

This paper was presented at the conference:

Oxford eResearch 2008

11-13 September 2008

University of Oxford

Conference website:

<http://www.oii.ox.ac.uk/microsites/eresearch08/index.cfm>

Conference papers collection:

<http://ora.ouls.ox.ac.uk/objects/uuid%3A64aa6f39-7e81-4d42-a008-ee2d7524bd67>

Conference organisers:

Oxford Internet Institute

University of Oxford

1 St Giles, Oxford OX1 3JS

<http://www.oii.ox.ac.uk/>

email: enquiries@oii.ox.ac.uk

Oxford e-Research Centre

University of Oxford

7 Keble Road, Oxford OX1 3QG

<http://www.oerc.ox.ac.uk/>



Transforming the Study of Australian Literature through a Collaborative eResearch Environment

Kerry Kilner¹, Anna Gerber², Jane Hunter²

¹AustLit, Faculty of Arts, The University of Queensland
²School of Information Technology and Electrical Engineering,
The University of Queensland, St Lucia, Australia

Abstract:

A unique collaborative eResearch programme has been transforming Australian literature research activities in Australia over the past eight years and is breaking new ground in the methodologies being used to interpret, analyse and disseminate data within a wide array of print culture and literary history projects. This paper will present a case study – illustrating how e-research is being enthusiastically embraced by literary scholars in Australia through the *AustLit*¹ portal. It will demonstrate the techniques that AustLit employs in data management, data analysis and collaborative authoring techniques, which ensure authority and reliability of content being developed. It will also describe how a sense of ownership has been established by the scholars who are participating in building AustLit, and how this has led to the widespread adoption and success of this collaborative eResearch environment.

Keywords: Australian Literature, eResearch, Annotation, Collaboration

1 Introduction and Background

In his recent article, ‘What’s Going on?’ [1], Willard McCarty discusses the “new thinking and new institutional structures” required to support and facilitate emergent practice in the humanities. What is required, he says, is the creation of an environment (virtual and institutional) where the imaginative exploration of new possibilities for divergent scholarly outcomes through computing in the humanities can occur. This type of imaginative exploration, unlike purely text-based research, relies on new forms of creative interdisciplinary collaboration between humanities academics and computer scientists. It also requires new ways of thinking about scholarly communication and recognised research outcomes and the forms it can take in a networked digital world. These are fundamental challenges to traditional humanities

In its own specific but significant way, the AustLit project is engaging with these challenges and new requirements and with emerging paradigms of interdisciplinary collaboration.

AustLit is a partnership between twelve Australian universities and the National Library of Australia. Providing authoritative information on hundreds of thousands of creative and critical Australian literature works relating to more than 100,000 Australian authors and literary organisations from 1780 to the present day, it is arguably the most important resource for scholars undertaking foundational research into many aspects of Australian literary heritage and print culture history [2].

¹ <http://www.austlit.org/>

Led by the University of Queensland, AustLit supports the activities of numerous research communities around Australia and, in some cases, internationally. AustLit indexes and describes Australian literature published in a range of print and electronic information sources. It also makes available selected critical articles and creative writing in full text. Its specific aims are to:

- to enhance and support research and learning in Australian literature;
- provide a discovery service for Australian literature research;
- integrate existing Australian literature infrastructure and related content;
- provide opportunities for online collaborative research activities.

AustLit serves the research, teaching and general information-seeking communities as both a source of basic and enhanced information and as a mechanism for the extraction and dissemination of research data which is becoming mandated by Australian funding bodies.

2 The Aus-e-Lit Project

Through the Australian National Collaborative Infrastructure Strategy (NCRIS), the National eResearch Architecture Taskforce (NeAT) was established in 2007 to identify a set of projects that will develop eResearch methods and tools for specific research communities and disciplines. One of the inaugural projects to be funded, is the Aus-e-Lit² project – a 2-3 year collaboration between the existing AustLit project, the UQ eResearch Lab and the ASAL (Association for the Study of Australian Literature).

The aim of the Aus-e-Lit project is to address the eResearch needs of researchers involved in the study of Australian literature and Australian print culture. The project will enhance and extend the existing AustLit web portal through the incorporation of the following services:

- data integration and search services
- empirical reporting and innovative visualization services
- specific support services for Research Sub-communities
- collaborative scholarly markup and annotation services
- compound object authoring, editing and publishing services

Figure 1 provides a high-level view of the anticipated extensions to the AustLit portal.

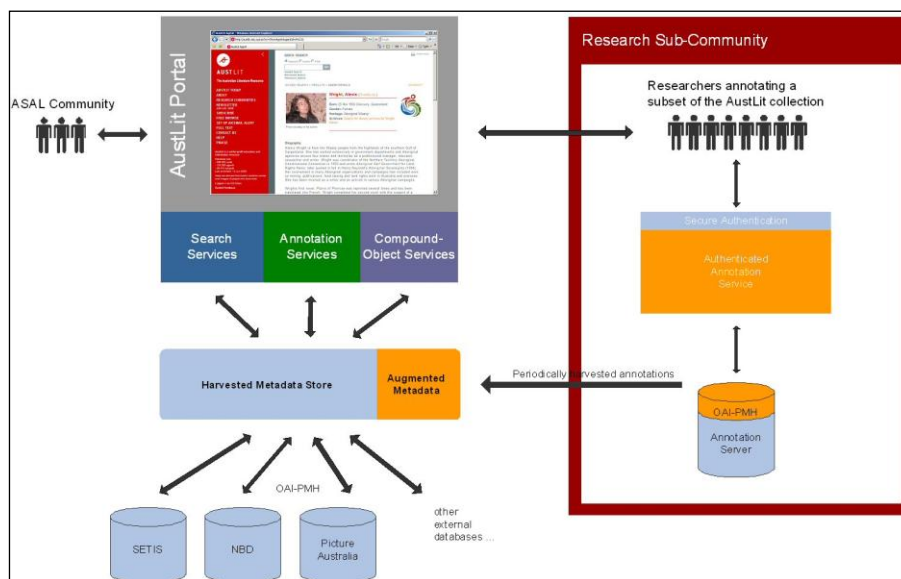


Figure 1: High Level View of the AustLit System Components

² <http://www.itee.uq.edu.au/~eresearch/projects/aus-e-lit/>

The remainder of this paper describes and discusses the specific requirements and details of the services under development.

3 Integrated Access to Federated Databases

AustLit currently provides users with web-based access to the following types of information:

- Constantly updated citations for monograph, journal and electronic publications by and about Australian authors;
- Biographical information about authors including pseudonyms and alternative names;
- Subject indexing of most works ;
- Links to available full text.

The majority of AustLit information is produced and updated by a team of expert librarians and bibliographic researchers, overseen by an Editorial Board of distinguished academics and information professionals.

However, significant value could potentially be added by incorporating related information stored in external, curated databases. Hence, one of the high priority activities currently being carried out through the Aus-e-Lit project is to expand the set of databases and collections, accessible and searchable through the AustLit Portal. In particular, the aim is to add more multi-media resources (full-text, biographies, artists and illustrators, performances, recordings and images) to the AustLit collection and to specify semantic associations within the collection through links to authors, works, subjects, events and places. Where possible, the plan is to store a local copy of the resource on the AustLit server but alternatively, the plan is to harvest the metadata (using OAI-PMH) to enable search and discovery of the related resource, stored on a remote server.

Discussions are currently underway with the custodians of the following collections to resolve data access agreements, IP issues and technological challenges associated with developing a federated search interface that incorporates these databases:

- [SETIS: Scholarly Electronic Text and Image Service](#)
- [Libraries Australia National Bibliographic Database \(NBD\)](#)
- [Register of Australian Archives and Manuscripts \(RAAM\)](#)
- [Australian Dictionary of Biography Online](#)
- [Dictionary of Australian Artists Online](#)
- [Australian Digital Theses DB](#)
- [People Australia](#)
- [Picture Australia](#)
- [PANDORA](#)
- [AusStage](#)

4 Empirical Reporting and Innovative Visualization Services

AustLit is central to an emerging trend in Australian literary research called the “new empiricism” as its large data-set (relatively-speaking) of over 1 million analysable data elements allows the types of research only dreamed about by earlier generations of scholars. It is now becoming possible to answer deceptively simple questions such as how many works of fiction were published in a particular period of Australia’s history or the comparative numbers of works for children by men and women and myriad other questions that reveal the value of literary culture to a nation across time.

Recent extensions to the AustLit portal include the ability to display the results of empirical and statistical analyses as graphical visualizations embedded in the Web browser e.g., number of works by genre, authors by country of birth etc. over the past 100 years. Figure 2 displays the results of an example empirical query – “*Show me a graph of the number of works of each genre*”, dynamically generated and displayed graphically via the Web Browser.

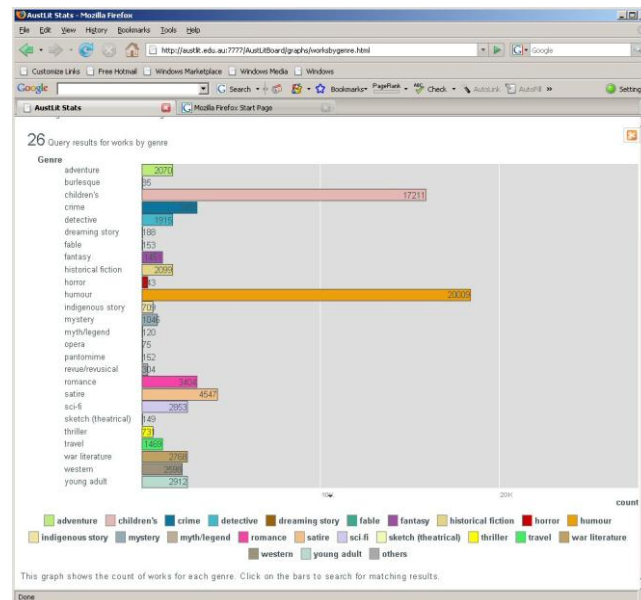


Figure 2: Graph showing works of Australian Literature of each genre

In addition, we have also been exploring the use of maps and timelines to display spatio-temporal and semantic relationships between authors and publications. Prototype queries and visualizations have been implemented using Simile Exhibit³ - an open source web application framework written in Javascript that enables the easy mash-up of data from various sources onto maps, timelines and scatter plots etc displayed through interactive web pages. Figure 3 is a Web-based visualization showing the birthplace and birthdate of authors associated with Australian Popular Theatre. The birthplace is displayed via a personal icon located on a map and the birthdate is displayed on the timeline at the bottom of the page, under the map.

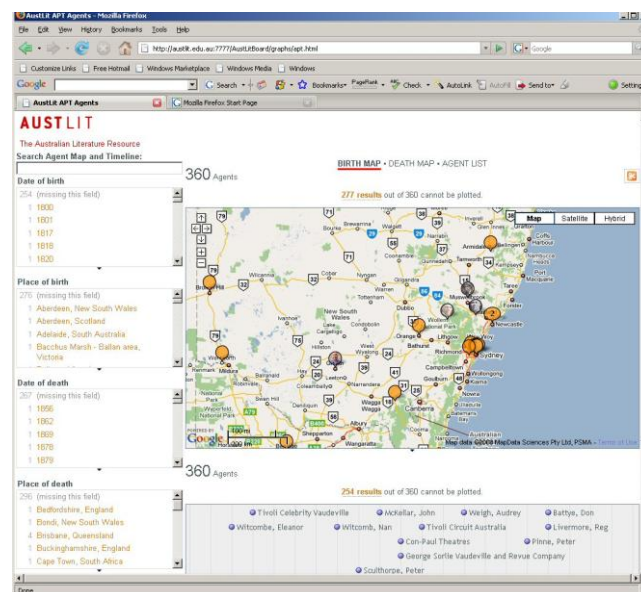


Figure 3: Map and Time-line based display of authors' birth places and dates

³ <http://simile.mit.edu/exhibit/>

5 Research Sub-Communities

AustLit also provides support for the establishment of research communities wanting to explore specific aspects of Australian literary and print culture. Researchers collaborate to identify subsets of the AustLit collection, of most relevance to their field. They then create highly annotated datasets of biographical, bibliographical and subject specific information or collections of full text to support their analytical work. These subsets, covering both broad and narrow subject areas, are published within AustLit as a subset of the whole AustLit dataset with introductions, accompanying essays and specialist scoped search functions⁴. In addition, within their virtual community space, research sub-communities are able to establish discussion forums, wikis and disseminate information about events and related projects. Currently supported research communities in AustLit, include:

- Literature for Children
- Black Words: Aboriginal and Torres Strait Islander Writers and Storytellers⁵
- Expatriate writers to Britain
- The Bibliography of Australian Literature
- 20th century Banned Books
- Multicultural Writers of Australia
- Australian Literary Responses to ‘Asia’
- Genre specialisations: Drama and Popular Theatre, Pulp Fiction, Magazines

6 Scholarly Markup and Annotation Services

Web-based collaborative annotation systems are designed to enable online communities of users to attach descriptive terms to Web resources as a way of organizing the content for future navigation, filtering or search. When applied to digital resources shared via the Web, annotations provide a very powerful collaborative tool - enabling the capture and dissemination of individuals’ and group opinions about particular digital resources.

Within the Aus-e-Lit project we will be developing services that enable communities of experts (sub-communities) to collaboratively select, tag and annotate specific digital resources, including full text documents, with keywords, notes, comments, interpretations, queries, links to related resources – and to share these annotations with their sub-community, thereby enriching the collection and enhancing the discovery services.

More specifically we will be focussing on providing the following:

- Tools to enable the attachment of tags (keywords) and annotations to resources within AustLit and the creation of responses to existing annotations;
- Support for the collaborative development of evolving ontologies (structured keyword lists) for specific sub-communities;
- Hosting of annotation servers for specific communities and the subsequent harvesting of annotations and social tags to enhance discovery services [3];
- Presentation, search and browse services for community annotations/tags;
- RSS feeds to notify communities of new annotations;
- Generation of tag clouds to display hot authors, hot topics and hot articles.

⁴ <http://www.austlit.edu.au/specialistDatasets>

⁵ <http://www.austlit.edu.au/specialistDatasets/BlackWords>

7 OAI-ORE Compound Object Authoring Services

Within the discipline of literature research and teaching, the ability to relate disparate digital resources in a standardized, machine-readable format has the potential to add significant value to distributed collections of literary resources. Such compound objects can be used to: track the lineage of derivative works which are based on a common concept or idea; or to relate disparate objects that are related to a common theme; or to encapsulate related digital resources for teaching purposes. For example, one might want to relate the original edition of Patrick White's *Voss* to the illustrated edition, a radio recording and a digital version of the film – and enable them to be retrieved and presented, with their relationships visualized, regardless of their location.

The Open Archives Initiative Object Reuse and Exchange (OAI-ORE) [4] is an international collaborative initiative, that aims to support the creation, management and dissemination of the new forms of composite digital resources being produced by eResearch and to make the information within these compound digital objects discoverable, machine-readable, interoperable and reusable.

Our objective is to develop a system that enables AustLit scholars to easily author, edit and publish compound information objects that comply with OAI-ORE - but which also represent the entities and relationships expressed in the IFLA Functional Requirements for Bibliographic Records (FRBR) (i.e., the concepts of *work*, *expression*, *manifestation* and *item*) [5]. The AustLit data model is based on the IFLA-FRBR model, extended to include *events*. To date we have developed a proof-of-concept prototype known as LORE (Literature Objects Re-use and Exchange).

Figure 4 shows the graphical user interface we have developed (as part of LORE) to enable the encapsulation of literary objects (discovered and retrieved via the AustLit search interface) within an OAI-ORE compliant compound object.

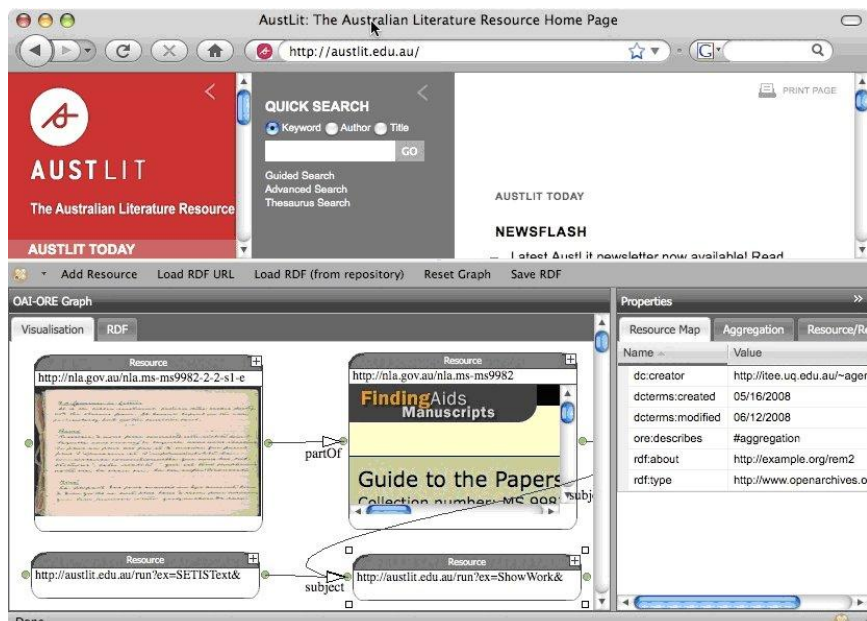


Figure 4: OAI-ORE Compound Object Authoring/Editing Interface

In the graphical representation, the *nodes* represent the individual resources aggregated via the resource map and the *arcs* represent typed relationships between the aggregated resources. The labels on the arcs are specified via a pull-down menu of predefined

relationship types. When complete, the compound objects may be saved (as a named graph in an RDF repository).

Further work and evaluation is required, but the feedback from the initial AustLit community has been enthusiastic. Future effort will focus on improving the robustness and usability of the system and implementing the following:

- Enabling OAI-ORE objects to be published within a Fedora repository.
- Enabling the discovery of compound objects through searches over the aggregated objects as well as over metadata terms and relationships.
- Enabling users to attach Creative Commons licenses to the compound objects, prior to publishing them, so the permitted types of re-use can be specified.
- Enabling the editing and re-use of OAI-ORE compound objects.

8 Conclusions

AustLit is an example of a successful arranged marriage between what many consider to be dire opposites: traditional bibliographical scholarship and new media technologies. One of the secrets to this happy marriage has been to highlight the value of each while not deriding the value of either. Engaging scholars from both camps has meant that we now have a universally approving and excited group of colleagues keen to engage with the collaborative research methodologies that Aus-e-Lit and related projects will enable.

However, the success of AustLit depends on more than just happy collaboration between humanities researchers and computer scientists. Additional aspects of AustLit that have ensured and will continue to ensure its success and sustainability include:

- Its low barrier subscription-based model that balances open access with sustainability;
- The commitment to a foundation of high quality data and metadata;
- The collaborative distributed nature of the content development and its peer review;
- The underlying data/metadata model (based on IFLA FRBR [1] and RDF) that ensures extensibility and support for evolving layers of knowledge and detailed provenance capture;
- Usage monitoring services that identify the frequency and extent of access, updates, additions and those areas of the database that are the focus of most recent activities.

Together all of these ingredients contribute to the generation of a showcase example of a successful eHumanities project – founded upon an evolving knowledge base built through collaborative and active participation by the communities it serves.

References

- [1] W.McCarty, “What’s Going On?” *Literary and Linguistic Computing*, 2008.
<http://staff.cch.kcl.ac.uk/~wmccarty/essays/McCarty,%20What's%20going%20on.pdf>
- [2] Kilner, K., *The AustLit Gateway and Scholarly Bibliography: A Specialist Implementation of the FRBR*. In: *Cataloguing and Classification Quarterly* 39.3/4 (2005)
- [3] J.Hunter, I.Khan, A.Gerber, *HarVANA – Harvesting Community Tags to Enrich Collection Metadata*, JCDL 2008, Pittsburgh, June 2008
- [4] Lagoze, C., Van de Sompel, H.: *Compound Information Objects: The OAI ORE Perspective* (2007) <http://www.openarchives.org/ore/>
- [5] IFLA Study Group on the Functional Requirements of Bibliographic Records, *Functional Requirements of Bibliographic Records: final report*, 1998