This master’s paper reports on a survey that gathered data on the current state of metadata creation within digital collections at the University of North Carolina at Chapel Hill (UNC-Chapel Hill) library system. The research was conducted in response to a request by the Metadata Caucus, a working group of librarians responsible for overseeing metadata creation for digital collections in the UNC-Chapel Hill library system. This survey was distributed to librarians and information professionals associated with digital collections at the UNC-CH library system. The following three overriding questions guided the study design: (1) What metadata standards are librarians using when creating digital collections?; (2) How did the librarians make these decisions regarding metadata creation?; and (3) What, if any, changes are the librarians considering regarding metadata creation and usage?

The paper summarizes background research and presents the research methods and findings. The survey yielded 23 survey responses that represented 24 digital collections in the UNC-Chapel Hill library system. The findings of this survey show that the creation of digital collections has risen steadily since 1996. Project managers for digital collections represented have made metadata creation a priority, although 14 respondents indicated that it was not a part of their job description. Metadata standards have been implemented for the majority of the collections, with the most popular standards being MARC, TEI, DC, and EAD. Project managers and librarians indicated that they most often choose metadata standards based on the materials that exist in their digital collections. Results indicated that they also look to other digital collections for ideas on which metadata standards to implement. Most metadata is created manually; and respondents would like more automation, in the future, during the creation process. Respondents indicated a desire to explore alternatives to CONTENTdm.

Headings:

Digital libraries
Metadata creation
Metadata standards
Metadata software
Metadata surveys
DOCUMENTING DIGITAL COLLECTIONS: SURVEYING METADATA STANDARDS USED BY THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL LIBRARY SYSTEM

by
Jennifer D. Joyner

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

Chapel Hill, North Carolina

July, 2008

Approved by

____________________________
Jane Greenberg
Acknowledgements

I would first like to thank the survey participants for their time and involvement in this study. Special thanks go to Natasha Smith, who introduced this project to me and offered guidance and support throughout its duration; Kristin Martin, who assisted in the development and implementation of this project; Josh Berkov, Adrienne MacKay, and Kristin Shaffer, who offered their ideas and support; and Tessa Sullivan, whose metadata survey served as a guide during the development of the survey used for this project. I would also like to thank Jane Greenberg for advising this project. Her guidance and encouragement have been instrumental to its success.
Table of Contents

Acknowledgements ................................................................. 1

List of Figures ................................................................. 3

List of Tables ................................................................. 4

I. Introduction ............................................................... 5

II. Review of the Literature ............................................... 7

III. Methods ................................................................. 18

IV. Results ................................................................. 22

V. Discussion ............................................................... 35

VI. Summary and Conclusion ............................................. 37

References ........................................................................ 40

Appendices ........................................................................ 42
   A. IRB Approval Letter ....................................................... 42
   B. Survey Consent and Survey ........................................... 44
   C. Recruitment Email ....................................................... 51
   D. Survey Email ............................................................. 53
   E. Reminder Email .......................................................... 55
   F. Interview Recruitment Email ........................................ 56
   G. Signed Consent Form for Interviews .............................. 57
   H. Interview Questions ..................................................... 61
List of Figures

Figure 1: Digital Collections' Dates of Creation .......................23
Figure 2: Materials Present in Digital Collections ..................24
Figure 3: Metadata Creation and Job Descriptions....................25
Figure 4: Types of Metadata Standards Being Used
  in Digital Collections ..............................................26
Figure 5: Reasons for Choosing Metadata Standards ...............27
Figure 6: Databases Used for Metadata Storage ....................28
Figure 7: Software Used in Metadata Creation ......................29
Figure 8: Automatic vs. Manual Metadata Creation ..................30
Figure 9: Points of Metadata Creation ............................31
Figure 10: Main Functions of Metadata ..............................32
Figure 11: Future Metadata Standards .................................33
List of Tables

Table 1: Digital Collections' Sources of Funding ...........................23
Table 2: Points of Metadata Creation -- Other Responses ..........31
I. Introduction

Books and archival materials are no longer confined to shelf space in libraries or repositories. Instead, digital library collections are becoming increasingly popular. At the University of North Carolina at Chapel Hill (UNC-Chapel Hill), the university library system recently created the Carolina Digital Library and Archives (CDLA). The purpose of the CDLA is to build a strong collection of resource material in digital form. In addition to housing Documenting the American South (http://docsouth.unc.edu), an established digital publishing program at UNC-Chapel Hill, the CDLA serves as the production center for new digital collections.

As with physical library collections, curators of digital libraries strive for their materials to serve users. Librarians’ use of metadata standards for digital library materials enhances collection accessibility, use, and preservation. With the CDLA beginning to take shape, members of the Metadata Caucus, a working group of librarians at UNC-Chapel Hill responsible for overseeing metadata creation for digital collections, are interested in evaluating the current state of metadata creation within digital collections at UNC-Chapel Hill in order to make suggestions for future changes. In order to assist the Metadata Caucus with their goal of assessing the digital collections at Carolina, a survey was created for those in charge of digital library collections at UNC-Chapel Hill. The primary goal of this survey was to determine how different departments of the library currently create metadata, as well as how the different library departments can work together to create a more cohesive approach to creating metadata. The Metadata Caucus
will benefit greatly by having up-to-date survey results outlining the metadata standards being used by digital collections at UNC-Chapel Hill. While the survey is limited to digital collections associated with the UNC-Chapel Hill library system, it is possible that in the future, the survey can be expanded to include the whole university, as well as to all digital projects underway at member institutions of the University of North Carolina system.

The survey questions addressed the following three research questions: First, what metadata standards do librarians use when creating digital collections? Second, how did the librarians decide to use these metadata standard(s)? Third, what, if any, changes are the librarians considering for the creation and use of metadata? Answers to these research questions uncover the current state of digital collections at UNC-Chapel Hill, which metadata schemas serve these collections, and the projected changes to these collections.

In addition to addressing the three research questions listed above, this survey served the purpose of this research. The purpose of this research has three parts. First, this research explored the recent scholarship concerning metadata and digital libraries. Ideas concerning metadata creation are constantly evolving; thus, it was important to digest the most recent scholarship concerning this topic. Second, the purpose of this research was to investigate the current state of digital collections associated with the UNC library. During the information-gathering process, an up-to-date list of all of the digital collections associated with the UNC library was compiled. Third, at the conclusion of this research project, the Metadata Caucus was able to make explicit statements regarding the metadata creation within digital collections associated with the UNC library.
II. Review of Literature

Mastering Metadata

Digital libraries use many metadata standards. Given the range, it is useful to understand the definition of metadata and the most-commonly used metadata schemes. Considering the number of available metadata standards, however, this is no easy task. Authors at Niso press offer the most useful definition of metadata. In Understanding Metadata, the authors stated that

[m]etadata is structured information that describes, explains, locates or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data or information about information (1).

The most common metadata schemes used in digital libraries are Dublin Core, the Text Encoding Initiative (TEI), Metadata Encoding and Transmission Standard (METS), Metadata Object Description Schema (MODS), the Encoded Archival Description (EAD), and the Learning Object Metadata Standard (IEEE LOM) (Niso Press 1-10). These are only a few of the hundreds of metadata schemes available for use in digital collections.

There has been some debate among scholars about the format of materials to which metadata applies. Some scholars argue that metadata is exclusive to electronic information. Other scholars, however, contend that metadata is simply an extension of actions that have been done previously, such as cataloging. Some metadata schemes can be used on digital and physical objects. Thus, the definition of metadata is not concrete (Greenberg 20).
El-Sherbini and Klim examined this issue in “Metadata and cataloging practices.” They suggested that even though the term metadata did once apply exclusively to “non-traditional formats,” the term now applies to “any kind of standardized descriptive information about resources, including non-digital formats” (El-Sherbini and Klim 238). They offered a glimpse into the most commonly used metadata schemes at the time of publishing. In addition, they provided a discussion of traditional cataloging practices and the current creation of metadata. For the sake of this research study, the term metadata applies solely to digital materials, as the survey subject pool only contained librarians associated with digital collections at UNC-Chapel Hill.

Despite the creation and implementation of many metadata schemes since the early 1990s, scholars have not extensively researched the “sum of these activities” (Greenberg 17). The Metadata Objectives, Domain, and Architectural Layout (MODAL) framework that Greenberg set forth in her article, “Understanding Metadata and Metadata Schemes,” may help researchers studying metadata schemes to gain a clearer understanding of the structure of the many metadata schemes available. Understanding the structure and use of the schemes can help creators of digital collections chose the right metadata standard for the digital objects in the collection.

The three components of the MODAL framework include the heart of all metadata schemes. The metadata objectives “identify the overall aims and goals of the scheme” and the principles are the “rules or means” for meeting those objectives (Greenberg 25). The domain is “the realm in which the scheme operates” (Greenberg 29). It includes the environmental domain, which is the community it serves; the object class domain, which is the grouping of similar objects by type; and the object format domain,
which is the object’s composition. The architectural layout describes the structural design of the scheme (Greenberg 31). Greenberg’s MODAL framework is a comprehensive guide to understanding the basics of the many metadata schemes available. Survey questions regarding the metadata schemes used at UNC-Chapel Hill digital collections were drawn from each of the three components of the MODAL framework.

*Metadata and Digital Library Design*

As digital libraries become more prevalent, scholars are learning more about the characteristics and components of digital libraries. Despite the growing numbers of digital libraries, only a few scholars have studied what characterizes a quality digital library. In “‘What is a good digital library?’ – A quality model for digital libraries,” Goncalves, Moreira, Fox, and Watson suggested that five concepts create a solid framework for a quality digital library. These five components comprise the 5s model: Streams, Structures, Spaces, Scenarios, and Societies (Goncalves et al 1416). By using these five components as a framework, digital libraries can ensure that the digital objects in their collections are accessible; are properly preserved through migration, emulation, wrapping or refreshing; are relevant to users’ information needs; are significant; and maintain their timeliness. Goncalves, et al, suggested that metadata format serves as part of the structure in the 5s framework. The greater the accuracy, completeness and conformance of the metadata formats for digital objects, the greater the quality of the digital library.

Goncalves, Moreira, Fox, and Watson conducted a focus group in 2007 to study the usability of the 5s model. Three librarians with digital library experience formed the
focus group. Goncalves, et al, sought to answer the following four questions about the 5s model:

1) Are you able to understand the 5s model?
2) How does it relate to (your) library world?
3) How do the proposed indicators relate to your practices in the library?
4) Would you be willing to apply these measures to your (digital) libraries?

While the users made various comments regarding the 5s model, members of the focus group devoted a great amount of time and attention to the issues of cataloging and collection completeness. Goncalves, et al, stated that “it was thought that in some cases, for instance catalogs based on the Dublin Core (15 attributes), this indicator (structure) makes sense” (Goncalves et al 1434).

Goncalves, Moreira, Fox, and Watson are not the only researchers who have stated the importance of a solid metadata structure for digital libraries. In their article “Context and Meaning: The Challenges of Metadata for a Digital Image Library within the University,” Attig, Copeland, and Pelikan linked metadata structure with accessibility. In their article, they stated that university faculty and staff are interested in using digital objects in their research and teaching (Attig, Copeland and Pelikan 251). In order for users to be able to access and use digital collections, however, solid metadata standards must in place (Attig, Copeland and Pelikan 253).

Attig, Copeland and Pelikan suggested that no metadata standard exists that will answer the needs of all digital collections. Indeed, Tony Gill asserted that “no single (meta)data schema fits all” (Attig, Copeland and Pelikan 253). Often, several metadata standards may be combined to fit the needs of a specific collection. In order for a digital collection to be useful and successful, it is necessary for the creators of those digital
collections to form a metadata schema “capable of both supporting the needs of all the disciples that comprise the university and interoperating with learning management systems throughout the university” (Attig, Copeland and Pelikan 259).

*Other Metadata Surveys*

Although it is clear that users find metadata schemes useful, one is left to wonder how often digital libraries and other creators of digital objects use metadata schemes to describe their collections. Fei Yu’s “Metadata Schema Used in OCLC Sampled Web Pages,” addressed this question. In his study, Yu examined 16,383 web pages from 200,000 OCLC sampled web pages in 2000. Yu intended to find which metadata schemes creators were using, as well as if one metadata scheme dominated the others. Yu found that only 8.19 web pages used meta tags; in addition, the only three types of tags used were Dublin Core, description tags, and keyword tags. Description tags and keyword tags were the most popular tags used on the sample web pages. Yu observed that the use of description and keyword tags increased from 1997 to 2000, which leads researchers to assume that this trend increased through this decade. Thus, with the use of meta tags increasing, digital libraries should be at the forefront of this trend in information sharing.

Yu’s research, while it does not specifically address digital libraries, points to one of the important aspects of the UNC-Chapel Hill CDLA survey: which metadata schemes are librarians using the most? It is important to see if Dublin Core, keyword tags, and description tags are still as prevalent in 2008 as they were in 2000. Thus, this study provides for an aspect of comparison.
The UNC-Chapel Hill CDLA survey gathered information regarding the interoperability of the digital collections at the University. Philip Hinder conducted a similar study in 2005. Hinder administered a questionnaire to Australian institutions (libraries, museums, and government agencies) with digital collections. He sent questionnaires to forty institutions that had been identified by the National Library of Australia as having digitization projects. Of these forty, nineteen responded (Hinder 289-290).

Hinder’s goal was to make concluding statements about the metadata standards being used by digital projects underway in Australia. More specifically, he wished to examine the interoperability of these digital collections. In the questionnaire, Hinder asked twenty-five questions related to metadata and interoperability. These questions included check-box answers, as well as some open-ended questions (Hinder 290).

Hinder found that the types of metadata standards being utilized in Australian digital collections did not vary greatly. Most digital collections used one of the following: MARC21, DublinCore (DC), Australian Government Locator Service (AGLS), and Encoded Archival Description (EAD). In addition, all but one institution applied more than one (Hinder 291). Respondents to the questionnaire asserted that established metadata standards were important to the creation and operability of their digital collections (Hinder, 295).

This study applies to the research question outlined above because it specifically addresses metadata standards being used in digital collections and projects. When surveying those involved with digital collections at the UNC-Chapel Hill library system, it was necessary to gather both factual and opinion-based information related to the
metadata standards and interoperability of the digital projects currently underway. In addition, the fact that most digital projects addressed in Hinder’s survey used more than one metadata standard created an area in which a comparison can be made. The questions Hinder raised in this study are similar to those raised in the study discussed in this paper.

Another similar survey, conducted in 2006, investigated digital projects underway at institutions that are members of the Association of Research Libraries (ARL). Kahl and Williams surveyed digital collections at English-speaking ARL libraries with the goal of evaluating digital collections at other universities in order to make suggestions for the creation of a digital collection at Illinois State University. Kahl and Williams conducted this research as members of the Digital Resources Management (DReM) task force at Illinois State University’s Milner Library.

This survey was comprised of two steps. First, Kahl and Williams searched digital collections only at twelve Big 10 and University of California system libraries. After this initial survey, Kahl and Williams expanded this search to include all academic ARL libraries that were English speaking. They examined a total of 111 academic institutions, which included a final total of 1117 digital projects.

When evaluating the digital collections at the above-listed institutions, Kahl and Williams sought to observe the accessibility, the ability to browse and search the collection, content, and metadata of the digital collections. Specifically, they sought to answer the following questions:

1. How were patrons able to access the digital projects? Were links available on the library home page?
2. Did patrons have the ability to browse and search the digital collection?
3. Was metadata provided?
4. What digital collections/projects were complete? What type of project was it? (365)
Kahl and Williams gathered this information both on the institution and the collection.

In total, Kahl and Williams found that of the 111 ARL institutions, 89 (80.2 percent) had digital collections or projects. Each library had an average of 12.6 digital projects, for a total of 1117 digital projects. From their findings, Kahl and Williams found that fewer than half of the digital projects were linked from the library’s main web page. Slightly more than half provided search options. Many were not listed in the library catalog. Overwhelmingly, though, the digital collections were open for viewing. Only 2.9 percent had restricted access (Kahl and Williams 366-367). From these findings, Kahl and Williams suggested the following for the digital projects at Illinois State University:

(1) The digital collection should have a link from the library’s main web page;
(2) The digital collection should have a search function;
(3) The collection should also have browsing options;
(4) The collections should use established metadata standards for all items in the collection.

While the authors chose ARL institutions because of their funding, it would also be useful to explore the digital collections of non-ARL member academic institutions. The same process could be taken from representative institutions, and the numbers could be compared with those of ARL institutions. These results could then support or dispute Kahn and Williams’ findings. In addition, Kahl and Williams dropped several data points during their study. These included the size of the collections, type of digital software, and the level of metadata provided for each digital project. While Kahl and Williams had valid reasons for dropping these data points, it would be useful for researchers to reexamine these points.
This particular study was of importance during the creation of a survey for the Digital Library at UNC because it provided a solid model for examining the digital projects and collections established or currently underway at the University. By following Kahl and Williams’ steps, it would be possible to make similar observations about the varied digital projects at UNC-Chapel Hill.

Another survey very useful to this project is the Automatic Metadata Generation Applications (AMeGA) Project survey, which was administered by Greenberg, Spurgin, and Crystal, endorsed by AMeGA Task Force members, and submitted to the Library of Congress in 2005. Greenberg, Spurgin and Crystal sought to meet these three goals: (1) Evaluate the ability of content creation software, integrated library systems, and automatic generation applications; (2) Administer a survey to metadata professionals to see how they view automatic metadata generation and which functionalities they would prefer to for automatic metadata generation applications to possess; and (3) Create a final report regarding the recommended functionalities for an automatic metadata generation application to be created by the Library of Congress (Greenberg, Spurgin and Crystal vi).

While the primary finding of this report was “that there is a disconnect between experimental research and application development,” the most useful portion of this research was that which corresponds to Goal 2 (Greenberg, Spurgin and Crystal vii). This survey sought to query metadata professionals on their experience with metadata, their current work with metadata and metadata creation, their knowledge and opinion of metadata generation, their knowledge and opinion of Dublin Core, and what they would like to see generated automatically from metadata generation applications (Greenberg, Spurgin and Crystal viii). From their results, which included 217 participants, they found
that metadata professionals are utilizing a wide variety of metadata standards, the most popular including Machine Readable Cataloging (MARC), Dublin Core (DC), Encoded Archival Description (EAD), Gateway to Educational Materials (GEM), Metadata Object Description Schema (MODS), Text Encoding Initiative (TEI), Institute of Electrical and Electronics Engineers standard (IEEE LOM), and the Government Information Locator Service (GILS) (Greenberg, Spurgin and Crystal viii, 23). Over 81 percent of participants used at least one or more systems for metadata creation within their profession, with participants using up to 7 systems (Greenberg, Spurgin and Crystal viii, 24). While participating metadata professionals preferred automatic metadata generation, over 96 percent felt it was necessary for humans to manually review, and possibly edit, that which was automatically generated (Greenberg, Spurgin and Crystal ix). Of the respondents, 63.4 percent had formal quality control activities in place to assure that metadata content complied with standards in place by that particular institution or the metadata standard(s) in place.

The survey reported on in this paper sought to query similar participants, namely, metadata professionals associated with digital objects that are online. While it did not seek to find out specifically about automatically generated software, it did inquire as to which metadata standards are used, the software used to create metadata regarding the digitized objects, and the methods used to create that metadata. Thus, the survey responses available in this report serve as points of comparison. This survey makes it possible to measure change with metadata standards over the past three years, including change in automatic metadata creation, number of metadata standards being used by metadata professionals within one or multiple organizations, and the types of metadata
standards or applications that are being utilized by metadata professionals. In addition, it is possible to compare how the metadata creation within digital collections at the UNC library system varies from or is similar to that of other institutions.
III. Methods

*General Description*

In order to obtain information regarding metadata creation and use at UNC-Chapel Hill digital collections, a survey was designed and administered online using the Qualtrics software, which is available through the Odum Institute of the University of North Carolina at Chapel Hill. The recipients of the survey were digital project managers whose digital collections are associated with the UNC library system. Methods for identifying these recipients are discussed below.

*Participants*

A survey is the most appropriate method for this particular research question because it one of the easiest, most readily available ways to gather comprehensive information about metadata creation and use. While it would be possible to view source code for most digital collection web pages, it would not provide full and accurate information about all metadata creation and usage at that particular digital collection. Thus, it is important to survey those most familiar with metadata creation and usage of that collection.
For this project, it was estimated that there would be no more than 30 subjects. The gender, ethnicity, age and race of the participants varied. It was required that all participants be over the age of 21 and professional librarians. In particular, the subject population consisted of librarians at UNC-Chapel Hill who work with digital library collections. The subjects were likely to be project managers of those digital collections or the person responsible for the publishing of that digital collection. It is important to note that each collection representative received one email announcing the survey. It is possible that one person may represent more than one collection. If this is the case, that person was required to complete one survey per collection.

Because of the small size of library digital collection staffs at UNC-Chapel Hill, it is possible that someone could deduce the identity of the participant, even though pseudonyms will be used in place of individuals’ names. To reduce this risk, all data, both internal to the library and that which may be made public during publishing, contains only pseudonyms for the participants’ names. Outside of library use, all collection names were de-identified. All participants were notified of this risk at the start of the survey and in the signed consent form.

The survey has a specific target population: the project manager or library personnel most familiar with digital collections at the UNC library system. To collect the names and contact information of these individuals, the Principal Investigator gathered a list of all of the digital collections at the UNC library system by searching the UNC-Chapel Hill library website (www.lib.unc.edu) for links to all digital collections and recorded information regarding the collections, the project managers, and their contact information. Secondly, the Principal Investigator asked Natasha Smith and Kristin
Martin, advisors on this project, for a list of all known digital collections and their project managers not included on the initial list. Using these two methods, I was able to procure the names of the digital collections and the individuals responsible for them.

Procedures

After receiving Institutional Review Board (IRB) approval, the Principal Investigator sent a recruitment email to identified individuals seven days before administering the survey to inform them of the intent of the research project. This email, as well as all other supporting documents, is available in Appendices A through E. The Principal Investigator sent the survey to the recipients on February 25, 2008. The recipients had two weeks to complete the survey. The Principal Investigator sent an email reminder to all recipients on March 3, 2008.

Before starting the survey, the participant was asked to view an information page. The purpose of this information page was to obtain consent from the participant. By clicking “yes” at the conclusion of this section, the participant implied consent and acknowledged that they are associated with a digital library collection at UNC-Chapel Hill. The questions sought information regarding the current use of metadata at the digital collection, the relationship of the digital collection to other existing digital collections, and future plans for metadata creation within the digital collection. The survey also asked for the identification of the project manager associated with that collection. All participants will be given the opportunity to review the final report at the conclusion of the study if the Principal Investigator chooses to publish the report. Any requests for removal of data will be honored.
It is important to note that this survey is part of a larger, ongoing project. This larger project includes an hour-long follow-up interview with survey participants to clarify and expand upon answered supplied in the survey. A question was included on the survey asking survey respondents of their interest in participating in the interviews. Participation in the follow-up interview will be strictly voluntary, and participants will be asked to sign another consent form. Each survey recipient who indicated interest in participating in a follow up interview will be sent an email to initiate conversation to set up a date and time for the interview. All supporting documents related to the interview are in Appendices F through H.

Survey Description

The initial, self-administered survey was digitized using Qualtrics software. Intellectually, the survey consisted of three sections. The first section identified the materials available in the collection, the characteristics of those materials, and the metadata creation policies and software used. The second section recorded information about the digital collection and its relationship to other digital collections. The third section inquired about the name of the digital collection, the funding and association of the digital collection, and the name of the project manager. In addition, the survey recipient was given an opportunity to state if were interested in participating in a follow-up interview. If they were not, by clicking no, they were taken to the end of the survey. All information in section 3 is optional. Each question in the survey was designed to show no bias toward any one metadata standard. In addition, the survey did not favor one technology above another.
IV. Results

About the Collections

The survey yielded 23 survey responses that were suitable for data analysis. The responses represented 24 collections. It is important to note that although the respondents were instructed to complete one survey per digital collection, in one case, multiple collections were represented in one survey, as each collection represented by the response had the same information. The respondent informed the principal investigator of this fact, and with the respondent’s permission, the principal investigator separated the response for the purpose of analyzing the data.

It is also important to note that collections 4 and 10 are the same collection; however, the respondents were describing two very different aspects of this digital collection. Collection 4 represents an internal database created for audio preservation purposes; Collection 10 represents the project, its public interface, and the database behind that interface. Since the responses were describing two very different aspects of this project, and two different project teams worked on the different aspects of the collection, they were considered separate collections.

The projects represented have creation dates than span over the last decade. The oldest collection was created in 1996, and a large number of collections are currently in various stages of implementation and publishing. Figure 1, shown below, shows the trend in digital library creation within the UNC library system. It is important to note that
respondents entered this information in a text box; in some cases, the respondent entered a date range. For this data, only the creation date (the first date listed) was used.

Figure 2: Digital Collections' Dates of Creation

The digital collections represented are funded primary through grants. There are some collections, however, that are funded privately or through general library or departmental funds. Some projects have multiple sources of funding. Table 1, shown below, displays the break down of the funding for these digital collections represented in this survey.

<table>
<thead>
<tr>
<th>Source of Funding</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Departmental</td>
<td>2</td>
</tr>
<tr>
<td>Grant (not specific)</td>
<td>2</td>
</tr>
<tr>
<td>LSTA grant</td>
<td>4</td>
</tr>
<tr>
<td>LSTA grant with local funding</td>
<td>1</td>
</tr>
<tr>
<td>IMLS grant</td>
<td>1</td>
</tr>
<tr>
<td>IMLS and library funding</td>
<td>2</td>
</tr>
<tr>
<td>NEH</td>
<td>1</td>
</tr>
<tr>
<td>Grant, &quot;National Digital Library Competition&quot; by the Library of Congress</td>
<td>2</td>
</tr>
<tr>
<td>Library funding</td>
<td>6</td>
</tr>
<tr>
<td>Library and University Funding</td>
<td>2</td>
</tr>
<tr>
<td>Private funding</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1: Digital Collections' Sources of Funding
The respondents represented digital collections that contain many different types of materials. Figure 2, shown below, illustrates the various types of materials represented in these digital collections. Manuscripts are the most popular items, followed by published books and maps.

Figure 3: Materials Present in Digital Collections

The Respondents

It is assumed that since the emails only went to individuals who were associated with the digital collections, the survey respondents were either project managers or librarians closely associated with that collection.
Of those surveyed, 10 respondents indicated that metadata creation was part of their job description. Largely, however, metadata creation was not part of their job description. As indicated in the Figure 3 below, 14 of the respondents indicated that creating metadata was not part of their job description.

![Figure 4: Is Metadata Creation Part of Your Job Description?](image)

**The State of Metadata Creation**

Despite the fact that creating metadata is not part of the job description for over half of the respondents, it is clear that metadata is being created for the digital collections associated with the UNC library system. According to respondents, metadata creation is well under way for digital collections. For their digital collections, 96 percent of the respondents reported that metadata is currently being created for their digital collections. Only one collection (Collection 13) indicated that metadata was not currently being created for the digital collection. Despite the fact that most collections are currently
creating metadata, it is not required for all collections. The majority, however, do require it. Indeed, 96 percent of the respondents indicated that metadata creation is required for the collection. Only 1 respondent (Collection 2) indicated that it is not required.

For those collections that do create metadata, respondents indicated that a wide variety of metadata standards are used. Figure 4, below, shows that most collections use Machine Readable Cataloging (MARC) (15 respondents or 63 percent) and Text Encoding Initiative (TEI) (12 respondents 50 percent). A fairly high percentage (38 percent) used metadata standards not listed on the survey. These standards include Thesauri of Graphic Material (TGM), Library of Congress Subject Headings (LCSH), Electronic Theses and Dissertations (ETDs) metadata standard and NISO Metadata for Images in XML Schema (MIX).

![Figure 5: Types of Metadata Standards Being Used in Digital Collections](image-url)
These standards were chosen for various reasons. Overwhelmingly, respondents indicated that these standards were chosen because they were best suited for the collection material. Also, respondents indicated that they used the standard because other collections with similar materials used that standard. The results are listed below in Figure 5.

**Figure 6: Reasons for Choosing Metadata Standards**

For those respondents who chose other, the primary response was that CONTENTdm maps to Dublin core, with 4 respondents listing this as their answer. The Respondent for Collection 3 noted that their metadata standards were selected because “[i]t was important to the project to use a nationally-recognized standard.”

When asked where their collections store metadata, 79 percent (19) of respondents listed that the metadata was stored in a database. Eleven (or 46 percent) noted that it was stored within the object, and 13 percent (3) responded that it was stored in a spreadsheet. Seven respondents listed areas other than those listed; these include
CONTENTdm (Collections 3, 5, 8, 9 and 13), in EAD and METS (Collection 7), and in METS and MODS files (Collection 10).

For those who indicated that the metadata was stored in a database, respondents were asked to indicate what type of database was used. Their responses are listed in Figure 6 below.

![Figure 7: Databases Used for Metadata Storage](image)

When asked what type of software they used to create metadata, respondents indicated that a wide variety of software programs were used. Their responses are shown in Figure 7 below.
For those who chose “other,” the primary response was Author/Editor SGML software (Collections 17, 18, 20, and 22). Other responses include Filmmaker Pro (Collection 1), Google Docs (collection 3 and 8), Google Spreadsheets (Collections 8 and 9), OCLC Connexion (Collection 6), and administrative web forms (PHP/Python) (Collection 14).

Twenty-one respondents answered the question regarding manual vs. automatic creation of metadata. From their responses, it is evident that those who work with UNC digital library collections primarily manually create metadata. Their responses are listed below in Figure 8.
Figure 9: Automatic vs. Manual Metadata Creation

Other responses include the following: “Mostly manual, but with a shell created through an automated script; mostly manual, but CONTENTdm supplies some technical metadata automatically; and, actual metadata is created manually, but the process for applying schemas will be automated by extracting necessary data from the database and formatting to the given specification.”

Respondents indicated that metadata is created during multiple stages of the digitization and online publication process. For most of the collections, metadata is created during the content creation process and before online publishing. For almost half of the collections, metadata creation happens after the online publication process. The breakdown is shown in Figure 9 below, and their text responses that clarified the “other” response are listed in Table 2.
Table 2: Points of Metadata Creation -- Other Responses

Respondents indicated that metadata serves a variety of purposes. When asked what they thought was the main function of the metadata assigned for the materials in their digital collections, respondents indicated that interoperability and resource discovery were the two main purposes of the metadata. Only two respondents noted that it was for preservation. Figure 10, below, illustrates their responses.
Three respondents listed alternate answers. In each case, the respondents indicated that they could only choose one answer. The formatting of the question indicated that they would be able to select all; however, the survey only allowed for them to choose one. These respondents indicated that they would have liked to have selected more, if not all, of the answers. One respondent wrote, “All of the above (can only click one button).”

**Future Metadata Developments**

While all respondents indicated that metadata creation had been or is currently being created for the digital collections, they also stated that there were metadata standards they would like to implement for their projects in the future. Overwhelmingly, respondents indicated that they would like to implement Metadata Object Description Schema (MODS) and Metadata Encoding and Transmission Standard (METS). Other popular standards include MARC and TEI headers. For those who chose to add a
standard not listed, they indicated that they were interested in Encoded Archival Context (EAC), ETDs, and MIX. The full range of responses is in Figure 11 below.

![Figure 12: Future Metadata Standards](image)

Respondents also indicated that they would like to see changes made when it comes to metadata creation. Three themes were present in their responses: more automation, more standardization, and a move towards MODS/METS. For some collections, this would mean using these standards place of CONTENTdm. Those who indicated that they would like experiment with alternatives to CONTENTdm noted that they would like to be sure that the metadata in CONTENTdm was safe and that “it could one day be successfully migrated into another content management system” (Collection...
8). The representative for Collection 8 also noted that content management systems might not be the way to go. The respondent wrote,

   I would like to see the metadata used in more dynamic ways, such as for making tag clouds, RSS feeds, and other special features. I would like to see a space for user-contributed metadata, such a commenting feature. These latter two points, however, are more like limitations of the content management platform (CONTENTdm).

Other respondents noted that the move towards METS and MODS is currently being planned (Collections 10, 11, and 12), which is a sign that project managers are realizing the advantages of metadata creation and are making its improvement a priority.
V. Discussion

Overall, the results of this survey were not surprising. The first documented digital collection (1996) corresponds with the birth of Documenting the American South (http://docsouth.unc.edu), a digital publishing initiative started by Natasha Smith and supported by the UNC library system. It is also not surprising that the number of digital collections created was at its highest in 2007, which was the year in which the Carolina Digital Library and Archives was formally created. Thus, the growth of digital collections is indicative of changes within the UNC library system.

Also, it was not surprising that respondents were most interested in implementing METS and MODS for their digital collections. Both are Library of Congress standards, and both are relatively new to the metadata community and have gained popularity, especially for use in digital libraries, during the past five years. Others, such as Dublin Core and MARC, are more established and have enjoyed popularity for a longer period of time. It is not surprising that these metadata professionals are interested in learning how newer metadata standards can serve their digital collections.

However, it was surprising to learn that metadata creation was not a part of many of these information professionals’ job descriptions. While 10 respondents indicated that metadata creation was part of their job descriptions, 14 indicated that it was not. This is surprising because the survey results reflect interest and awareness of metadata trends. Nineteen respondents indicated that they chose the metadata standards currently in use because it suited the collection material. In order to best assess which standard best suits
certain materials, it is necessary for information professional to research and examine numerous metadata standards. Thus, it is evident that respondents, even though they are not required to create metadata, are interested in doing so.

**Limitations**

While a survey was the best way to gather this data, it did have its limitations. Because project managers were required to provide some identifying information, they may have been unwilling to share unpolished ideas or unpopular opinions about metadata. Also, because it was necessary to provide some identifying information, some potential respondents may have been hesitant to respond and may have chosen not to participate.

In addition, it is possible that not all digital collections were represented in this study. While the Principal Investigator attempted to compile a complete list of digital projects and project managers, it is possible that past digital projects, newly-formed digital projects, or projects that do not yet have a web presence could have been overlooked. It is also very likely that representatives for some digital collections chose not to participate. Thus, those collections would not be represented in this study.
VI. Summary and Conclusion

The establishment of the Carolina Digital Library and Archives underscores the library system’s commitment to the creation of digital collections. As the CDLA builds more digital collections, the Metadata Caucus Working Group seeks to better understand the metadata schemes used for each established digital collection. This survey serves as a response to their request for a formalized gathering of information regarding metadata creation and usage at UNC library digital collections. This knowledge will help them improve the interoperability, accessibility, and longevity of these collections.

Currently, no other survey exists that outlines the metadata schemes in use by digital libraries at UNC-Chapel Hill. However, studies do exist in the library science community that survey digital library collections at other colleges, universities, and archival repositories. These studies served as the framework for the CDLA survey. Broadly, this survey contributes to the scholarship of metadata and digital libraries. Specifically, the results of this survey will prove useful to those involved in making decisions about these collections at the CDLA. In addition, the results of this survey will assist the CDLA at UNC-Chapel Hill as they make an in-house policy regarding metadata standards.

Those involved with the CDLA and those working with digital collections at UNC-Chapel Hill are not the only interested parties in such research. Other universities throughout the United States are also facing the question of which metadata standards to implement for their digital projects and collections. The librarians and information
professionals who assist these digital collections will be greatly aided by having a study that shows the workings of digital collections underway at another major university library system.

The findings of this survey show the creation of digital collections has risen steadily since 1996. The importance of digital collections at UNC-Chapel Hill is only magnified by the recent collection of the Carolina Digital Library and Archives.

Project managers for the represented digital collections have made metadata creation a priority, even if it is not necessarily part of their job description. Metadata standards are being implemented for almost all of the collections, with the most popular standards being MARC, TEI headers, DC and EAD. Project managers and other associated librarians have numerous reasons for choosing these standards; however, they most often choose the standard(s) based on materials that exist in the digital collections. Also, results indicate that they often look to other digital collections for ideas on which metadata standards to implement.

Survey results show that currently, most metadata is created manually. It is not surprising, then, that one of the requests that respondents listed on their “metadata wish list” is more automation during the metadata creation process. Also, respondents indicated a move away from CONTENTdm and towards implementation of METS and/or MODS. Indeed, project managers are wasting no time on achieving the goals on their “metadata wish list.” Respondents indicated that they are currently in the process of moving towards METS and/or MODS implementation.
Future Research

The immediate next step for this research project is to complete the second phase of the study: follow-up interviews with survey participants. Over 22 collection representatives indicated at the conclusion of their survey that they are interested in participating in a follow-up interview. These interviews will be conducted in July and August 2008, and the results will be recorded in another publication.

While this survey seeks only to query digital collections associated with the UNC library system, future researchers can use this survey to query new digital collections at UNC-Chapel Hill, digital collections throughout the UNC system, and digital collections affiliated with colleges and universities outside of the UNC system. This would allow for comparisons among various digital library collections associated with academic libraries.

In addition, this survey could be redistributed to the same participants at a later date to see how metadata creation and usage has changed over time. This would allow researchers to follow the trend from DublinCore and CONTENTdm to METS and MODS. Perhaps by the time this survey is re-administered, newer schemes will have become popular. This survey has the potential to serve UNC-Chapel Hill and other academic institutions both now and in the future.
References


Appendices

Appendix A: IRB Letter

OFFICE OF HUMAN RESEARCH ETHICS
Medical School Building 52
Memorial Farm Road
CB #7064
Chapel Hill, NC 27599-7064
(919) 966-3113
Website: ohsr.unc.edu
https://my.research.unc.edu for IRB status
Federalwide Assurance (FWA) #3301

To: Jennifer Joyner
Academic Affairs Library
CB:3360

From: Behavioral IRB

Authorized signature on behalf of IRB

Approval Date: 2/12/2008
Expiration Date of Approval: 2/10/2009

RE: Notice of IRB Approval by Expedited Review (under 45 CFR 46.110)
Submission Type: Initial
Expedited Category: 7. Surveys/interviews/focus groups, 6. Voice/image research recordings
Study #: 68-0147

Study Title: Documenting and Describing Digital Collections: Surveying Metadata Standards Used by the University of North Carolina at Chapel Hill Library System

This submission has been approved by the above IRB for the period indicated. It has been determined that the risk involved in this research is no more than minimal.

Study Description:

Purpose: To examine the state of metadata creation within digital collections associated with the UNC Chapel Hill library system.

Participants: 30 librarians at UNC Chapel Hill associated with digital collections and metadata creation.

Procedures: Administer on-line surveys and conduct interviews.

Regulatory and other findings:

This research meets criteria for a waiver of written (signed) consent according to 45 CFR 46.117(c)(2) for the survey component of this study only.

Investigator’s Responsibilities

Federal regulations require that all research be reviewed at least annually. It is the Principal Investigator’s responsibility to submit for renewal and obtain approval before the expiration date.
You may not continue any research activity beyond the expiration date without IRB approval. Failure to receive approval for continuation before the expiration date will result in automatic termination of the approval for this study on the expiration date.

When applicable, enclosed are stamped copies of approved consent documents and other recruitment materials. You must copy the stamped consent forms for use with subjects unless you have approval to do otherwise.

You are required to obtain IRB approval for any changes to any aspect of this study before they can be implemented (use the modification form at ohre.unc.edu/forms). Should any adverse event or unanticipated problem involving risks to subjects or others occur it must be reported immediately to the IRB using the adverse event form at the same web site.

******************************************************************************
Stuart Rennie, Ph.D.
Office of Human Research Ethics
Co-Chair, Behavioral Institutional Review Board
CB# 7097, Medical School, Bldg 52
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599-7097
aa-irb-chair@unc.edu
phone 919-966-3113; fax 919-966-7879
******************************************************************************

CC:
Jane Greenberg, School Of Info
Natalia Smith, Academic Affairs Library
Kristin Martin, Academic Affairs Library
Marcia Tauber, (School of Information and Library Science), Non-IRB Review Contact
Appendix B: Survey Consent and Survey

University of North Carolina-Chapel Hill
Consent to Participate in a Research Study
Adult Participants
Social Behavioral Form

IRB Study # _08-1047____________________
Consent Form Version Date: February 5, 2008

Title of Study: Documenting and Describing Digital Collections: Surveying Metadata Standards Used by the University of North Carolina at Chapel Hill Library System

Principal Investigator: Jennifer Joyner
UNC-Chapel Hill Department: School of Information and Library Science
Email Address: jdjoyner@email.unc.edu
Faculty Advisor: Dr. Jane Greenberg
Funding Source: None

Study Contact telephone number: 919-967-3971
Study Contact email: jdjoyner@email.unc.edu

What are some general things you should know about research studies?
You are being asked to take part in a research study. To join the study is voluntary. You may refuse to join, or you may withdraw your consent to be in the study, for any reason, without penalty.

Research studies are designed to obtain new knowledge. This new information may help people in the future. You may not receive any direct benefit from being in the research study. There also may be risks to being in research studies.

Details about this study are discussed below. It is important that you understand this information so that you can make an informed choice about being in this research study.

What is the purpose of this study?
The purpose of this research study is to learn about the state of metadata creation within digital collections associated with the UNC-CH library system. This research will gauge the current state of metadata creation and predict future trends. This information will help metadata librarians at UNC-Chapel Hill make suggestions for future changes regarding metadata creation.

You are being asked to be in the study because you have been identified as an UNC-Chapel Hill librarian who works directly with digital collections.
**How many people will take part in this study?**
If you decide to be in this study, you will be one of approximately 30 people in this research study.

**How long will your part in this study last?**
The online survey should take between fifteen and twenty minutes to complete. Participants will be given the opportunity to participate in follow-up interviews. The Principal Investigator will contact subjects who have expressed interest in the follow-up survey in early March 2008. Interviews will take approximately one hour. This hour includes the review of the signed consent form and a review of their initial survey with questions based on their previous answers. Survey participants are not required to participate in the follow-up interview.

**What will happen if you take part in the study?**
If you take part in this study, you will be asked to complete the following online survey, which consists of 25 questions. The survey should take no longer than 20 minutes to complete. The primary investigator will seek to protect the anonymity of the participant through the use of pseudonyms. All questions are optional. Collection names will be retained in internal reporting; however, participant names will be de-identified. For all reporting outside of the library, both participant names and collection names will be de-identified.

In the concluding section of this survey, you will be given the opportunity to express interest in a follow-up interview. Participation in the follow-up interview will be completely voluntary. This consent form applies only to this online survey. You will be asked to sign another consent form before participating in any addition part of this study.

**What are the possible benefits from being in this study?**
Research is designed to benefit society by gaining new knowledge. You may not benefit personally from being in this research study.

**What are the possible risks or discomforts involved from being in this study?**
There may be uncommon or previously unknown risks. You should report any problems to the researcher.

**How will your privacy be protected?**
Because of the small size of library digital collection staffs at UNC-Chapel Hill, it is possible that someone could deduce the identity of the participant, even though pseudonyms will be used in place of individuals’ names. To reduce this risk, all data, both internal to the library and that which may be made public during publishing, will contain only pseudonyms for the participants’ names. Outside of library use, all collection names will also be de-identified. Participants will be labeled as Participant 1, Participant 2, etc. Collections will be labeled collection A, Collection B, etc.

The Principal Investigator will create two separate reports using the data collected from
this survey. The first will include an internal report that will be shared with the Metadata Caucus, which is a working group of librarians at UNC-Chapel Hill. The second report will be made available for publishing outside of UNC-Chapel Hill. Before any report or publishing of data, participants will have an opportunity to review the report. Requests for deletion of identifying information will be honored.

All data collected will be stored on the Principal Investigator’s personal laptop, which is password protected. All data transmission will be performed on a secure email server. Only the Principal Investigator and research collaborators will have access to identifiable data.

Participants will not be identified in any report or publication about this study. Although every effort will be made to keep research records private, there may be times when federal or state law requires the disclosure of such records, including personal information. This is very unlikely, but if disclosure is ever required, UNC-Chapel Hill will take steps allowable by law to protect the privacy of personal information. In some cases, your information in this research study could be reviewed by representatives of the University, research sponsors, or government agencies for purposes such as quality control or safety.

**Will you receive anything for being in this study?**
You will not receive anything for taking part in this study.

**Will it cost you anything to be in this study?**
There will be no costs for being in the study

**What if you are a UNC employee?**
Taking part in this research is not a part of your University duties, and refusing will not affect your job. You will not be offered or receive any special job-related consideration if you take part in this research.

**What if you have questions about this study?**
You have the right to ask, and have answered, any questions you may have about this research. If you have questions, or concerns, you should contact the researchers listed on the first page of this form.

**What if you have questions about your rights as a research participant?**
All research on human volunteers is reviewed by a committee that works to protect your rights and welfare. If you have questions or concerns about your rights as a research subject you may contact, anonymously if you wish, the Institutional Review Board at 919-966-3113 or by email to IRB_subjects@unc.edu.

By clicking “I agree,” and participating in this survey, you imply that you consent to the information presented in this form. In addition, you confirm that you associated with a digital collection at UNC-Chapel Hill.
___ I agree          ___ I do not agree

1. What types of materials are in your collection: (For example, artifacts, manuscripts, photographs, maps, etc.)

2. Approximate number of digital objects in your collection:

3. Is metadata currently created for objects in your digital collection?
   - Yes
   - No

4. Is metadata creation required?
   - Yes
   - No

5. Is collection metadata available to the public?
   - Yes
   - No

6. If metadata is currently created, which of the following metadata schemas are used? Please select all that apply.
   - Dublin Core (DC)
   - EAD (Encoded Archival Description)
   - TEI (Text Encoding Initiative)
   - GILS (Government Information Locator Service)
   - IEEE (Institute of Electrical and Electronic Engineers, Inc.) Learning Object Metadata
   - ONIX (Online Information Exchange)
   - MARC
   - METS (Metadata Encoding Transmission Standard)
   - MODS (Metadata Object Description Schema)
   - MPEG Multimedia Metadata
   - FGDC Content Standard for Digital Geospatial Metadata (CSDGM)
   - ICPSR Data Document Initiative (DDI)
   - VRA (Visual Resources Association Core Categories)
   - Categories for the Description of Works of Art
   - Other (please list)

7. Why was it decided to use these metadata schemas? Please select all that apply.
   - Required by the department or funding agency.
   - Read about it
   - Word of mouth
- Similar digital collections use it
- Best suited to the collection material
- Other (please explain):

8. Where is the metadata stored? Please select all that apply.
- Within the object
- Database
- Spreadsheet
- Other

*If database was selected:*

8a. What type of database is used for metadata storage?

9. What metadata standard(s) would those working on this collection like to see implemented? Please select all that apply.
- Dublin Core (DC)
- EAD (Encoded Archival Description)
- TEI (Text Encoding Initiative)
- GILS (Government Information Locator Service)
- IEEE (Institute of Electrical and Electronic Engineers, Inc.) Learning Object Metadata
- ONIX (Online Information Exchange)
- MARC
- METS (Metadata Encoding Transmission Standard)
- MODS (Metadata Object Description Schema)
- MPEG Multimedia Metadata
- FGDC Content Standard for Digital Geospatial Metadata (CSDGM)
- ICPSR Data Document Initiative (DDI)
- VRA (Visual Resources Association Core Categories)
- Categories for the Description of Works of Art
- Other (please list)

10. When creating metadata for your digital collection, what software is used? Please select all that apply.
- Acrobat
- ContentDM
- Endeca
- Notetab
- Excel
- Oxygen
- Photoshop
- Word
- XMetal
- XSpy
- Other (please explain):
11. How is metadata created?
   - Automatically
   - Automatically, but with human editing
   - Created Manually
   - Other (please explain):

12. When is metadata created? Please select all that apply.
   - During the content creation process
   - After creation but before publishing online
   - During the online publishing process
   - After online publication
   - Other (please explain):

13. What is the primary function of the metadata created for this collection?
   - Interoperability
   - Resource discovery
   - Preservation
   - In-house use
   - Other (please explain):

14. Who makes metadata decisions for your digital collection?

15. Does the collection link to another digital collection at UNC? If so, which one(s)?
   - Yes (List):
   - No

16. Does the collection link to other digital collections outside of UNC? If so, which one(s)?
   - Yes (List):
   - No

17. What would you like to see happen with metadata creation within your digital collection?

18. Is assigning metadata part of your job description?
   - Yes
   - No

19. Name of digital collection:

20. Affiliation of digital collection:

21. Is the collection part of a larger collection? (Example: The Oral Histories of the American South collection is part of Documenting the American South)
   - Yes (List):
22. Date of creation:

23. Type of funding:
(Examples: grant, library funding, private funding, etc.)

24. Project manager:

25. The creator of this survey would like to conduct follow-up surveys with interested recipients. Are you available for a follow-up interview? Please select:

- Yes
- No

*If yes:*

26. If you selected yes, please list your contact information below:

*Must select two of three*

- Name
- Phone Number
- Email Address
Appendix C: Recruitment Email
February 20, 2008

Dear Librarian:

My name is Jennifer Joyner, and I am a graduate student in the School of Information and Library Science at the University of North Carolina at Chapel Hill. For my master’s paper, I am conducting a study on metadata creation within digital collections at UNC-Chapel Hill. You were selected for this study because you are a librarian at UNC-Chapel Hill who is associated with a digital collection. A total of 30 librarians from UNC-Chapel Hill have been chosen to participate in this study. Your participation is completely voluntary.

To participate in the study you would complete an online survey, which will be emailed to you on February 25, 2008. Completing the survey connotes your consent to be a participant in this study. This questionnaire is composed of questions addressing the metadata creation and use at your digital collection. The survey consists of three sections and includes 25 questions. Completion of the survey should take no longer than 20 minutes. All questions in section 3 are optional. At the conclusion of this survey, you will have the option to participate in a one-hour follow-up interview in March 2008. The follow-up interview is completely voluntary.

I will seek to protect the anonymity of each participant. You will be asked to put identifying information on the survey, but you will not be required to answer these questions. In all reporting, no participant will be identified by name. Each participant will be assigned a pseudonym. I will create two separate reports from this survey. The first will include an internal report that will be shared with the Metadata Caucus, which is a working group of librarians at UNC-Chapel Hill. All internal reporting will include the collection name but exclude all identifying information of the participant. The second report will be made available for publishing. For external reporting and publishing, all collections will be assigned a pseudonym. The only persons who will have access to these data are the investigator named on this letter and the academic advisors assisting the investigator. Please note that including identifying information in survey answers is completely optional.

You will receive the survey in five days. Because I want to encourage the participation of as many librarians as possible, I will send you a reminder email approximately 7 days after you receive the survey. All survey responses are due 14 days from receipt of the survey.

There are neither risks anticipated should you participate in this study nor any anticipated benefits from being involved with it. There is no cost to you or financial benefit for your participation.
You may contact me with any questions at (919) 967-3971 or by email (jdjoyner@email.unc.edu).

All research on human volunteers is reviewed by a committee that works to protect your rights and welfare. If you have questions or concerns about your rights as a research subject you may contact, anonymously if you wish, the Institutional Review Board at 919-966-3113 or by email to IRB_subjects@unc.edu.

Thank you for considering participation in this study.

Sincerely,

Jennifer Joyner
MSLS Candidate
School of Information and Library Science
Appendix D: Survey Email
February 25, 2008

Dear Librarian:

My name is Jennifer Joyner, and I am a graduate student in the School of Information and Library Science at the University of North Carolina at Chapel Hill. For my master’s paper, I am conducting a study on metadata creation within digital collections at UNC-Chapel Hill. You were selected for this study because you are a librarian at UNC-Chapel Hill who is associated with a digital collection. A total of 30 librarians from UNC-Chapel Hill have been chosen to participate in this study. Your participation is completely voluntary.

To participate in the study you would complete an online survey. The URL to the survey is located at the end of this message. Completing the survey connotes your consent to be a participant in this study. This questionnaire is composed of questions addressing the metadata creation and use at your digital collection. The survey consists of three sections and includes 25 questions. Completion of the survey should take no longer than 20 minutes. All questions are optional. At the conclusion of this survey, you will have the option to participate in a one-hour follow-up interview in March 2008. The follow-up interview is completely voluntary.

I will seek to protect the anonymity of each participant. You will be asked to put identifying information on the survey, but you will not be required to answer these questions. In all reporting, no participant will be identified by name. Each participant will be assigned a pseudonym. I will create two separate reports from this survey. The first will include an internal report that will be shared with the Metadata Caucus, which is a working group of librarians at UNC-Chapel Hill. All internal reporting will include the collection name but exclude all identifying information of the participant. The second report will be made available for publishing. For external reporting and publishing, all collections will be assigned a pseudonym. The only persons who will have access to these data are the investigator named on this letter and the academic advisors assisting the investigator. Please note that including identifying information in survey answers is completely optional.

Because I want to encourage the participation of as many librarians as possible, I will send you a reminder email 7 days from today. All survey responses are due in 14 days.

There are neither risks anticipated should you participate in this study nor any anticipated benefits from being involved with it. There is no cost to you or financial benefit for your participation.

You may contact me with any questions at (919) 967-3971 or by email (jdjoyner@email.unc.edu).
All research on human volunteers is reviewed by a committee that works to protect your rights and welfare. If you have questions or concerns about your rights as a research subject you may contact, anonymously if you wish, the Institutional Review Board at 919-966-3113 or by email to IRB_subjects@unc.edu.

Thank you for considering participation in this study.

Survey URL:
http://uncodum.qualtrics.com/SE?SID=SV_71d9l7RGfwOKAqE&SVID=Prod

Sincerely,

Jennifer Joyner
MSLS Candidate
School of Information and Library Science
Appendix E: Reminder Email

March 3, 2008

Dear Librarian:

My name is Jennifer Joyner, and I am a graduate student in the School of Information and Library Science at the University of North Carolina at Chapel Hill. For my master’s paper, I am conducting a study on metadata creation within digital collections at UNC-Chapel Hill. You were selected for this study because you are a librarian at UNC-Chapel Hill who is associated with a digital collection. You should have received the survey associated with this study one week ago. This is an email reminder that all survey responses are due in 7 days.

You may contact me with any questions at (919) 967-3971 or by email (jdjoyner@email.unc.edu).

All research on human volunteers is reviewed by a committee that works to protect your rights and welfare. If you have questions or concerns about your rights as a research subject you may contact, anonymously if you wish, the Institutional Review Board at 919-966-3113 or by email to IRB_subjects@unc.edu.

Thank you for considering participation in this study.

Sincerely,

Jennifer Joyner
MSLS Candidate
School of Information and Library Science
Appendix F: Interview Recruitment Email

Date

Dear Librarian:

My name is Jennifer Joyner, and I am a graduate student in the School of Information and Library Science at the University of North Carolina at Chapel Hill. For my master’s paper, I am conducting a study on metadata creation within digital collections at UNC-Chapel Hill. On your completed survey, you indicated that you are interested in participating in a follow-up interview. Please respond to this email to verify your interest. I will contact you within three days of receipt of your response to schedule a time and location for the interview.

This interview will take place in your work environment. If your work environment does not provide a closed-door, private environment where this interview can take place, please make a note of this in your response email. An alternate location will be provided.

You may contact me with any questions at (919) 967-3971 or by email (jdjoyner@email.unc.edu).

All research on human volunteers is reviewed by a committee that works to protect your rights and welfare. If you have questions or concerns about your rights as a research subject you may contact, anonymously if you wish, the Institutional Review Board at 919-966-3113 or by email to IRB_subjects@unc.edu.

Thank you for considering participation in this study.

Sincerely,

Jennifer Joyner
MSLS Candidate
School of Information and Library Science
Appendix G: Signed Consent Form for Interviews

University of North Carolina-Chapel Hill
Consent to Participate in a Research Study
Adult Participants
Social Behavioral Form

________________________________________________________________________

IRB Study # 08-0147
Consent Form Version Date: February 5, 2008

Title of Study: Documenting and Describing Digital Collections: Surveying Metadata Standards Used by the University of North Carolina at Chapel Hill Library System

Principal Investigator: Jennifer Joyner
UNC-Chapel Hill Department: School of Information and Library Science
UNC-Chapel Hill Phone number:  
Email Address: jdjoyner@email.unc.edu
Faculty Advisor: Dr. Jane Greenberg
Funding Source: None

Study Contact telephone number: 919-967-3971
Study Contact email: jdjoyner@email.unc.edu

What are some general things you should know about research studies?
You are being asked to take part in a research study. To join the study is voluntary. You may refuse to join, or you may withdraw your consent to be in the study, for any reason, without penalty.

Research studies are designed to obtain new knowledge. This new information may help people in the future. You may not receive any direct benefit from being in the research study. There also may be risks to being in research studies.

Details about this study are discussed below. It is important that you understand this information so that you can make an informed choice about being in this research study. You will be given a copy of this consent form. You should ask the researchers named above, or staff members who may assist them, any questions you have about this study at any time.

What is the purpose of this study?
The purpose of this research study is to learn about the state of metadata creation within digital collections associated with the UNC-CH library system. This research will gauge the current state of metadata creation and predict future trends. This information will help metadata librarians at UNC-Chapel Hill make suggestions for future changes regarding
metadata creation.

You are being asked to be in the study because you have been identified as an UNC-Chapel Hill librarian who works directly with digital collections.

**How many people will take part in this study?**
If you decide to be in this study, you will be one of approximately 30 people in this research study.

**How long will your part in this study last?**
The interview will last approximately one hour. There will be no follow-up after this interview.

**What will happen if you take part in the study?**
If you choose to take part in this study, you will be asked a series of ten questions. Questions will relate directly to your previous responses in the online survey you took in February 2008. This interview should take no longer than one hour. All questions are optional, and you may choose not to answer any questions you are asked. Collection names will be retained in internal reporting; however, participant names will be de-identified.

You will be audio taped during this interview. If at any time you would like to stop taping, please notify the interviewer.

**What are the possible benefits from being in this study?**
Research is designed to benefit society by gaining new knowledge. You may not benefit personally from being in this research study.

**What are the possible risks or discomforts involved from being in this study?**
There may be uncommon or previously unknown risks. You should report any problems to the researcher.

**How will your privacy be protected?**
Because of the small size of library digital collection staffs at UNC-Chapel Hill, it is possible that someone could deduce the identity of the participant, even though pseudonyms will be used in place of individuals’ names. To reduce this risk, all data, both internal to the library and that which may be made public during publishing, will contain only pseudonyms for the participants’ names. Outside of library use, all collection names will also be de-identified. Participants will be labeled as Participant 1, Participant 2, etc. Collections will be labeled collection A, Collection B, etc.

The Principal Investigator will create two separate reports. The first will include an internal report that will be shared with the Metadata Caucus, which is a working group of
librarians at UNC-Chapel Hill. The second report will be made available for publishing outside of UNC-Chapel Hill. Before any report or publishing of data, participants will have an opportunity to review the report. Requests for deletion of identifying information will be honored.

All data collected will be stored on the Principal Investigator’s personal laptop, which is password protected. All data transmission will be performed on a secure email server. Only the Principal Investigator and research advisors will have access to identifiable data.

Audiotapes from this interview session will be stored in a locked filing cabinet at the residence of the Principal Investigator. The interviews will not be transcribed verbatim. All research notes taken from these tapes will be stored separately from the tapes. Any written notes from this interview session will be transferred to the password-protected laptop and the written notes destroyed. Tapes will be kept for three years. They will be stored separate from any other data that could reveal the identity of the interview participant.

Participants will not be identified in any report or publication about this study. Although every effort will be made to keep research records private, there may be times when federal or state law requires the disclosure of such records, including personal information. This is very unlikely, but if disclosure is ever required, UNC-Chapel Hill will take steps allowable by law to protect the privacy of personal information. In some cases, your information in this research study could be reviewed by representatives of the University, research sponsors, or government agencies for purposes such as quality control or safety.

Check the line that best matches your choice:

_____ OK to record me during the study
_____ Not OK to record me during the study

**Will you receive anything for being in this study?**
You will not receive anything for taking part in this study.

**Will it cost you anything to be in this study?**
There will be no costs for being in the study

**What if you are a UNC employee?**
Taking part in this research is not a part of your University duties, and refusing will not affect your job. You will not be offered or receive any special job-related consideration if you take part in this research.

**What if you have questions about this study?**
You have the right to ask, and have answered, any questions you may have about this research. If you have questions, or concerns, you should contact the researchers listed on the first page of this form.

**What if you have questions about your rights as a research participant?**
All research on human volunteers is reviewed by a committee that works to protect your rights and welfare. If you have questions or concerns about your rights as a research subject you may contact, anonymously if you wish, the Institutional Review Board at 919-966-3113 or by email to IRB_subjects@unc.edu.

---

**Title of Study:** Documenting and Describing Digital Collections: Surveying Metadata Standards Used by the University of North Carolina at Chapel Hill Library System

**Principal Investigator:** Jennifer Joyner

**Participant’s Agreement:**
I have read the information provided above. I have asked all the questions I have at this time. I voluntarily agree to participate in this research study.

Signature of Research Participant  
Date

Printed Name of Research Participant

Signature of Person Obtaining Consent  
Date

Printed Name of Person Obtaining Consent
Appendix H: Interview Questions

Interviews will be based on survey responses. The following questions will be used as a guide for the interview.

1. You have indicated that <collection name here> collects <materials here>. What challenges does your collection face when assigning metadata to these materials?

2. You have indicated that your collection houses <number of digital objects here>. Do you anticipate a large growth for this collection in the next five years? If so, how will metadata help or hinder this growth?

3. You indicated that metadata creation <was or was not> required for your collection. Do you see this policy changing over the next five years? Has it changed recently?

4. You indicated that the following metadata schemas are used <list standards> and they were selected because <list reasons here>. How have these schemas affected metadata creation and use?

5. You indicated that the metadata is stored <list location here>. How has this process worked for your collection? Does it adequately meet the needs of the collection?

6. You indicated that those working on the collection would like to see the following metadata standards implemented <list standards>. What about these standards is appealing and makes them suitable for your digital collection?

7. Do you see the software used as having an impact on the metadata you create?

8. You indicated that the primary function for metadata creation was <answer here>. Is the metadata fulfilling its function?

9. You indicated that you would like to see <answer> happen with metadata creation within your digital collection. Can you expand on this?

10. You indicated that your collection <does/does not> link with other metadata collections. What do you see as the future for links between your collection and other collections? How does the metadata affect the ability of your collection to link with other collections?