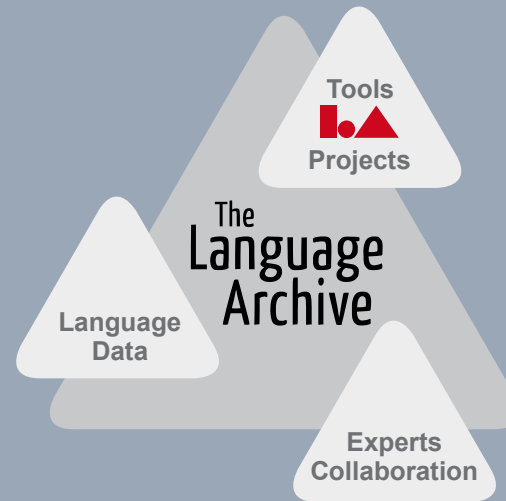


CLARIN

Common Language Resources and Technology Infrastructure



MAX-PLANCK-GESELLSCHAFT



Arbil

Structuring research data and writing metadata

Alexander König

The Language Archive - Max Planck Institute for Psycholinguistics
Nijmegen, The Netherlands



What is Arbil?



MAX-PLANCK-GESELLSCHAFT

Arbil

(short for “Archive Builder”)

is an application for organizing research data and associated metadata into a format appropriate for archiving.



Arbil Target Groups



MAX-PLANCK-GESELLSCHAFT

Arbil has been developed to service two different user groups

- 1) Documentary linguists from the MPI and the DoBeS projects
 - > IMDI support
- 2) the CLARIN community
 - > CMDI support



IMDI:

- Developed by MPI/TLA for DoBeS research data
- Designed to describe multi-media and multi-modal language resources.
- The TLA archive infrastructure and their various tools are based on IMDI.
- More information: <http://tla.mpi.nl/imdi-metadata>



CMDI:

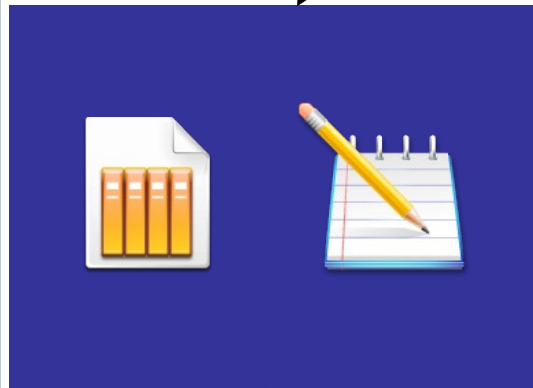
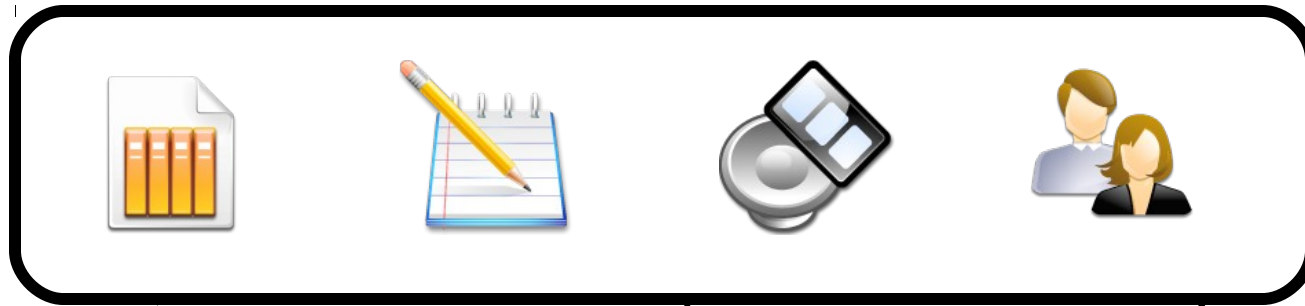
- **Component MetaData Infrastructure**
- Developed within the pan-European CLARIN (**C**ommon **L**anguage **R**esources and technology **I**nfrastructure) project
- Flexible building blocks
- Increased interoperability by referencing a data category registry
- More information: <http://www.clarin.eu/cmdi>



Build your own metadata schema



MAX-PLANCK-GESellschaft



project 1

project 2

project 3



Main Arbil Principles



MAX-PLANCK-GESellschaft

- Workflow focused
- Based on table views
- Drag and drop
- Bulk copy and paste
- Multiple undo and redo
- Usable offline



Arbil Features



MAX-PLANCK-GESELLSCHAFT

- You can edit multiple files by opening them as table rows (e.g. to compare certain values)
- If the metadata schema uses controlled vocabularies, Arbil provides them in a dropdown list
- You can save snippets of metadata as Favourites to reuse them later.
- You can drag resources (media, text) into the metadata instance and Arbil automatically populates some fields (size, file type)



Arbil User Interface



MAX-PLANCK-GESELLSCHAFT

Remote Corpus
View and import metadata from remote servers

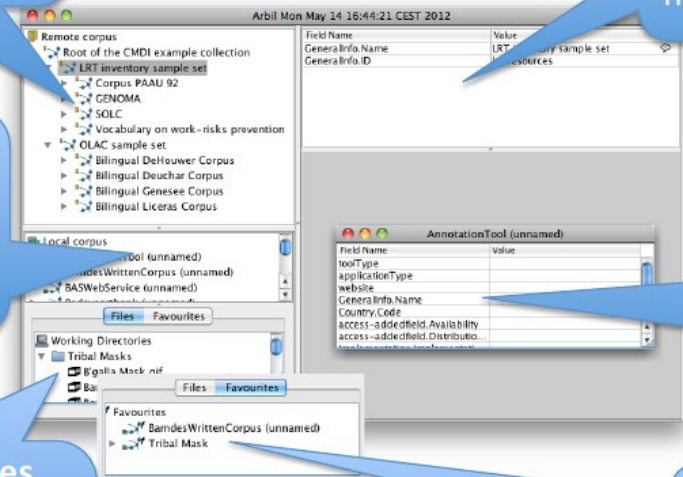
Local Corpus
All newly created metadata will be created here

Working Directories
Your data files can be browsed and associated with new metadata from here

Preview Table (optional)
The currently selected metadata.

Main Work Area
Multiple tables of metadata can be viewed and edited

Favourites
Frequently used metadata is saved here for easy replication

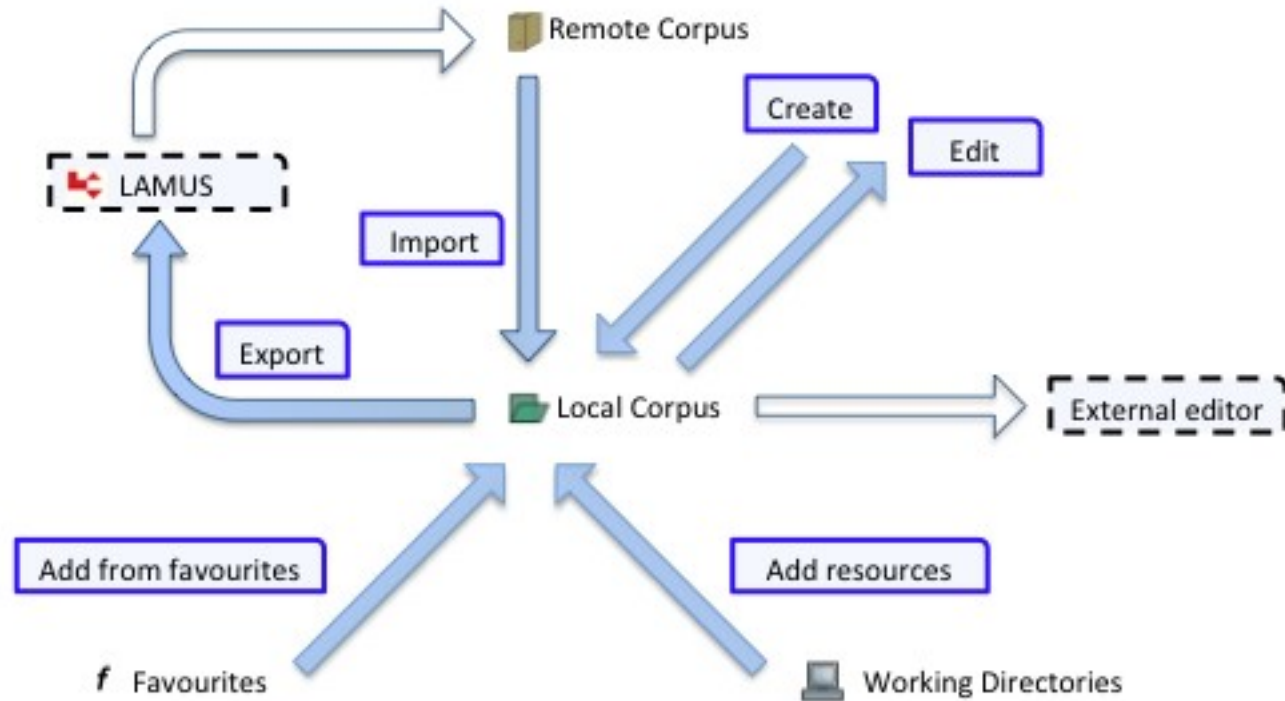




Arbil Workflow



MAX-PLANCK-GESELLSCHAFT





Alternatives to Arbil



MAX-PLANCK-GESELLSCHAFT

- CLARIN is getting bigger and some projects have already developed their own solutions that are better suited to their specific needs
- Two examples
 - a project already has their metadata in Excel tables and they don't want to copy it all to Arbil
 - Solution: Set up a script to convert Excel to CMDI



Alternatives to Arbil



MAX-PLANCK-GESELLSCHAFT

- Another project only has little metadata and doesn't want their researchers to go to the effort to learn Arbil
- Solution: ProFormA
 - simple HTML forms that generate CMDI



Alternatives to Arbil



MAX-PLANCK-GESellschaft

ProForm

Form-based CMDI-Editing

Create CMDI files with a NaLiDa profile

Create new CMDI-file

Please select the CMDI Profile

Please insert your preferred filename (no extension)

Create new CMDI files

Create new CMDI-file

Please insert the CMDI Profile ID

Please insert your preferred filename (no extension)

Edit CMDI files available via URL

URL specification

Please enter the URL to your CMDI file



Alternatives to Arbil



MAX-PLANCK-GESELLSCHAFT

- Both IMDI and CMDI are XML-based
- That means that you can always edit the plain XML
 - either using a dedicated editor like oXygen
 - or by writing scripts in your favourite scripting language (Perl, Python, etc.)



Arbil: **Strengths** and Weaknesses



MAX-PLANCK-GESELLSCHAFT

- + Intuitive GUI (table-based, drag'n'drop)
- + Usable online and offline
- + Easy bulk edit or search & replace
- + Java-based: runs on Windows, Linux, Mac
- + Open Source (GPL)



Arbil: Strengths and **Weaknesses**



MAX-PLANCK-GESELLSCHAFT

- Still a bit of a learning curve
- Has trouble dealing with lots of files or very big media files
- In some places still too focused on TLA archive
- Java-based: support on the client is waning



Get Arbil



MAX-PLANCK-GESELLSCHAFT

You can download Arbil from the TLA website

<http://tla.mpi.nl/tools/tla-tools/arbil/>

Providing you already have Java installed the webstart version is the fastest way to start

Alternately there are installers for

- Windows,
- Mac
- and Ubuntu (Debian).

Latest stable version



Run via webstart



Download the Windows installer



Download the Debian Linux package



Download the Mac installer



Arbil in Action



MAX-PLANCK-GESELLSCHAFT



Questions?



MAX-PLANCK-GESELLSCHAFT



<http://tla.mpi.nl/tools/tla-tools/arbil/>

Alexander.Koenig@mpi.nl