Office of Digital Collections and Research
University of Maryland, College Park

Best Practice Guidelines for Digital Collections

at University of Maryland Libraries

edited by

Susan Schreibman

authored by

*Yvonne Carignan, Janet Evander,
Gretchen Gueguen, Ann Hanlon,
Kate Murray, Jennifer Roper,
Tony Ross, Susan Schreibman*

libdcr@umd.edu

First Edition Released 6 February 2006

Revised June 2006

Second Edition Draft Released 4 May 2007

Preface to the Second Edition

This document was authored by Yvonne Carignan (Preservation), Janet Evander (Information Technology), Gretchen Gueguen (Digital Collections and Research), Ann Hanlon (Archives and Manuscripts), Kate Murray (Archives and Manuscripts and Digital Collections and Research), Jennifer Roper (Original Cataloging), and Susan Schreibman (Digital Collections and Research). The goal of the first edition was to introduce our colleagues in the University of Maryland Libraries to the opportunities and challenges of creating digital collections by addressing:

- current standards across media
- application of metadata
- issues of project management

With this second edition, we have more fully fulfilled that promise by addressing audio and moving image formats, as well as including sections on Digital Masters, User-Centered Design, and Web Authoring Guidelines.

As with the first edition, parts of this document are prescriptive, such as the minimum standard for master images. However much of it is descriptive, exploring the range of issues with which project teams must engage when designing and mounting digital collections. Perhaps the most important message of these Guidelines is that digital collections must be approached with a somewhat janus-faced attitude: on the one hand, standards should be followed as closely as possible; on the other hand, collections tend to have a unique look and feel and are conceived with particular project goals and user communities in mind. By virtue of being accessible on the Web, digital collections inevitably find new and unanticipated audiences. As these collections are conceived within a library environment, they are being developed with an eye to longevity and a commitment to maintaining them for as long as the library commits to maintain its analog resources. It is not only conceivable, but also inevitable, that these collections will undergo a series of migrations. To best prepare for these migrations, it is essential that standards be followed and that any deviation from standards be documented. Digital collections are like symphonies — each object akin to a single note: alone, an object may lack context, but, when brought into play against the hundreds or even thousands of other objects in the collection, harmony is produced in which the whole becomes greater than the sum of its parts.

The Office of Digital Collections and Research is a resource to help colleagues in the University of Maryland Libraries navigate the rapidly changing, acronym-laden, dynamic world of digital library initiatives. This document is intended both as an introduction and as a guide to help navigate this fast-paced domain.

Susan Schreibman, PhD
Assistant Dean
Head of Digital Collections and Research
College Park, Maryland, January 2007

Contents

1.0 Projects, Collections, and Objects 1
1.1 Digital Masters 2
2.0 Issues of Copyright and Permissions 4
2.1 Conditions of Use 4
2.1.1 Responsibilities to the Copyright Holder 4
2.1.2 Responsibilities to the User 4
2.2 Copyright Status as a Condition for Selection 4
2.3 Orphaned Works 5
3.0 Metadata 6
3.1 Metadata Mapping 6
3.2 Levels of Metadata 6
3.3 Types of Metadata 7
3.3.1 Descriptive Metadata 7
3.3.2 Administrative Metadata 7
3.3.3 Technical Metadata 7
3.3.4 Structural Metadata 8
3.3.5 Preservation Metadata 8
3.4 Controlled Vocabulary and Authority 8
3.5 Issues of Interoperability and Harvesting 9
3.5.1 Metadata Standards 9
3.5.2 MARC Records and Representation of Projects in the Catalog 9
4.0 Project Life Cycle 11
4.1 Selection/Collection Policy 11
4.2 Consideration of Original Materials 11
4.3 Workflow 11
4.4 Project Management and Staffing 12
4.5 Object Naming Conventions 12
4.5.1 File Naming Conventions 12
4.5.2 Persistent Identifiers 12
4.6 Digitization Options: In-house vs. Outsourcing 13
4.7 Budget 13
4.8 Quality Control: Ensuring the Quality of Digital Objects 13
4.9 Audience and Dissemination 14
4.10 Migration 14
4.11 Authenticity, Chain of Custody 15
5.0 User-Centered Design 16
5.1 Overview 16
5.2 Involving Users in the Design Work 16
5.2.1 Participatory Design 16
5.2.2 Personas 16
5.3 Card Sorting 16

5.4 Defining Users and Their Needs 17
5.4.1 Focus Groups 17
5.4.2 Individual Interviews 17
5.4.3 Surveys 17
5.5 Evaluating Sites Based On Usability Principles 17
5.5.1 Expert Evaluations 17
5.5.2 Usability Audits 17
5.6 Testing For Usability 18
5.6.1 Usability Testing 18
5.7 Conclusion 18
5.8 Further Reading 18
5.8.1 User-Centered Design 18
5.8.2 General Usability 18
5.8.3 Usability Testing 19
6.0 Web Authoring Guidelines 20
6.1 Content 20
6.1.1 Density of Text 21
6.1.2 Independence 21
6.1.3 Depth vs. Breadth 21
6.1.4 Parallelism 22
6.2 Design 22
6.2.1 Using CSS instead of writing style information in the code 22
6.2.2 Javascript server side includes 23
6.2.3 Masking emails from bots or spiders 23
6.2.4 Validating code 23
6.3 Administration 23
6.4 Becoming a Web Design Expert 24
7.0 Image Collections 25
7.1 File Formats 25
7.2 Color Mode and Bit Depth 26
7.3 Bit Depth 27
7.4 Resolution and File Size 27
7.5 Quality Control, testing, reference targets 27
7.6 Types of Inspection 28
8.0 Text Collections 29
8.1 Document Analysis 29
8.2 Considerations when working with full text 29
8.2.1 Structured or unstructured data 29
8.2.2 Corrected and uncorrected OCR 30
8.3 Full text mark up 30
8.4 Header vs. Body 31
8.5 Encoded Archival Description 31

8.5.1 Finding aids and their purpose 31
8.5.2 Arrangement of a finding aid 31
8.5.3 Levels of description 32
8.5.4 EAD and MARC 32
8.5.5 EAD@UMD 33
8.6 The Text Encoding Initiative 33
8.6.1 Encoding Levels 33
8.6.2 Collaboration 34
8.7 Creating the Electronic Text 34
9.0 Digital Audio and Moving Images 36
9.1 Special Considerations for Digitization of Analog Audio and Moving Images 36
9.2 Project Pre-work: Digital Audio and Moving Images 37
9.2.1 Time Management Needs 37
9.2.2 Condition of Original Materials 37
9.3 Digitization Of Audio Material 38
9.3.1 Required Digital End Products: Standard and Maximum 38
9.3.1.1 Standard 39
9.3.1.2 Maximum 39
9.4 Technical Policies 40
9.4.1 Policy on "Dead Air" in Original Recordings 40
9.4.2 Policy on Adding a Buffer to Digital Files 40
9.5 Digitization of Moving Images 40
9.5.1 Required Digital End Products: Minimum, Standard and Maximum 41
9.5.1.1 Standard 41
9.5.1.2 Maximum 41
9.5.1.3 Minimum 41
9.6 Quality Control for Digitized Audio 42
9.7 Quality Control for Digitized Moving Images 42
Appendix I: Public Domain Determination 43
Appendix II: Guidelines for Working with Original Documents 44
Appendix III: Steps in Usability Testing 47
Appendix IV: A Typology of Formats 49
Appendix V: Minimal Requirements for Creating Digital Images 52
Appendix VI: Quality Control for Images 53
Appendix VII: XML Examples 56
Sample TEI document 56
Sample EAD Document 59
Sample UMDM Record 65
Sample UMAM Record 66
Appendix VIII: Additional Audio Project Planning Tools 67
Appendix IX: Glossary 70
Appendix X: Bibliography 74

1.0 Projects, Collections, and Objects

The building blocks of digital library initiatives are objects that are incorporated into collections and are created as a result of projects. Since these three terms are used differently within the digital library community, this document defines them as follows:

Collections: A digital collection consists of a group of objects in digital format considered as a whole that demonstrates some identifiable organizing principle such as:

- A common theme or subject, format, or mode of publication
- Creation or collection by an individual (e.g. the works of an author) or corporate body
- A common source or origin

Digital collections may or may not be different in content from traditional collections. Depending on the purpose of a project, a digital collection might mimic an existing physical collection, or it could bring together items from multiple collections or multiple organizations. The key feature of digital collections is that they integrate discrete items, either through description and format or through the concept of the project itself.

To demonstrate just one example, a digital collection might comprise a selection of still images culled from several physical collections. The selection criteria, along with rules for description, and the packaging of the digital collection, integrates these objects into a single digital collection. Regardless of the final product, the criteria for development of a digital collection should be thoroughly fleshed out before bringing a project to fruition. For more on digital collection development, see section 4.0: Project Life Cycle.

Projects: A digital project includes the resources (human, computational, financial, material) that are brought together to create a digital collection. While projects are typically finite in duration, collections should not be. After the initial project phase, digital collections, like their physical counterparts, require ongoing curatorial attention.

A digital project may comprise one or more digital collections, or it could forego the convention of collections altogether. At project conception, project developers decide upon:

- The goals and purpose of the digital project
- Content, including the sources for that content
- Accessibility (web interface, copyright issues, etc)
- Audience
- The project's fit with other digital projects and initiatives, as well as Library priorities

These factors determine how decisions are made about collection development, what selection criteria are implemented, the degree of technical expertise required, whether there are copyright or privacy issues, the depth of description required, as well as the appropriate metadata format or formats. For example, a project designed to introduce Civil War materials to elementary school students might not involve decisions on copyright, as most materials are within the public domain. However, tailoring the project's web site to the appropriate audience might benefit from a user study and could very likely involve technical support either to build a new system and interface or to modify

an existing one. Another project might simply add a selection of digitized photographs of local historical sites to an existing database but could still require additional effort to determine rights information. This photograph project might benefit from further research to enrich the item-level description if the photographs have previously been described only at the collection or series level.

Objects: Conceptualizing a digital project involves understanding the building blocks — the digital objects that form a collection. Each digital object represents a discrete unit and is comprised of a digital file or files as well as descriptive metadata. Because of the nature of digital access, it is important to remember that, while a digital object may be part of a digital collection, it may not always be discovered within that context. This makes the metadata that describe the objects crucial and lends another dimension the implementation of digital projects. Digital objects begin life in one of two ways:

1. As a digitized file produced as a surrogate for materials that exist in analog format
2. As a "born digital" entity, with no analog counterpart

The National Information Standards Organization (NISO) has described digital objects as "equivalent conceptually to the items that may be found within library holdings, museum collections, and archival collections" (NISO 2004). This is true in the sense that a digital object is made up not only of a digital file but also of its accompanying metadata, just as a book in a library is "accompanied" by a catalog record.

Thus is it useful to think of simple and complex digital objects. A simple digital object may be a scanned photograph or a TEI-encoded text file. A complex digital object could be that single photograph (scanned as one digital file), along with its accompanying metadata. In other cases, a discrete analog item may be represented by multiple digital files. A slim pamphlet could be represented by four digital files – one for the cover, one for the title page, one for the text page, and one for the back cover. All four digital files, with their accompanying metadata, would constitute one complex digital object. The relationship between each digital file would be represented via metadata. For more on metadata, see section 3.0: Metadata.

1.1 DIGITAL MASTERS

Throughout this document the term "digital master" is frequently used. This term does not refer to a particular file format, but rather to a digital object that "most closely retains the significant attributes of the original." (Digital Library Federation Benchmark for Faithful Digital Reproductions) Masters are created to be long-lived and high quality to enable the production of versions for various uses. For example, if one were creating masters of a book, they might include a full text version of the text encoded in a standard XML format, such as the Text Encoding Initiative Guidelines. They might also include page images in an image format (TIFF or JPEG2000) of sufficient quality to print facsimiles. A project may have more than one master file. Archival or preservation masters adhere to specific file formats and are created under rigorous quality control standards. Archival masters may be created as substitutes for the original analog object. Print or submaster files, on the other hand, are used as a high quality digital object for the production of access or other copies. Submasters may be of the same file format as archival masters, but may, for example, be retouched for display purposes in a way that archival masters should not be.

Generally, a preservation master is protected in secure storage with good climate control and restricted use. Ideally, a preservation master is used only once, to create the submaster from which all user or access copies are created. Theoretically, then, a master might be used only once, or as infrequently as possible, serving only to create submasters. Many projects might not have the need

or economic resources to support both an archival master and a submaster. In this case, the archival master would be used to create derivative files. Original analog objects should also be preserved, no matter how fine the master copy is. There is always a chance that new technologies will result in improved capture and creation of even better quality master copies from originals (for a further discussion of these issues, see SoliNET Preservation Recording, Copying, and Storage Guidelines for Audio Tape Collections)

In reality, however, most research collections of analog and digital, text and audiovisual materials are faced with a complex scenario in which goals, availability, and finances, among other factors, may mean that masters and master copies do not exist as described above. The original may have already been lost. For example, perhaps a brittle book was microfilmed to low standards and then discarded. The microfilm might be the last, best copy, but a digital copy of the film might be much more legible. In the case of audiovisual formats, some titles may no longer survive in an original, but only as poor quality formats or as copies many generations removed from the original creation. For example, The UM Broadcasting Archives holds a unique Woody Allen "short" which survives only as a blurred copy on videotape. Even costly restoration cannot return the piece to the quality of a first generation original.

Besides the fact that some important items only exist in a less than ideal format, issues of intellectual property or donor wishes may result in the institution's creation of digital objects that are less than ideal master quality. For example, issues of copyright may prevent an institution from saving an archival master. In this case, a "print master" may be sufficient so that new access copies may be generated if an access copy is lost, deteriorates, or must be migrated to a newer access format.

Because of the reality described in these examples of master-less collections, "master copies" must include a broader definition than original or near-original quality. In these situations, a master copy can be viewed as the best copy available of a given item. Digital masters created under these conditions, will, as with preservation masters, not be used for access purposes, but be stored in an architecture which is digitally curated.

2.0 Issues of Copyright and Permissions

2.1 CONDITIONS OF USE

2.1.1 Responsibilities to the Copyright Holder

Conditions of use and copyright restrictions should be known prior to beginning a project. One should not assume that the University owns the intellectual property of an object among its physical holdings. Only objects that are in the public domain, whose intellectual property rights the University owns, or which the Library has explicit permission to use from the copyright holder should be freely integrated into a digital collection. In some circumstances the use of portions of objects may fall under Fair Use Guidelines. For any project which may involve copyright issues, University Legal Counsel should be consulted.

2.1.2 Responsibilities to the User

A clear and accurate account of the conditions of use of materials should be made available to users. If this information pertains to the entire collection, it should be captured in collection-level metadata in addition to item-level descriptions. In all instances, users should be informed of:

- Their rights to view and use the information and objects in the collection
- All applicable copyrights
- Restrictions on use
- How to obtain permission when use is restricted
- How to cite the resources for allowable uses

If special conditions exist for the display or viewing of the material (for example, if the terms of the agreement indicate watermarks on image reproductions), they must be honored. The project is responsible for including information for users to properly cite resources in the collection, including, at a minimum, the name of the resource, the resource handle or persistent identifier, the name of the project, and the project URL. This information is captured in the metadata, which is discussed in more detail in section 3.0.

United States Copyright Law (US Code, Title 17, section 107) includes provision for Fair Use. This allows the use of copyrighted materials for research, instruction, or private study without prior permission, as long as the original source is attributed. Any usage for commercial, display, or publication requires the prior permission of the copyright holder. Because of the nature of digital library projects, it is unlikely that Fair Use could be a justification for making digital collections publicly available. It could, however, permit the use of extracts in an online exhibition. If in doubt about this aspect of a digital project, advice must be sought from University Legal Counsel during the project planning process.

2.2 COPYRIGHT STATUS AS A CONDITION FOR SELECTION

A project may be particularly suited to digitization if materials are in the public domain or if explicit permission for use of copyrighted materials has been granted. If the copyright status is unknown or

the copyright holder cannot be found, the University may be at risk for copyright violation and a digitization project may not be considered. In matters of copyright, scope of distribution may also be a significant factor. For example, objects of broadcast quality video format may be created for preservation purposes, but a lesser quality digital file may be used for public distribution.

> *Viewed from any side, rights issues are rarely clear cut, and the rights policy related to any collection is more often a matter of risk management than one of absolute right and wrong* (NISO 2004).

NISO advises that project managers collect and maintain a record of rights holders and permissions granted to be able to document and justify actions if necessary. Evidence of this status should be available when projects are being considered for digitization and should take into consideration the extended provisions of the Digital Millennium Copyright Act (DMCA; See Appendix I: Public Domain Determination for further information).

2.3 ORPHANED WORKS

An interesting development in the area of copyright law is the notion of "orphaned" works. Copyrighted works are said to be "orphaned" when the owners of the copyright are extremely difficult or even impossible to locate. Since orphaned works may have been intentionally or unintentionally abandoned, using them does include some risk. This issue is potentially far-reaching as the Copyright Office states that "well less than half of all registered copyrighted works were renewed under the old copyright system" (Notice of Inquiry: Orphaned Works 2005).

The Copyright Office of the Library of Congress is currently reviewing the issue of orphaned works and has made a public inquiry into the matter and issued a report[1]. Future editions of this document will more fully integrate these findings.

The notion of orphaned works may open up opportunities to digitize library holdings in which copyright status is ambiguous. Until this issue is decided, however, careful consideration must be given to the status of orphaned works. Current copyright laws would be applicable in these cases, and a copyright holder may come forward and attempt litigation for misuse, even if the use falls under general Fair Use guidelines, as "the fair use defense is often too unpredictable as a general matter to remove the uncertainty in the user's mind" (Notice of Inquiry: Orphaned Works, 2005).

[1] http://www.copyright.gov/orphan/

3.0 Metadata

3.1 METADATA MAPPING

The University of Maryland Libraries has developed a standardized metadata scheme that will be applied to all new digital projects. Because many new digital projects will inherit the metadata of earlier projects, it is will be necessary to "map" or transfer the data still considered useful into this standardized metadata scheme. If a new project will inherit several older projects, multiple old metadata fields will need to be mapped.

In most cases, the metadata that already exists is in a non-standard format, idiosyncratic to that particular department or database. Project managers will need to work with a Metadata specialist associated with DCR to map this data into the University of Maryland Descriptive Metadata (UMDM) and the University of Maryland Administrative Metadata (UMAM) schemes. Some content from the old scheme will likely be discarded, while new content will likely be added in order to meet minimum metadata requirements.

3.2 LEVELS OF METADATA

Determining the granularity, or detail, of metadata description is essential when developing a digital project. The degree of granularity will not only determine how an object can be discovered by defining the search; it will also describe the types of relationships that exist between objects. Defining the relationships between objects is especially important in an online environment as users may be able to access objects through various paths and they may not discover the existence of related materials if the object metadata does not contain that information.

Not all online projects necessarily benefit from description at the most granular level, but some certainly do. The anticipated audience and the purpose of the collection, among other factors, will help determine levels of description. While some item level description is required for every digital object, some projects may believe that the minimum requirements meet their descriptive needs, while other projects may choose to provide more detailed or specialized information. In either case the minimum requirements for metadata must be met. Projects can choose to describe objects at the minimum requirement level or provide a greater detail of description, as necessary.

Keep in mind the collection context. The goal of the UM Libraries Digital Library Initiative is to house digital content in a single repository, providing the opportunity for objects to be discovered within the context of a collection or collections or from a search outside a collection context via a generic repository search screen. Thus, the relationships of objects are preserved through the metadata. The metadata must capture the name of the collection or project that unites those objects. This would enable a user to see, for example, all of the objects associated with a particular collection with one simple search.

Metadata is created not only for researchers using a collection but also for staff managing the collection. Information such as title, author, and subject is created to assist the user in finding and identifying a resource. Other metadata, such as technical properties of a digital object, is created to assist staff in managing that resource. It is essential to consider both the usefulness and impact of each of these types of metadata for both audiences, as well as the legal implications of providing that information in a vulnerable online environment.

3.3 TYPES OF METADATA

Metadata performs several specific functions in order to describe a digital object. These functions include:

- **Descriptive:** Facilitates discovery and describes intellectual content
- **Administrative:** Facilitates management of digital and analog resources
- **Technical:** Describes the technical aspects of the digital object
- **Structural:** Describes the relationships within a digital object
- **Preservation:** Supports long-term retention of the digital object and may overlap with technical, administrative, and structural metadata

3.3.1 Descriptive Metadata

The required descriptive elements for the Fedora system, found in the University of Maryland Descriptive Metadata (UMDM), encompass a range of information from basic elements such as title and subject to more advanced elements such as geographic or temporal coverage and relationships. The goal of successful metadata creation is to identify the main user group of a resource and define what types of information will help that group discover and make use of that resource. Be mindful not to limit metadata to known user needs, however, as there are also users whose needs are yet to be discovered. Digital resources are more broadly accessible than their analog counterparts, and it is important to improve the quality of that access by providing necessary description, even if that information may not seem significant to the primary, identified user group. For example, it is easy to imagine that a collection of Civil War diary entries might be created with an undergraduate class in mind, but, over time, might find a wider body of users, such as historians, genealogists, and Civil War buffs.

In the UMDM scheme, the descriptive metadata can also contain data about the creation and location of an analog resource from which a digital surrogate was created. For instance, a project may seek to provide access to an entire collection of digitized images, which are also held in analog format. In this case, the descriptive metadata may include information about the creation, and storage of both the analog and digital formats. However, if there is already a strong local tool for controlling the analog format, a project may only choose to track information related to the digital object through the administrative metadata.

3.3.2 Administrative Metadata

While descriptive metadata is produced to assist end users, administrative metadata is intended to facilitate management of resources for staff. Administrative metadata is regularly found in a record with descriptive metadata, and the two types of metadata often intersect with both user groups. Most often the type of administrative metadata provided is dictated by local needs. Administrative metadata must include a rights statement and technical data but can also track the creator of the digital resource, when and where it was created, as well as rights management issues, access requirements, or means of tracking the resource.

3.3.3 Technical Metadata

Technical metadata is necessary to ensure the usefulness of the digital object. It is often defined as a subset of administrative metadata. Technical metadata captures information about the digital

object itself, such as file size, file type or format, bit depth, compression ratio, etc. Some tools may make it possible to automatically capture some, if not all, technical metadata.

3.3.4 Structural Metadata

Structural metadata, as Caplan notes, "describes the internal organization of a resource" (Caplan 158). For example, a complex digital object such as a pamphlet or book may consist of more than one digital file. Structural metadata describes the relationship between the multiple digital files that make up that single digital object, including front and back covers, pages, and page order. Structural metadata might also include text markup that further describes the structure of a document, such as chapter information, chapter headings, and page breaks. For complex digital objects, the University of Maryland Libraries Digital Library Initiative uses the METS (Metadata Encoding and Transmission Standard) schema which is created dynamically by the Fedora repository.[2]

3.3.5 Preservation Metadata

According to the *NISO Framework of Guidance for Building Good Digital Collections*, preservation metadata is "a subset of administrative metadata aimed specifically at supporting the long-term retention of digital objects" (NISO 2004 27). Preservation metadata overlaps with technical and administrative metadata, detailing important information about the digital file, including any changes in the file over time and management history. Preservation metadata does not support discovery or use of digital files but rather their long term retention and use. The object meant to be preserved by preservation metadata is the preservation master digital object itself. Preservation metadata need not be created for the analog object or derivative digital objects created for access purposes, such as jpegs and thumbnails. In the case of complex digital objects (see section 1.0: Projects, Collections, and Objects), one preservation metadata record will be created to describe the entire complex object, rather than multiple records for each digital file.

3.4 CONTROLLED VOCABULARY AND AUTHORITY

For metadata to be effectively searched, authoritative forms of terms as well as controlled vocabularies should be used wherever possible. Users should not have to use multiple variations in terminology to search for the same concept (person, place, subject, etc.). Additionally, authoritative forms can strengthen links between disparate collections. Commonly used subject or genre thesauri include *Library of Congress Subject Headings*, *Thesaurus for Graphic Materials*, *Medical Subject Headings*, and the *RBMS Thesauri for Binding Terms, Printing & Publishing Evidence, Genre Terms, Provenance Evidence, Paper Terms, and Type Evidence*. Use of established thesauri for controlled vocabulary is recommended practice, although it is possible that none of the standard thesauri will fit a project's needs. In all cases, members of the project team should work closely with the Metadata Librarian to construct a useful, standardized vocabulary.

Authoritative forms of personal, corporate, and geographic names should also be used. The appearance of names can vary from publication to publication, and users should not be expected to search on all variations. Authoritative forms are an efficient way to collocate materials together. It is recommended practice to use the authority sources already employed by the Technical Services Division (TSD), so that users can apply their search techniques and knowledge from the library catalog to the Libraries' digital projects. The main authoritative source is the *Library of Congress*

[2] http://www.loc.gov/standards/mets/

Name Authority File (LC NAF), which is available and searchable online.[3] When a name does not appear in the *LC NAF*, an attempt should be made to investigate the proper form of that name. Resources such as the *Getty Institute's Union List of Artists Names*[4] and other biographies and bibliographies can be useful in determining an authoritative name form. In cases where a name does not appear in the *LC NAF* and the project managers feel that the name is significant enough for inclusion, the Metadata Librarian should be consulted about creating an authority record.

Descriptive elements are not the only metadata that should follow authority practice, as metadata that is created for administrative use also needs to be efficiently searchable. Elements such as UMDM's "mediaType," which defines the broad category that defines the material (e.g. image or sound), benefit from use of standardized vocabulary. In the case of the "mediaType" element, UM Libraries requires use of the *DCMI Type Vocabulary*, to ensure that constant terminology is used to differentiate between images, moving images, datasets, etc. Other examples of metadata components that benefit from use of controlled vocabulary include format, coverage, and policy statements. Consult with the Metadata Librarian to determine which fields require controlled vocabulary and the appropriate authorities to employ.

3.5 Issues of Interoperability and Harvesting

3.5.1 Metadata Standards

One of the first and most important questions to answer at the beginning of a digital project is "which metadata standard to employ?" Different schemes may be employed to support different projects. The UMDM will provide the descriptive metadata for all digital objects, but some digital projects might also be well served by the use of a complementary metadata standard. The Text Encoding Initiative (TEI)[5] is suitable for projects which seek to capture information about textual documents and is the standard for encoding full text documents in XML. If the information has a hierarchical structure, such as finding aids for Special Collections, the Encoded Archival Description (EAD)[6] may be a better match.

To determine which scheme is most well-suited to a particular project and to interoperability with other University of Maryland Libraries digital projects, as well as in the community at large, consult with DCR during the earliest stages of project planning.

3.5.2 MARC Records and Representation of Projects in the Catalog

Objects in a digital project should be represented, at some level, in the catalog. The catalog record need not be a one to one relationship with the digital objects, as the analog counterparts of many digital objects may exist as part of a larger group or collection that is represented by one MARC record. In cases where there is no catalog representation for material, in either digital or analog format, it is necessary to provide this access.

Like the decision on the level of granularity for the digital project metadata, a decision must be made regarding the granularity of the catalog record. In many cases, a one-to-many model is appropriate, with one catalog record acting as surrogate for many digital objects. In this approach, the catalog record is simply used as a hook into the digital project, providing only broad-based subject access.

---

[3] http://authorities.loc.gov/
[4] http://www.getty.edu/research/conducting_research/vocabularies/ulan/
[5] http://www.tei-c.org
[6] http://www.loc.gov/ead

The MARC record will allow for serendipitous identification of digitized objects from within the catalog for users who may not be aware that a digital collection exists. Other cases may call for a one-to-one relationship, in which the catalog records act as surrogates for individual objects. In this case, the focus will be more balanced between subject access and more elaborate description of an object. Deciding what type of MARC record is most appropriate should be made on a case by case basis, in consultation with the Technical Services Division. It should also be noted that, if the metadata created for the digital project is XML based, the creation of a MARC record can be automated, requiring minimal editing.

4.0 Project Life Cycle

4.1 SELECTION/COLLECTION POLICY

A digital collection should be the result of careful selection and collection policies. In addition to applicable copyright and technical considerations, the scope and nature of the materials, as well as the intended audience, should be taken into account. Materials selected for digitizing, as well as the resulting digital images or collections, should support the education, research, and service missions of the University of Maryland and fit within the Libraries' collection policy.

Preservation of valuable or unique objects may also provide compelling reasons to undertake a digitization project. Likewise, collections whose use will be increased by digitization will have a higher priority for selection. Digitization alone will not increase the use or value of materials. This value and utility must be demonstrated for digitization to take place.

For example, one might argue that digitizing the collection will provide greater access by allowing users direct access through browse and search modes or that making little-known or valuable works more available will widen their use. Projects may be favored which allow open access, in the first instance, to the entire campus community and then to the world. An argument may be made that by digitizing a collection the original materials are preserved or protected by providing an adequate surrogate. And lastly, an argument in favor of digitization is that new kinds of research may be enabled because of the nature of the electronic presentation.

Some document types may be more suited to digitization than others. This determination must be made on a case-by-case basis, but suitable types may include printed text, book illustrations, rare or damaged printed text, items that convey intrinsic information beyond the printed text, manuscripts, maps and architectural drawings, photographic prints and postcards, photographic transparencies and negatives, microformats, and audio and video formats.

4.2 CONSIDERATION OF ORIGINAL MATERIALS

Before making a digital surrogate of original materials that are part of UM Libraries' permanent collections, please notify the head of Preservation, the head of Brittle Materials, Reformatting and Deacidification, or the senior conservation technician so that they can ensure that the materials can be digitized without damage. Basic care and handling policies and procedures for all categories of original materials are in Appendix II: Guidelines for Working with Original Documents.

4.3 WORKFLOW

Before beginning the digitization process, workflow and environmental conditions should be considered. Scanner settings, screen calibrations, and environmental lighting, along with the organization of work processes and built-in quality checks will all be factors in the quality of the digitized files. As much as possible, these variables should be pre-determined to standardize workflow and quality.

4.4 PROJECT MANAGEMENT AND STAFFING

All digital library projects at the UM Libraries are coordinated through the office of Digital Collections and Research which, in turn, draws on the skills and expertise of individuals throughout the library as appropriate. Generally, the head of DCR, in conjunction with key personnel, will decide who will comprise the project team. Once a project is accepted by DCR, content holders should build project time into their Work Plans, or, indeed, the Work Plans of their Team or Unit, if it is anticipated that the project will draw on staff beyond the content expert or archivist.

Projects under DCR's supervision typically range from six months to two years, with five months to one year the optimal project timeframe. Of course, as mentioned elsewhere in this document, the project is expected to continue utilizing local resources after this initial start-up phase, with minimal input from the initial project team.

Individuals wishing to pursue a project with DCR should be aware that during this start-up phase there are periods of intense collaboration, with many aspects of the project moving forward simultaneously. During these periods, it is essential that all those involved, particularly those from the originating unit, be able to make work on the project a priority. This may involve negotiation with the indivudual's supervisor and possibly a decision to delay the start of the project so that other unit priorities are not impinged upon.

4.5 OBJECT NAMING CONVENTIONS

4.5.1 File Naming Conventions

File naming conventions must be established before digitization begins in conjunction with DCR and the Metadata Librarian. Identifiers should be persistent and unique and should not refer to an address, filepath, or URL as these can and will change. All objects created as part of a digital initiative at the UM Libraries will be prefixed by a two-four digit unique identifier helping to ensure the uniqueness of file names across the entire repository. Moreover, as recommended by NARA, filenames must:

- Be unique and consistently structured
- Take into account the maximum number of items to be scanned and reflect that in the number of digits used if following a numeric scheme
- Use leading 0s to facilitate sorting in numerical order if following a numeric scheme
- Not use an overly complex or lengthy naming scheme that is susceptible to human error during manual input
- Use lowercase letters and file extensions
- Use numbers and/or letters but not characters such as symbols or spaces that could cause complications across operating platforms

4.5.2 Persistent Identifiers

The creation of persistent identifiers generally follows one of two conventions. The first school of thought is that the identifier should be randomly generated, embedding no semantic meaning. The second school of thought advocates that unique identifiers should carry some meaning that is understandable to those working intimately with the collection.

For projects undertaken at the University of Maryland Libraries, identifiers will be randomly generated by the Fedora system. Individual objects will have a unique, sequentially generated persistent identifier (PID) of the form "umd:###", and this PID will be used for the persistent naming of the object within the Fedora system. The pre-Fedora ingest file name created using the above guidelines will be included in the metadata and used as an access point.

4.6 DIGITIZATION OPTIONS: IN-HOUSE VS. OUTSOURCING

Both in-house and outsourced alternatives should be considered when embarking on a digitization project. Whether to digitize in-house or via outsourcing depends on the scope, nature, fragility, and uniqueness of the materials, on the project budget, and on in-house resources. Although some materials may be unique and fragile, the project team may deem it necessary to outsource as the most time-effective way to undertake the project. On the other hand, a project with scarce financial resources may undertake to digitize over a longer period of time, using in-house resources as they become available.

Both options should be thoroughly investigated before undertaking a digitization project.

4.7 BUDGET

The project budget may be the one of the most important strategic documents for a project team to create. Not only should costs be considered for initial digitization and mounting in a repository, but a strategy must be in place for continuing the project after the grant-funded project phase or after the initial phase of working with DCR.

4.8 QUALITY CONTROL: ENSURING THE QUALITY OF DIGITAL OBJECTS[7]

Quality control (QC) "encompasses procedures and techniques to verify the quality, accuracy, and consistency of digital products" and is essential to all phases of a project life cycle to ensure that the product meets preset standards and goals" (Cornell University Library Research Group 2003). This guide recommends following the detailed advice found in *Moving Theory into Practice* for developing a quality control program (Kenney and Rieger 2000).

That advice calls for identifying products and goals, defining standards and measures for acceptable and unacceptable characteristics of the product, and deciding whether the product should be compared with originals or some intermediate. After these preliminary tasks are completed, decisions about the QC program include defining the scope of percentage of the objects that will be evaluated; choosing the methodology for evaluation; controlling the environment for QC, including configuring the hardware; evaluating system performance; documenting procedures and creating an inspection form; and performing the assessment itself.

The evaluation review process is an iterative one and should be begun early in the process to be able to make necessary changes. Evaluation guidelines and standards should be built into the project documentation so that objects created in future meet the same standard and are developed according to the same practice as the original objects.

---

[7] Sometimes a distinction is made between quality control and assessment, with the former referring to the vendor's assessment of the product to verify quality or the assessment of quality during the processes of creating the digital object. Quality assessment refers to verifying the quality of the digital product by institution staff. For the purposes of this guide, "quality control" will refer to all phases of quality assessment.

Primary objects, whether they are audio, video, image, or text files, should go through rigorous quality control according to standards based on this Best Practice document. Individual chapters on these formats contain more information about evaluations and quality control measures.

Metadata has a central role in processing, managing, accessing, and preserving digital collections. Because of the crucial role it plays in the life cycle of collections, metadata review should be an integral part of a quality control program. Metadata Quality Control can be done via system checks, manually, or a combination of the two. Quality Control should verify the following: data integrity, form and validity, accuracy of derived data, correctness of data, accuracy and completeness of components, and dynamic metadata. For more on metadata, see section 3.0: Metadata.

4.9 AUDIENCE AND DISSEMINATION

Digital collections should be developed with a perceived audience in mind. Collections created for a scholarly audience, for example, may have less contextualization and more sophisticated search interfaces than those developed for a high school or undergraduate audience. It is unlikely, particularly during the first phase of project conception and creation, that the collection could seamlessly serve multiple audiences. If the project seeks to address a K-12 usership, outside expertise, such as education specialists, should be consulted to assist at the project conception stage, as well as to develop lesson plans. Nevertheless, online collections inevitably find wider audiences than those for which the project was initially conceived, and project planning should, as much as possible, take into account the very real possibility of a multiplicity of audiences.

Once a collection is "finished," or at least ready for its first launch, dissemination should be broadly conceived, including announcements on listservs, presentations at professional meetings, linking from related sites, and flyers left at appropriate venues (conferences, public libraries, etc). Near the end of the initial phase of the project life-cycle, the project team should discuss possible publicity outlets, keeping this list on the development server so that announcements of enhancements to the site can be easily disseminated.

4.10 MIGRATION

Migration is an inevitable part of the life cycle of projects, collections, and objects. As technology and standards change, files will be migrated to new formats and systems. Although the particulars of this change cannot be anticipated, the preservation of a high-quality digital master is the best guarantee against unnecessary future re-digitization, if, indeed, the analog object survives some 50, 100, or 200 years into the future.

The inevitability of migration is one of the reasons why it is essential that digital objects and associated metadata adhere, as much as possible, to widely accepted standards. Any deviation from the standard should be documented in the project documentation, the project code, and associated metadata.

Digital collections will also, necessarily, be migrated, as technologies, and our expectations of digital collections change. One cannot presume relationships in a particular configuration of objects will be maintained in future re-configurations, thus objects and their associated metadata must be able to exist apart from the explicit relationships created by the collection interface, indices, and database.

4.11 AUTHENTICITY, CHAIN OF CUSTODY

Master files should represent the most faithful reproduction of the original materials possible. This file should be accompanied by documentation showing how the files have been protected from tampering or other undocumented change and the chain of custody of the objects. This documentation should include the technical procedures used to retain quality, integrity, authenticity, and reliability.

As the time period increases between an object's initial digitization, as the creators of the original object are no longer employed by the library, and as institutional memory fades, documentation of authenticity and chain of custody become even more essential as part of the record. As mentioned in the preceding section on migration, this information should not simply be relegated to documentation which may become separated from the objects to which they refer but should be embedded within the object and its associated metadata to the fullest extent possible.

5.0 User-Centered Design

5.1 OVERVIEW

In user-centered design, the needs of end users receive careful consideration throughout the life of a project. The process stresses the importance of considering audience at each stage of development and integrating user-centered design methods into the workflow. The philosophy behind the process is that teams must employ user-centered design methods is simple: projects must learn about users if they are going to succeed in building sites and applications that meet audience's needs.

The following is an overview of user-center design methods. To truly adopt this approach, the team should determine which combination of methods is most appropriate for each project. Whenever possible, usability testing should be part of the project.

The methods are divided into four main categories: involving users in the design work; defining users and their needs; evaluating sites based on usability principles; and testing for usability.

5.2 INVOLVING USERS IN THE DESIGN WORK

5.2.1 Participatory Design

Participatory design is when representative end users join the design team. Often this method is used only during the phase when the initial prototype is developed. The benefit is that it includes people on the team who think as users rather than as developers. One danger of this method is that users often becomes so aligned with the team as a whole that it is difficult for them to continue to bring an outsider's perspective.

5.2.2 Personas

A persona is a user archetype that helps the team to make decisions about a site's design and functionality by giving team members a concrete individual to think about during the design process. A persona is a fictional person that the team creates to reflect what is known about one of the key audience groups (sometimes that knowledge is gained from interviews, focus groups, or surveys). Typically, a team creates two or more personas to represent different audiences, while identifying one as the primary persona.

Helpful persona profiles include demographic information, levels of computer expertise, descriptions of the personas' needs for the particular site in development, and the goals and tasks they would have in mind when using the site.

5.3 CARD SORTING

Card sorting is a method for getting users to participate in determining the organization of a site, or sections of a site. Card sorting is useful when there is a lot of information to categorize and it is unclear what categories will make most sense to users.

A facilitator gives users (individually or working in teams) a stack of 50-100 index cards, with one topic from the Web site written on each card. Users group the cards based on which topics they think logically belong together. At the end, they name the categories they created.

5.4 DEFINING USERS AND THEIR NEEDS

5.4.1 Focus Groups

A focus group is a moderated discussion with 8-12 representative users. A moderator follows a script to ask specific questions. Focus groups are most useful early in the development process to answer broad questions about what users want from a site or to illuminate characteristics of users. They are not good for evaluating details of a design or determining if a design is usable.

5.4.2 Individual Interviews

Interviews are one-on-one discussions with individual users. They are similar to focus groups in that they are best for learning about user's characteristics and the broader desires they have for the site. They allow the opportunity to ask deeper follow-up questions without negotiating the group dynamics that are present with focus groups.

Interviews may be appropriate if the main users of the site are not members of the UM community but are available to meet with in person on occasion (at conferences, for example), or via telephone.

5.4.3 Surveys

Surveys are a standard set of written questions delivered by Web, e-mail, or traditional mail. They can reveal users' preferences and desires. They do not provide the opportunity to ask for clarification of responses, as is possible with focus groups and individual interviews.

Surveys are most often used early in the development cycle to better understand users, but may be used at any point, including after a site is launched to determine if the site is met with a positive response. Surveys are appropriate when a large sample size is important or when the majority of the users are not available in person.

5.5 EVALUATING SITES BASED ON USABILITY PRINCIPLES

5.5.1 Expert Evaluations

An expert evaluation is conducted by a usability specialist who is not involved in the project. Specialists rely on research-based usability principles and their own experience to evaluate the site. While experts can identify many usability issues, they are not able to fully predict how usable a design will be to the actual users of a particular site.

5.5.2 Usability Audits

A usability audit (also called a heuristic evaluation) is a comparison of the site's design with a checklist of standards. The standards can be based on a combination of general usability principles and specific principles for DCR projects. This is an inexpensive way to discover usability issues in a design. A few people evaluate the site individually and then compare notes to come up with recommendations. This can be a good way to tighten up a design before starting usability testing.

5.6 TESTING FOR USABILITY

5.6.1 Usability Testing

A usability test is a one-on-one method in which a facilitator gives a target user a set of tasks to complete using the site. This is a "think out loud" method in which the facilitator urges users to explain what they are thinking while using the site. Unlike focus groups, surveys, interviews, and other user information gathering methods, usability testing provides the opportunity to observe users directly interacting with a site.

Even if the target users of the site are too far away to make usability testing practical, it is often helpful to do local usability testing. If the subject matter of the site is not so arcane that it is only meaningful to the target users, local stand-ins can provide information on basic usability

Usability testing is an iterative process. The best approach is to run a session with 4-6 representative users (testing them one at a time), identify problems, improve the interface, and test again.

To be most effective (including cost effective) usability testing needs to be part of the entire development cycle. Ideally, testing begins at the earliest stages. Often this means testing prototypes before any markup is written or functionality is added. Identifying usability problems at such an early stage, enables the team to make changes before any staff hours are spent building the interface. For a list of steps in usability testing, see Appendix III: Steps in Usability Testing.

5.7 CONCLUSION

Most projects will benefit from using a combination of user-centered design methods. During the early stages of project planning, teams should determine which methods are practical and will best serve their goals. All the methods described here will elicit information about users' needs. However, to determine if people can actually use a site under development, it is important to watch people interact with that site. For this reason, usability testing should be included in projects whenever possible.

5.8 FURTHER READING

5.8.1 User-Centered Design

Usability.gov: Your Guide for Developing Usable and Useful Web Sites. http://usability.gov/

Usability Net. http://www.usabilitynet.org

5.8.2 General Usability

Health and Human Services Dept. (U.S.). *Research-Based Web Design & Usability Guidelines*. Available online: http://www.usability.gov/pdfs/guidelines.html

Krug, Steve. *Don't Make Me Think: A Common Sense Approach to Web Usability* (Second Edition). New Riders, 2005.

Nielson, Jakob. *Homepage Usability*. New Riders, 2002.

Nielson, Jakob. useit.com: Jakob Nielsen's Website

5.8.3 Usability Testing

Dumas, Joseph and Redish, Janice. *A Practical Guide to Usability Testing* (Revised Edition). Intellect, 1999.

Goto, Kelly. Usability Testing: Assess Your Site's Navigation & Structure. Online: http://gotomedia.com/downloads/goto_usability.pdf

Rubin, Jeffrey. *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests.* Wiley, 1994.

6.0 Web Authoring Guidelines

While DCR may cover a wide range of activities from digitization of collections to the creation of access tools or archival technologies, nearly all will require the creation of a supporting web presence as an organized and professional public face for the project. The web presence should be considered a component of the project itself as it is the main vehicle of transmission for the collection materials while providing important contextual information and background to users unfamiliar with the collection or project.

While many projects contain identifiable UM Libraries' branding elements as part of their overall site design, other projects may be more collaborative and ultimately reside outside the UM Libraries' web domain. This document is meant to augment the practices and procedures outlined in *UM Libraries' Web Best Practices*. It is strongly suggested that this document is read in addition to these guidelines before designing web content for a digital project.

These guidelines will cover three basic areas: content, design, and administration. While these three areas overlap in many ways, thinking of them somewhat independently can help to organize the development process and ensure that all these areas are addressed.

In addition, these guidelines cover the creation of both dynamic and static web content. Dynamic content is requested from a database via a special scripting language written into the coding. When the information is found in the database it is then translated into code and displayed. This dynamic content is refreshed every time the page is refreshed or a specific request is made (such as initiating a search, or updating the checkout page of a web vendor). Static content is that which doesn't change — the portions of a web page that will look the same every time the page is viewed. Both of these types of content can be fit into these best practices guidelines, remembering that dynamic content will need to accommodate variable content while static content will remain the same.

6.1 CONTENT

As with any type of information product, consideration of audience plays an integral role in defining not only the actual content but the organization of information (see section 5.0: User-Centered Design). Although web publishing facilitates much more universal access to materials than more traditional outlets, writing content for a specific audience, or for a hierarchy of audiences (for example, gearing most content towards students, with some portion aimed at faculty) keeps that content focused and comprehensible.

For example, a digital collection may be created highlighting Maryland history using materials from a collection on campus. The digital collection may include only a small portion of the materials in the analog collection, which the archival staff may anticipate will increase requests for the physical holdings among a regional scholarly audience and the portion of the general public involved in Maryland historical matters. A listing of hours of operation for the archive housing that collection then might be of importance to those audiences. These concerns should then influence the design of the structure of web pages, placing the information about hours in an easily findable location.

The web environment adds another layer of sophistication to structuring content and organization. The hyperlinked, non-linear nature of web pages affects how information is organized. In the scroll-down web environment, important information should appear closer to the top of the page, in a prominent position, where a user is more likely to see it before moving on to another link.

In the same sense, when thinking of the cluster of interlinked web pages that make up a web site, important pages should be easily navigated to from any page in the site or may be added to the navigation bar that appears on every page. Although the web pages may be organized into a hierarchy of layers (a home page with links to four main pages, which in turn have links to more specific information) users should be able to easily get to other sections of the site. Think of an example of an online media store. A highly important function, such as searching for a specific product, should be available to users as soon as they reach the site, whether that is through the home page or some other page retrieved by a search engine. Users should not have to click through a few pages listing new arrivals, shipping policies, or featured items before they are able to search for items they wish to purchase.

It is important to remember as well that web sites do not exist in a vacuum: they exist in the increasingly more established but still fast-changing and un-policed world of the Internet. Studies by the Stanford Persuasive Technology Lab have shown that users often judge the authority of resource based on its design, aesthetics, and organization rather than its content (Fogg 2002). Establishing authority through aesthetic, logical design as well as clear statements of responsibility, date of creation, and method of contact lends credence to content and reassures users that they can trust the information being relayed.

Other factors to consider when designing web content are as follows.

6.1.1 Density of Text

Study after study shows that users often do not read dense blocks of text online. Brevity is the soul of the web. The more that can be done to reduce the amount of text on the page and break it up into smaller chunks through bullets, short paragraphs, or layout and design, the more likely it is that users will be able to get the information they need from a resource. Obviously, many digital projects will be full-text conversion projects that may include essays, poetry, or even full-length books. When displaying these materials it may not be possible to break up textual content, but it can be presented in a less overwhelming manner through layout (see section 6.2: Design).

6.1.2 Independence

It is best to keep the information on each page focused on a single concept and introduce new themes or ideas on their own pages. The content of each page should be, in some sense, independent of the other pages on the site. At the same time, elements of navigation and design can give the user a quick overall context as to where in the site they are. For example, the names of the links in the navigation bar should correspond with heading at the top of page content. Breadcrumbs could be used to convey where the page being viewed sits in the hierarchy of web pages in the site. Logos, identities and statements of responsibility in the header or footer also give users a sense of the context in which they are viewing the individual page. All of these techniques help users remain oriented to the navigation of the site as well as provide context to those users that may have arrived at a particular page through a search engine or external link and did not necessarily begin navigating through the site from the home page.

6.1.3 Depth vs. Breadth

The drive to create a site of brief and independent web pages can have the unfortunate side effect of becoming too unwieldy. This is where good organization comes into play. Pages should be separated into hierarchical groups and presented with a logical structure. Navigational links on the home page and in the header or footer of interior pages should be as concise as possible without being too shallow. If a site has an abundance of pages, navigational links can be simplified with techniques

like roll-out menus or by listing links to specific resources on separate pages linked to the home page. But be careful, users may miss these extra links. Important information should be clearly linked and users should not have to navigate through too many links to get to it.

6.1.4 Parallelism

Navigational links should appear prominently on the home page as well as in header, footer, or sidebar navigation on interior pages. These links should be presented consistently throughout – the same words in the same order linking to the same information. The consistency of links, along with a consistent design and presence, is the glue that will form a series of web pages into a coherent web site.

6.2 DESIGN

Good design covers two basic objectives: aesthetics and accessibility. In addition, good design will be consistent and make future maintenance or additions to the site easier.

The term "accessibility" is commonly used to refer to accommodating those with disabilities or impairments through design. Section 508 of the Federal Rehabilitation Act was created to ensure that access to government information was not impaired for people with disabilities. Since being added to the Act in 1998, these requirements have become a de facto standard in web publishing. The government publishes a large amount of information relating to section 508 and its implementation on its web site (http://www.section508.gov/) including links to free tools verify that designs meet accessibility standards. Other sites that can help web authors incorporate accessible designs areW3C's Web Accessibility Initiative (WAI) (http://www.w3.org/WAI/Resources/) which also includes a list of tools for verification.

In addition to accessibility for those with disabilities, web authors should also consider the need to accommodate those with different resources and online connectivity. Like content decisions, the intended audience as well as the type of content is important in this consideration. The IRS, for example, spends a lot of time and effort to make sure that their online forms and applications are available to those with slow connections or older browsers since their intended audience, the American public, uses a wide spectrum of harware and software. On the other hand, if a project's main objective is to create an interactive resource with animation and gaming components, users with slow connections, older browsers, or unwilling to download specific commercial plugins like Flash or Quicktime may not be able to access it. The project has to weigh the importance of the limitations of audience resources, while striving to make as many components as possible simple and usable on multiple systems.

In addition, designs should be tested to ensure that they work on major browsers and platforms (for example, Internet Explorer on a PC implements designs differently than on a Mac; Mac platforms come with the Safari browser, while PCs do not). All browsers implement HTML and style elements slightly differently, and designers should be browser-independent as much as possible. Project managers may need to decide what browser versions and platforms will be supported if serious problems arise with browser compatibility.

6.2.1 Using CSS instead of writing style information in the code

Designing with an external cascading style sheet (CSS) instead of writing styles in the code has several benefits. On a theoretical level it divorces content from style. This makes it possible to alter one without affecting the other. On a practical level, this means that style changes only need to be

made once to affect an entire site introducing a level of consistency in style across pages. Templates for pages can be created and new content can be added quickly without having to recode all the style information. It also means that if the CSS is suppressed (and many users with visual impairments may prefer to view pages this way) the web page is still intelligible because the content appears in the order it should be read and headings, paragraphs, lines, etc. are all indicated correctly.

Using CSS for design means that <table> tags should only be used for the presentation of tabular information, not for layout. If information needs to be laid-out in specific areas of the page, tags should be used to indicate the blocks and their positioning should be indicated in the CSS. The s should be written into the page in the order in which they should be read.

6.2.2 Javascript server side includes

Using javascript to write server side includes, particularly for headers and footers, helps maintain consistency and eases maintenance of site pages. When using a server side include, files containing information that will be repeated on every page are written and stored. For example, a file might be written for the site header, or footer, or a sidebar that is repeated throughout pages. These are include files. They are written in HTML, or some web-readable code. Other pages are then written containing only a javascript reference to the include file where that information would occur. When these pages are viewed through a browser, the javascript code instructs the browser to find the include and display it along with the rest of the document.

Writing one file that will then be included in every page means that there is no chance that inconsistencies in the information will appear and changes, when needed, only have to be implemented once. The process of creating new pages is also easier since the included elements do not have to be recreated on every page.

6.2.3 Masking emails from bots or spiders

Email has become the primary form of contact for most professional matters. Having this information available on public websites, however, greatly increases the likelihood that the address will be harvested by some automatic spam script. To help avoid this, email addresses should always be listed in some way that they are understandable, but some manual intervention is needed before they can be used such as "Jsmith[@]umd.edu" or "JsmithATumdDOTedu."

Other techniques for masking email addresses from automatic harvesting, include using an image of the text for the link instead of the text itself and using a server script or JavaScript to avoid having a full email address in the code.

6.2.4 Validating code

Once the buildout of a web page is complete, code should be validated against the W3C code validation service (http://validator.w3.org/). This will ensure that all code is correct and adheres to W3C standards and recommendations.

6.3 ADMINISTRATION

Once a clear goal is formed for a web site and many of the content and design issues are worked out, site development will begin. Often the development and design will be an iterative process with layouts and ideas developed and tested and perhaps put aside for other layouts and ideas. When

development begins, proper guidelines for file naming and directory building should be followed. These issues are discussed in some detail in the *Web Best Practices*.

The maintenance of a web presence for a digital project will have a life long after the end of the original project. Resources should be committed to keeping these pages current. Links should be periodically checked as external resources might move or be taken down. Site content also may need to be updated on a periodic basis. If the digital project is seen as a one-time digital exhibition or product, further updates to the site may not be a part of the project design. More often, however, a site will need to be updated and adapted as more materials are added, scholarship in a field increases, or as developers adapt their content to an evolving user base.

6.4 BECOMING A WEB DESIGN EXPERT

These guidelines are intended as a jumping off point for the development of HTML pages to support digital projects. More detailed information about applying these concepts can be found by consulting a few resources.

The World Wide Web Consortium (W3C) (http://www.w3.org/) is an international organization committed to creating web guidelines and protocols. Their web site hosts numerous markup languages, protocols, and guidelines including HTML and CSS. They also host the Web Accessibility Initiative committed to enhancing access to the web for those with disabilities.

Section 508 (http://www.section508.gov/) is a government website dedicated to helping government agencies create web sites compliant with federal accessibility regulations. It contents information, guides, and tools to help all web authors create accessible design.

The Web Style Guide, 2nd Edition (http://www.webstyleguide.com/index.html) is a comprehensive and thorough guide to organizing content and writing good code for the web.

Stanford Web Credibility Research (http://credibility.standford.edu) brings together the research of Stanford University's Persuasive Technology Lab which has extensively studied how users determine what to believe and what not to believe on the web. The site includes research articles, tips for design and evaluation, and an extensive bibliography of other resources.

When designing pages for the University of Maryland Libraries, further guidelines can be found in the *Web Best Practices* (http://www.lib.umd.edu/itd/web/bestpractices/index.html) authored by both the Web Administration Committee and Web Services in the Information and Technology Division.

7.0 Image Collections

Image-based collections will present specific issues not pertinent to the creation of collections of other media types. Technical aspects of image projects require significant consideration and can greatly influence the outcome and workflow of projects. While image standards for digitization have a longer history than other media, varying standards still proliferate and are greatly influenced by project goals. The following sections outline various areas for discussion and decision before embarking on an image digitization project. These guidelines should be used in combination with the guidelines of the previous sections when planning image collections.

7.1 FILE FORMATS

Image formats are numerous and serve different purposes. Generally the same format will not be used as an archival master and as web deliverable because of differences in file size and because browsers can only display view certain formats. Archival masters should retain the largest amount of information possible and result in very large files that are not easily delivered to end users. Smaller files are usually compressed. This process reduces the information in an image file by selectively discarding pixels according to an algorithm calculated to make the loss less noticeable. Although the loss of the pixels may not be apparent to users, any future usage that would have required that information, such as creating an enlarged version, would be impossible.

General requirements for file formats for master images are:

- Ablility to preserve the resolution, bit-depth, color support, and metadata of a very rich image file
- Ability to be stored in uncompressed or compressed format using both lossless and lossy techniques
- A format that is open (non-proprietary) and well-documented, widely supported, and cross-platform compatible (Kenney and Rieger 2000)
- Ability to support derivation of access copies (NISO 2004)

Tagged Image File Format (TIFF) meets most of these requirements and remains the de facto standard for the following images: textual, graphic illustrations/artwork/originals, maps, plans, oversized, photographs, aerial photographs, and objects/artifacts. TIFF version 6 with Intel (Windows) byte order is recommended (NARA 2004).

The TIFF file is used as the archival master file (high resolution, high quality files from which surrogates are derived). Archival practices dictate that images may be corrected slightly to combat the natural degradation that a scanner will introduce and still be considered uncorrected. Inherently low quality images that are corrected to the point that they no longer reflect the original are considered corrected. Corrected files may be used as a submaster or as a surrogate. A digital project may decide to retain uncorrected or slightly corrected masters for archival purposes and create heavily corrected files for project use.

Smaller, web deliverable surrogates are usually presented in the Joint Photographic Experts Group (JPEG) format or Graphic Interchange Format (GIF). In general, the JPEG format is best for full size or mid-size surrogates while GIF is used for thumbnails due to its limited color palette options (more

about color palettes and modes can be found in section 7.2: Color Mode and Bit Depth). For more information on different data types and file formats, see Appendix IV: A Typology of Formats.

The JPEG2000 format may also be considered a master file in the future. This file can be either compressed or uncompressed and has a richer array of metadata embedded in the image file. Full browser support is not yet available, however, and the community of practice has yet to adopt it as a standard.

Regarding the status of standards for preserving electronic information: while use of guidelines produced by organizations like NARA or NISO should result in the creation of digital objects with potential for long term maintenance, standards for digital preservation have not yet been developed. Even the Association of Research Libraries' published position paper, "Recognizing Digitization as a Preservation Reformatting Method" (2004), highlights the uncertainty of digital preservation by listing the variety of digital preservation strategies still under development and the organizations with work underway on digital archiving (Atthur et.al. 2004). For good reason, there is not yet agreement in the preservation field on proven mechanisms for preserving digital information.

7.2 COLOR MODE AND BIT DEPTH

Assigning a color mode or profile to an image controls the way color information will be recorded and stored. The sophistication of the color palette that can be achieved with a given color mode can affect the color output of the digital image. For best results, this decision must be made before the digital image is created. Changes at a later stage can result in noticeable differences in values as they are translated from one color mode to another. Color modes fall into different categories based on the number of color values expressed and the amount of information used to express each value.

On one end of the spectrum are black and white or duotone profiles, which use only one bit to express that a value is either on or off, usually expressed as black or white (a more thorough explanation of how bits are used to express color is contained in the following section). This very simple color mode would only be suitable in instances such as page scans of modern books where rich physical details are not as important as clearly readable text.

More complex are the 8-bit grayscale and indexed color modes. These palettes offer 8 bits of information to express a color value, leading to a possibility of 256 colors or values. Some caution should be used with creating grayscale images, as many images that appear grayscale may actually just be a monochromatic image in another value, a shade of a bluish-gray, for example, rather than a true black to white scale of grays. A prime example of this phenomenon would be a sepia toned photograph. In other cases, black and white photographs may fade or contain other subtle color variations that would not be captured in a grayscale digital format. Therefore, use grayscale only when the image will not need retouching or may not be kept for archival purposes. If any retouching is anticipated, it is best to scan the image in a color profile and then convert to grayscale later if necessary. This will provide the best result in the retouching or editing process. Indexed color profiles are essentially limited and should only be used with simple graphics or small thumbnail files.

The most complex color modes are RGB and CMYK. Both of these options offer millions of possible color values. RGB stands for **R**ed, **G**reen, and **B**lue, the primary colors of light, and CMYK for **C**yan, **M**agenta, **Y**ellow, and blac**K**, the primary colors used in printing. Not all scanners and digital cameras will be able to create CMYK images, as this profile is mostly used for off-set printing. Monitors display using RGB color mode by default.

In addition, spot colors, such as the Pantone™ color systems, may be supported by various applications. Spot colors are special colors that do not fit into the other color modes. They should only be used when an outside vendor requests them for printing.

7.3 BIT DEPTH

Bit-depth refers to the number of bits that are used to express each pixel in the image. The simplest images use only one bit (1 or 0) to express either black or white, for each pixel. When color comes into play more bits are needed per pixel. Every color value is made up of a combination of red, green, and blue (in RGB color) or cyan, magenta, yellow, and black (in CMYK color). The simplest color profile is 8-bit, which uses eight bits to express a single color, leading to 256 possible values for each pixel. Color modes other than indexed color usually use 24 or 48 bits to express a color value, combining either 8 or 16 bit values for red, green, and blue to express a single color. The difference between 24 and 48 is misleading however: 48 bit color simply contains a duplication of the information in each pixel: two sets of identical bit values for each pixel have been included in the file.

The main importance of bit depth is to enable retouching and other file manipulations. Bit-depth redundancy does not serve any true archival purpose, but when using a program like PhotoShop to retouch or resize images, the final result will look smoother and more natural when the image has a higher bit-depth. When surrogates are made from an archival master, a higher bit-depth is preferred to prevent JPEG artifacts from appearing.

Recommended bit-depth and color profiles can be found in Appendix V: Minimal Requirements for Creating Digital Images.

7.4 RESOLUTION AND FILE SIZE

Resolution is measured in dots per inch, or dpi. Dots roughly correspond to pixels, which are the smallest units a computer monitor can understand. Common theory holds that the higher the resolution, or dpi, of an image, the better the quality of the image will be. This is partly true, but also overlooks the importance of file size, image size, print or display quality, and the overall number of pixels, all of which play a part in the quality of images.

Standards proliferate, but most archival quality image capture occurs at a measurement of at least 4,000 pixels on the longest side, and can go as high as 6,000 pixels. For an 8x10 image that would correspond to a range from 400 dpi to 600 dpi. For a smaller image, such as a slide, the resolution could be as high as 2400 dpi. Refer to Appendix V: Minimal Requirements for Creating Digital Images for suggested scanning resolutions for various formats and sizes.

7.5 QUALITY CONTROL, TESTING, REFERENCE TARGETS

Quality control (QC) is an integral component of creating digital content that will retain value and utility over time. QC encompasses procedures and techniques to verify the quality, accuracy, and consistency of digital products. The scope of the inspection should include not only an evaluation of the quality of the digital image files, but the accuracy and consistency of metadata, as well as the integrity of the storage media. The extent of the inspection must also be considered: a sampling frequency must be established for each aspect of the QC program. Recommended frequency is 100% of all image files and accompanying metadata; minimal requirement is 10% sampling of each image/metadata batch. The extent of inspection for each project should be decided in conjunction with DCR and the Preservation Department.

7.6 TYPES OF INSPECTION

The key factors in image quality assessment are resolution, color, and tone, and overall appearance. QC can be conducted by visual inspection of images on-screen or via printouts, although it is important to note that quality assessment, especially for tone and color, may be highly subjective and changeable according to the viewing environment and the characteristics of monitors and printers. The viewing environment and all links in the imaging system (including the scanner, monitor, and printer) should be carefully controlled. Image quality can also be judged through the use of technical targets (resolution, tone, and color) and increasingly through software.

Criteria for evaluation of digital images should be determined in advance of the digitization process. These evaluations should include technical targets to ensure that image projects meet a level of quality and suitability for future use. Quality assessments should cover the following areas:

- Completeness of the digital capture
- Format, compression, color mode, bit depth, and size of files
- Adequate capture of the proportion, color, orientation, and scale of the original items
- Correct and adequate metadata that consistently meets the standards created for the project
- Image quality, including tone, brightness, color accuracy, saturation, noise, artifacts, detail, sharpness, flare, and overall resolution.

Specific guidelines for carrying out these evaluations are found in Appendix VI: Quality Control for Images. However, no quality control process should be carried out without extensive consultation with DCR and the Preservation Department.

8.0 Text Collections

8.1 DOCUMENT ANALYSIS

Creating electronic full text presents a different set of challenges than creating digital images. Any full text project should begin with an analysis of the text to determine whether or not it is suited to digitization, what the best mode of digitization is, and the kinds of digital documents that can be created. The analysis should be based on factors such as the type of information contained within the text, the state of the original, the anticipated uses of the digital text, and available resources.

Another consideration is how the options of scanned images, rich markup, or a combination of the two fit with the project's goals and objectives. Is it important to provide a visual representation of the original text, or is encoding the text to make it more richly searchable a priority? Once a document is scanned it consists of an image or a series of images, and the words within that image cannot be searched since it is merely a snapshot of a document. Encoded (or marked up) text, however, can be searched, and the level of markup can provide different access points for searching. Encoded text alone, however, will not allow users to see how the original document looked. A project to digitize a 13th century illuminated manuscript may provide only page representations if the purpose is to present the artistry of illumination. Alternatively, a plan to digitize paper finding aids will favor encoded text, as searching the content of the document is more important than providing visual representations of the paper-based finding aid. Projects that feature both valuable text and visual information may deem both text markup and visual representation to be equally important.

In addition to these factors, the costs of undertaking such a project should be considered. While rich encoding may add much functionality and value to a digitized text, it is expensive in terms of time and commitment. This expense should be carefully considered before embarking on a project, weighing costs against the expected benefits of enhanced searchability and access.

These considerations must be made early in the project lifecycle as different techniques of digitization require different technologies and produce different outputs. Document analysis is used to help make these decisions. Determining the overriding features that the digital surrogate must capture, such as special formats, methods of organization, or any other visual, structural, or textual elements, can inform a decision about what types of markup to employ as well as which display features to utilize.

8.2 CONSIDERATIONS WHEN WORKING WITH FULL TEXT

8.2.1 Structured or unstructured data

An initial step to consider in document analysis is determining whether the data is structured or unstructured. Clearly, most data has some structure to it; the key is to determine whether the structure of the data is explicit or implicit. For example, data extracted from of a database or spreadsheet is structured. A computer can read such data as discrete, and the data can be manipulated based on this explicit structure. Implied structure, on the other hand, is found in cases where a human surveying the data recognizes and understands the structure, but a computer working with the same data neither recognizes it nor can it be manipulated in its raw form. In the case of unstructured data, such as digitized full text of correspondence, markup will provide a

framework around that data. This framework then allows a computer to understand and manipulate the data for display and search purposes.

8.2.2 Corrected and uncorrected OCR

When investigating scanning, the process should involve an evaluation of image-based scanning and Optical Character Recognition (OCR). OCR scans can provide a digital image as well as digitized text. OCR involves character recognition software that has an internal store of character shapes and when a digital image is created, this software converts the character shapes (text) found in the digital image into ASCII text. OCR scanning is most effective when used on clear copies of printed text created by a printer or printing press with standard fonts, a factor which tends to favor materials from the 20th century forward. Most current OCR software can also convert less-than clear copies, older, non-standard printed fonts, and even handwriting, but the error rate rises significantly with every factor that distances material from clean copies of standard fonts.

If OCR is a viable option for a project, a decision must be made between uncorrected or corrected OCR. Uncorrected OCR text is the software-derived digitized text as is, without correcting errors. This is a reasonable option when dealing with original text that meets the highest specifications for clear copy and standard font, as well as for full text that is utilized as an additional benefit to a project with another focus. If the project involves digitizing a large quantity of materials, checking for errors may prove too costly or time-consuming. For instance, corrected OCR is probably not a reasonable option when digitizing the entire run of a newspaper, not simply because of the scale of the project, but because of the higher tendency towards errors in the original text. In cases where literary or documentary editions are the focus, corrected OCR should be considered. In addition to the option of manually correcting OCR text by reading through and fixing errors, there are automated processes that can be used. Correction software can detect and correct OCR errors by using dictionaries, databases of common OCR errors, and study the distribution and use of other words throughout the document as a basis to determine errors.

8.3 FULL TEXT MARK UP

Text encoding is the gold standard for creating richly searchable full text. Text that has been encoded has been identified as having particular characteristics that will be expressed in search and display options. In an encoded document, the text is "tagged" with metadata within angle brackets: names, dates, geographic places, etc. are labeled with XML tags which end users typically do not see. However, the tags allow the search engine to narrowly focus queries. If structural data has been encoded (such as chapter sections, paragraphs, poetry stanzas, etc) users will be able to further narrow their searches to a particular portion of the document, for example, searching only the titles of poems.

There are several full text metadata standards that serve different purposes. Two of the most often used in the library are Encoded Archival Description (EAD)[1], which creates richly encoded finding aids, and the Text Encoding Initiative (TEI)[2], which was developed to encode literary and documentary texts, oral interviews, dictionaries, and bibliographies. Other schemas exist that may be suited to specialized materials, although they may be infrequently used in Library projects. For example DocBook[3], is used to describe books, articles, and technical documentation, and MATHML[4]

[1] http://www.loc.gov/ead/
[2] http://www.tei-c.org
[3] http://www.docbook.org/
[4] http://www.w3.org/Math/

is used for complex mathematical equations. See Appendix VII: XML Examples for an example of a TEI-encoded book review and an EAD-encoded finding aid.

8.4 HEADER VS. BODY

The specific details of TEI and EAD will be discussed in subsequent sections. Both standards, however, structure data into two major divisions, the "header" and the "body." These two sections allow for inclusion of the intellectual content and metadata in the same information package, "improving its identifiability, searchability, and interactivity" (Nellhaus 2001).

The header (in TEI this is captured in the <teiHeader> tags, in EAD, the <eadheader>) contains bibliographic information about source text, as well as the electronic document being created. A header usually includes information to identify the electronic document, its creator and publisher, bibliographic information about the original text, information about the encoding process and standards used, and a list of revisions to the text. Some parts of the header are required, others are optional, allowing the encoder to decide what level of metadata to provide.

The "body" of a TEI or EAD document is the version of the text that is being encoded augmented by information in the metadata. This metadata may include notes, scholarly editing, regularization of names, as well as the tags which represent structured divisions of the text, such as paragraphs and chapters. In TEI this information is captured within an element called <body> with many child elements for textual features. The <body> element is one of three major divisions of the <text> element which also includes <front> and <back> elements to encode information from the front or back matter. In EAD, an explicit <body> element is not used. Instead various sections are encoded separately immediately following the <eadheader>.

Dividing the text into these sections allows for efficient information retrieval, as only the original document text may be searched by utilizing the <body> element in a TEI document, for example, or ignoring the <eadheader> in an EAD document.

8.5 ENCODED ARCHIVAL DESCRIPTION

8.5.1 Finding aids and their purpose

Encoded Archival Description (EAD) was created to assist in the creation of electronic finding aids. A finding aid is a standard tool used for resource discovery in archival repositories. Archival collections often contain large amounts of materials which are best described at collection and sub-collection levels, as opposed to item description. For instance, while researchers interested in Theodore R. McKeldin may benefit from individual description of each item in a large collection of his personal papers, such item description is likely not a cost effective model for the repository trying to make that collection available. In trying to balance between access and description, typically archivists arrange the collection into smaller portions, arranged by format, subject, names, geography, or by other defining characteristics relevant to the collection. Furthermore, those portions may be further subdivided as appropriate, until the collection has been arranged in a coherent manner that will allow users to identify, at least on a macro level, its contents.

8.5.2 Arrangement of a finding aid

After the collection has been arranged in the most appropriate manner, the arranger then creates a finding aid, which is basically a textual representation of the coherent arrangement of the collection. The finding aid allows users both to identify specific material of interest as well as to get a general

sense of the collection as a whole. Finding aids are created to guide users through the collection as it is intellectually arranged rather than how it is physically housed. For storage purposes, a depiction of McKeldin Library drawn on oversized paper may be kept in a different physical location from a depiction on standard sized paper. In the finding aid, however, the user should find an entry indicating "drawings of McKeldin Library" and be led to both drawings from that entry, rather than being required to look in two finding aids or in two different parts of one finding aid to locate items that fit in traditional boxes and those that are oversized.

From the description above, it should be evident that archival arrangement is hierarchical, as large collections of materials are divided and subdivided into smaller, discrete groupings. Context would be lost if a sub-group was separated from its parent. Capturing the structure of the archival arrangement is integral to accurately representing that arrangement in text format. Created from within the archival community, EAD is a metadata standard that allows for the structural hierarchy of archival finding aids to be retained and displayed in a digital representation. By tagging in EAD, the finding aid becomes more richly searchable than is possible in a flat text document. The tagging allows users to limit searches to certain portions of the finding aid, such as the biographical and historical note or the contents listing, as well as limiting the search to subject, name, or geographical location, and so on. Thus the structure that provides the context for the material is retained while adding powerful full-text search capabilities that allow users to pinpoint the information they are seeking. Keep in mind that encoding with EAD may be an excellent way to make nontraditional finding aids such as inventories, registers, or listings available and searchable.

8.5.3 Levels of description

When preparing a finding aid, some consideration should be given to the level of description. It is possible to create a rich and informative EAD document that only describes at the collection level. Even at this level, the user can still find information about the source(s), scope, and basic content of the collection. If greater description is desired, the container levels can be used to hierarchically divide and subdivide the collection contents to a very fine level, including item level description. Cost-effectiveness and relative usefulness are factors in determining the generalness or specificity of the collection description.

8.5.4 EAD and MARC

Many archival collections are also represented in the library catalog through a MARC record. It should be noted, however, that while a MARC catalog record can be an excellent companion, it is not an appropriate substitute for an EAD finding aid. This is due in part to the difference in the conceptual level at which each metadata language exists. Archival description encompasses several different conceptual levels, whereas bibliographic description (represented by the MARC record) exists on one level. While both are metadata schemata designed to create a surrogate for a variety of materials, they do that task quite differently. The EAD finding aid creates a surrogate that is the equivalent of a model replica of the materials. The user can see the material as a whole, as well as get an in-depth glimpse into the structure and complexity of the material. In other words, it remains true to the structured arrangement. Alternatively, the MARC record provides the equivalent of a photograph of the material. The user can see the material as a whole, as well as a glimpse of the description, but the MARC record is flatter and less complex that the EAD surrogate. The MARC record can provide broad access to the collection's existence on the general subjects and names that apply across the collection, but will not be effective in allowing users to see more specific collection contents.

8.5.5 EAD@UMD

EAD finding aids for collections at the University of Maryland are made available to the public through the ArchivesUM portal.[5] EAD documents to be included in this database must currently be created using a customized Microsoft Access database and an EAD Converter program.

8.6 THE TEXT ENCODING INITIATIVE

The Text Encoding Initiative (TEI) was established in 1987 to create guidelines for the creation of electronic text. It publishes guidelines which are widely used in libraries, universities, museums, and several e-book and other commercial ventures for creating XML-encoded texts. The TEI Guidelines and the accompanying Document Type Definition (DTD), a key piece to creating an XML document, offer a set of tags that encoders can choose from as well as the option to create new tags. The guidelines allow for encoders to adapt the encoding specifically to the needs of their documents, rather than conform to a rigid standard. TEI was first expressed as SGML and in 2002 expressed as XML with the P4 Guidelines. The newest version, P5, is currently under development and is being released on a continuous basis. It supports XML schemas in addition to DTDs. Schemas essentially function as DTDs, but are expressed in XML, unlike DTDs which are written in a different language.

TEI documents consist of two main sections: a header and a body (for a description of the function of these two divisions, see the section 8.4: "Header vs. Body"). In addition, the TEI header may be used alone to create bibliographic records for use in a database. TEI DTDs, Schemas, and guidelines can be downloaded from the TEI Web site.[6] Customized P4 DTDs can be built using the TEI's "Pizza Chef."[7] TEI P5 DTDs and Schemas can be built and customized using "Roma."[8]

8.6.1 Encoding Levels

If the TEI is chosen for a project, some consideration should be given to the level of encoding desired. In 1999 a task force of representatives from six libraries issued "TEI Text Encoding in Libraries: Guidelines for Best Encoding Practices" (Friedland et. al. 1999), a document detailing different levels of encoding appropriate for different projects. Encoding levels range from fully automated conversion with no manual intervention (level one) to the creation of texts requiring "subject knowledge" to "encode semantic, linguistic, prosodic or other elements beyond a basic structural level" (level 5). Most of the levels include some automated and some manual input. The complexity of the hierarchical structure and the amount of textual elements identified and captured will determine the level of encoding used. For a full description of encoding levels, please refer to "TEI Encoding in Libraries: Guidelines for Best Encoding Practices."[9] A synopsis of the levels is included below:

> Level one is defined as fully automated conversion and encoding. This level of encoding will capture major divisions in the text, page break, and figures. The goal for this level is for encoded text to be subordinate to the image of the page. It is intended for use in large volume encoding or in cases where keyword searching will be sufficient or no manual intervention is desired in the encoding process. Encoding at this level does not preclude doing a higher level of encoding in the future.

---

[5] http://www.lib.umd.edu/archivesum/index.jsp
[6] http://www.tei-c.org
[7] http://www.tei-c.org/pizza.html
[8] http://tei.oucs.ox.ac.uk/Roma/
[9] http://www.diglib.org/standards/tei.htm

Level two is minimal encoding used to capture navigational markers such as text divisions or headings in addition to those features captured with level one encoding. In this type of encoding the electronic text could stand on its own. Level two encoding requires some human intervention to identify elements like headings and text divisions, although other encoding can be done automatically.

Level three involves simple analysis to denote hierarchy and typography without getting into content analysis. This kind of encoding can be done manually or by a conversion from HTML or a word processing document. With more tags to indicate typographical features as well as notes, figures, and front and back matter, this level of encoding has greater latitude for different types of display and searching than the previous levels. This level of encoding can stand alone without images since typographical features are captured making delivery and storage easier.

Level four includes basic content analysis capturing both the function of textual and structural elements as well as the nature of the content in addition to how it looks. This level doesn't necessarily capture all of the semantic, structural, or bibliographic features of the text but does allow for different tagging to be done for different genres (such a <l> tags for lines of poetry or <speaker> tags for drama) and supports a more sophisticated display and searching capability. It is ideal for texts that will be used for pedagogical or scholarly purposes because of its display and search capabilities. Since this level of encoding requires significant human intervention, the project team should consult with potential users of the collection to consider the costs in terms of time and staff resources versus usefulness versus the expected benefit in terms of enhanced accessibility.

Level five represents scholarly encoding and captures the rich array of textual, structural, and bibliographic elements. This level of encoding requires subject knowledge to encode the linguistic, prosodic, or semantic elements of the text. This level of encoding offers the richest scheme of metadata of any of the levels but requires a significant investment of resources. If a very rich level of encoding is desired and level five encoding is not possible, rich bibliographic information may be provided in the document header along with a lower level of encoding in the body. This will allow for subject access to information such as regularized personal names or titles, for example, but the encoding information will only appear in the header and not be tagged in the body.

8.6.2 Collaboration

Since TEI allows for significant scholarly editing and annotation, the possibility of working with a research or subject specialist provides for fruitful collaboration. Collaboration can have several benefits including increased awareness of the project by the user population and enhanced opportunities for grant funding from agencies such as the National Endowment for the Humanities, besides the general benefit of having an expert contribute to and add value to the library's online collection.

8.7 CREATING THE ELECTRONIC TEXT

Once a decision has been made regarding encoding schemes, levels of encoding, and the goals and scope of the project, the next step is to consider the different methods of creating the electronic text. One option is outsourcing all or part of the encoding or scanning or both. The decision to outsource should be based on an analysis of the time and resources available in-house in addition to the more long-term concerns of possible damage to fragile materials or a desire to educate staff for future

projects. The ALA has numerous online resources to provide information useful in making outsourcing decisions on its Web site including guides, policy statements, and checklists, offering insight into performing cost analyses.[10]

For in-house projects, employee time and skills, in addition to hardware and software costs should be factored. Both scanning and encoding will require input from skilled staff. Conducting a test run on a fraction of the overall project should determine if in-house production is efficient or even feasible. This test run could also provide the basis for specifications if a project is outsourced, as well as a benchmark for quality control assessment.

Finally, quality control should play a major part in considerations for production. Whether the work is done by a vendor or done in-house, goals should be set to measure the effectiveness of the digitization process. With a vendor this evaluation process should be written into the initial agreement allowing ample time to evaluate the service and product the vendor has delivered after an introductory period. When work is done in-house, evaluation should occur at specified periods throughout the work process to ensure that the project is proceeding as scheduled.

In either case, a good evaluation process will include an analysis of errors, statistics, and measurements. Before work on the project commences, thought should be given to the acceptable rate of errors, the desired levels of output, and the time required to complete the project. During the first review period, these measurements may be adjusted to more accurately reflect the work process. This is not an inconsiderable portion of the project life cycle, and the specific needs of a text project evaluation should be taken into consideration during the initial planning phase.

[10] http://www.ala.org/ala/oif/iftoolkits/outsourcing/Default2446.htm

9.0 Digital Audio and Moving Images

Historians Roy Rosenzweig and Daniel Cohen have called both audio and moving images "the most complex historical artifacts to digitize." (Rosenzweig) Enormous file sizes, time and labor intensive processes, and the instability of the original object all contribute to the difficulties inherent in digital reformatting of audio and moving images. There are, however, marked advantages to the digital reformatting of audio and moving image material that can compensate for the complexity of the process. These include fragile analog originals receiving less wear and tear due to repeated use, increased remote access to the content, improved intellectual access through appropriate metadata creation and increased flexibility for future use.

9.1 Special Considerations for Digitization of Analog Audio and Moving Images

Analog audiovisual materials have unique problems that often necessitate reformatting to another medium. For example, acetate audio tapes can develop sticky shed syndrome (SSS) due to a condition known as binder hydrolysis which renders them unplayable if not caught in time. Paper-backed audio tape is extremely fragile, and tears and creases can easily lead to clicks introduced during playback. Aluminum disc recordings are always at risk because the aluminum is so soft that it can be easily distorted by the wrong size or type of stylus. Magnetic audio and video tapes can be damaged by loose or broken parts in plastic cassettes cases. A magnetic tape's signal (the information carrier) is represented on a tape by the arrangement of the magnetic particles into a particular pattern. Strong magnetic fields can affect the signal on a tape, causing it to become unreadable or adding to errors in playback. Acetate motion picture film is susceptible to "vinegar syndrome," the chemical breakdown of the cellulose triacetate film base as evidenced by a strong vinegar odor.

But the inherently unstable physical condition of many analog audiovisual materials is not the only issue that necessitates reformatting. Unlike a photograph or text, for example, original analog audio and video recordings require a mediating technology in order for the user to perceive the moving images or to hear the recorded sound correctly. The original mediating technologies – in other words, the recording and playback technologies – are themselves endangered by technical obsolescence. The machines and spare parts for the wide variety of playback equipment required can be difficult to find. Moreover, the engineering skill to run and repair the machinery is also endangered. Machines that play 2 inch quadruplex video tape, popular in the 1970s, are no longer made and require specialized skills to run and maintain. The expert engineers of this era are retiring, taking their unique and much needed skills out of the marketplace. Furthermore, playback itself might endanger severely damaged audio and moving image media if the machinery is not properly cleaned and maintained.

Reformatting at risk audio and video analog material to digital is a generally accepted practice in the library and archive community although some institutions choose to reformat to analog mediums for a variety of reasons. Some institutions are simply more comfortable in the analog arena, others lack the hardware, software and technical skills to reformat to digital in-house, and others shy away from the significant overhead needed to sustain quality digital files for the long term. In most cases, the University of Maryland Libraries supports reformatting to digital when possible. Creating a series of digital files that meet certain needs limits repeated use of the fragile original analog material, although the analog material is generally retained (see section 1.1 on Digital Masters).The goal is to create a high quality digital file from the analog source item and keep this file as a master copy from

which any number of derivative files could be created to meet access or other needs. The form of the master file will depend on project goals, intellectual property issues, and the relative scarcity of the original. Specific information on file formats is outlined.

The reformatting of deteriorated motion picture film to digital is less accepted in the library and archive community. Many moving image professionals and film archivists advocate creating new motion picture film elements (such as negative and a variety of prints for projection), often at a professional film laboratory since very few institutional facilities can do this in-house. Digital BetaCam (i.e. DigiBeta) or BetaCamSP copies are often made from the new film elements and saved as "preservation masters" from which DVD or VHS access copies are created as needed. One major concern of performing any film-to-video transfer is the misalignment of frame rates from film to video which can result in image loss or continuity issues. Another is the relative stability of the polyester film base of the new film elements as compared to the fragility of magnetic video tape. However, many institutions do incorporate direct film-to-video transfer in-house using equipment such as an ELMO transfer unit or telecine machine, primarily because new film elements created at outside labs can be very expensive. There are risks involved, of course, including poor quality image and sound capture and the lack of color correction and balancing. While the content of the film may be still accessible as a digital file, direct film-to-video transfer risks introducing some compromise in the quality of the image and sound from the original analog film.

9.2 PROJECT PRE-WORK: DIGITAL AUDIO AND MOVING IMAGES

9.2.1 Time Management Needs

Transferring analog audio and moving images works to digital formats is a time- and labor-intensive process. With rare exceptions, the master analog to digital (A-D) transfer process happens in real time. This means it takes at least as long as the playing time of the recording to create the master file. This process is usually closely monitored by trained staff. Consider in time management planning that one hour of recorded audio will require about 1 ½ hours of dedicated computer and staff time, including time spent cleaning and rehousing the item, collecting metadata, preparing playback equipment, and processing digital files. Add to this time for any needed cleaning and rehousing of the analog item as well as basic playback equipment maintenance such as head cleaning or changing styli. Additional time must be factored in if the analog item needs more than routine cleaning. Other processes such as creating derivatives and uploading files can be batched and performed without close staff supervision.

9.2.2 Condition of Original Materials

Before beginning any audio and/or moving image analog-to-digital (A-D) transfer project, the condition of the original materials should be assessed to make sure they will not be damaged during the conversion process. Contact the AV Archivist for a condition analysis of the original material before undertaking any transfer work.

Some of the issues that might need to be addressed before attempting playback and/or transfer include:

- If needed, discs should be cleaned (preferably using a lint free cloth and an inert cleaning solution) to remove dirt, fingerprints and grime
- Discs should be inspected for scratches, breaks or missing pieces
- Plastic casings on cassettes may be broken or damaged and need to be replaced

- Polyester-based magnetic tape which displays signs of stickiness or shedding (known as Sticky Shed Syndrome or SSS) can not be played without remediation treatment. In severe cases of SSS, the item may not be playable at all
- Suspected mold or mildew must be removed
- Obvious broken splices on magnetic tape should be repaired
- Motion picture film should be inspected for dirt, broken splices and sprocket damage, and any necessary cleaning or repairs should be made before the transfer is performed. In addition, the film should have about 10 ft of inert polyester leader at the head and tail

In addition, plans should be made for appropriate rehousing of the material after the repair and transfer are complete.

9.3 DIGITIZATION OF AUDIO MATERIAL

Audio materials may be the easiest audiovisual format to digitize because, for the most part, analog sound quality and fidelity can be accurately reproduced in the digital format and even enhanced where appropriate. Analog audio materials do present some digitization challenges, however. It often takes significant time to determine the appropriate play back mechanism for open reel tape, for example. It is not possible to determine a tape's speed or the number of tracks just by looking at it. Nor is it possible to visually determine the actual length of the recording. A length of tape may be 1200 feet long but may only have ten minutes of recorded sound somewhere in the middle. Technicians performing the digital transfer must listen to the entire length of the tape — in real time — to be sure all the recording is captured. Fast-forwarding is not an accepted practice on archival tapes as it may cause the tape to stretch or break. Once the analog item has been correctly matched with the appropriate play back equipment however, the process of the digital transfer can be standardized.

The Guidelines outlined here should be observed for audio digitization and digital audio projects. These best practices include technical policies and guidelines for creating digital files. "Flat" transfer is encouraged in most cases. This means that the sound is transferred "as is" without correction or interference. The Libraries strive to replicate the same listening experience for the user of the digital materials as he would have using the analog material. Some cases call for some clean-up of the digital file for access purposes. In general, the master file is captured and saved unedited.

All original digital audio projects developed at the University of Maryland Libraries and using material owned by the University of Maryland Libraries should follow the best practices outlined here. However, new developments in the field of audio digitization and the changeable nature of technology mean that these practices and protocols will be changed and updated as necessary. At the time of this writing, these practices and protocols conform to established guidelines used throughout the academic library community and the Library of Congress (http://www.loc.gov/rr/mopic/avprot/audioSOW.html).

Special projects involving materials not owned by University of Maryland Libraries or other unique project-specific outcomes may follow other protocols with the knowledge and approval of DCR.

9.3.1 Required Digital End Products: Standard and Maximum

The University of Maryland Libraries have established two digital end product levels at which, under the circumstances outlined below, an audio digitization project is acceptable. Before any conversion

is undertaken, project leaders should confer with DCR staff to determine which digital end product level best suits the project needs.

Some things to consider when evaluating the end product level include:

- Fragility/condition and rarity of the analog originals
- Availability of digital storage space
- Estimated use
- Cultural significance
- Copyright status
- Specific project goals

Tools such as the University of Maryland Libraries Digital Preservation Policy Document and/or Digital Preservation Matrix (currently under development) will be used in making a determination.

The formats in this section are emerging, and this list of acceptable formats should be treated as provisional. Each original sound unit ("side," "cut,", "track" or other) will be reproduced as a set of two or three digital audio files.

9.3.1.1 Standard

This group of end products should be considered the normal operating procedures for new digital projects, resulting in both a high resolution master file and low resolution access file. These files will meet or exceed the needs of most digital projects created at UMD Libraries.

1. Master File: 48 kHz/24 bit or 44.1 kHz/16 bit WAVE file depending on the quality of the source recording
2. Access File: MP3 files are retained

9.3.1.2 Maximum

Special circumstances — such as forensic sound research, participation in a cooperative project requiring specific output — may require capturing the audio-to-digital transfer at a higher sampling frequency (96 kHz instead of the usual 48 kHz). Additionally, such projects may require the creation of a third end product, a submaster file that may be edited for better sound quality or otherwise altered in such a way as to make it significantly improved or different from the listening experience of the master file. In such cases, the submaster file can be used to generate the access file. The submaster file will not replace the unedited master file. Editing sound files for improved sound quality is extremely time consuming and requires the skills of experienced audio engineers. Moreover, it requires archiving more than one digital file. For these reasons, this level of end product should be considered the exception, not the rule, for UM Libraries digital audio projects.

1. Master File: 96 or 48 kHz/24 bit WAVE file
2. Submaster File: 44.1 kHz/16 bit edited WAVE file
3. Access File: MP3 files are retained

Additional information on audio digitization can be found in Appendix VIII: Additional Audio Project Planning Tools.

9.4 TECHNICAL POLICIES

9.4.1 Policy on "Dead Air" in Original Recordings

If an analog recording contains periods of "dead air" in which the microphone or other original recording device was on but not recording the primary subject, the "dead air" will be maintained in the digital file and noted in the metadata record. An example of this type of "dead air" includes background noise during an intermission of a musical performance or background noise in between interviews. However, in situations where the microphone or other recording device was not on in between recordings of the primary subject, this "dead air" in which nothing was recorded yet a length of tape exists between recordings will be cropped from the transferred digital file but noted in the metadata record. This is a common occurrence in oral histories, for example, when the interviewer begins recording on the second side of an audiocassette or open reel tape that has not been rewound to an endpoint and the recording doesn't begin until well into the second side.

9.4.2 Policy on Adding a Buffer to Digital Files

The digital recording should contain approximately five (5) seconds of "buffer" at each end of the transferred analog recording. Bracketing the transferred file in this way 1) demonstrates to the listener that no information has been cropped at the beginning and end of a file, and 2) demonstrates the noise level embedded in the digital recording from playback equipment or other means.

9.5 DIGITIZATION OF MOVING IMAGES

Currently, there are no widely accepted best practice guidelines for moving image digitization. Some reasons for this include the extremely large storage needs of uncompressed video files and constantly changing file formats which make moving image digitization a challenging undertaking. In his well known article "Preservation-Worthy Digital Video; or, How to Drive your Library into Chapter 11," Jerry McDonough explains some of the compromises that libraries must face when confronting the challenging needs of creating digital video:

> Large scale digitization of video at a level of quality suitable for preservation will be beyond the means of any but the well-funded of research institutions and a few large commercial entities. Smaller institutions will probably be forced to employ a mix of strategies based upon their available finances, including reformatting some materials to analog media, some to lossy digital, and perhaps a limited amount of extremely valuable and at-risk material to lossless digital formats (McDonough, 2004).

The University of Maryland Libraries is committed to creating and preserving high quality digital end products with no or minimal alteration to the analog original (which is retained in most cases). Currently, the UM Libraries follows the overall digital work plan used by the Library of Congress' American Memory Project.

All original digital moving image projects developed at the UM Libraries and using material owned by the UM Libraries should follow the best practices outlined in this document. Special projects involving materials not owned by UM Libraries or other unique project-specific outcomes may follow other protocols with the knowledge and approval of DCR.

9.5.1 Required Digital End Products: Minimum, Standard and Maximum

The University of Maryland Libraries have established three end-product levels at which, under the circumstances outlined below, moving image digitization projects are acceptable. Before any conversion is undertaken, project leaders should confer with DCR staff to determine which digital end product levels best suit the project needs.

The formats in this section are emerging, and this list of acceptable formats should be treated as provisional. Digital video is especially volatile and changeable and the file formats and specifications outlined below are subject to frequent review and revision. Particularly in the case of moving image digitization, consult DCR and the AV Archivist before embarking on a digitization project

Each original moving image unit ("program," "cut" or other) will be reproduced as a set of between one and three digital video files.

9.5.1.1 Standard

This should be considered as normal operating procedures for new digital projects resulting in both a high resolution master file and low resolution access file. These files will meet or exceed the needs of most digital projects that do not have a long term preservation motive but rather serve immediate access needs.

1. Master File: Keep the best quality file needed to regenerate access copies. At this time, this is MPEG 2
2. Access File: Low resolution access files are retained. At this time, this is RealMedia

9.5.1.2 Maximum

Some very rare, monetarily valuable or otherwise vital collections require the highest level of digital preservation. This very high level of digital end products has a long term preservation motive at heart and so the master files will be saved and maintained at the best quality that UM Libraries can support long term. Special circumstances may require saving the digital object in a higher quality or uncompressed lossless format and/or the creation of the third end product, an additional high resolution intermediate submaster. In such cases, the submaster file can be used to generate a low resolution access file. The submaster file will not replace the unedited master file. Editing video files for improved visual quality is extremely time consuming and requires the skills of experienced engineers. Moreover, it requires archiving a third digital file. For these reasons, this level of end product should be considered the exception, not the rule.

1. Master File: Keep the best quality file that UM Libraries can support long term
2. Submaster: Keep the best quality file needed to regenerate access copies. At this time, this is MPEG 2
3. Access File: Low resolution access files are retained. At this time, this is RealMedia

9.5.1.3 Minimum

This level is the minimum digital file creation for any digital project and should not be considered the norm. Only unusual circumstances, such as copyright issues or specific project goals, would dictate the lack of a high resolution master file being retained.

1. Master File: No high resolution master file retained after streaming service copy is created
2. Access File: Low resolution access files are retained. At this time, this is RealMedia

9.6 QUALITY CONTROL FOR DIGITIZED AUDIO

The best way to check for even quality in audio transfer is to actively listen to the entire recording start to finish. This is often not possible however due to time or staff limitations. The workflow for analog-to-digital transfer usually calls for a trained staff member to perform the transfer in real time. This means that the technician is actively moderating the transfer as well as noting appropriate content metadata if appropriate. Ideally, a second technician or engineer with "fresh ears" would follow up and perform quality control following established benchmarks. At University of Maryland Libraries, all digitized audio files should be sampled for sound quality. Technicians charged with quality control should listen for consistency in the audio quality at a number of points in the recording, listening for distortions in the sound, for proper playback speeds, and for artifacts such as hiss and hum. Technicians should specifically check that the volume levels are set correctly. Recordings should also be checked for completeness. Any inconsistencies in the original recordings or issues that arose as a result of the transfer process should be noted in the metadata.

9.7 QUALITY CONTROL FOR DIGITIZED MOVING IMAGES

All digitized moving images should be sampled for consistent quality. Technicians charged with quality control should watch for consistency in the video quality at a number of points in the recording, watching for distortions in the image and for artifacts such blurred, pixilated or jerking images, noise around sharp edges and correct color balance. Programs should also be checked for completeness. Any inconsistencies in the original program or issues that arose as a result of the transfer process should be noted in the metadata.

Appendix I: Public Domain Determination

The following table may be used as guide for determining public domain status (Sitts 2000 71). If, however, there is doubt about the copyright status of materials, a legal opinion should be sought before undertaking the digital project.

When Works Pass Into the Public Domain
Includes material from new Term Extension Act. PL 105-298

DATE OF WORK	PROTECTED FROM	TERM
Created 1-1 -78 or after	When work is fixed in tangible medium of expression	Life + 70 years[1](or if work is of corporate authorship, the shorter of 95 years from publication, or 120 years from creation)[2]
Published before 1923	In public domain	None
Published from 1923 -63	When published with notice[3]	28 years + could be renewed for 47 years, now extended by 20 years for a total renewal of 67 years. If not so renewed, now in public domain
Published from 1964 -77	When published with notice	28 years for first term; now automatic extension of 67 years for second term
Created before 1-1 –78 but not published	1-1 -78, the effective date of the 1976 Act which eliminated common law copyright	Life + 70 years or 12-31 -2002, whichever is greater
Created before 1-1 –78 but published between then and 12-31 –2002	1-1 -78, the effective date of the 1976 Act which eliminated common law copyright	Life + 70 years or 12-31 -2047 whichever is greater

Notes courtesy of Professor Tom Field, Franklin Pierce Law Center
Lolly Gassaway

[1] Term of joint works is measured by life of the longest-lived author.
[2] Works for hire, anonymous, and pseudonymous works also have this term. 17 U.S.C. § 302(c).
[3] Under the 1909 Act, works published without notice went into the public domain upon publication. Works published without notice between 1-1-78 and 3-1 -89, effective date of the Berne Convention Implementation Act, retained copyright only if, e.g., registration was made within five years. 17 U.S.C. § 405.

Appendix II: Guidelines for Working with Original Documents

In keeping with the University Libraries' Mission that includes both access and preservation to research materials, it is appropriate to increase access via digitization of analog library collections, but at the same time, to protect originals from damage. The following guidelines provide for a non-damaging environment for digitizing and specify procedures that ensure safe handling.[21]

GUIDELINES FOR CHOOSING TYPE OF SCANNER

Documents that can be safely scanned on a flat bed scanner

- Copy negatives
- Photographic prints in good condition
- Flat paper items in good condition and without friable media (such as pastels)
- Items sleeved in polyester film so that they can be handled
- Pamphlet and sheet music items that can safely be opened 180 degrees

Items that must be scanned by an overhead scanner

- Photos adhered to board mounts
- Anything that cannot safely be pressed flat safely
- Anything too large to fit on a flat bed scanner

If any damage begins to occur, stop work immediately and report the problem to Preservation faculty or staff. Do not undertake any taping, disbinding, or repairing. Please contact Preservation/Conservation staff if you have any questions.

GUIDELINES FOR DOCUMENT HANDLING

General Guidelines

- No food, drink, or tobacco products in work spaces
- Wash hands before handling materials
- Do not use pens, markers, or sharp objects near the materials
- Do not use rubber bands, paper clips, or self-sticking notes on the materials
- Prepare work spaces and surfaces before beginning
- Have sufficient work space cleared to handle safely all steps in the procedures. Normally, "sufficient work space" means a cleared area six times the dimension of the materials being scanned

[21] These guidelines borrow heavily from the Library of Congress National Digital Library Program and the Conservation Division, "Session on Care and Handling of Library Materials for Digital Scanning: Safe Handling of Library Materials – Review of Practices" workshop handout (1999).

- Do not place objects on top of the library collection materials
- If it is necessary to stack library materials, be sure to place the largest items on the bottom and limit the height of stacks
- Do not place items on the floor, near windows, or on heat/air conditioning registers
- When leaving the work area, close books and cover documents

Safe Handling of Books

- Support the sides of books that have been identified as having weak bindings or that cannot safely be opened a full 180 degrees
- Never force a spine open or apply hard or abrupt pressure
- Do not wet your fingers to turn the pages
- Turn pages by lifting the upper right corner and using your whole hand to support the page as you turn it
- Books with the following characteristics may be safe to invert on a flat bed scanner
    - Strong binding, flexible paper, gutter margins greater than 3/8 of an inch, smaller than 8 ½ X 11, less than 1 ½ inches thick, sewing strong and intact
    - OR, pages completely detached as loose leaves

Books with the following characteristics may not be safe to invert on a flat bed scanner and might require overhead scanning:

- Larger than 8 ½ X 11 inches
- Thicker than 1 ½ inches
- Weak binding or cover attachments
- Margins narrower than 3/8 inches
- Brittle or breaking pages (Check with Preservation librarian or conservator to be sure)

Safe Handling of documents and other flat materials

- Support single sheets with a more rigid backing such as a folder or a piece of board or polyester film that is as large or larger than the collection item
- If the item is fragile or brittle, Preservation staff will supply a polyester folder or encapsulation to protect the item during scanning. Because polyester film has static charge and sharp edges, to prevent a document from catching and tearing, be sure to open a polyester folder completely before inserting or removing a document
- Oversized items, if at all fragile, should be supported by a rigid board and sandwiched between two boards for support when being turned over
- Before unfolding a folded item, examine the edges for any sign of small tears or breaks from brittleness. If you find any such evidence, please contact Preservation staff for help before proceeding. Unfolding a brittle item could result in shattering or tearing it

Safe Handling of Photographs

- Do not touch the surfaces of photographic emulsions
- Wear clean, lint-free cotton gloves (or other material approved by Preservation staff) which may be obtained from Preservation staff before a photograph scanning project begins
- Do not try to flatten curled or curved photographs (contact Preservation staff for advice and assistance)

Appendix III: Steps in Usability Testing

1. Address the Big Issues
    - Define the site's goals.
    - Define and prioritize your target audience (e.g. undergraduates, faculty, history graduate students).
    - Identify the most important tasks you want users to be able to complete on your site.
    - List questions the development team is trying to resolve about the design of the interface.
2. Assign Roles
    - Test Writer (writes the usability test tasks and the facilitator's script)
    - Scheduler (reserves the room and schedules the participants)
    - Facilitator (guides the participants through the tests; this should not be someone closely involved in the design, authoring or development of the site)
    - Test Viewers (watches the test live and takes informal notes)
    - Report Writer (writes up test results and recommendations)
3. Prepare for the Test Session
    - Write tasks for the test and a script for the facilitator. Plan on each test taking approximately 45 minutes.
    - Recruit 4-6 participants from the target audience.
    - Reserve the usability lab, or other testing location.
    - Acquire incentives for testers (such as gift certificates to the campus bookstore).
    - Prepare waivers the test participants will sign. (You will need a form approved by the campus Institutional Review Board or IRB).
    - Test the software and hardware before the test begins. (ITD uses Morae Usability Testing Software.)
    - Confirm with developers that the test site will be available on the test date.
    - Run a practice test (to test the equipment, software, and tasks). Make any necessary adjustments.
4. Conduct the Test (Facilitator)
    - Have the participant sign a waiver.
    - Make the participant feel comfortable and encourage him/her to keep talking or "thinking out loud."
    - Spend 30-45 minutes allowing each person to work through the predefined tasks.
    - Record the test.
    - Thank the participant and give out the gift certificate or other incentive.

5. View the Test Live (Test Viewers)
    - Watch the test from a remote viewing station.
    - Take informal notes on problem areas.
6. Discuss Each Test (Facilitator and Test Viewers)
    - After each test, compare notes on issues observed. These observations will be the basis for the action list in the session report.
    - Identify any problems with the test questions, and revise the questions if necessary.
    - Attend to any technical problems that arose.
7. Report and Act On the Results (at the end of each group of tests in a session)
    - Write a report that includes a list of primary problems observed and possible action points.
    - If necessary, prepare video clips to share with interested parties.
    - Discuss the report and action points with interested parties and develop a final action list.
    - Make changes to the interface based on the action list.
8. Run as many more sessions (repeat steps 2-7) as deemed necessary for the project.

Appendix IV: A Typology of Formats

This table created by the NISO Advisory Group, lists file formats organized by a "typography that recognizes data types, and within data types, applications to which objects of that type may be put" (NISO 2004).

DATA TYPE	APPLIATIONS	FORMATS	GUIDELINES and REFERENCES
Alphanumeric data	Flat files; hierarchical or relational datasets	US-ASCII or UTF-8 text, or portable format files recognized as de facto standards (e.g. SAS or SPSS) with enough metadata to distinguish tables, rows, columns, etc.	For social science and historical datasets, see *Guide to Social Science Data Preparation and Archiving* (ICPSR, 2002} (http://www.icpsr.umich.edu/ACCESS/dpm.html) and *Digitising history, a guide to creating digital resources from historic documents* (HDS, 1999) (http://hds.essex.ac.uk/g2gp/digitising_history/index.asp).
Alphanumeric data	Encoded texts for networked presentation and exchange of text-based information	SGML, XML; use documented DTDs or Schema	
Alphanumeric data	Encoded texts for literary and linguistic content analysis	SGML, XML	*Text Encoding Initiative* (TEI) (http://www.tei-c.org). *Creating and documenting electronic texts* (OTA, 1999) (http://ota.ahds.ac.uk/documents/creating/) and *TEI text encoding in Libraries: Guidelines for Best Practice* (DLF, 1999) (http://www.diglib.org/standards/tei.htm).
Image data; bitonal, grayscale, and color page images of textual documents	Book or serial publication	Archival masters likely to be uncompressed, baseline TIFF files or lossless compressed JPEG2000 at color depth and pixilation appropriate for application. Derivative formats for access likely to vary depending on use.	National Archives and Records Administration. Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files—Raster Images (June 2004) (http://www.archives.gov/research_room/arc/arc_infor/techguide_raster_june2004.pdf). A consensus for minimum characteristics is Benchmark for faithful digital reproductions of monographs and serials, Version 1 (DLF, 2002) (http://www.diglib.org/standards/bmarkfin.htm). An example of one institution's local benchmarks: California Digital Library. Digital Image Format Standards (http://www.cdlib.org/news/pdf/CDLo,ageStd-2001.pdf).
Image data; bitonal, grayscale, and color page images of textual documents	Newspapers	Grayscale raster formats for masters, almost always supplemented with PDFs and OCR text for access and use	Library of Congress. *The National Digital Newspaper Program (NDNP) Technical Guidelines for Applicants* (http://www.loc.gov/ndnp/ndnp_techguide.pdf). OCLC Digitization and Preservation Resource Center. Click link to Newspaper Digitization (http://digitalcooperative.oclc.org/)
Scalable bit-mapped image data to support zooming and multiple resolution delivery	Maps, herbarium specimens, photographs, aerial photographs	Lossless compressed JPEG2000 files can be used as archival masters and lower quality/smaller size	

from a single file		access copies can be derived from them	
Audio	Music audio	Archival masters should consist of a linear PCM bit stream, which may be wrapped as an uncompressed WAVE or AIFF file. End-user delivery format options include MP3 (MPEG-1level 3), AAC, and RealAudio	This brief technical introduction to Digital Audio by the National Library of Canada provides useful explanations although it suggests the use of cleanup tools, a practice eschewed by most preservation reformatting programs (http://www.collectionscanada.ca/9/1/p1-248-e.html). The Harvard University Library *Digital Initiative Audio Reformatting* site has links to industry standards and will include project guidelines in the future (http://hul.harvard.edu/ldi/html/reformatting_audio.html). Essays that discuss concepts and practices include Carl Fleischhauer's paper on the *Library of Congress Digital Audio Preservation Project* (http:///www.arl.org/preserv/sound_savings_proceedings/fleischhauer.html) and *Sound Practice: A Report on the Best Practices for Digital Sound Meeting* (16 January 2001) at the Library of Congress (http://www.rlg.org/preserv/diginews/diginews/diginews5-2.html#feature3). A statement on ethics and practices has been published by the International Association of Sound and Audiovisual Archives (http://www.iasa-web.org/iasa0013.html)
Audio	Spoken work (e.g. oral histories)	See music audio above	HistoricalVoices.org (http://www.historicalvoices.org/), a project of the National Gallery of the Spoken Word (http:/www.ngsw.org/), includes best practices and research papers on digital speech and an oral history tutorial. The Spoken Word Project at Northwestern University is a good example of work on synchronizing transcripts and sound recordings (http:///www.at.northwestern.edu/spoken/).
Video (A/V)	Moving image content originally created as video or transferred from film	High resolution video files are huge and digital file formats for preservation quality video are immature. Therefore, at this time most organizations maintain their best archival copies of video content in media-dependent form. Preferred media-dependent formats contain a minimally compressed or uncompressed signal, e.g., DigiBeta, D1, or D5 tape. In a high bandwidth LAN, access copies may be high-bit rate	The *NINCH Guide to Good Practice in the Digital Representation of Cultural Heritage Materials* has a good chapter on audio/video capture and management (http:www.nyu.edu/its/humanities/ninchguide/). The Video Development Group (ViDe) provides information about digital video file creation (http://www.vide.net). Agnew, Grace. *Video on Demand: the Prospect and Promise for Libraries* in the *Encyclopedia of Library and Information Science* (New York: Marcel Dekker, 2004) gives an overview of digital video (http://www.dekker.com/serviet/product/productid/E-ELIS). Association of Moving Image Archivists. *Reformatting for Preservation: Understanding Tape Formats and Other Conversion Issues* (http://www.amianet.org/publication/resouces/guidelines/videofacts/reformatting.html). The Association of Moving Image Archivists (http://www.armianet.org/)is a non-profit professional association established to advance the field of moving image archiving. Many of the postings on the AMIA-L listserv

		MPEG-2or MPEG-4 files in larger picture sizes; for lower bandwidth applications and the Web, one may present lower-bit rate MPEG-4, RealVideo, or Quick Time formats with smaller picture sizes.	(http://www.amianet.org/amial/amial.html) are relevant to video archiving; the archive for the listserv may also be consulted: (http://lsv.uky.edu/archives/amia-l.html).
Video (A/V)	Capturing live performances	See above	Two draft guides from the Internet2/CNI Performance Archive and Retrieval Working Group: Capturing Live Performance Events, version 0.9 (2003) (http://arts.internet2.edu/files/erformace-capture(v09).pdf). Current Practice in Digital Asset Management, version 0.9 (2003) http://arts.internet2.edu/files/digital-asset-management(v09).pdf.
Miscellaneous	GIS	Alphanumeric data (e.g. as required to record coordinates), vector, and raster graphics (e.g. to represent maps)	*GIS. A guide to good practice* (ADS, 1998) (http://ads.ahds.ac.uk/project/goodguides/gis/index.html).

Appendix V: Minimal Requirements for Creating Digital Images

The following chart outlines only the minimal requirements for creating digital images derived from both the National Archives and Records Administration of the United States (2004) and the recommendations of Cornell University's Digital Preservation Policy Working Group (2001).

The size of the longest side of the original material is indicated along the top row. The type of material is indicated on the left. The proper resolution (or range of resolution) is found using those two variables. The color profile is indicated in the gray label for each section of materials.

	≥1.5"	**1.6" – 5.5"**	**5.6" – 11.5"**	**11.6" – 18"**	**<18"[1]**
Bitonal (black and white)					
Printed text *Any text printed with strong contrast (i.e. black on white) in a standard typographic font. Most bound books and periodicals will fall into this category.*	4000 dpi	1200 dpi	600 dpi	300 dpi	300 dpi
8-bit gray or 24-bit color					
Rare/damaged printed text *Any printed text that might have lower contrast, fading, or any other damage, or printed with any non-standard typographic font.*	2750 dpi – 4000 dpi	800 dpi – 1200 dpi	400dpi – 600 dpi	300 dpi	300 dpi
Book illustrations *Any illustration including photographs, drawing, or other reproduced artworks printed using a standard screening process. This does not apply to high-quality art reproductions.*	2750 dpi – 4000 dpi	800 dpi – 1200 dpi	400 dpi – 600 dpi	300 dpi	300 dpi
Manuscripts *Any text produced by non-mechanical means including any illustrated or illuminated manuscripts.*	2000 dpi – 3500 dpi	600 dpi – 1000 dpi	300 dpi – 500 dpi	300 dpi	300 dpi
24- or 16-bit color					
Photographic prints *A print of an image produced by any photographic process whether originally in black and white or color*	2750 dpi – 4000 dpi	800 dpi – 1200 dpi	400 dpi – 600 dpi	300 dpi	300 dpi
Graphic art *Original relief, intaglio, and planographic illustrations, maps*	2750 dpi – 4000 dpi	800 dpi – 1200 dpi	400 dpi – 600 dpi	300 dpi	300 dpi
Works of art on paper *Hand-produced works in any medium on paper, canvas, board, panel or any other flat surface*	2750 dpi – 4000 dpi	800 dpi – 1200 dpi	400 dpi – 600 dpi	300 dpi	300 dpi
Bitonal, 8-bit gray, or 24-bit color					
Transparencies *Any image produced on a transparent plastic film or medium, either in a negative or positive color or black and white image. Transparencies of texts may be scanned using a bitonal bit-depth, other images in 8- or 24-bit depth profiles*	2400 dpi – 3500 dpi	800 dpi – 1000 dpi	400 dpi – 500 dpi	300 dpi	300 dpi

[1] This assessment presumes that the oversize item will be used at its original size. If only a portion of the item is to be used, following the guidelines for materials at the size of the portion to be used.

Appendix VI: Quality Control for Images

1. Review documentation for the project to ensure that product goals are clear and it is possible to identify characteristics of an acceptable and unacceptable product.
2. Identify the products to be evaluated. "These might include master and derivative images, printouts, image databases, and accompanying metadata, including converted text and marked-up files" (Cornell University Library Research Group 2003).
3. Determine what the product will be measured against, originals or some intermediate, technical targets, as well as other specifications such as histograms.
4. Control the environment, including calibrating hardware. Kenney and Rieger advise attention to control of the following to control onscreen image quality inspection (2000 66-67):

    - Hardware configuration
    - Image-display software
    - Monitor set-up
    - Color quality control instruments and software
    - Color management
    - Viewing conditions
    - Human characteristics
5. Benchmarking: Verify the appropriateness of technical decisions using a sample of the material to be digitized. Keeping in mind the goal of the project (for example, the creation of a long-lived, high quality digital object), be sure to revise the technical decisions to meet appropriate specifications if the sample results are not satisfactory. When benchmarking is successfully established, keep in mind that the steps will be repeated throughout the project to maintain quality control for the rest of the project. Make sure that all the steps are documented and turned into check lists for easy repetition.
6. Establish the percentage of the product that will be reviewed for quality control. For University of Maryland Library projects, recommended frequency is 100% of all image files and accompanying metadata. Always review 100% of at least the first shipment or batch. The minimal requirement is 10% sampling of each image/metadata batch, assuming that the first shipment or batch is error free. 100% inspection is advisable for small projects. It is assumed that all mistakes will be corrected by those responsible for the errors. At any time, if the error rate rises above 1%, the product should be rescanned. Assessment of image quality adapted from the RLG Model Request for Proposal (RFP) for Digital Imaging Services (RLG 1997).

    - View technical targets and a sample of the digital images on-screen at full resolution using a high resolution monitor[1]

[1] Resolution and color targets should be scanned as part of the project. Common resolution targets include AIIM Scanner Test Chart#2, RIT Alphanumeric Resolution Test Object, IEEE Std 167A.1995, and IEEE Standard Facsimile Test Chart. Common color targets include Q13 and Q14 Kodak Color Separation Guide and

- Record technical readings on a Quality Control form
- Identify any missing/incomplete pages, pages out of sequence, and pages skewed, and evaluate the image quality of text and illustrations. Mark and compare any anomalies in image capture against project worksheets and/or the originals

7. For images consisting of text/line art, any or all of the following requirements must be exhibited when examining a 600 dpi paper printout without magnification:
    - Full reproduction of the page, with skew under 2% from the original
    - Sufficient contrast between text and background and uniform density across the image, consonant with original pages
    - Text legibility, including the smallest significant characters
    - Absence of darkened borders at page edges
    - Characters reproduced at the same size as the original, and individual line widths (thick, medium, and thin) rendered faithfully
    - Absence of wavy or distorted text
8. Magnification may be used to examine the edges and other defining characteristics of individual letters/illustrations. Under magnification the following text attributes are required:
    - Serifs and fine detail should be rendered faithfully
    - Individual letters should be clear and distinct
    - Adjacent letters should be separated
    - Open regions of characters should not be filled in
9. For illustrations and other graphics, the following attributes will be evaluated with or without magnification, as needed:
    - Capture of the range of tones contained in the original
    - Consistent rendering of detail in the light and dark portions of the image
    - Even gradations across the image
    - Absence of "noise" such as moiré patterns and other distorting elements
    - The presence of significant fine detail contained in the original

---

Grayscale Targets, Q60 Kodak Color Target, and Kodak Grayscale Charts. Keep in mind that fingerprints and scratches will compromise the usefulness of charts. Color targets are made with organic dyes and that these dyes break down as they age. Therefore, staff should wear gloves when handling the charts, and the charts must be replaced at regular intervals (NINCH 2002).

10. Aimpoints:

Aimpoint measurements shall be taken using either a 5x5 pixel or a 3x3 pixel sample. At least one set of three aimpoint measurements shall be taken per volume scanned.

Using the Kodak Q-13 Gray Scale as a reference target, aimpoint measurements shall be taken at the neutralized white point (W), neutralized midpoint (M) and neutralized black point (19). An alternate neutralized black point (B) may be used, if needed. Consult Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files—Raster Images (NARA 2004) for additional information.

Acceptable aimpoint ranges are:

Kodak Q-13	A	M	19	B
RGB Levels	242-242-242	104-104-104	12-12-12	24-24-24
% Black	5%	59%	95%	91%

When necessary (e.g., poor image capture of an illustration), unacceptable images will need to be re-scanned from the original volume and inserted into the proper image file sequence.

Appendix VII: XML Examples

SAMPLE TEI DOCUMENT

```
<?xml version="1.0" encoding="utf-8" ?>
<!DOCTYPE TEI.2 (View Source for full doctype...)>
<TEI.2 id="rev.nsn.002" TEIform="TEI.2">
<teiHeader>
   <fileDesc>
            <titleStmt>
                  <title type="main">Review of
                        <title level="m">Modern States Series:
                        <placeName>Italy</placeName></title>
                  </title>
                  <title type="sub">By <persName>R. Sencourt</persName></title>
                  <title type="version">A Machine Readable Version</title>
                  <author>Thomas MacGreevy</author>
                  <respStmt>
                        <resp>Creation of machine readable text by:</resp>
                        <name>Paula Murphy</name>
                        <resp>Proofed by:</resp>
                        <name>Susan Schreibman</name>
                        <resp>Header creation and markup by:</resp>
                        <name>Paula Murphy</name>
                  </respStmt>
            </titleStmt>
            <extent>6kb</extent>
            <publicationStmt>
                  <publisher>Susan Schreibman and Institute for Advanced Technology
                  in the Humanities</publisher>
                  <address>
                        <addrLine>University of Virginia</addrLine>
                        <addrLine>Charlottesville, VA</addrLine>
                        <addrLine>http://jefferson.village.virginia.edu/
                        </addrLine>
                  </address>
                  <date value="2000-05-02">2 May 2000</date>
                  <idno>rev.nsn.002.xml</idno>
                  <availability status="free" >
                        <p>This text is available only for the purpose of academic
                        teaching and research provided that this header is included in
                        its entirety with any copy distributed.</p>
                  </availability>
            </publicationStmt>
            <sourceDesc>
                  <biblStruct>
                        <analytic>
                              <title level="a">Review of Modern States Series: Italy.
                              By R. Sencourt.</title>
                              <author>Thomas MacGreevy</author>
                        </analytic>
                        <monogr>
```

```
                        <title level="j">The New Statesman and
                        Nation</title>
                        <imprint >
                                <date value="1938-11-26">26 November
                                1938</date>
                                <biblScope >p.890</biblScope>
                        </imprint>
                    </monogr>
                </biblStruct>
            </sourceDesc>
    </fileDesc>
    <encodingDesc>
        <editorialDecl>
            <normalization>
                <p>This text has not been normalised.</p>
            </normalization>
            <quotation>
                <p>Quotations marks in the text have been retained as in
                original.</p>
            </quotation>
            <hyphenation>
                <p>Line-end hyphenation has been normalised.</p>
            </hyphenation>
            <interpretation>
                <p>Titles of texts, foreign words, personal names, place and
                organisational names, and emphasised text have been encoded.
                Personal and organisational names which do not conform to the
                Thomas MacGreevy Archive list of standardised names have been
                standardised through the REG attribute.</p>
            </interpretation>
        </editorialDecl>
    </encodingDesc>
    <profileDesc>
        <langUsage>
            <language id="en">English</language>
            <language id="fr">French</language>
        </langUsage>
        <textClass>
            <keywords>
                <list type="keyword">
                    <item type="nationality">Italian</item>
                    <item type="subject">History</item>
                    <item type="date" >1900-1999</item>
                </list>
            </keywords>
            <keywords>
                <list type="collection">
                    <item type="bibliography">works_by</item>
                </list>
            </keywords>
            <classCode>BookReview</classCode>
        </textClass>
    </profileDesc>
```

```
</teiHeader>
<text>
  <body>
        <div0 type="BookReview" org="uniform" sample="complete" part="N">
              <pb n="890" />
              <head>
                    <title level="m" rend="bold">
                    Modern States Series:
                          <placeName>Italy</placeName>
                    </title>.
                    By <persName reg="case(smallcaps)">R.
                    Sencourt</persName>
                    <hi rend="italic"><orgName>Arrowsmith</orgName>
                    . 3s. 6d. </hi>
              </head>
              <p n="1">Since it is claimed that in the series to which this little book
              belongs, "due attention is given to the origins and to the continuity of the life
              of each people and of each state," one would have thought that the greatest
              period in the political as in the commercial and, above all, artistic life of
              <placeName>Florence</placeName>, that of the two and a half centuries
              between 1282 when it overthrew one tyranny and 1530 when another was
              imposed on it by the German head of the Holy Roman Empire, would come in
              for some discussion. But in the one of his eight chapters which
              <persName>Mr. Sencourt</persName> devotes to pre-Risorgimento
              <placeName>Italy</placeName>, he traces no connection between
              Florentine liberty and the Renaissance. One is not therefore surprised at his
              contempt for the <placeName>Italy</placeName> of the period between
              1870 and 1915 when, he says, "Administration was generally corrupt, political
              life dishonest, and government slipped into the hands of the <foreign
              rend="italic" lang="fr">canaille</foreign>. " He does not, however, deny
              that in the same period the cities had new water supplies, that
              <placeName>Rome</placeName> was cured of fever by draining the
              marshes and walling in the Tiber, that factories increased and that
              <placeName>Italy</placeName> enormously increased herincome. These
              things he attributes simply to the Italian genius for engineering.</p>
        </div0>
    </body>
</text>
</TEI.2>
```

SAMPLE EAD DOCUMENT


```
<?xml version="1.0" encoding="UTF-8" ?>
<ead relatedencoding="MARC21">
<eadheader audience="internal" countryencoding="iso3166-1" dateencoding="iso8601"
langencoding="iso639-2b" findaidstatus="edited-full-draft">
<eadid countrycode="iso3611-1" mainagencycode="MdU"> MdU.ead.histms.0173</eadid>
  <filedesc>
     <titlestmt>
        <titleproper>Guide to the Papers of John E. Rastall</titleproper>
        <author>Processed by Sarah Heim, October 2001.</author>
     </titlestmt>
     <publicationstmt>
        <publisher />
        <address>
           <addressline>Archives and Manuscripts Department, University of Maryland
           Libraries, Hornbake Library, College Park, MD 20742. Tel: 301-405-9058, Fax:
           301-314-2709, Email: archives-um@umd.edu</addressline>
              </address>
        <date>October 2001</date>
        <p>©University of Maryland Libraries. All Rights Reserved.</p>
     </publicationstmt>
  </filedesc>
  <profiledesc>
     <creation>EAD markup created using EAD database in Microsoft Access. Markup
     completed by Henry Allen, July 2004. Markup checked and verified by Jennie A.
     Levine, November 2004.
        <date>July 19, 2004</date>
     </creation>
     <langusage>Finding aid written in
        <language langcode="eng">English</language>
     </langusage>
  </profiledesc>
  </eadheader>
     <archdesc level="collection" type="combined">
        <did id="did_1415489940">
              <head>Brief Description of the Collection</head>
              <repository label="Repository">
                     <corpname encodinganalog="852$a">Historical
                     Manuscripts</corpname>
                     <address>
                            <addressline>Archives and Manuscripts Department, University
                            of Maryland Libraries, Hornbake Library, College Park, MD
                            20742. Tel: 301-405-9058, Fax: 301-314-2709, Email:
                            archives-um@umd.edu</addressline>
                     </address>
              </repository>
              <origination label="Papers/Records Created By">
                     <persname encodinganalog="100">Rastall, John E.</persname>
              </origination>
```

```
        <unittitle label="Title of the Collection" encodinganalog="245$a">Papers of
        John E. Rastall</unittitle>
        <unitdate type="inclusive" label="Dates of the Collection"
        encodinganalog="245$f">1861-1864</unitdate>
        <unitdate type="bulk" label="Bulk Dates" encodinganalog="245$g">1861-
        1864</unitdate>
        <physdesc label="Size of the Collection" encodinganalog="300$a">2 linear
        inches (128 items)</physdesc>
        <unitid label="Accession Number" encodinganalog="099">98-146</unitid>
        <physloc audience="internal">GR:36:A:91 J -</physloc>
        <abstract label="Short Description of Collection"
        encodinganalog="520$a">John E. Rastall was a Union Lieutenant with the
        First Regiment, Eastern Shore, Maryland Volunteers during the Civil War. The
        collection includes 128 letters written by Rastall to his family in Milwaukee,
        Wisconsin detailing his service in Virginia and Maryland, especially on the
        Eastern Shore.</abstract>
        <abstract type="civil">John E. Rastall was a Union Lieutenant with the First
        Regiment, Eastern Shore, Maryland Volunteers during the Civil War. The
        collection includes 128 letters written by Rastall to his family in Milwaukee,
        Wisconsin detailing his service in Virginia and Maryland, especially on the
        Eastern Shore.</abstract>
        <abstract type="family">Lieutenant John E. Rastell of Wisconsin was
        adjutant of the Union Army's First Eastern Shore Regiment of Infantry,
        Maryland Volunteers. Rastell's regiment was stationed in Salisbury, Maryland,
        during most of their Civil War service; their duty was to "pacify" the volatile,
        Southern-sympathizing Eastern Shore. Lieutenant Rastell's letters to his
        family describe his impressions of daily life in the Union Army; the impact of
        the Civil War on the citizens of Maryland, especially on the Eastern Shore;
        and his experiences at the Battle of Gettysburg.</abstract>
        <abstract type="geoges">John E. Rastall was a Union Lieutenant with the
        First Regiment, Eastern Shore, Maryland Volunteers during the Civil War. The
        collection includes 128 letters written by Rastall to his family in Milwaukee,
        Wisconsin detailing his service in Virginia and Maryland, especially on the
        Eastern Shore.</abstract>
    </did>
    <descgrp id="des_2136608943">
        <head>Important Information for Users of the Collection</head>
        <accessrestrict encodinganalog="506">
            <head>Use and Access to Collection</head>
            <p>There are no restricted files in this collection.</p>
        </accessrestrict>
        <acqinfo audience="external" encodinganalog="541">
            <head>Custodial History and Acquisition Information</head>
            <p>The University of Maryland Libraries purchased the Papers of John
            E. Rastall from manuscripts dealer Charles Apfelbaum in 1997.</p>
        </acqinfo>
        <prefercite encodinganalog="524">
            <head>Preferred Citation</head>
            <p>Papers of John E. Rastall, Special Collections, University of
            Maryland Libraries</p>
        </prefercite>
        <processinfo encodinganalog="583">
```

```
            <p>The letters were placed in acid-free folders and stored in an acid-
            free box.</p>
        </processinfo>
        <userestrict encodinganalog="540">
            <head>Duplication and Copyright Information</head>
            <p>Photocopies of original materials may be provided for a fee and at
            the discretion of the curator. Please see our <extref
            href="http://www.lib.umd.edu/mdrm/policies/duplication.html"
            show="new" actuate="onrequest">Duplication of Materials</extref>
            policy for more information.
            </p>
        </userestrict>
    </descgrp>
    <bioghist id="bio_44980194" encodinganalog="545">
        <head>Biography</head>
        <p>John Edward Rastall was born on July 23, 1840 in Cheltenham,
        Gloucestershire, England to Richard and Sarah Rastall. He had four brothers:
        Samuel, James, Richard (Dick), and Benjamin. The family emigrated to
        Milwaukee, Wisconsin in 1852 or 1853, where young Rastall learned printing
        through working at the Milwaukee <title render="italic">Sentinel</title> and
        at the <title render="italic">Beloit Herald</title> in Beloit, Wisconsin. Only a
        few years later, he was one of a group of Wisconsin abolitionists headed by E.
        G. Ross who went south to join the Free State Army in Kansas. Rastall made
        raids on slave-holding villages with riders led by Jim Lane and John Brown. He
        was captured by federal troops but escaped and made his way back to
        Milwaukee, where he worked at the <title render="italic">Sentinel</title>
        until the Civil War broke out in 1861. He enlisted immediately as a private in
        the Fifth Wisconsin Volunteer Infantry, Company B. In September 1861 he
        was discharged from the Fifth Wisconsin to accept a commission as a first
        lieutenant and adjutant of the newly organized First Maryland Eastern Shore
        Volunteers. He served with the First Maryland Eastern Shore from September
        1861 to October 1864. The regiment spent time at Salisbury, Point of Rocks,
        and Cambridge, Maryland, as well as at Fort McHenry and other sites around
        Baltimore. They saw action at the battle of Gettysburg. As adjutant, Rastall
        assisted Colonel James Wallace and later Colonel John R. Keene with
        administrative and clerical duties, and he officiated at court-martial
        proceedings.</p>
        <p>After the war, Rastall spent several years dividing his time between
        farming in Manistee, Michigan and printing in Milwaukee. In 1867 he married
        Miss Fannie Hawley, the daughter of a Milwaukee dry-goods merchant. In the
        early 1870s he returned to Kansas with his wife. From 1876 to 1877 he
        published the <title render="italic">Junction City Union</title> in Junction
        City, Davis County. In 1877 the couple settled in Burlingame, Osage County,
        where they remained for many years while Rastall published the <title
        render="italic">Osage County Chronicle</title>. Beginning in 1881, Rastall
        served in the Kansas state legislature. Fannie Rastall was also active in the
        community and used her influential position in the local chapter of the
        Women's Christian Temperance Union to help establish an Industrial School
        for Girls in Beloit, Kansas. </p>
        <p>Rastall later moved to Washington, D. C., where he worked in the United
        States Government Printing Office until his retirement, probably in the late
        1910s. Fannie died in Manchester, Vermont in 1920. Rastall returned to
```

```
        Wisconsin for medical treatment at the Wisconsin Veterans' Home in King in
        1923 but resided in Washington until his death in 1927.</p>
    </bioghist>
    <scopecontent id="sc_1416093302" encodinganalog="520">
        <head>Scope and Content of Collection</head>
        <p>The Papers of John E. Rastall consist of 128 letters written between
        September 1861 and October 1864 by John E. Rastall to his parents and
        brothers in Milwaukee, Wisconsin. Of particular interest are discussions of
        both the military and social aspects of army life, as well as descriptions of
        how Union and Confederate sympathies were expressed by civilians in
        Maryland during the war.</p>
    </scopecontent>
    <arrangement id="arr_1209999377" encodinganalog="351">
        <p>The collection is organized as one series.</p>
        <list>
            <item>Series 1: Correspondence</item>
        </list>
    </arrangement>
    <relatedmaterial id="rm_1307126084" encodinganalog="544">
        <head>Related Material</head>
        <p>The Wisconsin Historical Society Archives holds an autobiographical
        sketch Rastall wrote in 1924 and three 1923 articles from Wisconsin
        newspapers. Additional information can be found in <title
        render="italic">History of the State of Kansas</title> by William G. Cutler
        (Chicago: A. T. Andreas, 1883), at the Kansas State Library <archref
        href="http://skyways.lib.ks.us">http://skyways.lib.ks.us</archref> and
        <archref
        href="http://www.theleefamily.org">http://www.theleefamily.org</archref>
        maintained by Monty Lee. All web addresses were valid as of October 2001.
        </p>
        <p>For other related archival and manuscript collections, please see the
        following <archref xpointer="rguide">subject guides</archref></p>
    </relatedmaterial>
    <controlaccess id="ca_1663547175">
        <head>Selected Search Terms</head>
        <p>This collection is indexed under the following headings in the University
        of Maryland Libraries' <extref
        href="http://catalog.umd.edu/">Catalog</extref>. Researchers desiring
        related materials about these topics, names, or places may search the
        Catalog using these headings.</p>
    <controlaccess>
        <head>Subjects</head>
            <persname role="subject" encodinganalog="600">Rastall, John
            Edward, -- 1840-1927 -- Correspondence</persname>
            <corpname role="subject" encodinganalog="610">United States. --
            Army. -- Maryland Infantry Regiment, 1st (1861-1865)</corpname>
            <geogname role="subject" encodinganalog="651">United States --
            History -- Civil War, 1861-1865 -- Personal narratives,
            Union</geogname>
            <geogname role="subject" encodinganalog="651">Maryland --
            History -- Civil War, 1861-1865 -- Personal narratives</geogname>
        </controlaccess>
    </controlaccess>
```

```
<dsc type="analyticover" id="dsc_1940908697">
    <head>Contents of Collection</head>
    <c01 level="series" id="series1.a">
        <did>
            <unittitle>Correspondence</unittitle>
            <unitdate>1861-1864</unitdate>
            <physdesc>128 items</physdesc>
        </did>
        <scopecontent>
            <p>Series I consists of letters sent from John E. Rastall to his
            parents and brothers. The letters discuss the social and
            professional aspects of life in the military, as well as current
            events and specific battles. Rastall describes his regiment's role
            in the battle of Gettysburg, speculates on the possible effect of
            the draft on his brothers in Wisconsin, and offers his
            impressions of prominent people such as General McClellan and
            Mary Todd Lincoln. He also reminisces about life at home in
            Milwaukee and describes civilian parties, meetings, and
            excursions he enjoyed in Maryland. The subjects addressed also
            include deserters from the First Maryland, Rastall's attempts to
            secure a promotion, his role in the battle of Gettysburg, court-
            martials, the draft, and the impressment of free blacks into the
            Union army. The arrangement is chronological.</p>
        </scopecontent>
    </c01>
</dsc>
<dsc type="in-depth" id="dsc_1317900819">
    <c01 level="series" id="series1">
        <did>
            <unittitle>Correspondence</unittitle>
            <unitdate>1861-1864</unitdate>
            <physdesc>128 items</physdesc>
        </did>
        <c02 level="file">
            <did>
                <container id="box1.1" type="box">1</container>
                <container parent="box1.1"
                type="folder">1.0</container>
                <unittitle>Correspondence</unittitle>
                <unitdate>September 1861-October 1864</unitdate>
            </did>
        </c02>
        <c02 level="file">
            <did>
                <container parent="box1.1"
                type="folder">2.0</container>
                <unittitle>Correspondence</unittitle>
                <unitdate>September 1861-October 1864</unitdate>
            </did>
        </c02>
        <c02 level="file">
            <did>
```

```
                <container parent="box1.1"
                type="folder">3.0</container>
                <unittitle>Correspondence</unittitle>
                <unitdate>September 1861-October 1864</unitdate>
            </did>
        </c02>
        <c02 level="file">
            <did>
                <container parent="box1.1"
                type="folder">4.0</container>
                <unittitle>Correspondence</unittitle>
                <unitdate>September 1861-October 1864</unitdate>
            </did>
        </c02>
        <c02 level="file">
            <did>
                <container parent="box1.1"
                type="folder">5.0</container>
                <unittitle>Correspondence</unittitle>
                <unitdate>September 1861-October 1864</unitdate>
            </did>
        </c02>
    </c01>
  </dsc>
 </archdesc>
</ead>
```

SAMPLE UMDM RECORD


```
<descMeta>
        <pid>TMP:00000001</pid>
        <mediaType type="image">
              <form type="analog">Trade cards</form>
        </mediaType>
        <title>Ph. J. Lauber's Restaurant, Centennial Grounds</title>
        <agent type="provider">
              <corpName>Lehman & Bolton Phila</corpName>
        </agent>
        <agent type="creator">
              <persName>Ph. J. Lauber</persName>
        </agent>
        <covPlace>
              <geogName>Philadelphia (Pa.)</geogName>
        </covPlace>
        <covTime>
              <century era="ad">1801-1900</century>
              <date era="ad">1876</date>
        </covTime>
        <culture>American</culture>
        <description>view of Lauber's Centennial Restaurant</description>
        <subject type="topical">Restaurants--Philadelphia (PA)</subject>
        <subject type="topical">Exhibitors</subject>
        <subject type="genre">Buildings</subject>
        <identifier>Call Number: 1876-phi-08-0002</identifier>
        <identifier>Box H</identifier>
        <physDesc>
              <size>6.1 x 11 cm</size>
              <color>black and white</color>
        </physDesc>
        <relationships>
              <relation type="isPartOf" label="Fair">Centennial Exhibition (1876 :
              Philadelphia, Pa.)</relation>
              <relation type="isPartOf">Treasury of World's Fair Art &
              Architecture</relation>
              <relation type="isPartOf" label="08">Ephemera</relation>
        </relationships>
        <repository>
              <corpName>Art & Architecture Libraries</corpName>
        </repository>
        <rights>The textual information and images contained in this website are provided
        for educational purposes only. Textual information and/or images may not be
        borrowed or reproduced beyond educational use without obtaining permission from
        the copyright holder. Reproduction in any form is subject to the copyright law of the
        United States. Please refer to sources on intellectual property, copyright, and fair
        use for further information.</rights>
</descMeta>
```

SAMPLE UMAM RECORD


```
<adminMeta>
        <pid>TMP:00000002</pid>
        <identifier>1876-phi-08-0002-0001.jpg</identifier>
        <digiProv>
              <date>2002-06-12</date>
              <agent type="creator">
                    <persName>Paul Hammer</persName>
              </agent>
              <description>Created</description>
        </digiProv>
        <adminRights>
              <policy>
                    <access>Public</access>
              </policy>
        </adminRights>
        <technical>
              <image>
                    <fileSize>121529 bytes</fileSize>
                    <format>
                          <mimeType>jpeg</mimeType>
                          <compression>Lossy</compression>
                          <colorSpace>Color</colorSpace>
                    </format>
                    <spatialMetrics>
                          <imageWidth>1000 pixels</imageWidth>
                          <imageLength>572 pixels</imageLength>
                          <imageResolution>75.0063500127 pixels</imageResolution>
                          <sourceX>0</sourceX>
                          <sourceY>0</sourceY>
                    </spatialMetrics>
              </image>
        </technical>
</adminMeta>
```

Appendix VIII: Additional Audio Project Planning Tools

DIGITAL AUDIO FILE SPECIFICATIONS

Master File

The master file is the digital preservation master file, created for long term retention. Because it is the digital preservation master, this file should conform to the parameters listed below. These parameters may be amended as research and standards advance.

Bitstream: Uncompressed pulse code modulation (PCM).

Configuration: Monophonic or stereo depending upon the characteristics of the source item.

Sampling frequency: 96 or 48 kHz depending upon characteristics of source item.

Bit depth: 24 bit or 16 bit depending upon characteristics of source item.

File format: WAVE (.wav) or Broadcast WAVE (.bwf).

Enhancement: No cleanup, or minimal cleanup (such as volume normalization, cropping dead air, etc) as agreed to during pre-project planning.

Filename structure: Follow DCR Best Practice Guidelines for file naming.

Submaster File

Bitstream: Uncompressed PCM.

Configuration: Monophonic or stereo depending upon characteristics of source item.

Sampling frequency: 44.1 kHz.

Bit depth: 16 bit.

File format: WAVE (.wav).

Enhancement: Minimal, or in some cases significant, cleanup possible as agreed to during pre-project planning.

Filename structure: Follow DCR Best Practice Guidelines for file naming.

Access Service File

Bitstream: MP3

Configuration: Monophonic or stereo depending upon characteristics of source item

Quality: Data rate of 256 or 128 kilobits/second with analysis and recommendations as agreed to during pre-project planning

File format: MP3

Enhancement: No cleanup, or minimal cleanup (such as volume normalization, cropping dead air, etc) as agreed to during pre-project planning UNLESS the lower fidelity file is created from a modified submaster file.

Filename structure: Follow DCR Best Practice Guidelines for file naming.

NOTE: There is a licensing fee associated with creating MP3 files because it is a proprietary format. Open source options such as Ogg Vorbis (http://wiki.xiph.org/index.php/VorbisStreams) maybe explored in the future.

BRIEF OVERVIEW OF AUDIO RECORDING FORMATS

Format	Description	Years in Use
Wax Cylinder Records	2- or 4-minute formats, wax or wax compound	1888– 1929
Recordable Disc Records (Direct or Acetate Discs)	7", 12", or 16", recorded at 33 or 78 revolutions per minute (rpm). Generally vinyl on a paper, glass, or metal base.	1899– 1960s
Recording Wire	Spooled wire, usually in 15- to 30-minute lengths, one direction only	c. 1945– 1955
Open Reel Recording Tape	1/4"– 2", 3"– 10 1/2" reels, 1 7/8– 30 inches per second (IPS) speeds. Multiple tracks possible.	c. 1945– Present
Compact Cassette	1/8" tape in hard case, 1 7/8 IPS format	1965– Present
Microcassette / Minicassette	Very small 2-4 cm cassette tapes	1977– Present
Digital disk, MP3, and other digital recorders	Audio recorded directly in digital files to optical disks or internal hard drives	2000– Present

Table adapted from: http://www.cdpheritage.org/digital/audio/documents/CDPDABP_1-2.pdf

ESTIMATING RECORDING TIME OF OPEN REEL TAPES

Time Chart for Reel-to-Reel Audio Tape (Approximate)										
* Reel Diameter	Tape Length	Single Mono Track			2 Track			4 Track Discrete		
		1 7/8 ips	3 3/4 ips	7 1/2 ips	1 7/8 ips	3 3/4 ips	7 1/2 ips	1 7/8 ips	3 3/4 ips	7 1/2 ips
3"	300 ft	30 min	15 min	7.5 min	1 hour	30 min	15 min	2 hours	1 hour	30 min
5"	600 ft	1 hours	30 min	15 min	2 hours	1 hours	30 min	4 hours	2 hours	1 hour
5"	900 ft	1.5 hours	45 min	22 min	3 hours	1.5 hours	45 min	6 hours	3 hours	1.5 hours
7"	1200 ft	2 hours	1 hour	30 min	4 hours	2 hours	1 hour	8 hours	4 hours	2 hours
7"	1800 ft	3 hours	1.5 hours	45 min	6 hours	3 hours	1.5 hours	12 hours	6 hours	3 hours
10"	2400 ft	4 hours	2 hours	1 hour	8 hours	4 hours	2 hours	16 hours	8 hours	4 hours
10"	3600 ft	6 hours	3 hours	1.5 hours	12 hours	6 hours	3 hours	24 hours	12 hours	6 hours

Estimating Digital File Storage Needs

One hour of recorded stereo audio transferred at 48 kHz and 24 bit will result in a file size of approximately 1GB.

The chart below outlines approximate storage needs for the different mandatory end products.

Minutes	Type of End Product	Sample Rate (kHz)	Bit depth	Channel: mono or stereo	File Size MB	File Size GB
60	Master – musical recordings	98	24	mono	1009	1.01
60	Master – musical recordings	98	24	stereo	2018	2.02
60	Master – non musical recordings	48	24	mono	494	.49
60	Master – non musical recordings	48	24	stereo	988	.99
60	Higher Fidelity Submaster	44.1	16	mono	301	.30
60	Higher Fidelity Submaster	44.1	16	stereo	605	.61
Minutes	**Type of End Product**	**Compression Rate**	**Bit Rate**	**Channel: mono or stereo**	**File Size MB**	**File Size GB**
60	Low Fidelity Listening Access Copy	12:1	128 kbits/sec	stereo	57	.057
60	Low Fidelity Listening Access Copy	6:1	256 kbits/sec	stereo	115	.11

Appendix IX: Glossary

access copy: Copies of a digital file that are made available for users. These copies may be of a lower resolution or smaller file size to facilitate quick downloading.

administrative metadata: Metadata retained to track the creation and maintenance of a digital object. The University of Maryland Administrative Metadata standard required information about the creation of the object and technical details of the datastream.

artifact: A visual effect unintentionally introduced into a digital image in the course of digitizing or during imaging software manipulation.

bit-depth: The number of bits that are used to express the color of a pixel. Each pixel can contain up to three channels for color, each of which can be expressed in up to 8 bits.

born digital: An object is "born digital" when there is no analog "hard copy" counterpart. For example, a photo taken with a digital camera.

byte order: The order in which digital information is stored. When creating digital images different byte orders can be defined. UM Libraries' recommendation is to use the Intel (Windows) byte order.

CMYK: A color model comprised of the four colors cyan, magenta, yellow, and black. This is the color model used for offset printing.

collection: A group of objects in digital format that, when considered as a whole, demonstrates some identifiable organizing principle.

color mode: Refers to the palette of colors available for a digital image. The number of separate colors available in each palette is determined by the bit-depth of the pixels and, when applicable, the channels of color used. Examples include RGB, CMYK, Grayscale, etc.

controlled vocabulary: A defined set of words or phrases used to describe an object. Controlled vocabularies are used in tandem with metadata schemas to create metadata records for digital objects.

cross-platform: Refers to the capability of software or hardware to run identically on different platforms. Many applications for Windows and the Macintosh, for example, now produce binary-compatible files, which means that users can switch from one platform to the other without converting their data to a new format.

datasets: Any data that is organized in a defined set. Lists, tables, and databases are examples.

DCR: Digital Collections and Research.

derived data: Any data that is added to an object through a process or computation.

digital capture: Refers to the digitizing of an object. For example, an archival photograph which has been scanned and now exists as a TIFF has been "captured", and that file is the photograph's "digital capture."

digital master: A digital object that "most closely retains the significant attributes of the original." (Digital Library Federation Benchmark for Faithful Digital Reproductions) Masters are created to be long-lived and high quality to enable the production of versions for various uses

DMCA: Digital Millennium Copyright Act

dpi: dots per inch. A measurement of the scanning resolution of an image or the quality of an output device. Can mean the number of dots a printer can print per inch or the number a monitor can display.

DTD: Document Type Definition.

dynamic metadata: Any metadata added to a record by an automated process such as scripting.

EAD: Encoded Archival Description.

fair use guidelines: The criteria by which uses of copyrighted materials may be considered "fair use" as outlined by United States Copyright Law (US Code, Title 17, section 107).

file extensions: These are the letters appearing after the period at the end of a computer filename. They help the computer identify the file format and what software should be used to interact with it. Examples include: .doc — .jpg — .gif — .xl — .js and so on.

filepath: The location of a file in a server directory

flare: Unwanted reflection of a photographic lens which can cause optical effects in photographs.

GIF: Graphic Interchange Format.

granularity: Level of detail in a metadata record.

identifier: Any information used to differentiate one object from another. Identifiers could be file names, persistent identifiers assigned by a digital repository, or a metadata field.

ITD: Information Technology Division.

JPEG: A file format for images defined by the Joint Photographic Experts Group

JPEG2000: This is a new standard for JPEG file compression which uses so called "wavelet" technology to produce higher-quality images with greater compression.

lossless: A file compression technique for digital objects in which file size is reduced but all information in the original file is recoverable. This type of compression algorithm can reduce file size up to 50%

lossy: A file compression technique for digital objects in which information lost is unrecoverable. This type of compression algorithm can reduce file size more than 50%.

MARC record: MAchine Readable Cataloging. A widely used format for storing bibliographic information in library catalogs.

markup: Refers to text encoded using a markup language such as HTML (HyperText Markup Language) or XML (eXtensible Markup Language).

metadata: Data about other data. Metadata is used to describe what an object is and how it was created, changed or described. Types of metadata include descriptive, administrative, technical, structural, preservation.

metadata scheme: The format used to store metadata about an object. The UM Libraries require that the University of Maryland Descriptive Metadata (UMDM) and Administrative Metadata (UMAM) schemas be used for all digital object added to the digital repository.

METS: Metadata Encoding and Transmission Standard. A standard for encoding descriptive, administrative, and structural metadata about objects within a digital library, expressed using XML. METS is the emerging national standard for wrapping digital library materials. It is being developed by the Digital Library Federation (DLF) and is maintained by the Library of Congress.

microformat: an analogue copy of a print resource greatly reduced in size to conserve storage space. Common formats are microfilm, microfiche, or microcard. Not to be confused with the open source data format standard.

migration: The transfer of digital objects from one hardware or software configuration to another, or from one generation of computer technology to a subsequent generation.

NARA: National Archives and Records Administration.

NISO: National Information Standards Organization.

noise: Extraneous or unidentifiable data introduced into a digital object in the course of digitization that are due to imperfections or sensitivities in the capture device.

object: A digital object is the digital representation of one analog item. A digital object may be composed of several digital files or datastreams which together represent one analog counterpart. An example might be a single digital audio file of an early musical recording, or multiple digital image files that comprise one book.

OCR: Optical Character Recognition.

persistent identifier: An identifier that will stay with a digital object despite any moves to a different storage location, manipulation of the datastream, or addition of other datastreams in the composition of the object.

pid: persistent identifier.

pixel: Derived from "picture element," the smallest element of data of a digitized image. A pixel represents one point in a grid that covers the image and conveys information about the color and location of that point.

preservation metadata: Information captured to ensure the long term retention and preservation of a digital object.

project: The endeavor of creating a digital collection including all its corresponding resources.

public domain: Material that is available to the public because it is out of copyright or unable to be copyrighted, such as government publications or ideas.

QC: quality control.

resolution: An indication of the quality of an image. For digital images, resolution is calculated in dots of color per inch of print size (dpi). When referring to monitor displays, resolution refers to the amount of information (in pixels) included in the horizontal and vertical axes. In video, resolution is expressed in an aspect ratio along with horizontal and vertical pixels.

RGB: A color model comprised of the three primary colors red, green, and blue. This is the color model used for computer monitors and television sets.

saturation: The strength of purity of a color, this represents the amount of gray in proportion to the hue, measured on a scale from 0-100% where 0% = entirely gray and 100% = no gray present.

structural metadata: Information that describes the organization of a digital object.

surrogate: An item which takes the place of, or fills the role of, another item.

technical metadata: Information that describes the technical properties of a specific digital object type.

TEI: Text Encoding Initiative.

TIFF: Tagged Image File Format.

thumbnail: A small version of an image which acts as a link a larger version of the same image.

TSD: Technical Services Division.

UMAM: University of Maryland Administrative Metadata.

UMDM: University of Maryland Descriptive Metadata.

URL: Uniform Resource Locator. This is an identifier which allows one to locate or access an electronic resource via the internet. Although it is most frequently used to mean a website address starting with "http://" (such as http://www.lib.umd.edu), there are many other schemes that fall under this group, including ftp, mailto, irc, file, etc.

watermark: A pattern of bits inserted into a digital image, audio or video file that identifies the file's copyright information (author, rights, etc.). The purpose of digital watermarks is to provide copyright protection for intellectual property that's in digital format.

Appendix X: Bibliography

Arms, Caroline R. and Carl Fleischhauer. Sustainability of Digital Formats: Planning for Library of Congress Collections. Online: http://www.digitalpreservation.gov/formats/index.shtml

Association of Research Libraries (ARL). (2004). *Recognizing Digitization as a Preservation Reformatting Method (Prepared for the ARL Preservation of Research Library Materials Committee).* Retrieved November 18, 2005 from http://www.arl.org/preserv/digit_final.html

Atthur, K., Byrnbe, S., Long, E., Montori, C., & Nadler, J. (2004) *Recognizing Digitization as a Preservation Reformatting Method*. Washington, DC: Association of Research Libraries. Retrieved May 1, 2005 from http://www.arl.org/preserv/digit_final.html.

Bates, M. J. (2002). "The Cascade of Interactions in the Digital Library." *Information Processing & Management, 38*(3), 381-400.

Caplan, Priscilla. (2003) *Metadata Fundamentals for all Librarians*. American Library Association: Chicago, p. 158.

California Digital Library. (2001). *Digital Object Standard: Metadata, Content and Encoding.* Retrieved November 18, 2005 from http://www.cdlib.org/news/pdf/CDLObjectStd-2001.pdf

Collaborative Digitization Program (CDP). Digital Audio Working Group Digital Audio Best Practices Version 2.0 November 2005. Online: http://www.cdpheritage.org/digital/audio/documents/CDPDABP_1-2.pdf

Cornell University Library. (2001). *Report of the Digital Preservation Policy Working Group on Establishing a Central Depository for Preserving Digital Image Collections; Part 1: Responsibilities of Transferee*. Retrieved November 18, 2005 from http://www.library.cornell.edu/preservation/IMLS/image_deposit_guidelines.pdf

Cornell University Library Research Department. (2003). *Moving Theory into Practice Digital Imaging Tutorial*. Retrieved on November 11, 2005 from http://www.library.cornell.edu/preservation/tutorial/quality/quality-01.html.

Digital Library Federation Benchmark Working Group (2001-2002). (2002). *Benchmark for Faithful Digital Reproductions of Monographs and Serials. Version 1*. Retrieved November 18, 2005 from http://purl.oclc.org/DLF/benchrepro0212

Dumas, Joseph and Redish, Janice. *A Practical Guide to Usability Testing* (Revised Edition). Intellect, 1999.

Goto, Kelly. Usability Testing: Assess Your Site's Navigation & Structure. Online: http://gotomedia.com/downloads/goto_usability.pdf

Fogg, B.J. (May 2002). "Stanford Guidelines for Web Credibility." A Research Summary from the Stanford Persuasive Technology Lab. Stanford University. www.webcredibility.org/guidelines

Friedland, L., Kushigan, N., Powell, C., Seaman, D., Smith, N., & Willett, P. (1999). *TEI Text Encoding in Libraries: Guidelines for Best Encoding Practices, version 1.0 (July 30, 1999)*. Retrieved November 21, 2005 from http://www.diglib.org/standards/tei.htm.

Health and Human Services Dept. (U.S.). *Research-Based Web Design & Usability Guidelines*. Available online: http://www.usability.gov/pdfs/guidelines.html

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

Krug, Steve. *Don't Make Me Think: A Common Sense Approach to Web Usability* (Second Edition). New Riders, 2005.

Library of Congress. American Memory. How to view. Online: http://rs6.loc.gov/ammem/help/view.html#video

Library of Congress. Digital Audio-Visual Preservation Prototyping Projects: Appendix 5: Special Consideration for Digital Video and Audio. Online: http://www.loc.gov/rr/mopic/avprot/rfq5.html

_____. Illustrative Example of a Statement of Work Typical Elements for Use in a Statement of Work for the Digital Conversion of Sound Recordings and Related Documents. Online: http://www.loc.gov/rr/mopic/avprot/audioSOW.html

Library of Congress National Digital Library Program and Conservation Division. (1999). "Session on Care and Handling of Library Materials for Digital Scanning: Safe Handling of Library Materials – Review of Practices." Washington DC: Library of Congress. Retrieved December 13, 2005 from http://memory.loc.gov/ammem/techdocs/conserv83199b.html.

McDonough, Jerry. Preservation-Worthy Digital Video, or How to Drive Your Library into Chapter 11. Presented at the Electronic Media Group Annual Meeting of the American Institute for Conservation of Historic and Artistic Works Portland, Oregon June 13, 2004. Online: http://aic.stanford.edu/sg/emg/library/pdf/mcdonough/McDonough-EMG2004.pdf

National Digital Information Infrastructure and Preservation Program (NDIIPP). Sustainability of Digital Formats Planning for Library of Congress Collections. Online: http://www.digitalpreservation.gov/formats/index.shtml

National Information Standards Organization (NISO). (2004). A *Framework of Guidance for Building Good Digital Collections, 2nd ed. 2004.* Retrieved November 18, 2005 from http://www.niso.org/framework/framework2.pdf

_____. (2003). Data Dictionary: Technical Metadata for Still Images. Beta Release. Retrieved November 18, 2005 from http://www.niso.org/standards/resources/Z39_87_trial_use.pdf

National Initiative for a Networked Cultural Heritage (NINCH). (2002). *The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials, 2002*. Retrieved on November 11, 2005 from http://www.nyu.edu/its/humanities/ninchguide/

Notice of Inquiry: Orphaned Works, 70 Fed. Reg. 3739.

Nellhaus, T. (2001). XML, TEI, and Digital Libraries in the Humanities. *Portal, 1*(3), 257-77.

Nielson, Jakob. *Homepage Usability*. New Riders, 2002.

Nielson, Jakob. useit.com: Jakob Nielsen's Website

The Research Libraries Group, Inc. (RLG). (1997). *RLG Model Request for Proposal (RFP) for Digital Imaging Services*. Washington, DC: The Research Libraries Group, Inc. Retrieved on November 11, 2005 from http://www.rlg.org/preserv/RLGModelRFP.pdf.

Pearson, Glenn. Towards a New Meta-Standard for Long-Term Digital Video Preservation - A Quick Overview. December 7, 2005. Online: http://archive.nlm.nih.gov/VideoArchivists2005/maf-review.html

Rosenzweig, Roy and Daniel Cohen. Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web. Philadelphia: University of Pennsylvania Press, 2005. online: http://chnm.gmu.edu/digitalhistory/

Rubin, Jeffrey. *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests.* Wiley, 1994.

Sitts, M.K. (ed.) (2000). *Handbook for Digital Projects: A Management Tool for Preservation and Access*. Andover, Massachusetts: Northeast Document Conservation Center.

SoliNET Preservation Recording, Copying, and Storage Guidelines for Audio Tape Collections http://www.solinet.net/preservation/leaflets/leaflets_templ.cfm?doc_id=789

Texas Commission on the Arts Videotape Identification and Assessment Guide. Online: http://www.arts.state.tx.us/video/glossary.asp

U.S. National Archives and Records Administration (NARA). (2004). *Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images.* Retrieved on November 18, 2005 from http://www.archives.gov/research_room/arc/arc_info/techguide_raster_june2004.pdf

Usability.gov: Your Guide for Developing Usable and Useful Web Sites. http://usability.gov/

Usability Net. http://www.usabilitynet.org