

Collaborative Integration and Annotation Services for Australian Literature Communities

Overview *Summarise the context and objectives of this project.*

The aim of this project is to provide services to address the eResearch needs of researchers involved in the study of Australian Literature. In particular it will focus on cyber-infrastructure to satisfy the needs of the ASAL (Association for the Study of Australian Literature) community. The aim of this project is to provide this comparatively under-funded research community with four new classes of services:

- data integration services;
- collaborative annotation services;
- compound object authoring, editing and publishing services;
- collaborative workspace creation services for sub-communities.

Users *Identify the research communities and resource providers that this project serves; and the potential number of users. This should include some NCRIS capabilities or other data federating or collaborating research groups, and any institutions that will participate through setting requirements for or steering this project.*

This project will primarily focus on serving those communities who study Australian Literature and Australian print culture. In particular it will work with the members of ASAL (Association for the Study of Australian Literature), the peak body for the discipline, who include individuals and research groups from the following:

- Schools of English, Humanities and Media at Australian Universities;
- AustLit: The Australian Literature Resource;
- Centres for Australian Studies world-wide;
- Indigenous studies centres in Australia;
- Australian Academy of the Humanities;
- [American Association of Australian Literary Studies](#);
- [SETIS](#) – the Scholarly Electronic Text and Image Service;
- [ARC Cultural Research Network](#);
- [European Association for Studies of Australia](#);
- [British Australian Studies Association](#);
- [Association for Commonwealth Literature and Language Studies](#);
- Australian Scholarly Editions Centre;
- [Centre for Australian Language and Literature Studies](#);
- [National Centre for Australian Studies](#);
- [The Westerly Centre for Australian and Asian Literature](#).

These organisations collectively cover virtually all academics, researchers and postgraduate students working in the field. The AustLit Web portal and online collection is the key resource for this field and will be the conduit for the project's facilities.

Other communities and resource providers that will participate include: the National Library of Australia, the University Libraries (Fryer Library, UQ), State Libraries, Australian Broadcasting Commission, National Film and Sound Archive. International communities who will be engaged to participate include: OAI-ORE, The Arts and Humanities Data Service (AHDS) in the UK, The British Library and the Library of Congress.

Online databases and catalogues that will be federated in the first phase will include: AustLit, SETIS, LibrariesAustralia National Bibliographic Database (NBD) and Australian Digital Theses DB. Additional databases to be included in future phases include: [Australian Dictionary of Biography](#)

Online, [Dictionary of Australian Artists Online](#) (UNSW), PeopleAustralia, PictureAustralia, [Newspapers Online](#), [National Registry of Audiovisual Collections](#), [OzLife](#), [Project Gutenberg Australia](#), colonial (biographical) databases in most states, the Australian War Memorial biographical databases,

Needs *Describe the needs of the research communities or resource providers that this project seeks to address.*

The members of ASAL and the Advisory Board of the AustLit project have identified three primary requirements with respect to eResearch services:

1. The ability to seamlessly search across a small number of databases and archives to retrieve reliable information on a particular author, topic or publication. The envisaged service is a secure web portal that federates search across:
 - o AustLit, <http://www.austlit.edu.au/>
 - o SETIS: Scholarly Electronic Text and Image Service
 - o LibrariesAustralia National Bibliographic Database (NBD),
 - o Australian Digital Theses DB <http://adt.caul.edu.au/>
 - o AusStage - <http://www.ausstage.edu.au/>
 - o [Dictionary of Australian Artists Online](#)
 - o [Australian Dictionary of Biography](#)
 - o NLA's PeopleAustralia, PictureAustralia
 - o [Project Gutenberg Australia](#),
2. The ability for closed communities of experts (sub-communities) to collaboratively tag/annotate 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, enrich the collection and enhance the discovery services. Examples of such sub-communities are shown here: <http://www.austlit.edu.au/specialistDatasets> and include topics such as: *children's literature; expatriate authors; indigenous literature*. Such communities have identified filtered subsets of the complete AustLit collection, that are relevant to their particular specialty.
3. The ability for community members to create/author Compound Objects (OAI-ORE compliant), publish them, edit them and search and retrieve them. OAI-ORE objects relate multiple resources from disparate databases within a single compound object – through RDF named graphs. They are particularly useful as online learning objects for teaching and training.

Services *Describe the result of the project in terms of the service(s) that will be implemented and demonstrated by the project and which could be operated in an ongoing fashion; and the proposed operator of each service.*

The services described below will be developed by the UQ eResearch group (under Prof Jane Hunter) working in close consultation with the AustLit staff (Kerry Kilner) and AustLit users. After comprehensive evaluation, refinement and deployment, the AustLit-specific services will be hosted and operated by the UQ Library (which currently hosts the AustLit server). The more generic services that may be deployed in disciplines other than AustLit will be hosted, supported and operated initially by UQ's eResearch group and in the long term by UQ's IT Services.

Year 1: Data Integration, Search and Reporting Services. These will comprise:

- OAI-PMH metadata harvesting services – that periodically harvest metadata from the specified external repositories and aggregate the metadata with the AustLit metadata repository. This service will be developed by the eResearch Group at UQ and deployed for each external collection to be added to the Federated search interface.

- Ability to easily specify a new external collection to be added – the location of the server, the schema of the metadata to be harvested and the frequency at which updates should be harvested.
- Metadata Schema/Ontology Registry - enabling the submission, validation, description, version control of metadata schemas associated with external collections and ontologies and terminologies associated with specialist sub-communities or topics
- Federated Search Portal for ASAL that supports:
 - Searching across metadata terms as well as full-text searching across corpuses
 - Search, retrieval and presentation of records, text and images (thumbnails)
 - Quantitative reporting services –responses to queries such as: give me a bar graph showing how many short stories have been published by women writers in the 1980s that include the phrase “a dingo ate my baby”

Year 2: Secure, Collaborative Annotation services. Authenticated users need to be able to add layers of annotations to original texts in order to create scholarly editions. This requires the following services:

- Secure annotation creation, editing and attachment services;
- Annotation presentation, browse and search services that enable users to:
 - See who has attached which annotation when and the sequence of responses.
 - Search across the content of the annotations or for annotations by particular individuals
 - RSS notification services – so users are notified of new annotations on particular texts or by particular individuals
- Annotation storage and servers (with authenticated access control)
- Annotation harvesting and aggregation services

In addition, will also be investigated as a tool for creating collaborative workspaces for sub-communities (e.g., expatriate authors, or indigenous literature).

Deleted: in Year 2, Sakai

Year 3: Compound Object Authoring, Editing and Publishing services. These services will be available through the AustLit Web portal and enable users to:

- Create OAI-ORE compliant compound objects that encapsulate a set of URIs to related objects plus typed relationships between them (as RDF graphs)
- Named RDF graph visualization and editing tool
- Metadata attachment and publishing service for OAI-ORE objects – enabling their upload to a Fedora repository
- Search, retrieval and presentation interface for OAI-ORE objects based on both metadata terms plus component objects.

NeAT Characteristics

eResearch effect *What changes in behaviour and activity are expected from the project that will demonstrate the broader adoption of eResearch practice?*

The overall effect of the proposed services will be to expedite the widespread adoption of eResearch practises by humanities researchers in Australia.

The project will enhance the richness of the existing AustLit online research facility, and enhance the interoperability, discoverability and re-use of the resources. The proposed services will enable easier sharing of resources and capture of knowledge within the communities and sub-communities that use the AustLit online resource collection – through novel user-driven technologies.

The proposed services will enable the collaborative authoring of new content to be delivered through AustLit in a more efficient and streamlined manner. The establishment of specialist sub-communities working on specific sub-sets of resources using a set of services tailored for that sub-community will facilitate greater collaboration and richer, deeper research outcomes.

Broader adoption *Which additional communities, resource providers or organisations would also be expected to benefit from the provision of the same or similar services should the project succeed?*

The services developed within this proposed NEAT project would also be of benefit to:

- NCRIS 5.12 Marine Sciences and Climate
- Social Sciences, ASSDA
- Atlas of Living Australia, Terrestrial Ecosystem Research Network (TERN),
- NCRIS 5.3 Microscopy and Microanalysis
- NCRIS 5.8 Bio-security

Value adding *Identify the components of the project that could be based around generic technologies or be implemented through shared services for which the project would provide an exemplar use case or requirement set.*

- Metadata harvesting and mapping services
- Metadata schema, ontology and controlled vocabulary registry
- Data integration and search services
- Collaborative Annotation and tagging services
- Compound Object Authoring and Publishing services

Standardisation *Describe the global technology development or standardisation work that will be adopted, adapted or extended within the project and any risk reduction available by collaboration with similar activities occurring elsewhere in the world.*

- W3C XML Schemas and RDF Schemas
- W3C RDF and OWL ontologies
- W3C Annotea protocol
- Dublin Core, ANZLIC
- JISC IEMSR (UKOLN - Emma Tonkin, Rachel Heery, Liz Lyons)
- NSDL Metadata Registry (Dianne Hillman, Cornell)
- eXtended Metadata Registry (XMDR) Project (Bruce Bargmeyer, LBNL)
- ISO/IEC 11179 Metadata Registries (MDR)
- Metadata Open Forum <http://metadataopenforum.org/>
- OAI Object Re-use and Exchange ORE - <http://www.openarchives.org/ore/>

Project Scoping

Key Participants *Name any Pfc components, any NCRIS capabilities, or any other institutions or groups that will need to be involved in the project planning and execution.*

The main community that this project will serve is ASAL (Association for the Study of Australian Literature). ASAL includes individuals and research groups from the following:

- Schools of English, Humanities and Media at the majority of Australian Universities
- AustLit: The [Australian Literature Resource](#)
- Australian Academy of the Humanities
- Australian Studies Centres worldwide
- [American Association of Australian Literary Studies](#)

- [European Association for Studies of Australia](#)
- [British Australian Studies Association](#)
- [Association for Commonwealth Literature and Language Studies](#)
- [Australian Scholarly Editions Centre](#)
- [Centre for Australian Language and Literature Studies](#)
- [National Centre for Australian Studies](#)
- [The Westerly Centre for Australian and Asian Literature](#)
- [Australian Public Intellectual Network \(API\)](#)

Other communities and resource providers that will participate include: National Library, National Archives, University Libraries (Fryer Library, UQ), State Libraries, Australian Broadcasting Commission, National Film and Sound Archive. International communities who will participate include: OAI-ORE, The Arts and Humanities Data Service (AHDS) in the UK, The British Library and the Library of Congress.

Governance

Outline the arrangements proposed to manage the contributions and user interaction. Eg: a project managed by ARCS, ANDS, or under another NCRIS capability governance, or by subcontract to a named lead agency, or a new J/V between the parties.

Lead Agency – the University of Queensland

Service Provider/Host:

- the **UQ Library** will host the expanded AustLit services (the existing AustLit server and associated services are developed by AustLit/UQ eResearch staff and hosted by the UQ library)
- the **eResearch Group at UQ** will develop, test and deploy the proposed services. After comprehensive system and user evaluation, the AustLit-specific services will be hosted by the UQ library
- More generic eResearch services will be initially supported, maintained and hosted by the eResearch group at UQ but in time they will move to **IT Services at UQ**

Steering Committee

- Kerry Kilner (AustLit)
- Jane Hunter (UQ eResearch)
- Ross Wilkinson (ANDS)
- Paul Coddington (ARCS)

Advisory Committee (includes the Steering Committee as well as):

- Elizabeth McMahon, ASAL President
- Dr Philip Mead –ASAL
- Keith Webster, UQ Librarian
- Ross Coleman, SETIS, University of Sydney
- Paul Egert – University of NSW
- Margie Burn, Warwick Cathro - NLA
- Gavan McCarthy – eScholarship Research centre
- Willard McCarty – Professor of Humanities Computing, Kings College, London
- Carl Lagoze – OAI-ORE
- Sheila Anderson – Director, Arts and Humanities Data Service

Project Scale *Identify the overall scale expected in the project, eg. 1 to 3 years, total effort in any year, and nominate any parties that have indicated a willingness to participate through providing resources. (funded or in-kind, people or facilities).*

The AustLit eResearch project will be a 3 year project that builds extensively on existing investments as well as significant inkind and cash contributions from both AustLit and the Uni. of Queensland.

- Builds on AustLit ARC LIEF (Linkage Infrastructure Equipment and Facilities) grant (2006-2008) and AustLit enhancement and augmentation project.
 - o In-kind [from existing software that has been developed from this project](#)
 - o [In-kind of 20% of Kerry Kilner's time and 10% of Jane Hunter's time](#)
- Builds on current Carrick Institute Project – Australian Literature Teaching Survey and the Teaching Experience Database
 - o In-kind \$100,000 pa 2007-2009
- Builds on DART/ARCHER/PILIN/RUBRIC investments in
 - o Semantic Web-based data integration and RDF Triple store development
 - o Annotation services
 - o OAI-ORE services
- Leverages QCIF and State Govt funding - \$35,000 pa (hardware and storage)
- **\$105K cash over 3 years (\$35,000 pa) from UQ.**
- UQ Library - \$20,000 pa for hosting the server/services

← Formatted: Bullets and Numbering

A vital component of the project will be outreach and promotion into the scholarly community. Hence a Project Officer – HEW 6 – who can liaise, administer the project, communicate outcomes and opportunities to the sector will be essential. In addition, we are requesting two full-time software developers.

The project will be undertaken as a 3 year project. EFT per annum are estimated as follows:

Description of staff	Participant	In-kind UQ+AustLit	Cash UQ	NeAT
Project Leadership	AustLit + UQ eResearch	1.0		
Admin + Outreach and education	Project manager		0.5	0.5
Systems Development	2 software engineers			2.0

Estimated total costs on a per annum basis are as follows:

Description of resource (staff, facilities, etc)	AustLit	University of Qld	QCIF	NeAT
Total staffing as above (average \$100k per EFT)	\$50,000 inkind	\$50,000 pa inkind \$35,000 pa cash		\$250,000
Hardware Storage arrays and maintenance			\$15,000 \$20,000	
Travel (national and international)				\$15,000
TOTAL	\$150,000pa	\$185,000pa	\$35,000pa	\$265,000 pa

Major Steps

Identify the key steps that will be visible to users as the services develop. Note that some observable deliverable is needed every half year and projects may be reviewed based on the achievement of these steps.

Year 1 – integrate 5 key databases, build and support federated search interface: AustLit, SETIS, , Australian Dictionary of Biography Online, Dictionary of Australian Artists Online, National Library services - National Bibliographic Database (NBD), People Australia, Picture Australia, PANDORA, and Australian Digital Theses DB.

Year 2 – develop, test and deploy annotation services and extend number of integrated databases

Year 3 – provide compound object authoring tools and services

Dependencies

Identify dependencies that exist to activities or developments external to the project.

Related national and international dependencies include:

- Metadata Schema Registry (ANDS)
- Collections Registry (ANDS)
- Collaborative Annotation Services (DART, ARCS)
- User Authentication (Shibboleth/MAMS)
- OAI-ORE Initiative (NSF, Carl Lagoze)

This project also builds on an existing online demonstrator of a Metadata Schema Registry

- Authenticated Web Interface to an online repository of metadata schemas and application profiles (Shibboleth for authentication)
- Users can: Submit/register and validate schemas (and example instances); Describe metadata schemas (or versions); Search, browse, retrieve and edit metadata schemas
- Existing codebase and demonstrator developed in Python, PHP, Apache, XSLT, MySQL, eXist, Redland RDF Library
- Completed surveys, scoping studies
 - “Survey of Metadata Registries”, Suzanne Little, 2006
Evaluation of JISC IEMSR software

It also builds on the following existing tools and services:

- W3C Annotea annotation/tagging system – an open source server and model
- Shibboleth and XACML – authentication and access control standards
- OAI/ORE – Open Archives Initiative for harvesting annotations
- PILIN – identifier services
- Connotea – open source social bookmarking tool for scientific publications and citations
- Vannotea, Co-annota, AnnoCryst – collaborative open source annotation/tagging services developed within the UQ eResearch group through GrangeNet and DART funding
- AustLit project <http://autlit.edu.au/>
- Annotation Harvesting project – being funded by PILIN/ARROW projects

It also builds on the following established international linkages:

- JISC IEMSR
- NSDL Metadata Registry (Dianne Hillman, Cornell)
- eXtended Metadata Registry (XMDR) Project (Bruce Bargmeyer, LBNL)
- ISO/IEC 11179 Metadata Registries (MDR)
- Metadata Open Forum <http://metadataopenforum.org/>
- OAI Object Re-use and Exchange ORE – Jane Hunter is on the international advisory committee and is the DEST liaison <http://www.openarchives.org/ore/>

- AHDS (Arts and Humanities Data Service)
- Humanities Computing Kings College, London
- UKOLN
- TextGrid – <http://www.textgrid.de/> - a community grid for the humanities
- Ben Lund, Tony Hammond, Nature Publishing, Connotea
- Marja-Riita Koivunen, W3C Annotea
- Martin Doerr, CIDOC CRM
- Dolores Iorizzo, Imperial College, London
- Donatella Castelli, Rosemary Russell - DRIVER project

References

1. J.Hunter, I.Khan, A.Gerber “[HarVANA – Harvesting Community Tags to Enrich Collection Metadata](#)”, submitted to JCDL 2008, June 2008
2. J. Hunter, I. Khan, R. Chernich and A. Gerber. *Open Repositories 2.0: Harvesting Community Annotations to Enhance Discovery services*, Open Repositories 2008, April 2008, Southampton
3. K. Cheung, J. Hunter, A. Lashtabeg, J. Drennan. “[SCOPE - A Scientific Compound Object Publishing and Editing System](#)”, 3rd International Digital Curation Conference, Washington DC, Dec 11-13, 2007
4. J. Hunter, M. Henderson, I. Khan, “[Collaborative Annotation of 3D Crvstallographic Models](#)”, *ACS Journal of Chemical Information and Modeling (JCIM)*, 47(5). September 2007. doi:10.1021/ci700173y
5. R. Schroeter, J. Hunter. “[Annotating Relationships between Multiple Mixed-media Digital Objects by Extending Annotea](#)”, *Proceedings of the 4th European Semantic Web Conference, ESWC 2007*. Innsbruck, Austria. June 2007. pp 533-548. doi:10.1007/978-3-540-72667-8_38
6. R. Schroeter, J. Hunter, J. Guerin, I. Khan and M. Henderson. “[A Synchronous Multimedia Annotation System for Secure Collaboratories](#)” *2nd IEEE International Conference on E-Science and Grid Computing (eScience 2006)*. Amsterdam, Netherlands. December 2006.
7. S.Little, “[A Survey of Metadata Schema Registries](#)”, Technical Report, January, 2006
8. J. Hunter, R. Schroeter, B. Koopman, and M. Henderson. “[Using the Semantic Grid to Build Bridges between Museums and Indigenous Communities](#)” *Global Grid Forum: Semantic Grid Applications Workshop*. Honolulu, Hawaii. June 2004. pp 46 – 60
9. M. Doerr, J. Hunter, and C. Lagoze. “[Towards a Core Ontology for Information Integration](#)” *Journal of Digital Information*, 4 (1). April 2003