Repositories and Archives as Nodes in a Grid

Paul Trilsbeek, Sebastian Drude, Bruna Franchetto, Jürg Gasche, Lucia Golluscio, Carlos Levinho, Victor Miyakawa, Frank Seifart, Daan Broeder, Peter Wittenburg

Already in the second year of the DOBES project the goal was formulated to promote the setup of "regional and local" archives. There are two major reasons, amongst others, for such a step: (1) Long-term data preservation requires a wide distribution of data. (2) Data should be made accessible according to the *Live* Archives ideas close to the places where they were recorded or processed. In the last two years, full-fledged regional repositories were set up at six different places. They are under full control of the local experts and store the corresponding resources using the Language Archiving Technology (LAT) developed at the MPI. In particular for the three South-American repositories in Iquitos, Buenos Aires and Rio de Janeiro, there is the intention to turn them into national or regional centers for language documentation. All these repositories are self-standing nodes, but currently only the MPI repository can offer archiving functions, i.e. it can take care of long-term preservation aspects and store backend versions. All other repositories could offer similar functionality, depending mainly on local efforts and funding. In fact, this grid could be extended by others using equipment that ranges from simple PCs at remote places to, at the high end, big computer centers that store backup versions. Currently, the LAT software provides the basis for exchange of data, easy integration and accessibility of the local centers. A component that takes care of dynamically and symmetrically exchanging data in case of new versions or extensions is in progress. It is important that the repository owners determine which data will be exchanged with whom and that the resource owners have the control over who is allowed to access one of the copies. Since the repository setup is very cost-efficient (the software is free and open source and MPI is offering basic support to keep things running), there are already seven new requests for other repositories. The amount of these requests require to look for additional funding, in particular for the necessary training.

Appendix A: Requirements for a local repository:
- a server with sufficient storage capacity and based on Suse Linux (~6000 Eur)
- formal contracts with institutions with whom some form of exchange is intended
- an Internet connection with a reasonable capacity (for remote places disk transfers could be applied)
- installation of the LAT software (could be done remotely)
- availability of a local expert with UNIX knowledge and experience with linguistic collections who is responsible and invests some of his/her time (depends on the intentions)
- interested and committed researchers

Appendix B: Current LAT Installations
- IIAP Iquitos
- CONICET Buenos Aires
- Museo d'Indio Rio de Janeiro
- Linguistic Dep. Kiel
- Linguistic Dep Lund
- SOAS London
- MPI for Psycholinguistics