The presentation and long-term preservation of the oral histories in the Houston Asian American Archive

Amanda Focke, Certified Archivist and Digital Archives Specialist, Fondren Library, Rice University

Abstract

The complex, multi-part interviews in the Houston Asian American Archive are created by a rotating group of students, with training, direction, and oversight provided by Dr. Anne Chao (HAAA Program Manager, a part-time position), and support by the Fondren Library. These digital oral histories include full transcripts, consent forms, photographs, and associated analog ephemera, as well as files which provide time-syncing and indexing data to enhance user interaction with the interviews. Tracking of these interviews as they are developed into their multiple parts uses cloud-based online tools and storage options that are available to anyone, while preservation of the final online versions and access to them are enabled via Rice University's institutional repository. **This article will explore the complex system used to provide robust online preservation and access to complex oral histories using undergraduate student work, part-time faculty and library support, cloud-based tools and storage, and space in an institutional repository. Challenges in maintaining this system will also be explored.**

Introduction

The oral history interviews in HAAA are publicly available in Rice University's institutional repository at scholarship.rice.edu. This space, controlled by Fondren Library, provides robust, long-term back-up and reliable public access in a solid, warehouse-style format. In an additional online presence with a focus on a more creative and user-friendly format, the HAAA interviews at haaa.rice.edu, controlled by the Chao Center for Asian Studies, draw upon the files available at scholarship.rice.edu, presenting them in a much more flexible manner.

The interviews include multiple elements, such as transcripts, consent forms, photographs, and associated analog ephemera, as well as time-syncing and indexing files to enhance user interaction with the interviews. This time-syncing and indexing is created by students via the web-based Oral History Metadata Synchronizer (OHMS) tool developed by Dr. Doug Boyd at the Louie B. Nunn Center for Oral History, University of Kentucky Libraries.

On average, twenty-five interviews are created per year, with six active interns working between five and ten hours per week at any given time. Dr. Anne Chao, Program Director for the HAAA, spends approximately 25% of her time training students, developing community connections for potential interviews, communicating with interviewees and interns, and giving HAAA-related

presentations at community events and academic conferences. As archivist for the HAAA program, I spend approximately 5-10% of my time preparing interviews for upload, uploading interviews and OHMS files, filing physical parts of interviews in the HAAA archival collection, collaborating with students on maintenance of training documents and file storage locations, and giving HAAA-related presentations at community events and academic conferences. Fondren Library provides additional support by hosting the institutional repository at Rice University, where the HAAA interviews are preserved, and by hosting the Omeka platform for the haaa.rice.edu website. Rice University's Central IT Department also lends support by maintaining 500GB of back-up server space.

The workflow and tracking of these interviews as they are developed into their multiple parts use cloud-based online tools and storage options that are available to anyone (Box file storage and Google Drive). Online preservation of the final versions and access to them are enabled via Rice University's institutional repository, as well as via the more user-friendly haaa.rice.edu website.
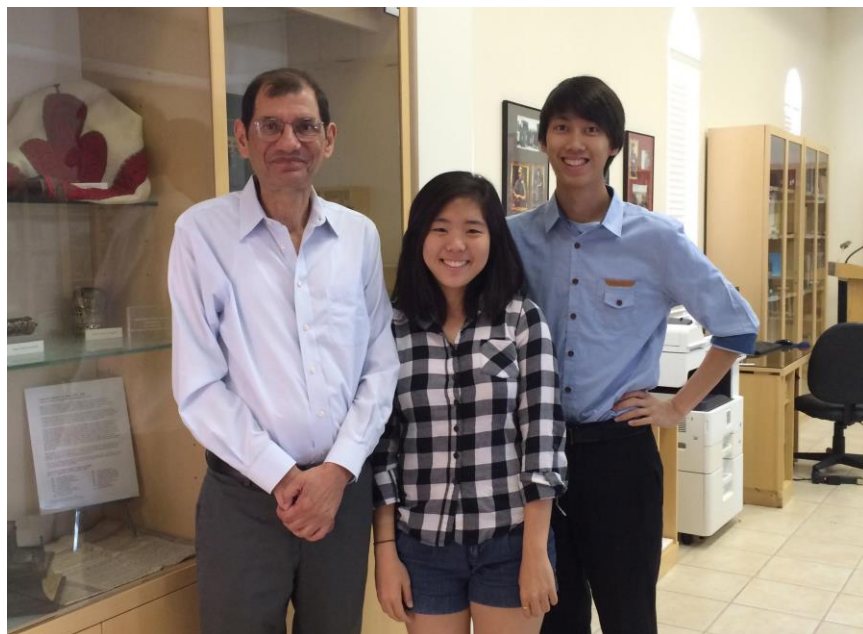
Interview with Sarosh J.H. Manekshaw by Linda Heeyoung Park and Gabriel Wang, June 23, 2014. Houston Asian American Archives oral history interviews, MS 573, Woodson Research Center, Fondren Library, Rice University. http://hdl.handle.net/1911/77551

Undergraduate student work makes this program possible and offers students valuable experience in communication, research, detailed transcription, close adherence to a template, and the abstraction of concepts from interviews to serve as index topics. Some students are able to participate as interns in the HAAA program during all four years of their undergraduate careers and develop significant expertise in this process, serving as mentors to fellow interns.

Students are trained in oral history techniques by Dr. Chao, who also provides a framework of typical interview questions and connects students with prospective interviewees. Students contact these community members, invite them to be interviewed at a time and place of their choice, and provide them with interview consent forms and questionnaires. Interviewees complete the questionnaires and return them to the students, helping them to tailor their interviews. Students also conduct some basic research concerning the interviewees or certain aspects of their lives. For example, if an interviewee worked as a health care professional at the HOPE Clinic, the student would firstly develop a basic understanding of the nature and mission of this clinic.

The consent form includes an introduction to the HAAA program and information concerning what interviewees should expect. The form allows interviewees to provide specific consent in relation to how the interview material is collected and used. For example, they may consent to being photographed, or to having their interviews documented in written, audio or video formats, displayed physically, quoted on the HAAA website, or archived at the Woodson Research Center. They may also consent to the recordings, transcripts, and photographs being made available online at scholarship.rice.edu, and to the interviews being used by others for research purposes. Interviewees may choose to consent to some or all of these options. In a very small number of cases, interviewees have requested anonymity, while others have requested that embargoes be placed on their interviews, pushing their online publication dates forward to particular times in the future. While most elements of the interviews are shared publicly online, each consent form and the accompanying questionnaire are stored in the administrative (non-public) side of the repository and in the physical boxes used for the interviews, which also contain printed versions of the transcripts and photographs, as well as any other ephemera. The questionnaires include addresses and other kinds of personal information which help students to frame their interviews but are not shared online.

Students start out at a pay rate of $12 per hour for their work and enjoy flexible hours, with the ability to work on most of their tasks from any location. Students are recognized for their work, their names as interviewers being included in the transcripts and metadata.

With new interns joining the HAAA program every year, training is an ongoing activity. Full team meetings each semester introduce participants to each other and ensure that everyone is aware of and has access to project documentation. Google Drive was selected as a centrally accessible location for storing training manuals as well as the main tracking spreadsheet for processing the interviews. In the HAAA Google Drive, there are documents describing transcription and use of the time-syncing application, known as the Oral History Metadata Synchronizer (OHMS), as well as a blank consent form/questionnaire, example questions, a transcript template, and an example of a thank you note.

In 2018, student Priscilla Li moved beyond her standard tasks, re-framing the HAAA intern training documentation from several Google Docs into a single 51-page PowerPoint presentation. This was a significant improvement to the documentation, bringing together in one location steps from the entire process, from initial contact with interviewees, through the

creation of transcripts and time syncing, to the creation of interviewee pages on the haaa.rice.edu website. The program now has one training document which can be used and updated annually as needed. The document includes links to more detailed instructions which can be accessed as required.

HAAA intern training guide, by Priscilla Li.
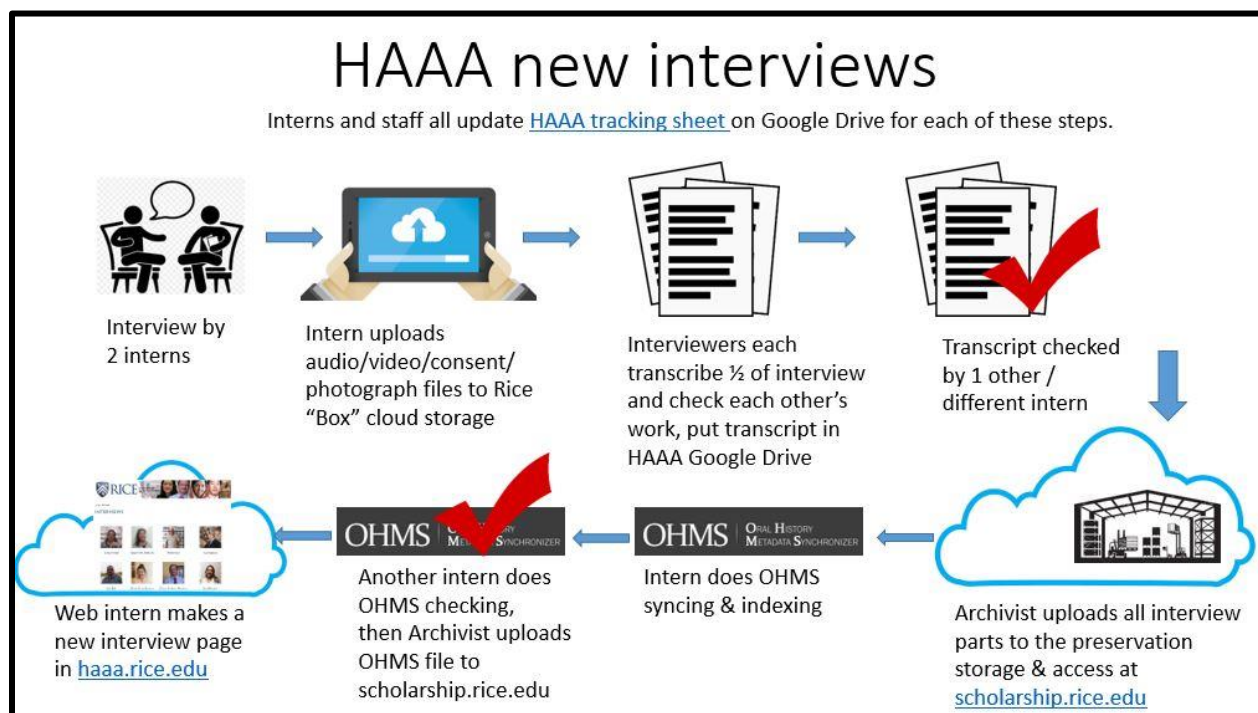
This pictographic overview explains the high level steps taken for each interview, and serves as an introduction to how the interviews are processed, where the steps happen, and who does them.

The main tracking spreadsheet for the interviews, located in the HAAA Google Drive, is the key tool for ensuring that each interview is handled in a consistent and timely manner in accordance with the steps in the HAAA program. Each interview has a dedicated row in the spreadsheet, with columns for the interviewee's name, the interview date, and the interview identifier number, as well as for tracking all of the steps shown in the high level pictograph above. Interns and the program archivist update the spreadsheet each time a step is completed, permitting subsequent participants to see where they should begin their work. For example, a student who has conducted a particular interview uploads the recordings, photographs and consent form to a folder in Rice University's "Box" cloud storage, including a link to that folder in the tracking spreadsheet. The interviewer creates the first draft of the transcript and updates the tracking sheet, with the draft transcript filed in the Google Drive, as per the training protocol. From there, fellow interns are able to see from the tracking sheet that the draft of the transcript is present, and that transcript checking can therefore begin. Once the transcript has been signed off in the tracking sheet as having been checked by interns who did not participate in the interview, the archivist can see that the interview is ready to be uploaded to the institutional repository. Participants continue working in this manner until all of the steps have been completed in relation to the interview: transcription; upload to public site; time-syncing and indexing via the OHMS; and, finally, construction of a dedicated page on the haaa.rice.edu website.

Once the interns have finalized the transcript, the archivist completes the final steps before uploading the files to the institutional repository. Descriptive metadata is written for each interview in accordance with the Dublin Core metadata schema. Required fields include: name of interviewee; names of interviewers; unique identifier number; date of interview; description;

length of interview; interview language; publisher; rights; subject; digitization specifications; and genre. Transcripts are converted to Adobe PDF format and have metadata embedded in the PDF, including the interview name, identifier number, copyright status and rights statement. This helps to maintain a sense of the context in which each interview was conducted, even after the transcript has been downloaded by users and is no longer situated in its original context as part of the online metadata. Guidelines for this metadata are available at Fondren Library's digital project wiki.

All of the digital parts of an interview, such as photographs, audio and video records, are allocated names that incorporate that interview's unique identifier as the root of their file names. For example, for interview wrc05143, the transcript, audio file, video file and consent form would respectively be named: wrc05143_transcript.pdf; wrc05143.mp3; wrc05143.mp4; and wrc05143_consent.pdf. Master files, such as the master audio file in WAV format, can be filed on the administrative side, and can be easily accessed as needed by library staff.

Once an interview has been successfully uploaded to scholarship.rice.edu, interns can begin working on time syncing and indexing using the Oral History Metadata Synchronizer (OHMS) tool. The OHMS tool requires that audio and video files be already accessible via a public URL, meaning that OHMS work cannot begin until the interviews have been uploaded. Interns begin by listening closely to the interviews in order to sync the audio or video to the transcript at one-minute time intervals. It is possible to use the OHMS even without a full or even partial transcript, and also to use it simply for its indexing functionality. However, the HAAA program has chosen to create and publish full transcripts, the texts of which are fully searchable using any web browser, in order to enhance the 'findability' of the interviews. Full transcripts in the OHMS allow researchers to search transcripts and click in order to jump to a particular location in a recording.

OHMS transcript view, with search box on the right, and audio time syncing to one-minute intervals on the left. (View this interview online)

Indexing in the OHMS allows researchers to browse high-level topics within each interview. Students define the topics based on the transcripts, and their indexing choices are reviewed for clarity and completeness by a fellow intern before the OHMS display goes live.

OHMS index view, with search box on the right, and high-level topics in the interview featured in the main window. (View this interview online)

A more creative and flexible web presence for the HAAA interviews at haaa.rice.edu is housed on an Omeka Classic platform hosted by Fondren Library's Digital Scholarship Services (Fondren DSS) Department. Omeka (https://omeka.org/) is an open source web publishing platform for sharing digital collections developed by the Roy Rosenzweig Center for History and New Media, George Mason University, and the Corporation for Digital Scholarship. Fondren Library's DSS administers the Omeka presence, which is hosted and maintained for version upgrades by Reclaim Hosting. A page is created for each interview, and includes a photo, a short bio, a link to the transcript, and the OHMS player. This Omeka site does not actually store the audio or video files for each interview; rather, it stores a photograph file for each interviewee and includes links to the OHMS file, which is found on the scholarship.rice.edu site. This means that the HAAA program only needs to maintain one set of files for each interview, but is able to present it on two sites for creative purposes. HAAA program interns and staff have creative control over the sites in relation to the pages and exhibits, as well as the fine tuning of site navigation. Requests must be made to Fondren DSS for changes, such as for the addition of plug-ins for new functionalities or modifications to the design theme.

Back-up environments for the systems which support the long-term preservation of and access to the HAAA interviews follow the "three golden rules" of IT, with three copies maintained in at least two storage formats, one of which is at a geographically separate location or offsite. All of the HAAA interviews are found on the scholarship.rice.edu site, which runs on a DSpace platform and provides nightly file-level fixity checks. One mirror copy of the content is saved

locally nightly, a second copy is stored in Amazon's low-cost Glacier storage at thirty-day intervals, and a third copy is stored in DuraCloud's relatively higher-cost, higher-service environment at six monthly intervals. These back-ups are maintained by Fondren Library's DSS and Central IT at Rice University. Copies of the interview files, which are initially uploaded by interns to the Box cloud environment, are also copied by the archivist to HAAA's 500 GB server on campus to fill the gap between interviews going online and them being backed up at thirty-day and sixty-day intervals by Glacier and DuraCloud respectively. Snapshots in time of the HAAA tracking sheet from the HAAA Google Drive are also filed on the HAAA 500GB server as a precaution. The Omeka-based haaa.rice.edu website is backed up by weekly crawls of the Archive-It tool, which copies it to the Wayback Machine (see https://web.archive.org/web/*/haaa.rice.edu), renowned for its system of creating four geographically diverse copies of web content. Because the Omeka site does not hold any of the actual interview files, this grab of the Omeka page text and images is sufficient.

While it would be possible to store and present the HAAA interviews on the Omeka-based haaa.rice.edu site only, this would require significant back-up procedures to be implemented. For example, if the files were found only on Omeka, the institution would need to explore two other environments for back-up purposes. The scholarship.rice.edu environment is rigorously backed up, while the Omeka haaa.rice.edu is less so.

In the early days of the program, interviewees often donated flyers, business cards, and other small ephemera at the time of their interviews. These materials would be described by the archivist in the archival finding aid and filed in the archival boxes, which also contained the transcripts and consent forms. As time went by, the number of donations began to grow, eventually requiring separate housing and description. As of 2018, there were almost twenty such archival collections, featuring items contributed by members of the Chinese-American, Filipino-American, and Zoroastrian communities in the Houston area. Highlights of these collections are listed in an online overview, each collection also having its own detailed archival collection guide written by the archivist. In many cases, interviews were conducted with the donors of these materials, and such items provide incredible context to each of their recorded life stories. For example, Glenda Joe's Houston Asian American Festival Association records include thousands of festival images recorded over many years, as well as items related to her community advocacy efforts regarding discrimination against Asians in Houston. The Chinese American Citizens Alliance (CACA) Houston Lodge records document the local activities of this national service organization via extensive photographs, ephemera, directories, and event programs. Joanna Po's nursing career awards collection connects closely with her interview, highlighting an example of one highly recognized Filipino-American nurse in Houston.

Anne Chao and I have attended and hosted many community events in order to encourage community members to participate in additional interviews and contribute archival materials. Such events have included a Chinese American Citizens Alliance group tour and library speaking engagements, a Filipino American library tour and luncheon, several Chinese Baptist Church presentations, and oral history training at the Zoroastrian Community Center, for

example. Such events provide important opportunities to build a sense of community and raise awareness of the program.

Challenges

With the complexity of procedures involved in the processing of these interviews, and with so many people involved on a rotating basis over the years, there are many challenges involved in maintaining consistency. Student interns may employ slightly different transcription styles from one other. Over the years, different program managers have provided different styles of intern training. At the outset of Dr. Chao's tenure as Program Manager, she launched a team initiative to comb through all of the existing transcripts, starting from the very beginning of the program and moving forward, checking that consistency in style and spelling had been maintained, ensuring that words had been transcribed correctly, and identifying other issues. There are cases where the transcriptionist struggled to understand the accent of the interviewee, or was simply unfamiliar with the words being used, and consequently included his or her best guesses at what was being said. Solutions for such problems include: employing a detailed transcript template; guiding transcriptionists to ask colleagues for help in understanding words; or, as a last resort for unintelligible words, using an ellipsis. The process of closely reviewing all of the circa 160 transcripts over a two-year period required a temporary hiatus in the creation of new interviews, but resulted in a significant improvement in the quality of the transcripts. Another benefit of this process has been an improvement in program documentation, since consistent training for all participants inspires uniform quality for subsequent interviews.

Over the years, with technological changes happening so quickly, the program has used different methods for the submission of all of the various parts of interviews to the archives. In the early days, CD-ROMs containing the audio files and transcripts were submitted in person, along with printed consent forms. Thumb drives have also been used. In more recent years, there has been a shift to the use of a strictly cloud-based storage system for the submission of all of the parts of the interviews to the archives. Additionally, one intern has been tasked with the checking of interviews for 'final completeness' before alerting the archivist by email that they are ready to be processed for upload. This system is more organized, more centrally accessible to all team members, and has greatly economized the archivist's time when working with the interviews.

Conclusion

Rice University is fortunate to have the resources to support the HAAA program, enabling it to fund interns, engage a part time director and a part-time archivist, benefit from the support of the university's IT staff, and have access to various software, hardware, and cloud vendor environments. Oral history programs come in many shapes and sizes. The environment at Rice that supports these interviews is a complex one, but we believe that it is the best system that we can achieve in order to ensure reliable long term preservation and access while at the same time maintaining creative flexibility.