Disaster recovery plan for The Language Archive
June 2015

The Language Archive has not yet undergone a formal risk assessment, e.g. making use of the DRAMBORA method, nor does it currently have a formally described elaborate disaster recovery plan. There are however procedures in place that will be followed in case archived resources would be lost or become inaccessible. This document describes these procedures and thereby serves as a preliminary disaster recovery plan until a more thorough and elaborate plan has been worked out.

The storage system of the Technical Group of the Max Planck Institute for Psycholinguistics (MPI-PL), which holds the primary copies of the archived data of The Language Archive, is set up to minimize the risk of complete loss of all copies. The disk arrays and the tape library that together form the storage part of the SAM-QFS Hierarchical Storage Manager system are located in different parts of the building, such that fire or flood are unlikely to damage all copies at the same time.

The Technical Group of the MPI-PL maintains an IT security handbook, the content of which is prescribed by the IT security officer of the Max Planck Society. This handbook contains a section on disaster recovery of data. In the event that all local copies of the stored data are lost due to fire, flood, other natural disaster or malicious acts, copies need to be recovered from the primary external backup location, the Max Planck Computing and Data Facility (MPCDF) in Garching, Germany. In such a scenario, a significant downtime is to be expected. Once the physical facility in Nijmegen has recovered from the disaster that has occurred or has moved into a different temporary facility, the restoration of the stored data can begin. Depending on which proportion of the data needs to be recovered, either a dedicated gigabit network connection will be deployed to copy the materials over the network, or physical storage media (hard disk drives or magnetic tape) will be sent from the MPCDF in Garching to the MPI-PL Nijmegen.

Since the procedure described above could take months, depending on the kind of disaster, and since priority will likely be given to recover data that is needed immediately, such as user home directories and workspaces rather than archived materials, a separate failover strategy for the archive is in place in order to make the archived materials accessible (read-only) within a few working days. This is done by deploying a virtual machine at the MPCDF in Garching that runs The Language Archive’s repository solution and has the storage volume mounted on which the replicated files are stored. This solution has been tested as a proof of concept. In order to have an almost instant failover solution in the future, a synchronised read-only version of a new repository software that is being developed for The Language Archive will be continuously running at the MPCDF Garching and possibly also at the second backup location, the GWGD in Göttingen, Germany, once it is ready for production use (Second half of 2016).