Integrating TEI and EAD

Chris Turner
LEADERS Project,
School of Library, Archive and Information Studies,
University College London

Copyright, UCL

LEADERS: Linking EAD to Electronically Retrievable Sources
Linking EAD to Electronically Retrievable Sources

- Developing a generic computer-based toolset that integrates EAD encoded finding aids with TEI encoded transcripts and digitised images of archival material.
- Text Encoding Initiative (TEI)
  - For archival source materials
- Encoded Archival Description (EAD)
  - For finding aids
- NISO Metadata for Images in XML (MIX) Schema
  - For digital images metadata
TEI for Archives: Research Methodology

• Analysis of commonly occurring structures, features and contents found within a range of different types of archive source material

  • Material held in UCL’s Special Collections and Record Office
  • Expect to analyse material held in other archival repositories to validate initial findings and uncover as wide range of encoding challenges as possible
TEI for Archives: Preliminary findings

• ‘Transcription of Primary Sources’ tag set in TEI can deal with a wide range of encoding challenges inherent in archival material:
  – Complex additions, deletions and corrections
  – Gaps within and damage to the text
  – Changes in document hands, style and character of writing
TEI for Archives: Preliminary findings

• Need to explore encoding options for:
  – ‘overlaid data’
  – Textual and numerical data presented in complex tables
  – Formulae and mathematical expressions within the text

• Examining TEI and other DTDs specifically built to handle these structures and features
Exploring solutions for encoding ‘overlaid data’

‘overlaid data’: when an underlying layer of data is used as the basic structure onto which further data [other layer(s)] is applied

– Accounts and registers
– Address books
– Calendars
– Questionnaires and forms
Example of ‘overlaid data’
TEI/EAD Integration: Overlaps

• Overlaps between EAD and TEI occur in relation to metadata that:
  – Identifies, locates and describes the creation of the original archive document
  – Describes the physical characteristics of the original object
  – Provides contextual information about the creator of the original object and the participants within the object
  – Interprets/describes the data in the object
TEI/EAD Integration: Objectives

- Avoidance of repetition of information within EAD and TEI
- More complete description of an object’s physical characteristics
- Standardisation of contextual information regarding individuals/organisations
TEI/EAD Integration: search and retrieval issues

- Elements that interpret/categorise the actual data within the object

<table>
<thead>
<tr>
<th>EAD Elements</th>
<th>Overlapping TEI elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name of immediate parent</td>
<td>Name of immediate parent element(s)</td>
</tr>
<tr>
<td>child elements</td>
<td>Name of child elements</td>
</tr>
<tr>
<td>&lt;controlaccess&gt;</td>
<td>&lt;profileDesc&gt; within &lt;teiHeader&gt;</td>
</tr>
<tr>
<td>&lt;genreform&gt;</td>
<td>&lt;keywords&gt;&lt;classcode&gt;&lt;classref&gt;</td>
</tr>
<tr>
<td>&lt;geogname&gt;</td>
<td>&lt;textDesc&gt; within &lt;profileDesc&gt; within &lt;teiHeader&gt;</td>
</tr>
<tr>
<td>&lt;persname&gt;</td>
<td>&lt;channel&gt; &lt;constitution&gt; &lt;domain&gt; &lt;factuality&gt; &lt;preparadness&gt; &lt;purpose&gt;</td>
</tr>
<tr>
<td>&lt;famname&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;corpname&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;occupation&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;subject&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;date&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;function&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;scopecontent&gt;</td>
<td></td>
</tr>
<tr>
<td>N/A</td>
<td>&lt;text&gt; &lt;name&gt;&lt;date&gt;</td>
</tr>
</tbody>
</table>
LEADERS toolset

- **XML Schema**
  - Use of namespaces
  - Schema will provide a generic and re-usable means to encode resources

- **XSLT/CSS**
  - Style sheets will provide a means to manipulate, transform and present the encoded resources, thus supporting re-purposing of encoded materials

- **WSDL/SOAP**
  - Incorporating ‘self-describing services’ will allow multiple applications to be constructed to make different use of the encoded materials
Sample application

- Demonstrator – a sample application to show what can be produced/generated from the encoded materials and the toolset
- Basic search and retrieval and alternative presentations to show the possibilities of TEI/EAD encoded resources
- Used to gather feedback from users which will guide us in further design and development