompanying MAPDI database tool was expanded to include the capture of collection-level Dublin Core discovery metadata as well as preservation metadata. Results of the survey indicate that smaller repositories may be in need of far greater assistance in their preservation efforts, and that they frequently have to place practicality before perfect practice.

Headings:

Metadata

Digital Preservation

Digital Libraries

Table of Contents

Introduction

> *There is concern that smaller institutions are at a great disadvantage. Production capacity and the creation and sustenance of architectures that are necessary for new opportunities are seldom encountered outside of large organizations [1, p.17].*

A great deal has now been written about the critical need for preservation metadata in the midst of the digital revolution facing cultural heritage institutions [1-4]. While the problems of media degradation, software and hardware obsolescence, and other technical issues involved in digital asset management have yet to be solved, experts widely agree on the vital role of preservation metadata. The subsequent steps necessitated by this conclusion have been uncertain, but these, too, are now beginning to take shape. As they do, it is important that the industry leaders – professional organizations and cultural heritage institutions with the means to forge ahead in digital preservation – create metadata schemes and data collection tools applicable to a wide range of heritage institutions.[1] Few small- and medium-sized repositories, however, are participating in preservation metadata development efforts because of a lack of

---

[1] This is happening with institutions such as the Library of Congress (LoC), Research Libraries Group (RLG), and universities including Michigan, Virginia, and the California system. These organizations are creating metadata tools such as METS, CEDARS, and the Nordic template for Dublin Core.

awareness, funds, or expertise. Grand schemas and detailed technology applications may well be ideal, but if preservation efforts cannot be scaled down to fit organizational profits, the long-term viability of digital assets in most of the country's repositories will be at risk. Such repositories may waste valuable resources on ineffective measures, or do nothing at all to preserve digital assets.

Officially a joint project of the State Library of North Carolina and the Special Collections Library of Duke University, North Carolina Exploring Cultural Heritage Online acts as a centralizing force for all North Carolina repositories that maintain "a non-living collection"[5]. NC ECHO strives "to build a statewide framework for digitization and [address] a full-range of digitization needs of the state's cultural collecting agencies"[5]. From its creation, NC ECHO has made it part of its mission not just to include repositories of all sizes, but to leverage the institutional knowledge of a statewide network to help smaller institutions keep pace with the cultural heritage community at large.

In 2002, NC ECHO formed the Preservation Metadata Working Group (PMWG) to examine the issues of preservation metadata for NC ECHO partner repositories and the emerging standards in the industry. As part of this effort, the PMWG produced the Metadata for Administration and Preservation of Digital Images (MAPDI) schema and its accompanying metadata capture database system. These tools are an attempt to provide a feasible and practical

implementation of preservation metadata for resource-challenged North Carolina repositories.

Study Overview

This study set out to determine the actual status of preservation metadata awareness and use in NC ECHO member repositories, and to get their feedback on the MAPDI schema developed by the NC ECHO working group. Through an online survey, respondents indicated that their awareness of preservation metadata was indeed minimal. While most respondents indicated past or anticipated involvement in digitization projects, the capture of preservation metadata did not figure prominently in their project architecture. Their review of the MAPDI schema suggests it successfully rendered its aim of simplicity but that the most technical elements of the schema still exceed respondents' current technical awareness.

Background

Concern for digital preservation has arisen in response to the fragility of the massive amount of digital materials now produced as a normal course of business. Whether or not digitization of analog artifacts constitutes true preservation of the original artifacts is still under some debate. What is clear is that all digital objects, be they born digital or the product of digitization, need on-going preservation from their creation to ensure that they survive the ravages of

time. In fact, the projected life span of digital materials, if left unmanaged, is far shorter than that of paper-based or other traditional materials.

Digital preservation is far more than the technical survival of bits and bytes. Hedstrom defines digital preservation as "the planning, resource allocation, and application of preservation methods and technologies necessary to ensure that digital information of continuing value remains accessible and usable" [6]. Day points out that this definition necessitates assessing digital preservation as not just a technical problem, but also an organizational challenge [7]. Coordinating the capture and management of the necessary information requires participation and support from the entire organization to be effective.

Three primary methods have been identified for the preservation of digital assets: preservation of technology through emulation and technology museums; migration of data to new formats as old ones become obsolete; and preservation of the digital object complete with its presentation information (encapsulation) [6,8]. All of these rely heavily on the use of preservation metadata, although the specific data may differ substantially to suit the method selected. Migration has emerged as a slight favorite [9], but future technological advances may yet make emulation and encapsulation more viable.

The most simplistic definition of metadata, and the most commonly used, is "data about data." In the context of digital assets, however, the term metadata is

usually broken down into specific types and uses of data about data. The Institute for Museum and Library Studies' (IMLS) "*A Framework of Guidance for Building Good Digital Collections*" [10] describes metadata as providing three basic kinds of information about digital objects - content, context, and structure. These types of data "are commonly known as descriptive, administrative, and structural [metadata], respectively" [10 p.14 of online printout]. The National Initiative for a Networked Cultural Heritage (NINCH) says that descriptive metadata "describes and identifies information resources, to facilitate searching, retrieval, and management" [11]. Common examples of descriptive metadata including bibliographic information such as author, title, and subject headings. Structural metadata "describes the internal structure of digital resources and the relationships between their parts" [11]. According to this three-part framework, administrative metadata "[helps] collection managers keep track of objects for such purposes as file management, rights management, and preservation" [11], thus placing preservation metadata as a subset of administrative metadata. Gilliland-Swetland [12] breaks down metadata further into five equal categories: administrative, descriptive, preservation, technical, and use. Regardless of the specific definition used, good metadata describes resources at varying levels of aggregation, and its content depends on the needs and use of the systems and users accessing it.

Having identified many of the issues facing the management and preservation of digital resources [3, 4, 13], what we know best is that no one knows exactly what

future technology will look like or what impact it will have on today's efforts. Collecting metadata is a way to manage that uncertainty as best we can with the information available today.

Metadata Models & Schemas

The Open Archive Information System (OAIS) is an over-arching reference model developed by the Consultative Committee for Space Data Systems (CCSDS) for the long-term preservation of digital data. The reference model, now an ISO standard, "establishes a common framework of terms and concepts which comprise an Open Archive Information System (OAIS). It allows existing and future archives to be more meaningfully compared and contrasted. It provides a basis for further standardization within an archival context" [14, Foreword, p. iii]. The OAIS model was initiated as a project for the space data community, but organizations concerned with the long-term preservation of digital information have since co-opted it for use in many different industries. The OAIS model is a theoretical construct to be used as a guide for designing archives and information architectures that can share information through a standardized structure. The CCSDS stresses that OAIS is a set of recommendations, not rules, and that the unique needs of each organization will necessitate customization.

OAIS was approved as an ISO standard (ISO 14721:2002) and the blue book released in January 2002. Long before its official ISO approval, however, the

archival community began to take up the OAIS model as a strong candidate for synchronizing digital archive management efforts. Metadata has always played a critical role in the OAIS model, although as a reference model it does not recommend specific schemas. For example, the earliest draft [15] of the reference model put forth by the Computer Science Corporation and NASA in 1995 included metadata management as one of six primary system components, and one that would interact with all of the other primary components.

OAIS provides two detailed models: the Informational Model and the Functional Model. The Informational Model consists of three kinds of information packages made from four types of information objects. Critical to the informational model is the Designated Community and its knowledge base. The Designated Community is defined as “an identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities” [14, p. 1-10]. The knowledge base is the set of information assumed to be incorporated by the Designated Community and allows them to understand or interpret information.

An information object is the resulting combination of the Designated Community’s knowledge base, a data object (physical, like a blood sample, or digital, like a TIFF file), and the representation information for that data object. The representation information should be appropriate for the Designated Community’s knowledge base.

Information objects can be assigned to one of four information object types: Content Information, Preservation Description Information, Packaging Information, and Descriptive Information. These are assembled to produce three types of information packages: the Submission Information Package (SIP), the Archive Information Package (AIP), and the Dissemination Information Package (DIP). The actual data involved in each of these packages may overlap and information objects can be used and re-used as necessary to best meet the needs of a given package structure.

In order to be applicable to the broadest possible range of information, the OAIS reference model is highly generalized. The report clearly states that the description of the functional entities as provided in the reference model "is not to be taken as a recommended design or implementation" [14, p. 4-3]. The use of the Designated Community and Knowledge Base constructs gives organizations a great deal of flexibility to design systems that are customized to their needs and resources, but this flexibility may be obscured by the depth and breadth of the details in the OAIS model. One criticism of the OAIS model has been that its purely theoretical content has made it inaccessible to many professionals who lack the time, energy, and expertise to deconstruct the reference model into an applicable, practical architecture. As Tibbo states, "OAIS is a very high level model and…its translation into a working system takes a great deal of effort and time. The digital archiving community must produce modularized,

interchangeable, and portable tools that function within an OAIS environment before fully operable implementations will be possible" [1].

Practical implementation guidelines and standards are a strong first step towards such tools, and fortunately these are beginning to emerge. In 2002, the Online Computer Library Center (OCLC) and the Research Libraries Group (RLG) published a report focused on the Preservation Description Information object [16]. More recently, they published *Preservation Metadata and the OAIS Information Model* [14]. Given the complexity of the initial OAIS report, these reports, and in particular the latest on preservation metadata, are a much-needed translation of the reference model (or at least, a significant portion of it). This translation and the implementation guidelines will hopefully allow more repositories to take the first steps toward OAIS-compliant systems.

Organizations around the globe have developed metadata schemas and tools since the early 1990's but very few, if any, could be considered as accepted standards. Some initiatives have focused on text, others on images, some on description, others on preservation. The Dublin Core element set is probably the most widely recognized metadata schema for describing digital resources. [17]. As one of the most commonly known and implemented metadata schemas, Dublin Core comes the closest to being an industry standard. It is a very simple

schema[2] aimed primarily at description and resource discovery. It is usually implemented at the collection level and not for individual digital objects.

Librarians have been creating descriptive metadata for years by using the MARC standard for cataloging. This is probably the most widely adopted metadata schema in use today, even though many librarians might not realize that bibliographic cataloging equates to descriptive metadata. The adoption of the MARC format has revolutionized the library catalog by enabling machine-readable cataloging and thus, the creation of online catalog systems. However, because the MARC format has its foundations in traditional AACR2 cataloging, it is not particularly good at handling electronic resources. For example, traditional cataloging is based on the dichotomy of monographs and serials. Electronic resources like web sites do not fall neatly into either of these categories.

The Metadata Encoding and Transmission Standard (METS) is by far one of the most robust schemas. This extremely detailed schema incorporates all three metadata types (descriptive, administrative, and structural) to describe digital objects using the XML language schema[3]. A particular strength of the METS schema is its ability to describe formats other than digital images (e.g., digital audio and video), and its ability to document the relationship between multiple

---

[2] Full information on the Dublin Core Metadata Initiative is available at http://dublincore.org.

[3] An overview and tutorial on the METS schema is available at www.loc.gov/standards/mets/METSOverview.html.

digital objects that are part of a single, larger digital object. For example, a digital library might have the digital text of a book (transcription), a digital audio file of the book being read aloud, and digital images from scans of the book's cover or illustrations. METS provides the functionality to document the relationship of all these items as being part of a single object. METS is a Digital Library Federation initiative maintained by the Library of Congress, and it is compatible with the OAIS reference model.

The Cedars (CURL exemplars in digital archives) project "[aimed] to address strategic, methodological, and practical issues relating to digital preservation" [18]. The project was divided to examine three different aspects of digital preservation – digital preservation standards and techniques, collection development and rights management, and preservation metadata – but was primarily focused on "born digital" resources [7]. The part of the project looking at preservation metadata, however, did release a metadata schema based on the OAIS model [19].

Day has produced a useful review of digital preservation efforts [8]. It covers international efforts in the realm of digital preservation ranging from documents management to metadata schemas. It includes the RLG Working Group on the Preservation Issues on Metadata, which released their final report in 1998. Their work focused only on describing digital images, and used the Dublin Core and the USMARC-based core record standard as its foundation for developing a

sixteen-element set “deemed crucial for the continued viability of a digital master file”[9]. Also covered in Day’s review are InterPARES (International Research on Permanent Authentic Records in Electronic Systems)[4], which is focused on the issues of authenticity in the digital archive; NEDLIB (Networked European Deposit Library), which has developed a metadata schema revolving around the issue of technological obsolescence; and the many activities in the Australian libraries and archives community. The National Library of Australia has supported PADI (Preserving Access to Digital Information)[5] and the incorporation of the PANDORA (Preserving and Accessing Networked DOcumentary Resources of Australia) Archive[6].

The National Library of New Zealand (NLNZ) has made significant first steps towards incorporating digital asset management into the core business practices of the library. Just a few months ago they published a report on their preservation metadata efforts, the aims of which closely mirror the NC ECHO PMWG’s: “to strike a balance between the principles expressed in the OAIS Information Model and the practicalities of implementing a working set of preservation metadata” [20]. The NLNZ model has more elements than the MAPDI model, broken down into four main entities, and uses the XML schema language. Of particular note is their mission “to move digital preservation into a business-as-usual

---

[4] More information available at http://www.interpares.org.

[5] More information available at http://www.nla.gov.au/padi/index.html

[6] More information available at http://pandora.nla.gov.au/index.html

framework…away from describing the requirements of digitial preservation as ‘problematic,’" acknowledging that “the risk of such rhetoric is that digital preservation continues to be perceived outside the norms of business processes” [20]. This represents a significant shift in institutional thinking, and as a guiding philosophy it can serve repositories of all sizes equally well.

Beyond Metadata

Identifying threats to the accessibility, context, and stability of digital information is a strong first step towards ensuring the long-term viability of digital objects. Having identified the threats, organizations have produced numerous frameworks and schemas, each purporting to be a candidate for an industry standard, yet still very few can be identified as accepted standards. The absence of accepted standards leaves organizations without clear guidance for their digitization and metadata collection efforts. This kind of uncertainty is greatly magnified in organizations that are not on the cutting edge of these developments, and indeed, are probably not even following the developments. Without clear guidelines, organizations that undertake digitization efforts do so without assurance that their efforts will have long-term benefits, or even that the digital objects they create will be viable assets in a few years’ time.

The cultural heritage sector is becoming increasingly attuned to the jeopardy facing resource-challenged repositories as they venture into digital asset creation and management. The NLNZ report notes that most efforts in preservation have

been theoretical in nature which “poses some risks for organisations needing to implement preservation metadata schemas sooner rather than later” [20]. Many repositories are hanging back in their metadata efforts waiting for a clear standard, or at least a documented success story, to emerge, and those that are not waiting, my be forging ahead with insufficient data. If industry leaders have not had definitive success, the challenge for smaller repositories is that much greater. In her discussion of Ross and Gow’s *Digital Archaeology: The Recovery of Digital Materials at Risk*, Tibbo points out that the stories of digital fragility and data loss come from “major institutions that should logically be expected to have better resources, knowledge, and motivation to preserve digital information than many smaller organizations. Undoubtedly, the day-to-day preservation situation in smaller, less well-heeled institutions is far worse” [1].

The Council on Library and Information Resources’ (CLIR) report, *Building and Sustaining Digital Collections: Models for Libraries and Museums* [21] , documents an excellent cross-discipline approach to the study of digital libraries and their development. In 2001, CLIR and the National Initiative for a Networked Cultural Heritage (NINCH) brought museum and library senior executives together with “business and legal experts, technologists, and funders to discuss the challenges that cultural institutions face when putting collections online and to identify some models for sustainability that support the core missions and do not conflict with the internal cultures of nonprofit entities”, [21]. Along with anticipated issues, such as longevity of storage mediums, and access and rights

management, new issues, and new thoughts on old issues, also arose. Leaders recognized that “it is difficult for large, relatively well-funded nonprofit museums and libraries to devise business models that promise to be sustainable. This raises serious concerns about the fate of small and medium-sized institutions to have appropriate space on the Web” [21, p. 17]. Among the possible solutions was the benefit which not-for-profit organizations could gain from following for-profit business practices.

> [The participants] believed strongly that nonprofits must be as ‘businesslike’ as any entity that wants to succeed. Any other attitude is no longer feasible, let alone desirable. Doing business in the digital realm, whether for profit or not, demands large amounts of capital, new skills, and a new organizational culture. The assumption that commercial organizations are better managed, and need to be so, is not only false but dangerous [21, p. 13].

While the report was focused mainly on business models (both for-profit and non-profit), their discussions covered all aspects of a digital library, and specifically included the importance of industry standards as a way to bring all repositories forward in an efficient effort and to mitigate the professional risks inherent in a digitization project. If metadata schemas can be standardized and incorporated into projects, and these projects can attain greater success levels through collaboration and support, then the long-term viability of the digital objects created can be guaranteed.

Repositories of all sizes are beginning to understand that digitization is an inevitable part of their future, but there are many challenges to entering the field. The cost of equipment and training, the labor and time invested in the process

and long-term management of digital assets makes digitization costly on many levels. Metadata is frequently the most expensive part of the digitization process. Automated metadata production efforts are still in their infancy so producing high-quality metadata requires intensive man-hours. Inappropriate or incomplete metadata can be just as expensive. It may require significant editing or re-creating, and in some cases (such as revision history or context information), re-creation is not possible, rendering the information unusable. Finding the right balance between cost and coverage is a tricky problem, and one that does not have any immutable rules of governance.

Nonetheless, administrators are trying to move forward. As stated in the summary of CLIR's report, "even officials from public institutions, burdened by the need to maintain the ill-defined 'public trust,' agreed that not to take risk is itself a risky strategy. They are looking for ways to manage the risk intelligently" [21, p.15]. Organizations like NC ECHO can serve a crucial role as a guide and unifying force for the hundreds of repositories looking for a partner to share in the risk and responsibility of digitization.

Metadata for the Administration and Preservation of Digital Images (MAPDI)

In the summer of 2002, NC ECHO employed the School of Information and Library Science (SILS) of the University of North Carolina at Chapel Hill to form the Preservation and Administration Metadata Working Group. The task of the working group was to develop a preservation metadata schema that would be

realistic and feasible to implement for a wide range of repositories across the state. The preservation metadata schema would follow industry standards and established best practices (when available) and would be added to its published guidelines on digitization. Of particular concern, given NC ECHO's mission of inclusion, were the needs of small repositories and their ability to participate in preservation metadata efforts in a meaningful way. This necessitated a manageable schema, both in size and scope, as well as a tool for the capture and management of that metadata. "Small" or "medium" repositories were not officially defined but were generally agreed to be the resource-challenged institutions that make up the majority of the NC ECHO partners. Anecdotal evidence described dozens, if not hundreds, of repositories as "lone arrangers," institutions with one or two full-time staff and little reliable technology support.

A simple software application would be a major asset to a small repository all on its own, while a repository with greater financial resources and technological prowess could use that same tool as a basic starting point on which to build. The Working Group used this philosophy as a guideline in developing the database that would accompany the preservation metadata schema. Similar thinking guided the construction of the schema itself. With the fear that most repositories were doing nothing at all in the way of preservation metadata, the goal was to provide a schema of the bare essentials. Such a simplified schema would be manageable for smaller repositories to implement with a minimal learning curve, and once tackled, it would hopefully encourage them to investigate expanding

their participation in metadata collection with the use of other more robust schemas. Larger repositories could build upon the basic schema with other schemas or their own customized metadata, or they might have the resources and abilities to implement a more robust schema from the beginning.

The task of devising a preservation metadata scheme was greatly simplified by focusing only on digital images. NC ECHO determined that digitizing analog images was the predominant type of digitization project their partners undertook. In addition, it was recognized that developing a comprehensive scheme for all digital object formats was beyond the scope of the Working Group's time and resources. "Born digital" images were also included in the calculations. With this framework in mind, the Working Group reviewed the prominent metadata schemes and implementations of the time, including Dublin Core, METS, CEDARS, the OAIS model, VRA, the Colorado Digital Library, and the California Digital Library.

The Visual Resources Association, a group of image management professionals, primarily in the art community. In February 2002, the VRA released version 3.0 of their Core Categories metadata schema. The principle challenge for image management professionals in describing their collections was differentiating between the original work of art or architecture and the image representation of that work in their collection. The VRA Core Categories resolved this dilemma with

a single element called Record Type, which specifies whether the metadata record describes the actual work or a representation of that work.[7]

The Working Group rejected Dublin Core and the VRA core categories because both schemas dealt primarily with descriptive metadata and not preservation metadata. Their elements were aimed strictly on descriptive metadata for resource discovery and were far too simplistic for the needs of long-term preservation.

The Cedars project and METS both provided important breadth and depth to the preservation metadata considerations, but were too large and complicated to be tackled by most repositories. The METS schema contains hundreds of elements, and many of the administrative and preservation elements are highly technical. In addition, those elements are placed within XML wrappers, so even if we selected only a small portion of the elements, repositories would have to be familiar with XML. Even if repositories could obtain software tools that generate XML from database records, some knowledge of XML would be required to take advantage of the METS documentation (which, naturally, is largely given in XML), not to mention for customization, troubleshooting, and expansion of the schema once implemented in the repository. This does not in any way disparage the strength

---

[7] A full description of the VRA Core Categories metadata schema is available at http://www.vraweb.org/vracore3.htm.

and quality of the METS framework; it simply makes it impractical for the majority of repositories that the working group targeted.

The California and Colorado digital libraries were both examined at length, not only for their specific preservation metadata uses, but for their innovative information networks. As centralized, statewide initiatives, they have many strengths and valuable experiences to share with NC ECHO, and in fact, NC ECHO is based largely on the Colorado model. The California Digital Library is a program of the University of California libraries. Their Digital Image Format Standards report [22] is based largely on the Making of America (MOA) II project.[8] The administrative portion of their metadata model was thorough but not overwhelming, and had many elements common to other schemas. The CDL was where the METS model first started, and so, not surprisingly, it also recommends an XML structure. They reference the availability of XML-generating tools created by the MOA II project as a way to avoid hand-coding of the XML metadata, but, as with METS, the use of XML seemed prohibitive for the steep learning curve it entails. The Colorado Digital Library is a very similar model to NC ECHO, including repositories of different kinds from around the state. The Colorado metadata model also used elements common to other models and its simplicity coincided with the aims of the working group but it

[8] More information on MOA II available at http://sunsite.berkeley.edu/moa2/.

focuses on descriptive elements and did not discuss preservation metadata at the time of MAPDI development.

It was apparent that there was no single model that could meet the goals of the Working Group. The working group hypothesized that the repositories relying on NC ECHO for a metadata solution had most likely not yet begun to collect any administrative metadata for their digital images, and that these repositories were probably the most resource-challenged in terms of technical training and support. Thus, they were likely to take the schema recommended by NC ECHO as an “out of the box“ solution to their metadata initiatives (at least initially) without much customization or expansion. With this framework in mind, most of the schemas the working group reviewed were either too descriptive or too highly generalized to be repurposed as a general preservation solution. Other schemas were too complex to be realistically adopted by the resource-challenged repositories the Working Group aimed to accommodate. So as numerous organizations have done before, the Working Group set about combining the strongest parts of the models at hand with an eye to best practices and practicality. There were several administrative elements that were common to most schemas and capture of those elements was taken to be *de facto* best practice. Determining technical metadata elements was a challenge, however, as the group sought to balance simplicity with thorough preservation information. Figure 1 provides the element list for MAPDI and basic justification information for why an organization should

capture that data. It also indicates the requirement status for each element. Most elements are suggested, but not required.

Figure 1.

	Field	Use*	Justification
Identifying the Repository	Repository ID	S	For collaborative projects, this field ensures that items and collections will be uniquely identified across repositories.
Identifying the Digital Image	Item ID	R	Database system generates this identifier.
	Alternate Item ID	S	If the repository already has a system for uniquely identifying digital objects this will allow data linking between systems.
	Item Title	S	The most descriptive way to identify an image and a strong retrieval key.
	Collection ID	R	This ID is necessary to bring the items within a digital collection together within the repository's database.
	Collection Title	O	Most users and curators will identify the collection by this title.
Source Material	Source ID	S	Links the digital image to its original, whether the original is digital or analog.
	Source Type	S	Provides clarification beyond the value "Image," which is recommended for use in Dublin Core when describing photographs.
	Source Creation Date	S	Assists repositories to monitor when content enters the public domain.
	Unit of Measurement	S	
	Physical Dimensions of Source	S	Useful to researchers and allows comparison of size of digital object to its source object.
Creation of the Digital Object	Physical Dimensions of Area Scanned	S	When presented with Physical Dimensions of Source, gives some sense of proportion of the digital image with respect to original. Mainly of value to the user.
	Creation Date	R	
	Digital Creator	O	
	Capture Hardware	S	Will assist in determining generation of item.
	.accessories	O	In digital photography, light source and lenses can be relevant to the digital image produced.
	Capture Software	S	
	.settings	S	
	Image Manipulation Software	*S*	
Creation of the	.settings	S	

	Field	Use*	Justification
Digital Object	Resolution	S	Strong quality indicator of image.
	Compression	S	
	.type	S	Indicates compression algorithm used, which may be important to future migration or display.
	.degree	S	Indicates compression algorithm used, which may be important to future migration or display.
	Dimensions of Digital Object	S	Important for complete display of image.
	Bit Depth	S	Quality indicator.
	Controls	S	
	.color bar/gray scale	O	Important to ensure tonal quality of the digital image and any future derivatives.
Creation of the Digital Object	.control target	O	Important to ensure resolution quality of the digital image and any future derivatives.
	Color Space	O	Most images made for use in digital displays are in RGB. Images that are made for use in printing (brochures, ads, etc.) are usually in CMYK. Digital masters stored for print use may be in CMYK. Other options include YCbCr or CIE Lab.
	Watermark	O	Watermarks can have repercussions for future use or migration.
	File Format	R	While the file format often can be derived from the filename extension, providing it here allows for much faster searching and indexing within the database.
	Filename	R	
	Digital Master	R	Digital masters should be identified for preservation purposes.
Revisions	Revision History	O	This allows the repository to track changes to a single image over a long period of time. This could provide valuable information about migrations to other file formats, size changes, exposure changes, etc.
Rights Management	Repository Copyright	S	The creator of a digital image automatically owns the digital image, but not necessarily the copyright to the content of the image.
	Standard Rights	S	Most digital images will fall under a standard rights and distribution restriction policy of the repository.

*** R = Required, S = Strongly Suggested, O = Optional**

The final preservation metadata schema crafted by the Working Group (see Appendix B), called Metadata for the Administration and Preservation of Digital

Images (MAPDI) was presented to NC ECHO in September 2002. While the Working Group was confident that the schema incorporated the basic information necessary to facilitate the most rudimentary management of digital images, it could not be sure if it had met its goal of accessibility to resource-challenged repositories. The study that follows set out to determine where North Carolina repositories are in their efforts to digitize collections, how they are incorporating preservation metadata (if at all), and to get their assessment of the MAPDI schema. Efforts were also made to expand the accompanying database tool to include the capture of collection-level Dublin Core metadata. In reviewing existing schemas, the Working Group was acutely aware of all that was omitted from the final MAPDI schema. The complexity and detail of a schema like METS shows just how much data it is possible to collect, but the Working Group was certain that METS-like detail and technicality was beyond the scope of most repositories. Following this, the author hypothesized that this survey would reveal knowledge of preservation metadata to be low and that repositories would still find the MAPDI schema too complex.

Methodology

The Survey

The survey was divided into two parts. The first part collected general information about the repository and its exposure to digitization and preservation metadata. Questions investigated the size of the repository, the number of employees, the size of their digitization projects (if any), and their awareness of preservation

metadata and metadata schemas. The second part asked them to review the MAPDI schema and provide their impressions and suggestions.[9] The MAPDI schema was available for download from the survey web site.

The author randomly selected twenty-five repositories from the online list of NC ECHO partner institutions (approximately 800 in all) to participate in this research. The only requirement for participation was the listing of an email address in the contact information, as organizations without email were determined to be unlikely candidates for digitization projects. NC ECHO was the ideal population sample because of its restriction to North Carolina and its specific mission of "[building] a statewide framework for digitization." Also of importance was NC ECHO's inclusion of repositories of all sizes, and the fact that MAPDI was created for NC ECHO member repositories. The small sample size was selected due to time constraints.

Each of the twenty-five repositories selected received an initial telephone invitation to participate in the study. The brief telephone conversation included a description of the study's purpose and format. Subjects were given the option of filling out the survey online or on paper. All subjects who agreed to participate selected the online option and received an email confirming their participation and providing the survey URL. A follow-up email was sent approximately one

---

[9] See Appendix A for survey questions.

week later to encourage those who had not yet completed the survey to do so. Since the survey was completely anonymous, the follow-up email was sent to all participants.

Results

Sixteen subjects gave verbal confirmation of their intention to participate. There were eight Part 1 surveys completed and eight Part 2 surveys completed (not everyone that filled out Part 1 completed Part 2 and vice versa). All surveys were filled out online. Not all respondents answered all the questions.

Clearly, digitization is an issue for the repositories surveyed. Of the eight respondents, 87.5% indicated plans to digitize or begin digitizing materials in the next year, and 62.5% indicated they had digitized materials in the past.

The size of the repositories was generally very small, with two notable exceptions (see Figure 2). Elimination of the two largest repositories, which were significantly larger than the others, yielded an average of 1.5 employees. While the study sample was too small to be statistically representative, Kevin Cherry, a former project manager for NC ECHO indicated that this coincided with his experience with NC ECHO repositories. Nonetheless, 50% (3) of the

Figure 2

How many employees are there in your repository?
3
1
1.25
4
15
1
1 working with preservation
45

respondents said they considered their repository to be medium-sized, in terms of resources. Thirty-three and a third percent (2) of the respondents identified their repository as small. This average size indicates that the sample coincides with the Working Group's earlier estimation of repository size distribution within NC ECHO. This also makes the sample well suited to review the MAPDI schema, as they generally match the repository demographic that was envisioned during the schema creation.

The size of the repository did not always correlate to the number of employees or the respondent's assessment of their repository's resources (Figure 3). This could be indicative of the different kinds of repositories that belong to NC ECHO, but it could also indicate that the survey questions needed to be more specific.

Figure 3

What is the size (in items, volumes, etc) of your repository?	**In terms of resources, do you consider yours to be a small, medium, or large repository?**	How many employees are there in your repository?
100,00	Medium	45
875 linear feet manuscripts; 13 linear feet photographs; 1000 artifact, textile, and oversize items; 800 reels of microfilm; 8500 titles (monographs and periodicals)	Small	1.25
Unknown at this time	Small	1 working with preservation
5,000,000 plus	Large	1
40,000	Medium	4
40,000	Medium	15
500-600 individual items including collections, University records, groups, and rare books. This number may be on the low side.	Small	1
Approximately 18,000	Medium	3

As expected, the respondents' knowledge of metadata schemas was not broad. MARC, Dublin Core, and EAD were the most widely recognized (Figure 4). No participants indicated that they had "heard of or used" METS, VRA, CEDARS, PANDORA, or the NISO Draft Data Dictionary[10]. Almost half of the respondents did not answer this question at all, leaving open the possibility that they had no knowledge of any of the schemas listed.

Figure 4

Which of the following metadata schemas have you heard of or used: (check all that apply)

	Response Percent	Response Total
Dublin Core	80%	4
METS	0%	0
VRA	0%	0
CEDARS	0%	0
Making of America 2	20%	1
NISO Draft Data Dictionary	0%	0
MARC	100%	5
EAD	60%	3
PANDORA	0%	0
Other (please specify)	0%	0

All but one of the respondents answered the question, "What does the term "preservation metadata" mean to you?" Several of the responses were vague or indicated uncertainty, others associated the term with a very specific resource type, such as "rare photographs and paper documents" (Figure 5).

[10] The NISO Data Dictionary is a technical metadata schema for digital images developed by the National Information Standards Organization and AIIM International. The full specification is available in PDF format at http://www.niso.org/standards/resources/Z39_87_trial_use.pdf.

Figure 5

	What does the term "preservation metadata" mean to you?
RESPONSES	A means of storing information, which had previously stored [*sic*] on paper, in an electronic format that can then be easily accessed and protected from physical damage.
	I see it as the metadata that is used in the creation of electronic finding aids. I am not sure if this is the exact meaning, just my understanding of it.
	Preservation metadata refers to the digital recording of fragile, distressed and rare photographs and paper documents.
	Well organized and searchable information (tagged according to set institutional standards) pertaining to the cataloging details of an item (provenance, extant, and subject information) that is in a machine readable, stable, and easily migratable format.
	Although there may be many different definitions, we are a repository of archaeological data which includes artifacts, site data (cultural and environmental), records of investigations, maps, reports, photographs, etc. Metadata concerns how the information contained in our repository was acquired, its accuracy and its origin. My main concern is with site files and the accompanying paper records, which will be the focus of my answers in this questionnaire.
	Detailed information concerning observations or classifications of preservation data
	Data about digital images that allows users in the future know how [*sic*] the digital images were created.

Respondents' analyses of the challenges facing metadata collection were varied. Respondents were asked to estimate how challenging a series of issues were for their repository, using a scale of 1 to 5, 5 being the most challenging. Time and Man-power received the most consensus, with 62% ranking these issues as a 5 and a 4, respectively (Figure 6), and both with an average ranking of 4.375.

Figure 6

There are many challenges to collecting metadata. Please rate the following challenges for your organization (select one rating for each challenge)
1 = Very Challenging, 5 = Least Challenging

	Very Challenging 5	4	3	2	1	Not a Challenge	Average ranking
Time	62% (5)	12% (1)	25% (2)	0% (0)	0% (0)	0% (0)	4.375
Training	0% (0)	50% (4)	38% (3)	12% (1)	0% (0)	0% (0)	3.375
Equipment	25% (2)	38% (3)	25% (2)	0% (0)	12% (1)	0% (0)	3.000
Man-power	38% (3)	62% (5)	0% (0)	0% (0)	0% (0)	0% (0)	4.375
Cost	38% (3)	38% (3)	25% (2)	0% (0)	0% (0)	0% (0)	4.125

Given the respondents' apparent lack of familiarity with schemas as indicated in question three, it is interesting that Training had the lowest average challenge rating at 3.375, and that none of the respondents ranked Training at the highest

challenge level (5). Nonetheless, 50% of the respondents did rate Training as a 4 on the challenge scale. Among those who ranked Training as a 4 (see Figure 7), Man-power had the highest average challenge level at 4.25, with Time and Equipment not far behind at an even 4. That is a substantial increase in the average ranking of Equipment, which had an average ranking of 3 across all respondents. This indicates the varied demographics even within resource-challenged institutions and could suggest a relationship between certain resource challenges. Even more interesting is the fact that Cost had the lowest average ranking. Given that the other four challenges can usually be met if funds are sufficient, then it is possible that repositories are not making that connection, or else that cost is somehow a different kind of challenge.

Figure 7

There are many challenges to collecting metadata. Please rate the following challenges for your organization (select one rating for each challenge)
1 = Very Challenging, 5 = Least Challenging

	Very Challenging 5	**4**	**3**	**2**	**1**	**Not a Challenge**	Average Rank
Time	50% (2)	0% (0)	50% (2)	0% (0)	0% (0)	0% (0)	4.00
Training	0% (0)	100% (4)	0% (0)	0% (0)	0% (0)	0% (0)	4.00
Equipment	50% (2)	0% (0)	50% (2)	0% (0)	0% (0)	0% (0)	4.00
Man-power	25% (1)	75% (3)	0% (0)	0% (0)	0% (0)	0% (0)	4.25
Cost	25% (1)	25% (1)	50% (2)	0% (0)	0% (0)	0% (0)	3.75

Budgetary commitment is a good indicator of an institution's focus on a given service. Acquisition and preservation typically receive line items in a library budget because they are recognized as vital core elements of the organization's mission. Clearly, digitization has not yet been institutionalized to this level: more than half (57.1%) of the respondents indicated that they did not have a digitization budget, either from their institution or from grants. One respondent

(14.3%) indicated a budget between $10,001 and $30,000, and the remainder of the respondents (28.6%) indicated a budget under $10,000.

The size of the collections that repositories have digitized or plan to digitize was fairly evenly distributed across the categories provided (see Figure 8). This suggests that tools for digitization projects probably will need to be scalable to accommodate a wide range of project sizes.

Figure 8

How many items (approximately) did you or will you digitize (select one):

	Response Percent	Response Total
<50	28.6%	2
50-100	14.3%	1
101-200	14.3%	1
201-500	0%	0
501-1000	14.3%	1
>1000	28.6%	2
Total Respondents 7		
(skipped this question: 1)		

The last question in Part 1 of the survey asked subjects how they had or would store their preservation metadata. The vast majority, 87.5%, indicated they used a database of some kind. The Working Group's decision to use MS Access appears validated by the fact that 66.67% of these respondents indicated the use of MS Access as their storage database. This does not necessarily signify the appropriateness of MS Access as a long-term storage solution, but does confirm its popularity and familiarity in repositories.

Part 2 – Review of MAPDI

Part 2 of the survey asked the participants to review the MAPDI model and then comment on their impressions. The model (see Appendix A) included not only a

list of the elements, but example data and a justification for the collection of that data for each element.

Assessment of the schema's complexity was divided between only two of the five possible answers. It is encouraging that the majority of respondents, 66.7%, rated the MAPDI schema as "just right." The remainder, 33.3%, rated the schema as "a little too complicated." While this bodes well for the overall appeal of the schema, the results are somewhat misleading given the lack of general metadata knowledge held by the respondents. Ideally, administrators would select the MAPDI schema as educated consumers, in full knowledge of what it does and does not include compared to other schemas. This would probably ensure greater long-term satisfaction with the schema. And indeed, the respondents were aware of this themselves, as evidenced by their responses to the question, "How does this schema compare to others you are familiar with?" (Figure9)

Figure 9

How does this schema compare to others you are familiar with?	
RESPONSES	I'm not familiar enough with any other schema or this one to compare.
	From what I can tell, it follows closely to the schemas I am familiar with and use.
	Don't feel I know enough to really compare at this point.
	More comprehensive. We are digitizing maps with several types of data on them, but no objects or art works.
	NA
	This is my first foray into this type of project. I have no basis for analysis.
	More technically detailed.

The majority of respondents were positive but noncommittal when asked if they would use the MAPDI schema for their preservation metadata gathering. This was attributed largely to the respondents' unfamiliarity with preservation metadata in general. Only one respondent (14.29%) indicated conclusively that

he or she would not use the schema, stating “it is a little too lengthy for our purposes and we already have an image database that provides a similar function.”

As anticipated, some of the more technical elements of the schema (color space, bit depth, and the elements of controls) were unrecognized by almost all the participants. This is discouraging, given that leading digitization handbooks [23, 3,] cover these concepts and their use. On the other hand, it lends support to the Working Group’s belief that advanced technical components could be a barrier to proper preservation metadata collection. All of the responses are available in Figure 10.

Figure 10

Which elements are you already familiar with (from other schemas, previous digitization work, research, etc)? (check all that apply)

	Response Percent	Response Total
Repository ID	75%	6
Item ID	100%	8
Alternate Item ID	87.5%	7
Item Title	100%	8
Collection ID	100%	8
Collection Title	87.5%	7
Source ID	87.5%	7
Source Type	87.5%	7
Source Creation Date	75%	6
Unit of Measurement	87.5%	7
Physical Dimensions of Source	75%	6
Physical Dimensions of Area Scanned	62.5%	5
Creation Date	100%	8
Digital Creator	62.5%	5
Capture Hardware	50%	4
Capture Hardware .accessories	37.5%	3
Capture Software	50%	4
Capture Software.settings	37.5%	3
Image Manipulation Software	50%	4
Image Manipulation Software.settings	37.5%	3
Resolution	75%	6
Compression	62.5%	5
Compression.type	37.5%	3
Compression.degree	12.5%	1
Dimensions of Digital Object	37.5%	3
Bit Depth	12.5%	1
Controls	25%	2
Controls.color bar/gray scale	25%	2
Controls.control target	12.5%	1
Color Space	25%	2
Watermark	62.5%	5
File Format	100%	8
Filename	100%	8
Digital Master	37.5%	3
Revision History	37.5%	3
Repository Copyright	50%	4
Standard Rights	37.5%	3
Total Respondents 8		

The second question of Part 2 of the survey asked for the participants to comment on which areas of the schema were unclear to them following their review of the model (Figure 11). All six of the participants who responded to the second question indicated uncertainty about the bit depth and color space elements; 83% indicated they were unclear on the definition of both these terms. The use of controls was the second most unfamiliar area, with 83% of the respondents indicating uncertainty on two or more of the defined levels (unclear on definition, unsure of how to implement, and unsure of purpose).

Figure 11

Which elements are unclear to you? (check all that apply)				
	Unsure of the Definition	**Unsure of the Purpose**	**Unsure How to Implement**	**Response Total**
Repository ID	0% (0)	0% (0)	0% (0)	0
Item ID	0% (0)	0% (0)	0% (0)	0
Alternate Item ID	0% (0)	0% (0)	0% (0)	0
Item Title	0% (0)	0% (0)	0% (0)	0
Collection ID	0% (0)	0% (0)	0% (0)	0
Collection Title	0% (0)	0% (0)	0% (0)	0
Source ID	0% (0)	0% (0)	0% (0)	0
Source Type	0% (0)	0% (0)	0% (0)	0
Source Creation Date	0% (0)	0% (0)	0% (0)	0
Unit of Measurement	0% (0)	100% (1)	100% (1)	1
Physical Dimensions of Source	100% (1)	0% (0)	0% (0)	1
Physical Dimensions of Area Scanned	0% (0)	0% (0)	0% (0)	0
Creation Date	0% (0)	0% (0)	0% (0)	0
Digital Creator	100% (2)	0% (0)	0% (0)	2
Capture Hardware	67% (2)	0% (0)	33% (1)	3
Capture Hardware.accessories	67% (2)	0% (0)	33% (1)	3
Capture Software	67% (2)	0% (0)	33% (1)	3
Capture Software.settings	67% (2)	0% (0)	33% (1)	3
Image Manipulation Software	100% (2)	0% (0)	0% (0)	2
Image Manipulation Software.settings	100% (2)	0% (0)	0% (0)	2
Resolution	0% (0)	50% (1)	50% (1)	2
Compression	0% (0)	50% (1)	50% (1)	2
Compression.type	0% (0)	50% (1)	100% (2)	2
Compression.degree	33% (1)	67% (2)	100% (3)	3
Dimensions of Digital Object	0% (0)	100% (2)	50% (1)	2
Bit Depth	83% (5)	67% (4)	50% (3)	6
Controls	80% (4)	60% (3)	80% (4)	5
Controls.color bar/gray scale	60% (3)	80% (4)	80% (4)	5
Controls.control target	80% (4)	60% (3)	80% (4)	5
Color Space	83% (5)	50% (3)	67% (4)	6
Watermark	50% (1)	50% (1)	100% (2)	2
File Format	0% (0)	0% (0)	0% (0)	0
Filename	0% (0)	0% (0)	0% (0)	0
Digital Master	100% (1)	100% (1)	100% (1)	1
Revision History	100% (2)	50% (1)	50% (1)	2
Repository Copyright	100% (2)	50% (1)	50% (1)	2
Standard Rights	100% (2)	50% (1)	50% (1)	2
Total Respondents 6 (skipped this question: 2)				

Unfortunately, only one respondent chose to answer the last question, which was an open-ended invitation for comments about the MAPDI schema. This is probably a flaw in the survey design and does not indicate a total lack of opinion on the schema; people probably do not want to extend the survey experience with extensive commentary so close to the end. Additional questions of a more specific nature would probably have obtained a better response rate.

.

Conclusions

The results suggest that preservation metadata is not being collected to the desired depth and breadth, and in fact, that not enough is known about preservation metadata by the people who should be collecting it. Although most repositories indicated plans for a digitization project, most had limited awareness of schemas in use in the industry. They also suggest that professionals are constrained as much by organizational policy as by resources. This further reinforces past studies [7,21] that more needs to be done to increase awareness of digital preservation issues, not just among practitioners, but in the ranks of upper management policy decision-makers. A larger study of more repositories, their metadata awareness, and their technical skill levels would provide more complete guidance for future digital preservation efforts.

The survey may not have fulfilled its full potential. When querying the respondents about their understanding of the elements in the schema (Part 2,

question 2), if they are unsure of the definition, it stands to reason they are probably unsure of the purpose of collecting that information and uncertain how to implement the collecting. For a more informed analysis of the MAPDI schema, Part 2 of the survey could have been restricted to only repositories that planned to digitize images.

Future Studies

Additional work needs to be done in the area of training and information distribution for digitization guidelines and preservation metadata standards. Such educational resources do exist, but they may not be reaching the audience that needs them most. North Carolina is not unique in the wide-ranging sizes of its repositories; small to medium repositories across the country may have similar training and resource needs. In light of recommendations for a centralized database of digitization efforts and best practices [21], future studies might examine where “lone arrangers” currently seek out information, the exact nature of the information they are searching for, how they would best like to access that information, and how they assess the feasibility of adopting guidelines and standards for their own repository.

Field testing of the accompanying MAPDI database application was beyond the scope of the current study, mainly because a thorough evaluation of the system would require at least several months of real-world evaluation. Extensive field testing of the MAPDI application is nonetheless necessary to determine if the

appropriate balance between simplicity and usefulness has been struck. Ideally, a handful of repositories of varying size (small to medium, to match the original target demographic) would adopt the application and test the ease with which they can alter the database for their purposes. It might be advisable to undertake a separate study of the application documentation (Appendix C) first. In this way problems with the documentation and training can be isolated from possible problems with the application.

Further development of the MAPDI database application would do well to include documentation for assisting repositories in building OAIS-compliant Archival Information Packages, Submission Information Packages, Dissemination Information Packages. While it would not be possible to anticipate the myriad combinations of legacy systems and data structures in use, more generalized guidelines on connecting to other databases and examples of using MAPDI and DC elements in SIPs, DIPs, and AIPs could be very helpful. Enabling organizations to make tangible step towards interoperability and information exchange within the OAIS network might provide valuable reinforcement for their efforts.

The MAPDI schema does not address metadata regarding digital authentication. For the time, authentication methods remain highly technical, such as checksums and digital watermarks, which are beyond the current means (financially and technically) of most small and medium repositories. Without special attention to

this inaccessibility, authenticating digital assets may prove impossible for thousands of items digitized in the past few years and near future.

Resources

1. *Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects*. 2002, OCLC Online Computer Library, Inc. Research Libraries Group. p. 51.
2. Tibbo, H.R., *On the Nature and Importance of Archiving in the Digital Age*. Advances in Computers. Vol. 57. 2003: Elsevier Science. 67.
3. Besser, H., *Section IX Digital Longevity*, in *Handbook for Digital Projects: A Management Tool for Preservation and Access*, NEDCC. p. 11 (printed from online).
4. Conway, P., *The Relevance of Preservation in a Digital World*, Northeast Document Conservation Center (NEDCC). p. 10 (printed from online).
5. *NC ECHO Official Website*. Retrieved from the World Wide Web: www.ncecho.org
6. Hedstrom, M. (1997). Digital preservation: a time bomb for Digital Libraries. *Computers and the Humanities, 31*(3), 189-202.
7. Day, M. *Issues and Approaches to Preservation Metadata*. in *Joint RLG and NPO Preservation Conference*: Research Libraries Group.
8. Day, M. *Metdata for Digital Preservation: A Review of Recent Developments*. in *5th European Conference on Research and Avanced Technology for Digital Libraries, ECDL 2001*. 2001. Darmstadt, Germany. Retrieved from the World Wide Web: http://www.ukoln.ac.uk/metadata/presentations/ecdl2001-day/paper.html
9. *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*, The Commission on Preservation and Access The Research Libraries Group.
10. *A Framework of Guidance for Building Good Digital Collections* (2001). IMLS. Retrieved, from the World Wide Web: http://www.imls.gov/scripts/text.cgi?/pubs/forumframework.htm
11. *The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials* (2002). [Online Guide]. Humanities Advanced Technology and Information Institute (HATII), University of Glasgow
12. Gilliland-Swetland, A., *Defining Metadata*, in *Introduction to Metadata: Pathways to Digital Information*, M. Baca, Editor. 1998, Getty Information Institute.
13. Waugh, A., et al. *Preserving Digital Information Forever*. in *Digital Libraries*. 2000. San Antonio, TX.

14. *Reference Model for an Open Archival Information System (OAIS)* (January 2002). Consultative Committee for Space Data Systems. Retrieved, from the World Wide Web: http://wwwclassic.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf

15. Reich, L. and D. Sawyer, *Draft Archiving Reference Model, Version 1*. 1995, Computer Sciences Corporation NASA.

16. *A Recommendation for Preservation Description Information* (2002, April 2002). The OCLC/RLG Working Group on Preservation Metadata. Retrieved, from the World Wide Web: http://www.oclc.org/research/pmwg/pres_desc_info.pdf

17. *Dublin Core Metadata Initiative Website*. Retrieved, from the World Wide Web: http://dublincore.org/index.shtml

18. *The Cedars Project Report* (2001). Cedars Project. Retrieved, 2003, from the World Wide Web: http://www.leeds.ac.uk/cedars/pubconf/papers/projectReports/cedarsrepmar01exec.html

19. Russel, K., Sergeant, D., Stone, A., Weinberger, E., & Day, M. (2000). Metadata for Digital Preservation. Cedars Project. Retrieved, from the World Wide Web: http://www.leeds.ac.uk/cedars/metadata.html

20. Searle, S., & Thompson, D. (2003). Preservation Metadata: Pragmatic First Steps at the National Library of New Zealand. D-Lib Magazine, 9(4), 9.

21. Building and Sustaining Digital Collections: Models for Libraries and Museums. (2001). Washington, DC: Council on Library and Information Resources.

22. Retrieved from the World Wide Web: http://www.cdl.ucop.edu/about/publications/CDLImageStd-2001.pdf

22. *California Digital Library Digital Image Format Standards* (2001). California Digital Library. Retrieved, from the World Wide Web: http://www.cdl.ucop.edu/about/publications/CDLImageStd-2001.pdf

23. Frey, F. S., & Reilly, J. M. (1999). Digital Imaging for Photographic Collections: Image Permanence Institute, Rochester Institute of Technology

Bibliography

Besser, H. Section IX Digital Longevity, *Handbook for Digital Projects: A Management Tool for Preservation and Access* (pp. 11 (printed from online)): NEDCC.

Building a National Strategy for Digital Preservation: Issues in Digital Media Archiving. (2002). Washington, DC: Council on Library and Information Resources, Library of Congress.

Building and Sustaining Digital Collections: Models for Libraries and Museums. (2001). Washington, DC: Council on Library and Information Resources.

California Digital Library Digital Image Format Standards (2001). California Digital Library. Retrieved, from the World Wide Web: http://www.cdl.ucop.edu/about/publications/CDLImageStd-2001.pdf

The Cedars Project Report (2001). Cedars Project. Retrieved, 2003, from the World Wide Web: http://www.leeds.ac.uk/cedars/pubconf/papers/projectReports/cedarsrepmar01exec.html

Conway, P. *The Relevance of Preservation in a Digital World*: Northeast Document Conservation Center (NEDCC).

Day, M. (199). *Issues and Approaches to Preservation Metadata.* Paper presented at the Joint RLG and NPO Preservation Conference.

Day, M. (2001). *Metdata for Digital Preservation: A Review of Recent Developments.* Paper presented at the 5th European Conference on Research and Avanced Technology for Digital Libraries, ECDL 2001, Darmstadt, Germany.

Dekkers, M., & Weibel, S. L. (2002). Dublin Core Metadata Initiative Progress Report and Workplan for 2002. *D-Lib Magazine, 8*(2), 11.

Dublin Core Metadata Initiative Website. Retrieved, from the World Wide Web: http://dublincore.org/index.shtml

Duff, W. M. (2001). Evaluating Metadata on a Metalevel. *Archival Science, 1*, 285-294.

Duval, E., Hodgins, W., Sutton, S., & Weibel, S. L. (2002). Metadata Principles and Practicalities. *D-Lib Magazine*(April 2002), 15.

EAD Tag Library for Version 1.0 (1998). Society of American Archivists Library of Congress. Retrieved, from the World Wide Web: http://lcweb.loc.gov/ead/tglib/tlhome.html

A Framework of Guidance for Building Good Digital Collections (2001). IMLS. Retrieved, from the World Wide Web: http://www.imls.gov/scripts/text.cgi?/pubs/forumframework.htm

Frey, F. S., & Reilly, J. M. (1999). *Digital Imaging for Photographic Collections*: Image Permanence Institute, Rochester Institute of Technology.

Gilliland-Swetland, A. (1998). Defining Metadata. In M. Baca (Ed.), *Introduction to Metadata: Pathways to Digital Information*: Getty Information Institute.

Greenstein, D., & Thorin, S. E. (2002). *The Digital Library: A Biography*. Washington, DC: Digital Library Federation, Council on Library and Information Resources.

Hedstrom, M. (1997). Digital preservation: a time bomb for Digital Libraries. *Computers and the Humanities, 31*(3), 189-202.

Lin, H. (2000). Fluency with Information Technology. *Government Information Quarterly, 17*(1), 69-76.

Lupovici, C., & Masanes, J. (2000). Metadata for Long-term Preservation. *NEDLIB*(1.0), 25.

Metadata for Digital Preservation: The Cedars Project Outline Specification Draft for Public Consultation (2002). Cedars Project. Retrieved, 2003, from the World Wide Web: http://www.leeds.ac.uk/cedars/documents/Metadata/cedars.html

Metadata Standards Framework: Preservation Metadata Schema (2000). National Library of New Zealand Te Puna Mätauranga o Aotearoa. Retrieved, from the World Wide Web: http://www.natlib.govt.nz/en/whatsnew/4initiatives.html#meta

METS: An Overview and Tutorial (2001). [Official Website]. Library of Congress. Retrieved, from the World Wide Web: www.loc.gov/standards/mets/METSOverview.html

NC ECHO Official Website. Retrieved, from the World Wide Web: http://www.ncecho.org/about.html

The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials (2002). [Online Guide]. Humanities Advanced Technology and Information Institute (HATII), University of Glasgow

National Initiative for a Networked Cultural Heritage. Retrieved, 2003, from the World Wide Web: http://www.nyu.edu/its/humanities//ninchguide/

Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects. (2002). OCLC Online Computer Library, Inc., Research Libraries Group.

Preserving Digital Information: Report of the Task Force on Archiving of Digital Information. The Commission on Preservation and Access, The Research Libraries Group.

A Recommendation for Preservation Description Information (2002, April 2002). The OCLC/RLG Working Group on Preservation Metadata. Retrieved, from the World Wide Web: http://www.oclc.org/research/pmwg/pres_desc_info.pdf

Reference Model for an Open Archival Information System (OAIS) (January 2002). Consultative Committee for Space Data Systems. Retrieved, from the World Wide Web: http://wwwclassic.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf

Reich, L., & Sawyer, D. (1995). *Draft Archiving Reference Model, Version 1* [URL]. Computer Sciences Corporation, NASA. Retrieved from the World Wide Web: http://ssdoo.gsfc.nasa.gov/nost/isoas/us01/p004.html

Rothenberg, J. (2000). *An Experiment in Using Emulation to Preserve Digital Publications*: The Koninklijke Bibliotheek, Den Haag.

Russel, K., Sergeant, D., Stone, A., Weinberger, E., & Day, M. (2000). *Metadata for Digital Preservation*. Cedars Project. Retrieved, from the World Wide Web: http://www.leeds.ac.uk/cedars/metadata.html

Searle, S., & Thompson, D. (2003). Preservation Metadata: Pragmatic First Steps at the National Library of New Zealand. *D-Lib Magazine, 9*(4), 9.

Tibbo, H. R. (2003). *On the Nature and Importance of Archiving in the Digital Age* (Vol. 57): Elsevier Science.

Waugh, A., Wilkinson, R., Hill, B., & Dell'oro, J. (2000, 2000). *Preserving Digital Information Forever.* Paper presented at the Digital Libraries, San Antonio, TX.

Appendix A – The Survey

This is Part I of the survey. All questions are optional and responses are confidential.

1) What does the term “preservation metadata” mean to you?

2) There are many challenges to collecting metadata. Please rate the following challenges for your organization. (select one rating for each challenge listed)

	Very Challenging 1	**2**	**3**	**4**	**Least Challenging 5**	**Not a Challenge**
Time	○	○	○	○	○	○
Training	○	○	○	○	○	○
Equipment	○	○	○	○	○	○
Man-power	○	○	○	○	○	○
Cost	○	○	○	○	○	○
Other	○	○	○	○	○	○
Please specify:						

3) Which of the following metadata schemas have you heard of or used: (check all that apply)

- □ Dublin Core
- □ METS
- □ VRA
- □ CEDARS
- □ Making of America 2
- □ NISO Draft Data Dictionary
- □ MARC
- □ EAD
- □ PANDORA
- □ Other (please specify) _____

4) What is the size (in items, volumes, etc.) of your repository? _____

5) How many employees are there in your repository? _____

6) What is your job title? _____

7) In terms of resources, do you consider yourself a small, medium, or large repository?

○ small ○ medium ○ large

8) Have you undertaken a digitization project in the past?

○ Yes ○ No

9) Do you plan to digitize (or begin digitizing) material within the next year?

○ Yes ○ No

If you answered <u>Yes</u> to question 7 or 8

10) Approximately how large is your digitization budget, either from the institution or through grants? (select one)

- ☐ Don't have one
- ☐ < $10,000
- ☐ $10,000-$30,000
- ☐ $30,001-$60,000
- ☐ $60,001-$100,000
- ☐ $100,001-$400,000
- ☐ >$400,000

11) How many items (approximately) did you or will you digitize (select one):

- ☐ < 50
- ☐ 50-100
- ☐ 101-200
- ☐ 201-500
- ☐ 501-1000
- ☐ >1000

12) Did you or will you capture any preservation metadata for the digitized materials?

○ Yes ○ No

If you answered <u>Yes</u> to question 11

13) Did you or will you adhere to a standard metadata scheme or create your own (either completely unique or a modification of existing schemas)? (select one)

☐ Use standard schema

☐ Our own schema – completely unique

☐ Modified existing schema(s)
Please specify
_____

14) Who was or will be entering the metadata? (check all that apply)

☐ student worker(s)

☐ volunteer

☐ staff with significant experience or training in cataloging or metadata

☐ staff lacking significant experience or training in cataloging or metadata

15) How did you or will you store the metadata? (check all that apply)

☐ Spreadsheet (e.g., MS Excel)

☐ Database (e.g., MS Access, FoxPro, Oracle, MySQL)
Please specify

_____

☐ Word processing (e.g., MS Word, Word Perfect)

☐ Other _____

Part 2
Please review the Metadata for Administration and Preservation of Digital Images (MAPDI) Schema and then answer the questions below. Attach additional pages if necessary. All questions are optional and responses are confidential.

1) Which elements are you already familiar with (from other schemas, previous digitization work, research, etc)? (check all that apply)

□ Repository ID	□ Image Manipulation Software
□ Item ID	□ Image Manipulation Software.settings
□ Alternate Item ID	□ Resolution
□ Item Title	□ Compression
□ Collection ID	□ Compression.type
□ Collection Title	□ Compression.degree
□ Source ID	□ Dimensions of Digital Object
□ Source Type	□ Bit Depth
□ Source Creation Date	□ Controls
□ Unit of Measurement	□ Controls.color bar/gray scale
□ Physical Dimensions of Source	□ Controls.control target
□ Physical Dimensions of Area Scanned	□ Color Space
□ Creation Date	□ Watermark
□ Digital Creator	□ File Format
□ Capture Hardware	□ Filename
□ Capture Hardware.accessories	□ Digital Master
□ Capture Software	□ Revision History
□ Capture Software.settings	□ Repository Copyright
	□ Standard Rights

2) Which elements are unclear to you? (check all that apply)

	Unsure of the Definition	Unsure of the Purpose	Unsure How to Implement
Field	□	□	□
Repository ID	□	□	□
Item ID	□	□	□
Alternate Item ID	□	□	□
Item Title	□	□	□
Collection ID	□	□	□
Collection Title	□	□	□
Source ID	□	□	□
Source Type	□	□	□
Source Creation Date	□	□	□
Unit of Measurement	□	□	□
Physical Dimensions of Source	□	□	□
Physical Dimensions of Area Scanned	□	□	□
Creation Date	□	□	□
Digital Creator	□	□	□
Capture Hardware	□	□	□
.accessories	□	□	□
Capture Software	□	□	□
.settings	□	□	□
Image Manipulation Software	□	□	□
.settings	□	□	□
Resolution	□	□	□
Compression	□	□	□
.type	□	□	□
.degree	□	□	□
Dimensions of Digital Object	□	□	□
Bit Depth	□	□	□
Controls	□	□	□
.color bar/gray scale	□	□	□
.control target	□	□	□
Color Space	□	□	□
Watermark	□	□	□
File Format	□	□	□
Filename	□	□	□
Digital Master	□	□	□
Revision History	□	□	□
Repository Copyright	□	□	□
Standard Rights	□	□	□

3) How does this schema compare to others you are familiar with?

4) How complicated is this schema, on a scale of 1 to 5? (select one)

Too Simplistic	Almost Enough	Just Right	A Little Too Complicated	Overwhelmingly Complex
○	○	○	○	○

5) Would you use the MAPDI schema for your preservation metadata? Why or why not?

6) Would you be more likely to use the MAPDI schema if it came with a database application for storing and reporting on the data? (select one)

Absolutely	Probably	Maybe	Probably Not	Definitely Not
○	○	○	○	○

7) Please provide any additional comments on the schema that you think would be helpful.

Appendix B – MAPDI Element Data Definitions

	Field	Definition	Example	Justification	Use*	Notes
Identifying the Repository	**Repository ID**	Unique identifier for the repository.	US-UNC-CH	For collaborative projects, this field ensures that items and collections will be uniquely identified across repositories.	S	NC ECHO participants are encouraged to use ISIL standard (ISO/DIS 15511). It is a 16-character variable length code to uniquely identify libraries and related organizations.
Identifying the Digital Image (IDI)	**Item ID**	The unique identifier for the digital object generated by the database system.	125	Database system generates this identifier.	R	Do not enter a value in this field. This number is automatically generated by the system (e.g., Access).
	Alternate Item ID	A unique identifier generated from another system used within the repository.	1711_imgAC	If the repository already has a system for uniquely identifying digital objects this will allow data linking between systems.	S	Only enter a value in this field if your repository has an established method for creating unique identifiers for digital images.
	Item Title	The natural language title of the digital object.	At schoolhouse in Sunburst 1911	The most descriptive way to identify an image and a strong retrieval key.	S	Title is generally taken from the back of the photograph when available. Indicate "None" if title cannot be identified.

* R = Required; S = Strongly Recommended; O = Optional

	Field	Definition	Example	Justification	Use*	Notes
	Collection ID	A unique identifier for the digital collection to which the digital image belongs.	Cameron 8912; MC 35	This ID is necessary to bring the items within a digital collection together within the repository's database.	R	Should be unique within the repository. Can be alphanumeric. Natural language (e.g., a title) is discouraged. Many digital images may belong to the same digital collection, but no two digital collections should have the same Collection ID.
IDI	**Collection Title**	The natural language title of the collection.	Carl Alwin Schenck Collection, 1890 – 1959	Most users and curators will identify the collection by this title.	O	Should be unique within the repository. No two collections may have exactly the same name.
Source Material	**Source ID**	The unique identifier for the source material from which the digital image was created.	468a.1 c1901	Links the digital image to its original, whether the original is digital or analog.	S	Could be a call number
	Source Type	Term that indicates the general format of the source.	Photograph	Provides clarification beyond the value "Image," which is recommended for use in Dublin Core when describing photographs.	S	The value for this field should be derived from the *Thesaurus of Graphic Materials* or similar standardized controlled vocabulary.
	Source Creation Date	The creation/publication date of the source object.	04/22/1955	Assists repositories to monitor when content enters the public domain.	S	January 1 of the copyright year is sufficient if the exact date is unknown.

* R = Required; S = Strongly Recommended; O = Optional

	Field	Definition	Example	Justification	Use*	Notes
	Unit of Measurement	The unit of measurement used for Physical Dimensions of Source and Physical Dimensions of Area Scanned.	cm		S	Best if repository reports data in a consistent fashion using either metric or non-metric measurements for all sources. Should always supply a value for Unit of Measurement if Physical Dimensions of Source or Physical Dimensions of Area Scanned are filled out.
	Physical Dimensions of Source	Height x Width.	7.25 x 4.25	Useful to researchers and allows comparison of size of digital object to its source object.	S	
Creation of the Digital Object (CDO)	**Physical Dimensions of Area Scanned**	Height x Width of the area of the source material that is actually scanned or photographed.	4 x 4	When presented with Physical Dimensions of Source, gives some sense of proportion of the digital image with respect to original. Mainly of value to the user.	S	Important for times when only a portion of the source item is scanned.
	Creation Date	Full date of creation of the digital object.	11/12/1997		R	
	Digital Creator	Creator (Individual) of the digital object.	Jane Smith		O	If the digital object was acquired from another repository, use the name of the creating organization.

* R = Required; S = Strongly Recommended; O = Optional

	Field	Definition	Example	Justification	Use*	Notes
	Capture Hardware	The hardware used to capture the digital image. Usually a scanner, but could also be a digital camera.	Scanner: UMAX Powerlook III; Digital Camera: Pentax 180R	Will assist in determining generation of item.	S	Provide make and model number whenever possible.
	.accessories	Any hardware accessories, such as a special digital camera lens, or lights used.	[none]	In digital photography, light source and lenses can be relevant to the digital image produced.	O	Provide make and model whenever possible. Qualifier of Capture Hardware.
	Capture Software	The name and version of the software used to capture the digital image.	MagicScan V4.4; HP Precision Scan		S	This is usually the scanning software provided with the scanner. This is not the software used to manipulate the image after capture such as Adobe Photoshop. Provide version whenever possible.
	.settings	Any settings used in the creation of the image, such as exposure, color balance, or resizing.	Sharp B & W		S	Qualifier of Capture Software
	Image Manipulation Software	The name and version of the software used to manipulate the digital image after capture.	Photoshop 7.0		*S*	Recommended when applicable.
Creation of the Digital Object	**.settings**	Any settings used in the manipulation of the image, such as exposure, color balance, or resizing.			S	Recommended when applicable.
	Resolution	Resolution of the final digital image, in dots per inch (dpi)	600	Strong quality indicator of image.	S	
	Compression	Yes/no field indicating whether or not digital image was compressed.	No		S	

* R = Required; S = Strongly Recommended; O = Optional

	Field	Definition	Example	Justification	Use*	Notes
	.type	Indicates type of compression, e.g., JPEG, LZW	JPEG	Indicates compression algorithm used, which may be important to future migration or display.	S	Qualifier of Compression
	.degree	Indicates the level of compression. For example, JPG compression may be represented in different image editors as "Compressed to 70%" or "Medium" or "8/12" (read as “8 out of 12”).	70%	Indicates compression algorithm used, which may be important to future migration or display.	S	Qualifier of Compression
	Dimensions of Digital Object	Height x Width in pixels	800 x 600	Important for complete display of image.	S	Indicates the size of the digital object relative to display settings.
	Bit Depth	The bit depth of the digital image. Standard values are 1 (black and white); 2-8 (grayscale); 24, 32, 48 (color).	24 bit	Quality indicator.	S	Some scanners capture “extra” bits – e.g., 10 bits for grayscale and 30 for color to allow for misregistration and scanner “noise.”
	Controls	Yes/No field indicating if controls to ensure color and size accuracy were used.	Yes		S	
	.color bar/gray scale	The color bar(s) or grayscale bar(s) used during image capture.	Kodak Q13 Color Separation Guide and Gray Scale.	Important to ensure tonal quality of the digital image and any future derivatives.	O	Provide make and model whenever possible. Qualifier of Controls.

* R = Required; S = Strongly Recommended; O = Optional

	Field	Definition	Example	Justification	Use*	Notes
Creation of the Digital Object	**.control target**	The control target(s) used during image capture.	AIIM Scanning Test Chart #2, RIT Alphanum. Resolution Test Object RT-1-71	Important to ensure resolution quality of the digital image and any future derivatives.	O	Provide make and model whenever possible.
	Color Space	Color space refers to the base palette of the image	RGB	Most images made for use in digital displays are in RGB. Images that are made for use in printing (brochures, ads, etc.) are usually in CMYK. Digital masters stored for print use may be in CMYK. Other options include YCbCr or CIE Lab.	O	Standard values are RGB or CMYK.
	Watermark	Yes/No field indicating use of a watermark in the digital object.	No	Watermarks can have repercussions for future use or migration.	O	Watermarks are embedded in digital images and identify them as belonging to the repository or collection. Watermarks may or may not be visible when viewing the image.

* R = Required; S = Strongly Recommended; O = Optional

	Field	Definition	Example	Justification	Use*	Notes
	File Format	The file format of the digital image.	TIF	While the file format often can be derived from the filename extension, providing it here allows for much faster searching and indexing within the database.	R	Use of [MIME] Internet Media Types from http://www.isi.edu/in-notes/iana/assignments/media-types/media-types recommended. Standard formats include JPG, TIF, and GIF.
	Filename	Filename of the digital object including file extension.	blea4ad2.jpg		R	Repository should establish file-naming protocols.
CDO	**Digital Master**	Yes/No field indicating whether or not the digital object is the digital master.	Yes	Digital masters should be identified for preservation purposes.	R	
Revisions	**Revision History**	Repeatable field for notating any changes to the digital object after its creation.	Cropped image to 790 x 583 to remove empty space.	This allows the repository to track changes to a single image over a long period of time. This could provide valuable information about migrations to other file formats, size changes, exposure changes, etc.	O	

* R = Required; S = Strongly Recommended; O = Optional

	Field	Definition	Example	Justification	Use*	Notes
Rights Management	**Repository Copyright**	Yes/No field indicating whether the repository own the copyright to the content of the digital object.	No	The creator of a digital image automatically owns the digital image, but not necessarily the copyright to the content of the image.	S	If the repository does not own the copyright to the content of a digital image, the content should either be in the public domain or the repository should have documented permission for the creation and display of that image as from a donor.
	Standard Rights	Checkbox that indicates digital image adheres to the repository's standard rights and distribution restriction policy.	Yes	Most digital images will fall under a standard rights and distribution restriction policy of the repository.	S	Check this box if the digital image adheres to this standard rights policy.

Appendix C – MAPDI Application Documentation

I. INTRODUCTION

MAPDI is designed to assist you in managing information about the digital images in your collections. The term “Administrative and Preservation Metadata” simply means information about the digital images that will be useful for long term maintenance of the collection, from an administrative and preservation point of view. Over time, technical advances, whether in the industry at large or simply an upgrade in your office, will necessitate changes to your digital objects, such as migration to a new digital format, or transfer to a new medium (e.g., from floppy disk to CD ROM). The goal is to meet the challenges of migration and refreshing without jeopardizing the aims of preservation. This system is intended to help you with those changes and preservation efforts by describing the technical aspects of the digital images in your repository.

This means that MAPDI is not intended as a catalog of your digital objects (notice that there is no place to describe the subject of the digital images), nor is it intended to replace any catalog, registration, or collection management software you may already have. Descriptive information is information about the subject of the image, such as a person’s name if the image were a picture of a person. This information has no bearing on your efforts to preserve the digital image itself, and that is why it is not captured in this system.

I.1 OVERVIEW

Within this system, you will have the ability to add, edit, and delete records describing your digital images. In most cases, one record represents one digital image. So if the instructions refer to locating a record, that means locating the record within the system that represents a particular digital image. You will also be able to run reports that will give you information about a specific image, or general information about an entire digital collection.

MAPDI was built for the express purpose of serving repositories whose resources may be extremely limited, either in actual funds or in technical training and support. Microsoft Access was selected for its ease of use, immense user support base, and availability to state employees, who support a large portion of the NC ECHO repositories. MAPDI was built using the bare essentials of MS Access. There are no modules and only basic VB code generated by the Wizards within MS Access. If that last sentence made no sense to you, you are in the majority of our target audience. Just know that we did not put in a lot of bells and whistles for you to get tangled up in if and when you decide you need to make changes to the system. That being said, this means that there are not a lot of bells and whistles that could make the system a little more efficient to use. This was the tradeoff we had to make. However, if you ever have someone available to you who is familiar with MS Access, or if you get a small grant to expand your technical infrastructure, there are many small but significant changes that could be made to enhance the system and we encourage you to do so (see the FAQs for more info on changing the system).

I.2 Before you Begin

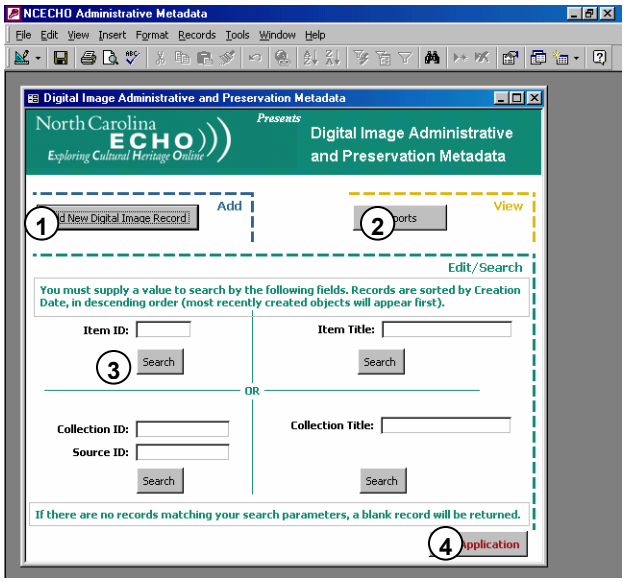
It is strongly recommended that each repository review this system and the documentation, and establish repository guidelines before using the system. By having your own metadata plan, you will be able to ensure the integrity of your data and its value in the future. The metadata plan should include things like controlled vocabulary terms and a list of the fields the repository needs to accurately preserve and administer the digital collections. Items like Repository ID, Collection ID, and Collection Title should have a repository-designated value so that all users of the system know the proper value to use, or how to find it. In addition, the repository should decide which fields it wants to require. For maximum flexibility, the database only truly requires five fields (specified in field list). The system will not allow you to complete a record without information in these five fields. However, there are many other fields that are strongly recommended for complete coverage and maximum value in the future. This

means that they are important, but not all repositories will have the technology or the training to provide all the fields, and so they are technically optional. As a repository, however, you can establish guidelines for your users mandating any fields that your work process can provide. For example, if your repository is doing its own scanning, then it should be no problem to provide the Capture Hardware and Capture Software information. If your items already have an official title, then by all means, require users to provide that data in the Item Title field. Because the system can not enforce use of the required fields that each repository decides upon, it is vital that repository management convey the importance of these fields and provide training on how to obtain the necessary information.

1.3 Installing the Database

The database file (the file with the .mdb extension which you obtained from NC ECHO) should be installed in a central location. If you plan to have multiple users using the system at the same time, or from different computers, you **must** have a computer network. Install the file in one location on the network, grant other users access to it, and then create a shortcut to that single file on each of machines from which your users will be entering data. **Do not** install the file on more than one machine. This will create separate, *unconnected* instances of the database. Your data will be divided amongst the different installations and they will not be able to talk to each other (at least, not easily, and it would require a knowledgeable network or MS Access technician).

II. Searching and Points of Access



The Main Screen

The Main Screen can be thought of like the home page of a web site. It is the first screen you will see when you open the application, and it is the access point for all functions.

1 – Add New Digital Image Record – Click this button to add a new digital image. A blank Object Screen will appear.

2 – Reports – Click this button to access the Reports Screen.

3 – Searching – The lower section of the Main Screen provides four different ways to search for a record or records with a certain characteristic. The different search methods are separated by teal lines.

The first method, searching by Item ID, uses the number automatically assigned to each record by the system. The Item ID field requires an exact match between

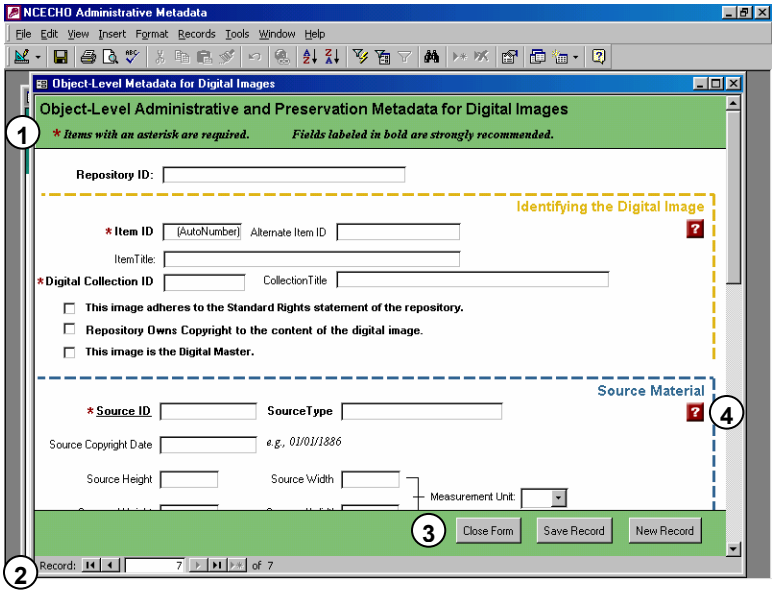
the search value you provide and the value stored in the system. For example, if you enter the number 3498, and there is no record where the Item ID equals 3498, you will get no results (i.e., a blank Objects screen). Even if you enter the number 3, and there are records with the Item ID value of 13 and 34 but not simply 3, the system will return no records.

The other search parameters - Item Title, Collection ID, Source ID, and Source Title – all use partial matching (usually referred to as wildcard searches). For example, if you are looking for all the images in a given collection and you know only that the word “house” appears somewhere in the Collection Title, you can enter “house,” and all the records that have the word “house” anywhere in the Collection Title will be returned. So images belonging to the “Houses of the Southeast” collection and images belonging to the “North Carolina House of Representatives” collection would both be returned. For even broader results, if you entered the letter “r” in the Item Title search field, you would get every digital image that has an “r” anywhere in the Image Title.

The Collection ID/Source ID search fields are paired together to allow you to search by the Collection ID or the Source ID or both. If you provide a search term for one field but not the other, the system will ignore the parameter you did not provide. However, if you provide a search term for both Collection ID and Source ID, the system will look for a record that matches both criteria.

4 – Exit Application – This button will completely close the application and exit MS Access. Any other windows within the application that are open (e.g., reports or search results) will be closed.

III. Entering and Editing Object-Level Data



The Object Screen

The screens for entering information and editing information look exactly the same. If you are adding a new record, the form will be empty when the screen appears. If you are editing an existing record, the form will be populated with the existing information. This screen is referred to as the Object Screen because it contains the majority of the information for a single digital object.

The Object Screen is broken up into four main areas: Identifying the Digital Image, Source Material, Creation of the Digital Image, and Revision History.

1 – Required Fields – Required fields are indicated with a red asterisk. All five required fields must be filled out in order to save the record or move on to another operation. The Item ID field has a red asterisk, but this value will always be supplied by the system, so you don't need to provide it.

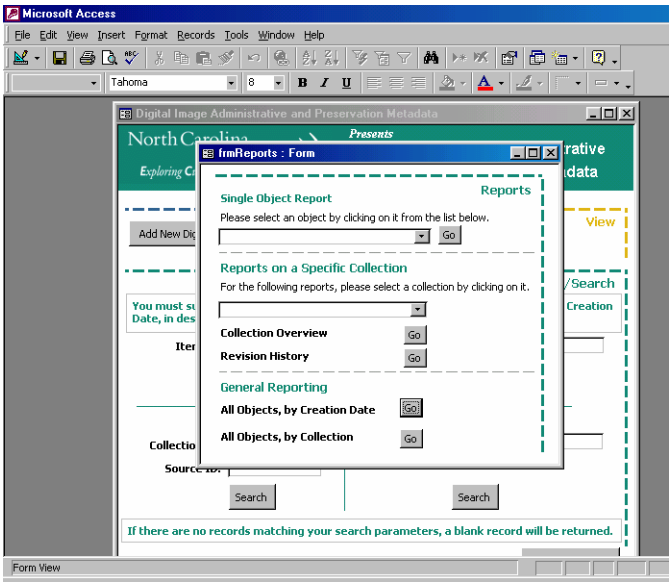
2 – Record Navigation – The Object Screen displays information for only one record at a time, so if you do a search that returns more than one record, you will need to navigate among the records. The buttons at the bottom of the screen enable you to do that. The number that appears in the white box tells you which record you are viewing out of the total number of records returned. In the picture above, we are on the 7th record out of 7 records returned by a search.

The ▸ and ▸| buttons take you to the next and last record, respectively. The |◂ and ◂ buttons take you to the previous and first record, respectively. The ▸| button opens a new (blank) record. If a button is "grayed out", that function is not available at that time. For example, if you do a search that returns more than one record, the |◂ button will appear gray when the Object Screen first loads because you are already on the first available record. The system automatically saves any changes you make to a record when you move on to another record, or when you close the screen.

3 – Function Buttons – The buttons at the lower right-hand side of the Object screen are provided simply as an added convenience. While it is hoped that their large size and prominent placement will prove convenient, each of their functions can be performed another way. As explained above, the Record Navigation buttons provide a way to create a new record. The Object screen can always be closed by clicking on the standard Windows ✕ box in the upper right-hand corner of the window. And any record will be saved (provided the required fields are not blank) when you move to a new record or close the Object Screen.

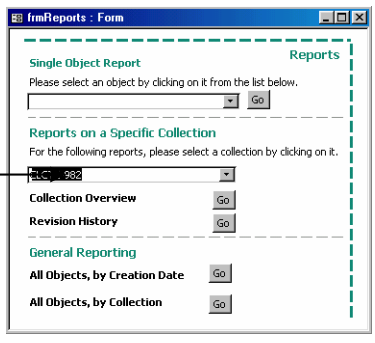
4 – Helpful Information –The  button, which appears at the top of each main section, is a link to additional information. A help screen will pop up with a description of each field in the section. You may also always refer to the documentation for additional help. There is a separate help screen for each section of the Object Screen.

IV. RUNNING REPORTS



The Reports Screen

The Reports Screen allows you to run the reports describing your data from the object level to collection summaries. A sample of each report is provided in Appendix A. When the screen first loads, the drop down boxes will appear blank. When you click on the ▼ button (at the end of each drop down box) the available options will appear. To view the report, simply click the Go button next to the desired report. For the reports that use a drop down box, if you do not make a selection from the drop down box, the report will not function. When you click on an option, that option will appear highlighted in the box, as in the image below.



Highlighted

V. Points to Remember

V.1 Garbage In, Garbage Out

This system can only be as good as the data that you put into it. It is important to remember when using this system that one of the main purposes of collecting this data is to be able to retrieve it later, usually searching for records that meet certain criteria. In order to get data that is reliable, all of the users of the system must adhere to the same guidelines and rules for entering information. The computer works at the level of 1's and 0's, Yes or No. It has no way of interpreting abbreviations or acronyms. So, if some of your users are entering "Southern Quilting Association" and some are entering "Southern Quilting Assoc." and still others are entering "SQA," when you want to find all of the images that are part of the Southern Quilting Association, you're going to have a very difficult time. That is why it is important that your repository establish a controlled vocabulary, or a list of acceptable terms for use in all fields that identify a collection or software. It does not matter what terms you use, as long as everyone using the system uses the same terms (this is where the metadata plan your repository developed before using the system comes into play). In this way you can be sure that your searches, and more specifically your data, accurately reflect your holdings. The phrase "garbage in, garbage out" is a popular phrase at computer help desks and refers to the fact that the computer can only return what you put into it. So if you want accurate data, you have to make sure you enter data accurately.

V.2 Backup Backup Backup

If you have ever lost a valuable document to a computer glitch, then maybe you don't need reminding. If you haven't, consider yourself one of the fortunate few, but heed this warning: if you do not back up your data, you *will* lose it. Computers malfunction for a variety of reasons, and all computers will fail at some point in time. It could be just software or it could be hardware. Either way, the component

that fails could be one that makes it impossible to recover your data. You won't get a warning or a grace period. It will just happen. Your data will simply be gone, vanished or damaged beyond recovery. So do yourself a favor and take five minutes on a regular basis to back up your data. It is only one file, and while it can get very large over time, the bigger it gets the more awful it would be if you lost everything, right? The process of backing up requires simply making a copy of the database file and placing that copy in a separate place. The separate place should be a storage medium such as a floppy disk or CD, but if that is not feasible, at the very least make sure that the backup and the working copy of the system (the one you actually use) are on different machines. How frequently you back up your data is up to you and is usually a compromise between the consequences of losing the data entered since the last backup if the computer failed, and the amount of time required for more frequent backups. If you are only using the system to enter a few records a week, you may be able to run the backups every other week. If you are using the system every day, or are entering dozens or more records per week, you may want to consider weekly backups. Whatever you decide, be diligent. The day you decide to skip the backup may be the day your computer comes down with a data-erasing virus.

It is also a good idea to make a backup of the database prior to any major changes, such as attempts to tweak with the inner workings, or compacting, as discussed below.

V.3 COMPACTING

One thing you can do to keep the size of your database file (the file with the .mdb extension which you obtained from NC ECHO) from growing wildly out of control is to compact it on a regular basis. Just using the database creates a certain amount of digital clutter within the system, which, in layman's terms, inflates the file size and eventually causes problems like slower performance or even data loss. Compacting is a feature of MS Access for the express purpose of keeping things tidy. You should follow the instructions outlined by MS Access in its

documentation or help files, but we will give you the short version here, just to show you how simple it is. First, make sure that no one else has the system open. Now make a backup of the database file, just in case. Next, open the system (i.e., your original file, not the backup). From the menu across the top, select Tools, then Database Utilities, then Compact and Repair. If everything goes smoothly, the system will “blink”, i.e., the Main Screen will disappear for a split second and then reappear. In the rare event that this process generates some sort of error, have someone from your IT support assist you, or simply revert to your backup.

VI. Frequently Asked Questions (FAQs)

Q: I want to make some changes to the way the system functions.
A: You are more than welcome to make changes to this system once you have installed it. That is one of the reasons the system was designed as simply as possible. However, please be aware that there is no technical support for this product. Be sure that you, or whoever is working on the system, has adequate technical knowledge to make those changes. And as always, make a backup before you begin making changes.

Q: Is there any way for the system to remember the values I entered for the last record, so I don't have to keep re-typing the same information?
A: Unfortunately, with this version that capability is not there. We recognize the value of such a feature and hope to include that in future versions of the system.

Q: What do I do if I don't have all the required information for a digital image record?
A: The system will not let you save a record without all of the required information. If you have already started on a new record and the system will not let you proceed because it keeps asking for the information you don't have, hit the Esc key. This will cancel the new record you have started.

Q: I entered some search information, but when I clicked the Search button, all I got was a blank Objects Screen.
A: This means that the system did not find any records that matched your search criteria. See the section in this documentation regarding searching under The Main Screen.

Q: I clicked on the Search button under Item ID and got a weird syntax error message.

A: You did not provide a number in the ItemID search parameter field. This field cannot be left blank for an Item ID search.

Q: I deleted an entry from the Revision History, and all the entries prior to that one disappeared. What happened?

A: When you delete a record from Revision History, the system rolls that selected record to the top of the screen and asks you to confirm that you really and truly want to delete that record. If you say yes, the record is deleted and the next available record moves to the top of the screen. All the other records are still there. Just hit the up arrow on your keyboard or use your mouse to scroll up and all the old records should reappear.

Q: How do I print a report?

A: There are a few different ways to print. 1) After running the report, right click on the report and you will see the Print option. 2) Select File from the MS Access menu at the very top of the screen. Then select Print. 3) Ctrl-P will also bring up the printing options.

Q: How do I locate version information for Capture Software or Image Manipulation Software?

A: Version information usually appears briefly when the program first loads. For Windows users, it also may appear under the Help menu as “About [the application name].” Since this information does not change very often, the name and version would be good information to include in your repository’s metadata plan.

Q: How did you decide which fields to have available in the system, and which ones to make required?

A: A group of archivists from North Carolina’s repositories and leading universities reviewed the predominant metadata standards in use by organizations capturing metadata around the world. Some of the schemas and organizations evaluated were:

California Digital Library
CEDARS
Colorado Digitization Project
Dublin Core
Harvard Univ. Digital Repository Services
JISC Image Digitization Initiative (JIDI)
Metadata Encoding and Transmission Standard (METS)
OCLC Framework (OAIS model)
Research Libraries Group
VRA Core Categories

For more information on the fields available, be sure to review the MAPDI Data Elements Definitions, which provides descriptions, justification, and examples.