


Data Curation Profile – History / Sustainable Development

Profile Author	Megan Sapp Nelson / Katherine Beavis
Author's Institution	Purdue University
Contact	Megan Sapp Nelson (msn@purdue.edu)
Researcher(s) Interviewed	<Withheld>
Researcher's Institution	Purdue University
Date of Creation	02/15/2013
Date of Last Update	
Version of the Tool	1.0
Version of the Content	1.0
Discipline / Sub-Discipline	History / Sustainable Development
Sources of Information	An interview conducted on November 4, 2011 (duration of 2 hours).
Notes	
URL	http://docs.lib.purdue.edu/dcp
Licensing	Attribution-NonCommercial-ShareAlike 3.0 Unported 

Section 1 - Brief summary of data curation needs

The researcher's Uganda Oral History Project is centered on conducting interviews from the Nnindye community. The researcher believes that these interviews will be of interest to other researchers, practitioners, and the general public. However, the researcher does yet know what the best means are for making the data accessible. Because of the historical component of her research, she needs her data to be preserved indefinitely. During the data collection process, the researcher places high priority on the ability to back up her data. Given that electricity and internet are not always readily available, the researcher needs to have a backup method that is reliable and allows her co-collaborators in the United States to access the data.

Additionally, the researcher expressed needs related to hiring individuals within Uganda to transcribe and translate the interviews. She stated that in the United States, Luganda translators are scarce and expensive. Also, because the researcher is unaware of existing metadata standards for oral history and is unfamiliar with how to apply the coding to her data, the researcher may need to employ a graduate student who is knowledgeable of applying codes to aid in this process.

Section 2 - Overview of the research

2.1 - Research area focus

The researcher is working on a project to establish a longitudinal data set that will look at the long term impact of sustainable development in the Nnindye community located in the Mpigi District in Uganda. She will conduct oral history interviews with community members and gather data on cultural, medical, medicinal, and agricultural practices. She plans to use the data to gauge the success of development projects and to gain a better understanding of cultural change as it occurs. At the time of the interview, this project was still in the planning phase.

2.2 - Intended audiences

The researcher indicated that the intended audience includes a variety of people. Specifically, she listed people within and outside of her field, such as development practitioners, historians of development, and anthropologists of development; undergraduate and graduate students at universities who are in similar fields; and the communities with whom she conducted the interviews. Additionally, she mentioned that policymakers and development practitioners may find the data useful.

2.3 - Funding sources

The researcher's project is funded by a Wenner-Gren anthropology grant. She also is looking into other sources of funding such as a National Endowment for the Humanities (NEH) collaborative research grant, a Clinical and Translational Sciences Institute (CTSI) grant, or an American Council of Learned Societies (ACLS) collaborative research fellowship. Because the researcher has not yet applied for most of these grants, she is unfamiliar with each agency's requirements for data management.

Section 3 - Data kinds and stages

3.1 - Data narrative

The researcher plans to collect her data in Summer 2013 in the Nnindye community located in the Mpigi District in Uganda. She will begin with a pilot study, collecting the data from a preliminary sample.

The first stage in the data lifecycle involves collecting the raw data. The data will consist of recorded oral interviews that the researcher will conduct. Photos and videos also will be included as a part of the raw data. The interviews will be conducted in the Luganda language. Interviews will be recorded using digital recorders. Photographs and videos will be taken with digital cameras and digital video cameras. The researcher anticipates conducting 20 to 25 interviews for the pilot study. She anticipates that the audio and video recording and photographs will be saved in common standardized formats (e.g., WMV, JPEG).

In the second stage of the data lifecycle, the interviews will be transcribed and then translated. Stage Two will need to occur in Uganda due to the high cost of getting Luganda translations outside of Uganda. She plans to recruit undergraduate or graduate students from a Ugandan university to do the transcriptions and translations. The transcripts and translations will be saved as Word documents. Ideally, the researcher hopes to have the interviews first transcribed into Luganda and then translated. She worries that if the interviews are only translated and then transcribed, information could be lost in translation and she would not be able to refer back to a Luganda version. Thus, there will be two Word files for each interview: one in the original Luganda and the other translated to English.

In the third stage of the data lifecycle, the researcher and the other principal investigators (PIs) will code the translated interviews. She is considering using ethnographic software to code them.

Themes will be identified within and between the interviews. At this stage, new files containing the coding would be created for each file and a master file would also be created.

The researcher is unsure of what would occur in the fourth stage of the data lifecycle. She sees the fourth stage as the point when she would attach metadata files to the transcripts in order to make the data searchable. However, it was noted that after this is complete, she would begin to produce derivative works.

3.2 – The data table

Data Stage	Output	# of Files / Typical Size	Format	Other / Notes
Primary Data				
Raw	Recorded interviews; photographs; videos	250 files / unknown	wmv; jpeg	Each interview may contain multiple files as the researcher anticipates technological failure (e.g., batteries running out) and needing to stop and start a new recording.
Processed	Transcribed interview in Luganda; translated transcription into English	~50 files / unknown	.doc or other text file	
Analyzed	Coded transcriptions	~25 files / unknown	.doc or text file	The researcher is unsure what software will be used to code the interviews
Finalized				As the project has not yet begun, the researcher was uncertain of what specific activities would need to be performed to finalize the data.

Note: The data specifically designated by the scientist to make publicly available are indicated by the rows shaded in gray (the “processed” row is shaded here as an example). Empty cells represent cases in which information was not collected or the scientist could not provide a response.

3.3. - Target data for sharing

The raw and processed data will be shared with the researcher’s immediate collaborators. After publication, the researcher wants to share her data in any of its various forms with anyone. (She noted that the other two PIs would need to also agree with this before sharing the data.) Specifically, the researcher would like the original recordings to be available to the interviewed community members and their families. Although there may not be immediate interest, she believes that in the future (e.g., 20 years from now), community members may have interest in the recordings as historical documents or to hear what their relatives said. She believes it is preferable to have the recordings available, as opposed to the transcripts, because many people in the community are illiterate and thus the transcripts would be of little interest. The researcher acknowledges that there may be some confidentiality considerations that could limit the sharing of these recordings.

The researcher also believes that the original recordings in Luganda might be of more immediate use to students and faculty at the university, individuals who are doing agricultural development-related work, and researchers or workers within the community outreach program. She would also want the Luganda transcripts available to these people. The researcher believes that students and faculty at Ugandan universities also might find the transcripts useful.

The researcher would like to share the English versions of the transcripts and the coding files with the above groups of people. In addition, she believes that policy makers, development practitioners and scholars (i.e., historians of development and anthropologists of development), and undergraduate students in development programs would have an interest in the data.

3.4 - Value of the data

The researcher identified several potential uses for the data. Personally, the researcher wants to use the data to examine the long term impact of sustainable development in the specific Ugandan community.

The researcher believes that her data could be useful to university students for the purposes of using the data as case studies and analyzing it themselves. Thus, it could be used as a means of teaching students to do similar work. Additionally, the analyzed data could be useful to policy makers. She can also see her data being archived as be part of an oral history project, such that any person could read about it (e.g., “history buffs”). The researcher also suggested that in the future, the data could also be used as a way of legitimizing oral history as an important medium for collecting history.

3.5 - Contextual narrative

One important consideration for the researcher in terms of the researcher's workflow is the ability to find individuals who can transcribe the Luganda interviews. According to the researcher, having the interviews transcribed in the US would be very expensive. She hopes to find undergraduate and graduate students at one of the universities in Uganda to transcribe and translate the interviews.

Section 4 - Intellectual property context and information

4.1 - Data owner(s)

The owners of the data are the researcher and the other two principle investigators. The researcher did not provide additional information about her collaborators.

4.2 – Stakeholders

The Wenner-Gren Foundation is a stake holder for the data. If the researcher received the NEH grant, the NEH would be a stakeholder.

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

The researcher is unsure of funding agency requirements. However, she does note that there are confidentiality concerns regarding participants who wish their identities to remain anonymous or confidential. The procedures regarding confidentiality will be addressed in her IRB application.

4.4 - Attribution

Attribution was not discussed with the researcher.

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

5.1 - Overview of data organization and description (metadata)

The researcher plans to use detailed annotations to go along with the transcripts. The annotations could be helpful for people looking through the various interviews, such that they

would be able to read a summary of discussed issues to decide if they wanted to proceed to the full transcript. For her and her collaborators' use, she would also need a codebook. She also plans to use column headings to organize information about the interviews, such as date, time, weather, and age of interviewee. The researcher did not describe any particular formats for this data.

5.2 - Formal standards used

The researcher is not currently using any formal metadata standards. The researcher was not familiar with the standards available for oral history research. The interviewer shared two examples of formal metadata standards (e.g., TEI) and the researcher was interested in learning more about them and applying them to her research.

5.3 - Locally developed standards

The researcher does not have any established local standards.

5.4 - Crosswalks

Crosswalks were not discussed.

5.5 - Documentation of data organization/description

The researcher does not seem to have any specific documentation practices in place regarding the description or organization of the data.

Section 6 - Ingest / Transfer

Ingest and transfer were not discussed.

Section 7 – Sharing & Access

7.1 - Willingness / Motivations to share

The researcher is enthusiastic about sharing her data with others to make use of it themselves because she views the data as having multiple uses and purposes. She is open to sharing data from any of the stages once she has published her initial findings.

7.2 - Embargo

The researcher estimates that she would need a period of approximately three to seven years before publicly releasing the data so that she would have time to write and publish her own findings first.

7.3 - Access control

The researcher stated that she would put limitations on videos and photographs without permissions due to confidentiality issues. However, there are no groups that she would exclude from accessing the rest of the data set.

7.4 Secondary (Mirror) site

The use of a secondary storage site for the annotated data set is a low priority for the researcher so long as it is kept in a trusted digital repository.

Section 8 - Discovery

The researcher desires that her data be discoverable to other researchers and professionals as well as to individuals who may have a general interest in the data. She stated that she would be interested in having her data be part of a broad oral history project through which anyone could access the data. However, how individuals would actually discover the data was not discussed.

Section 9 - Tools

The tools needed for the researcher to carry out the project include laptops, digital recorders, digital cameras, digital video cameras, transcription software, an external hard drive, and batteries. For other people to use the data, they would need internet access, Microsoft Excel, Microsoft Word (or other compatible office package), and a media viewer for pictures and video. Additionally, the researcher sees it as important to be able to link the data set to visualization and analytical tools as important. The researcher does not want to allow the ability to comment on or annotate her data set.

Section 10 – Linking / Interoperability

Linking and interoperability were not discussed with the researcher.

Section 11 - Measuring Impact

11.1 - Usage statistics & other identified metrics

The researcher believes that it could be important to know the usage statistics of how many people have accessed her data.

11.2 - Gathering information about users

The researcher rates the ability to gather information about the individuals who are accessing or using the data as not a priority.

Section 12 – Data Management

12.1 - Security / Back-ups

The researcher plans to initially download the data from the recording equipment onto a “master” computer and an external hard drive. She stated that her other PIs will also have these master computers on which the data would also be saved. She also plans to burn the data to CDs or DVDs so that the data is saved in multiple places. The researcher plans to back up her data on a daily basis.

12.2 - Secondary storage sites

The researcher desires a secondary storage site for the data set in a different geographic location, that is, one in Uganda and one in the US. The storage site in the US would allow the researcher and her co-collaborators to both view and work on the data simultaneously.

12.3 - Version control

The researcher plans to use version control for the transcripts of her interviews and coding in terms of the changes she and her co-PIs make to the files.

Section 13 - Preservation

13.1 - Duration of preservation

The researcher views that the most important aspects of her data to preserve are the recorded interviews, photos and videos, the Luganda transcripts, and the English translations. She desires that the data be preserved indefinitely.

13.2 - Data provenance

Data provenance was not discussed.

13.3 - Data audits

The researcher rated the ability to audit the data set to ensure its structural integrity over time as being high priority. Because the data is based on original interviews, it is important that the content stay the same over time.

13.4 - Format migration

The researcher considers data migration to be a medium priority. She has some concerns with individuals in Uganda being able to access the data because there are typically delays in technological advances compared to the United States.

Section 14 – Personnel

Not used in this profile.