



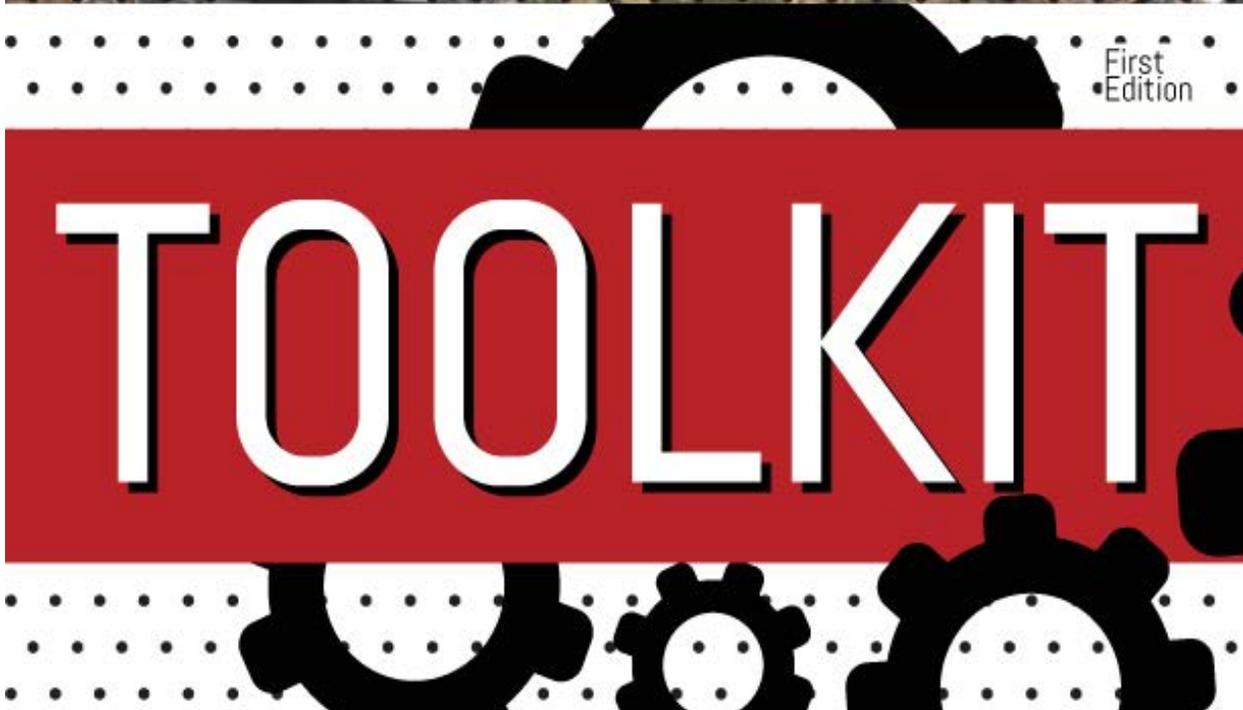
OH LA

Oral History in the Liberal Arts

**The
Oral History
in the Liberal Arts**

First
Edition

TOOLKIT



Archiving Oral Histories from Start to Finish

Jenna Nolt
Digital Initiatives Librarian
Kenyon College

Last updated 7/5/17

Table of Contents

1. Introduction

1.1. Planning like a Librarian

2. Data Collection

2.1. Metadata

2.2. Starting the Interview

2.3. Expanding the Scope

2.4. Supporting Materials

3. File Transfer & Processing

3.1. File Formatting

3.2. File Transfer

3.3. Naming Files & Folders

3.4. Backup

3.5. File Editing & Merging

4. Transcription, Captioning, and Indexing

4.1. General Transcription Guidelines

4.2. Transcription Software

4.3. Indexing with OHMS

5. Hosting, Access, & Digital Preservation

5.1. Content Management Systems

5.2. Streaming

5.3. Institutional Repositories & Digital Preservation

1. Introduction

1.1. Planning like a Librarian

This toolkit takes a technical and archival approach to collecting, processing, hosting, accessing, and preserving oral histories. It is intended as a guide for oral history practitioners and collaborating library and IT partners. It was created as part of the [OHLA \(Oral History in the Liberal Arts\)](#) initiative, supported by the [GLCA Expanding Collaboration Initiative](#) and funded by the [Andrew W. Mellon Foundation](#).

This toolkit lays out the processes through which oral historians can simplify their technical workflows and insure that their interviews can be accessed and preserved long-term. Collecting basic information like name spellings during interviews can prevent mistakes and avoid time-consuming research later on. Utilizing best practices for files makes the transfer, editing, and archiving processes easier. Using transcription and indexing software makes the interviews more usable to a wider audience. Understanding what's necessary for digital preservation can be the difference between long-term availability, and a project that disappears within a few years. Here are some of the questions we'll help you consider:

- What information should you collect with your interviews?
- How will you transfer, convert, merge and clean your A/V clips?
- Will the interviews be fully transcribed, partially transcribed, indexed, or some combination thereof?
- If you are creating an online collection, where will it go?
- How long should/will your collection exist?

2. Data Collection

2.1. Metadata

Metadata is, simply put, data about data. For the purposes of this document, when we talk about metadata, we are referring to descriptive metadata. Descriptive metadata describes and identifies resources. It consists of things like interview title, names of interviewee and interviewer, location of interview, date of interview, keywords, etc. This metadata is essential

for organizing your interviews as you process, upload, access, and preserve them. We created the [Interview Metadata form](#) to make it easy. This form records metadata that will be useful regardless of your workflow. It can be easily converted to comply with [Dublin Core](#) and other common metadata standards and is intended to work well with a variety of systems, including the [Oral History Metadata Synchronizer \(OHMS\)](#), which is discussed in 4.3.

2.2. Starting the Interview

The archiving process begins the moment a recording begins. At the start of each interview, the interviewer should state the following:

- Project title
- Name(s) of interviewee(s)
- Name(s) of interviewer(s)
- Location of interview
- Date
- A few words regarding the expected content of the interview

For example:

“This is Jenna Nolt with the Family Histories project, and I am here interviewing Jane Nolt on March 18, 2016. Jane is going to tell me about her experiences growing up in the rural Midwest in the 1930s.”

While this is metadata that should go into the [Interview Metadata form](#) as well, it is important to include the basic context in the interview file itself. This makes it easy to differentiate interviews during the editing and upload processes, and if an interview file is ever separated from the information around it, it will still be possible to contextualize it.

2.3. Expanding the Scope

Regardless of the primary purpose of your interviews, including certain kinds of information will make them more searchable now, and more relevant to new research approaches in the future.

- If the interviewee uses unfamiliar words or confusing phrases, ask them to explain the meanings.

- If the interviewee uses unusual proper nouns (like names or place names), ask them to spell them out.
- Try to include specific information like dates (date of birth, date of marriage, date of moving to a new place, etc.), place names, etc.

This kind of information opens up the interview for the purposes of genealogical research, geolocation, and chronolocation. It can also expand the relevance of the interview into areas like linguistics, folklore, and many others.

2.4. Supporting Materials

In addition to the interviews at the heart of oral history, most projects can benefit from supporting materials, which can be used to contextualize, enrich, and promote the project. This can often include:

- Photographs
- Letters
- News articles
- Physical items that can be filmed or photographed

These materials might be introduced naturally during the course of the interview. If they are, it is good to be prepared with a scanner, cell phone, etc. to capture them. Most smartphones will take a high-resolution picture, but be careful not to hold it at an angle. Portable scanners are another option. If it isn't possible or convenient to scan or capture images of these materials, it is recommended that they be described in sufficient detail during the interview so as to enable them to be requested at another time.

There can also be additional materials that do not come up during the interview, but emerge as a result of further research. Some possibilities are:

- Maps
- Census Data
- Directories
- Legal Documents
- Genealogical Information

These materials might be available through search engines, institutional databases, or local sources. These can be useful resources in developing an online collection or website around your interviews.

3. *Data Collection*

3.1. File Formatting

If your recording device has options for file format, choose a format intended for digital preservation. Appropriate formats are **AIFF or WAV** formats for audio files, and **AVI or Quicktime Movie** formats for video. As a general rule, prioritize a lossless format (a format that preserves all data from the file) over a lossy format (one where some data is not saved). It is also important not to compress the original files. Lossiness and compression both make files harder to preserve over time. For more information, refer to the [Library of Congress Sustainability of Digital Formats site](#). The format you choose on your recording device is intended for the initial, preservation copy of your files. In 3.5., we discuss exporting a lossy, compressed format as an access copy.

3.2. File Transfer

The process of transferring files from your recording device to your computer differs depending on the device, but these are the three most common methods:

1. SD or micro-SD card
2. USB connection
3. Direct to cloud storage

Saving directly to cloud storage is preferable for several reasons. Cloud storage is automatically backed up in multiple places and checked for fixity, so it is less like to become corrupted than files on a local device. In addition, it makes the files easily available to anyone else working on the project. The only potential disadvantages are that it requires a decent internet connection, and that not all recording devices have the option of saving to cloud storage. We recommend using Google Drive, which offers 25 GBs of storage free with every account, and if you have institutional access, you have unlimited storage. Saving to cloud storage is an in-app process, setting it up depends on the device and recording software you are using.

When transferring and storing files via SD, USB, or any non-cloud method, it is important to keep copies of your files on at least two separate hard drives, such as your computer and an external hard drive or thumb drive. Otherwise, you risk file loss and corruption. You will want to store copies of your files locally for editing purposes, and if you don't have internet access in the field. However, we only recommend using local storage for the short term, and moving your files to cloud storage as soon as possible.

♥ NOTE: Depending on recording equipment, format, and length of the interview, the final result might be broken up into multiple files when you transfer it to your computer. This may happen automatically, to prevent files from being so large that they are difficult to use. In that case, you will need to merge the files, which is covered in 3.5.

♥ Tip for Windows Users: To transfer large files more quickly and safely, download [Teracopy](#). The Mac operating system has a better default file transfer program than Windows, so an additional program is not necessary.

3.3. Naming Files and Folders

Start by creating a folder for each interview, using the following conventions:

NameofProject_LastNameofIntervieweeFirstNameofInterviewee

For example: CommunityWithin_BookerGeorge

This folder should contain all of the A/V recordings of George Booker's interview, as well as any images, documents, etc. associated with Booker, and a completed copy of the [Interview Metadata Form](#).

Name the interview files within your folder using the following conventions:

NameofProject_LastNameofIntervieweeFirstNameofInterviewee_Part#

For example: CommunityWithin_BookerGeorge_Part1

If you have other files associated with the interview, name them as well, using the structure of NameofProject_LastNameofIntervieweeFirstNameofInterviewee, and adding a brief description of the file to the end of the name.

Do NOT use spaces or special characters in your file or folder names. These conventions allow the file and folder names to be both human and machine readable, and each file will be individually identifiable, even if separated from the other component parts.

It is important to have a consistent file naming structure. If you need to make choices about naming structure beyond what is outlined above, document those choices and keep that documentation in the folder with your files.

3.4. Backup

After renaming your files, copy them from your workstation to another storage device, such as cloud storage, or an external hard drive (see 3.2.). These backup files are important in case something goes wrong during the editing, upload, or access process.

Most institutions provide some type of storage, however, the capacity and reliability varies. It may be network drive or cloud storage. Network storage is good for storing large files like A/V because it is local, which provides good upload and download speed. Cloud storage tends to have slower file transfer speeds, but is still a good option. If you are unsure what your institution offers, or whether it is appropriate to use as backup for your interviews, contact your IT department.

3.5. File Editing and Merging

Most recordings benefit from at least minor editing, such as noise correction, removing silences, etc. With video interviews, you might choose to add a title screen, subtitles, and credits. As mentioned in 3.2, larger files are sometimes automatically divided up by your recording device, in which case you will need to merge them. Below are some editing options that are free to use:

	Mac	PC
Audio	Audacity Ocenaudio GarageBand	Audacity Ocenaudio
Video	iMovie	Windows Movie Maker

The above products are well-established and comparatively easy to learn, with widely available tutorials.

Professional options such as Final Cut Pro, Camtasia, and the Adobe suite are also available, but they are expensive and complex. They are powerful tools intended for professional grade audio and video creation and editing, and are generally unnecessary for basic oral history processing. Due to the cost and complexity, we do not recommend these programs unless you already have access to them, and are familiar with how to use them.

After you finish editing your project, you will export a completed file. Unlike your original, preservation file, this should be in a lossy format, such as an MP3 or MP4. This is your access copy, which will stream better than your larger preservation file. Save a copy of this file into the folder you created in 3.3.

4. Transcription, Captioning, and Indexing

4.1. General Transcription Guidelines

Some software requires specific formatting, however, we have developed a set of best practices for transcription over the course of a number of oral history projects:

- Identify the interviewee(s) and interviewer(s) by their full names at the beginning of the interview. If this information is not available, use whatever information is available. If someone is not identified, call them Unknown. If you don't know the names of multiple people, call them Unknown 1, Unknown 2, etc.
- After the first identification, identify individuals by initials.
- Do not use quotation marks.
- Do not transcribe "um" or other meaningless filler noises.
- When someone laughs, coughs, or does a significant hand gesture like pointing, write it as (laughs), (coughs), (waves), etc. These can be integrated into the sentence, and

include their own punctuation. If the expression comes at the beginning of the sentence, capitalize it within the parentheses.

An example of a transcription created using these guidelines can be found [here](#).

4.2. Transcription Software

Manual transcription is a time-consuming process; based on prior projects, we estimate that it takes approximately six times as long as the recording to create a rough draft of a transcription, not including quality checking. There are a variety of software options that can speed up this process, and these options are expanding and improving rapidly as the technology advances. Below are some of the options we have explored.

[Pop Up Archive](#) is an online service that allows you to upload an audio or video file, download an automated transcription with or without timestamps, and stream and view collections with an embeddable tool. They offer one hour of free speech-to-text transcription. OHLA-funded projects currently have access to unlimited basic processing, and we are actively using this service. The quality of the automated transcription relies heavily on the quality of the audio. For more information, see [this blog post](#).

[Youtube](#) has a free automatic transcription option. Learn all about how to use it in [our tutorial](#).

[oTranscribe](#) is a free, open source web application that streamlines manual transcription. It puts the a/v file and the transcription in one window so the transcriber doesn't have to switch back and forth, uses hot keys to rewind and fast-forward, creates interactive timestamps to navigate the transcript with, and automatically saves every second.

[Dragon Speech Recognition Software](#) is a downloadable program that can be locally installed on a Mac or PC. Cost varies based on version and license. Dragon is intended primarily for dictation, and requires the user to set up a profile based on their voice during the installation. After the creation of the profile, files can be loaded into Dragon for processing. The challenge here is that Dragon "learns" and increases its accuracy based on an individual profile's voice, which makes it less accurate for oral history transcriptions where there are multiple speakers. We tested this product, but chose to go with Pop Up Archive for the reasons outlined in the blog post linked above.

[3Play Media](#) offers video captioning, transcription, and subtitling services, but the cost is substantial, starting at \$3 per minute on a pay-as-you-go plan. This is outside the budget of any of our past or current projects, so we have no direct experience with this service.

4.3. Indexing with OHMS

The [Oral History Metadata Synchronizer \(OHMS\)](#) is free, open source software developed by the Louie B. Nunn Center for Oral History at the University of Kentucky. This innovative tool provides a platform for indexing and time stamping interviews with segment-level metadata and/or full transcriptions. To use OHMS, you will need:

1. The [OHMS Application](#) is accessed via a username and password provided by the Nunn Center. Here you can add metadata such as a title, names of interviewer(s)/interviewee(s), description, etc., as well as uploading an existing transcript, transcribing an interview within the interface, and/or “indexing” the interview. More information is available [on their website](#). **The OHMS Application does not store or stream the interview file.** The interview file must be uploaded to a streaming platform, and linked into the OHMS manager via a URL (see 2).
2. A software service to store and stream the interview from. As of June 2017, the following options are compatible with the OHMS Application: [YouTube](#), [Brightcove](#), [Kaltura](#), [Soundcloud](#), or a hosted streaming server.

If you want your OHMS-indexed interview to be publicly accessible, you will also need the following pieces:

3. An installation of the [OHMS Viewer](#). This is free software that can be installed on 3rd party web host by you or your institutional IT department.
4. A content management system to present your collection, see 5.1. for more information.

OHLA-funded projects have automatic access to the suite of tools outlined above. Please access our OHMS tips and video tutorials at [ohla.info](#).

5. Hosting, Access, and Digital Preservation

5.1. Content Management Systems

To make your project accessible you'll need to use a content management system, which will present the embedded interviews within the context you've created for them, which consists of any transcriptions, indexing, metadata, and associated images or writing. All funded OHLA projects are given space on [OHLA's Wordpress site](#).

An alternate content management system that we recommend is [Omeka](#). Omeka allows you to create individual objects with associated metadata that you can access and organize in a variety of ways. So, for example, each interview might be an object with metadata fields like a title, interviewees, keywords, etc. Then the interviews could be organized into exhibits around themes (like cooking, home remedies, religion, etc.). This option offers a lot of flexibility for curation and organization. OHLA does not currently offer a hosted instance of Omeka, however, we can offer technical support, and if you are interested in exploring this option, please let us know.

The limitation to content management systems is that they are not designed to be permanent. There is no mechanism in place to handle changing technologies, or file format changes over time. If you are interested in preserving your project for the long term, you will need to consider digital preservation, and may want to look into the option of using an institutional repository.

5.2. Streaming

Content management systems usually provide their own streaming service (both Wordpress and Omeka do). Streaming is the ability to play audio/video files in the browser, rather than having to download and open them with a separate program on your device. Other streaming options include Internet Archive and YouTube. Using Internet Archive to store your interviews can be a good backup option, because they will maintain the files in perpetuity, whereas services supported in content management systems may not be updated regularly, or may end up unsupported in the future.

5.3. Institutional Repositories & Digital Preservation

Institutional repositories (IRs) are online archives for collecting, accessing, and preserving the intellectual output of an institution. Most colleges and universities have them, but the types of materials they include vary by institution, and may or may not include oral histories.

There are several advantages to using IRs. They are typically optimized to be as findable as possible, which maximizes the visibility of your collection. In addition, your collection appears within the context of the institution's research, so users will recognize it as an academic, authoritative source. Many IRs will do your uploading for you, and at the very least will provide training and troubleshooting support. Additionally, collections are maintained long-term by the institution, meaning you won't have to fix broken links, switch streaming services, or convert file formats to make sure your collections stays accessible over time.

The disadvantages of IRs are that they typically have less options than content management systems to customize the look and feel of your collection, and as mentioned above, not all institutional repositories can or will accept oral histories. As a result, you may consider showcasing your project on a content management system, and looking into the possibility of using an IR to retain your information long term.

For example, OHLA projects that come out of Kenyon College are showcased on the OHLA site, but the files will also be stored in Kenyon's IR [Digital Kenyon](#). Digital Kenyon is maintained by Kenyon's library, and it is the responsibility of the library to insure the access and preservation of the contents of Digital Kenyon in perpetuity. Digital Kenyon is a complex technology stack built on [Digital Commons](#) repository software and backed up with [bepress Archive](#). The underlying software architecture of Digital Kenyon will change over time, but the contents will continue to be accessible as technology advances.