Authoring, Editing and Visualizing Compound Objects for Literary Scholarship

Anna Gerber, Jane Hunter

School of Information Technology and Electrical Engineering
The University of Queensland
{agerber, jane}@itee.uq.edu.au

This paper presents LORE (Literature Object Re-use and Exchange), a light-weight tool designed to enable scholars and teachers of literature to author, edit and publish OAI-ORE-compliant compound information objects that encapsulate related digital resources and bibliographic records. LORE provides a graphical user interface for creating, labelling and visualizing typed relationships between individual objects using terms from a bibliographic ontology based on the IFLA FRBR. After creating a compound object, users can attach metadata and publish it to a repository (as an RDF graph) where it can be searched, retrieved, edited and re-used by others. LORE has been developed in the context of the Australian Literature Resource project (AustLit) and hence focuses on compound objects for teaching and research within the Australian literary studies community. However it can easily be tailored to support the creation of compound objects for literary and bibliographic research more generally.

1 Introduction

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, to relate objects around a common theme, or to encapsulate related digital resources for teaching purposes. For example, one might want to relate the original edition of *Follow the Rabbit-Proof Fence* to the illustrated edition, a radio recording and a digital version of the film – and to retrieve and present these resources, with their relationships visualized, regardless of their location. Our objective is to provide a software tool to enable such encapsulation and subsequent re-use and visualization, by building on the efforts of two previous digital library initiatives:

- The IFLA Functional Requirements for Bibliographic Records (IFLA, 1998)

- The OAI-Object Reuse and Exchange (OAI, 2008)

FRBR is a recommendation of the International Federation of Library Associations and Institutions (IFLA) to restructure catalogue databases to reflect the conceptual structure of information resources. It uses an entity-relationship model of metadata for bibliographic resources that supports four levels of representation: work, expression, manifestation and item. It also supports three groups of entities: products of intellectual or artistic endeavour (publications); entities responsible for intellectual or artistic content (a person or organisation); and entities that serve as subjects of intellectual or artistic endeavour (concept, object, event, and place).

The Open Archives Initiative Object Reuse and Exchange (OAI-ORE) is an international collaborative initiative, focusing on a framework for the exchange of information about Digital Objects between cooperating repositories, registries and services. OAI-ORE 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 objects discoverable, machine-readable, interoperable and reusable. Resource Maps and their component resources are all web resources which can be identified and unambiguously referenced by HTTP URI handles, thus providing a basis for reuse and exchange. Named Graphs (Jeremy et al, 2005) are endorsed by the OAI-ORE initiative as a means of publishing compound digital objects that clearly states their logical boundaries (Lagoze, 2007). They do this in a way that is discipline-independent, but that also provides hooks to include rich semantics, metadata, ontologies and rules. In the terms of the OAI-ORE, compound objects correspond to ORE *Aggregations*, and the Named Graphs that describe them to ORE *Resource Maps*.

Our hypothesis is that OAI-ORE Resource Maps provide the ideal mechanism for representing literary compound objects that encapsulate the entities and relationships expressed by the IFLA FRBR. To test our hypothesis, we are working with the Australian literature studies community through AustLit, as part of the Aus-e-Lit project.

2 Objectives

AustLit is a collaboration led by the University of Queensland, between twelve Australian universities, the National Library of Australia and the Australian Research Council. It provides the peak resource of bibliographic data for scholars undertaking research into Australian literary heritage and print culture. Within AustLit, research activities are undertaken by communities of researchers focusing on particular topics, regions, genres or of other special interests. AustLit currently supports sixteen different research communities, including Black Words, Children's Literature, Expatriate and Popular Fiction.

The AustLit data model is based on the IFLA FRBR (Kilner, 2005), making it ideal for evaluating LORE. Figure 1 shows an AustLit work record with correspondences to FRBR entities. Within AustLit, each *work* record presents information relating to a FRBR *Work*. The terms *version* and *publication* are used interchangeably with FRBR's *Expression* and *Manifestation*. AustLit does not record any details about *Items*.

The AustLit data model extends FRBR with event modelling, based on the ABC Harmony (Lagoze and Hunter, 2001) and INDECS models. For example, *Works* have a *Creation Event* with associated attributes to record the *Date* and *Agents* (people or organisations) who were involved e.g. the authors, illustrators, translators etc. The AustLit data model also adds attributes and relationships that are used to record information for specific research groups, for example, the *Banned in Australia* research community uses the *isBanned* attribute plus additional relationships to indicate the responsible *Agent*, and notes and dates associated with the censorship of the Work. By examining the entities and relationships in the AustLit database, we created an OWL ontology that provided the basis for the LORE ontology.

Figure 1. AustLit work record

Authoring and editing of AustLit records is restricted to AustLit staff and a few key members of the research sub-communities who have been trained to use the complex data entry interface. Research communities cannot create their own additions or extensions to the data model to record specialized research data – they must request changes to be made to the underlying AustLit database on their behalf. As the amount of specialized research activity within AustLit has increased, the proliferation of additions to the shared underlying data model has increased. This has increased the complexity of the AustLit user interface, making it even less accessible to scholars who have not been trained in its use.

Hence the core aims of the work described in this paper were to provide easy-to-use tools that can be seamlessly integrated within existing research practices through the AustLit Web Portal and that enable:

- collaborative authoring of scholarly compound objects using standardised formats;

- the publishing of compound objects in open access repositories so they can be readily shared, re-used and edited;

- the discovery and re-use of these compound objects through the attachment of simple metadata;

- the expression of relationships between literary resources (including the lineage of derived intellectual products);

- the visualization of relationships through intuitive graphical user interfaces and presentation formats;

- standardized metadata terms for describing compound objects and a controlled set of relationship types for describing the relationships between the components that make up a compound object.

3 Related Work

A number of previous efforts have applied OAI-ORE to specific scientific disciplines to encapsulate experimental data and results. These include: Foresite (2008), eChemistry (Van Noorden, 2008), UIUC (Cole, 2008) and SCOPE (Cheung et al, 2007). Although CULTOS (2003) uses RDF to represent multimedia and hypertext presentations for e-Humanities applications, it does not combine OAI-ORE and IFLA-FRBR to capture or label the precise relationships between entities. Also relevant is an overview of previous implementations and applications of IFLA FRBR, provided by Babeu (2008).

Although a significant past focus of eHumanities tools development has been on scholarly markup and annotation tools to attach interpretations to individual objects or parts of objects (e.g., paragraphs within an article) (Schreibman et al, 2004) . LORE takes the annotation paradigm a step further, enabling authors to annotate links between multiple resources of mixed media type, with tags from an ontology.

Previous efforts focusing on the visualization of OAI-ORE compound objects have developed node and arc diagrams for visualizing Resource Maps and their evolution over time. Van De Sompel & Lagoze (2007) conducted an experiment that used the Internet Archive's Wayback Machine to archive and display changes to Resource Maps over time. This work uses Webdot and GraphViz to visualize Resource Maps on the Web.
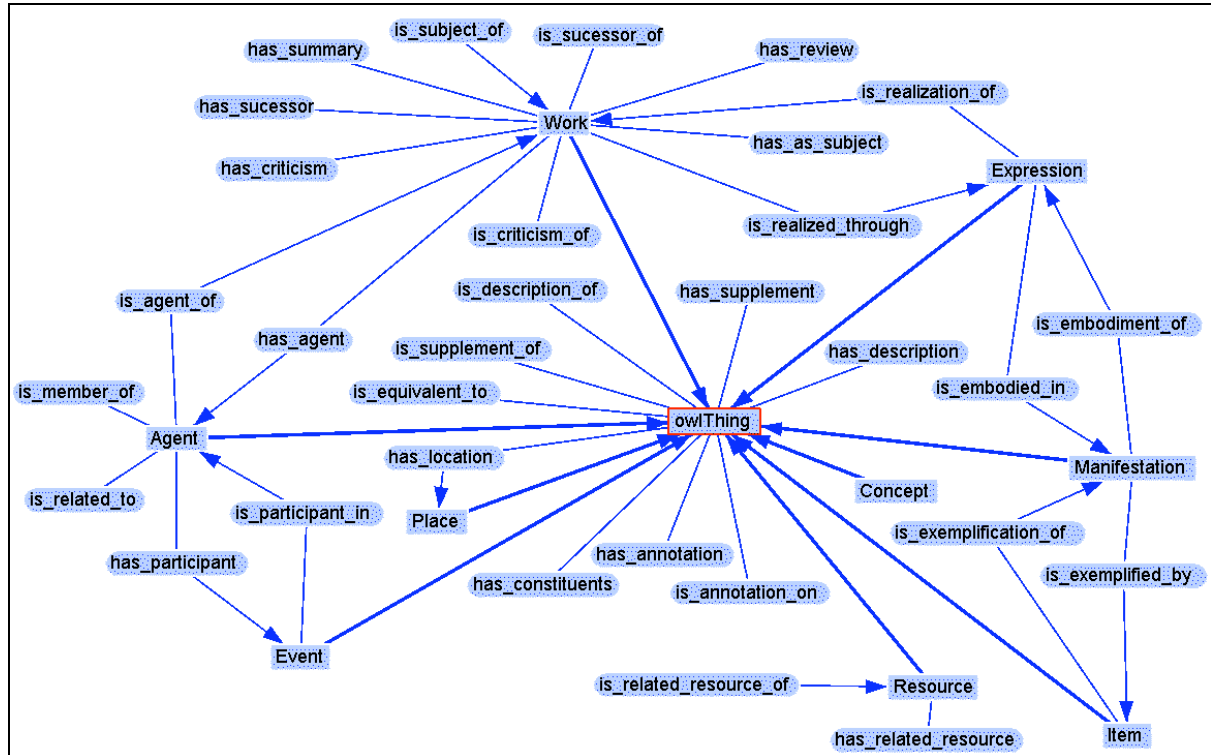
4 Implementation

LORE is implemented as a Mozilla Firefox extension using AJAX technologies. The User Interface (UI) elements are implemented using XUL (XML User Interface Language). LORE stores and queries compound objects via RESTful web services on a Sesame 2 RDF data store running as a Tomcat Java Servlet. A Sesame context (the equivalent of a Named Graph) is created to represent each compound object. The use of Named Graphs allows intra-aggregation relationships and metadata about aggregated resources to be stored and retrieved along with the ORE Resource Map. We set up Apache2 redirect rules to ensure that an RDF/XML representation that describes the Named Graph is located at the URI that identifies each Resource Map.

The internal relationship types, as well as metadata terms that can be applied to aggregated objects are specified by a domain-specific OWL ontology. This ontology was developed by extending the existing AustLit ontology with relationships identified through consultation with AustLit researchers. Figure 2 shows the key classes and properties from the relationship ontology.

The LORE relationship ontology is a simplified version of AustLit's implementation of IFLA FRBR, containing just the *Agent*, *Concept, Event, Work*, *Expression*, *Manifestation*, *Item*, *Place* and *Resource* classes. The *Event* class in this ontology represents events such as writers' festivals and workshops rather than events associated with the lifecycle of FRBR entities from the AustLit data model. Some of the relationships defined by the relationship ontology can be expressed using the

AustLit ontology, but are not easy to apply using LORE. An example is the *is_agent_of* property and its sub-properties such as *is_author_of*. The AustLit data model only allows authors be attached to a Work's Creation Event rather than directly to a Work. The simplified LORE relationship ontology allows users to specify an authorship relationship directly between a Work and Agent.

Figure 2. Subset of LORE relationship ontology

The LORE relationship ontology is presented in more detail in Appendix A. The LORE relationship ontology is provided as the default ontology; however users can configure LORE to use any OWL ontology by modifying the user preferences.
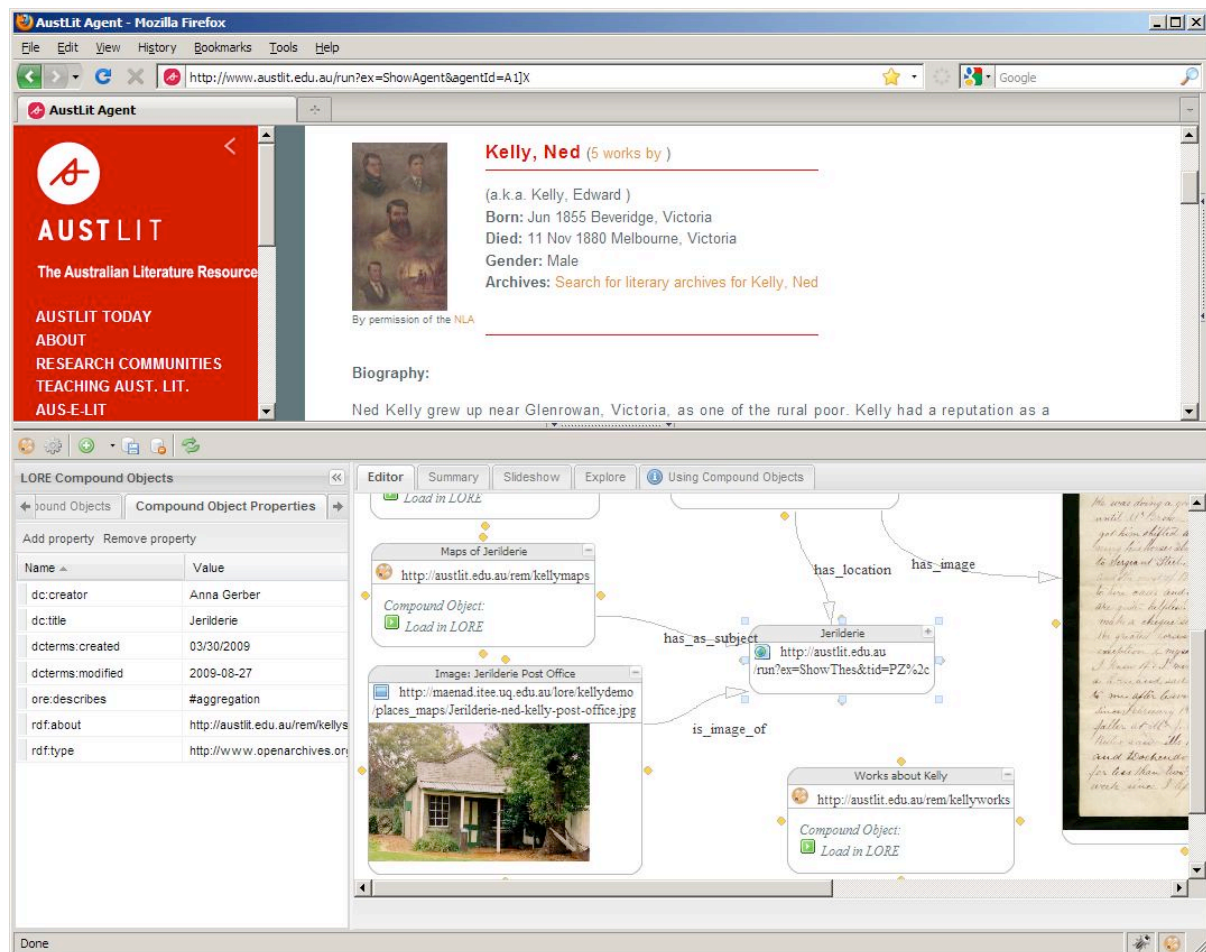
LORE can export to RDF/XML, TriG format or Fedora Object XML (FOXML) files (FedoraCommons, 2009). This allows compound objects to be used with other RDF-enabled tools or to be ingested into a Fedora repository. When exporting to Fedora, a FOXML file is created describing a Fedora object with a data stream of type 'R' for each resource. A Dublin Core (DC) data stream within the object stores metadata for the resource (such as dc:format). The resource map is represented by a Fedora object with a DC data stream for metadata and with a RELS-EXT data stream to store the ore:aggregates relationships as well as any other relationships between resources.

5 User Interface

Figure 3 illustrates the display of an OAI-ORE resource map within LORE's graphical editor. *Nodes* represent the individual resources that are aggregated within the resource map and *arcs* represent the typed relationships between them.

Each node displays the URI and an interactive preview of the resource that it represents. Clicking on the URI opens the resource in the main browser window. Node previews can be collapsed to

conserve screen space, as shown for the *Jerilderie* node in the centre of Figure 3, or resized to display more content, as shown for the image of the *Jerilderie Letter* on the right-hand side of Figure 3. This allows users to view and interact with aggregated resources directly from within LORE rather than having to load them individually via the main browser window.
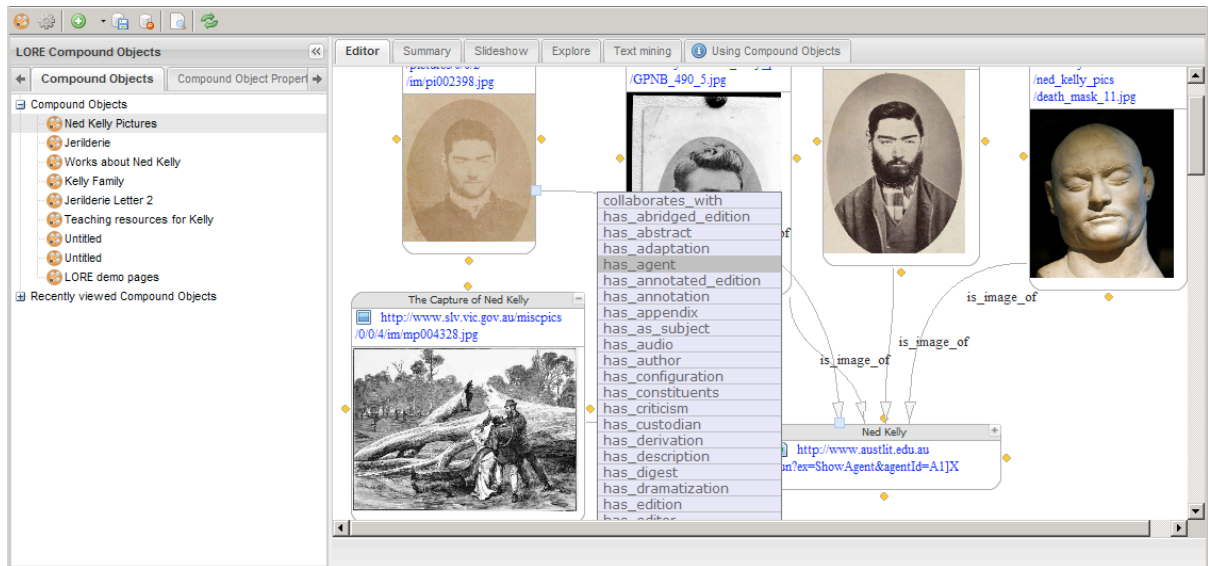

Figure 3. Editing a compound object about Ned Kelly's *Jerilderie Letter*

New resources to be added to the resource map are discovered via the Web Browser (top panel). Clicking on the OAI-ORE logo in the status bar toggles the editor's visibility, so that the full browser window can be used for resource discovery, whilst the resource map under construction remains accessible but hidden during the browsing session. LORE can also be toggled shut or resized via the split bar between the main browser and LORE. A resource loaded into the browser can be added to the resource map via browser context menus or the via the 'Add Node' button in LORE's toolbar. The toolbar also provides options to save or delete compound objects from the repository as well as import and export from local files.

Metadata about the OAI-ORE resource map and aggregated resources is displayed and can be added to or edited via the *Properties* panel. This is displayed to the left of the graphical editor. The metadata terms that are supported include: OAI-ORE, DC and DCMI Metadata Terms (DCMI, 2008), selected terms from FOAF (FOAF, 2007), as well as any datatype properties from the user-specified domain ontology.

Relationships can be asserted between resources by dragging between the connection points to create an arc between nodes. The type of the relationship is indicated by the label on the arc.
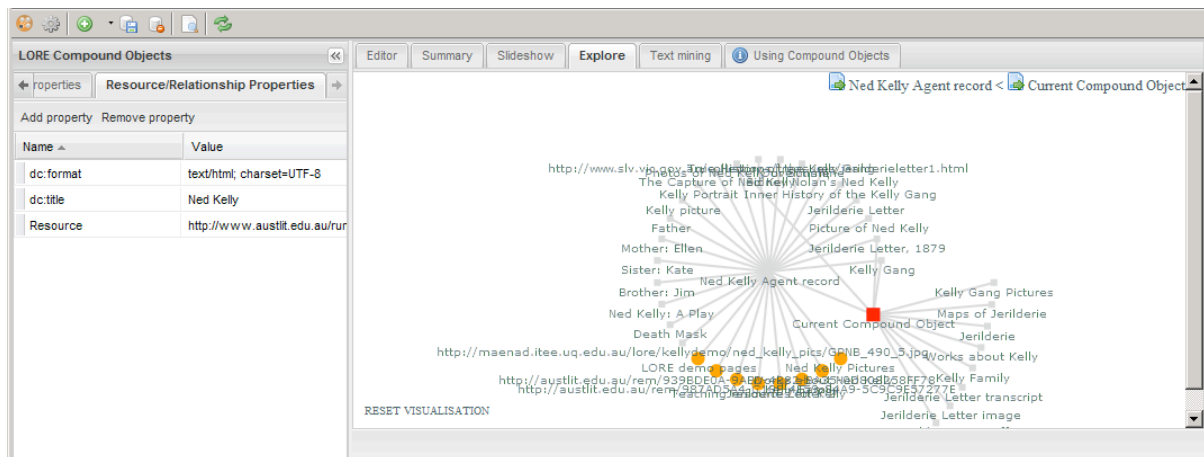
New arcs are created with the type dc:relation by default. However the relationship type can be changed by editing the properties of the arc (via the properties panel) or by selecting from the arc *context menu*, which is populated by the object properties from the domain ontology, as shown in Figure 4. LORE uses the type of the property from the OWL ontology to determine whether to display the property as a bidirectional or unidirectional property; if it is a symmetric property it is displayed with an arrow on both ends of the arc.

Figure 4. Editing relationships between resources

Compound objects can be discovered via automated searches triggered by browser activity. For example, when the user navigates to a URL in the Web Browser, compound objects that aggregate the resource identified by that URL are displayed in the search results in the left-hand panel, also shown in Figure 4. The user can load compound objects from the search results by double-clicking on them. Right-clicking on a compound object from the list of results displays a context menu that allows users to perform other actions, such as adding it as a nested node within another compound object. The *Maps of Jerilderie* and *Works about Kelly* nodes shown in Figure 3 are both examples of nested compound objects. As compound objects are web-accessible resources identified by a URI, they are treated no differently to any other resource from the perspective of OAI-ORE.

LORE uses XSLT to generate a variety of views of the compound objects dynamically. Figure 5 shows the Explore view, which allows the user to navigate the relationships between the current compound object and its aggregated resources. This view is generated by transforming the XML results from SPARQL queries on the repository to a JSON format that can be rendered using the JavaScript Information Visualization Toolkit. This view shows all relationships from the repository, including those both internal and external to the current compound object. Clicking on each resource in this view expands the graph to show more relationships. The Explore view does not display previews for each resource, thus it provides a more compact and faster way to discover compound objects – as opposed to loading each compound object into the graphical editor.

Figure 5. Exploring relationships via the Explore view

Other generated views provided by LORE include: an HTML summary of the current compound object, and a slideshow that displays a preview of each resource alongside key metadata fields such as title, description and rights. We have also created basic mappings to generate SMIL multimedia presentations from mixed-media compound objects, using the relationships between resources to determine where and when to display the related information in the presentation. This builds on previous work (Little, 2002) that maps semantic relationships to spatiotemporal relationships within presentations.

6 Discussion

User feedback has been acquired through user trials involving researchers from a number of the AustLit research communities, as well as independent academics who are regular users of AustLit. During these trials, LORE was used to create compound objects representing research trails, to capture pathways through source materials, for tracking the lineage of derivative works, and to relate disparate resources around a common theme for the purposes of teaching.

The features that the researchers ranked most highly were the interactive node previews and the direct integration of the editor with the browser. They were also very enthusiastic about the generated slideshows and multimedia presentations; however they would like to be able to customize the rules used to generate the visualizations, and would also like to be able to generate more conventional scholarly and document formats, such as reference lists. Our trial users found adding resources to compound objects to be intuitive; however they requested the ability to add resources in batches, for example from AustLit search results, bookmarks or browser history.

When using the AustLit ontology directly, users found applying the ontology terms to be more complex than using the simplified relationship ontology. However, the number of relationships shown in the arc context menu for both ontologies is overwhelming for most users. Presenting the relationships in a menu that uses the sub-property hierarchy from the ontology rather than as a flat list in alphabetical order may be more intuitive. Alternatively, allowing users to type ahead, and to have the UI only show matching terms may be effective. Other strategies for addressing this issue could include adding more semantic checks to the UI to assist users in applying the ontology terms, or tailoring the domain ontologies based on community needs and understanding. Enabling discovery of additional ontologies via a metadata schema registry may also assist individual users to locate existing ontologies that better suit their needs.

Many users were concerned about copyright issues, as they assumed that we were storing copies of the resources that were added to compound objects. They requested more obvious attribution of the source of each resource in the presentation formats and to make it clearer that we only record a link to and user-entered metadata for each resource. The requirement that resources to be added to compound objects must be accessible online also presented an issue during the trials. To add a new resource, users must first publish it, for example by manually uploading it to their institutional repository. Users have requested that this process be made simpler, with the ability to upload new content directly from within LORE. In addition, some objects that exist within institutional repositories only have local identifiers, and resources found on the web that have URIs may not have persistent URIs, which will result in errors with resources failing to load if they are moved or deleted. We may need to incorporate a service to assign persistent URIs to objects that don't already have them, and use archives such as the Internet Archive or PANDORA.

ORE Resource Maps may aggregate concepts or non-information entities (as opposed to information resources) provided they follow the principles of linked data, i.e. that they are accessible on the web via dereferenceable URIs. When using LORE to relate things like *Works, Agents* or other abstract concepts from the AustLit thesaurus, the AustLit record and thesaurus pages can be considered to be proxies for the underlying objects. However, *Expressions* and *Manifestations* are displayed on the Web only within the context of a *Work* record page, making it difficult for users to attach metadata or relationships to those objects. Also, because LORE uses the URI displayed in the browser address bar, it is also possible to aggregate resources with essentially the same content via different URIs. Examples include the address of a work record with and without "www" at the start, or the URI of a search that returns that work as a single result. This has been a source of confusion for some of our users and makes querying the repository for related resources more difficult. We are currently experimenting with the use of RDFa embedded in AustLit records and search result pages to ensure that AustLit resources are identified using consistent URIs, and to make it easier for our users to attach metadata and relationships to embedded objects.

7 Future Work

To make it easier for users to provide new content to include in compound objects, we intend to add upload support via SWORD (SWORD, 2009) or by using the Fedora API directly. We are also planning to provide better support for querying and exporting to Fedora using the REST API, and are investigating mapping a compound object to a single Fedora object with a data stream for each contained resource and the relationships stored in the RELS-INT data stream, rather than multiple objects with the relationships stored in RELS-EXT.

In addition to LORE, the Aus-e-Lit project has developed an in-browser annotation client for attaching Annotea-compliant (W3C, 2005) annotations to web accessible resources. This annotation client has been bundled into the LORE Firefox extension. Future work will involve better integration between these tools to enable the publishing of Scholarly Editions as compound objects that encapsulate different versions of a text, the annotations on the text, and scholarly commentary. Additional presentation rules need to be defined that will enable such compound objects to be exported to TEI documents that use scholarly editing conventions.

Future work will also focus on enabling users to attach creative commons and other licenses to their compound objects. We also plan to integrate a rules engine to infer additional relationships between aggregated resources and to enhance the editor. For example, transitive closures can be

exploited to reconnect nodes that are connected indirectly, if an intermediate node is deleted or to disable relationships or properties that are not applicable to the selected resource.

8 Conclusions

In this paper, we describe LORE, a light-weight tool for authoring, editing and visualizing OAI-ORE compliant compound objects that use the IFLA FRBR model to represent bibliographic relationships. Response to date from the literature research community indicates that the on-going development and evaluation of LORE will provide an essential contribution to the cyber-infrastructure requirements of Australian literary scholars locally, as well as literary scholars globally.

Acknowledgements

References

AustLit. (2008). AustLit: The Australian Literature Resource. http://austlit.edu.au (accessed 11 November 2008).

Babeu, A. (2008). Building a "FRBR-Inspired" Catalog: The Perseus Digital Library Experience. http://www.perseus.tufts.edu/~ababeu/PerseusFRBRExperiment.pdf (accessed 11 November 2008).

Cheung, K., Hunter, J., Lashtabeg, A., Drennan, J. (2007). SCOPE - A Scientific Compound Object Publishing and Editing System, *3rd International Digital Curation Conference*, Washington DC.

Cole, T.W. (2008). OAI-ORE experiments at the University of Illinois Library at Urbana-Champaign. http://www.openarchives.org/ore/meetings/Soton/Cole-OAI-ORE-Roll-Out-OR08.pdf (accessed 11 November 2008).

CULTOS. (2003). http://www.cultos.org/ (accessed 11 November 2008).

DCMI. (2008). DCMI Metadata Terms. http://dublincore.org/documents/dcmi-terms/ (accessed 11 November 2008).

FedoraCommons. (2009) Introduction to FOXML. http://www.fedora-commons.org/confluence/display/FCR30/Introduction+to+FOXML (accessed 14 September 2009)

FOAF. (2007). FOAF Vocabulary Specification 0.91. http://xmlns.com/foaf/spec/ (accessed 11 November 2008).

FORESITE. (2008). http://foresite.cheshire3.org/ (accessed 11 November 2008).

IFLA. (1998). Functional requirements for bibliographic records (FRBR): Final report. http://www.ifla.org/VII/s13/frbr/frbr.pdf (accessed 11 November 2008).

Jeremy, J.C., et al. (2005). Named graphs, provenance and trust, *14th international conference on World Wide Web*. ACM Press, Chiba, Japan

Kilner, K. (2005). The AustLit Gateway and Scholarly Bibliography: A Specialist Implementation of the FRBR. *Cataloguing and Classification Quarterly.* 39:3/4.

Lagoze C. and Hunter J. (2001). "The ABC Ontology and Model (v3.0)" Journal of Digital Information. 2(2). November 2001.

Lagoze, C., Van de Sompel, H. (2007). Compound Information Objects: The OAIORE Perspective. http://www.openarchives.org/ore/documents/CompoundObjects-200705.html (accessed 11 November 2008).

Little, S.. Geurts, J. and Hunter, J. (2002) "The Dynamic Generation of Intelligent Multimedia Presentations through Semantic Inferencing", Proceedings of ECDL 2002, Rome, September 2002

OAI. (2008). Open Archives Initiative - Object Reuse and Exchange. http://www.openarchives.org/ore/ (accessed 11 November 2008).

Schreibman S., Siemens R., Unsworth J. (2004) "A Companion to Digital Humanities, Oxford: Blackwell, 2004. http://www.digitalhumanities.org/companion/

SWORD. (2009) http://www.swordapp.org/ (accessed 14 September 2009)

Van De Sompel, H., Lagoze C., (2007) Interoperability for the Discovery, Use and Re-Use of Units of Scholarly Communication, CTWatch Quarterly, Vol 3, Number 3, August 2007 http://www.ctwatch.org/quarterly/multimedia/11/ORE_prototype-demo/ (accessed 31 March 2009)

Van Noorden, R. (2008). Microsoft Ventures into Open Access Chemistry. *Chemistry World*. http://www.rsc.org/chemistryworld/News/2008/January/29010803.asp (accessed 11 November 2008).

W3C. (2005) Annotea Project. http://www.w3.org/2001/Annotea/ (accessed 14 September 2009)

Appendix A: Object Properties from the LORE Ontology

Domain	Property	Range
Work	has_agent/is_agent_of • has_author/is_author_of • has_editor/is_editor_of • has_illustrator/is_illustrator_of • has_owner/is_owner_of ○ has_custodian/is_custodian_of • has_producer/is_producer_of • has_publisher/is_publisher_of • has_translator/is_translator_of	Agent
	•	
Work	has_annotation/is_annotation_on	Resource
Work	has_as_subject/is_subject_of	Concept or Object or Place or Event
Work	has_configuration/is_reconfiguration_of	Work or Expression or Manifestation
Thing	has_constituents/is_aggregate_of	Thing
Resource	has_copy/is_copy_of	Resource
Work	has_criticism/is_criticism_of	Work or Expression or Manifestation

Work
or
Expression
or
Manifestation
	has_derivation/is_derived_from
	• has_adaptation/is_adaptation_of
	   ○ has_dramatization/is_dramatization_of
	   ○ has_novelization/is_novelization_of
	   ○ has_screenplay/is_screenplay_of
	• has_edition/is_edition_of
	   ○ has_abridged_edition/is_abridged_edition_of
	   ○ has_annotated_edition/is_annotated_edition_of
	   ○ has_expurgated_edition/is_expurgated_edition_of
	   ○ has_illustrated_edition/is_illustrated_edition_of
	• has_imitation/is_imitation_of
	   ○ has_parody/is_parody_of
	• has_performance/is_performance_of
	• has_representation/is_representation_of
	• has_revision/is_revision_of
	• has_translation/is_translation_of
	• has_version/is_version_of
		Work
		or
		Expression
		or
		Manifestation

Thing	has_description/is_description_of • has_metadata/is_metadata_for	Resource
Resource	has_location/is_location_of	Place
Work	has_part/is_part_of • has_excerpt/is_excerpt_from • has_subset/is_subset_of	Work or Expression or Manifestation
Agent	has_participant/is_participant_in	Event
Thing	has_related_resource/is_related_resource_of • has_audio/is_audio_of • has_image/is_image_of	Thing

	• has_related_papers/papers_relating_to • has_transcript/is_transcript_of • has_video/is_video_of	
Work	has_reproduction/is_reproduction_of	Work or Expession or Manifestation
Work	has_review/is_review_of	Work or Expession or Manifestation
Work	has_sucessor/is_successor_of • has_sequel/is_sequel_of	Work or Expession or Manifestation
Work or Expression or Manifestation	has_summary/is_summary_of • has_abstract/is_abstract_of • has_digest/is_digest_of • has_table_of_contents/is_table_of_contents_of	Resource
Work or Expression or Manifestation	has_supplement/is_supplement_of • has_appendix/is_appendix_of • has_glossary/is_glossary_of • has_index/is_index_of	Resource
Work or Agent	influences/is_influenced_by • references/is_referenced_by	Work or Agent
Expression	Is_embodied_in/is_embodiment_of	Manifestation
Thing	is_equivalent_to	Thing
Item	is_exemplification_of/is_exemplified_by	Manifestation
Agent	is_member_of	Agent
Expression	is_realization_of/is_realized_through	Work
Agent	is_related_to • collaborates_with ○ is_co_author_with • is_family_member_of • is_mentor_of/is_mentored_by • is_acquaintance_of	Agent