

Language Documentation and Description

ISSN 1740-6234

This article appears in: *Language Documentation and Description, vol 12: Special Issue on Language Documentation and Archiving*. Editors: David Nathan & Peter K. Austin

Archiving for the community: Engaging local archives in language documentation projects

JOSHUA WILBUR

Cite this article: Joshua Wilbur (2014). Archiving for the community: Engaging local archives in language documentation projects. In David Nathan & Peter K. Austin (eds) *Language Documentation and Description, vol 12: Special Issue on Language Documentation and Archiving*. London: SOAS. pp. 85-102

Link to this article: <http://www.elpublishing.org/PID/139>

This electronic version first published: July 2014



This article is published under a Creative Commons License CC-BY-NC (Attribution-NonCommercial). The licence permits users to use, reproduce, disseminate or display the article provided that the author is attributed as the original creator and that the reuse is restricted to non-commercial purposes i.e. research or educational use. See <http://creativecommons.org/licenses/by-nc/4.0/>

EL Publishing

For more EL Publishing articles and services:

Website:	http://www.elpublishing.org
Terms of use:	http://www.elpublishing.org/terms
Submissions:	http://www.elpublishing.org/submissions

Archiving for the community: Engaging local archives in language documentation projects

Joshua Wilbur

University of Freiburg

1. Introduction

An increased awareness of documentary linguists' ethical responsibilities towards the communities where they work (cf. e.g. Himmelmann 1998, Grinevald 2003, Austin and Grenoble 2007, Czaykowska-Higgins 2009, Macri 2010, Rice 2010), particularly where languages are endangered, has heightened the importance of providing access to linguistic materials not only to the international academic community, but also to members of the language communities themselves. A local archiving institution (such as a local library or museum) can play a vital role in ensuring such access both now and in the future because it can potentially overcome certain technical and language-based hurdles encountered by some language communities. However, a number of challenges exist in archiving locally due to lack of technical infrastructure and experience, and limited prioritization of language materials.

In the following discussion, I describe my own experiences gained in working with local archiving institutions in northern Sweden while carrying out the Pite Saami Documentation Project.¹ First, this should provide an impression of how a variety of local institutions can serve as repositories of digital language materials. Second, these experiences will establish a foundation to discuss some practical aspects of fulfilling the ethical responsibility to provide documentation materials to local communities. In doing so, I outline ideas about the implementation of this for other (endangered) language documentation projects.

The paper is structured as follows: Section 2 provides a theoretical background concerning why archiving locally is important, Section 3

¹ The Pite Saami Documentation Project (saami.uni-freiburg.de/psdp/) was funded by the Endangered Languages Documentation Programme (ELDP, part of the Hans Rausing Endangered Languages Project) from 2008 through 2011, and again from 2013 through 2014. Some project materials are available from the Endangered Languages Archive (ELAR) at elar.soas.ac.uk/deposit/0053 [accessed 2013-08-12]

describes details of my own experiences in the field, and Section 4 summarizes the situation and presents factors to be considered for linguists planning to archive materials locally.

2. Background

The relatively recent rise in awareness of the urgent need for documenting endangered languages has been accompanied by a change in attitude of researchers towards the relevant speech communities. In somewhat exaggerated terms, a shift has been taking place from considering speakers to be objects of linguistic study for the scientific community,² to respecting members of a speech community as human beings and actively including them in the process of documenting their language.³ Indeed, Himmelmann (1998: 172) even insists that ‘the interests and rights of contributors and the speech community should take precedence over scientific interests’. Rice (2010: 27) indicates that this shift has in fact occurred, at least for most researchers: ‘The notion that the linguist has responsibility to the community has become a fundamental principle of research’.

One manifestation of this shift can be seen in the redefinition of the target audience of language documentation so that it not only includes researchers from various academic disciplines, but also considers speech community members as equally valid users of (endangered) language documentation, albeit with different needs and expectations than linguists and other academics. As Austin and Grenoble (2007: 16) point out, ‘the very design of the documentation project needs to take into account ... the ways in which they [stakeholders] will or will not be able to access the corpus’.

² Cf. for instance Samarin’s 1967 guide to linguistics fieldwork in which the author does acknowledge that to ‘some degree the linguist is obliged to his helpers [speakers] to meet their expectations’ (p. 16), but concludes that it is ‘a matter of personal decision’ (p. 17) how or if this happens. Aside from this paragraph, no mention is made of supporting language communities or including them in documentation or description efforts.

³ Grinevald (2003: 70) argues for an approach which gives ‘the speakers their rightful central place’ in endangered language documentation. More recently, this was further illustrated by the inclusion of the word ‘values’ equally with ‘practice’ in the title of Grenoble and Furbee 2010, not to mention chapters dealing with the linguist’s responsibilities to the community (Rice 2010) and ethics in language documentation (Macri 2010). See also Czaykowska-Higgins (2009) for a thorough presentation and discussion of approaches to cooperation between linguists and language community members.

In other words, a (documentary) linguist's obligation to support the community extends to providing the language community with the resulting language documentation materials. Some attempts in this direction have been made in recent years. For instance, the DoBeS⁴ initiative created regional archives to, among other things, make 'the speech community and the local linguists more involved and more interested in the archive' by having the data 'stored nearby' in relative terms.⁵ Projects funded by the Hans Rausing Endangered Languages Project are required to go a step further; their funding application advice recommends that language documentation materials be 'accessible to and usable by members of the language community' (ELDP 2013: 4), and applicants are thus strongly encouraged to 'deposit materials with an appropriate local body such as an archive, library, or museum' (ELDP 2013: 7).

Contemporary language documentation and archiving is primarily digital. One significant advantage of digital archiving is that it can allow interested individuals who are not physically present at the archive to access digital data via the internet. Indeed, as Nathan (2013: 25) points out, a 'digital documentation archive has to be more than a data repository. It has to find ways to preserve diverse materials *and disseminate* (or publish) *them to a variety of stakeholders* [my emphasis]'. With this in mind, I use the term 'archiving' in the present discussion not only to refer to the preservation of language materials, but also to highlight the process of accessing them; indeed, in the present discussion, my emphasis is on access.

Global access to language materials initially seems restricted only by the extent of access to the internet; wherever there is internet, there can be access. However, there are three particularly important prerequisites for effective access:

- a) technical infrastructure
- b) users' technical skills
- c) the access interface language and users' skills in it

⁴ Dokumentation bedrohter Sprachen, funded by the Volkswagen Foundation.

⁵ See dobes.mpi.nl/archive_info/regional_archives [accessed 2013-08-12]

With this in mind, it is useful to divide potential users into two groups: members of the international academic community, and members of the local language community.⁶ Most members of the former group have high-speed internet access from their home institutions which allow them to conveniently access a wealth of digitized linguistic data.⁷ Furthermore, many academics have at least reading knowledge of English, which is currently required in order to understand most international archives' websites. However, the same generalizations concerning access to digital archives cannot be made for the latter group. The lack of a (state-of-the-art) telecommunications infrastructure in many countries in which (endangered) language communities are located makes access difficult or completely rules it out (but see also Garrett's contribution to this volume). However, even in well-networked communities such as the one at my own field site in Swedish Lapland, further barriers to access by members of the language community include a lack of relevant computer skills and insufficient English language skills.

Several large web-based archives are aware of the obstacle that a lack of language skills can create, and attempts are planned or already being made to reduce this barrier. For instance, the website providing access to the DoBeS archive⁸ includes localizations in Bahasa Indonesian, Spanish, French, Portuguese and Russian. The search interface, menu items and other standard interface text is in these languages; however, the metadata targeted by the search engine are not necessarily in these languages. As an example: a search for 'ours' (French for 'bear') in the French localization, or for 'медведь' (Russian for 'bear') in the Russian localization, do not return any hits, while a search for the English word 'bear' in any of the interfaces returns 37 hits. For such localizations to really be useful to local communities, more than just navigational texts in the interface must be in these languages. One solution is to have individual depositors/projects include metadata in the relevant language of wider communication or even in the documented language (assuming an orthography is available), and include these in the searchable

⁶ Other user groups may also exist, such as the general public; however, the two groups mentioned here are typically the main target audiences for language documentation.

⁷ While digital archiving and web-based access often go hand-in-hand, they are not necessarily co-existent. For instance, the Leipzig Endangered Languages Archive (LELA) indicates that one of its 'primary function[s]' is the 'dissemination of those materials to members' of the speaker communities; however, there are 'no plans at present for creating an open web-based resource dissemination system' (LELA 2013).

⁸ See dobes.mpi.nl/

data.⁹ Clearly, this would require a large effort, and would be particularly dependent on the field worker or other contact with the local language community for each deposit.¹⁰

It is precisely as a result of the lack of a technological infrastructure as well as computer and language skills within communities that local archiving institutions can potentially make up for language community members' lack of access to web-based archives. Local archives can provide community users with a portal to archived materials in a language they understand, and assist them with unfamiliar digital technologies. In addition, they hold an additional copy of the data and thus strengthen preservation of the materials.

However, local archiving does not come without its challenges. One cannot assume that local archives have the same experience and provide the same infrastructure and services that international language archives do. For instance, local archive staff may not be familiar with digital formats, they may not be particularly interested in language documentation, or they may lack adequate funding to support archiving of digital language materials. Local archiving infrastructures may have difficulty ingesting digital and/or linguistic data and implementing a data access protocol.

The following sections highlight some of my experiences in collaborating with three local archives on archiving language documentation materials.

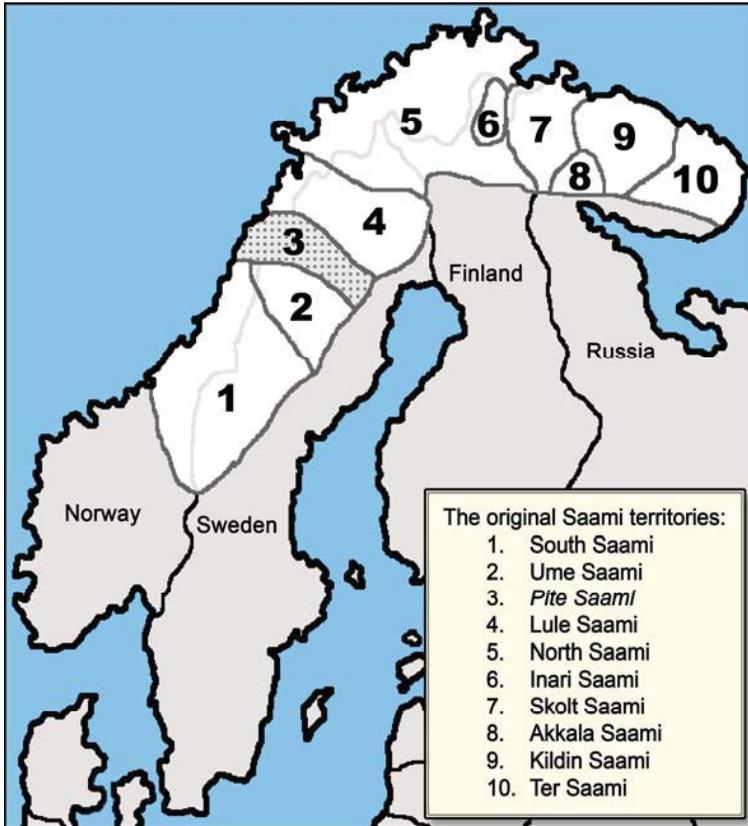
3. Engaging local archiving institutions

Initially, I provide a brief background of the project and materials considered here. The language Pite Saami (Uralic) is one of ten Saami languages which form a dialect continuum across northern Scandinavia, Finland and far northwestern Russia, as illustrated in Figure 1. It is currently spoken by around 30 individuals in and around the Arjeplog municipality in Swedish Lapland.

⁹ Some deposits at the Endangered Languages Archive (ELAR) do this already. For instance, the Cha'pala collection (elar.soas.ac.uk/deposit/0114) provides some metadata in Spanish, and the Pingjiang traditional love songs collections has descriptions in Chinese (elar.soas.ac.uk/deposit/0079) [both accessed 2013-08-12].

¹⁰ Initiatives to provide interfaces in more than just English are certainly a step in the right direction, and the individuals/institutions behind such decisions deserve much credit for beginning to undertake this task.

Figure 1: Map of the Saami language territories, highlighting the approximate area in which Pite Saami was traditionally spoken



As part of the Pite Saami Documentation Project (PSDP), I have collected a spoken-language corpus of approximately 28,000 transcribed words with over 36 hours of audio and video recordings. In fulfillment of the funder's requirements and in support of the Pite Saami community, I have been working with three local institutions in northern Sweden to arrange for archiving of the project's digital language materials.

Here, I use the term 'local' in a very broad sense to mean those institutions which are potentially more accessible to the extended language community and potentially interested local laypeople, in this case from northern Scandinavia; this is in contrast to the international academic community (from linguistics and other fields) that large archives, such as those mentioned

above, serve best. The three Swedish institutions described below (DAUM, Åjtte, Silvermuseet) in fact act at national, regional and municipal levels, respectively.

3.1. The local archiving institutions – three portraits

Below, I provide short overviews of the three local archiving institutions I have worked with. The variety of opportunities and structures that an archive depositor may be faced with, even within just one project in one country, should become clear.

3.1.1. DAUM

The official English translation of *Dialekt-, ortnamns och folkminnesarkivet i Umeå* (DAUM) is ‘The Department of Dialectology, Onomastics and Folklore Research in Umeå.’¹¹ It is responsible for preserving ‘non-material cultural heritage by, on a scientific basis, maintaining and mediating the knowledge of dialects, folklore, folk music, place names and person names’ (DAUM 2013, my translation) for the two Swedish counties of Norrbotten and Västerbotten, and is thus responsible for the territory in which Pite Saami is spoken. DAUM has, among other things, a collection of researchers’ field notes, 5000 hours of recorded language materials and a reference library on the languages, dialects, place names and music of the region. This includes a number of Pite Saami legacy materials. As part of the Swedish state language agency,¹² DAUM is state-funded. Currently, there is a limited on-line search engine in Swedish which allows for simple word searches of the catalogue; only a select set of the metadata is available online, and at the time of writing, none of the materials from PSDP on deposit have been added to this database. However, a new catalogue is being developed, and an on-line version is scheduled to be released in the near future; this will hopefully increase discoverability for the collection. Access to the materials themselves is only possible on-site in Umeå, which is about 350 kilometers from the center of the language community in Arjeplog. Otherwise, copies of digital materials can be ordered from the staff for a small administration/materials fee. No significant access protocol is in place, although the relatively low level of discoverability and

¹¹ For more information see www.sofi.se/servlet/GetDoc?meta_id=1196 [accessed 2013-08-12].

¹² This governmental agency is called *Institutet för språk och folkminnen* ‘The Institute for Language and Folklore’

the relative difficulty of acquiring materials acts as a sort of filter, effectively enabling access only by seriously interested end-users. As a result, however, no distinction is necessarily made between groups of end-users, such as between members of the research/academic community and members of the language community.

3.1.2. *Ájtte*: Svenskt Fjäll- och Samemuseum

Ájtte is the Swedish museum responsible for Saami peoples and culture as well as natural and cultural aspects of the Swedish mountains.¹³ It is situated in Jokkmokk, about 200 kilometers from Arjeplog by car, and also serves as a tourist information center for the mountains of Swedish Lapland. There is a permanent exhibition focusing mostly on the Saami people, as well as space for temporary exhibits. The museum is located directly next door to *Samernas utbildningscentrum*, ‘the Saami Educational Center’. *Ájtte* is funded partly by the regional and local governments, but mainly by a foundation set up specifically for the museum as compensation for the damming of a local river for hydroelectric power. The museum has a library and archive called *Ája*.¹⁴ The archive contains pictures, researchers’ notes, archived documents from various Saami associations, and some language recordings relating mostly to the Saami people, including Saami-language programs from *Sveriges Radio*, the state-run radio station. Materials from the library and archive can only be accessed on-site. There is no web-based interface to the vast majority of materials, although most of the catalogue is accessible on-line. Copies of digitized materials can be ordered from the staff for a small administration/materials fee. No access protocol is in place, although the lack of web access does require interested parties to be physically present in Jokkmokk to make use of archived materials.

3.1.3. Silvermuseet

Silvermuseet, or the Silver Museum,¹⁵ is the local museum for the Arjeplog municipality, the main village in Pite Saami territory. It is named for the importance of silver mining in the region in the past. In addition to serving as a local prehistoric and modern history museum, including a section of the

¹³ The name *Ájtte* is a Lule Saami word which refers to a traditional Saami storage building or shed; the museum’s website is www.ajtte.com/ [accessed 2013-08-12]

¹⁴ *Ája* is the Lule Saami word for a spring or source of water.

¹⁵ The museum’s website is www.silvermuseet.se/ [accessed 2013-08-12]

permanent exhibition on local Saami culture, the museum specializes in subarctic landscape research. It also functions as the local tourist office. The Silver Museum is run by the Arjeplog municipality, but also receives funding from Norrbotten county and the European Union. The museum's archive contains photographic and textual documents relating to Arjeplog, but has never archived linguistic materials in the past, and no general access protocol is in place.

3.2. Experiences with archiving locally

As should be clear from the short descriptions above, the three local archiving institutions are very different from one another, especially in their familiarity with and prioritizing of linguistic materials preservation, as well as how they provide access to such materials beyond just having a language in common with Saami community members (i.e. Swedish). As a result, the opportunities and challenges I have encountered differ across the institutions. Based on the sum of my experiences with them, I would recommend that a documentary linguist looking for a local archive consider the following factors:

- physical accessibility of the archive for the language community: How far is the archive from the community? How easily can community members reach the archive?
- prioritizing the archiving of linguistic materials: Is the staff willing and/or allowed to dedicate time and energy to archiving these specific materials?
- financing the archiving of linguistic materials: Is the institution willing and/or allowed to use its non-human resources to archive these specific materials?
- technical infrastructure: Is the staff familiar with archiving digital materials? Does the institution have the appropriate equipment?
- awareness of format migration: Is the staff aware of and able to deal with the challenges of format migration?
- implementation of a digital access protocol: Does an access protocol already exist? Can it incorporate the relevant restrictions? If not, can one be created?
- terms of deposit: To what extent can the collection be used by the institution? Can copies be made of the materials? Can the collection be used commercially? How long are the terms valid?
- discoverability: Will potential target groups – particularly the language community – be aware of the existence of the collection at the archiving institution?

It goes without saying that the relative significance and prioritization of each of these factors will depend on the goals of any particular documentation project and its community context. In the following discussion, I describe my own experiences in navigating these factors with each of the institutions.

Meetings with the three local institutions' directors and archivists took place in May 2011; occasional email contact has also further supported coordination since then, albeit rather sluggishly. The project materials that I initially handed over consisted of linguistic data gathered in fifteen different recording sessions covering a variety of language-use genres. In total, there were approximately 11 gigabytes of audio, video and image files as well as various annotation and metadata files. This collection was chosen not only as an initial representative sample of recordings, but was also meant to serve as a test-run for further archiving at each institution. Annotations were created using the ELAN software,¹⁶ while metadata were provided in XML-structured plain-text files. As a further attempt at ensuring accessibility to the annotation data in the future and/or for users unfamiliar with ELAN or XML, plain-text (.txt) versions of transcriptions and annotations (created using the 'Export As...Interlinear Text' function in ELAN) were also deposited. The data was provided to the archives on DVDs and SD-cards as data carriers and was then copied onto a hard drive or server at the individual archives. Access permission levels based on the levels implemented at ELAR¹⁷ were explained and expected to be adhered to. Any commercial use by any individual or other entity was expressly ruled out. Finally, I made clear that I intend to continue archiving Pite Saami materials from other sessions in the future as these become finished, pending successful integration of the initial collection into the respective archives.

Subsequent experiences with each of the three archiving institutions are detailed below.

3.2.1. DAUM

In accordance with its role as a language archive responsible for the Saami languages and dialects in Sweden, DAUM was particularly interested in receiving Pite Saami linguistic materials. Their interest in my recordings and

¹⁶ ELAN is free, open-source, xml-based software for linguistic annotation created by the Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands; it is available at tla.mpi.nl/tools/tla-tools/elan/ [accessed 2013-08-12].

¹⁷ For a detailed description of ELAR's access protocol see www.elar-archive.org/using-elar/access-protocol.php [accessed 2013-08-12].

transcriptions was never in doubt; the head archivist was very pleased to receive the materials. Furthermore, there was never any question as to whether DAUM's budget could be applied to support archiving the materials. Since the materials were already in digital format, human resources were required only to add them to the collection. DAUM never doubted that there would be sufficient space on their servers, despite the relatively large volume of video files. Although DAUM had no previous experience with video data, the director has since indicated that it has not been a problem to integrate these into their system. The director is somewhat familiar with metadata in XML format and has limited experience with the ELAN program; this facilitated my explanations of digital aspects of the materials.

The DAUM archive has its own catalogue of metadata for archived materials, and the director said that my metadata could either be imported directly or would be entered by hand by an employee. I also provided a copy of the FileMaker Pro database that I use for metadata and inventorizing within the PSDP project because the director indicated that the archive staff were also partially familiar with that software.

We agreed that, based on the access permission levels I specified, copies of materials could be provided to interested individuals for the minimal administration and materials fee that DAUM normally charges.¹⁸ Any published reference to the materials should cite myself and the PSDP project. Ultimately, this was just a 'gentlemen's agreement', and no written form was drawn up. At the time of writing, the materials have been added to the archive, but are not yet discoverable off-site because on-line access to the updated version of the catalogue is not yet available.

3.2.2. Ájtte

The archive at Ájtte in Jokkmokk was also pleased to receive Pite Saami materials, particularly since the Pite Saami people fall into the museum's area of responsibility. While the archive does not specialize in linguistic materials, the archivist does have some experience with audio recordings, particularly of traditional Saami music; access to these audio materials is available from computers set up in the museum's library. An IT specialist also present at the initial meeting with the archivist indicated that archiving the digital materials

¹⁸ Currently, this fee is SEK 120 per CD (approximately EUR 14) plus shipping; however, 'under special circumstances, the department head may decide to grant a discount' (Institutet för språk och folkminnen GD-beslut (Decision of the General Director) no. 05/2013, p. 4, my translation).

and importing metadata would not be a problem. No one at Ájtte was familiar with XML structured data, nor had anyone ever used ELAN, so I provided very brief introductions to these topics, and pointed out that annotations were also provided in plain, human-readable text files. The archivist was particularly interested in ensuring that school children studying aspects of Saami culture would be able to access the materials, and that the museum could use these in carrying out its own activities. However, because there is no system for implementing access levels, we agreed that materials were only allowed to be accessed on-site, and moreover that copies could only be made with my express permission. If materials are permitted to be copied, then the museum is allowed to charge an administration/materials fee.¹⁹ A binding written agreement to the above effect was drawn up and signed. The agreement is stated to be valid for fifty years, after which time the museum will become responsible for further administration of access rights.

At the time of writing, the materials have not yet been added to the database and are not available via the on-line catalogue. Furthermore, Ájtte has not publicized the existence of the materials in any way to my knowledge, nor is there any mention of them on the museum's website.

3.2.3. Silvermuseet

The local museum in Arjeplog, Silver Museum, is physically the closest of the three institutions to the Pite Saami language community. Because of its location and the linguistic skills of its staff, access to the materials and the technical literacy required to make use of them can be provided in Swedish, which is a significant advantage for language community members without sufficient English or technical skills. From my perspective as a documentary linguist, Silver Museum is therefore the highest priority for short, medium and long-term archiving because it offers greater opportunities for the materials to be utilized by the community. Thus, this museum has the most potential to fulfill my ethical obligations towards the language community, as discussed at the beginning of this paper. However, archiving language documentation materials here turned out to be the most challenging.

First of all, and in contrast to DAUM and Ájtte, the Silver Museum has no official responsibility for the (Pite) Saami language or culture, and thus was not under any obligation to consider including Pite Saami language materials. Secondly, before my deposit, the museum's archive consisted of only paper documents and photographs, and did not include any digital

¹⁹ Currently, this fee is SEK 150 per CD (approximately EUR 18) plus shipping.

materials, linguistic or otherwise. It was thus no surprise that it was necessary for me to give the museum's archivist an introduction to ELAN and XML-structured data.

Nevertheless, the museum's director and archivist both agreed that the museum would be a good place for my materials and were eager to find a solution. An initial request to increase the museum's share of the municipality's server capacity in order to store the additional 20-30 gigabytes of data, however, was denied on the grounds that the increased server size would lead to 'an extra cost for the museum' (my translation of email correspondence with the museum). Because the Silver Museum is administratively a part of the Arjeplog municipality, the museum's server is hosted and financed by the municipality and the proposed increase of server usage had to be authorized by the municipality. The decision to deny this increase thus came from outside/above the museum staff. As I am not familiar with the internal politics of the municipality and the museum, I cannot say to what extent the museum archivist and/or the museum director could have appealed this decision.

The museum's initial response could have been left at that; however, upon my insistence we were able to reach an agreement acceptable to all involved, under the circumstances. We agreed to circumvent the municipal authority by storing the data on a non-networked hard drive in the museum library enabling on-site access only. This arrangement has a number of disadvantages from an archivist's point of view. There is only one copy of the materials; no regular backup is made; and any required migration to newer technologies (including the hard drive medium) will likely be left to the awareness, abilities and the whim of the local archivist. On the other hand, local access to the materials is guaranteed, at least for the short term. Because such local and immediate access to the materials is the main motivation for depositing them at Silver Museum, this solution is acceptable under the circumstances. It is possible that future users from the language community will have to refer to one of the other archives (including ELAR in London) for access. Younger generations of the language community are internet-savvy and, if access to and familiarity with the internet continues to become more and more common, future loss of this very local access is unlikely to lead to a complete loss of access to the materials for the language community. In addition, loss of the materials at the Silver Museum will not mean a permanent loss because multiple copies are held elsewhere.

When depositing the materials at Silver Museum, the terms and conditions of archiving were drawn up by the archivist. These stipulate that the museum may use the materials in its own activities and for research, but no copies will be made without the depositor's express permission for 50

years after the deposit was registered. Any commercial use of the materials is expressly forbidden for all time.

Currently, the Silver Museum has not publicized the existence of the collection, but in early autumn 2013 the museum staff intends to include a description on its website, including summaries which I have provided for each section of the deposit²⁰. Language community members are thus currently only aware of the existence of the materials at the Silver Museum because I informed them directly in the course of discussing permission to archive recordings. Technical problems with installing the ELAN software and accessing the local media files had initially prevented the archivist from attempting to work with the files in my absence. During a field trip in mid-2013, I was able to resolve these and provide the museum staff with a basic introduction to using ELAN to view and navigate the documentation media and transcriptions. The staff intends to allow interested visitors to view the collection using ELAN. Hopefully, the first users will access the collection at Silver Museum in the very near future.

4. Discussion and summary

Modern technology makes it fairly simple for most linguists and other researchers to access primary linguistic data without leaving home. This is particularly valuable for accessing otherwise hard-to-reach data from endangered languages. However, many local language communities – the ultimate source of the data – have no or only partial access to materials in which they are principal stakeholders. Researchers have an ethical obligation to provide communities with the results of research projects as best they can, conveyed by technical interfaces that community members can understand. One method of fulfilling this obligation is archiving at the local level. However, local archives may not share the priorities, experience or technical infrastructure that international language archives do.

In this discussion, I have described my own experiences with attempting to archive the results of the Pite Saami Documentation Project at three ‘local’ institutions so as to ensure access to the materials for the

²⁰ Update: in mid-November 2013, Museum staff put a short announcement (in Swedish) on the museum website noting the existence of the project materials in the museum's library.

language community both today and in the future. I chose three institutions at the national, regional and municipal levels, all of which communicate in Swedish or one of the Saami languages, and thus provide an understandable interface for potential users from the language community. However, this language asset alone is not sufficient to ensure access to materials; a number of other factors should also be considered, as discussed above and summarized in Table 1.

Table 1: Factors to consider in local archiving of language materials, and their relevance to the local archives considered for this study

<i>factors</i>	<i>local archives</i>		
	DAUM	Ájtte	Silverbmuseet
1. physical distance from language community	350 km	200 km	0 km
2. prioritizing of linguistic data	yes	not necessarily	no
3. financing for linguistic data	yes	not necessarily	no
4. extant technical infrastructure	yes	not necessarily	no
5. awareness of format migration	yes	no	no
6. digital access protocol	no	no	no
7. terms of deposit	verbal	written	written
8. discoverability of materials	low	low	low

As a national agency responsible for language archiving, DAUM was familiar with my requirements, had the technical infrastructure required, and was aware of format migration to ensure future preservation and access. The Saami linguistics community in Sweden is familiar with DAUM, and should

have easy access to the materials this way. Surprisingly, though, no official written archiving agreement was suggested by or concluded with DAUM.

The regional Saami museum *Ájtte* was somewhat less familiar with the preservation of language materials, but easily adapted their archive to be able to ingest the project's materials, and should thus be able to provide access to interested Saami individuals and institutions from across *Sápmi*, the Saami homeland.

The most local institution, Silver Museum, was a high priority due to its location in the 'heart' of Pite Saami territory, as materials located there have the most realistic chances of reaching interested community members who rely on face-to-face interaction with the archive, as opposed to on-line access. However, budgeting matters and the lower priority for language materials were potential hurdles. These were overcome by agreeing to keep materials only on a local hard drive. While the long-term preservation of the materials in this way is not guaranteed and the solution defies good archiving practices in several respects, the existence of the same materials at ELAR, DAUM and *Ájtte* lessens the severity of this problem.

My experiences, particularly with Silver Museum, have shown that a good deal of patience is required in working with local archives, as well as awareness of the fact that archiving digital language materials locally should be considered a time-consuming undertaking, particularly if visits to the field site cannot be made regularly. Perseverance by both the linguist and local archivists may be required to attain the goal of archiving locally.

Furthermore, discovery of the materials is not guaranteed by their mere existence in these archives, and it remains to be seen if they will actively promote the existence of the language materials they have now been entrusted with. At the time of writing, little has been done by the archives/museums to make the materials discoverable (however, see footnote 20).. Nonetheless, an up-coming version of the DAUM catalogue should have the materials listed, and the Silver Museum website will soon mention the collection; *Ájtte*, on the other hand, has not actively promoted the existence of the materials so far. The extent to which documentary linguists can influence this kind of situation will vary between institutions and individual linguists, but linguists should certainly make efforts to encourage and support discoverability at local archives. Otherwise, word-of-mouth and chance contact may be the only ways to reach potential users from the language community now and in the future.

In conclusion, the choice of where to archive endangered language materials to ensure access for the language community is important in fulfilling linguists' ethical obligations. While large international archives focus, among other things, on preservation and access for professional, technically savvy users, local archives have a unique ability to reach the

language community, particularly for users without the language skills and/or the technological know-how or equipment required to access on-line archives. In making a choice, the following factors should be considered: physical distance from the language community, levels of prioritization and resourcing for archiving linguistic data, the extent of the existing technical infrastructure, awareness of format migration, existence of an access protocol or willingness to implement one, the formulation of the terms of deposit, and the ability and willingness to make the collection discoverable.

There are two additional and positive side-effects of local archiving worth mentioning here. First, the chances of long-term preservation are improved by increasing the number of copies in existence. Second, archiving locally could help the language community to better understand what archiving endangered language materials entails, and thereby potentially facilitate trust towards archives and researchers in general by providing a tangible example.²¹ Finally, one essential final factor concerns documentary linguists themselves: they should be flexible, patient and persistent in finding a solution to ensuring local archiving of materials, including being prepared to act as a kind of language archiving consultant for dealing with technical skills such as file formats, software and best practices.

References

- Austin, Peter K. and Lenore A. Grenoble. 2007. Current trends in language documentation. In Peter K. Austin (ed.) *Language Documentation and Description*, Vol. 4, 12-25. London: SOAS.
- Czaykowska-Higgins, Ewa. 2009. Research Models, Community Engagement, and Linguistic Fieldwork: Reflections on Working within Canadian Indigenous Communities. *Language Documentation and Conservation*, 3(1), 15–50.
- DAUM, Dialekt-, ortnamns och folkminnesarkivet i Umeå, website. www.sofi.se/1198 [accessed 2013-03-30]
- DoBeS, Dokumentation bedrohter Sprachen, website. www.mpi.nl/ [accessed 2013-03-30]
- ELDP. 2013. Major Documentation Project Information Pack: Grant Round 2013. eldp.net/application/files/infopacks/MDP_Archive.zip [accessed 2013-10-31]
- Grenoble, Lenore A. and Louanna Furbee (eds.) 2010. *Language Documentation: Practice and values*. Amsterdam: John Benjamins Publishing Co.

²¹ I am indebted to an anonymous reviewer for pointing out this aspect. See also the discussion in Gardiner and Thorpe's paper in this volume.

- Grinevald, Colette. 2003. Speakers and documentation of endangered languages. In Peter K. Austin (ed.) *Language Documentation and Description*, Vol. 1, 52-72. London: SOAS.
- Himmelmann, Nikolas. 1998. Documentation and Descriptive Linguistics. *Linguistics*, 36, 161-195.
- LELA, Leipzig Endangered Languages Archive, website. www.eva.mpg.de/lingua/resources/lela.php [accessed 2013-03-30]
- Macri, Martha J. 2010. Language documentation: Whose ethics? In Lenore A. Grenoble and Louanna Furbee (eds.), *Language Documentation: Practice and values*, 37-47. Amsterdam: John Benjamins Publishing Co.
- Nathan, David. 2013. Access and accessibility at ELAR, a social networking archive for endangered languages documentation. In Mark Turin, Claire Wheeler and Eleanor Wilkinson (eds.), *Oral literature in the digital age: archiving orality and connecting with communities*, 21-40. Cambridge: Open Book Publishers.
- Rice, Keren. 2010. The linguist's responsibilities to the community of speakers: Community-based research. In Lenore A. Grenoble and Louanna Furbee (eds.), 25-36. *Language Documentation: Practice and values*. Amsterdam: John Benjamins Publishing Co.
- Samarin, William J. 1967. *Field linguistics: a guide to linguistic field work*. New York: Holt, Rinehart and Winston.