

## Case Study

# Clemson University Thurmond Speeches Series Indexing Project

JAMES CROSS AND MARSHA MCCURLEY

**Abstract:** The Speeches series of the Senator Strom Thurmond Collection contains over 100 cubic feet of material spanning more than 50 years of public service. Because of the need for improved access, the Special Collections Unit and the Cataloging Unit of the Clemson University Libraries initiated the Thurmond Speeches Indexing Project, using the AMC MARC format to index the series. In this article, the authors describe the planning and implementation of the project and discuss some of the lessons learned, as well as future developments that might occur in the index.

*About the author:* James Cross has been the Thurmond archivist for the Special Collections Unit of the Clemson University Libraries since 1987. He was an archivist at the Richard B. Russell Memorial Library of the University of Georgia Libraries from 1984 to 1987. Marsha McCurley has been the head of the Cataloging Unit of the Libraries since 1986. Prior to 1986 she was head of automation/cataloging at Augusta College and retrospective conversion cataloger at Clemson.

THE THURMOND SPEECHES INDEXING PROJECT is the prototype for a file using NOTIS, a bibliographic utility, to provide item- and folder-level access to a selected group of archival materials (in this case, Senator Strom Thurmond's speeches). The AMC MARC format is used to input these materials. Cooperation among archivists, catalogers, and systems personnel has resulted in a file for which full subject and archival series access is available, as well as a customized biographical/authority file for researchers and archivists to use under NOTIS 5.0. Additional access by title, date, place, group presented to, and type of speech is also provided. The file is loaded as a separate NOTIS institution and will be accessed via a menu option screen.

## Background

The driving force behind the index was the library administration's interest in doing a project that used the existing local cataloging system, NOTIS, to provide access beyond the collection- series-level description and cataloging already done by Special Collections and Cataloging for the main on-line catalog. The decision to proceed was an administrative one, made by the dean of the Libraries. The administration hoped that, in time, all the collections in the Special Collections Unit could be accessed in the same fashion.

The reasons for choosing the Speeches series for this project were purely pragmatic. This is not to say that the project participants were unmindful of the historical and literary value of the series. The material in Speeches reflects the thoughts, opinions, and philosophy of Senator Thurmond. They can be used to document the senator's public views on a variety of subjects important to understanding America in the twentieth century, among them civil rights, states' rights, segregation, military preparedness, anticommunism, foreign policy, and trade.

Researchers also use the series as a surrogate for other collection series that are unavailable for use. For example, Nadine Cohodas, author of *Strom Thurmond and the Politics of Southern Change*, used the Speeches series for her research on the portion of her book concerned with the 1948 states' rights campaign and as a way of illustrating the senator's stand on civil rights during this career. Literary researchers can use the series to compare Thurmond's speaking style and delivery with other politicians of his era or to investigate the changes in political speech making over time.

One of the pragmatic reasons for choosing the Speeches series was availability. Although the bulk of the Thurmond Collection arrived at Clemson in 1982 and processing of the collection started shortly thereafter, in 1987 (when the present Thurmond archivist arrived) there were no manuscript series open for research use. Only two or three series were at the point where preliminary descriptions had been drafted. Even after opening series for research use became a priority, only five series were available for the project once the administration decided that the project would go forward. The Speeches series was one of those available.

It was felt that indexing speeches would prove to be simpler than indexing correspondence, since a speech is written to cover a limited number of specific subjects, whereas this may or may not be true of correspondence. Accuracy in the description of the speeches was another area of concern, and the hope was that the indexing project could address it. The number of folders in the series (often more than one hundred in an archives box) made it impossible to verify the accuracy of the folder titles. Invariably there would be corrections to the folder information concerning the speeches (title, date, locations, etc.). The creation of the folder-title list for the series took several months—any corrections

would require a significant effort to re-edit the list. The indexing project provided not only a means by which these corrections could be made but also the possibility that the file itself could be used to create a new folder-title list once the corrections were complete.

The major reason for choosing the Speeches series for indexing, however, was the need for improved access to the series due to its arrangement. The Speeches series is organized into two subseries: General File and Originals. The General File contains the "record" copy of a speech, whereas Originals contains speeches Thurmond actually used. Thus, the speeches show additions, deletions, corrections, and markings indicating emphasis, pauses, and phrasing. The Speeches series contains not only speech material—such as addresses, remarks, and statements on the Senate floor—speeches, and statements, but also articles (by and about Thurmond), questions and answers (interviews), "Reports to the People" (newsletters), telegrams, telephone logs, and a small amount of correspondence. This material is arranged chronologically overall, then by type of speech (address, statement, etc.) within each year, and then chronologically within each type. Retrieval is difficult, since the archivist or researcher must know both the type of speech of the desired item *and* its date. In addition, the two subseries of speeches have approximately a 40 percent overlap of material. An automated index would serve to bring these items together.

The catalyst for the project was the propitious provision in 1988 for Roger Leemhuis, a professor from the History Department, as the Thurmond Speeches indexer. Until this occurred, Special Collections was unable to go forward with the project because the unit could not spare a staff member to provide detailed subject analysis for the material. Unfortunately, only three weeks separated the announcement that Leemhuis would be the indexer

and the time he started work—three weeks to develop the forms and procedures he would need.

## Development

It must be emphasized that this file is considered an index, not a catalog, because it provides more in-depth subject analysis than conventional cataloging. The word *index* is used here by the authors in the sense of "not a catalog," rather than in the definitional sense of "a systematic guide to the contents of a file, document or group of documents, consisting of an ordered arrangement of terms or other symbols representing the contents and references, code numbers, page numbers, etc., for accessing the contents."<sup>1</sup> The decision to use the term *index* in this way was made early in the project to prevent confusion between the Libraries' on-line catalog, which requires strict adherence to cataloging standards because the records are eventually entered into OCLC, and the index, which is a local database and not tied to a national bibliographic utility (or its input standards). No research was done on indexing standards or approaches, since the short period available to actually develop the project before the indexer's arrival did not allow the luxury of research. Nor are any of the participants in the project trained as indexers.

One of the first tasks faced by the staff in both Special Collections and Cataloging was to understand each other's terminology. For example, the archival definition of the word *series* ("file units or documents arranged in accordance with a filing system or maintained as a unit because they result from the same accumulation or filing process, the same function, or the same activity; have a particular form; or because of

<sup>1</sup>Heartsill Young, ed., *ALA Glossary of Library and Information Science* (Chicago: American Library Association, 1983), 116.

Figure 1. AMC Cataloging Worksheet

Archivist's name:

CONTROL NO. (035): MSS

SERIES

MAIN ENTRY (1XX):

TITLE (245):

FORM:

INCLUSIVE DATES:

BULK DATES:

EXTENT (300): SIZE:

PKGING UNIT:

ACCOMPANYING MATERIALS:

DIMENSIONS:

ORGANIZATION/SERIES (351):

ARRANGEMENT:

ARCHIVAL LEVEL:

LINKING NOTE (580; series only; must have 773):

GENERAL NOTE (500):

ADDITIONAL PHYS. FORM (530):

REPRODUCTION NOTE (533):

TYPE:

PLACE:

AGENCY:

DATE OF REPRO:

LOCATION OF ORIGINALS/DUPS (535):

CUSTODIAN:

POSTAL ADDRESS:

COUNTRY:

TEL:

BIOGRAPHICAL/HISTORICAL NOTE (545):

SUMMARY/SCOPE (520):

SOURCE:

(continued on page 356)

some other relationship arising out of their creation, receipt or use”<sup>2</sup> is quite different from the definition used by catalogers (“a group of separate items related to one another by the fact that each item bears, in addition to its own title proper, a collective

title applying to the group as a whole.”)<sup>3</sup> The issue of file integrity was also an important one in the early stages of the project, especially to the Cataloging staff—it was understood from the start that the index would never be combined with any other file or database, so that some localizing and adjustment of cataloging stan-

<sup>2</sup>Lewis J. Bellardo and Lynn Lady Bellardo, *A Glossary for Archivists, Manuscript Curators, and Records Managers*, Archival Fundamental Series (Chicago: Society of American Archivists, 1992), 32.

<sup>3</sup>Michael Gorman and Paul W. Winkler, ed., *Anglo-American Cataloging Rules*, 2nd ed., 1988 rev. (Chicago: American Library Association, 1988), 622.

Figure 1. AMC Cataloging Worksheet (*continued from page 355*)

<b><u>PROVENANCE</u> (561):</b>	
<b>ACQUISITION SOURCE</b> (541):	<b>RECEIVED FROM:</b>
ADDRESS:	
STATUS:	DATE:
<b>RESTRICTIONS</b> (506):	
<b><u>FINDING AIDS</u></b> (555):	
<b>CASE FILE CHAR.</b> (565):	
<b>PUBS. NOTE</b> (581):	
<b>*ACTIONS</b> (583):	
<b>REFERENCES</b> (510): CITATIONS:	
LOCATION:	
<b>PREFERRED CITATION</b> (524):	
<b>USE/REPRODUCTION TERMS</b> (540):	
<b>ASSOCIATED MATERIALS</b> (544):	
<b><u>SUBJECTS</u></b> (6XX) (prioritize):	
<b><u>GENRE/FORM TERMS &amp; sources</u></b> (655):	
<b>OCCUPATION TERM</b> (656):	
<b>FUNCTION TERM</b> (657):	
<b><u>ADDED NAME ENTRIES</u></b> (7XX) (prioritize):	
<b>PLACE ADDED ENTRY</b> (752):	
<b>HOST ITEM ENTRY</b> (773; series only; must have 580):	

dards could be done to meet the special needs of the index. Once the definitional process was completed, work began on designing the indexing form, the screen format for the index and the procedures for the actual indexing.

The indexing form developed was a variation of the AMC worksheet (see figure 1) already in use for description of collections and series entered into the main on-line catalog. The pattern of patron requests of speeches from other collections determined the access points needed, as did sugges-

tions from both the Special Collections and Cataloging staffs. Modification in the usage of MARC tags provided some additional access; for example, the manuscript control number was input into the unused thesis note field in order to get keyword access for the folder numbers. This resulted in the form shown in figure 2.

Two types of documents in the Speeches series posed special problems. Newsletters, entitled "Reports to the People," and press releases after 1971 were batched in folders, rather than treated as individual items, as



Figure 3. Newsletter Indexing Form

<b>SPEECHES INDEXING FORM – Reports to the People</b>	
<b>CONTROL NO. (035):</b> Mss 100-11A- <b>AUTHOR (1XX):</b> Thurmond, Strom, 1902– <b>TITLE (245):</b>    	<b>LOCATION:</b> Special Collections
<b>DATE:</b> _____	
<b>PLACE (260):</b> _____	
<b>NOTE 1 (502):</b> Mss 100 11A	
<b>NOTE 2 (502):</b> _____	
<b>LINKING NOTE (580):</b> Part of: Strom Thurmond Reports to the People	
<b>PREFERRED CITATION (524):</b> Mss 100, Strom Thurmond Collection, Speeches series, General File subseries, Special Collections, Clemson University Libraries, Clemson, SC 29634	
<b>PUBS. NOTE (581):</b> _____	
<b>SUBJECTS (6XX):</b>    	
<b>ADDED NAME ENTRIES (7XX):</b> _____	
<b>HOST ITEM ENTRY (773):</b> Strom Thurmond Reports to the People	
<b>DATE:</b> _____	

of indexing forms for each folder. The releases are now indexed at the folder level; individual indexing continues for the newsletters.

Another important decision to be made was how to subject-analyze the material in the series. Time constraints did not allow the construction of a thesaurus unique to the index, so Library of Congress Subject Headings (LCSH) were chosen as the basis for subject indexing. In addition to being a convenient solution for the immediate problem, this choice also allowed those in-

involved in the project to take advantage of the on-line authority control and maintenance of subject headings available in NOTIS.

Most important, from the user's point of view, it eliminated the need to learn an entirely new vocabulary to use the index. This minimized the problems researchers would face when switching from the libraries' main on-line catalog to the index. Although the indexer hews closely to the LCSH, some subject headings have been created just for the index; these local sub-

**Figure 4. Press Release Indexing Form**

**SPEECHES INDEXING FORM — Press Releases**

**CONTROL NO.** (035): Mss 100-11A- **LOCATION:** Special Collections  
**AUTHOR** (1XX): Thurmond, Strom, 1902-  
**TITLE** (245):

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

**DATE:** \_\_\_\_\_

**PLACE** (260): \_\_\_\_\_

**NOTE 1** (502): Mss 100 11A

**NOTE 2** (502): \_\_\_\_\_

**LINKING NOTE** (580): Part of: Strom Thurmond Press Releases

**DATE:** \_\_\_\_\_

**PREFERRED CITATION** (524): Mss 100, Strom Thurmond Collection, Speeches series, General File subseries, Special Collections, Clemson University Libraries, Clemson, SC 29634

**PUBS. NOTE** (581): \_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_

**SUBJECTS** (6XX):

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

**ADDED NAME ENTRIES** (7XX): \_\_\_\_\_

\_\_\_\_\_

**HOST ITEM ENTRY** (773): Strom Thurmond Press Releases

**DATE:** \_\_\_\_\_

ject headings concern major Supreme Court decisions.

Once all the decisions were made, a small test file was created to test those decisions against reality. Some modifications were made in the indexing instructions as a result of this testing. For example, the manuscript number originally had dashes in it to separate different components (collection number, series number/subseries letter, folder number, e.g., Mss 100-11A-101) from one another. The NOTIS system was unable to sort these numbers correctly.

so the dashes were replaced by spaces. Indexing sheets are continuously reviewed, and there are occasional meetings between the Special Collections and Cataloging staffs to resolve problems and revise procedures.

### Current Procedure

The indexer devotes approximately thirty hours a week to indexing the speeches and creating the worksheet descriptions. He also checks the on-line au-



thority file and provides authority records for the index. The indexer uses his knowledge of American history to find the proper subject terms in the LCSH. The basic criteria for determining the assignment of a subject term is whether the item provides substantial information on that subject. The indexer attempts to include all legitimate alternative search strategies when choosing the headings without necessarily including all possible variations of the heading. He also tries to be as specific as possible (e.g., historic buildings in Charleston would be assigned the term "Historic buildings—Charleston, SC" rather than the broader term "Historic buildings"), although grouped press releases are of necessity assigned broader terms. At this point in the project, the indexer has established a consistent set of terms that he uses for items covering certain subjects, such as the Vietnam War.

Important Supreme Court cases discussed in an item are indexed under the name of the case. Reciprocal headings are used when assigning terms covering foreign relations between the United States and another country. For example, if the relationship between Cuba and the United States were the topic of a speech, the terms "United States—Foreign relations—Cuba" and "Cuba—Foreign relations—United States" would be used. The names or individuals are used as subject terms if the person is a head of state, an official of cabinet rank or an important office (like the FBI or NASA), a prominent member of Congress, a prominent South Carolinian, an important member of the senator's staff (such as the legislative director), a member of the senator's family, or an individual whose papers are also at Clemson. All others are assigned a broad subject heading (e.g., "United States—Officials and employees") when possible.

If a new authority record needs to be created for an individual's name or corporate body, the indexer provides the in-

formation to cataloging staff for the creation of that record, using the same forms used by the archivists to provide authority control for the collection and series records in the main catalog (see figures 5 and 6). Like the authority records in the main catalog for the collection- and series-level cataloging entries, these authority records, not the AMC record itself, provide the biographical or organizational history information. This allows circumvention of limitations on the record size that presently exist in NOTIS.

Worksheets are converted into the AMC MARC format and input by a Special Collections library technical assistant who was trained by the Cataloging staff. The process of entering the records into the index is accelerated by NOTIS's capacity to derive records from "template" records. The technical assistant creates a record and, if a similar type of record later needs to be entered into the database, can derive from the previous record and make the appropriate changes. The Cataloging staff then reviews the bibliographic records and creates authority records to match them. Because the index is in a separate NOTIS institution, it has a separate authority file and indexes. No circulation or order/pay/receipt records are used in this file. Archival collection and series records are also included in this file, providing some archival hierarchical arrangement. The index currently contains over 3,800 bibliographic records and over 1,200 authority records.

### Lessons Learned

A number of very important lessons were learned during the planning and implementation process. One of the first is that the development of a prototype is not something one should try to accomplish in a three-week period if one can avoid it. If more time had been available, a more systematic approach would have been possible. Research on indexing standards and

**Figure 5. Personal Name Authority Record Form**

**SPECIAL COLLECTIONS**  
**AUTHORITY FILE RECORD CREATION/AMENDMENT REQUEST**  
***PERSONAL NAMES***

Date: \_\_\_\_\_  
Requester's  
Name: \_\_\_\_\_

**This request is for (check one):**

**CREATION** of New Authority Record \_\_\_\_\_  
**AMENDMENT** of Existing Authority Record \_\_\_\_\_

**Manuscript No. (where name appears):** \_\_\_\_\_

**Name ("best-known" form):** \_\_\_\_\_  
\_\_\_\_\_

**Birth/Death dates (if known):** \_\_\_\_\_

**Other forms of name:** \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

**Biographical information:** \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

Return completed form to Cataloging

approaches would have been undertaken and some of the mid-course corrections needed during the life of the project might have been avoided. Fortunately, the short development time apparently has not impaired the utility of the project.

The project required close cooperation among archivists, catalogers, and programmers. The catalogers and programmers know the system best and are in a position to make suggestions based on that knowledge. These in turn enable the archivist to make more effective use of the system and

to avoid some of its pitfalls. Catalogers are also in a position to use their knowledge of subject analysis and cataloging conventions to assist in the design of an index such as the Speeches Index. Archivists know their collections best; only they can decide the best way to describe the material. Above all, there is a need for flexibility and creativity on everyone's part in a project such as this one, because one is "pushing the limits." Different and innovative ways of looking at the format or the options provided in the system enable staff to overcome obstacles

Figure 6. Corporate Name Authority Record Form

<b>SPECIAL COLLECTIONS</b> <b>AUTHORITY FILE RECORD CREATION/AMENDMENT REQUEST</b> <b>CORPORATE NAMES</b>	
	<b>Date:</b> _____ <b>Requester's</b> <b>Name:</b> _____
<b>This request is for (check one):</b> <div style="display: flex; justify-content: space-around; margin-top: 5px;"> <span><b>CREATION</b> of New Authority Record _____</span> <span><b>AMENDMENT</b> of Existing Authority Record _____</span> </div>	
<b>Manuscript No. (where name appears):</b> _____	
<b>Name ("best-known" form):</b> _____ _____	
<b>Other forms of name:</b> _____ _____ _____	
<b>Organizational history information:</b> _____ _____ _____ _____ _____ _____ _____	
Return completed form to Cataloging	

and the system to provide the type of access desired.

### Recent Developments and the Future

By early 1991 enough records had been entered to begin beta-testing the index. At the same time, a statistical report program, SAS (Statistical Analysis System), was run in an attempt to create a container list for the series from the index by extracting the folder number, title and date, place, and type of speech from their respective fields and arranging them in proper sequence on the list. Although the SAS program was

largely successful in creating the container list, the results of this experiment and the beta-testing indicated that changes needed to be made in some of the fields to ensure proper results for certain types of searches and to create a "clean" copy of the container list. For example, a date field was created in the subject (6XX) field so that searches by date would be done using a subject search rather than a keyword search, since the latter resulted in more entries being generated than were valid, due to the retrieval of dates in the 580 and 773 fields.

Inputting new index records ceased while the records already in the index were corrected. Retrospective conversion (RECON) of all records then in the database was completed in May 1992. A second printing of the container list showed only a few remaining typographical errors, which were easily corrected; some editorial work remains to be done by the systems staff. The input of new indexing records resumed; to date, speeches up to the year 1969 are in the database. Indexing and the creation of new authority records continued during the RECON project—half of the General File subseries is indexed, covering the years 1935 to 1973. As noted above, individual input of press releases has ceased.

Two hurdles remain before the index be-

comes fully operational. The index will be included on a new local menu interface that allows public access to multiple institutional databases such as the index. In addition, our library is currently awaiting the implementation of NOTIS 5.1.1, which restores the capability of allowing separate institutional files to have individual guide screens, help screens, and displays; these had been taken away by NOTIS 5.0.

With full implementation of the index as a separate institution and the completion of screen and display design, the index can be made accessible for patron use. Both the researcher and the library staff will then have a powerful tool that will provide access into the series at a variety of entry points, thus saving time and increasing the utility of Senator Thurmond's speeches.