ARBIL, a tool for organising corpus/language data

Peter Withers
Max Planck Institute for Psycholinguistics, Nijmegen
Peter.Withers@mpi.nl
2011-01-17
Outline

- What is ARBIL
- What is metadata
- Why was ARBIL written
- The user interface
- Workflow
- Adding metadata
- Visualising and using the metadata
- Exporting the metadata
What is ARBIL

- ARBIL is an application for organising research data and associated metadata into a format appropriate for archiving.
- There are many features in ARBIL that enable users to view and edit their data.
- The data can be viewed side by side in tables and bulk edited in the same table.
- ARBIL is designed so that it can be used offline in remote locations.
- The data can be entered at any stage in part or as a whole.
What is Metadata?

Metadata is information that describes the content of the files you are working with (the data)

This includes:
- Who collected the data
- Where was it collected
- What the data contains

This information can then be used so that you and others (if permission is granted) can find the data in the future.

(most archives allow you to search the metadata but not the data itself)
The development of ARBIL has stemmed from discussions with many researchers at the MPI and members of the DOBES community and the experience gained from previous metadata editors developed at the MPI over the last ten years.

ARBIL has been developed over the last two years and contains many features in order to fulfill the wide ranging needs expressed by its users, while maintaining the functionality of the previous editors. ARBIL continues to be actively developed to extend these features further.
ARBIL is essentially an XML editor
Reads and writes to XML
Validates via the schema
Generates XML from the schema
Specialised Functions

- ARBIL differs from a standard XML editor in that is has functionality specialised for editing metadata.
- Whereas if a user hand edits the XML files there are likely to be difficult to detect errors.
- The attaching resource files is handled consistently.
- Both IMDI and CMDI formats are supported.
- Workflow focused
- Table view
- Drag and drop
- Bulk copy and paste
- Multiple undo and redo
- Resource file preview
- Customisable columns
Getting to know the interface

Remote Corpus
View and import metadata from remote servers

Local Corpus
All newly created metadata will be created here

Working Directories
Your resource 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
- Export
Configuring ARBIL Templates

- Depending on how you will use ARBIL you may want to select specific templates. For instance to use a specific Clarin profile you must enable it in the templates dialogue. This can also be preconfigured by your system administrator.

The selected profile will be available via the add menu.
Metadata files are created in the “Local Corpus” via the “Add” menu.

At each level of the tree the appropriate sub components are offered in the “Add” menu.
Resources can be attached to a session by dragging
Or by adding a resource node then browsing for the file.
Entering Data

- The data can be edited in any table.
- Tables can contain one or many separate nodes.
- Data can be edited in bulk across the table.
- Controlled vocabularies are offered as a dropdown.
- Data can be compared across multiple rows at a time.
- The table data can be copied for use in external applications.
Some fields can be entered in multiple languages
The full language list is very long so it should be condensed
Your selected languages will be shown in the table dropdown
The long field editor allows text that is too long to display in a table to be viewed and edited.

- Fields with multiple values can be viewed.
- Key names (when present) can be edited.
The table can either show separate documentation languages a condensed column or as a separate column for each language.

<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
<th>Description [ISO6]</th>
<th>Description [ISO6]</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010-07-13</td>
<td>Demo Session</td>
<td>This is a demo ...</td>
<td>Dit is een demo ...</td>
</tr>
<tr>
<td></td>
<td>Standard ...</td>
<td>Unspecified</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010-07-13</td>
<td>&lt;multiple values&gt;</td>
</tr>
<tr>
<td></td>
<td>Unspecified</td>
</tr>
</tbody>
</table>
Organising Data

- Columns can be hidden to simplify the information displayed in a table.
- Sets of these column views can be saved and easily selected for display.
- A default column view can be selected so that new tables show your preferred columns.

<table>
<thead>
<tr>
<th>Title</th>
<th>Data</th>
<th>Desc</th>
<th>Location</th>
<th>Location</th>
<th>Location Address</th>
<th>Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSL_JF</td>
<td>ECHO SSL JF</td>
<td>June 2003</td>
<td>Unspecified</td>
<td>Unspecified</td>
<td>ECHO</td>
<td>ECHO</td>
</tr>
<tr>
<td>SSL_JF</td>
<td>ECHO SSL JF</td>
<td>June 2003</td>
<td>Copy</td>
<td>Paste</td>
<td>ECHO</td>
<td>ECHO</td>
</tr>
<tr>
<td>SSL_JF</td>
<td>ECHO SSL JF</td>
<td>June 2003</td>
<td>Open in Long Field Editor</td>
<td>Im Univar</td>
<td>ECHO</td>
<td>ECHO</td>
</tr>
</tbody>
</table>

Save Column View

- Enter a name to save this Column View as DemoView
  - Save Column View
  - Edit this Column View
  - Show Only Current Columns
  - Clear Cell Highlight
  - Highlight Matching Cells
  - Copy Cell to Whole Column
  - Collapse Child Nodes

<table>
<thead>
<tr>
<th>Title</th>
<th>Data</th>
<th>Project Name</th>
<th>Project Title</th>
<th>Project Id</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSL_JF</td>
<td>June 2003</td>
<td>ECHO SSL JF</td>
<td>The elephant's boy and the wolf</td>
<td>ECHO</td>
</tr>
<tr>
<td>SSL_JF</td>
<td>June 2003</td>
<td>ECHO SSL JF</td>
<td>The tortoise and the hare</td>
<td>ECHO</td>
</tr>
<tr>
<td>SSL_JF</td>
<td>June 2003</td>
<td>ECHO SSL JF</td>
<td>The lion and the mouse</td>
<td>ECHO</td>
</tr>
<tr>
<td>SSL_JF</td>
<td>June 2003</td>
<td>ECHO SSL JF</td>
<td>The boy friends and the bear</td>
<td>ECHO</td>
</tr>
<tr>
<td>SSL_JF</td>
<td>June 2003</td>
<td>ECHO SSL JF</td>
<td>The dog and his reflection</td>
<td>ECHO</td>
</tr>
<tr>
<td>SSL_JF</td>
<td>June 2003</td>
<td>ECHO SSL JF</td>
<td>The shepherd's boy and the wolf</td>
<td>ECHO</td>
</tr>
<tr>
<td>SSL_JF</td>
<td>June 2003</td>
<td>ECHO SSL JF</td>
<td>The elephant's boy and the wolf</td>
<td>ECHO</td>
</tr>
<tr>
<td>SSL_JF</td>
<td>June 2003</td>
<td>ECHO SSL JF</td>
<td>The tortoise and the hare</td>
<td>ECHO</td>
</tr>
<tr>
<td>SSL_JF</td>
<td>June 2003</td>
<td>ECHO SSL JF</td>
<td>The lion and the mouse</td>
<td>ECHO</td>
</tr>
<tr>
<td>SSL_JF</td>
<td>June 2003</td>
<td>ECHO SSL JF</td>
<td>The boy friends and the bear</td>
<td>ECHO</td>
</tr>
<tr>
<td>SSL_JF</td>
<td>June 2003</td>
<td>ECHO SSL JF</td>
<td>The dog and his reflection</td>
<td>ECHO</td>
</tr>
<tr>
<td>SSL_JF</td>
<td>June 2003</td>
<td>ECHO SSL JF</td>
<td>The shepherd's boy and the wolf</td>
<td>ECHO</td>
</tr>
</tbody>
</table>
Visualising the Data

- Matching text can be highlighted.
- Sub-nodes can be shown in the parents table as a cell to give a visual indication of their presence and to give easy access to them and to sort the table by them.
- The full metadata of the sub-nodes can be displayed in a separate table just like any other metadata.
- All columns can be used to sort the rows even the sub-node columns
- The table can be searched for specific text
- Selected cells can have the found text substituted
The local metadata tree can be searched.

Multiple parameters can be used for the search.

The search results are displayed in a table that has all the functionality of the other tables in ARBIL.

Multiple different branches can be searched at one time.
Exporting

- Entire branches can be exported as discrete files
- IMDI files can be exported to HTML
- The exported branches can be uploaded into LAMUS
Installing ARBIL

- There is a link to ARBIL on the MPI website http://www.lat-mpi.eu/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).