

Data Curation Profile – History/Cultural Heritage

Profile Author	Bonnie Finn
Author's Institution	University of Tennessee, Knoxville
Contact	Bonnie Finn, Graduate Research Assistant School of Information Sciences University of Tennessee, Knoxville bfinn1@vols.utk.edu
Researcher (s) Interviewed	Professor of History
Researcher's Institution	Maryville College
Date of Creation	March 24, 2017
Data of Last Update	
Version of the Tool	1.0
Version of the Content	1.0
Discipline/Sub-Discipline	History/Cultural Heritage
Sources of Information	An interview conducted on March 20, 2017
Notes	The profile was completed prior to completion of oral history interviews, which are in progress. As this project is still in the working phase, the profile may be adapted as needed for the continuing project.
URL	
Licensing	Creative Commons Attribution-NonCommercial 4.0

Section 1 – Brief summary of data curation needs

The researcher is conducting oral history interviews of people who attended or worked at a Girl Scout camp in the Tennessee Smoky Mountains between 1929-1959. She hopes to offer the interviews and transcripts for public access as part of Smoky Mountain cultural heritage. She is open to ways to make them available. As she is still in the interview process and few interviews have been transcribed, the project is in the planning stages. Some of the existing transcripts have been done by Maryville College history interns. Back-ups for the data are being done regularly by the researcher and conversations with the Maryville

College IT department have briefly discussed later storage of the interviews on the college server.

Section 2 – Overview of the research

2.1 – Research area focus

The researcher's Camp Margaret Townsend Oral History Project consists of recorded interviews, transcripts, and donated photos of people who attended or worked at the Camp Margaret Townsend, a Girl Scout camp in the Great Smoky Mountains of Tennessee. It operated from 1929-1959. Its location is now part of the Great Smoky Mountain National Park. The CEO of Tremont Institute approached the researcher with twelve names and asked her to collect the oral stories. Since that time, more names have been added for a total of 36 people to contact. As this project is in its first stages and interviews are being gathered and transcribed, the researcher is currently looking for a long-term plan for the collection. Public access to the completed collection is a primary focus, thus tools that support this goal will be explored and considered. The data needs to be preserved indefinitely as cultural heritage. Currently, interns from Maryville College assist with the transcriptions. Storage will likely be at Maryville College on a server, however she is open to repository suggestions.

2.2 – Intended audiences

Potential audiences for this project cover several spectrums. Past and present Girl Scouts, especially those that attended the camp will be interested. Other audiences include those with a love for local East Tennessee history and the Smoky Mountains. Interested stakeholders of the final project include the audiences, as well as the Tremont Institute, Maryville College, and the Girl Scout Headquarters. Currently, the data is not being shared, as all the interviews are not complete.

2.3 – Funding Sources

One small grant was received a couple years ago to pay for recorders and microphones. However, the cost of the project is minimal and consists mostly of the researcher's time. There may be future storage costs, which will be explored at a later date.

Section 3 – Data kinds and stages

3.1 – Data narrative

The data lifecycle is in the beginning stages. In the first stage, a total of 36 interviews will be conducted, only 18 have been completed. The transcription process, stage two, currently has 9 completely transcribed interviews. This part of the process will take the longest, but alternatives are being examined for expediting public access. OHMS may offer an alternative for the audio interviews, while the transcriptions are being completed.

The third stage will also be a lengthy process. Organizing the files, scanning, photos, importing files to a system, and labeling them with subjects or keywords and other metadata will be time-consuming. Any photos donated by the interviewees will also be scanned. The data will be prepared for upload to an online host. Depending on the availability of interns, the researcher may complete most of the work.

The fourth stage will consist of making the data public. An online host will be located, possibly on the Maryville College website through the library. Other copies of the data will be given to the Tanasi Girl Scout Headquarters and Tremont Institute in Townsend, Tennessee. The data will be licensed under Creative Commons, most likely as Creative Commons Attribution-NonCommercial. Data will include MP3s, scanned photos, and documents, such as obituaries. Organizational tools are not fully conceived or developed at this time.

The fifth stage is not reliant on the fourth. The researcher hopes to publish a book with the collected stories of camp life, organized by themes, such as animal stories. The book will include photos of the most beloved names that occur within the interviews. Content will be about 150 pages, consisting of an Introduction, conclusion, and five chapters.

3.2 – The data table

Data Stage	Output	# of Files/Typical Size	Format	Other/Notes
Primary Data				
Collecting Names, Interviews	36 names	12-180.5mb	MP3	only 18 interviews have been conducted to date
Transcription, scanning, other documents	9	16-108 kb	doc, pdf, email, photographs	photos, copies of obituaries also included
Organization & Importing, marking times, metadata	n/a	n/a	MP3, pdf, tif, considering Soundcloud, OHMS (Oral History Metadata Synchronization)	This stage has not been defined yet and is still under consideration.
Making data public online and at Tanasi HQ	n/a	n/a	MP3, pdf, tif	Still in planning stages, but considering Creative Commons License
Book	n/a	150 pages	5-7 chapter book	collected stories of camp life, organized by theme, with pictures

3.3 – Target data for sharing

Prior to organization, the researcher would limit accessibility of the data to collaborators and Maryville College. She would also accept assistance during the third organizational stage from others in her field. Once the data is collected, organized, and published, she would make it available to anyone interested in the history.

3.4 – Value of the data

The data would be valuable for the regional Girl Scout volunteers and administrators, as it celebrates a point in their history that is important to them. Local historians, especially those interested in the Smoky Mountains would also place a high value on the data. It also highlights the importance of oral histories for cultural heritage.

3.5 – Contextual narrative

Since the project is still in the initial stage, specific information regarding size and format of most of the data is not yet available. In its raw form, it would be difficult to make the data available to the public, so the organizational stage will be a vital part of the project. Once tools are identified and tested, more information about the organization of the data will be forthcoming.

Section 4 – Intellectual property context and information

4.1 – Data owner(s)

The primary owner of the data is the researcher, as the interview subjects have given permission for their stories to be collected and published by the researcher.

4.2 – Stakeholders

Stakeholders include Girl Scout Tanasi Headquarters, Maryville College, Smoky Mountain National Park, Tremont Institute, past and present Girl Scouts, and anyone with an interest in local history.

4.3 – Terms of use (conditions for access and (re) use)

Consent forms have been signed and the interviewees have been informed that their information will be freely shared. They were advised to withhold information they wished to keep private. The final data is to be shared with the public.

4.4 – Attribution

When others use the data in research, the researcher should receive citation credit. She also wants the ability to cite the data in her own publications.

Section 5 – Organization and description of data (incl. metadata)

5.1 – Overview of data organization and description (metadata)

At this point, the data is stored in files on the researcher's computer, her cloud, and an external hard drive in its raw form. However, tools for oral histories will be examined to determine if metadata can be added at various points of the interviews. It is also a high priority to make the data available in multiple formats.

5.2 – Formal standards used

Since the ability to make the data accessible is important, the researcher will check the Smithsonian and other cultural heritage sites for metadata terminology.

5.3 – Locally developed standards

None.

5.4 – Crosswalks

No crosswalks are needed.

5.5 – Documentation of data organization/description

Documentation of data organization is not in place at this time.

Section 6 – Ingest/Transfer

Although the data has not yet been deposited in a repository, the researcher is willing to consider it once the interviews are done and the data is organized. She and the IT director of Maryville College have also briefly discussed storing the data on the Maryville College server, with access through the college's library website. However, until the third stage is complete, she does not want the data to be transferred to the server or ingested into a repository.

Section 7 – Sharing and Access

7.1 – Willingness/Motivations to share

The researcher is willing to share all data stages with her immediate collaborators. After the initial stage, she would share with others at Maryville College. Beginning with the third stage, the researcher would open the data to others in her field. After completion of the fourth organizational stage, data would be shared with all other interested people.

7.2 – Embargo

No embargo needed.

7.3 – Access control

The purpose of this project is to share the data with interested people. Thus, few restrictions will be placed on access. However, it is a high priority for her to be able to cite the data in her publications and for others to cite the data if they use it in their own research.

7.4 – Secondary (Mirror) site

A secondary mirror site for off-line access is not a priority.

Section 8 – Discovery

The researcher envisions others finding her organized data through search engines, or links from related websites such as the Girl Scout Tanasi Council, or Tremont Institute. It is a high priority to be easily discoverable for the public in general and by using internet search engines. Although she sees that her data could be valuable for researchers, it is a lower priority to her. Thus, a link located on related websites or at Maryville College makes the most sense, while the actual location of the data is not as important.

Section 9 – Tools

To generate the data, the researcher used a digital voice recorder for in-person interviews. When phone interviews were necessary, NoNotes, a smartphone app for call recording was used. A scanning app for smartphones was used to collect physical documents. Once the data was collected, a computer, external hard drive, and scanner were all utilized. Software used with the current collection consists of Microsoft Word and Excel. In the future, sound editing software, such as Audacity, will be used. The researcher will also look into the use of OHMS (Oral History Metadata Synchronization tool), by the University of Kentucky. This tool enhances access to oral history and allows creation and indexing of metadata in the player.

Section 10 – Linking/Interoperability

The ability to merge with other data or connect to publications is a low priority. The only publications to possibly consider would be the Girl Scout Magazine, or the Smoky Mountain National Parks publications. However, in the future, it would be good for the tools used to access the data to support the use of web services APIs, which would make the data more discoverable.

Section 11 – Measuring Impact

11.1 – Usage statistics and other identified metrics

Not a priority.

11.2 – Gathering information about users

Although it is a low priority, the researcher thought it would be nice to gather information about those who access the data, to see who might be using or viewing it.

Section 12 – Data Management

12.1 – Security/Back-ups

Currently, interviews, transcripts, and digital images are saved on a laptop and backed up on an external hard drive after each new interview or every 10 days. Cloud storage is also being considered for raw files. Security measures are not necessary, as the data is not confidential.

12.2 – Secondary storage sites

The completed project should be backed up and stored in a different geographical location, to protect loss from natural disasters, or other damaging possibilities. Curation to maintain version control for the final data is also under consideration.

12.3 – Version control

Version control is a high priority for the researcher. It is important to keep the data accessible for preservation of cultural heritage.

Section 13 – Preservation

13.1 – Duration of preservation

As part of the Smoky Mountain and Girl Scout cultural heritage, the data should be preserved indefinitely. The most important parts of the data to manage and preserve are the recordings, transcripts in edited forms, and the published book.

13.2 – Data provenance

Documentation of all changes made over time to the data should be tracked. This includes keeping copies of the original files before alteration and editing.

13.3 – Data audits

Data audits were not discussed.

13.4 – Format migration

Structural integrity is a high priority, thus the ability to migrate the data into accessible formats is vital. Currently, this is not an issue, as the data is in MP3, doc, and pdf formats. However, moving forward, it will be necessary to consider the available formats.

Section 14 – Personnel

14.1 – Primary data contact (data author or designate)

Data client and contact info

Nancy Locklin-Sofer, Professor of History

Maryville College, Maryville, Tennessee

14.2 – Data steward (ex. Library/archive personnel)

Not currently assigned

14.3 – Campus IT contact

John Berry, Director of IT,

Maryville College, Maryville, Tennessee

14.4 – Other contacts

n/a