LORE: An Open Source Research Tool for Australian Literary Scholars

Anna Gerber

UQ ITEE eResearch Lab

http://itee.uq.edu.au/~eresearch

DIAS-B Danno
AustLit is a partnership between the National Library of Australia and twelve universities.

Research activities focused around communities.

Aus-e-Lit services extend AustLit web portal.
The Aus-e-Lit project is developing digital research tools and services for the Australian Literary Studies Community:

- Full Text Search
- Federated Search
- Data Aggregation and Visualization
- Compound Object Authoring
- Annotations

LORE
Compound Object Authoring

Allow scholars to:

- **Gather** web resources, organise and tag them, describe links between them and make notes about them.
- **Publish** this information to shared repositories as compound objects.
- **Discover** and reuse compound objects that others have created about related topics and web resources.
- **Present** information from compound objects as slideshows and diagrams.
Annotations

➡ Enable scholarly annotation of textual documents, web pages and images:
- Free-text comments, notes, analysis etc
- Threaded discussions
- Textual variations
- Fine-grained attachment of metadata, corrections and tags to augment AustLit records

➡ Allow annotations to be shared and re-published e.g. in electronic editions
LORE: Literature Object Reuse and Exchange

- LORE
  - GPL 3.0
  - Firefox Extension
  - Javascript/HTML/XUL

- Annotation server (Danno)
  - GPL 3.0
  - Java / Tomcat

http://sourceforge.net/projects/metadata-net/
Other considerations

- Publish data online, facilitate discovery, re-use
- Allow data to be re-purposed or migrated to other systems: avoid silos
- Tools must be able to be installed in an institutional environment (e.g. library) by non-privileged, non-technical users
- Scholars work with digitised texts, manuscripts and other digital resources distributed online in archives & institutional repositories
- Users preferred a browser-based solution

Open Source + Open Data

Browser extension
Linked Data

- Available on the web
- Open licence
- Machine readable, structured data
- Non-proprietary format
- Standards e.g. W3C's RDF
- Link data to provide context

Semantic Web Standards

Resource Description Framework (RDF)

Web Ontology Language (OWL)

- Define vocabularies with formal semantics
- LORE can be customised by specifying alternate ontologies to provide metadata terms used in compound objects and annotations
Object Reuse and Exchange

- Open Archives Initiative Object Reuse and Exchange (OAI-ORE): http://openarchives.org/ore/

- Collaborative initiative focused on developing standardized, interoperable, machine-readable mechanisms to express compound objects on the web

- Resource maps
  - express a set of resources that are aggregated by a compound object
  - express metadata and relationships between aggregated resources
  - may be published as RDF Named Graphs
ORE Compound Object

OAI-ORE Named Graphs/Resource Maps:
- Define set of components
- Metadata attached to compound object
- Typed Relationships between components
- Different views of the compound object

http://austlit.edu.au/rem/...
AustLit Ontology

- Entities represent products of intellectual or artistic endeavour (Publications)
  - Work
  - Expression
  - Manifestation
  - Item
- Responsible Persons and Organisations (Agents)
- Subjects (concepts, objects, events, places)
- Based on IFLA Functional Requirements for Bibliographical Records (FRBR) standard
LORE relationship ontology
Open Annotation Collaboration

- Aims to enable the sharing and interoperability of scholarly annotations across annotation clients, collections, media types, applications and architectures.

- Provides an RDF-based model for annotations on web resources.

- Aus-e-Lit Annotation client uses OAC model for scholarly annotations.

http://www.openannotation.org/
Annotating Textual Variation
Variation Annotation extends OAC Annotation to allow complex annotation bodies using OAI-ORE

- Relates selections from variant texts
- Records variation location, date and agent responsible
- Records notes
Community Contribution

- Allow the research community to contribute to and comment on AustLit content through annotation.

- Existing web annotation approaches:
  - rely on W3C media fragments or XPointers to identify the context of an annotation.
  - are unstable when presentation mark-up is dynamically generated.

- We annotate the underlying data entities, which can be encoded in HTML record pages using RDFa.
Identifying AustLit Entities

The Drover's Wife

RDFa

```html
<div typeof="austlit:Work" id="CMCg" about="#CMCg" property="austlit:topicID" content="CMCg">
  <span rel="austlit:hasTitle">
    <span typeof="austlit:Title" id="SPJi" about="#SPJi" property="austlit:Ftle">
      The Drover's Wife
    </span>
  </span>
  <span rel="austlit:form">
    <span typeof="austlit:Form" about="http://austlit.../ns#i$" property="austlit:topicName">
      short story
    </span>
  </span>
</div>
```

```
<spn rel="austlit:hasTitle">
  The Drover's Wife
</spn></spn>
```

```
<spn rel="austlit:form">
  short story
</spn></spn>
```

```
<spn rel="austlit:usedInput">
  Two versions:
</spn></spn>
```

```
<spn rel="austlit:producedOutput">
    Find the full text at University of Queensland Library.
  - Short Stories in Prose and Verse: Selected Work, Poetry, Prose, Short Story.
    Author: Lawson, Henry (birth name: Larsen, Henry (Lawson)) (a k a. H. L.)
    Sydney, New South Wales: Louisa Lawson, [1894]. (pp.25-39)
</spn></spn>
```

Work

Agent

Subject

Expression

Manifestation
Annotating AustLit Entities

Tags from AustLit thesaurus
Other LORE use cases

- Teaching
  Relating disparate resources around a common theme

- Scholarly Editions
  Compound objects encapsulating different versions of a text, annotations and scholarly commentary

- Research Trails
  Recording research sources and notes

- Lineage
  Tracking the lineage of derivative works and ideas
Open Data Challenges

- Persistent identification of resources found online
- Need for archives
  - Copyright issues
- Trust and quality
Firefox Add-on Development Challenges

- Flaky developer tools
- Difficult to debug and test
- API updates
- Security!
- HTML 5 vs XUL extensions:
  - Chromium
  - Jetpack project's Add-on SDK
Further information

Anna Gerber
a.gerber@uq.edu.au

LORE is released under the GPL 3.0

http://sourceforge.net/projects/metadata-net/
http://itee.uq.edu.au/~eresearch/
http://www.austlit.edu.au/
Acknowledgments

Aus-e-Lit
Collaborative Integration and Annotation Services for Australian Literature Communities

Professor Jane Hunter, eResearch Lab, Information Technology and Electrical Engineering (ITEE), UQ
Kerry Kilner, Director of AustLit Research and Publications, English, Media Studies and Art History (EMSAH), UQ
Dr Roger Osborne, Aus-e-Lit Project Manager, EMSAH
Anna Gerber, Senior Software Engineer, eResearch Lab, ITEE
Damien Ayers, Software Engineer, eResearch Lab, ITEE

Aus-e-Lit is building on the eResearch infrastructure for Australian literary studies at AustLit: The Australian Literature Resource. Aus-e-Lit is funded by the National Collaborative Research Infrastructure Strategy (NCRIS) through the National eResearch Infrastructure Task Force (NeAT), and by the University of Queensland.