New Developments in Arbil Metadata Manager

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

• What Arbil Does
• The Target Workflow
• Metadata Formats
• The User Interface
• Workflow
• Metadata Display
• Table Views
• Controlled Vocabularies
• Using Favourites
• Creating Metadata for a Resource
• Searching the Metadata
• Find / Replace and Highlighting Cells
• Installing Arbil
• Arbil in the Future
  – Marbil
  – YAAS prototype
What Arbil Does

• ARBIL is an application for organising 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.
• 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.
• Many metadata files can be bulk edited in a single table.
• With the exception of accessing the remote archive, all features are available offline such as in remote field sites.
• You can enter the data as it becomes available, there is no mandatory order.
The Target Workflow

• Arbil was designed for the DoBeS community
• There was a specific workflow in use
• The users were trained in that workflow
• Curation and offline use was a core need
• Outside of the DoBeS community there are many possible workflows and different needs
• To achieve this Arbil will need to be more flexible yet still guide the user
Metadata Formats

• Both IMDI and Clarin metadata formats are supported.
• Other XML formats can potentially be supported by the use of custom templates or schema files.
Metadata Formats

IMDI (ISLE Meta Data Initiative) is a metadata format for linguistic data

- Has been around for about a decade
- Targeted towards multimodal / multimedia
- 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

Clarin metadata is a flexible format with ISOCAT

- It is a more recent metadata format
- Each profile has its own schema files
- Developed as a part of the Clarin EU project
- Can be adapted to suit specific project needs
- The metadata fields and layout are customisable
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

Peter Withers, Paradisec RR 3/12/2013

New Developments in Arbil Metadata Manager
- 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

Workflow

Create and Edit

Backup

Export

Peter Withers, Paradisec RRR
3/12/2013

New Developments in Arbil Metadata Manager
Metadata Display

• 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.
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 the entire favourite can be used or just the desired sections added to existing metadata
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.
Installing Arbil

- There is a link to ARBIL on the MPI 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).
- The manual and user guide are also available for download on the same page.
Arbil in the Future

• MArbil
  – Simplified interface
  – Less workflow specific

• YAAS
  – Online search
  – Potential mobile application
MArbil

- Highly simplified interface
- Minimal workflow requirements
- The local cache will not be relevant
- The local corpus can be any directory/s
- The metadata files reside next to the data file
- The OS’s file browser is the main entry point
- Search widget:
  - Immediate use of the metadata
  - Gathers the data/metadata into one interface

Peter Withers, Paradisec RRR
3/12/2013

New Developments in Arbil Metadata Manager
YAAS prototype (yet another archive search)

- Web based archive search
- Based on Arbil (shares components)
- Features:
  - Corpus tree
  - Data tables
  - Faceted tree
  - Search results tree
  - Branch based search

New Developments in Arbil Metadata Manager
Peter Withers, Paradisec RRR
3/12/2013
**YAAS prototype** (yet another archive search)

- XML database like KinOath
- Rendered in HTML5
- Can also be compiled into mobile application
- Potentially a cloud based design in the future

**YAAS Web Prototype**
Translations of Arbil

• Arbil has been translated into a few languages
  – English
  – Spanish
  – Italian
  – German

• These languages are known to our current student assistants
  – Other languages will be added when possible
  – As the student assistants change our ability to support a given language will also change

• Launchpad (or other)
  – Provides a translation tool
  – Facilitates community based translations
  – Any language with sufficient community interest can therefore be maintained
Conclusion

• Arbil has been developed with a strong focus on the workflow of the DoBeS community.
• Future work is aimed at a wider community.
• 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.