

4.2.2 Archiving: plan for it now

A major difference between the newer methods in language documentation and those that preceded them is the use of digital media. As pointed out above, digital data has no proven longevity, and format changes can potentially lead to orphaned data that can no longer be read by available software. Therefore, the sooner your data can be stored in a dedicated repository, with trained staff and an institutional commitment to protect and migrate your data, the better. A proper archive will describe your data adequately, assign persistent identifiers, and share data in several locations to ensure its long-term accessibility. Archiving is an essential part of any language documentation work in which a major motivation is to create an enduring record of the language. Unfortunately, your computer is not an archive. An external hard drive, as crucial as it is for day-to-day backup, is not an archive, nor is a website from which you make multimedia files available for download.

Many people think of archiving as the last step in a research program, to be done years after the fieldwork is complete, after everything has been transcribed and mined in service of analysis—essentially, something to be done after we are ‘finished’ with the data. Here, we advocate the exact opposite: data should be archived immediately and often (subject, of course, to your having resolved any ethical issues that may have arisen about what you recorded, and whether the speakers have agreed to have their recordings archived—see Rice, Chapter 18 below, for further discussion of ethical issues).

It is quite possible—and becoming increasingly common—to archive recordings periodically from the field as they are collected. Then, whenever you finish a transcription, archive it. When you write up an analysis, archive it. As your lexical database grows, archive it. If you make a children's book from your data, archive it. Think of your archived collection as another backup of your entire corpus, only this one is stored in a remote offsite location and maintained by professionals. If the idea of sending your data to another location before it is ready to be made available to the general public makes you uncomfortable, you can temporarily restrict access to certain parts of your collection.¹³

Establishing a relationship with an archive before you leave for the field ensures that the staff can work with you from the beginning of your project to secure the long-term preservation of your collection. Finding out what kind of metadata and file naming conventions the archive uses is a good idea, because this will guide how you set up your metadata database. Be sure to keep track of the information the archive requires, as well as any extra information that you will find helpful that is not necessarily requested by the archive. You can also make some arrangements for regular deposits of your data from the field, either via direct upload to the archive or by mailing DVDs or a small hard drive.