

TLA's Mission

- digitize and archive language resources
- support access to language resources
- develop tools, services and infrastructures
- set up of regional archives worldwide
- organize education and training activities
- give help and support

Max Planck Institute for Psycholinguistics

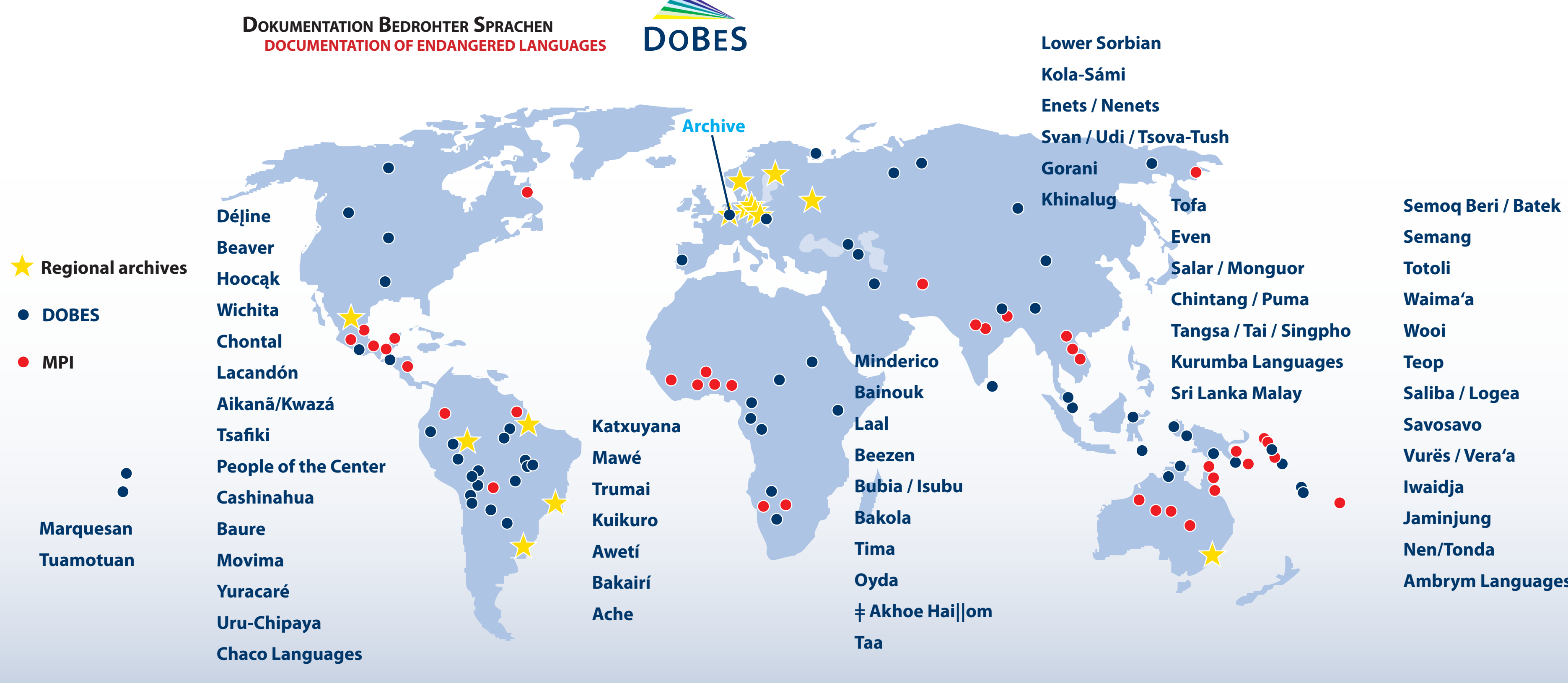
TLA is jointly funded by the Max-Planck-Society, the Berlin-Brandenburg Academy of Sciences and the Royal Netherlands Academy of Arts and Sciences

With substantial contributions by the Volkswagen-Foundation, the European Commission, the German Ministry for Education and Research, the Dutch Science Foundation and the Max Planck Institute for Psycholinguistics.

The Language Archive
Max Planck Institute for Psycholinguistics
 P.O. Box 310, 6500 AH Nijmegen
 Wundtlaan 1, 6525 XD Nijmegen
 The Netherlands
 Phone: (+31) (0)24 - 352 19 11
 Fax: (+31) (0)24 - 352 12 13
 eMail: tla@mpi.nl
 www.mpi.nl/tla



DOKUMENTATION BEDROHTER SPRACHTEN
 DOCUMENTATION OF ENDANGERED LANGUAGES
DOBES

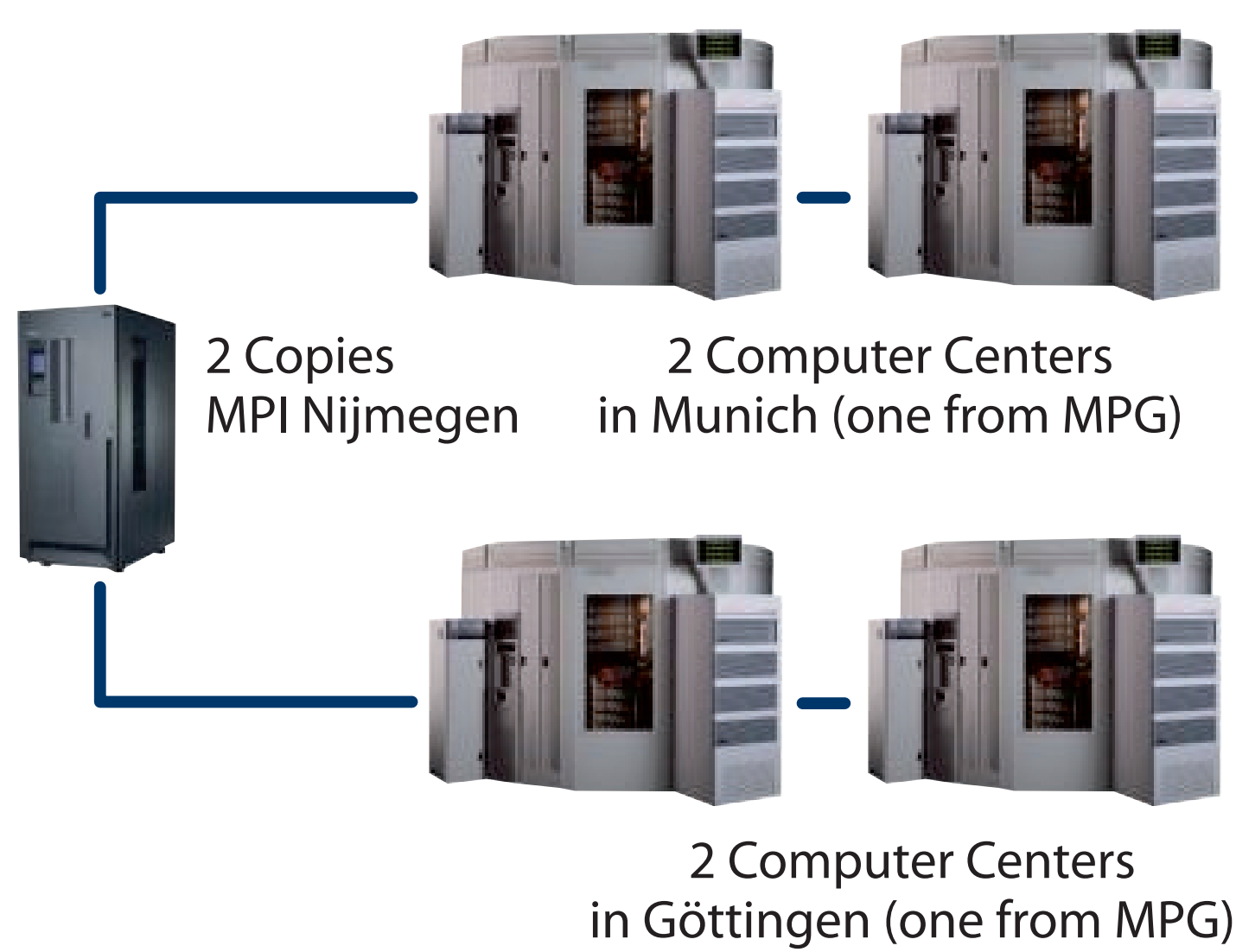


State of the Archive

- 60+ terabyte, 500.000+ files
- 73.000+ metadata sessions
- 20.000+ hours audio/video recordings
- 60.000+ annotation files
- 4.5 million+ annotated segments
- 45+ lexica
- speech, multimodal, acquisition, multilingual, language and cognition, brain imaging, ethnological and other data

Archive

TLA builds on a large archive of language resources, including primary data (multimedia recordings), secondary data (annotation, lexica, comments, etc.), and metadata. To prevent its loss, the Archive is copied to various locations including a growing number of regional archives, preserving relations, contexts and provenance information.



To take care of the interpretability of data in the long run, adherence to standards and a continuous curation procedure are very important. Access to the data in the spirit of the **Live Archives** idea and regulated by a code of conduct and other agreements is guaranteed to those who have access permissions to the individual resources which are defined in four levels (fully open to closed) by the depositors.

Besides the fieldwork data of about 60 DOBES projects, TLA continues to digitize and archive an increasing amount of other language related data. Currently there are data on more than 200 languages in the archive.

Activities

TLA is involved in a number of initiatives devoted to the archiving of digital language data, to the improvement of technologies to create, manage and access language data, and to the construction of infrastructures that facilitate cross-institutional and cross-corpora access. The resulting infrastructures will allow researchers to build virtual collections and workflows to improve data access in the direction of eHumanities usage scenarios. TLA also contributes to standardization in ISO TC37/SC4 (www.tc37sc4.org) which aims at facilitating interoperability in the language resources domain.

Past Projects: MUMIS, INTERA, ISLE, LIRICS, DAM-LR (all EC), CGN (NWO), HARVE, INTER, ROR (all MPG), REPLIX, (DEISA, CLARIN-EU). Running Projects: DOBES (VWS), CLARIN (NL, DE), DASISH, INNETH, CLARA, EUDAT (all EC), AVATech, (MPG-FhG), RELISH (DFG/NEH).

Technology

The LAT software suite, started in 2000 with the multimedia annotation tool ELAN and the IMDI metadata infrastructure, covers about 15 components and tools. It is continuously being debugged, adapted and extended.

It includes tools for Resource Creation & Organization (ELAN, LEXUS, IMDI/CMDI, ARBIL, AV Recognizers), tools for Management, Upload & Infrastructure (LAMUS, IMDI/CMDI, AMS, COSIX, HANDLE, REPLIX), and tools for basic and complex resource access (IMDI/CMDI, VLO, ANNEX, IMEX, LEXUS, GIS, TROVA, VICOS).

Data Life Cycle Support

