



MAX-PLANCK-GESELLSCHAFT



berlin-brandenburgische  
AKADEMIE DER WISSENSCHAFTEN



KONINKLIJKE NEDERLANDSE  
AKADEMIE VAN WETENSCHAPPEN

# TLA @ MPI-PL

## The Language Archive at the Max Planck Institute for Psycholinguistics

Sebastian Drude

*RELISH meets LOEWE, Frankfurt, 2011-10-10*



- The new unit at the MPI-PL in Nijmegen
- Data archive and digitization
- Tools and projects
- Experts and collaborations
- Summary: Mission



- The new unit at the MPI-PL in Nijmegen
- Data archive and digitization
- Tools and projects
- Experts and collaborations
- Summary: Mission



# The new unit: History



MAX-PLANCK-GESELLSCHAFT

- Technical Group at MPI-PL has a long history
- ~1990ies: Challenge of archiving of digital data
- DoBeS ca. 2000 – ca. 2016
- From 2000 on: development of “Language Archiving Technology”
- Need for sustainability of the archive and its activities
- September 2011: new unit was founded



# The new unit: Funding



MAX-PLANCK-GESellschaft

Basic funding by bi-national cooperation:

- MPI Nijmegen (W. Klein) 3
- Max Planck Society 2.5
- Berlin Brandenburg Academy of Sciences 1
- Dutch Academy of Sciences 1

Extended funding by external projects will continue to play an important role



# The new unit: Perspective



MAX-PLANCK-GESELLSCHAFT

- Funding guarantees 7 fixed positions
- Current funding is for 5 years
- Evaluation of quality of work after 3 years
- Purpose: Clear orientation to long-term persistence of data and software



# The new unit: People



MAX-PLANCK-GESellschaft

- |                    |                        |    |
|--------------------|------------------------|----|
| • Wolfgang Klein   | Scientific Director    | 3y |
| • Peter Wittenburg | Head of Unit           | 2y |
| • Daan Broeder     | Deputy Head            |    |
| • Sebastian Drude  | Scientific Coordinator |    |

## **Core Team and responsibilities:**

- |                    |                               |
|--------------------|-------------------------------|
| • Peter Wittenburg | management, projects          |
| • Daan Broeder     | technology, archive software  |
| • Sebastian Drude  | bridge to users, project mng. |
| • Paul Trilsbeek   | archive, digitization         |
| • Han Sloetjes     | annotation software           |
| • Menzo Windhouwer | knowledge software            |



# The new unit: People



MAX-PLANCK-GESELLSCHAFT

archive and digitization

- Paul, Nick, Alexander, Florian, Bastiaan, Rob, Monica, Lucas

archive software (LAMUS, AMS, IMDI, CMDI, ARBIL)

- Daan, Patrick, Guilherme, Mariano, Dieter, Peter Withers, Willem

annotation software (ELAN, ANNEX, IMEX, TROVA, Lat-Bridge)

- Han, Eric, Aarthy, Lari

knowledge software (LEXUS, VICOS, ISOcat)

- Menzo, André, Herman, N.N.

AV-detectors

- Przemek, Eric, Hugo, Binyam, Peter

fluctuation in 2009-2011

- Claus (VICOS)
- Thomas (ANNEX), Marc (LEXUS), Jacqueline (LEXUS/VICOS, Arch), Huib (LEXUS)





- The new unit at the MPI-PL in Nijmegen
- **Data archive and digitization**
- Tools and projects
- Experts and collaborations
- Summary: Mission



## State of Archive:

- about 60 Terabyte of well-described resources
- about 20.000 hours of digitized audio/video recordings
- about 73.000 metadata described sessions
- about 4.5 million annotated segments
- about 47 lexica



## Resources of about 200 languages

- about 60 DOBES teams are documenting more than 80 endangered Languages
- acquisition, speech, multimodal, multilingual, language and cognition, brain imaging, ethnological and other data
- We will continue to accept to digitize and archive other valuable data

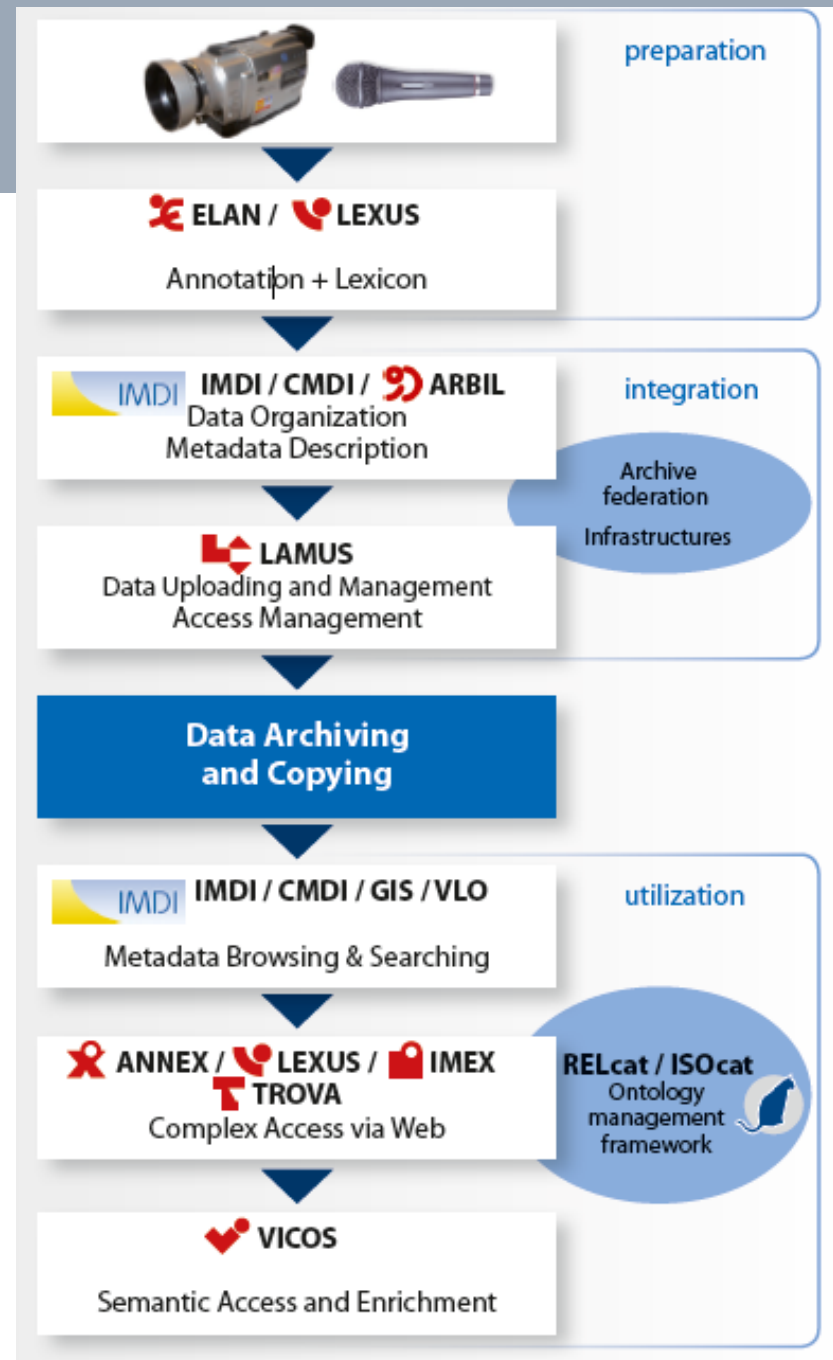


- The new unit at the MPI-PL in Nijmegen
- Data archive and digitization
- **Tools and projects**
- Experts and collaborations
- Summary: Mission



# Tools and projects

- Full lifecycle support
- New metadata format: CMDI (flexible, components can be chosen)
- ARBIL supports CMDI
- AVATech: modules for automatic A/V recognition





# Software state



MAX-PLANCK-GESELLSCHAFT

	state	
LAMUS	mature	much archive federation activities
AMS	mature	no developments
IMDI	oldish	will be phased out stepwise
ARBIL	mature	will be extended
CMDI	in progress	will replace IMDI tools
ELAN	mature	being extended by AV detectors
ANNEX	mature	extensions not clear
IMEX	mature	no developments
TROVA	mature	will be extended in AVATech
LEXUS	redesign	complete redesign in progress
VICOS	redesign	improved integrations in progress
ISOcat	mature	some further developments
Bridge	started	LEXUS-ELAN/ANNEX interaction started
ADDIT	taken out	



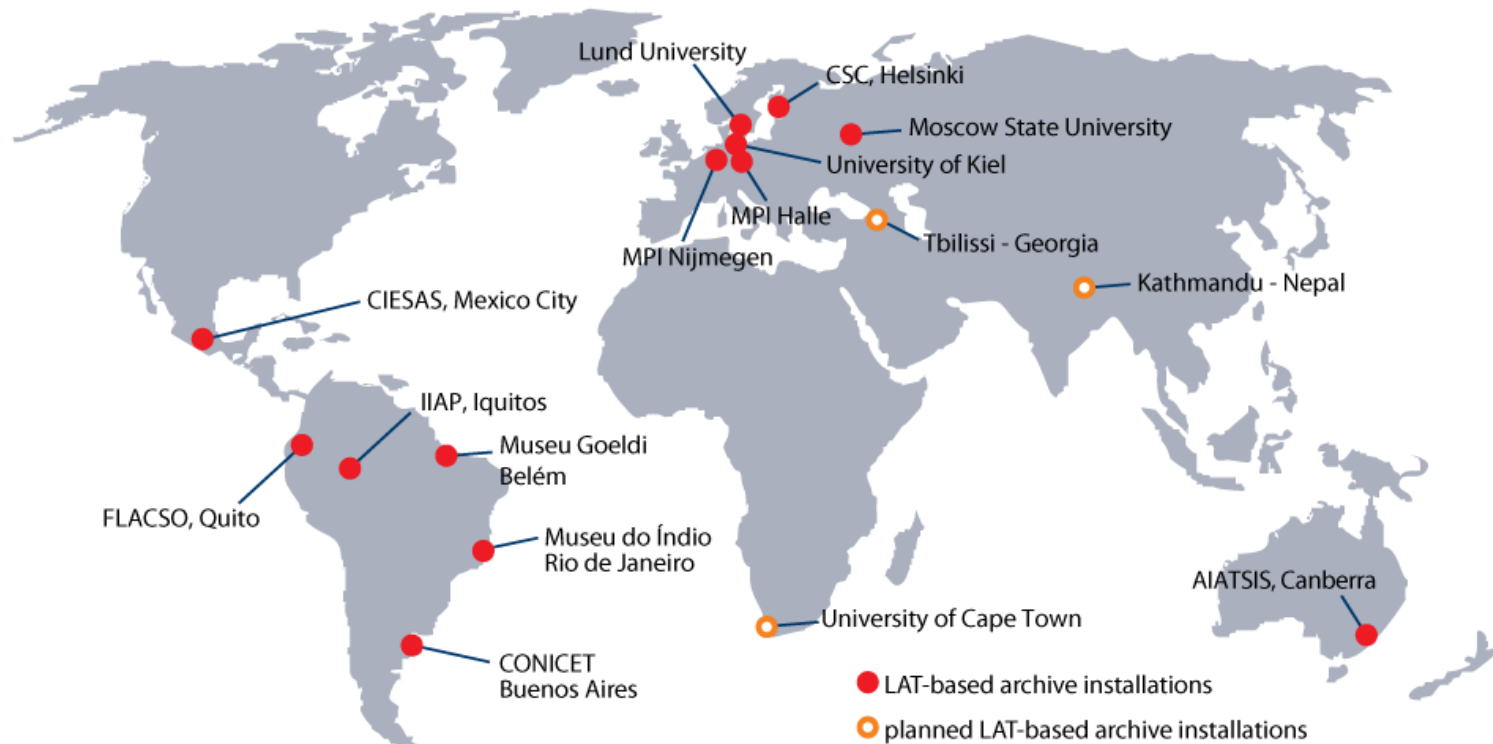
- The new unit at the MPI-PL in Nijmegen
- Data archive and digitization
- Tools and projects
- **Experts and collaborations**
- Summary: Mission



# LAT Archiving Software Installations



MAX-PLANCK-GESELLSCHAFT



Regional Archives Initiative: Collaboration of TLA/MPI-PL with other organizations interested in EL archiving.

Now also larger organizations use TLA LAT archiving software.

- Encourage local resource collecting & archiving
- Come to an exchange of information & data





# Collaborative projects



MAX-PLANCK-GESELLSCHAFT

- ISLE, MUMIS, ECHO, INTERA, DAM-LR, DoBeS, CGN, CLARIN EU, CLARIN NL, CLARIN D, CLARA, Inter, HARVE, AVATech, REPLIX, RELISH
- EUDAT, DASISH, INNET, DAITE CORII
- Standardization work in ISO
  - ISO DCR, MPI-PL is registration
  - Standards: 12620 DCR, 24618 PISA, 24622 CMDI

CLARIN EU, NL, D

- ESFRI roadmap project
- Single virtual domain for LR and LT
- Follow-up in *DASISH*

## Why Infrastructure projects?

- Data is the treasure for empirical research
- Allow our researchers access to the best data & tools
- Fair exchange of efforts
- Important to embed our archive in the wider world



- The new unit at the MPI-PL in Nijmegen
- Data archive and digitization
- Tools and projects
- Experts and collaborations
- **Summary: Mission**



## Summary: Mission



MAX-PLANCK-GESellschaft

- Digitize and archive valuable language resources deposited by any researcher worldwide
- Provide access to archived language resources while respecting legal and ethical restrictions
- Develop tools, services and infrastructures that can be freely used for research purposes
- Set up regional archives worldwide to foster language documentation and revitalization
- Advance international standardization and infrastructure development
- Organize and participate in educational / training courses and give help/support



MAX-PLANCK-GESELLSCHAFT



berlin-brandenburgische  
AKADEMIE DER WISSENSCHAFTEN



KONINKLIJKE NEDERLANDSE  
AKADEMIE VAN WETENSCHAPPEN

# TLA @ MPI-PL

## The Language Archive at the Max Planck Institute for Psycholinguistics

Sebastian Drude

*RELISH meets LOEWE, Frankfurt, 2011-10-10*