Annotating Across Resource Formats Within Electronic Scholarly Editions

Damien Ayers, Anna Gerber, Roger Osborne, Jane Hunter

The University of Queensland, Brisbane, Australia,
{d.ayers | a.gerber | r.osborne | j.hunter}@uq.edu.au

INTRODUCTION

The AustESE project is developing an extensible, open source Workbench to support the production of electronic scholarly editions. Annotations are an important aspect of that workflow. A common practice amongst scholars since the age of print, annotation is now considered an important ‘scholarly primitive,’ previously consigned to personal comments on single items. Annotation in the digital age has been extended to facilitate collaborative analysis of distributed, multimedia digital resources. The AustESE Workbench adopts a component-based architecture to integrate several scholarly editing tools. This means that resources such as transcriptions and images of versions of literary works are displayed via a variety of views within the Workbench. The presentation of these resources (e.g., rendering them from a TEI/XML document) may be different for each view. For example, for the collation tools, markup is displayed to indicate textual variation between selected versions (additions, deletions, transposition). Most annotations within an electronic edition refer to parts of resources such as individual words or paragraphs. To support annotation across the different versions, tools and views encapsulated within the Workbench, we need to be able to capture, describe and display annotation selections consistently across all representations of a resource (including across different formats, e.g., HTML or PDF). Existing web annotation tools such as LORE typically use XPointers to represent selections. This approach is too fragile for our purposes as XPointers depend on the underlying markup structure to be static and cannot address multiple document formats. This abstract describes how we have overcome this limitation to by basing the AustESE Annotation tools on the Open Annotation data model.

OPEN ANNOTATION ACROSS RESOURCE FORMATS

We have adopted the Open Annotation Data Model [1] to represent annotations for AustESE. Unlike existing approaches of embedding annotations directly in TEI/XML representing a version, OA annotations are stored separately to the resources themselves, so that annotations can be attached without compromising the integrity of the original resource, and can potentially target multiple resources. Figure 1 shows an annotation that provides a choice of OA selectors to identify a segment of a transcription of a version. The target source is a URI identifying the targeted version that is not tied to any particular representation. The values for the TextQuoteSelector and TextPositionSelector are calculated based on the textual content of the version regardless of format, and the content is normalized so that characters such as line breaks, which vary between representations, are stripped out. Adding the austese:RangeSelector improves performance, as the XPointers do not need to be recalculated for the original view in which the annotation was attached, while the general selectors based on offsets and text quotes can be used for other views.

IMPLEMENTATION

OKNF’s Annotator provides a solid foundation for our annotation tool. It provides a cross-browser compatible UI for selecting regions of text, and for creating and viewing annotations; along with a robust extension mechanism. We have extended this tool by creating several plugins to support the specific annotation requirements of scholarly editors within AustESE. The plugins that we have developed are outlined in Table 1. The user interface for creating and editing Annotations is shown in Figure 2.

---

**Figure 1: Open Annotation model for Textual Note**

[Diagram showing the Open Annotation model for Textual Note]

---

**Table 1:** Plugins developed for AustESE

<table>
<thead>
<tr>
<th>Plugin Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TextQuoteSelector</td>
<td>Selector for quotes, supports multiple formats</td>
</tr>
<tr>
<td>TextPositionSelector</td>
<td>Selector for positions within text</td>
</tr>
<tr>
<td>RangeSelector</td>
<td>Selector for ranges of text</td>
</tr>
<tr>
<td>AustESE:RangeSelector</td>
<td>Performance-enhanced version of RangeSelector</td>
</tr>
</tbody>
</table>

---

**Figure 2:** User interface for creating and editing Annotations

---

20–25 Oct | 2013

Brisbane | Australia

7th eResearch Australasia Conference
Table 1: AustESE Annotator Plugins

<table>
<thead>
<tr>
<th>Plugin</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LoreStore</td>
<td>Maps the internal JSON data structure used to represent annotations in the Annotator system to an equivalent OA JSON-LD serialization. Also implements a protocol for communicating with our OA-compliant annotation repository.</td>
</tr>
<tr>
<td>Image</td>
<td>Enables the selection and annotation of rectangular regions of images. This plugin identifies an image using HTML5 data attributes that associated with each image rather than by the image filename. This allows us to display annotations across all views or renderings of the same base image. The annotation selection boundaries are stored as relative percentages of the image width and height, rather than absolute pixel values, since the images may be served to the browser in different sizes.</td>
</tr>
<tr>
<td>CharacterRange</td>
<td>This plugin calculates character offsets and text selections used to identify the part of the (text-based) resource being annotated across representations.</td>
</tr>
<tr>
<td>Motivation</td>
<td>We have extended the default set of OA Motivations that can be used to categorize annotations for filtering or search purposes. This Plugin allows users to attach these custom Motivations (e.g. ExplanatoryNote, TextualNote, Glossary) to annotations.</td>
</tr>
<tr>
<td>Reply</td>
<td>Enables discussion between editors, scholars and readers.</td>
</tr>
<tr>
<td>Prov</td>
<td>Records and displays provenance (creator, creation time) for an Annotation.</td>
</tr>
</tbody>
</table>

We have developed a Drupal module that integrates the Annotator tool and our plugins for any of the views within the AustESE Workbench. This module also includes discovery pages for annotations, which allow searching by keyword, creator, target and motivation. It also provides Atom feeds so that users can subscribe to Annotations by creator, target, subject, tag etc. Annotations are stored using our secure lorestore repository, which provides a REST API for storing, searching, retrieving and validating annotations. By default lorestore can create and manage user accounts, however as AustESE users are already managed by Drupal, we extended the repository to provide an option to delegate all authentication to Drupal so that permissions for annotations can be managed using the same roles and groups as the other tools within the Workbench.

![Figure 2: AustESE Annotation tool](image)

DISCUSSION AND FUTURE WORK

Our approach allows scholarly editors to attach and view annotations through any view or tool available within the AustESE Workbench. The Hypothesis [2] project, which also extends OKFN’s Annotator, has been working on similar solutions to allow annotations on the same article to be visible when it is published in different formats online. Like our system, their approach stores a unique identifier for the resource, with a copy of the annotated text, and then attempts to use a fuzzy matching algorithm to re-attach the annotation if the underlying text changes. We intend to integrate their fuzzy matching algorithm into our tools to enable annotation across different representations of dynamically changing resources. This will also enable us to display annotations across different versions of a work, by applying fuzzy matching and integrating it with the Multi-Version-Documents used to represent versions within AustESE.

All of our annotation tools have been developed under an open source license and have been designed to be generic enough to be used by other digital humanities projects. We will be contributing relevant plugins back upstream to the OKFN Annotator project to help grow the supporting community.

AustESE is funded by the Australian National eResearch Collaboration Tools and Resources (NeCTAR) eResearch Tools program

REFERENCES

ABOUT THE AUTHORS

**Damien Ayers** is a software engineer working on the AustESE project, developing eResearch tools to support the collaborative authoring and management of electronic scholarly editions. Previously he developed a new online catalogue for the UQ Anthropology Museum, with follow up work in collections management for the Archaeology division of the School of Social Sciences. He is also developing annotation tools and services based on the OA data model to support annotations in the context of collaborative development of scholarly editions. From March 2009 – July 2011, Damien worked as a software engineer for the Aus-e-Lit project.

**Anna Gerber** is the senior software engineer for the AustESE project, developing eResearch tools to support the collaborative authoring and management of electronic scholarly editions. She is also part of the Open Annotation Collaboration, developing annotation tools and services that can be used to evaluate and demonstrate the applicability of the OA data model in the context of collaborative development of scholarly editions. From July 2008 – July 2011, Anna was the senior software engineer for the Aus-e-Lit project. She had previously worked on the HarvANA: Harvesting and Aggregating Networked Annotations and AnnoCryst for PyMOL projects. Prior to joining the eResearch Group at The University of Queensland School of Information Technology and Electrical Engineering, Anna was a Research Scientist at DSTC from 2000 – 2005, focusing on Enterprise Modeling.

**Dr Roger Osborne** has published widely in the fields of book history, print culture and textual criticism. He completed a PhD at the UNSW in 2000 and was a Postdoctoral Fellow in the Australian Studies Centre, University of Queensland, from 2004-2007. He was Project Manager of the Aus-e-Lit Project from 2008-2011. He is co-editor of the Cambridge Edition of Joseph Conrad’s *Under Western Eyes* (forthcoming 2013) and as the 2011 Nancy Keesing Fellow at the State Library of New South Wales, he is working towards an electronic edition of Joseph Furphy’s *Such is Life*. Roger is also conducting preliminary research for an edition of Joseph Conrad’s modernist classic, *Nostromo*.

**Prof Jane Hunter** is the Director of the eResearch Group at the School of ITEE, the University of Queensland. She was also a Chief Investigator on the Open Annotation Collaboration funded by the Andrew Mellon Foundation and the technical lead of the AustESE project. She is also the Vice-President of the Australasian Association for Digital Humanities.