1. Saami languages

Pite Saami is spoken by ca. 25 people in Arjeplog district in Lapland, Sweden.

The Kola Saami languages Kildin (<300 speakers), Ter (<20 speakers), Skolt (<20 speakers in Russia, <300 in Finland) are all spoken on the Kola Peninsula in Russia and adjacent parts of Finland.

2. Language documentation

Endangered languages – here Saami – must be documented before it is too late.

However,
- Only gathering language data does not result in useful documentation
- Storing language data is not the same as creating an archive
- Digitization alone is not sufficient preservation.

"Language documentation is a lasting, multipurpose record of a language." (Himmelmann 2006:1)

With this in mind, it becomes obvious that the biggest problems are:
- the lack of information about the existence and relevance of recordings (missing metadata)
- their unaccessibility (missing annotations and/or physical unavailability)
- as well as the question of long-term storage and preservations (use of open, non-proprietary formats and ensuring longevity of data carriers).

4. Technology and Workflow

Recording (audio, possibly video)
create basic (cataloging) metadata
transfer audio/video material to computer
edit video
create ELAN file
divide ELAN into annotation units
create basic transcription and translation
parse, interlinearize, supplement dictionary
retype into ELAN
get archiving permission from participants
complete deposit
archive deposit
explore Toolbox

5. Annotation and presentation in ELAN

With the help of ELAN, audio and video files of linguistic events can be linked visually to different kinds of annotations. Basic corpus searches can also be performed.

All PSDP and KSDP recordings are annotated in ELAN and accessed via the respective archives. However, because ELAN files are simple text files in xml, the annotations and time alignment are accessible even without ELAN.

6. Enrichment of legacy data

Annotations and detailed metadata can be added to older linguistic materials in accordance with modern archiving standards. Here is an example of a Kildin Saami audio recording stored at KOTUS (picture on the left) which was annotated and archived according to KSDP’s workflow (picture on the right). Annotations (in ELAN) and enriched metadata are stored at the DOBES archive and can be linked to the original KOTUS recording.

7. Links and References

- DoBeS=Documentation of Endangered Languages: http://www.mpi.nl/DOBES
- DoBeS Archive: http://corpus1.mpi.nl/ds/imdi_browser/
- ELAN: http://www.lat-mpi.eu/tools/elan/
- ELAR: The Endangered Languages Archive: http://elar.soas.ac.uk/catalogue/
- HRELP: http://www.hrelp.org/
- KSDP=Kola Saami Documentation Project: http://www2.hu-berlin.de/ksdp/
- KOTUS=Kotimaisten kielten tutkimuskeskus: http://www.kotus.fi/
- PSDP=Pite Saami Documentation Project: http://www2.hu-berlin.de/psdp/
- Toolbox: http://www.sil.org/computing/toolbox/