# The Oral History Metadata Synchronizer (OHMS) Enhancing Discoverability for Oral History Collections

www.stephaniesapienza.com/ohms-workshop

STEPHANIE SAPIENZA

MARYLAND INSTITUTE FOR TECHNOLOGY IN THE HUMANITIES (MITH) <u>SAPIENZA@UMD.EDU</u> @SAPIENZA77



An oral history is a unique form of primary historical document. I would argue that among primary documents, they are the most humanistic kind. They are personal and sometimes intimate, and reveal aspects of history and the individual experience in a way that is uniquely human. If you merely want evidentiary value and facts, you can simply read a transcript (when they exist). But to hear a person speaking the actual words lends a certain amount of intimacy and understanding that reading text cannot provide. Ideally, you can use the text from a tape log, summary or transcript as a sort of scaffold, to get into the interview at the exact point that's relevant to your research.

But even though oral history collections are fairly ubiquitous in all sorts of cultural and historical repositories, we're only starting to develop ways to 'scaffold' them in ways which allow us to get at the subjects and access points we need. An audiovisual historical document of any kind, but especially an oral history, inherently comes with what Sarah Florini has called a 'temporal commitment barrier.' Meaning, we don't have to time to listen to the whole thing in order to know if it contains the information we need for our research.



The Oral History Metadata Synchronizer (OHMS) is a tool that starts to solve some of these problems. OHMS was developed at the Louis B. Nunn Center for Oral Histories at the University of Kentucky. It allows you to import metadata from your oral history collection at varied levels of granularity. You can use it to create an extremely rich, descriptive index of each section which includes subjects and keywords, GPS coordinates, hyperlinks to outside resources, and a synced transcript. Or, if you have a lot of oral histories and only a little time (and you're more of a More Product, Less Process kind of person) you can import a large batch of metadata about your collection at once at a skeletal level and create very basic index points with only crucial information for broad level access.



In this workshop we're going to look at some examples of OHMS in action, including some different levels of indexing and descriptive granularity. We'll look at how completed collections are bundled and made accessible by a couple of institutions. We'll look briefly at the application's back end indexing and transcript syncing user interfaces so you can get a sense of how to use OHMS. And then we'll do some hands-on work.



OHMS Application: Hosted by University of Kentucky Libraries. You must acquire an OHMS Repository account (free), and conduct all your indexing, syncing and QC within the application on their servers. After the work is complete, you can export XML files, which you then upload into your CMS and with which you can generate URLs. You can also export a .CSV document which is meant to be used separately (but in conjunction with) the exported XML, for importing into your CMS (such as Omeka or ContentDM).

OHMS Viewer: You install the Viewer on your own server instance or CMS. You upload the exported XMLs here and generate URLs for each. If you want to create a browse and search interface or a finding aid, you can use exported CSV file from the OHMS Application to harvest metadata and link to the individual elements.

# OHMS Viewer Examples

- ▶ Viewer Example (YouTube): Interview with Steve Zahn, May 29, 2013
- Viewer Example (SoundCloud): <u>The Wisdom Project Podcast</u> <u>Episode #004: Martin Luther King Jr., Malcolm X, and Robert Penn</u> <u>Warren</u> (University of Kentucky Libraries)
- Viewer Example (Omeka with embedded Brightcove): Interview with Beulah Collins, August 1, 1983
- Viewer Example (HTML5): <u>Interview with Fred Noe, November 14,</u> 2013 (University of Kentucky Libraries)



<u>The Athens Oral History Project</u> - Uses EAD to package metadata exported from OHMS in CSV format, in conjunction with separately exported XML for each OHMS Viewer page.

<u>University of Wisconsin – Milwaukee</u> (Society of American Archivists Oral History Project) - Uses ContentDM to package metadata from OHMS in CSV format, in conjunction with separately exported XML for each OHMS Viewer page.



The indexing process is about making right-sized decisions about the representation of your interview data. The archival field trends towards collection-level processing and descriptive metadata at the series level. Item level processing with hefty descriptive data represents the richest use of professional time, and the richest possible resource. In happy medium land between those two is indexing. With OHMS, you'll create 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.'



Level 1 - Either an index or a transcript, but usually not both. Minimal indexing with just questions as partial transcripts, and one small set of keywords and subjects applied to all sections.

Level 2: Example: Interview with Lemuel LaRoche, January 12, 2015 (University of Georgia)

Level 3: Most granular, with custom keyword and/or subject thesauri attached to index and multiple keywords/subjects assigned.

Substantial text or transcript excerpts in the 'partial transcript' fields. Includes summaries.

Example: Oral History Interview with Patricia Middleton Briscoe from the Lakeland Community Heritage Project (MITH/University of Maryland)



The OHMS Application allows you upload a plain text file of your transcript. It can contain time code, but does not need to as it will utilize its own sync functionality to match select moments in the interviews to their corresponding text in one-minute increments. The sync tool operates with a duo of beeps. The first (low) beep indicates that it is playing the first 10 seconds of a one minute increment. You find the text on the transcript and follow along with your cursor as the 10 seconds runs down. Then you hear a second (high) beep, which means you should click on the word being spoken at this moment. This creates a time stamp for that minute, and the tool immediately jumps a minute forward and repeats the process. This takes a small bit of practice, but once you've gotten a handle on it, it's much like a game and goes very quickly.

TIP: Use Control+F as soon as you hear the low beep to orient yourself as to where you should follow with your cursor to mark the sync point.



You can create custom keyword thesauri in a simple one-column CSV document and upload it into the OHMS Application. In the Metadata module for an interview, you can then set the keyword thesaurus to that uploaded taxonomy. Then, when you're in the Index module and are adding keywords, they will auto-populate as you type. You can use an OHMS CSV template for importing the initial metadata for large groups of interviews at once, dramatically speeding up your workflow. This requires mapping your metadata from its original source into the template, but the field elements are fairly straightforward.

For workflow optimization (when working with a team), each module - Metadata, Index, Sync - has a pull-down menu indicating its status towards completion. If working on your own, you can simply jump from 'In Progress' to 'Completed.' If you're working with a team, a team member can use 'Ready for QC' and 'Active QC' to indicate that they're working on reviewing the module. Once all modules are in 'Completed' status, you'll see that the Status column will change from 'In Progress' to 'Completed' for the whole record, which means it's ready to export. There's also a Notes field which functions like a bug or ticket tracker for issues or questions that need to be flagged or resolved within a team. Once they're resolved, they disappear.



If you've been given your own individual OHMS User account with your personal email, log into the OHMS Application at <u>https://ohms.uky.edu</u>.

If not, click on the link to the list of (7) account logins, choose one, and use it to log in. Please write your name next to the account you've claimed. You may work in groups. At this point, all of us are logged into the same repository with separate user accounts. This means we'll have to split up which person or group does which activity.

### Activity 1: Upload a Thesaurus

GOAL: Upload a customized controlled vocabulary as a keyword thesaurus.

- 1. Download the CSV thesaurus of keywords for Lakeland Community History Project (linked to from workshop site, slide #5);
- 2. Click on the Thesaurus Manager. Choose 'Add New Thesaurus,' add a name for it, and under 'Type,' choose Keywords;
- 3. Select the thesaurus CSV file from your computer, and click save. You should now see it at the bottom along with the linked data LOC Subject Headings thesaurus.

#### Activity 2 : Create a new Interview

GOAL: Create a new, single interview for Pamela Boardley.

- 1. Download the Pamela Boardley Oral History Metadata and Tape Log from Google Drive (linked to from workshop site, slide #6);
- 2. Go to the Interview Manager tab and click on New, which will lead you to a blank Metadata Editor interface;
- Use the data from the above Google Drive document to cut and paste the correct elements into the right fields. Make sure to populate all the required fields with a red asterisk, including title and media format (Audio or Video). <u>You must include the Media URL for the Vimeo link,</u> and the embed code.
- 4. Click the pulldown menu option to assign the newly-uploaded custom thesaurus for keywords, and the Library of Congress Subject Headings thesaurus for subjects. Click Save.

### Activity 3 : Create index tags

GOAL: Create a new index with 3-4 tags for the Pamela Boardley interview.

- 1. In the Interview Manager, click on the Index link in the record for the newly-created Pamela Boardley interview;
- Use/refer to the Pamela Boardley Oral History Metadata and Tape Log you downloaded to determine how the index sections should be broken up. Press play on the embedded SoundCloud player and scrub to that time stamp in the interview;
- 3. Click 'Tag Now' to access the indexing user interface pop-up window;
- 4. Create a segment title which describes the basic premise of that segment, and cut and paste the summary from the tape log into either the Partial Transcript or Segment Synopsis sections;
- 5. Assign a few subjects and keywords as you type, they should auto-populate as they pull from the assigned thesauri. Click Save.

### Activity 4 : Import a batch of new interviews

GOAL: Use the OHMS CSV import template with mapped metadata to import a small collection as a batch.

- 1. Download the mapped CSV import template for a historic public radio series, 'Comment on a Minority' from Google Drive (linked to from workshop site, slide #8);
- 2. Go to the Import Interviews tab. Click 'Choose File,' and upload the CSV document;
- Go back to the Interview Manager tab. You should now see several newly-created items listed there. Click on the Metadata link for one to check that the metadata was properly mapped.



Attach the streaming media information to the John Boggs episode to connect it to the SoundCloud File. Click on the Metadata module, click on 'Audio' for media format, and then paste the Media URL and embed code (both are in the Google Doc). Click save. Now open the transcript module for the John Boggs episode, and upload the .txt file to attach it to the interview. Click save. Once you see the 'Sync' module link turn blue, click on it. Try to sync the first 3 or 4 minutes. The audio file will start playing immediately. At the low beep, click Ctrl+F and find the text being spoken. Follow with your cursor as the last ten seconds of that minute play. When you hear the higher beep, click on the word being spoken at that time. The sync module will assign that minute marker to that spot and jump to the next minute marker. You don't have to sync the entire transcript, but once you've done what you can, click Save.

#### Activity 5 : Sync the first 3-4 minutes of an OHMS interview

GOAL: Attach streaming media file information to one of the batch-imported interviews, upload a transcript and sync the first few minutes.

#### (Continued from previous slide)

- 4. In the Interview Manager, click on the Transcript link for the Boggs record. Upload the transcript .txt file to attach it to the interview. Click Save.
- 5. Once you see the 'Sync' module link turn blue, click on it.
- 6. Try to sync the first 3 or 4 minutes. The audio file will start playing immediately. At the low beep, click Ctrl+F and find the text being spoken. Follow with your cursor as the last ten seconds of that minute play. When you hear the higher beep, click on the word being spoken at that time. The sync module will assign that minute marker to that spot and jump to the next minute marker. You don't have to sync the entire transcript, but once you've done what you can, click Save.

Attach the streaming media information to the John Boggs episode to connect it to the SoundCloud File. Click on the Metadata module, click on 'Audio' for media format, and then paste the Media URL and embed code (both are in the Google Doc). Click save. Now open the transcript module for the John Boggs episode, and upload the .txt file to attach it to the interview. Click save. Once you see the 'Sync' module link turn blue, click on it. Try to sync the first 3 or 4 minutes. The audio file will start playing immediately. At the low beep, click Ctrl+F and find the text being spoken. Follow with your cursor as the last ten seconds of that minute play. When you hear the higher beep, click on the word being spoken at that time. The sync module will assign that minute marker to that spot and jump to the next minute marker. You don't have to sync the entire transcript, but once you've done what you can, click Save.