

Indexing Interviews in OHMS: An Overview

Louie B. Nunn Center for Oral History, University of Kentucky Libraries

Since its inception, efficient access to oral history in an archival setting has been challenging. Oral history involves audio or video recording on time-based media. Unlike print or photographic resources, it is cumbersome for users or researchers to visually scan an oral history for content unless a transcript is created textually representing the recorded event. In the absence of a verbatim transcript, users must listen to each moment in an interview in order to determine the potential for relevant information discovery. Until speech recognition and artificial intelligence technologies become more accessible and affordable, transcription remains difficult and expensive.

When conducting fieldwork or interviews in an analog recording context (audio cassette and reel to reel tape), oral historians, folklorists, anthropologists and other related practitioners would sequentially “log” their tapes on paper to create efficient ways to get back to the moments that they determined were “important” or “useful.” These logs would include a variety of metadata types, including partial transcripts, titles, keywords and descriptions, which all corresponded to a counter number. Analog technologies lacked true time code and were measured by a counter that was usually machine specific. While these tape logs were essential for the researcher who was writing a book or an article, this usefulness rarely transferred into an archival context. As formats were migrated, the counter numbers (which were specific to each machine) failed to maintain consistency. As a result, archivists and researchers utilized the tape logs instead as the source for comprehensive item-level metadata. However, the archivist/user/researcher still had to manually scan the audio/video in order to access the corresponding segments represented in the log.

OHMS creates a method to log an oral history interview with descriptive metadata that corresponds to true timecode in a way that translates to the user side of the oral history process. In his article [OHMS: Enhancing Access to Oral History for Free](#), OHMS creator and Nunn Center Director Doug Boyd poses a useful analogy pertaining to OHMS indexing and the user experience:

If you were to go to the grocery store and ask the manager, “Where can I find the Cheerios?,” the manager would, most likely, smile and say “aisle ten.” At that point, you would, upon identifying the location of aisle ten, proceed down aisle

ten until you were at the specific location of the Cheerios. In many ways, this is how I view the role of indexing an oral history interview.¹

Another useful analogy for thinking about indexing using OHMS, from a user perspective, is to think about a book. The OHMS index provides the user/researcher the browse function of a book's table of contents, while simultaneously providing the granular and precise searchability and the navigational interface of a book's index. The result is a user experience that makes an indexed interview browsable and searchable in ways that link the user's search terms to corresponding moments in the audio or video interview.

The primary goal when indexing oral history interviews in OHMS is to create a useful and affordable user interface for oral histories. It is equally important to create a dynamic framework that optimizes the effectiveness and efficiency of information discovery in both single interviews as well as in oral history collections of different sizes.

Indexing is a subjective process and can vary greatly in breadth and depth. The Nunn Center has worked out three levels of indexing, which correspond to the depth and breadth of our desired outcome. Level one indexes make use of fewer metadata fields while also requiring less detail for each field. Level one segments include a title and keywords. A level two index provides a good balance between the sparse information of a level one index and the detailed, longer process of creating a level three index. Level two indexes contain a title, keywords and/or subjects, partial transcripts and a brief synopsis. A level three index is the most detailed level, providing users with the greatest amount of information. Level three indexes make use of all available metadata fields.

An index is only as good as the indexer and the effectiveness of the segment or story-level metadata being created. The following is a guide to assist indexers in maximizing the effectiveness and future usefulness of the oral history indexes they create in OHMS. This is not a rulebook; it is a model containing guidelines we have developed at the Nunn Center over the past year. Our intention is to update this guide in an ongoing fashion.

If you would like to see an example of a strong OHMS index, view the interview with Buffalo Trace Distillery [Master Distiller Elmer T. Lee](#).

What gets Indexed?

¹ [Douglas A. Boyd. "OHMS: Enhancing Access to Oral History for Free" Oral History Review, Volume 40, Issue 1, Summer: Fall 2013.](#)

Each interview contains hundreds of possible access points that could potentially prove useful to researchers some day. Indexing involves making good decisions about the representation of data in an interview. Most professional efforts in the archival community focus on creating collection or series level descriptive metadata, and when the resources warrant it, at the item level. OHMS provides a framework for creating access points within the item by constructing searchable, descriptive metadata packages at the segment level that correlate to a corresponding moment in the audio or video recording.

For each segment created in an OHMS index, the indexer can create information in the following fields:

- Time Stamp
- Partial Transcript
- Segment Title (required)
- Subjects (semi-colon delimited)
- Keywords (semi-colon delimited)
- Segment Synopsis
- GPS Coordinates
- GPS Description
- Hyperlink
- Link Description

Oral history interviews are not just random recordings; they are structured interviews that result in a dynamic interaction between interviewer and interviewee. Often, interviews conducted as part of a larger project will contain a natural structure that is driven by emerging patterns created by the interviewer. For example, interviews from the Nunn Center's From Combat to Kentucky Oral History Project (a collection of interviews with veterans returning from Iraq and Afghanistan who are enrolled in higher education) follow a similar structure from interview to interview:

- Interview introduction
- Biographical information
- Making the decision to join the military
- Enlistment
- Boot camp, training and specialization
- Deployment
- Numerous stories documenting their deployment experiences
- Transition out of the military
- Transition into higher education
- Conclusion

In an archival context, indexing an oral history interview is about effectively documenting and describing interview transitions. Our goal is to provide the future user of the interview with a descriptive search and browse framework. Continuing the book analogy, the indexer creates chapter marks that correlate with these natural transitions, and titles and describes each chapter in the book. Understanding the interview's structure is crucial for effectively documenting these transitions.

- In an oral history interview, questions often indicate the locations of subject transitions.
- By choosing to index (or not to index) a moment in an oral history interview, the indexer is controlling significance. This is a very powerful position to be in since the index you create will frame all future navigation of that interview. It is useful to ask yourself the following questions when deciding whether to index a moment (or not):
 - Does it represent a major topic within this segment of the interview?
 - Will inclusion of this topic be useful to future researchers?
 - Does this segment contain unique, compelling, or interesting content that, even if discussed briefly, stands out in the interview?
 - Does the content correlate to major historical events such as the Great Depression or World War II?
 - Will the creation of an index point for this segment represent this interview effectively to a user, researcher or even a search engine?
 - Have I created index points that will help a user navigate this interview?
 - Have I represented the content effectively?

An OHMS index is not a substitution for a transcript. The OHMS index creates a structured and descriptive navigational framework for effectively and efficiently engaging with an oral history interview in an online archival context. The transcript provides text-based, keyword searchability of an interview's verbatim content. In an OHMS index, you have the opportunity to map natural language to concepts. For example, an individual could talk about living under segregation for much of a 3-hour interview, yet never mention the word "segregation." This would require the user of a searchable transcript to make a series of educated guesses about words or phrases that might relate to the topic of segregation in order to determine the interview's relevance to their information needs. A good index will tell the user that the interview contains a discussion about segregation, even if the word "segregation" is never mentioned in that interview.

How to Index: The Mechanics of Indexing in OHMS

This section has been taken from the *OHMS: Getting Started Guide*. For more details, please consult that guide.

Activating an interview's indexing module for the first time will bring up a view such as this one:



In order to begin indexing, you must press the “Play” button  on the player. The player must be playing in order for you to create an index point, which is created by pressing the “Tag Now” button at the appropriate moment. When you press “Tag Now” you are presented with a series of empty descriptive fields for you to fill out.

Tag Data

* Indicates a required field.

Interview:

Timestamp:

Partial Transcript:

Segment Title:

Current Thesaurus: **Standard Titles**

Subjects:

Subjects should be separated by a semi-colon.
Current Thesaurus: **LCSH-SpokeDB**

Keywords:

Keywords should be separated by a semi-colon.
Current Thesaurus: **SPOKE**

Segment Synopsis:

GPS:

Coordinates should be specified in XX.XXX, YY.YYY format where X is north or south and Y is east or west.

Hyperlink:

The indexer can control the player within the tagging module. The nature of indexing warrants that the indexer will, inevitably, tag a segment late. In other words, you don't know something is important until after you have already heard it. For this reason, the player backtracks a few seconds each time the tagging module is activated.

Tag Data

* Indicates a required field.

Interview:

Timestamp:

⏪ ⏮ ⏭ ⏹ ⏩ UPDATE TIME

When creating an OHMS index, the time stamp is critical. You are creating metadata that corresponds to a particular time code. As a result, you will need to make sure that placement of the time stamp (the precise location in the interview where you want the segment to begin) is exactly where you want it to be. This can be done by controlling the player in the tagging module.



Creating an Index

OHMS Metadata fields at the segment level include:

- Time Stamp
- Partial Transcript
- Segment Title (required)
- Subjects (semi-colon delimited)
- Keywords (semi-colon delimited)
- Segment Synopsis
- GPS Coordinates
- GPS Description
- Hyperlink
- Link Description

Although we have identified a function for each index field, each institution should discuss specific policies regarding the particular use of these fields.

Time Stamp

This is created as soon as you press the “Tag Now” button, and will determine, on the user side, the location of the starting point for the audio or video segment. However, you will need to adjust your time stamp so that your segment correlates to the appropriate time. Place the time stamp where you think is the most appropriate entry point to a segment from the user’s perspective. Often, a segment beginning corresponds to an interviewer’s question. Be careful that you place the time stamp in a comfortable location, such as at the beginning of a question or answer, rather than mid sentence.

Partial Transcript

The partial transcript field can be used to convey the verbatim transcription of the recorded segment. This can be important because it creates an orientation point for the user of an oral history index. For example, a segment could be titled “Segregation” but the narrator may not address this topic immediately. Even minimal population of this field will transform the user experience. Although some may utilize this tool to fully transcribe the segments, the Nunn Center transcribes only the first 140 or so characters of the segment, following the Nunn Center’s transcribing standards. For example, in a segment titled “*Major events in the history of the Wild Turkey Distillery*” from an interview with Master Distiller Jimmy Russell:

Sample Segment Transcript:

YOUNG: Well tell me about your distillery. What makes it unique?

RUSSELL: Well, this distillery--you know, we just built a new distillery--the first new distillery, completely, that's been built in Kentucky for, uh, gosh, I don't know how many--it's been a long, long time since comp--every--everybody's had to remodel and expand because the bourbon business got too good. We were across the road here, and we sit right on top of the Kentucky River and US Highway Sixty-two. And we'd expanded as far as we could expand. So, we had this property up here on the hill, and we come up here and built a complete new distillery. Now, we haven't changed anything--the way we make it. We're still using the same formula, the same--all I can

vouch for, for our yeast, it's fifty-nine years old. It was--the yeast--it was here when I got here, and we're still using the same yeast culture.

Transcript Indicated in the OHMS "Partial Transcript" Field:

Well tell me about your distillery. What makes it unique?

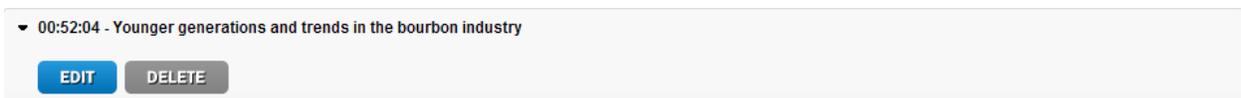
Often the verbal interaction being reflected in the partial transcript will involve multiple speakers. The Nunn Center has chosen to transcribe only the first speaker's dialogue, although other repositories may choose different transcript structures that identify speakers when appropriate. For example:

Young: Well tell me about your distillery. What makes it unique?
Russell: Well, this distillery--you know, we just built a new distillery--

If you have placed the time stamp in a potentially awkward place for partial transcription (several people talking at once), you can adjust the time stamp and make up for the content by utilizing additional fields in the index. For usability purposes, we strive to adjust the time stamp to correspond to a single speaker for this field.

Segment Title

Segment titles are required. It is best to use descriptive titles that act as chapter titles for the interview. The titles in the OHMS Viewer function as a table of contents, offering the user a quick glimpse of the contents of the interview. When creating a title for a segment in OHMS, it is helpful to do so in a way that assumes that if a user/researcher never opens the title tab to explore content further, they would still understand the essence of the content of a segment and of the overall interview. The title of an OHMS index is the primary access point for browsing the contents of an interview.



- Be as succinct and descriptive as possible. Use descriptive titles that clearly and plainly represent the content of this segment to the future user.

Example Title: *Distillery activity* is better represented as *Wild Turkey's bourbon making process*

- Do not try to be clever or use ambiguous or esoteric references in titles.

Example Title: *Good things come in small packages* is better represented as Title: *The development of the Blanton's brand and other small-batch bourbons*

- Avoid using slang.
- If you struggle with creating titles, ask yourself the following questions:

What is the main point of this segment of the interview?

How would the topic most effectively be represented to a contemporary audience?

Does your title require esoteric information about the topic in order to understand the content it represents?

Does the title you chose require clarification?

- The Nunn Center begins indexes with the title *Interview introduction*, if there is an introductory portion of the interview. Similarly, if there is a concluding portion of the interview, the final segment is titled *Interview conclusion*.
- The Nunn Center only capitalizes the first word of the title and the proper nouns present in the title.
- You can create title root words such as "Childhood" that represent parts of a longer story or major categories that repeat. They are then qualified by adding a -- and a narrower, more specific term or phrase. For example:

Title: Childhood--Growing up in Frankfort, Kentucky

Title: Childhood--Family

Title: Childhood--Elementary school

- It is unnecessary to use the interviewee's names in segment titles

Title: *Russell becoming a Master Distiller* should be *Becoming a Master Distiller*

Title: *Russell's personal accomplishments* should be *Personal accomplishments*

- The following are examples of good titles we have seen:

Title: Biographical background and education

Title: Roach's parents (generally interviewee's names do not need to be included in titles but can be useful for clarification when referring to their family, friends, etc. [i.e., "Roach's parents" instead of "Parents"])

Title: Early teaching career

Title: Race relations in Frankfort

Title: Becoming the Vice President

- The following are examples of titles that could be improved:

Title: Narrow escape

Title: Fisticuffs

Title: A story about Lincoln and Seward (Should be "A story about President Lincoln and Secretary of State Seward")

Title: Becoming "the Veep"

Subjects

This field is one of two OHMS indexing fields designed for the entry of descriptive terms that represent the content of each segment. These are crucial for providing effective searchability. The Nunn Center uses this field primarily for *Library of Congress Subject Headings*, though it may be used for any controlled vocabulary. A lesson on using *LCSH* is beyond the scope of this guide. However, *Library of Congress Subject Headings* can be vague and general at the interview level, and the segment level often demands even greater specificity. Nevertheless, we have found them invaluable for interoperability of metadata. The indexer's goal in creating subjects and keywords is to provide representative and descriptive terms for searchability, but it is also to map natural language to concepts. As mentioned earlier, a narrator/interviewee can talk for 3 hours about living under segregation without ever saying the word "segregation." It is important for the indexer to use effective descriptors, such as the *Library of Congress Subject Heading*:

African Americans--Segregation--Kentucky

- This field allows for multiple entries, which are separated by a semi-colon. It can be used in conjunction with a thesaurus to control or suggest terms to the indexer (see the *OHMS: Getting Started Guide*).
- The Nunn Center strictly controls the use of this field. If a term is not an auto-suggested term, it can be either added to the thesaurus before using or added as a keyword.

- If an indexer is not familiar with cataloging standards, we recommend that they use the keywords field exclusively.

▼ 00:00:00 - First job at the Wild Turkey Distillery

[EDIT](#) [DELETE](#)

Partial Transcript: This is Al Young interviewing Wild Turkey's Master Distiller Jimmy Russell on September the 30th, 2013 for the Kentucky Bourbon Tales oral history project.

Segment Synopsis: Jimmy Russell is introduced. He talks about his family and about growing up in Lawrenceburg, Kentucky. He discusses how he came to work at the Wild Turkey Distillery and his first job in the quality control department.

Subjects: Distilleries--Kentucky;Whiskey industry--Kentucky;Family-owned business enterprises.;Quality control.;Families;Distillers.;Distillation.

Keywords

Like the subjects field in OHMS, the keywords field contains descriptive terms that represent the content of each segment. These are crucial for searchability. The Nunn Center uses the keywords field for any term that represents or describes the content in the segment. As with the subjects field, the indexer's goal in creating of keywords is to provide representative and descriptive terms for searchability, and also to map natural language to concepts. Unlike the subject field, keywords are specific and local terms that effectively communicate the contents of the segment. Using the segregation example referenced in previous sections, accompanying keywords might include:

Restaurants; Water fountains; Lunch Counters; Bathrooms

Users will often search by keywords, so the longer and more complicated our keywords are, the less effective the index will be.

- Unlike subjects, we use this field liberally. We encourage as much specificity as the indexer deems necessary to effectively communicate the contents of the segment. This should be balanced with brevity in order to maintain searchability.
- When tagging keywords, try to use words that the interviewee uses. If it is a somewhat esoteric term, such as a nickname, a regional term, or an unusual phrase, put it in quotation marks.
- Quotation marks should also be used for terms that are used by the interviewee but are no longer appropriate in a cataloging or metadata context. This is important to retain the original context of an interview and also deploys good metadata practice. For example, interviewees in some of our older interviews, such as those in the Robert Penn Warren Civil Rights Oral History Project, make frequent use of the word "negro". This is no longer an appropriate term for African Americans in a metadata context. Using quotation marks here indicates that the term is used within the interview, but is not one assigned by the indexer.
- Keywords are an excellent way to balance out the generality of controlled vocabulary by using more specific terms. Believe it or not, until recently there was no Library of

Congress Subject Heading for “Bourbon.” The keywords you choose can be an effective way to compensate for an oversight such as this one.

- This field allows for multiple entries, separated by a semi-colon, and can be used in conjunction with a thesaurus to control or suggest terms to the indexer (see the *OHMS: Getting Started Guide*). The Nunn Center uses this field for local terms and tags as well as project specific vocabularies.
- The Nunn Center recommends making keywords as short and descriptive as possible, unless, of course, they are proper nouns.
- In general, keywords are only one or two words long. For example, “making bourbon in Kentucky” is descriptive, but so long that it is unlikely to be searched together as a phrase. Instead, break this concept into keywords like “Bourbon”, “Kentucky” and “Distilling”. It is far more likely for a researcher to search one of these terms than that they would use this specific phrase.
- Try to make keywords as specific as you can. For example, the word “fun” might be used by the interviewee, but it is too vague and subjective to be a useful search term. Instead, if possible you should use terms related to the activity they are describing.

▼ 00:08:24 - Brand names and advertising campaigns

[EDIT](#) [DELETE](#)

Partial Transcript: Well, you know, from time to time, you know, some things change and some things remain the same. But was Wild Turkey's American Honey a stroke of genius or just plain luck?

Segment Synopsis: Russell talks about his involvement in the creation of Wild Turkey American Honey and discusses their original target market for the product. He talks about Wild Turkey's "Give 'em the bird" advertising slogan. He discusses the name changes the distillery has gone through over the years and how the name "Wild Turkey" was chosen.

Subjects: Target marketing.;Branding (Marketing);Advertising.;Consumers.;Sales promotion.;Whiskey industry--Kentucky;Distilleries--Kentucky;Prohibition.

Keywords: Promotional campaigns;Advertisements;"Give 'em the bird";Wild Turkey brand name;Wild turkey hunts;Labels;Wild Turkey American Honey;Women;Lawrenceburg (Ky.);Saffell Distillery;Anderson County Distilling Company;Austin Nichols;101 proof;Jake turkey

More on Subjects and Keywords

Make sure your subjects and keywords in an OHMS index work together. Both fields are searchable, so there is no reason to represent the same content in both. The descriptive terms you choose should thoroughly represent the content of the segment. As with the interview or collection level, choose words that accurately and comprehensively describe the segment. When possible, we recommend that you provide indexers with relevant controlled vocabularies (even for keywords) to ensure consistency between interviews and projects.

Segment Synopsis

The segment synopsis is designed to contain a descriptive statement about the segment. Synopses are used to further clarify what the segment is about. The segment synopsis is

particularly useful to express concepts or topics that are too complex to be conveyed by keywords alone. A good synopsis complements the other descriptive metadata fields in the OHMS index. For example, if the title of the segment is “Growing up during the Great Depression”, an effective synopsis might “Johnson reminisces about her childhood and the hardships her family experienced. She talks about her parents’ struggle to make ends meet and ways that economic hardship affected her as a child.”

- The Nunn Center typically refers to the interviewee and interviewer by last name in a synopsis. It is too informal to refer to them by their first names, and may be confusing. If both the interviewer and interviewee have the same last name, a first initial should be used to distinguish the two (i.e., A. Kelly and D. Kelly). If the interviewees have both the same last name and first initial, full names should be used (i.e., William Marshall and Willard Marshall).
- Synopses should also be relatively short, typically no longer than a sentence or two. They provide a brief summary, but should not be a substitute for listening to the segment.
- The synopsis field in OHMS is very versatile. It can also be used to provide additional comments or details such as the presence of technical problems in the original recording. These comments should be placed in brackets. Repositories may wish to create a standardized list of these commonly used phrases in order to maintain consistency.
- In general, use synopses to clarify what is stated in the title and describe some of its nuances.

▼ 00:14:02 - Involvement of consumers in the bourbon industry

[EDIT](#) [DELETE](#)

Partial Transcript: Well now the first time I met you, I met you at the old Seagram plant over on, uh, the other part of Anderson County.

Segment Synopsis: Russell talks about traveling and speaking with consumers in the bourbon industry and how they have become more educated about bourbon over the years. He talks about the location in the warehouses that produces the best tasting bourbon. He talks about the popularity of the Kentucky Bourbon Trail and Wild Turkey's expanding tourism industry.

GPS Coordinates

This field connects a user from the oral history segment to a location on Google Maps, though the institution can change the map resource if desired. Coordinates are entered in the format "XX.XXX, YY.YYY" where X is latitude (north or south) and Y is longitude (east or west). Only one set of coordinates per segment is allowed at this time. There must be a space following the comma.

GPS Description

This field serves as a label for the GPS coordinates you have provided. Each institution will need to determine standards for how they want to textually describe the GPS coordinates, but we typically put it in the standard Library of Congress format for places. The text entered into this field will serve as the linkable text in the OHMS viewer. The image below is how the GPS coordinates and text will appear in the OHMS viewer:

GPS: Wild Turkey Distillery (Lawrenceburg, Ky.)
Map Coordinates: 38.0379712, -84.8496304

Hyperlink

The hyperlink field allows a hyperlink to be present, connecting the oral history segment to an external resource of any type. Only one hyperlink is allowed per segment at this time. The full hyperlink must be entered into this field:

Hyperlink:

Link Description:

Hyperlink Description

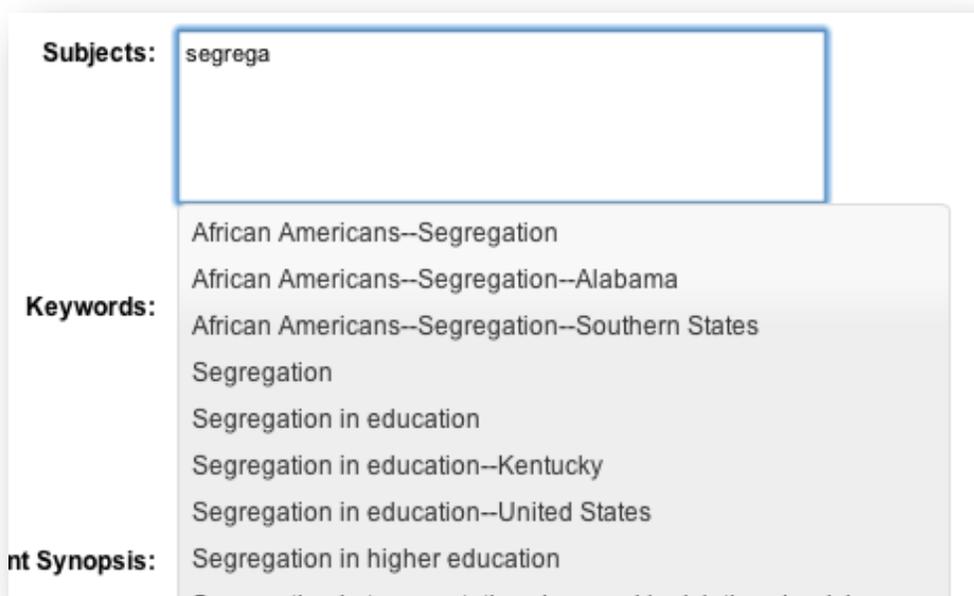
This field serves as a label for the hyperlink that you have provided. Each institution will need to determine standards for how they want to textually describe these hyperlinks. As with GPS coordinates, the text entered into this field will serve as the linkable text in the OHMS viewer (as opposed to the raw hyperlink).

Hyperlink: Image of Jimmy Russell outside of the Wild Turkey Distillery warehouses

Using Controlled Vocabularies and Thesauri

Title, subjects and keywords fields in OHMS can correlate to an uploaded thesaurus of terms. Each field can be assigned a different thesaurus for a particular interview. If a thesaurus has been uploaded and properly assigned in the metadata manager, terms will be autofilled or suggested based on a partial keying of letters. If the indexer begins typing “segrega” in the subjects field, all of the terms in the assigned thesaurus containing “segrega” will be suggested. The indexer then selects the term desired, and the term is dropped into the subjects field.

See the thesaurus manager section in the *OHMS: Getting Started Guide* for more information on uploading and assigning a thesaurus.



Concluding Thoughts

An effective segment in an OHMS index should utilize fields so that each metadata type works together to convey an overall meaning. Repetition of terms in these fields can be useful. For example, you could use the term “Segregation” both in a title and as a keyword. However, remember that the fields should also complement each other, with different types of terms in each section. If the term “Segregation” is in the title, then also listed as a subject or a keyword, the term does not necessarily need to be used in other fields as well.

Additionally, the fields in an OHMS index are meant to be flexible. Although the Nunn Center utilizes the subjects field for Library of Congress authorized headings, and the

keywords field as descriptive tags without an authorized source, we are open to utilizing these fields in alternative ways, depending on the particular project. For example, our Haiti Memory Project contains interviews conducted in Haitian Kreyòl. Because of this, we decided to use the subjects field for descriptive terms in English (both LC and local terms) and the keywords field for descriptive terms in Haitian Kreyòl. We have also utilized the synopsis field to convey non-descriptive notes pertaining to the segment, as needed.