

8-5-2002

A Metadata Approach to Preservation of Digital Resources: The University of North Texas Libraries' Experience

Daniel Gelaw Alemneh

Samantha K. Hastings

University of South Carolina - Columbia, hastings@sc.edu

Cathy Hartman

Follow this and additional works at: http://scholarcommons.sc.edu/libsci_facpub



Part of the [Library and Information Science Commons](#)

Publication Info

First Monday, Volume 7, Issue 8, 2002. <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/981/902> ©2002
First Monday

This Article is brought to you for free and open access by the Library and information Science, School of at Scholar Commons. It has been accepted for inclusion in Faculty Publications by an authorized administrator of Scholar Commons. For more information, please contact SCHOLARC@mailbox.sc.edu.

First Monday, Volume 7, Number 8 - 5 August 2002

- [Home](#)
- [About](#)
- [Log In](#)
- [Register](#)
- [Search](#)
- [Current](#)
- [Archives](#)
- [SUBMISSIONS](#)
-

[Home](#) > [Volume 7, Number 8 - 5 August 2002](#) > [Gelaw](#)



A Metadata Approach to Preservation
of Digital Resources:
The University of North Texas
Libraries' Experience

by Daniel Gelaw Alemneh,
Samantha Kelly Hastings,
and Cathy Nelson Hartman

Abstract

A Metadata Approach to Preservation of Digital Resources: The University of North Texas Libraries' Experience by Daniel Gelaw Alemneh, Samantha Kelly Hastings, and Cathy Nelson Hartman

Preserving long-term access to digital information resources is one of the key challenges facing libraries and information centers today. The University of North Texas (UNT) Libraries has entered into partnership agreements with federal and state agencies to ensure permanent storage and public access to a variety of government information sources. As digital resource preservation encompasses a wide variety of interrelated activities, the UNT Libraries are taking a phased approach to ensure the long-term access to its digital resources. Formulation of preservation policy and creation of preservation metadata for electronic files and digital collections are among the most important steps. This paper discusses the issues related to digital resources preservation and demonstrates the role of preservation metadata in facilitating the preservation activities in general. In particular, it describes the efforts being made by the UNT libraries to ensure the long-term access and preservation of various digital information resources.

Contents

[Introduction](#)
[Metadata for Digital Resources Management](#)
[Digital Initiatives at UNT Libraries](#)
[Preservation Metadata Requirement Analysis](#)
[Issues and Challenges](#)
[Lessons Learned](#)
[Conclusion](#)

FM
Volume 7, Number 8 - 5 August
2002

[Table of Contents](#)

Reading Tools

A Metadata Approa...

Gelaw, Hastings, Hartman

- [Abstract](#)
- [Review policy](#)
- [About the author](#)
- [How to cite item](#)
- [Indexing metadata](#)
- [Print version](#)
- Notify colleague*
- Email the author*

Related items

- [Author's work](#)
- [Government policy](#)
- [Book](#)
- [Book reviews](#)
- [Dissertations](#)
- [Online forums](#)
- [Quotations](#)
- [Resources](#)
- [Media reports](#)
- [Web search](#)

Search journal

All

[Close](#)

* Requires [registration](#)

Introduction

It is now common knowledge that digital information is fragile in ways that differ from traditional technologies, such as paper or microfilm. The fact that information is increasingly stored in digital form, has led to an accelerated search for effective methods of managing electronic information resources. The huge and ever expanding multiple sources of information on the Web normally contain special formatting and are produced with a variety of software in different versions. If the original digital resource is not "born digital", it may be a digital representation or digital surrogate of the physical medium, e.g. a page of text, an object, a painting, a photograph, a sculpture, a song, a movie, etc.

The persistence of digital information resources is an important factor for any digital library development. Addressing the preservation and long-term access issues for digital resources is one of the key challenges facing libraries and information centers today. In order to make sense of the high heterogeneity that exists among digital resources, a growing body of research has attempted to deal with the problems associated with the volume and nature of information on the Web and to look into ways to achieve consensus on a standard.

Metadata for Digital Resources Management

Metadata is a set of attributes used to describe an object. In reviewing the library and information science literature of the past few years, there is no shortage of views of the significant role of metadata in meeting the most pressing needs and challenges of digital resource management. A number of researchers (Moen, 2001; Waibel, 2001; Besser, 2000; Sutton, 1999; Zeng, 1999; among others), agreed that the underlying principle for metadata is to link and integrate heterogeneous, multi-platform, massive digital information collections that are contributed by different institutions into a single unified resource so these digital repositories are accessible by anyone, from anyplace, at anytime.

A number of metadata initiatives provide detailed and descriptive information about a digital resource to facilitate discovery by users. Resource description is essentially about describing information resources using a standard framework or set of principles. But because of the specific nature of heterogeneous digital resources, describing digital resources in a consistent fashion may not be an easy task, and in some cases, it is a complex process. Those concerned with digital information management all regard metadata as an essential component of the evolving networked information environment, but each of these communities view metadata with notably different perspectives.

Current Metadata Initiatives

Metadata standards come from various professional community efforts to support many needs in the digital environment. The literature reveals that different communities view metadata in significantly different contexts. The recent (2001) report from Research Libraries Group (RLG), comprised of key stakeholders from a variety of institutions, affirmed the fact that no single metadata standard can be expected to accommodate the needs of all communities. Although some projects, such as Dublin Core (DC) have tried to develop a coherent set of metadata schemes that can work for wide range of communities, they have not yet provided a complete description or solution for all types of digital information resources.

There is a great diversity of perspectives on various aspects of metadata issues. For instance, librarians have used machine-readable cataloguing (MARC) since the 1960's to identify, describe and provide access to their collections. However, what worked well for libraries may not work in other environments. Similarly, the basic metadata required for describing an image or work of art or non-text objects will bear a strong resemblance to the metadata that describes traditional print documents.

However, some significantly different extra elements will be required for a complete description of non-text images and multi-media resources. In light of this, some formats of metadata have been developed specifically for use in certain fields of study or type of information source.

Different communities have developed their own organizational and descriptive standards for accessing, arranging, and administering their specific digital collections, such as: the Medical Record Metadata for health professionals, Government Information Locator Service (GILS) metadata for describing government information resources, Visual Resources Association (VRA) Core Categories for describing visual information resources, and many more.

A number of commentators (e.g. Moen, 2001; Besser, 2000; and, Sutton, 1999) are optimistic that the core element set will be as minimal as possible. Thus, the core element set meanings will be easy to understand by most users and the element set will be flexible enough for description of diversified resources in a wide range of subject areas. Of course, the various previous efforts provide ways of describing digital resources to facilitate interoperability among resource discovery tools. Further developments in Resource Description Framework (RDF), Extensible Markup Language (XML), and Z39.50 may also provide means for integrating diverse metadata-based resources. For instance, the most recent work at the Library of Congress (LC) - Metadata Encoding and Transmission Standard (METS) schema - provides a flexible mechanism for encoding descriptive, administrative, and structural metadata for a digital library object, and for expressing the complex links between these various forms of metadata.

As indicated by Mullen (2001), most metadata initiatives have focused on resource discovery and to make it easier for people to find all of the information they need. Although such standards and structures are the most important steps in the development of the Web to avoid a chaotic repository of information, they do not guarantee continual long-term access to digital resources.

Preservation Metadata

For years, information centers preserved important electronic resources by transferring the files at regular intervals to the latest new information carriers available. As described by Besser (2001), refreshing a file involves periodically moving a file from one physical storage medium to another to avoid the physical decay or the obsolescence of that medium. Similarly, refreshing of files involves periodically moving files from one file encoding format to another that is usable in the current computing environment. But with multi-media digital resources (unlike in print) restoring in digital format may not be possible without the original software or hardware.

Preserving digital resources is made difficult by the fact that digital resources can only be read by software. This would mean that in order to ensure long-term access to digital resources, we need to preserve all the software, hardware, and operating systems on which the software ran. However, with the current quick obsolescence of information technologies, such an approach may not be feasible. Furthermore, inadequate media longevity is one of the issues. For instance, optical disks are expected to have a physical lifetime of up to 30 years but even a life expectancy of 30 years for storage media far exceeds the lifespan of hardware and software. Considering the ever-growing global Internet traffic, another problem is the mass of data and the need to compress it for efficient storage and transmission. However, compression sometime causes loss of data. It is also likely that repeated transfers over years from one carrier to another may cause data loss. This raises a number of issues including copyright, authenticity, and reliability.

Evidently, sustainable solutions to preserve digital resources are not yet available and are still being tested by various communities. Unlike the traditional notion of 'preservation,' (which refers to conservation and permanent preservation of the material or its information content), the ideal digital preservation activities would ensure that digital resources in all formats would be accessible as long as necessary. As described by Chapman (2001), if the objectives of digital preservation strategies were to preserve the artifact only, regardless of usability, longevity would be measured according to the lifespan of an object stored in a given environment. A number of researchers defined digital preservation in a variety of ways and present their views on how digital preservation might be achieved. According to RLG/OCLC's more specific definition, "Digital preservation refers to the series of managed activities necessary to ensure continued access to and preservation of digital materials."

It is clear that digital preservation is a critical issue, calling for measures that go beyond permanent archiving and all stakeholders agree that digital resource preservation encompasses a wide variety of interrelated activities. According to the RLG report, the problem of preserving digital sources is compounded by the fact that most of the sources do not have proper

descriptions. Similarly, Besser (2001) stated that such multi-format resources created by differing software require detail descriptions of the technical environment needed to view the digital resources.

Despite the fact that most metadata research gives more emphasis on resource discovery, a small breakthrough has been achieved in the last couple of years for preservation issues. A growing number of efforts to perfect the digital preservation methods by various organizations and agencies include: Reference Model for an Open Archival Information System (OAIS), CEDARS (CURL, Consortium of University Research Libraries, Exemplars in Digital Archives-UK), National Library of Australia (NLA), and RLG/OCLC (Research Library Group) to name but a few. These high-level preservation metadata initiatives provide much needed information required to manage the long-term preservation of digital resources.

As indicated by Besser (2000), preservation metadata is a strategy to provide sufficient technical information about the resources and to support the two primary strategies for preservation of digital resources, migration (transfer of digital resources from one generation to a subsequent generation) and emulation (developing techniques for imitating obsolete systems on future generations of computer.) Besser asserts that properly used metadata facilitate the long-term access of the digital resources by explaining the technical environment needed to view the work, including applications and version numbers needed, decompression schemes, other files that need to be linked to it, among others.

Digital Initiatives at UNT Libraries

During the past few years, the University of North Texas (UNT) Libraries, Government Documents Department made efforts to preserve various federal and state government information resources by forming partnerships with state and federal agencies. The various digital projects and undertakings in Government Documents Department include the *Cybercemetery*, the *Nineteenth Century Texas Law Online*, and the *Texas Register* project.

Figure 1 depicts one of the UNT Libraries' Digital Projects Page available at <http://govinfo.library.unt.edu/>. As the name *Cybercemetery* indicates, the UNT Libraries collect the digital publications from "deceased" federal agencies and preserve them for current and future public access. Furthermore, various digitization undertakings at the Government Documents Department made available selected local resources such as the *Texas Criminal Justice Statistical Reports*. These, together with the *Texas Register* and the *Nineteenth Century Texas Law Online* projects can be cited as part of the UNT Libraries, Government Documents Department's initiatives to preserve state historical and agency publications. The UNT Libraries are also currently involved in a project with the Texas State Library and Archives Commission to preserve current state electronic publications.



Figure 1. Cybercemetery Digital Project Page: preserving "deceased" U.S. government digital publications.

Access, Use, and Preservation

In complying with the fundamental principle of free public access to government information resources, the goal of the UNT Government Documents Department's digital projects and initiatives is to make government information resources accessible for exploration and search by anyone, from anywhere, at anytime.

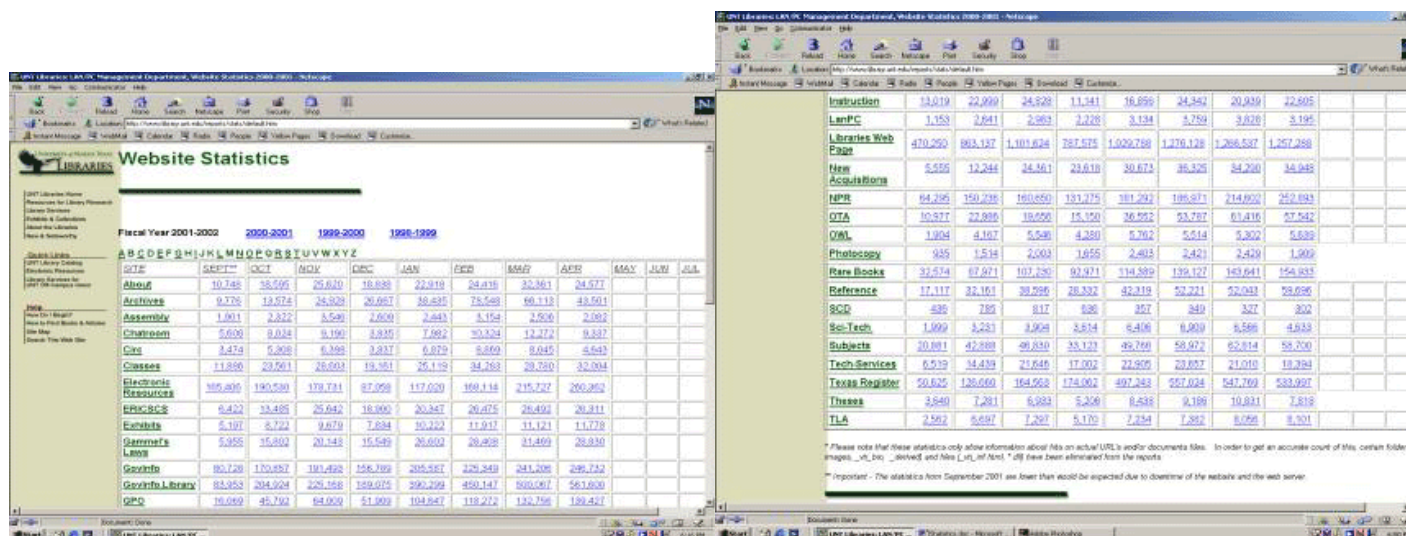


Figure 2: UNT Libraries Overall Web Usage Statistics (<http://www.library.unt.edu/reports/stats/default.htm>)

As can be seen from [Figure 2](#) above, each digital resource has its own user statistics. If we scrutinize the usage statistics very closely, there is a nearly linear increase in the monthly percentage of users for the Government Information Connection digital collections (*Gammel's Laws, GovInfo, GPO, NPR, OTA, Texas Register*). For instance, *Texas Register* is one of the UNT libraries popular electronic resources that became available online in 2001. Since then, the user base has grown dramatically, with the current number of hits more than sixteen times the hits in the early months. Usage grew from just over 30,000 (in January, 2001) to over 530,000 (in April 2002). Each digital information collection attracted a lot of users from all over the world. As the Libraries' digital collections grew, the need to address the issue of preserving long-term access to these resources became evident.

In light of this, the UNT Libraries are taking a phased approach to prolonging the usable life of the libraries' digital resources. Formulation of preservation policy and creation of preservation metadata for electronic files and digital collections are among the most important steps in UNT Libraries' preservation initiatives. In view of that, the Digitization Workgroup was charged to recommend a plan that will ensure long-term future access to the UNT Libraries' electronic information resources. The Workgroup is reviewing the different types of the UNT Libraries electronic resources. These include:

- State records;
- The online library catalog system;
- Personal and shared directories;
- The Libraries' Web site; and,
- The Libraries' digital collections.

Preservation Metadata Requirement Analysis

Metadata is a key factor for ensuring the long-term access of digital resources. There is a continuous need for extending the existing metadata element set to be able to describe all available digital resources. During the past two years, the UNT libraries reviewed several metadata initiatives to build an element set appropriate for its digital collections while monitoring the RLG/OCLC efforts toward building a standard metadata element set.

In addressing the issues of identifying specific metadata requirements, UNT Libraries attempted to assess the specific characteristics of the existing digital resources. In the preliminary needs assessment, the following issues, among others, were considered:

- Specific creation features and production life cycle of the digital information resources: Structural Type (Text, image, Audio, Video etc.), Integrity Issues, etc.;
- Users' information seeking behavior: [who, what, how ...];
- UNT's objectives [to ensure longevity]; and,
- Current standards and future trends [Interoperability (mapping) with current best practices and international standards plus complying with federal and state requirements].

Life Cycle Assessment of the Digital Resources

As indicated by the National Library of Australia (NLA) report (1999), to manage digital collections or individual items one needs to have a clear understanding of one's digital collection. Documentation has always played a key role in preservation practice and there are many instances where documentation provided the only information about processes and changes that had been applied and might need to be corrected.

In this regard, all available digital resources creation manuals, guidelines, and reports at the Government Documents Department were reviewed and modified accordingly. Sample documents, (Texas Register digital collection creation processes report and ACIR procedure manual) can be viewed at <http://texinfo.library.unt.edu/texasregister/text/report/TX-Reg-Report-2002.doc> and <http://www.library.unt.edu/gpo/ACIR/document/ACIR-procedure.doc>, respectively. Those documents provide detailed information about the creation history and complete life cycle of the digital resources. The preliminary resource assessment and evaluations assisted us in identifying the specific characteristics and requirements of the available digital resources.

Based on the thorough assessment of the available digital resources, attempts have been made to review current best practices and standards to represent a range of relevant fields. The review pays particular attention to the preservation and management metadata sets, which are needed to support various preservation approaches including migration and emulation.

The work at NLA developed a practical model for dealing with the immediate threat of disappearing digital objects, and established a workable distributed archive. Similarly, a number of projects and researches - such as OAIS (Open Archival Information System), CEDARS (CURL Exemplars in Digital Archives), NEDLIB (Networked European Deposit Library), and others - have investigated options for dealing with long-term preservation challenges.

Based on the preliminary survey of the existing digital collection and a detailed review of current best practices, we chose to base our recommendation of preservation metadata on a synthesis of various preservation metadata until the OCLC/RLG (2001) completes a national standard.

The Draft Metadata Architecture

The extensive literature review revealed that effective metadata is our best way of minimizing the risk of digital resources becoming inaccessible. Metadata, to be most valuable, both for the users and owners, needs to be consistently maintained throughout the process. Creating documentation that governs and informs the metadata creation steps and procedures in a consistent and uniform manner is among the most important steps in metadata creation. The detailed workflow and user guide document provides procedural information required to create metadata with examples for different file formats. Since the metadata assigned to an item entirely depends on the metadata creators' definition of the work, the detailed user guide also provides rules, syntax and descriptive information to identify the source of information for each element.

The following chart ([Figure 3](#)) illustrates the basic structure of the UNT Libraries draft preservation metadata contents. A detailed description of the recommended preservation metadata elements can be found in [Appendix II](#) of this paper.

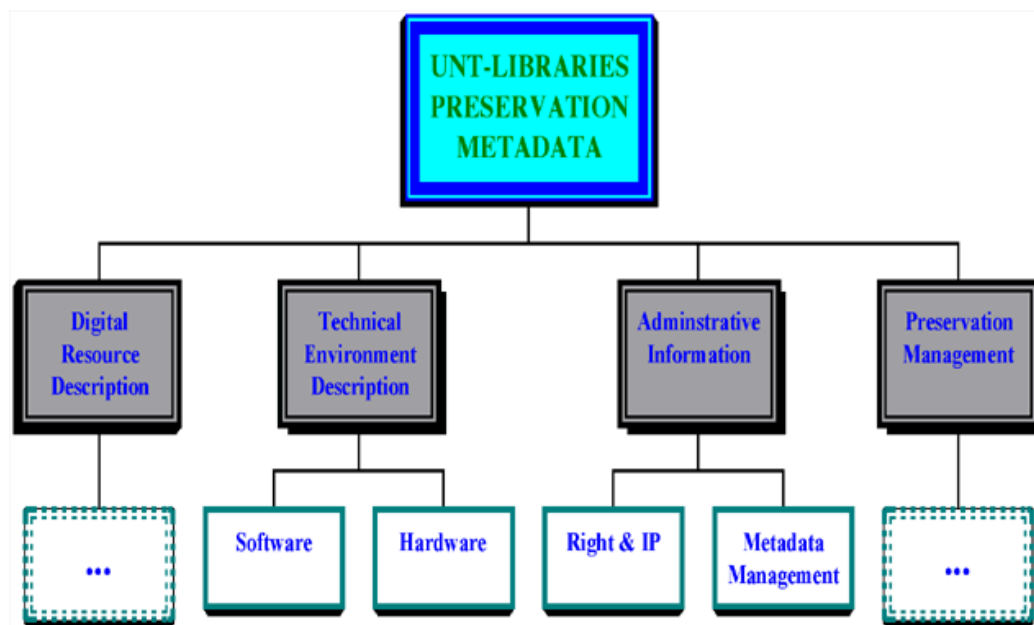


Figure 3. UNT Libraries' Preservation Metadata Structure.

The following table ([Table 1](#)) describes the subheadings of each metadata elements.

Table 1: Preservation Metadata Elements' Subheadings.

Name of Sub-Heading	Description	Remark
Element Name	Name of the element	The element name may not be identical to the name of the origin
Sub-element	Indicate existence of sub-elements, labels	
Origin(s)	Source of the element	Mapping to other metadata standards will start soon
Definition	Further explanation for clarification of the purpose of the element.	Further described in "Description" and "Comment" subheadings
Description	A brief statement that defines the concept of the category	
Required	Indicate whether the element value is mandatory or optional	Yes/No
Repeatable	Indicate if the element is repeatable or not repeatable	Yes/No
Example	Local examples	
Comment	Notes to clarify exceptions	

As can be seen from the sample element description in [Table 2](#) below, each of the identified metadata elements are described under separate subheadings. For a complete list of recommended elements, see also [Appendix II](#).

Table 2: Sample Preservation Metadata Element Description.

Element Name	Access Inhibitors
Origin	NLA
Sub-element	
Definition	Description of any features of the digital resources intended to inhibit access
Purpose Definition	Without this information the DR may not be able to be accessed, copied or migrated
Required	No
Repeatable	Yes
Example	Encryption, watermarking, digital signature, password protection, etc.
Comment	This information may be placed in the Documentation linked to the DR

Metadata Creation Workflow

Most preservation metadata project managers acknowledged that the best practice is to create the metadata at the information creation stage. Hodge (2000) recognized that creation is where long-term archiving and preservation must start. The metadata routinely collected at the point of creation would be relatively easy, consistent, reliable, and automatic. Of course, the preservation and archiving process is made more efficient when the creators provide an indication of the long-term value attached to the information resources. More importantly, attention would be paid to issues of consistency in the process of metadata creation in the very beginning of the information life cycle.

Much of the preservation metadata continues to be created "by hand" and after-the-fact. This problem is coupled with the fact that metadata creation is not sufficiently incorporated into the tools for the creation of elements' record to rely solely on the creation process. As standards groups and vendors move to incorporate XML and RDF architectures in their word processing and database products, the creation of metadata as part of the origination of the object will be easier.

The following diagram illustrates the logical steps in creating metadata tags for digital resources in general. As can be seen in [Figure 4](#), metadata can be incorporated into the digital resources (step 3-1), and/or can be stored in repositories separate from the resources it describes (step 3-2). When the metadata have been saved in their appropriate location, the process of metadata creation is considered to be complete.

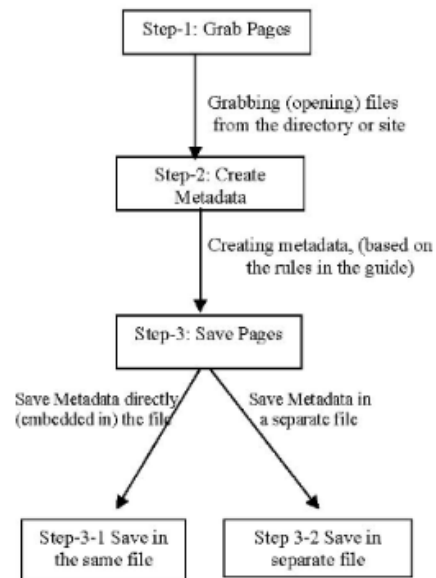


Figure 4. Metadata Creation Steps.

Metadata Creation and Editing Tool

There are various metadata creation tools and wizards available. For the purpose of testing and demonstrating our prototype, we selected the NoteTab Light program (<http://www.notetab.com>).

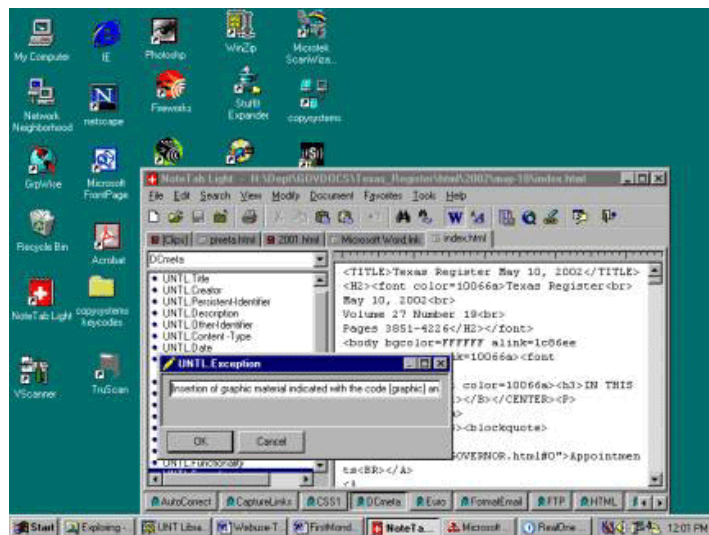


Figure 5. Customized NoteTab Light Metadata Creation Tool.

This freeware version of NoteTab metadata creation tool allows us to add and modify metadata elements and also copy metadata values either to be embedded in resources or to maintain in a different repository. This tool reduces the need for editors and data enterers to learn the syntax of the metadata.

Issues and Challenges

We are just at the beginning stages. We plan to develop a prototype for experiments and demonstrate the feasibility of preservation metadata at the UNT Libraries. Our initial prototype testing will be limited to the Government Documents' digital resources. The files and Web pages will be modified to include the recommended metadata elements.

Technical Issues

During the life cycle of digital resources, there are a series of processes that require various sets of hardware and software infrastructures. Similarly, as described in the metadata creation workflow, there is a series of managed activities that determine the appropriate hardware and software technologies to be used at each step of the preservation metadata creation process. These include:

- identifying the appropriate metadata creation tool,
- appropriate means for creating a metadata repository database,
- appropriate indexing and harvesting software and search engines to use,
- designing several interfaces for field searches and related considerations.

Lessons Learned


A pretest activity will allow us to determine the resources required to implement a comprehensive preservation metadata project. Due to issues of cost, compliance, and heterogeneity, we found the fewer elements that are required the better (provided that such minimum mandatory metadata elements would not have any consequences on the preservation activities). In addition, the quality control analysis involves various levels of assessment, including examining the metadata records for consistency, reliability, adequacy, etc. We plan to work on a detailed analysis of the costs and benefits for the recommended metadata elements, including the amount of time, and the level of skills required to create and manage a successful preservation metadata system. We have created a set of questions that provide scenarios about the issues and challenges of implementing the recommended preservation metadata system. These include:

- Is the preservation metadata system easy to use?
- Does the User Guide include a clear set of rules?
- Is it feasible to develop controlled vocabulary lists from the many files on the libraries' server to represent content and do so adequately?
- Is it feasible to consider creating the default values in the metadata creation tool for some of the mandatory fields?
- Is the preservation metadata system supported by the existing UNT Libraries' search engine topology? (Adaptability of existing schema?)
- Is the preservation metadata interoperable with current and future international standards? (Semantic, structural, and

syntactical mapping issues?)

Conclusion

Like many others, UNT Libraries realize that being digital does not mean being accessible. Access to digital resources through descriptive metadata is only short-term. Preservation metadata plays a significant role in facilitating preservation decisions, detects preservation threats and provides measures for minimizing risks to long-term access. We anticipate that the management, storage and serving of large datasets will be greatly improved by the use of preservation metadata management tools.

Finally, we will evaluate and assess the practical application of the whole process of metadata creation workflow and user guide documents. We expect a tremendous amount of discussion from all stakeholders regarding the types of metadata elements most useful to a specific requirement. Based on the feedback and input from the field, the preliminary versions will be reviewed and modified. Of course, the real test will be in the efficiency of our first migration. 

About the Authors

Daniel Gelaw Alemneh is currently a doctoral student in information science, with a digital imaging specialty, at the University of North Texas (UNT). He is an IMLS fellow from Ethiopia, and received a Post-Master's Certification in Digital Image Management from the UNT in August 2000. Prior to that, he earned his Master's Degree in Library and Information Management from the University of Sheffield, U.K. Mr. Alemneh is employed as a Super Graduate Library Assistant in the Government Documents Department, and works on various digitization projects.

E-mail: dalemneh@library.unt.edu

Dr. S.K. Hastings joined the faculty at the UNT in 1995. She is very active in state and national professional associations. Dr. Hastings has served as a resource person and presented a number of papers at various professional meetings and conference programs including Curricula Development for Multimedia Librarians, Standards for Museum Information Managers and Index Access Points in the Retrieval of Digital Art Images, the changing role of information professionals/digital managers. Dr. Hastings continues to research problems associated with the access, retrieval, and preservation of digital images, with particular emphasis on designing information communities for the 3D environment. She is principal investigator for a federally funded IMLS Library- Museum-University Collaboration project. You may view her various projects at

<http://www.courses.unt.edu/shastings/>.

E-mail: hastings@lis.admin.unt.edu

Cathy Nelson Hartman is the head of the Government Documents Departments at the UNT Libraries. She has been very active in state and national professional associations and serving as chairperson for a number of work groups, committees, and taskforces at state and national level. In addition, she has served as a resource person and presented a number of papers in various professional meetings including Computer in Libraries, ALA, Texas Library Association, and Depository Library Conferences. She has published a number of articles on digital resource management issues. Ms. Hartman is a successful grant recipient and she is project manager for several digital projects including the *Cybercemetery*, the *Texas Register*, and others. You may also view her various projects at <http://www.library.unt.edu/govinfo/>.

E-mail: chartman@library.unt.edu

References

- H. Besser, 2001. "Digital preservation of moving image material?" Accepted for publication in *The Moving Image*, at <http://www.gseis.ucla.edu/~howard/Papers/amia-longevity.html>, accessed 16 November 2001.
- H. Besser, 2000. "Digital longevity," In: *Handbook for Digital Projects: a Management Tool for Preservation and Access*. Andover, Mass.: Northeast Document Conservation Center.
- S. Chapman, 2001. "What is digital preservation?" *OCLC Symposium: Digital past, digital future: An Introduction to digital preservation*, at <http://www.oclc.org/events/presentations/symposium/chapman.shtm>, accessed 9 May 2002.
- G. Hodge, 2000. "Best practices for digital archiving: an information life cycle approach," *D-Lib Magazine*, volume 6, number 1, at <http://www.dlib.org/dlib/january00/01hodge.html>, accessed 18 January 2002.
- W. Moen, 2001. "The Metadata approach to accessing government information," [Electronic version], *Government Information Quarterly*, volume 18, pp. 155-165.
- A. Mullen, 2001. "GILS metadata initiatives at the state level," *Government Information Quarterly*, volume 18, pp. 167-180.
- National Library of Australia, 1999. *Preservation metadata for digital collections*. Exposure draft, at <http://www.nla.gov.au/preserve/pmeta.html>, accessed 24 October 2001.
- OCLC/RLG, 2001. *Preservation metadata for digital objects: A Review of the state of the art*. A White Paper by the OCLC/RLG Working Group on Preservation Metadata, at http://www.oclc.org/digitalpreservation/presmeta_wp.pdf, accessed 14 January 2002.
- Reference Model for an Open Archival Information System (OAIS), n.d. "Draft Recommendation for Space Data System Standards," at <http://www.ccsds.org/RP9905/650x0r1.pdf>, accessed 26 January 2002.
- S. Sutton, 1999. "Conceptual design and deployment of a metadata framework for educational resources on the Internet," *Journal of the American Society for Information Science*, volume 50, pp. 1182-1192.
- G. Waibel, 2001. "Produce, publish and preserve: A Holistic approach to digital assets management," at <http://www.bampfa.berkeley.edu/moac/imaging/index.html>, accessed 19 September 2001.
- M. Zeng, 1999. "Metadata elements for object description and representation: A Case report from a digitized historical fashion collection project," *Journal of the American Society for Information Science*, volume 50, pp. 1193-1208.

Appendix 1. Acronyms used in this paper, with selected Web addresses

- | | |
|------|---|
| AACR | <p>Anglo-American Cataloguing Rules.
 http://www.ala.org/editions/updates/aacr2/</p> <p>The AACR second edition, 1988 revision (AACR2) is used in the preparation of bibliographic records by OCLC participants as well as by most libraries in the United States. Requests for changes in the rules go to the American Library Association (ALA), Association for Library Collections and Technical Services (ALCTS), Committee on Cataloging: Description and Access (CC:DA). CC:DA submits proposals for changes in the rules to the Joint Steering Committee for Revision of AACR (JSC). This international body, after appropriate consultation with all countries involved, issues changes to the rules.</p> |
| ACIR | <p>Advisory Commission on Intergovernmental Relations.
 http://www.library.unt.edu/gpo/ACIR/</p> <p>The ACIR is a permanent, independent, bipartisan agency that was established under U.S. Public Law 86-380 in 1959 to study and consider the federal government's intergovernmental relationships and the nation's intergovernmental machination.</p> |

AMICO	<p>Art Museum Image Consortium. http://www.amn.org/AMICO/</p> <p>The AMICO is a not-for-profit organization of institutions with collections of art, collaborating to enable educational use of museum multimedia.</p>
ANR	<p>Access to Network Resources. http://ukoln.ac.uk/elib/lists/anr.html</p> <p>ANR is part of the Electronic Libraries Programme (eLib), which established by the Joint Information Systems Committee (JISC) (U.K.). The main aim of the eLib programme, through its projects, is to engage the Higher Education community in developing and shaping the implementation of the electronic library.</p>
ANSI	<p>American National Standards Institute. http://www.ansi.org/default.asp</p> <p>ANSI is a private, non-profit organization founded on 18 October 1918. The Institute's mission is to enhance both the global competitiveness of U.S. business and the U.S. quality of life by promoting and facilitating voluntary consensus standards and conformity assessment systems, and safeguarding their integrity.</p>
CAMiLEON	<p>Creative Archiving at Michigan and Leeds: Emulating the Old on the New. http://www.si.umich.edu/CAMiLEON/</p> <p>CAMiLEON is a research project that is investigating emulation as a digital preservation strategy. The project is a collaborative effort of researchers at the School of Information, University of Michigan (USA) and the University of Leeds (U.K.).</p>
CEDARS	<p>CURL Exemplars in Digital Archives, U.K. http://www.leeds.ac.uk/cedars/</p> <p>CEDARS began in April 1998 and ended in March 2002. Its broad objective was to explore digital preservation issues. These range through acquiring digital objects, their long-term retention, sufficient description, and eventual access.</p>
CDWA	<p>Categories for the Description of Works of Art http://www.getty.edu/gri/standard/cdwa</p> <p>CDWA is a product of the Art Information Task Force (AITF), which encouraged dialog between art historians, art information professionals, and information providers so that together they could develop guidelines for describing works of art, architecture, groups of objects, and visual and textual surrogates. The Categories describe the content of art databases by articulating a conceptual framework for describing and accessing information about objects and images. They also provide a framework to which existing art information systems can be mapped and upon which new systems can be developed.</p>
CIMI	<p>Computer Interchange of Museum Information http://www.cimi.org</p> <p>CIMI is a consortium of cultural heritage institutions and organizations that work together to bring rich cultural information to the widest possible audience.</p>
CURL	<p>Consortium of University Research Libraries, U.K. http://www.curl.ac.uk/</p> <p>CURL's main objective is to promote, maintain and improve library resources for research, learning and teaching in research-led universities in U.K.</p>
DC	<p>Dublin Core http://purl.oclc.org/dc/</p> <p>The DC Metadata Initiative is an open forum engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models. Its activities include consensus-driven working groups, global workshops, conferences, standards liaison, and educational efforts to promote widespread acceptance of metadata standards and practices. See also OCLC.</p>
DOI	<p>Digital Object Identifier http://www.doi.org/</p> <p>The DOI is a system for identifying and exchanging intellectual property in the digital environment. It provides a framework for managing intellectual content, for linking customers with content suppliers, for facilitating electronic</p>

commerce, and enabling automated copyright management for all types of media.

FGDC

Federal Geographic Data Committee

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

The FGDC coordinates the development of the National Spatial Data Infrastructure (NSDI). The NSDI encompasses policies, standards, and procedures for organizations to cooperatively produce and share geographic data.

GILS

Government Information Locator Service

http://www.access.gpo.gov/su_docs/gils/

The GILS is an effort to identify, locate, and describe publicly available Federal and state information resources. GILS is a decentralized collection of agency-based information locators using network technology and international standards to direct users to relevant information resources within the Federal Government.

ISO

International Organization for Standardization

<http://www.iso.ch/iso/en/ISOOnline.frontpage>

The ISO is a worldwide federation of national standards bodies from some 140 countries, established in 1947. The mission of ISO is to promote the development of standardization and related activities in the world with a view to facilitating the international exchange of goods and services, and to developing cooperation in the spheres of intellectual, scientific, technological and economic activity.

MARC

Machine Readable Cataloging

<http://www.loc.gov/marc/>

The MARC formats are ANSI/NISO, (Z39.20) standards for the representation and communication of bibliographic and related information in machine-readable form.

METS

Metadata Encoding and Transmission Standard

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

The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium. The standard is maintained in the Network Development and MARC Standards Office of the Library of Congress, and is being developed as an initiative of the Digital Library Federation.

MOAII

The Making of America II

<http://sunsite.berkeley.edu/MOA2/>

The Making of America II is a Digital Library Federation project to create a proposed digital library object standard by encoding defined descriptive, administrative and structural metadata, along with the primary content, inside a digital library object.

NISO

National Information Standards Organization

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

NISO founded in 1939 as a non-profit association accredited by the American National Standards Institute (ANSI), identifies, develops, maintains, and publishes technical standards to manage information in our changing and ever-more digital environment. NISO standards apply both traditional and new technologies to the full range of information-related needs, including retrieval, re-purposing, storage, metadata, and preservation.

NLA

National Library of Australia

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

The NLA is among the pioneer institutions in digital preservation research. Its preservation activities' page provides links to its various initiatives and projects. The set of preservation metadata developed by NLA is invaluable resource.

OAIS

Open Archival Information System

<http://www.ccsds.org/RP9905/>

The OAIS Reference Model has been of great value in providing a comprehensive and consistent frame of reference that encompasses many of the issues surrounding the creation of digital repositories. See also CCSDS.

OCLC

Online Computer Library Center

<http://www.oclc.org>

OCLC is a nonprofit membership organization serving 41,000 libraries in 82 countries and territories around the world. Its mission is to further access to the world's information and reduce library costs by offering services for

libraries and their users. OCLC is the leading global library cooperative, helping libraries serve people by providing economical access to knowledge through innovation and collaboration.

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

PANDORA is an archive of the National Collection of Australian Online Publications copied with the publisher's permission and preserved and made available for the future.

RDF Resource Description Framework
<http://www.w3.org/RDF/>

The Resource Description Framework (RDF) integrates a variety of applications from library catalogs and world-wide directories to syndication and aggregation of news, software, and content to personal collections of music, photos, and events using XML as an interchange syntax. The RDF specifications provide a lightweight ontology system to support the exchange of knowledge on the Web.

RLG The Research Libraries Group
<http://www.rlg.org>

TRLG is a not-for-profit membership corporation of over 160 universities, national libraries, archives, historical societies, and other institutions with remarkable collections for research and learning. Rooted in collaborative work that addresses members' shared goals for these collections, RLG develops and operates information resources used by members and nonmembers around the world.

SGML Standard Generalized Markup Language.
<http://www.w3.org/MarkUp/SGML/>

SGML is an international standard for the definition of device-independent, system-independent methods of representing texts in electronic form. See also XML.

TRAIL Texas Records and Information Locator Service
<http://www.tsl.state.tx.us/trail/>

TRAIL provides access to Texas State government information contained in electronic publications. TRAIL facilitates ready access to the information for Texas citizens and other users.

VRA Visual Resource Association
<http://www.vraweb.org/>

The VRA is a firmly established association with over 600 active members and devoted to advancing knowledge, research, and education in the field of visual information resources.

XML Extensible Markup Language
<http://www.w3.org/XML/>

The XML is the universal format for structured documents and data on the Web. It is a simple, very flexible text format derived from SGML (ISO 8879). XML is also playing an increasingly important role in the exchange of a wide variety of data on the Web. The base specifications are XML 1.0, W3C Recommendation February 1998, and Namespaces, January 1999.

W3C World Wide Web Consortium
<http://www.w3.org/>

The W3C develops interoperable technologies (specifications, guidelines, software, and tools) to lead the Web to its full potential as a forum for information, commerce, communication, and collective understanding.

Z39.50 refers to the International Standard, ISO 23950: "Information Retrieval (Z39.50): Application Service Definition and Protocol Specification", and to ANSI/NISO Z39.50.
<http://www.loc.gov/z3950/agency/>

Appendix 2. Draft Preservation Metadata Elements

Element's						

Name	Origin	Sub-Element	Definition	Purpose	Example	Comments	
						Note	Remarks
Digital Resource Description							
Title	OAIS (Reference Information)		This element will record the name given to the resource. Typically, a Title will be a name by which the resource is formally known.	Any form of the title used as an alternative to the formal title of the resource.	<i>Texas Register:</i> Volume 27 Number 8		
Creator	OAIS (Reference Information)		This element will record the entity primarily responsible for making the content of the resource.	An Author or Creator include a person, an organization or a service.	The Office of the Texas Secretary of State.		
Date			This element will record a date associated with an event in the life cycle of the resources. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 [W3CDTF] and follows the YYYY-MM-DD format. But if the full date is unknown, month and year (YYYY-MM) or just year (YYYY) may be used.	The date the resource was created or became available in its current form, or the date that it was last modified. Therefore, the following qualifiers may be used: - <i>created</i> - creation date of the resource - <i>modified</i> - modification date of the resource - <i>issued</i> - date on which the resource was made formally available.	2002-02-25.		
Persistent	OAIS (Reference Information)		An identifier or 'permanent name' for a digital	An identifier or 'permanent name' for an object that identifies it uniquely,	URL, ISBN, ISSN, File-Name,		

Identifier	Information) NLA (Persistent Identifier)		resource that identifies it uniquely and persistently.	enables links to metadata about it, and to other objects related to it.	...	
Other (Metadata) Identifier			This element will record different kinds of identifiers such as Local Control Numbers.	These identifiers will serve as pointers to the metadata information in the local system.	III catalogue address, OCLC no, ...	
Original Content Type	Cedars (Object Origin)		Contains a description of the original information resource prior to created (or modified) as the current digital form.	It describe the physical nature of the original item from which the digitized content was produced.		Use this element if the original is not born Digital or different from the current format.
Relation	NLA (Relationships) -OAIS - Context Information (Relation)		Specifies any other information resources, which were judged, to be significantly related to the digital resource being described necessary for preservation management. It also enables a digital resource to be linked to earlier or later editions of it, other forms of it, to its metadata, and other objects, including finding aids.	It is essential to maintaining a history of the change of a related digital information source.	<ul style="list-style-type: none"> -Is part of a higher aggregation, e.g., this is part (section) of [Identifier] - Contains the lower component (repeatable), e.g., contains [Unique Identifier] - Relation to the primary digital resources, e.g., this is the html version of [Unique Identifier of primary digital resources] - Related to accompanying information resources, e.g., accompanied by map [Unique Identifier of the accompanying information resources] - Linked to previous and/or following in a migration sequence, e.g., was migrated from or to [Unique Identifier] 	
Structural type	NLA		Class of digital object represented by the digital resource.	Choice of appropriate preservation strategy depends on knowing structural type.	Still image, sound, text, data base, Web document, executable program, etc. (List of MIME types may serve as a useful reference).	

Technical infrastructure (of complex digital resource)	NLA		Internal Structure of complex digital resource: i.e. an enumeration of the components of a complex object, along with their interrelationships.	Managing preservation requires managing the structure of complex digital resource as well as their components.	- e.g.-1. Web page (consists of one ASCII HTML file, along with three embedded static GIF files and one embedded audio WAV file) - e.g.-2. CD-ROM containing 22 files (14.gif image files, 3.wav audio files, 3.txt files and 2.ex executables assembled in accordance with ISO 9660.		
File description	NLA		Technical Specifications of the digital resource(s) or file(s) comprising a content Data Object.	Describe type-specific metadata essential for managing preservation.	File-format and version, resolution, dimensions in pixels; color palette; compression, other-info, e.g. Image (TIFF v.4.0, 600 dpi, dimensions in pixels; color palette; compression algorithms).	This metadata should apply to file formats which are used to directly render or access content, rather than file formats which are used for storage convenience (e.g., ZIP or TAR files).	Depending on the local requirements and type of dr, it can further breakdown for various classes of DR.
Installation requirements*	NLA		Any specialized procedures needed to install an object.	Enables access to objects with special installation requirements.	DR is in the form of a ZIP file, which must be unpacked and stored on local hard drive in a specified directory tree prior to use; computer must be re-booted after installation, etc.		* Information pertaining to IR (with System Requirements) may be placed in the Documentation linked to the DR.
File-Size	RLG/OCLC (size) NLA (storage Information)		Describes details of the storage requirements of the digital resources (in bytes).	Necessary for managing the object within the archive system. E.g. migration of storage media to CD-ROM might require this information, since standard	Size of DR: 1.4 MB.	This element is important for dissemination purposes as some versions of Windows cannot accept files greater	

				CD-ROMs have a maximum capacity of 700 MB.		then 2 GB.	
Access Inhibitors*	NLA		Description of any features of the digital resources intended to inhibit access.	Without this information the DR may not be able to be accessed, copied or migrated.	Encryption, watermarking, digital signature, password protection, etc.		This information may be placed in the Documentation linked to the DR.
Access Facilitators*	NLA		Description of any system or method used to enhance access to information within the content of the digital resource, which need to be maintained in successive generations.	Enable the aids and facilitators to be taken into account in any preservation process.	Time markers in audio or video files, navigational links in a hypertext document, CD type ID points linked to file, Metadata description, etc.		Information about Access Facilitators can be placed in the Documentation linked to the DR.
Exception	NLA (Quirks)		Any characteristic that may appear as a loss in functionality or change in the look and feel of the digital resource resulting from the preservation processes and procedures.	Help to assess the success (or otherwise) of preservation strategies, and prevent time being spent on trying to solve problems that were inherent in the object at the time the strategy was applied.	- Web page: has been migrated from HTML to PDF (as a result, hyperlinks are broken; embedded JavaScript application no longer functional). - The shockwave files could not be captured from the source document.	This element describes peculiarity or exceptions that occur as a result of digitization, migration, and other processes in the preservation cycle. See also <i>Functionality</i> .	Exceptions Vs. Functionality: - one as the "negative" of the other. Functionality metadata records all of the attributes, which still exist in the current instance of the DR. Conversely; the Exceptions (NLA's Quirks) metadata lists all of the attributes, which no longer exists as part of the current DR. [I.e. E+F=original (all) attributes].
Functionality	RLG/OCLC		Description of any functional or "look and feel" attributes of the rendered digital resource (in regards to its current manifestation).	Enumerate the set of functional properties exhibited by the digital resource relative to the current stage of the	- Web page: contains an interactive JavaScript application and embedded animations.	See also <i>Exception</i> .	

				preservation cycle.			
Documentation	RLG/OCLC		Supporting documentation necessary/useful for display and/or interpretation of the digital resource.	Link the digital resource to supporting documentation useful for rendering and understanding its content.	Manual, Procedure, Glossary, etc.	Sub-Element	
	RLG/OCLC	Location	Location of Documentation.	Link the digital resource to the document.	URL, File name, etc.		

Technical Environment Descriptions

2.1 Software

Access Application (Software Name and version)	- RLG/OCLC (Display/Access Application) - CEDARS (partly Render/analyze engine, Input formats, output formats)		Identification of software program capable of displaying or accessing the content of the digital resource.	Translate the archived byte stream into human-readable content.	- Internet Explorer 6.0, - Adobe Acrobat Reader 4.0, etc.		Specify if it is the minimum or recommended environment?
	RLG/OCLC	Location	Location of the Access Application needed to display and/or access the digital resource's content.	Link digital resource to compatible Display/Access Application.	This, (description of where the required Access Application can be obtained), may take the form of anything ranging from manufacturer information, to a pointer (e.g. URL) to the location of where the Access Application can be directly obtained (e.g., via download, or through the archive itself).		
				Identify operating environment		According to NEDLIB, for e.g. Windows	

Operating System (Name and Version)	NEDLIB		Name/designation and version of the Operating System or software platform upon which rendering programs operate.	used by the rendering programs of the DR and also distinguish between different versions of an operating environment, which could potentially impact the ability to access the DR.	Windows (Windows 3.1, Windows 95, Windows 98, Windows ME), Linux, Apple, Solaris, etc.	NT is a general operating environment, characterized perhaps by a particular look and feel and set of functionality. Windows NT 4.0, however, is a specific implementation of the Windows NT environment.	- Specify if it is the minimum or recommended environment. - NEDLIB recommends creating metadata for OS name and OS Version separately.
	RLG/OCLC	Location	Location of working copy of the Operating System.	Link DR to compatible Operating System.	URL to download OS from manufacturer, or from a digital repository holding an archived copy of the OS. Also could be include the location of an emulator for this environment.		
		Documentation Location	Location of supporting documentation useful for operation or use of the OS.	Link the OS metadata to supporting documentation useful for operation.	E.g. URL of Users' Manual, Glossary, etc.		

2.2 Hardware

Microprocessor Requirements	NEDLIB		Description of Microprocessor specifications necessary to operate the content of the DR's software environment.	Ensure that users' obtain sufficient processing power to run the software necessary to display the DR.	Could be a general specification (e.g. 333Mz), or a particular microprocessor (e.g. Intel Pentium II 333 Mz).		
	RLG/OCLC	Documentation Location	Location of supporting documentation useful for operation or use	Link the Microprocessor metadata to supporting documentation	E.g. URL of Users' Manual, Glossary, etc.		

			of the Microprocessor.	useful for operation.			
Storage Information	NLA		Description of any permanent storage resources necessary for the operation of the software environment and or rendering of the digital resource.	Ensure that users' obtain sufficient storage resources to display/render the digital resource.	User must have 33 MB of hard disk space free in order to install/run the software environment.		
	RLG/OCLC	Documentation Location	Location of supporting documentation useful for operation or use of the Microprocessor.	Link the storage metadata to supporting documentation useful for operation.	E.g. URL of Users' Manual, Glossary, etc.		
Peripheral Requirements	NEDLIB		Description of additional equipment needed to render/display of the digital resource.	Describe the complete set of physical resources necessary to access to object's content.	Sound card, speakers, a monitor with a particular resolution, CD-ROM drive, etc.		
	RLG/OCLC	Documentation Location	Location of supporting documentation useful for operation or use of Peripherals.	Link the Peripherals metadata to supporting documentation useful for operation.	E.g. URL of Users' Manual, Glossary, etc.		
Location of Hardware	RLG/OCLC		Location of the physical devices needed to render the digital resource.	Link DR to compatible Hardware Environment.	Description of where the required Hardware Environment can be obtained. This may take the form of anything ranging from contact information for a "technology museum" to the location of emulation programs (perhaps	This may take the form of an enumeration of all possible environments.	We may choose to describe only a minimum or (recommended) hardware environment.

					maintained by the UNT itself).		
Alteration history	- OAI Provenance Information (Modification history) - NLA (Process)		This element documents what has happened to a particular digital resource. It describes any changes made, from the time of creation of the digital resource.	All relevant details of any process applied to a digital resource, including specific settings or actions that were required to produce the digital resources. This information is essential to document what preservation methods have been applied to the digital resources and how the various copies or formats of digital resources might differ from each other.	This field probably store information such as the element was disintegrated into its integral parts or change in Format.	This element will have sub elements such as: <i>-Action:-</i> describes what was done to change the original digital resource. <i>-Policy Applied:-</i> this element will serve as a pointer to existing policies relating to system processes like migrations.	The series of linked records pertaining to the digital resources builds up a change history over time.
Preservation Metadata Creation Date			The date on which the preservation metadata record was created.	We may record the most recent date on which the preservation metadata was updated.	YYYY-MM-DD format.		
Preservation Metadata Creator	NLA (<i>Record Creator</i>)		The names of individuals who have contributed data to this metadata record.	This element record responsibility for the metadata creation and/or alteration.		System-generated log could be one way of recording preservation metadata creator information.	

Note			Any other information relevant to the preservation of the digital resources.	This element will serve as a catch all note field.	Free form text.	Not encourage to use this element.	
------	--	--	--	--	-----------------	------------------------------------	--

Editorial history

Paper received 17 May 2002; accepted 22 July 2002.

[Contents](#) [Index](#)

Copyright ©2002, First Monday

Copyright ©2002, Daniel Gelaw Alemneh

Copyright ©2002, Samantha Kelly Hastings

Copyright ©2002, Cathy Nelson Hartman

A Metadata Approach to Preservation of Digital Resources: The University of North Texas Libraries' Experience by Daniel Gelaw Alemneh, Samantha Kelly Hastings, and Cathy Nelson Hartman
First Monday, volume 7, number 8 (August 2002),
URL: http://firstmonday.org/issues/issue7_8/alemneh/index.html