

## **Report of the DELAMAN Costing Case Study**

Third INNET Conference

September 5-6, 2014

Hungarian Academy of Sciences, Budapest

### **Executive Summary**

DELAMAN member archives estimated the costs of archiving two sample deposits of language documentation data. Comparing these costs to the cost of funding a documentation project which would generate this amount of data, the costs of archiving can be estimated to 8% of the total direct costs of the funded project. DELAMAN proposes that grantees and grantors use this 8% figure as a more simplified way to calculate archiving costs which better reflect the nature of archiving as basic infrastructure for endangered language research.

### **Background**

Since its founding in 2003 the Digital Endangered Languages and Musics Archiving Network (DELAMAN) has been the premier professional body overseeing the development of standards for digital language and ethnographic archives. Its governing board represents a diverse range of archives from across several continents who share a common commitment to best practices in digital language archiving. In September 2014 DELAMAN board member met under the auspices of the Innovative Networking for Infrastructure for Endangered Languages (INNET), a project funded by the Seventh Framework Program of the European Commission, to discuss costing and sustainable funding of endangered language archives. Member archives estimated archiving costs for specific case studies using their individual cost models. The results of these costing case studies were compared to develop an overall recommendation reflect the cost of archiving endangered languages data.

### **Archive Costing**

There are real costs associated long-term archiving of digital language data; however, quantifying these costs is notoriously difficult. Individual archive deposits vary greatly in format and organization, making it difficult to estimate the cost of archiving a particular item. Documentary linguistics produces a diverse range of data; there is no “typical” archive deposit. Also, archiving activities are usually embedded within larger institutional responsibilities, and it can be difficult to tease out particular costs from these larger institutional commitments.

DELAMAN archives were asked to consider the costs of ingesting and preserving two fictitious archival deposits: one consisting of analog data and the other consisting of digital data. Depending on the chosen cost model, these costs could be split between recurring and non-recurring costs (Some archives choose to calculate long-term storage costs on an annual basis whereas others calculate a fixed sum up front and assume storage prices to drop to a negligible amount within a certain number of years).

For the purposes of these case studies the collections can be assumed to be well-formed; that is, the depositor has read the requirements of the archive and subscribes to ARCHIVISM, the notion that preparing a collection for archiving is a necessary activity for the depositor to engage in.

Collection 1: Analog

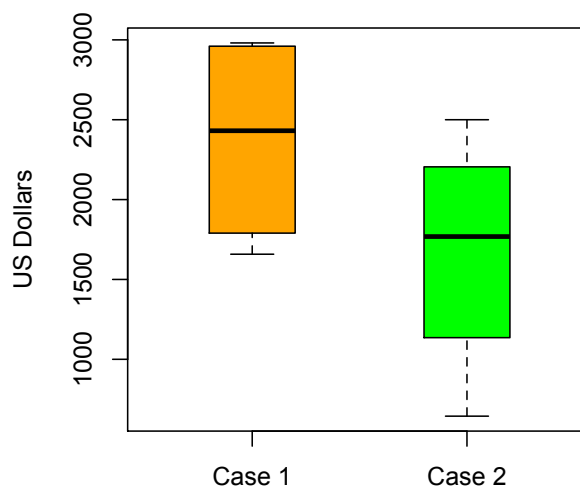
20 audio-cassette tapes in good condition with handwritten transcripts of 5 of them. 200 photographs (on paper) with a list of their contents (on paper). A spreadsheet listing the contents of the audio tapes.

Collection 2: Digital

150 digital audio files (uncompressed WAV, 16bit/48 kHz, average length 15 minutes), 20 video files (mp4/H.264, HD, average length 30 minutes) , 250 digital images, a spreadsheet listing of the contents of all of these. EAF transcripts of all the video and 100 of the audio files with names that match their source file. Total size of the collection: 150 GB.

There is significant variation in costs among the various DELAMAN archives (Figure 1). This variation can largely be accounted for in terms of the different funding models employed by the archives. For example, some archives have access to digital storage at no cost through their institutional affiliations, while other archives must budget for storage costs as a line item. Similarly, some archives can bundle some labor costs as part of ongoing expenses, while others must budget all labor costs associated with ingest and digitization. However, in spite of this variation there is a large overlap between the case studies in the neighborhood of \$2000. Thus, this number provides a good rough estimate for the cost of archiving collections of this size.

Figure 1: Range of costs across DELAMAN archives for each case study. Colored bars indicate range of one standard deviation from the mean.



Not all documentation projects will produce equal amounts of digital data. The amount of data in the case studies might be typical of that produced by a twelve-month DEL Fellowship award, whereas a typical DEL grant with an overall budget of \$75,000 might produce twice as much

data, compiled over two years. We can use these figures to calculate the cost of archiving based on a percentage of the grant award amount. If we assume a 50% indirect cost rate, then a grant award of \$75,000 has a total direct cost of \$50,000. Cost of archiving the documentation produced by this award would be twice that of the case studies, or \$4000. Thus, the cost of archiving data produced by this typical DEL grant is 8% of the direct costs.

### Recommendation

Our recommendation is that archiving costs for documentation be calculated as a percentage of direct costs. Further, we suggest a figure of 8% of direct costs as a reasonable estimate based on current data collection rates. Calculating archiving costs as a percentage has several advantages:

- Using a percentage is much more straightforward for the grantee, since it is often impossible to know in advance just how much data will be produced by a project. Although many proposals do include estimates of number of hours of recordings and/or transcriptions to be produced, these are often highly speculative and bear little relation to the actual project outcome.
- Using a percentage simplifies the budgeting process for the grantor, since the cost of archiving is known in advance, based on the number of dollars allocated to documentation projects.
- Using a percentage is easier for the receiving archive, since it allows the archive to budget for ongoing, recurring costs. Although individual grant projects come and go, grant dollars awarded for language documentation in a particular region remain relatively steady. Knowing these figures, regional language archives can better budget for anticipated income from documentation grants.
- Using a percentage better reflects the nature of archiving as infrastructure. Just as physical infrastructure is funded indirectly through facilities and administration costs as a percentage of direct cost, archiving can be similar budgeted as a simple percentage.

It should be stressed that while the case studies used in this analysis focus on ingestion of materials, the real costs of archiving are ongoing and recurring. For example, a language archive must maintain minimal staffing regardless of whether they are currently in the process of ingesting materials or not. Although additional temporary staff can be brought on to assist with specific projects, the core staff must be continuously funded. Similarly, storage costs persist indefinitely. Some archives included an estimate for long-term storage costs in their cost model; this was achieved by front-loading the costs for a fixed period of 5-20 years. No archive directly budgeted storage costs beyond this time horizon. Notably, archives did not amortize fixed, recurring costs other than digital storage.

The percentage model proposed here recognizes that the costs of language archiving are not always directly related to the amount or format of the materials to be archived. There are many hidden costs which are very difficult to estimate in archive cost models. By costing archiving based on a percentage of direct costs in the grant budget we can ensure the long-term maintenance of archival infrastructure for endangered languages.



## Links

- DELAMAN, [delaman.org](http://delaman.org)
- INNET, [innet-project.eu](http://innet-project.eu)