Metadata Management with Arbil

Peter Withers, Twan Goosen, and Sebastian Drude
The TLA Language Archive,
Max Planck Institute for Psycholinguistics Wundtlaan 1, 6525 XD, Nijmegen, The Netherlands
peter.withers@mpi.nl
Overview

• What Arbil Does
• Metadata Formats
• The User Interface
• Workflow
• Configuring ARBIL
• Metadata Display
• Table Views
• Controlled Vocabularies
• Using Favourites
• Creating Metadata for a Resource
• Searching the Metadata
• Find / Replace and Highlighting Cells
• Conclusion
What Arbil Does

• ARBIL is an application for organizing your research data with metadata such that it can be archived as opposed to buried.
• Metadata structures can be created, viewed and edited.
• Multiple metadata files can be edited at the same time. Many metadata files can be bulk edited in a single table.
• You can enter the data as it becomes available, there is no mandatory order.
• The local metadata tree can be searched with multiple parameters.
• The data files can be launched in the associated applications such as ELAN or Media Players.
• With the exception of accessing the remote archive, all features are available offline such as in remote field sites.
Both **IMDI** and **CMDI** (CLARIN metadata) are supported.

**Other XML formats** can potentially be supported by the use of custom templates or schema files.
IMDI (ISLE Meta Data Initiative) is a metadata format for linguistic data

- Has been around for about a decade
- Targets multimodal/multimedia linguistic resources
- Has a single schema file with fixed metadata fields and optional key fields
- Is used for instance throughout the Language Archive in Nijmegen and associated archives
CMDI is CLARIN’s flexible component based format

- Developed as a part of the CLARIN EU project
- Each profile has its own schema file
- Can be adapted to suit specific project needs
- Metadata fields and layout are customisable
- Semantic interoperability through ISOCAT
The User Interface

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
Workflow

- Create and edit metadata
- Import for offline use
- Import as favourites
- Insert from favourites
- Save as favourite
- Add resource files
- Backup to USB Disk
- Export

Withers, Goosen, & Drude (MPI-PL)
1/3/2013 ICLDC
Configuring ARBIL

- When Arbil is first started, you will be asked which metadata format to use.
- Depending on how you intend to use ARBIL you will want to select IMDI templates or Clarin profiles.
- You can change later in the templates & profiles dialogue.
- This can also be preconfigured by your system administrator.

The selected profile will be available via the add menu.
• The metadata is displayed in tables and trees, which allow an overview of the metadata and the ability to populate and update many metadata sections in bulk.

Metadata files are shown in the tree

Multiple metadata files or subsections can be viewed in a single table

Invalid metadata will be highlighted in red

Metadata are shown as separate subnodes
Table Views

• When multiple metadata files or subsections are viewed in a single table, they are each shown in a separate row.

• When a table shows a single metadata file or subsection, it is shown in the long view.

• Individual fields can be edited in the long field view, which allows each field to be edited sequentially.
Controlled Vocabularies

- Clarin vocabularies are read from the schema file.
- IMDI vocabularies are read from an XML file.
- Both are provided as dropdown lists in the table when editing.
Using Favourites

- Favourites are snippets of prefilled metadata instances
- New metadata instances can be created from them, leaving only the specific details to be edited
- Either entire instances can be used or just the desired sections of an instance added to an existing instance
Creating Metadata for a Resource

- Create a metadata instance
- Add a data file
- Add a metadata subsection
- Add a field
Searching the Metadata

• Searches in Arbil are tree based.
• Multiple search parameters can be entered.
• The search results are shown in a standard table with the usual editing facilities.
Find / Replace and Highlighting Cells

- The table can be searched for specific text.
- Selected cells can have the found text substituted.
- Matching table cells can also be highlighted.
Current developments

• **Internationalisation**, providing support for languages other than English in the user interface

• Advanced, high performance **metadata search**, both locally and on the server

• Extensibility through **plugins and modules**
Plugins and modules

• Flexible functional extensions through plugin framework available in current release
• Arbil itself is being modularised, sharing components with KinOath already
• Certain plugins may evolve into built-in Arbil modules
• Arbil functionality becomes available for other tools, such as a web-based archive search
Plugin extensions

- **Plugin example:** favourites import/export

Available as a plugin soon!
Conclusions

• Arbil has been developed with a strong focus on workflow and usability.
• User can view and edit metadata without mandating any sequence of entry.
• Warnings will be shown, if the metadata does not comply with the requirements.
• It is hoped that the features of Arbil will lead towards the recording of metadata at an earlier stage resulting in greater detail and better quality of that metadata.
Conclusions

• Metadata can be managed for resources in all humanities

• A plugin framework allows for flexible extension of Arbil functionalities

• Arbil functionality will become available in other applications, local and on the web

• New features and improvements are in active development
Installing Arbil

- There is a link to ARBIL on the MPI website [http://tla.mpi.nl/tools/arbil](http://tla.mpi.nl/tools/arbil)
- Providing you already have Java installed the Web Start version is the fastest way to start
- Alternately there are installers for Windows, Mac and Ubuntu (Debian).
- The manual and user guide are also available for download on the same page.