The Evolution of a Documents Digitization Project
within the American Library Association's New Members Round Table

Anna M. Ferris
University of Colorado at Boulder

William W. Armstrong
Louisiana State University

ABSTRACT

This article describes an exploratory exercise in knowledge management and database development undertaken by a task force of volunteer members of the New Members Round Table (NMRT) of the American Library Association. The authors will show how the group defined its mission, established priorities, drew up procedural guidelines, and developed a searchable database to serve the organizational needs of their fellow NMRT members while making use of the basic skills and equipment available to the group at that time. The steps in this exercise are replicable, provided that volunteers can be identified within an organization who are willing to commit their time and resources to a project as challenging as this one proved to be.

INTRODUCTION

In January 2001, the Secretary of the New Members Round Table¹(NMRT) of the American Library Association (ALA) proposed several new projects for NMRT members to explore. One of these projects called for volunteers to start up the Documents Digitization Task Force (DDTF) whose charge would be to digitize past NMRT reports (that were stored at the ALA Archives at the University of Illinois in Urbana-Champaign) and to make them accessible to NMRT members via the Web.

In this article, the authors will focus on the work of the DDTF, primarily on the practical aspects of converting paper documents to an electronic format (i.e., scanning, editing, reformatting) and the prerequisites of developing a Web-enabled archival database in order to bring these documents within reach of every NMRT member. They will discuss the specific details that led to the successful completion of this project as well as the wider issues surrounding it, such as knowledge management and organizational memory. It is the authors' hope that, by offering a practical example of how this group was able to provide links to NMRT's organizational memory, other groups will explore similar ways of accessing the hidden resources stored within their own organizations.

At around the same time as the DDTF was being organized, the NMRT Archives Committee was given the charge of updating the *NMRT Archives Policy* that dealt with the preservation of materials for the NMRT archives.² Previously, the Archives Committee's main function had involved the harvesting of paper reports and documents from committee chairs and officers and forwarding them in file boxes to the ALA Archives. But with a notable increase in the number of electronic documents being produced within NMRT, the Archives Committee was asked to draft a new policy³ that would address the preservation of all materials in a variety of electronic formats (not just the newer, digitally-born documents, but the older, archived materials as well). It was inevitable that the Archives Committee and the DDTF should work together on this pressing issue since the tasks of both committees were highly interrelated and close cooperation and communication were required. Specifically, it was up to the Archives Committee to deal with the theoretical aspects of managing electronic documents by setting policies, assessing preservation requirements, and exploring technological developments; it was up to the DDTF to concentrate on the more practical aspects of digitization as described above. The culmination of the work of

these two committees would result in a successful records management system that, today, continues to grow and evolve as it begins to meet the needs of the entire NMRT organization.

ABOUT NMRT

NMRT is a membership unit of ALA whose main purpose is to provide a common forum where new members (librarians, students, paraprofessionals, etc.) can discuss ideas of mutual interest and where potential leaders in the profession can develop. NMRT is also a creative and energetic organization where opportunities abound for new members to team up with more experienced members on innovative projects that are relevant to the library profession in general. NMRT members come from diverse backgrounds and are affiliated with different types of libraries, from public to academic to special libraries. At the time of this writing, there are 1,374 members of NMRT.

The NMRT Executive Board is made up of ten officers. One of the responsibilities of being an officer is the supervision of various committees within NMRT. Currently, there are 32 committees, which include seven ad hoc groups. Given the size and constitution of this organization, one would be right to suppose that a considerable amount of information is documented by NMRT members during a one-year period.

There are two official conferences held each year for which NMRT members must prepare—the Midwinter and Annual ALA conferences, usually held in January and June, respectively. Most of these preparations encompass the following types of activities: 1) the governance of the organization itself as manifested in the development of administrative policies, the maintenance of the official handbook, and the supervision of the various committees; 2) organizing special events at each year's annual meeting which involves nearly all of the committees and officers; and 3) the creation and implementation of various programs throughout the year. In addition, the NMRT Executive Board may convene electronic board meetings, pre-arranged or impromptu, via email whenever a pressing issue requires discussion or resolution. On an individual level, each NMRT officer and committee chair is responsible for maintaining current records of official activities. This entails submitting three reports (planning, progress, and final) to the executive board during the course of the year they are serving as officer or chair.

The following list shows the estimated number of documents generated within NMRT during a one-year period (Figure 1). These documents—chronicles of the discussions and

accomplishments of many participants in a wide variety of activities—represent a type of legacy that has been consigned by conscientious members of NMRT to their future colleagues and successors ever since the organization was established in 1931 as the Junior Members Round Table (JMRT).

7	Executive Board Minutes (2 MW board meetings, 3 Annual board meetings, 2 Eboard meetings [at least])
30	Planning, Progress, & Final Reports (submitted by NMRT officers)
96	Planning, Progress, & Final Reports (submitted by NMRT committee chairs)
4	Miscellaneous policies, procedural guidelines, forms or templates
137	Number of documents per year (estimated to within 3-4)

(Figure 1) Estimated NMRT documents generated during a one-year period

NMRT is very much like other professional organizations in the following ways: 1) the activities of its members generate large numbers of documents that record not only the decisions that have been made but the processes by which these decisions were formed; 2) its success can be measured by its ability to make and implement good decisions as well as by its production of accurate documents; 3) its success is determined by how efficiently these documents—tangible evidence of an organizational memory—can be shared and accessed (i.e., identified, retrieved, and applied) by the organization’s members in their attempts to continue making the best decisions possible.

WHY DIGITIZE?

While there are many similarities between NMRT and other professional organizations, it is important to point out the factors that may distinguish NMRT from other groups initiating digitization projects. Some of the specific reasons NMRT chose to explore a document digitization project in the first place are 1) there was a pressing need to provide access to NMRT’s historic documents; 2) the exercise was well-suited to the expertise of librarians; 3) the project served the mission of NMRT; and 4) technology was an effective way to offset the disadvantages caused by being, in some respects, a “virtual organization,” as NMRT is, par excellence.

To Provide Access

Developments in technology have had a tremendous impact on how organizational memory is being created, delivered, stored, and shared. But even as technology has brought us welcomed advancements in the way information is organized and used, it has also managed to underscore a compelling fact. There are still vast amounts of valuable historical documents in paper format

stored away in remote storage cabinets that can be retrieved only by a select few within an organization such as NMRT. Converting these documents to a digital format will make them readily accessible to all members of the organization, especially the officers, chairs, and committee members who will benefit from them the most. This is by far the most important reason NMRT established the DDTF.

To Utilize the Expertise of Librarians

NMRT members are uniquely suited to participate in a documents digitization project because, as librarians, they already have an informed perspective on how to deal with issues of knowledge management.⁴ Daniel D. Stuhlman (2001) provides a succinct explanation as to why this is so:

Librarians by education, training, experience and personality are highly qualified to help ... organizations manage knowledge. . . . Knowledge management requires the wisdom of experience, the organization ability of a professional cataloger, the insight of a psychologist, the interview ability of a journalist, and organizational ability of a professional manager.

To Contribute to the Mission of NMRT

By proposing a project as intricate and challenging as the documents digitization project, NMRT officers were offering their members an excellent opportunity to make a contribution to the whole organization. This was also an ideal way to help further the goals and objectives of NMRT's mission which, in part, is "To help those who have been association members less than ten years become actively involved ... in the profession ... [by providing] formal opportunities for involvement and/or training for professional ... experiences."⁵

Also, NMRT members were an ideal group to recruit for such a project. As dedicated professionals, many share an enthusiasm and a willingness to learn new tasks that make volunteering their time to NMRT a worthwhile experience. This was an important factor in the ultimate success of the project to preserve NMRT's organizational memory.

To Offset the Disadvantages of Being a "Virtual Organization"

NMRT can be considered a "virtual organization," which, in short, means that it is 1) a network of individuals connected by a common purpose, existing without any physical offices or headquarters and 2) a constantly changing group of people, made up of small groups with short-term goals and time limitations (Schwartz, Divitini, Brasethrik, 2000). Among the obvious drawbacks of being a "virtual organization" is the sense of discontinuity that arises from the

frequent changes in administrative personnel and the membership in general, all of whom reside in different places across the United States. What is more, since membership within NMRT is limited to a ten-year period, the very persons who are knowledgeable enough to help fill in the organizational “memory gaps” become ineligible to actively participate in committee work or discussions. Another drawback is the inefficiency that results from having limited access to useful information and records that are widely scattered across the country, making them exceedingly difficult to locate and retrieve.

PHASE I: PROJECT PRELIMINARIES

The Initial Charge

As stated in the NMRT secretary’s “Call for Participation,” posted on the NMRT list in January 2001, the provisional charge for a potential digitization project was:

DIGITIZE RECENT COMMITTEE REPORTS⁶ . . . [W]e should try to post the past 3-5 years of committee and officers reports on the Web for easy access: a.) Gather reports; b.) Scan and edit the files for accuracy; c.) Recommend/implement printing and access enhancements including the use of metatags or any relevant technology in compliance with NMRT Web publishing and ALA policies.

At this early stage of the process, the type of documents, the limited range (three-five years’ worth), and the procedures to be followed were merely offered as suggestions since the exact details of what was to be digitized and the steps required to do this would have to be left up to the discretion and experience of the people who volunteered. A six-month deadline was also proposed for the project but could not be enforced until more specific tasks and priorities were identified. Needless to say, the scope of this initial charge would be refined once more details (i.e., the answers to who, what, when, and why) began to emerge.

Refining the Scope

An obvious first step in the process of digitization was to identify the documents to be scanned. After considering the NMRT secretary’s suggestion to use committee and officer reports for the project, it became apparent that locating and gathering three to five years’ worth of reports would be a difficult task and the number of reports would be prohibitive for an exploratory exercise in digitization. As seen in the list of documents generated by NMRT (Figure 1), each of the 32 committees produces three reports annually (planning, progress, and final

reports) as does each of the ten officers. For a five-year period, this amounted to over 630 reports. Assuming that each report averaged at least three pages (or more) in length, the number of pages to be digitized would come close to 2,000.

On the other hand, based on what was available in the NMRT Archives, the minutes to executive board meetings held at the ALA Midwinter and Annual conferences (two to three meetings at midwinter and three meetings at annual) were easier to locate and represented a more manageable number: 25 reports (three pages each) over a five-year period came to approximately 375 pages. The amount was found to be manageable enough for the scope of the project to be increased to include all of the executive board minutes dating back as far as 1978 and up to 1998—a 20-year span that represented approximately 100 reports, or 300 pages. (Post-1998 reports were already available in electronic format via ALA's Web site and would be dealt with later.) With the range and type of documents identified, it was necessary to proceed to the next step in the process.

Identifying Volunteers

Ten people answered the call for volunteers posted on the NMRT list. All ten were encouraged to introduce themselves to the others via email and to share information about their background in librarianship. Their knowledge of digitization (if any), their interest or expectations for the project, their access to scanning equipment, and any other issues they could think of that might be relevant to the project were also discussed. As it turned out, only four volunteers were eventually able to commit their time to join the DDTF. An interesting combination of new and experienced librarians came together from a variety of backgrounds: a chemistry librarian with database and digitization experience, a catalog librarian with no digitization experience (these first two were to become the task force co-chairs), a reference librarian with digitization experience, and a library school student with some scanning experience. Each of the four volunteers had access to scanners and OCR software.

Procuring the Documents

With the scope of the project resolved and the task force members identified, the next step was to set about procuring copies of the executive board minutes themselves. This involved enlisting the staff at the ALA Archives at the University of Illinois at Urbana-Champaign to have them pull the files out of storage. Fortunately, a rudimentary finding aid was available that provided the location of the specific boxes and folders containing minutes from each conference

during a particular year. The finding aids in Figures 2 and 3 were supplied by the ALA Archives. An Excel checklist was created by the DDTF co-chairs to help keep track of the files by year, conference (summer/midwinter), and meeting type (Board/orientation/membership) (Figure 4). While the checklist was a particularly useful tool for the DDTF, it would also help the ALA Archives staff as they retrieved and photocopied the files that would subsequently be forwarded to the DDTF in care of the NMRT secretary.

Secretary's File, 1978-1995

Record Series Number: 46/1/7
Volume: 1 cubic feet
Arrangement: Chronological

Description: Secretary's File includes correspondence (1979-85, 1993-95), and agendas and minutes for Executive Committee meetings held at the Midwinter and Annual Conferences of ALA (1978-95). The series also includes semi-annual reports prepared by NMRT officers, director's, committee chairs, and liaisons (1978-95).

[A finding aid \(box and folder listing\) or related website is available here.](#)

(Figure 2) Finding aid for NMRT Secretary's File

46/1/7
 New Members Round Table
 Executive Committee
 Secretary's File, 1978-93

Box 1:

Correspondence
 1979-85
 1993-94
 1994-95

Executive Board-Minutes & Reports
 Midwinter Conference, 1978
 Annual Conference, 1978
 Midwinter Conference, 1979
 Annual Conference, 1979

(Figure 3) Excerpted folder listing from Secretary's File # 46/1/7

	A	B	C	D	E	F
1	Document Name	Possible Location	Box	Status		Comments
2		Folder		Found	Copied	
3	1981 Annual Board Minutes I			✓	✓	
4	1981 Annual Board Minutes II			✓	✓	
5	1981 Board Orientation Minutes	June 30, 1981 (11:30-12:30 pm)		✓	✓	
6	1981 Annual Membership Meeting	June 30, 1981		✓	✓	
7						
8						
9	1980 Midwinter Board Minutes I			✓	✓	Needs typing
10	1980 Midwinter Board Minutes II			✓	✓	"
11	1980 Midwinter Board Minutes III			✓	✓	"

(Figure 4) Excel checklist

Standardizing the Format (Function vs. Form)

Upon closer inspection of some of the older sets of minutes, it was found that many of the photocopies received from the ALA Archives were unscannable, primarily because these earlier minutes had originally been printed on dot-matrix printers that produced very poor quality originals. A decision had to be made regarding the best way to handle these.

The poor quality photocopies brought into focus one of the biggest challenges the DDTF had to face at the outset—how to resolve the question of *function* vs. *form*. Was its mission to reproduce faithfully the exact image (*form*) of the original document being converted? Or was the mission to communicate the content (*function*) of the original document? After consulting with the executive board, the DDTF concluded that *function* was the principal goal of the project (i.e., preserving the textual content was more important than preserving the image of the text). This decision was to have immediate and practical ramifications. Capturing the textual content of each set of minutes would require the DDTF to divide up its workload between two separate tasks: the scanning of the better quality minutes using Optical Character Recognition (OCR)⁷ software and the keying in of the poorer quality minutes by hand. The existence of these unscannable minutes opened up an opportunity for the DDTF to involve more volunteers in the digitization project. Other members without scanners could now be recruited to help re-key the poorer quality copies using whatever word processing programs they had at their disposal.

Establishing Hardware/Software Requirements

Hardware/software requirements grew logically out of the DDTF's decision to make the textual content, not the image of the text, their primary concern. The further decision to make each set of minutes searchable (discussed in more detail in the section on Phase II) required the production of a pure text version of the original document and the use of OCR scanning software to capture and extract the original text in its entirety. Because the OCR software available at that time was not best suited for reading and recognizing inferior images, extensive editing was often required.

Besides a pure text version of the original document, it would also be necessary to create a version that could be easily used and read on the Web. The DDTF initially considered saving the scanned minutes as Portable Document Format (PDF)⁸ image files but ultimately decided not to for several reasons: 1) the DDTF was concerned with retaining the textual content of each

document, not the facsimile image; 2) image file conversion yields files that are larger in size, which could possibly render this format inaccessible to people with older machines and slower connections; 3) PDF image files are not searchable. HTML (Hypertext Markup Language) format was chosen as the best format to provide Web operability and readability for the following reasons: 1) ease of conversion; 2) small file sizes; 3) readability of HTML by all browsers; and most important, 4) searchability.

The DDTF advocated the use of Microsoft Word (Version 6 or higher) as the word processing program to use for saving the scanned documents in a variety of digital formats. It provided near universality and the convenience of converting documents into HTML format with a click of the mouse. The use of Microsoft Word would, however, require an additional step in the conversion process, i.e., saving all the scanned minutes files in Rich Text Format (RTF)⁹ prior to saving them as pure text documents and HTML files.

The basic hardware/software requirements that DDTF members were expected to use were 1) a flatbed scanner with OCR scanning software for the optimal capturing of the text of the minutes (optional); 2) a personal computer equipped with Microsoft Word (6.0 or higher) for optimal word processing and digital conversion capabilities (required); and 3) an electronic mail system (required) for sending the converted files as attachments to Louisiana State University (LSU) where they would be uploaded to the host server.¹⁰

As was to be expected, the type of scanning equipment used varied with each volunteer. Among the scanners used were Hewlett Packard ScanJet ADF, Minolta PS 700, and Epson TWAIN PRO. Some of the scanning programs used were OmniPage Pro (version 10.0 & 11), Adobe Acrobat 4, HP PrecisionScan Pro, Acrobat Capture, TextBridge, and Pagis Pro.

Identifying the Tasks

Once the photocopied minutes were received from the ALA Archives, the DDTF co-chairs were able to document the specific steps that task force members were to follow throughout the entire digitization process. These steps were as follows:

1. Triage

- Organize the reports into one or two-year batches;
- Photocopy each report;
- Distribute good print quality reports to members with scanners;
- Distribute poor print quality reports for keying in by members without scanners.

2. Scanning

- Capture the contents of each page using OCR image editing software;

- Recommended settings: B&W image, 12 bit (Grayscale), 300 resolution;
- Match the scanned page to the original to make sure all text has been captured;
- Save the scanned file as an RTF file;
- Assign a filename (do not use caps or spaces).

3. Proofing & OCR Editing

- Clean up the document (removing unwanted non-textual characteristics or blemishes)
(*An example of such blemishes appears in the photocopied minutes in Appendix B*);
- Correct identifiable errors in spelling or layout;
- Use consistent formatting for each document, such as margin size, fonts, etc.

4. Exporting

- Open RTF file in Word and save as “Web page” (HTML);
- Use RTF file as source for document text to be inserted into the database record for this document;
- Send HTML file to the database administrator for uploading onto the Web server.

Establishing Format and Description Standards

While one of the most important aspects of digitizing the NMRT minutes focused on preserving the textual content of each document (as opposed to the exact image of the text), it was still important to maintain consistency in the way the minutes would be viewed by NMRT members at the end of an online search. Wherever possible, task force members converting paper documents to a digital format were asked to apply the following standards:

- Letter-size paper format
- One-inch margins all around
- Preferred font: New Times Roman
- Eliminate columns (type data in paragraph format)
- Remove paragraph indentations (justify all paragraphs to the left)

The reason for applying these rules was to develop a simple process that would result in a consistent and readable end-product every time. They were also intended to keep task force members from going to elaborate lengths to reformat each set of minutes. Beyond these simple formatting rules, the DDTF would also develop more detailed procedural guidelines (discussed in the section on Phase III) to help illustrate the step-by-step process involved in creating a unique metadata record for each digitized document.

PHASE II: DATABASE CONSTRUCTION

Concurrent with the preliminary arrangements shown in Phase I, the DDTF was proceeding with the second part of its charge, the construction of a searchable archival database. This was accomplished through close communication with the Archives Committee, which was

establishing standards based on the kind of access the DDTF would be providing. The decisions being made by the DDTF in this second phase would also help inform some of the final steps in the scanning procedures, such as filename assignment and the formulation of URLs for file transmission.

In short, the object of Phase II was to find an efficient means of making the NMRT minutes available via the Web. It was agreed that, when dealing with a large number of documents representing many different types of reports, the most useful way of providing access was through a Web-enabled database—a database that not only provided a link to the full-text document but also had the capacity to conduct a full-text word search of that document.¹¹

The decision to use an online database meant that individual metadata records would have to be created to supply the descriptive information about each digitized document. The next logical steps, then, were 1) to select a metadata standard that would be suitable to use for the new archives database, and 2) to settle on the specific criteria—the database architecture, the user interface, the controlled vocabulary—that would effectively help NMRT members retrieve the documents they were seeking.

It was during this phase of the project that the DDTF began to appreciate the long-range implications and possibilities of the work ahead of them. Though it was constructing a database to accommodate one type of document (i.e., scanned minutes), the DDTF saw this as an opportunity to create a multipurpose Web application, flexible enough to include virtually all NMRT records. This realization helped provide a focus for the Archives Committee as they worked on the new *NMRT Archives Policy for Electronic Documents*, and, at the same time, provided a basic framework for the decisions the DDTF would need to make.

Metadata Issues

Based on the Archives Committee's research findings and deliberations on metadata standards,¹² the DDTF chose to use the Dublin Core (DC) metadata schema for the following reasons: 1) DC is a universally recognized metadata standard; 2) the Dublin Core Metadata Initiative's (DCMI) set of fifteen elements is approved by the American National Standards Institute (ANSI); 3) the elements are easily understood and applied; 4) the elements cover the most essential information needed to describe a digital item; and 5) the list of elements can be tailored to specific local needs—an especially appealing feature for the DDTF.

The DDTF chose to make use of 18 elements (three locally assigned) that would enable

NMRT members to search by a variety of categories.¹³ The 15 DCMI elements adapted by the DDTF were *Title*, *Creator* (NMRT label: *Author/Creator*), *Subject* (NMRT label: *Keywords*), *Description*, *Publisher*, *Contributor* (NMRT label: *Other Contributor*), *Date*, *Resource Type*, *Format*, *Resource Identifier*, *Language*, *Source*, *Relation*, *Coverage*, and *Rights Management*. The three elements created for NMRT's local needs were *Status* (for designating whether a document was a "draft" or "final" copy), *Addressee Name* (used to identify the person(s) to whom an email document was addressed), and *URL* (a replication of *Resource Identifier* intended to facilitate the data entry process). An additional cut and paste feature was created to provide an insertion point for entering the pure text version of the document into the database record for indexing purposes.

For the archives database, one of the most important characteristics of the Dublin Core metadata schema was its maximum flexibility.¹⁴ Certain elements, mentioned above, could be designated as optional elements in order to accommodate a variety of document types. For example, the *Relation* element plays an essential role in showing a corresponding relationship (usually represented as a URL) between two documents (say, when a policy statement refers to a supplemental report or has an addendum connected to it). In the case of the NMRT minutes, the *Relation* element could be left blank because there were no corresponding documents involved. This flexibility was one of the reasons the archives database developed into a successful all-inclusive records management system that is meeting the needs of the entire NMRT organization.

Database Criteria

Database Application

Any number of relational database applications could have been used for this project, but *DBMan SQL*¹⁵ was chosen for the following reasons: 1) it provided an easy means for creating the underlying database; 2) it provided easily customizable Perl¹⁶ scripts for creating the necessary Web front end for the administration panel and an end-user interface; 3) it is compatible with Apache running on a Unix platform; and 4) the database developers, working out of LSU, already had access to the program.

Two User Interfaces

The Web site architecture for the archives database was divided into two distinct user interfaces: 1) the administrative panel, which was to be used exclusively for creating metadata

records and inputting data; and 2) the patron-search panel, which was to be used by NMRT members and the general public for basic searching purposes. Each panel consisted of an HTML form with fields corresponding to the components by which each document was indexed.

1) The Administrative Panel: This interface provides access to the database templates for creating and maintaining the metadata records. The Main Menu screen (Figure 5) is the first step to accessing the templates. (Access is permitted only via an authorized user I.D. and password.)

The screenshot shows a web interface titled "Database Manager: Main Menu". Below the title bar, the text "Main Menu" is centered. Underneath, it says "Permissions: View Add Delete Modify Admin". At the bottom, there is a navigation bar with links: | Home | Add | View | Delete | Modify | List All | Admin | Log Off |.

(Figure 5) Main Menu to administrative panel templates

A select group of functions is displayed in a navigation bar at the bottom of the Main Menu (e.g., *Add*, *View*, *Delete*, *Modify*, among other management functions). This design feature provides consistent and easy-to-use navigation throughout all aspects of the administrative interface (see Appendix A to view the *Add a New Record* template).

2) The Public Panel: This interface provides two search options, a *Basic Search*¹⁷ screen (Figure 6) and an *Advanced Search*¹⁸ screen. The DDTF chose to create two patron-search interfaces so that members could have the option of searching from a simple keyword search screen (*Basic*) or from an array of separately-indexed fields (*Advanced*).

The screenshot shows a web interface titled "Search the NMRT Archives Database". Below the title, there are links for "HELP TIPS", "Basic Search", and "HELP TIPS". The main search area has a text input field, a "Keyword Search" button, and a "View Records" button. Below this, there is a "Search Settings:" section. It states: "Boolean Operator is 'OR'." and "Truncation is automatically employed (e.g. 'minor' retrieves 'minority OR 'minorities')." There is a text input field with "25" and the label "Max. Returned Hits". Below that, there is a "Sort By:" dropdown menu and a "Sort Order:" dropdown menu with "Ascending" selected. At the bottom, there is a link "Browse Records: LIST ALL".

(Figure 6) Basic Search interface

To avoid any confusion that might arise from not knowing how to search a particular field, a detailed *Help* page was created to show definitions of each field and to spell out the appropriate use of each field within the scope of the archives database. Tips were also provided to help in the search process (e.g., offering suggestions on boolean operators and truncation/phrase searching). Both the *Basic Search* and *Advanced Search* screens have a separate Help page.

Controlled Vocabulary

In formulating the criteria for a user interface, it was important for the DDTF to develop a controlled vocabulary of locally-assigned terms to ensure consistency in the way task force members entered the data, thereby enhancing the search and retrieval process for everyone. The following areas were the particular places where a controlled vocabulary approach was required: 1) the assigning of filenames for the converted documents; 2) *Resource Type* designations; 3) *Format* designations; and 4) specific descriptions supplied in the guidelines for data entry.

1) Filenames: Each time a document was converted, the resulting file (saved in either Word, RTF, or HTML format) was assigned a name that would be mnemonic and yet unique enough to identify the contents of that file. In order for the assigning of filenames to be as systematic as possible and to avoid duplication when names were being assigned by different individuals, task force members were instructed to supply a name that indicated the conference, the year, and the particular board meeting (for example: *midwinter_86-1* was the name given to the first board meeting of the 1986 Midwinter conference; similarly, *annual_93-3* referred to the third board meeting of the 1993 Annual conference). The one proviso was that neither capital letters nor blank spaces (underscores were preferred) be used in the filename.

2) Resource Type: To accommodate the various kinds of official documents that could be produced within NMRT, a pre-made pull-down list of document types was attached to the *Resource Type* element (Figure 7). Task force members could select from such document types as Constitution and Bylaws, Statement of Function, Membership List, Report, and Minutes. (The authorized list of resource types is given in *appendix C* of the *NMRT Archives Policy for Electronic Documents*.)

(Figure 7) Pull-down list of Resource Types

3) Format: Since so much depends on the nature of the documents being entered and/or searched in the archives database—that is to say, whether they are in HTML or PDF format—an authorized list of standard formats was also provided via a pull-down feature attached to the *Format* element (Figure 8). (The authorized list of standard formats is given in *appendix D* of the *NMRT Archives Policy for Electronic Documents*.)

(Figure 8) Pull-down list of Standard Formats

4) Descriptions in Guidelines for Data Entry: Finally, since there were so many people involved in entering data into the archives database, sample statements were given in the procedural guidelines for the wording to be used in the *Description* field of each set of minutes. The statements were intended to facilitate the data entry process for task force members and, more importantly, to enhance the search and retrieval process for end users. The following is a sample statement used in the *Description* field for NMRT minutes: “Minutes of the [1st, 2nd, 3rd?] [year] [MW/Annual?] meetings of the NMRT Executive Board held at the ALA in [place] recording business conducted by the board, including personnel actions, resolutions, and policy decisions.”

There was also a sample statement provided in the procedural guidelines that showed task force members how to formulate the pre-approved Uniform Resource Locator (URL) string. This special URL, consisting of the source document filename (in box) plus the server path, serves to

identify the file once it is loaded onto the Web server at Louisiana State University: “USE http://www.lib.lsu.edu/ALA/nmrt/midwinter_86-1.htm FOR URL”.

PHASE III: IMPLEMENTATION

Phase III involved a complete synthesis of the steps taken in Phase I (project preliminaries) and Phase II (database construction). The objective of this final phase of the project was to begin the actual record creation, indexing, and data entry of the newly converted NMRT minutes into the archives database. But before this could be done, it was necessary to create a complete set of standard procedures and guidelines for the inputting process of the many documents at hand. Without an official set of procedural guidelines, there would have been major problems with consistency since, with the exception of the co-chairs, none of the task force members had ever used the administrative interface of the database before and were unfamiliar with the intricacies of metadata creation and full-text indexing. For these reasons, every step in the inputting process had to be clearly explained and documented by the co-chairs. The resulting document is entitled *Procedures for Inputting Records into the Database* (2002).

These procedures were developed to provide detailed instructions on the processing of *scanned* documents (in this case, the NMRT board minutes from 1978 to 1998). Specifically, these guidelines contained information on how to log into the administrative interface; how to add a record to the database; how to describe a document concisely in the *Description* field; how to enter the data that go in the *Title*, *Author*, and *Date* fields; and how to formulate the URL to identify each document.

Additional instructions were needed to provide guidance on how to cut and paste the pure text document into the record, a procedure that was not at all intuitive. As touched upon earlier in the section on Phase I, the scanned documents had to be saved as RTF files first to facilitate the conversion into both pure text and HTML formats. While RTF files are quite functional, they still contain formatting codes that cannot be read by the database, hence the need for the pure text conversion. To rectify this, a somewhat cumbersome, yet effective, method was used that involved opening the RTF file in WordPad, saving it as “text only,” closing the file, and reopening it again in WordPad.¹⁹ (These steps were necessary to remove all of Microsoft’s extra hidden codes.) The resulting pure text file was then ready to use and could be cut/copied and pasted directly into the allotted field of the record. Once this was done and the record was added

to the database, the entire text within this field could be indexed and searched via *Keyword*.

Another important development during this phase of the project concerned the file-naming convention discussed in the section on Phase II. It soon became apparent that it would be more practical to have the co-chairs assign the filenames that were to be used for the scanned documents prior to their being distributed to the task force members who were inputting the data. This proved to be the best way to help each person work as independently as possible and, in so doing, avoid any slowdowns. Since the project co-chairs already knew what the exact path to the source document file was to be on the Web server ([http://lib.lsu.edu/ALA/nmrt/...](http://lib.lsu.edu/ALA/nmrt/)), a pre-approved file-naming convention was put in place enabling the task force members to complete each URL by simply adding the filename to the above string without any unnecessary delays. Once this was accomplished and every relevant field filled in, the record was completed by clicking the *Add Record* button at the base of the template screen. The final step to the data entry process was to forward an email attachment of the minutes, now HTML (Web-ready) file(s), to the contact person at LSU for loading onto the Web server where they would henceforth reside.

It was the full immersion of the task force members in Phase III of the project that represented the last step in completing the DDTF's mission and goals. The end result of the project was the conversion of 95 sets of NMRT minutes from print format into HTML files during the period June 2001-June 2002 and the indexing of each set of minutes (via their metadata records) in the new online archives database. The newly created database allowed users to search each record (by using basic keyword or multiple search criteria) and retrieve the full-text version of the document they were seeking. (Appendix B shows the evolution of a set of minutes from start to finish—from paper to digital.)

Recognizing the significance of what had been accomplished with the documents digitization project (at last a direct link to their organizational memory had been made), the NMRT Executive Board recommended that the work of the DDTF be extended under the aegis of the Archives Committee, the most logical committee to oversee its activities from that point. Its work successfully completed, the DDTF was merged with the NMRT Archives Committee in the summer of 2002 following the ALA Annual conference.

RECOMMENDATIONS

As discussed at the beginning of this article, the documents digitization project was initially conceived as an exploratory exercise in dealing with a number of current issues (i.e., scanning/digitization, knowledge management, and database development). That the project's completion was such a success can be directly attributed to certain factors that could be feasibly replicated with future digitization projects. Below are some recommendations for those groups interested in exploring a similar project:

- 1) Know your limitations: the size of the project will depend greatly on the number of volunteers and the type of equipment available.
- 2) The group should have (at least) two people acting as coordinators who have some specialized knowledge, one of database construction, the other of basic cataloging.
- 3) Ideally, these two coordinators should be the people responsible for making decisions, handling the delegation of tasks, and loading the converted files onto the host server.
- 4) Someone in the group should have direct access to or actual possession of the items to be digitized.
- 5) Every person involved in the project does not necessarily have to take part in every step of the process.
- 6) Create explicit guidelines on how to perform each task.
- 7) When assigning the tasks, make sure each person receives the appropriate documentation (guidelines) so s/he does not have to stop to ask unnecessary questions.

BENEFITS

Beyond the very tangible benefit of having a searchable database to access archived documents, there were other, more subtle benefits resulting from the digitization project that had just as favorable an impact on the NMRT organization as a whole:

- 1) Minimal cost: most task force members were able to make use of equipment and services (i.e., mailing facilities) they used at their workplace. The time and effort spent contributing to a national professional organization is viewed as a service requirement for many librarians by their respective libraries or institutions.
- 2) No copyright restrictions: the documents involved were the property of NMRT.
- 3) Lessons learned: everyone who took part in the project learned new tasks and concepts they might never have learned under other circumstances.

- 4) Networking promoted: the new relationships and colleagues found as a result of this project will play an important part in each task force member's career for years to come.

THE NEXT STEPS

In 2002, with a secure set of digitized board minutes accessible via the new archives database, the recently expanded Archives Committee (merged with the DDTF) felt they could focus their attention on the assortment of reports and documents being produced by the various NMRT committees. (The chart shown in Figure 1 shows that this represented approximately 130 documents on average each year.) The biggest difference, however, between these reports and the old executive board minutes was that all these new reports were born digital and thereby did not have to undergo the paper-to-digital conversion. Taking this into consideration, the Archives Committee worked with the NMRT secretary to investigate procedures by which the committee chairs and officers could submit their reports in a way that could be easily harvested, indexed, and uploaded to the server.²⁰

Another set of procedures was developed, *Procedures for Inputting Reports into the Database* (2002) which specifically shows Archives Committee members how to create metadata records for these electronic documents. These procedures for electronic documents differ from the earlier set of procedures for scanned minutes in two primary ways: 1) the conversion process has been simplified (no longer requiring the RTF-to-HTML step due to OCR software no longer being used), and 2) the descriptive content of each field was significantly changed since the data provided in reports were quite different from those found in board minutes. Once again, consistency was essential for facilitating retrieval of the documents at a future date.

As of this writing (Fall 2004), the creation and indexing of records for committee/officer reports has proceeded very smoothly. Currently there are 190 reports in the archives database, extending from August 2002-August 2004, and more continue to be added at a rate of 120+ reports per year. In all, there are 316 records in the archives database. Of these, 106 are sets of executive board minutes dating from 1978-2003. Other miscellaneous documents include seven forms or templates and five statements of function (primarily documents dealing with procedural guidelines).

Now that policies and procedures for scanned minutes have been established, as well as procedures for gathering and indexing current electronic reports and current electronic minutes as

they are submitted, the next project will most likely be a return to the digitization of older committee reports. There was general agreement at the 2004 annual conference that it would be extremely useful to explore another scanning/digitization project of the past committee reports presently in our possession. These represent approximately 321 planning, progress, and final reports in paper format dating back to 1996. These are invaluable sources of information for current (and future) committee chairs and officers who will be looking for an historical perspective on those activities that concern their respective committees and positions.

CONCLUSION

As a result of the work undertaken by the DDTF and the Archives Committee, NMRT was able to expand what began as an exploratory digitization project into a vital records management system. In so doing, it has taken control of how its own past, present, and future pronouncements and discussions are chronicled through its various policies and official reports. The archives database produced by the volunteer members of the DDTF and the Archives Committee has effectively brought NMRT's paper archives (i.e., the recorded accounts of past achievements, decisions, and policies within the organization) to the fingertips of every single member of NMRT, not just to the members of the executive board. By providing online accessibility and searchability to a disparate set of historical documents, the archives database has helped to bring NMRT's organizational memory to the fore. As a result, it is having (and will continue to have) a major effect on current (and certainly future) committee operations, especially on the decision-making abilities of officers and committee chairs.

An added benefit of this project has been the capability of using this same archives database to collect, organize, and make available the most recent documents being produced in electronic formats. The pervasiveness of digitally-born documents within NMRT presented the Archives Committee with an excellent opportunity to explore the management of electronic records from inception to preservation or deletion, whether they were considered permanent or temporary depending on the appropriate retention cycle of each document.

NMRT has recently discovered that while it is important for an organization to preserve its past, it is just as important to provide ready access to it. This was the dilemma that preoccupied NMRT (as well as any organization contemplating a digitization project) at the outset of this project: how best to accomplish the archiving and accessing of both paper and electronic records.

NMRT's decision, ultimately, was to straddle the fence. On one side, it has chosen to preserve its past documents in paper format by archiving these in the traditional manner. On the other, it has chosen to preserve the text of these same documents along with the newest electronically-formatted documents by making their content available online via the archives database. While the older documents had to be converted to a digital format, this procedure will no longer be required once the majority of these documents have been successfully converted. Today, virtually all the documents being produced by members of NMRT are being submitted in MS Word or HTML format. As a result, it has now become necessary, ironically, to convert the electronic documents into paper format—printing them so that they can be forwarded to the physical archives and stored just as they are being preserved in the electronic archives database. So, in this surprising twist, it is the paper version that has now become the copy and the electronic version that has become the original!

Given this quantum switch in the original paper-to-digital format, it will become necessary for groups handling such documents to develop their own set of procedures and guidelines for records management. Indeed, the NMRT Archives Committee has established its own policies and guidelines for handling electronic documents and for maintaining their integrity over time. It is to be expected that these policies and guidelines will continue to evolve as the need arises and as more experience is gained. But now that the process is underway for NMRT, it has proven to be an exceedingly worthwhile endeavor. The success of the documents digitization project and the archives database has shown that the creation of an electronic repository and retrieval system for an organization's historical and contemporary records is not only useful but essential to the best interests of the organization. That it should be done is an unqualified yes. How it should be done must be left up to the organization considering such a project. In this paper, we have offered a description of the process that has worked well for NMRT (i.e., a group of motivated volunteers making use of the technical resources available to most organizations). The questions remaining to be answered by any group revolve around the nature of the organization in question. What are its specific needs? Is function more important than format? It certainly was for NMRT insofar as access to the content of the documents was more important than access to the facsimile of the original paper document. In the case of electronic documents, this question may never emerge at all. But this is one example of the many questions that will need to be considered by any organization contemplating such a project. The answers will help determine the specific manner

in which that organization's collective memory will come to life through a records management program. In essence, this was the exact outcome of NMRT's documents digitization project—the restoration of organizational memory through the merging of the past, present, and future into the organization's ongoing, dynamic life.

NOTES

¹ In January 2001, the NMRT Secretary was Joseph Yue (Assistant Professor, Associate Faculty Director, William M. White Business Library, University of Colorado at Boulder). He was elected NMRT vice president in 2001-02 and president in 2002-03.

² The *NMRT Archives Policy* (revised in 2002) can be found at http://www.lib.lsu.edu/ALA/nmrt/nmrt_archives_policy022202.pdf.

³ The new *NMRT Archives Policy for Electronic Documents* can be found at http://www.lib.lsu.edu/ALA/nmrt/nmrt_elecdocs-policy073101.pdf

⁴ As defined by Chou & Chow, "Knowledge management is concerned with the effective management of enterprise knowledge, the knowledge that an organization lives by and is built upon."

⁵ *ALA Handbook of Organization 2003-2004*, p. 140.

⁶ In this article, we use the terms *digitization* and *digitize* when referring to the conversion of analog information into a digital format. A more useful description is "Synonymous with scanning, it is the conversion from printed paper, film, or some other media, to an electronic form where the page is represented as either black and white dots, or color or grayscale pixels." (<http://www.princetonimaging.com/scanning/glossary/>)

⁷ Optical Character Recognition software is used to convert images of letters and numbers into computer-recognizable codes such as ASCII or Unicode.

⁸ Portable Document Format software uses PostScript programming language to describe what a page actually looks like.

⁹ Rich Text Format is the internal markup language used by Microsoft Word and understood by dozens of other word processors. RTF is a universal file format that pervades practically every desktop. (<http://www.oreilly.com/catalog/rtfpg/desc.HTML>)

¹⁰ Jennifer Cargill, Dean of LSU Libraries, was very supportive of the project and agreed to allow the use of the LSU Libraries Web server to house both the database and the archival files.

¹¹ As noted earlier in *Software Requisites*, the searchability function requires that files be converted into a pure text format before being saved as HTML files.

¹² The consulted resources are provided in the bibliography of the *NMRT Archives Policy for Electronic Documents*.

¹³ The metadata elements used by the DDTF are listed in appendix B of the *NMRT Archives Policy for Electronic Documents* at <http://www.lib.lsu.edu/ALA/nmrt/nmrtelecdocs-policy073101.pdf>

¹⁴ Another important aspect of using metadata schemas—consistency—is addressed later on in Phase III.

¹⁵ DBMan SQL is an “off-the-shelf” database application for Windows and Unix, using Perl as the underlying language for manipulating data in relational databases.

¹⁶ Perl is the Web scripting language used to generate the HTML interfaces for the administrative panel and for the search results.

¹⁷ The *Basic Search* interface is accessible at <http://www.lib.lsu.edu/ALA/nmrt/basic.html>

¹⁸ The *Advanced Search* interface is accessible at: <http://www.lib.lsu.edu/ALA/nmrt/advanced.html>

¹⁹ Any pure text editor, such as WordPad or Notepad, could probably be used in this capacity.

²⁰ Although the process has not yet been automated, a form of standardization has been introduced with the creation of specific templates that are to be used when submitting official reports.

REFERENCES

- ALA handbook of organization 2003-2004*. (2003). Chicago, IL: American Library Association.
- Chou, S. T., & Chow, E. T. (2000). Essential factors in knowledge management with COTS products. In D.G. Schwartz, M. Divitini, & T. Brasethvik (Eds.), *Internet-based organizational memory and knowledge management* (pp. 42-56). Hershey, PA: Idea Group Publishing.
- The Dublin Core metadata initiative element set*. [On-line]. Available: <http://dublincore.org/documents/dces/>.
- NMRT archives policy* (2002). [On-line]. Available: http://www.lib.lsu.edu/ALA/nmrt/nmrt_archives_policy022202.pdf.
- NMRT archives policy for electronic documents* (2001). [On-line]. Available: http://www.lib.lsu.edu/ALA/nmrt/nmrt_elecdocs-policy073101.pdf
- PrincetonImaging.com. [On-line]. Available: <http://www.princetonimaging.com/scanning/glossary/>
- Procedures for inputting records into the database* (2003). [On-line]. Available: http://www.lib.lsu.edu/ALA/nmrt/input_procedures_records.pdf
- Procedures for inputting reports into the database* (2002). [On-line]. Available: http://www.lib.lsu.edu/ALA/nmrt/input_procedures_reports.pdf
- RTF pocket guide* (2003). [On-line]. Available: <http://www.oreilly.com/catalog/rtfpg/desc.html>.
- Schwartz, D. G., Divitini, M., & Brasethvik, T. (2000). On knowledge management in the internet age. In D.G. Schwartz, M. Divitini, & T. Brasethvik (Eds.), *Internet-based organizational memory and knowledge management* (pp. 1-23). Hershey, PA: Idea Group Publishing.
- Stuhlman, Daniel D. (2001). *Knowledge management experts* [On-line]. Available: <http://home.earthlink.net/~ddstuhlman/kmexpert.htm>

APPENDIX A

Database Template (Administrative Panel)

Database Manager: Add a New Record

Add a New Record

Title:	<input type="text"/>
Author/Creator:	<input type="text"/>
Subject	<input type="text"/>
Keywords:	<input type="text"/>
Description:	<input type="text"/>
Publisher:	<input type="text"/>
Other	<input type="text"/>
Contributor:	<input type="text"/>
Date (YYYY-MM-DD):	<input type="text"/>
Resource Type:	<input type="text"/>
Format:	<input type="text"/>
Resource Identifier:	<input type="text"/>
Language:	☐ English ☐ French ☐ German
Source:	<input type="text"/>
Relation:	<input type="text"/>
Coverage:	<input type="text"/>
Rights Management:	<input type="text"/>
Addressee Name:	<input type="text"/>
Status:	<input type="text"/>
Url:	<input type="text"/>
Cut and paste document text here:	<div><div></div><div></div></div>

Add Record

Reset Form

| [Home](#) | [Add](#) | [View](#) | [Delete](#) | [Modify](#) | [List All](#) | [Admin](#) | [Log Off](#) |

APPENDIX B

Excerpt of photocopied minutes:

American Library Association JUNIOR MEMBERS ROUND TABLE Executive Board Minutes	January 8, 1979 9:30 a.m.-12:30 p.m.
The ALA/JMRT Executive Board meeting was called to order by President Joyce Wyngaarden at 9:40 a.m.	
EXECUTIVE COMMITTEE MEMBERS PRESENT Sharon Stewart, Beth Bingham, Jon Scheer, Pam Bonnell, Gail Terwilliger, Sue Stroyan, Nancy Bettencourt, Janis Bruwelheide, Carolyn Jeffries, Kippy Jagoe, Celia Moore, Bob Halcums, Kristen Dahles, Tom Carter, Stan Bustetter, Ann Scott, David Warren, Carrie Alford, Janice Beck, Ann Vidor, Blanche Wyson, Jan Dickens, Anthony Benoit, Cary Frederick and Margaret Stewart.	
GUESTS: Karen Stoudenmire, John A. Mooman, Mary B. Mayer, Barbara Tooker, Larry Iram, G. David Anderson, and Al Tweedy.	

Example of full metadata record in database template:

Title:	ALA/JMRT Executive Board Meeting 1
Author/Creator:	Margaret Stewart
Description:	Minutes of the first of two JMRT Executive Board meetings held at the 1979 ALA Midwinter Conference in Washington, DC.
Keywords:	minutes, midwinter meeting, Executive Board
Publisher:	ALA, JMRT/NMRT
Other Contributor:	
Date (YYYY-MM-DD):	1979-01-08
Resource Type:	Minutes
Format:	html
Resource ID:	
Source:	paper document converted to rtf/html
Language:	English
Relation:	
Coverage:	
Rights Management:	open access
Addressee Name:	
Status:	final copy of official minutes
Text:	
URL:	Link

Excerpt of digitized minutes (accessed from "Link" button):

American Library Association JUNIOR MEMBERS ROUND TABLE Executive Board Minutes January 8, 1979 9:30 a.m. - 12:30 p.m.
The ALA/JMRT Executive Board meeting was called to order by President Joyce Wyngaarden at 9:40 a.m.
EXECUTIVE COMMITTEE MEMBERS PRESENT Sharon Stewart, Beth Bingham, Jon Scheer, Pam Bonnell, Gail Terwilliger, Sue Stroyan, Nancy Bettencourt, Janis Bruwelheide, Carolyn Jeffries, Kippy Jagoe, Celia Moore, Bob Halcums, Kristen Dahles, Tom Carter, Stan Bustetter, Ann Scott, David Warren, Carrie Alford, Janice Beck, Ann Vidor, Blanche Wyson, Jan Dickens, Anthony Benoit, Cary Frederick and Margaret Stewart.
GUESTS Karen Stoudenmire, John A. Mooman, Mary B. Mayer, Barbara Tooker, Larry Iram, G. David Anderson, and Al Tweedy.