

DOBES

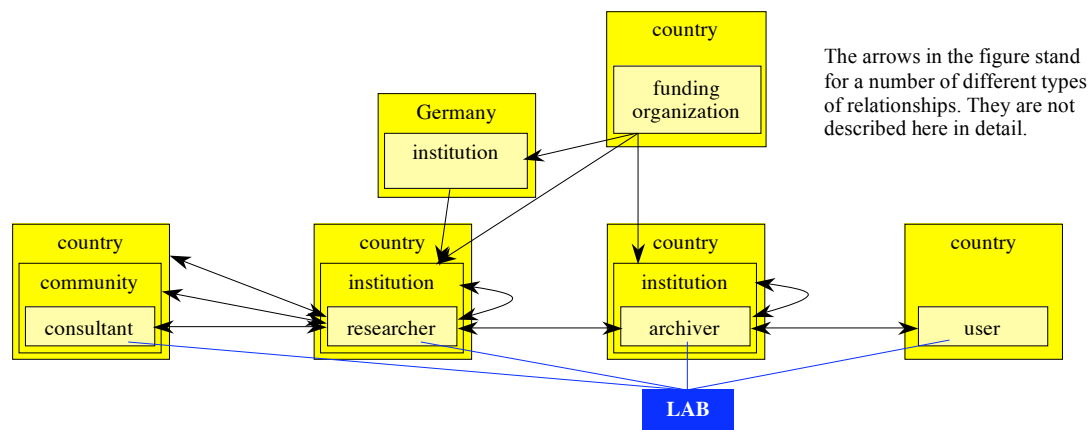
Data Access and Protection Rules *DAPR-V2*

Peter Wittenburg, 19.7.2005

General Statements

This document describes the policies for accessing and protecting the DOBES data as they apply to the archivist.

As indicated in the figure different parties are involved in DOBES program with different roles:



The acting teams in DOBES (indicated by a light yellow color) are embedded in organizational structures all defining legal frameworks of different nature. To define a common basis for all persons and organizations involved in DOBES a Code of Conduct was defined and all are bound to adhere to its rules. Also all users of the DOBES material will have to accept the Code of Conduct as their basis.

Also the archivist is bound to the rules of the Code of Conduct. If the DOBES archive changes its institutional base the transfer to a new institution can only be carried out once the new institution accepts the DOBES Code of Conduct as its operational basis and when it accepts all existing contracts. The archivist falls under the regulations of the European Data Protection law.

Linguistic Data Types

The DOBES archive will contain various primary resources such as audio/video recordings. The copyright for this material in the DOBES archive rests with the creators, i.e., the depositors and the consultants. The archivist will reserve all rights on the archived material on behalf of these creators. The archivist has the right to archive the material.

In addition, the archive will contain all sorts of secondary resources generated by the field researchers such as annotations¹, dictionaries, grammar descriptions etc. The copyright for this material rests with the responsible researcher. It is agreed within DOBES to make this material available to the interested public at least after a period of three years. Nevertheless, it will happen that even access to such resources must be restricted due to ethical reasons.

The archive is organized and searchable with the help of metadata descriptions. All metadata descriptions are openly accessible, since they only inform users about the sort of data that is stored in the archive. Metadata information such as participant names may be anonymized.

General Access Rules

- By default, all archived material except the samples on the DOBES demo web-site and the metadata descriptions are not openly available and access therefore will be restricted.
- The depositor always has unrestricted access to the material he/she has recorded or generated.
- The recorded person(s) have a right to access the corresponding material in the archive. In general, it will be the researcher setting the access rights such that the native speaker community can access the resources.
- In DOBES the depositor is the central link with respect to accessing the data during the time of the DOBES funded project. He/she will inform the archivist concerning the terms of access or set the access rights. The actual usage of the material will be monitored and documented at a web-site. All DOBES funded depositors must complete a final access statement at the end of their funded research period, indicating access rights principles.
- Persons interested in getting access to the archived material have to describe the nature of their intended usage and to accept the Code of Conduct and the copyright rules.
- Every copy and access action has to be documented except for the open examples on the web-site and the metadata transcriptions.
- The responsible researcher can specify the name of a delegate that will take over all responsibilities including the rights to determine access policies.
- In unclear cases or where there is a dispute about access issues the archivist will consult the Linguistic Advisory Board² for its advice.
- The responsible depositor can specify the name of a delegate that will take over responsibility.

Technical Access Rules

- The archive data will be accessible in three ways:
 - video, audio, images, and texts will be stored in open and well-documented formats such that users who have got the rights to access data will be able to use tools of their own choice;
 - specialized and free-to-use software will be made available to browse and search in the XML-based IMDI metadata descriptions and to operate on the multimedia documents;

¹ Annotations are used here in a more general sense as introduced by Liberman and Bird.

² <http://www.mpi.nl/DOBES/INFOpages/program/advisoryBoard.html>

- in addition presentation versions will be made available such that metadata and the resources can be viewed with the help of normal web-browsers
- In general the data will be available via the web only. In special cases the archivist may create copies on disks, DVDs, tapes etc and distribute them by mail. In particular, the archivist would like to support efforts to set up local data centers in the speech communities. The archivist will take efforts to also offer more simple user interfaces.
- Independent of the terms of access users are bound to adhere to the copyright statements and the Code of Conduct. At least three levels of access will be distinguished:
 - There will be openly accessible information such as the samples on the DOBES web-site and the metadata descriptions.
 - There will be material which is available on request by users.
 - There may be material which is not accessible except to the depositors and consultants.

Data Protection Steps

- Given the task to preserve the treasures of cultural heritage for future generations the primary concern of data protection is to improve the probability of long life times.
- Long-life time³ will be assured by
 - storing multiple copies at the archivist's site;
 - storing at least one copy not in the same building;
 - creating dynamic copies at other institutes with great expertise in long-term data storage and trying to support data GRID initiatives to distribute the data worldwide;
 - perform a migration of the data to new storage media technology dependent on the advancement of technology
 - storing the data in "open formats" where possible
- The protection against un-authorized usage will be done with a password mechanism. Passwords will have limited life-times.
- Copyright statements will protect all material at the DOBES web site.
- To all visual material in the DOBES archive a logo can be added on the request of the depositor.

³ The DOBES archivist has taken steps in all mentioned areas. It maintains two copies of the data at the MPI in two different locations and with different technology; it stores dynamic copies at three other locations (Munich, Göttingen, Leipzig); it upgraded its storage system completely after 5 years to modern technology; it participates in GRID-like projects with the goal to create further dynamic copies. It managed to get institutional backing for 50 years for all archived data from the Max-Planck-Society. Contracts with all copyholders are signed that they have to handle the data according to the rules.