

Regional Archives and Community Portals

Paul Trilsbeek and Dieter van Uytvanck, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

The Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands, hosts one of the largest digital archives of linguistic resources. Many of those resources are unique recordings of languages that are highly in danger of becoming extinct or in some cases are already no longer spoken. The MPI archive is the central archive for the DOBES²² endangered languages documentation projects funded by the German Volkswagen Foundation²³.

Most linguists working with endangered languages see it as an important aspect of their documentation work to return recorded material back to the speech community in some form, either for language revitalization efforts or just as a moral obligation to return to the communities what belongs to them. The medium of choice for returning material to the communities very much depends on the technology that is available within a particular community and can vary from printed books to CDs/DVDs to online material via the internet. David Nathan calls this particular dissemination of material in various forms mobilisation (Nathan 2006).

Preparing such material for any medium in a useful form requires a lot of knowledge of the particular community and besides that involves a lot of editing and design work. In the case of rich interactive multi-media presentations specific skills are also required to technically implement these presentations, which often means that the linguists need to seek help of technicians to collaborate on the creation of such presentations. For a large archive such as the one at MPI, it is practically impossible to offer the creation of specialized rich multi-media presentations as a standard service.

The MPI archive however does provide two services to facilitate the “repatriation” of linguistic and cultural material to those communities that have internet access at their disposal. One of these services is the establishment of so-called “regional archives”; the other service is to help with the creation of specialized community web-portals.

Regional Archives

The MPI has been involved in the creation of technical infrastructure for digital archiving for about 10 years now. During this time, a framework has been developed that consists of a number of components that can roughly be categorized in three groups:

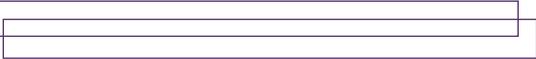
- Tools for linguists to work with their primary data
- Tools for archiving the resources including the creation of metadata descriptions
- Tools for accessing archived material in various ways via the web (browsing, searching, visualizing)

Using this same framework, called the Language Archiving Technology (LAT) framework²⁴,

²² <http://www.mpi.nl/DOBES>

²³ <http://www.volkswagenstiftung.de>

²⁴ <http://www.lat-mpi.eu>



the MPI started an initiative to set up regional archives in locations closer to the areas where the recordings were originally made. This serves a number of purposes:

- Speeding up regional internet access to the archived material that is relevant for that area
- Giving the local researchers an autonomous archive in which they can store whatever resources they want and optionally provide copies to the central archive in Nijmegen to profit from its long-term preservation strategies
- Breaking the almost colonial trend of archiving cultural heritage far away in western archives, which may result in more involvement and interest in archiving cultural heritage from within the communities themselves
- Making the institutions that host the regional archives part of an international grid of archives, which may facilitate further international collaboration, funding requests etc.

These regional archives cannot be placed just anywhere, there are some criteria that need to be fulfilled in order to make them possible. First of all there needs to be a reasonable technological infrastructure present with a reliable broadband internet connection. Also there needs to be local staff who on the one hand can keep the server running from a technical point of view and on the other hand can perform the tasks of an archive manager, helping depositors to put material into the archive and to organize the archive in a consistent way. This generally means that the regional archives are placed within research institutes, museums, or other governmental organizations that can provide the necessary infrastructure.

The procedure of establishing these regional archives varies, but always involves the purchase of a server with sufficient storage capacity. Either the MPI purchases this server in the Netherlands, configures it completely and sends it over to the hosting institution, or the hosting institution purchases the server locally and connects it to the internet and then installation and configuration happens remotely over the internet from Nijmegen. After the server has been installed generally an archive manager from MPI visits the location to give training to both the technicians/archive managers as well as the people who will be working with the archive material. Until now, 6 regional archives have been set up and another 5 are planned for 2009 (Fig 1).

Besides for the regional archives, the LAT framework has also been used within a European project on distributed access management (DAM-LR²⁵). One important component still missing from the LAT framework is a synchronization component to automatically synchronize data between different LAT archives. This component is being worked on, but until now the synchronization needs to be done manually.

²⁵ <http://www.dam-lr.eu>

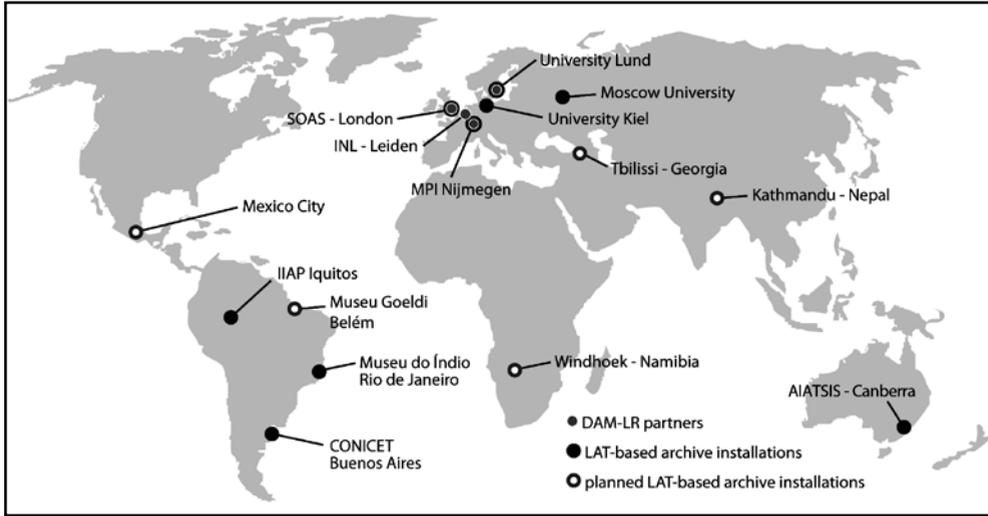


Fig. 1 Established and planned LAT-based archive installations.

Community Portals

The LAT archiving framework makes all archived resources available via the web for authorized users. The standard user interfaces offer extensive browse and search facilities within the metadata catalog, which are suitable for linguists or other specialist users who know what they are looking for, but are less ideal for the general public and the speech communities. To make material available online in a form that is more suitable to the speech communities, the MPI together with some of the DOBES documentation projects have started to develop customized community web portals. The idea behind these portals is that a selection of archived material is presented within an attractive and easy to use website, possibly also in the language of the community. Since internet access and access to computers in remote areas are becoming more common in certain parts of the world, this way of returning cultural material to the communities is becoming a feasible option.

One key feature of the community portals is that they are not static web sites, but dynamic sites that make use of the archived resources and the metadata catalog. Metadata queries are defined within the portal in order to present certain categories of resources that are in the archive. Because these queries are performed on the fly whenever a user clicks on a certain link in the portal, new additions to the archive will automatically become part of the portal if the metadata descriptions match any of the defined queries. This makes the portals almost maintenance free once they have been set up.

The querying of the metadata catalog from an external web portal is made possible through the implementation of a REST (Representational State Transfer) Web Service (Fielding and Taylor 2002). This Web Service makes use of a fairly simple protocol to send a number of query parameters as arguments in a URL to the server and then returns the search results in XML format. This XML result can then be further transformed into nicely formatted lists within the particular dynamic web environment that is used for the portal. In the case of the MPI portals, they are implemented in a content management system called Plone²⁶. Plone

²⁶ <http://plone.org>

is an open source content management system written in the Python²⁷ language, it has a very active developers' community working on all sorts of extensions and it is fairly easy to create customized content types for specific purposes. This makes it an easy platform to integrate dynamic content coming from Web Services such as our search Web Service.

One of the most important aspects of the community portals is the layout and graphical design. For this, a lot of input is needed from the researchers working on the languages since they have more knowledge of what works well for their particular community. On the other hand, only few researchers are also highly skilled web developers or graphic designers, so generally a compromise needs to be made between a design proposal from the researchers, some common usability principles for web site design and possibly some extra design work from a graphic designer. The researchers also need to propose the particular metadata queries that they want to be present in the portal.

The first community portal that was implemented was the Beaver (Dane-zaa) community portal (Fig. 2). Beaver is a language spoken in British Columbia and Alberta in Canada by some 150 remaining speakers. The documentation team has come up with a graphical design that contains images that are of cultural significance to the Beaver community and they have also selected categories of resources that are important in the Beaver culture as well as categories of resources that can be used for educational purposes. The portal is currently still being filled with more content and will be finalized in the next few months.

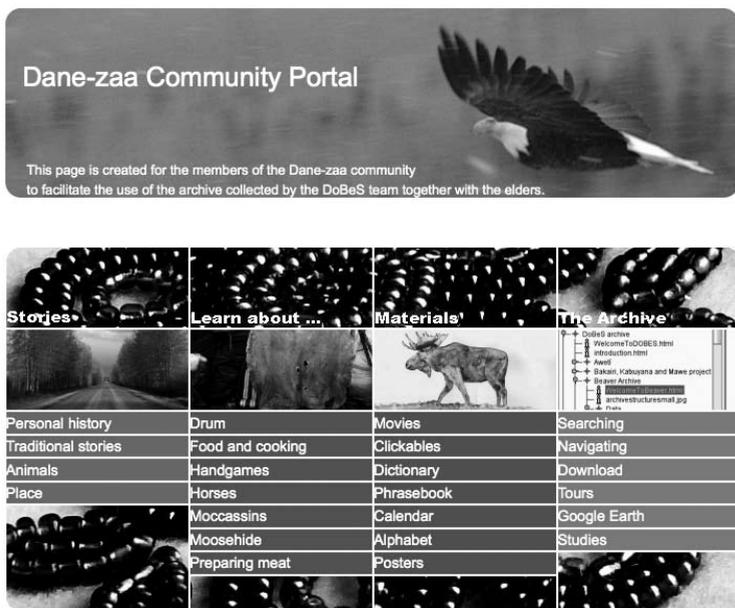


Fig. 2 Front page of the Beaver (Dana-zaa) community portal as designed by the Beaver documentation team²⁸

Obviously for many regions of the world the presentation of material via the web is not yet the most effective way to “repatriate” cultural heritage. The OLCAP project (McConvell

²⁷ <http://python.org/>

²⁸ <http://www.mpi.nl/DOBES/projects/beaver>



and Lee 2008) that is being carried out at the Australian Institute of Aboriginal and Torres Strait Islander Studies in Canberra, Australia, is a pilot project to investigate the extent to which this is true for a number of areas in northern Australia. These areas have access to the internet but the bandwidth is limited and internet access is only available to the community members in places with a public function such as the language center in Katherine, Northern Territory and the library in Lockhart River, Cape York Peninsula, Queensland. Within the OLCAP project, a regional LAT archive has been set up as well as two community portals using the same methodology as described above. Initial results have shown that the available bandwidth in the pilot locations at this moment is good enough for presenting audio material but is not adequate for presenting video material of reasonable quality. Whether the communities in the pilot project appreciate this way of accessing their recordings still remains to be seen, however with the advances being made in broadband internet access in remote areas and the success of web sites such as YouTube also among young indigenous people, one can expect that some years from now the internet will play an important role in returning recordings about endangered languages and cultures back to the communities.

A community portal can either be combined with a regional LAT archive installation, such as in the case of the OLCAP project, or it can be configured to access the archived resources at the central archive in Nijmegen.

Conclusions

The establishment of regional archives can be an important step to help preserve recordings of disappearing languages and cultures. Having an archive present near to the areas where the languages are being recorded will improve access to the resources and may increase interest in documenting endangered languages and cultures both within and outside of the language communities.

Creating community web-portals is becoming a more viable option for returning recordings of endangered languages and cultures to their communities. The advances in broadband internet access in remote locations and the familiarization of members of the language communities with websites such as YouTube will only increase the viability of this medium.

References

Fielding, Roy T. and Taylor, Richard N. (2002), "Principled Design of the Modern Web Architecture" (PDF), *ACM Transactions on Internet Technology (TOIT)* (New York: Association for Computing Machinery) 2 (2): 115–150, doi:10.1145/514183.514185

McConvell, Patrick and Lee, Jason (2008), "OLCAP: The On-Line Languages Community Access Pilot Project in Northern Australia", *International Association of Sound and Audiovisual Archives Annual Conference 2008*, Sydney, Australia.

Nathan, David (2006), "Thick interfaces: mobilising language documentation". In Jost Gippert, Nikolaus Himmelmann and Ulrike Mosel (eds.), *Essentials of language documentation*. Berlin: Mouton de Gruyter. 363-379