LORE: A Compound Object Authoring and Publishing Tool for the Australian Literature Studies Community

Anna Gerber[1], Jane Hunter[1]

[1] University of Queensland
St Lucia, Queensland, Australia
(617) 3365 1092
{agerber, jane}@itee.uq.edu.au

Abstract. This paper presents LORE (Literature Object Re-use and Exchange), a light-weight tool which is designed to allow scholars and teachers of Australian literature to author, edit and publish compound information objects encapsulating related digital resources and bibliographic records. LORE enables users to easily create OAI-ORE-compliant compound objects, which build on the IFLA FRBR model, and also enables them to describe and publish them to an RDF repository as Named Graphs. Using the tool, literary scholars can create typed relationships between individual atomic objects using terms from a bibliographic ontology and can attach metadata to the compound object. This paper describes the implementation and user interface of the LORE tool, as developed within the context of an ongoing case study being conducted in collaboration with AustLit: The Australian Literature Resource, which focuses on compound objects for teaching and research within the Australian literature studies community.

Keywords: OAI-ORE, IFLA FRBR, Bibliographic Ontologies, Named Graphs

1 Introduction and Background

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 *Follow the Rabbit-Proof Fence* to the illustrated edition, a radio recording and a digital version of the film based on the novel – and enable them to be retrieved and presented, with their relationships visualized, regardless of their location.

Our objective is to provide a tool to enable such an 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 (FRBR) [1]
- The OAI-Object Reuse and Exchange (OAI-ORE) [2]

The ability to easily share and exchange such compound mixed-media digital objects will facilitate collaborative eScholarship and discussion amongst researchers of Australian literature.

The IFLA Functional Requirements for Bibliographic Records (FRBR) is a 1998 recommendation of the International Federation of Library Associations and Institutions (IFLA) to restructure catalog 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); those entities responsible for intellectual or artistic content (a person or corporate body); and entities that serve as subjects of intellectual or artistic endeavor (concept, object, event, and place).

The Open Archives Initiative Object Reuse and Exchange (OAI-ORE) is an international collaborative initiative, focusing on an interoperability 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 compound digital objects discoverable, machine-readable, interoperable and reusable.

Named Graphs [3] are endorsed by the OAI-ORE initiative [4] as a means of publishing compound digital objects that clearly states their logical boundaries. When applied to compound objects, the nodes in the Named Graph correspond to the individual aggregated resources, and the arcs correspond to typed relationships between those resources. In the terms of the OAI-ORE, compound objects correspond to ORE aggregations, and the Named Graphs that describe them to ORE resource maps. Resource maps and their component nodes and arcs are all web resources which can be identified and unambiguously referenced by HTTP URI handles, thus providing a basis for reuse and exchange. Our hypothesis is that OAI-ORE Named Graphs provide the ideal mechanism for representing literary compound objects that encapsulate the entities and relationships expressed by the IFLA FRBR. They do this in a way that is discipline-independent but provides hooks to include rich semantics, metadata and discipline-specific vocabularies, ontologies and rules.

In developing LORE, we aim to apply OAI-ORE to eScholarship within the discipline of Australian Literature through a case study involving the creation, exchange and re-use of compound digital objects for the purposes of teaching and research within the Australian literature studies community. In addition, the LORE services enable users to label the nodes and arcs within an OAI-ORE compound object using an ontology of classes and relationships which is based on the IFLA FRBR, but extended to support new types of entities and relationships, specific to certain subcommunities.

2 Case Study

AustLit [5] is a non-profit collaboration between the National Library of Australia and twelve Universities. It provides an important resource for scholars undertaking

research into many aspects of Australian literary heritage and print culture history. AustLit serves the research, teaching and general information-seeking communities as a source of information about Australian literary works and the people and organizations involved in their creation or publication (agents), and also as a mechanism for the extraction and dissemination of research data.

AustLit supports the activities of numerous distributed research sub-communities around Australia and the world. These communities include: Black Words, focusing on Australian Indigenous literary cultures and traditions; Australian Popular Theatre, focusing on variety theatre of the 19th and early 20th Century; Australian Responses to Asia, focusing on Australian creative works about or referring to Asia; and many others, each built around an area of specialist literary research.

The AustLit data model is based on IFLA FRBR. The data model adopts the core work, expression and manifestation record structure, with enhancements to represent additional metadata required by AustLit's specialist research communities. AustLit's implementation of IFLA FRBR has also been extended with event-awareness based on the Harmony ABC ontology [6], to represent events such as the creation of a work, realization of an expression or embodiment of a manifestation and the associated agents and event metadata [7]. The records are stored using a highly normalized Oracle database schema based on the structure of RDF and Topic Maps [8].

3 Objectives

In the current AustLit system, authoring and editing of 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. Sub-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 specialized research activity within the AustLit community has increased, the proliferation of additions to the shared underlying data model has increased the complexity of the AustLit user interface (UI) for all users, making it even less accessible to scholars who have not been trained in its use.

Hence, our primary objective is to provide an intuitive method for the AustLit community to collaboratively author scholarly content that can be integrated with existing research practices and tools, including the AustLit web portal. More specifically, we aim to develop an easy-to-use, light-weight, in-browser tool to enable literary scholars to easily author machine-processable and human-understandable compound objects in a standardized format. Additional objectives are:

- to enable the authors to enter metadata describing these compound objects to enable their easy discovery and re-use;
- to enable the publishing of these compound objects in open access repositories so they can be readily shared and re-used;
- to enable the lineage of derived intellectual products to be documented and visualized through these compound objects.

4 Related Work

The SCOPE system was developed by Cheung et al [9] specifically to enable the authoring and publishing of Scientific Compound Objects – and to document the provenance of related scientific outcomes (e.g., data, models, publications). Our aim is to develop a similar tool specifically for scholars of literature and related products. Although some previous work has been done using RDF to represent multimedia and hypertext presentations for e-Humanities applications (e.g., CULTOS [10]), this work does not combine the advantageous features of both OAI-ORE and IFLA-FRBR to capture or label the precise relationships between entities. An overview of other implementations and applications of IFLA FRBR is provided in [11]. Of particular relevance are extensions to the Greenstone digital library software, to provide a visual interface to enable librarians to enrich digital libraries with FRBR data [12], and an alerting service [13]. Existing efforts to implement OAI-ORE for eScience and eScholarship include FORSITE [14], eChemistry [15] and UIUC [16].

A significant focus of e-Humanities tools development has been on scholarly mark-up and annotation tools to attach interpretations to individual objects or parts of objects (e.g., a paragraph in a journal article). LORE takes the paradigm of annotation a step further, enabling authors to annotate or tag the relationships between multiple objects or resources with tags from a specific ontology (an extended OWL version of the IFLA FRBR model).

5 Implementation

In order to address our aims of providing a light-weight tool that can be assimilated with the AustLit web portal and existing research practices, we have adopted an implementation approach based on Web 2.0 technologies. LORE is implemented as a Mozilla Firefox extension, and a standalone version can also be installed on any web server to provide cross-browser support. Both versions are implemented using AJAX technologies (Asynchronous JavaScript and XML). UI elements in the Firefox version are implemented in XUL. The Firefox extension provides some advantages over the standalone version, as it can be customized to suit individual research needs via user preferences; it allows us to bypass security sandboxing that might be imposed on server-hosted JavaScript within institutional environments; and the extension framework provides an ideal mechanism for software distribution and updates which is familiar and accessible to our target community. LORE uses RESTful web services on a Sesame 2 RDF data store, running as a Tomcat Java Servlet, for storing and querying Named Graphs representing the compound objects.

The types for intra-aggregation relationships as well as metadata terms for aggregated objects are specified via an OWL ontology, which is configured at start-up from the user preferences. By examining all of the topics and topic relationships from the AustLit database, we developed an OWL ontology that describes the existing AustLit data model, which is based on IFLA FRBR. We use this as the default ontology in our case study. The AustLit ontology is presented in Appendix A. We also tested LORE with the FRBR RDF model [17] as well as other Bibliographic ontologies [18].

6 User Interface

Figure 1 shows the editing interface provided by the LORE Firefox extension. OAI-ORE resource maps are displayed in a graphical form, as in Figure 1, as well as RDF/XML, shown in Figure 2, which may be selected and copied for use with any RDF-enabled system. In the graphical representation, the *nodes* represent the individual atomic resources aggregated via the resource map and the *arcs* represent typed relationships between the aggregated resources.

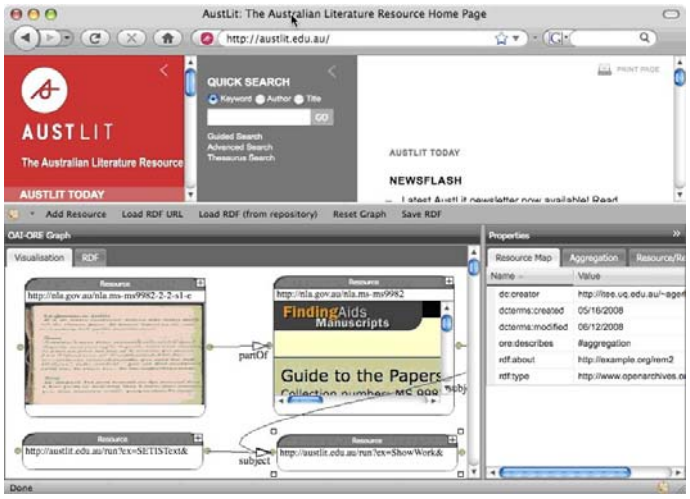

Fig. 1. Compound object editing interface: compound object about Patrick White's *Voss*

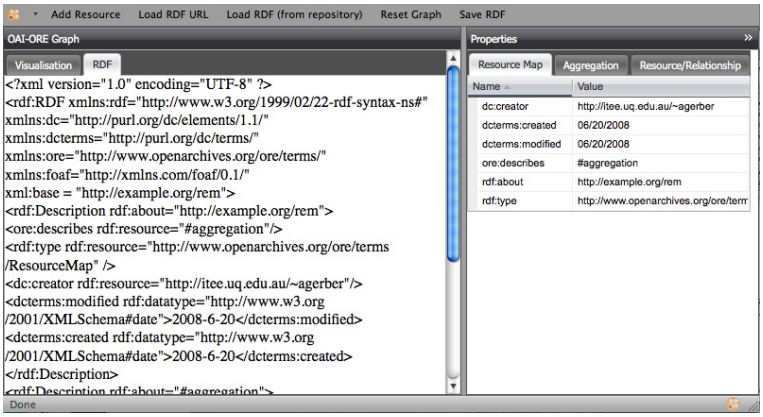

Fig. 2. RDF/XML view of resource map

Each node in the graphical view provides an interactive preview of the resource that it represents, allowing resources to be easily visually distinguished. The preview can be collapsed so that only the URI of a resource is visible, to conserve screen space

in larger resource maps, as seen in the bottom two nodes of Figure 1. When a node is expanded, it may be resized by clicking and dragging with the mouse on the node border. This feature means that the preview can be made large enough for the user to view and interact with the resource directly from within the LORE editor rather than having to individually load aggregated resources via the main browser window.

Metadata about each resource aggregated by the resource map is displayed and can be edited via the *Resource/Relationship* tab in the properties view in the right-hand panel when the resource is selected. The other tabs in the properties sidebar allow metadata about the resource map and the aggregation that it represents to be specified.

An ontology that encodes the relationships and metadata terms specific to the application domain can be specified in the tool preferences. For our case study, we use this preference to load bibliographic ontologies including FRBR and our OWL representation of the AustLit data model. The preferences are also used to configure default values for the creator and the repository where compound objects are stored.

By default, only the metadata required by OAI-ORE is displayed in the properties view when creating a new compound object. A user may specify additional metadata by selecting from the properties menu, as shown in Figure 3. The metadata terms selectable from the menu are those from Dublin Core [19] and DCMI Metadata Terms [20], plus optional terms specified by OAI-ORE, as well as selected terms from FOAF [21] and for resources, datatype properties from the domain ontology.

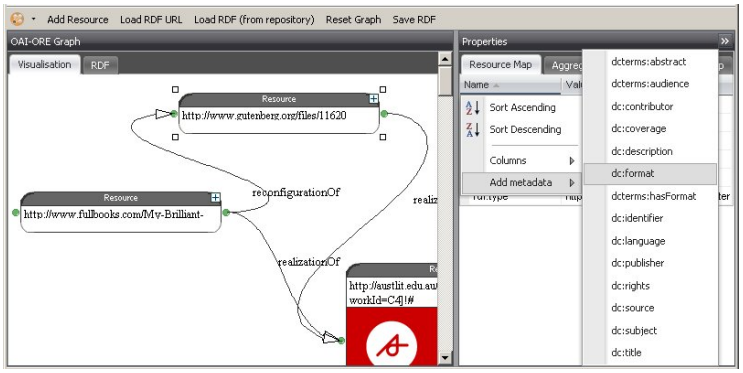

Fig. 3. Additional metadata menu

The types for the relationships represented by the arcs in the visualization are indicated by labels on each arc, and in the *Resource/Relationship* properties view. The type of a relationship can be changed by editing in the properties view or by selecting from the context menu of each arc. The arc context menu types are populated by the object properties from the domain ontology. Figure 4 shows some of the object properties from the AustLit ontology populating the arc context menu.

Resources to be added to the resource map may be discovered by browsing or searching via the Firefox browser. Clicking on the OAI-ORE logo in the status bar toggles the editor between hidden and visible, so that the full browser window can be used for resource discovery, whilst the resource map being constructed or modified remains readily accessible throughout the browsing session. A resource loaded in the browser can be added to the resource map by selecting *Add Resource* from the menu.

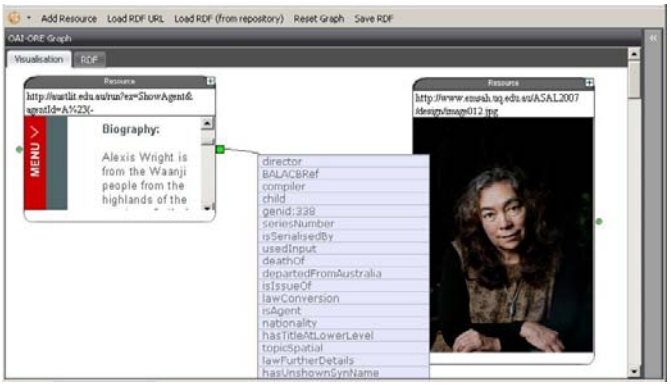

Fig. 4. Resource relationship types from the AustLit ontology

The menu also provides options for saving and loading compound objects. Resource maps that have been created or modified using LORE are saved as named graphs in the RDF repository specified in the preferences by selecting *Save RDF*. Resource maps can be loaded directly from a URL locating an RDF/XML representation of the resource map by selecting *Load RDF URL*. Alternatively, resource maps can be loaded from the default RDF repository, by means of querying the repository for a named graph by its identifier through the *Load RDF (from repository)* menu option.

7 Discussion

7.1 User Interface Feedback

Feedback from the AustLit researchers with whom we have been collaborating has been very positive. They particularly liked the interactive node previews, the direct integration of the editor with the browser through the Firefox extension and the ease with which they could customize the relationship types and metadata supported by the editor by loading different domain ontologies. They would like to see support for additional types of arcs with distinguishing visualizations, e.g. differently shaped line decorators or arrow types, use of colors or line types, to highlight specific relationships, as well as explicit support for bi-directional relationships. The ability to identify a relationship context within a resource (for example, to relate a section of a document to a region within an image) was flagged as a priority. Some usability issues with the current interface were also identified, for example, the UI for adding additional metadata is not intuitive and will require redesign, and some of the terminology used in the UI is too technical for our target users. Feedback so far has been collected through walkthroughs and user testing involving three AustLit researchers. More extensive usability testing will be conducted as part of Aus-e-Lit.

7.2 Considerations Arising from the Case Study

The case study has raised the following issues for further consideration:

Objects can only be added to a compound object in the LORE editor if they can be loaded from a URL in the web browser. This is consistent with OAI-ORE, as it is designed for creating aggregations of web-accessible resources. However, not all objects that researchers may want to encapsulate in a compound object will have URIs that resolve to a web resource (URIs may be used purely as identifiers not as resource locators), and not all objects can be identified by URIs, for example objects that exist within institutional repositories using local identifiers.

Because the IFLA FRBR ontology (and by extension, the AustLit ontology) is quite complex, it can be difficult for a Literary scholar who is not an expert in this model to apply the metadata terms and relationship types from the ontology to relate resources. For example, while they may understand the distinction between a FRBR *manifestation* and *item* in the case of a physical publication such as a book, they may not understand how these concepts apply to a digital resource, and may apply relationships or metadata from multiple levels (work, expression, manifestation or item) to a single resource. Strategies for addressing this issue could include adding more semantic checks to the UI to assist users in applying the ontology terms correctly, or simplifying the domain ontologies based on community needs and understanding.

7.3 Limitations and Future Work

The LORE tool that has been developed is a proof-of-concept prototype that demonstrates how OAI-ORE can be used to author, edit and publish compound objects directly from within the familiar interface of the web browser. Continued development of LORE is part of the Aus-e-Lit project, which aims to address the eResearch needs of scholars of Australian Literature. Further effort will be undertaken as part of this project to improve the robustness and usability of the system and to overcome existing limitations including:

- Fedora [22] support is being implemented so that the OAI-ORE objects can be published to a Fedora repository in addition to the existing RDF datastore functionality.
- We plan to add a search and retrieval interface, to enable discovery of compound objects through searches over the aggregated objects as well as over metadata terms and relationships. In addition to searches based on user-input, searches will be automated based on browser activity, for example, if the user navigates to a URL in the web browser, compound objects that aggregate the resource identified by that URL will also be displayed in the search results interface.
- LORE currently only supports uni-directional relationships. We will investigate distinguishing bi-directional, symmetric, transitive and reflexive relationships from the domain ontology within the editing UI.
- A rule engine can be used to infer additional indirect relationships between aggregated objects. For example, if a transitive property is asserted between objects A and B and also between objects B and C, we can infer that the same property exists indirectly between objects A and C. We intend to investigate how semantic inferencing capabilities can improve LORE's compound object search and editing UI.

- The current implementation only allows a single domain ontology for the metadata terms and relationship types to be configured, and requires the user to configure that ontology directly by URL. This is not ideal for our target users, and we intend to incorporate ontology discovery by enabling querying a Metadata Schema Repository.
- We intend enabling users to attach CreativeCommons licenses to the compound objects, prior to publishing them, so the permitted types of re-use can be specified.

The types of relationships that may be created between aggregated objects, and metadata that may be attached to those objects, as specified in the AustLit ontology will need to be refined and extended in specialized ontologies for use by AustLit's specialist research sub-communities. We plan to evaluate these ontologies and the LORE tool more thoroughly through user feedback and case studies taken from the research sub-communities and from within the wider Australian literature studies community.

8 Conclusions

In this paper, we describe LORE, a light-weight tool for authoring and publishing OAI-ORE compliant compound objects that use the IFLA FRBR model to represent bibliographic relationships. LORE enables literary scholars to create, exchange and re-use compound objects for the purposes of teaching and research, to describe works, agents and related digital resources, plus their associated metadata and typed relationships. The continued development and evaluation of LORE in the context of the case study will provide an essential component of the cyber-infrastructure requirements of the Australian literary studies community, as well as other literary scholars globally.

Acknowledgements
Aus-e-Lit is funded by DEST through the National eResearch Architecture Taskforce (NeAT), part of the National Collaborative Research Infrastructure Strategy (NCRIS).

References

1. Functional requirements for bibliographic records: Final report, (1998) http://www.ifla.org/VII/s13/frbr/frbr.pdf
2. Open Archives Initiative - Object Reuse and Exchange (OAI-ORE), http://www.openarchives.org/ore/
3. Jeremy, J.C., et al.: Named graphs, provenance and trust, In: 14th International Conference on World Wide Web, pp.613--622. ACM Press, Chiba, Japan (2005)
4. Lagoze, C., Van de Sompel, H.: Compound Information Objects: The OAIORE Perspective, (2007) http://www.openarchives.org/ore/documents/CompoundObjects-200705.html
5. AustLit: The Australian Literature Resource, http://austlit.edu.au
6. Doerr, M., Hunter, J., Lagoze, C.: Towards a Core Ontology for Information Integration In: Journal of Digital Information, vol. 4, no. 1 (Apr 2003)
7. Kilner, K.: The AustLit Gateway and Scholarly Bibliography: A Specialist Implementation of the FRBR. In: Cataloguing and Classification Quarterly, vol. 39, no. 3/4 (2005)
8. Fitch, K.: Taking RDF And Topic Maps Seriously. In: AusWeb 2002, The Eighth Australian World Wide Web Conference, Queensland, Australia (2002)

9. Cheung, K., Hunter, J., Lashtabeg, A., Drennan, J.: SCOPE - A Scientific Compound Object Publishing and Editing System, In: 3rd International Digital Curation Conference, Washington DC, (2007).
10. CULTOS, http://www.cultos.org/
11. Babeu, A.: Building a "FRBR-Inspired" Catalog: The Perseus Digital Library Experience (Jan 2008) http://www.perseus.tufts.edu/~ababeu/PerseusFRBRExperiment.pdf
12. Buchanan, G.: FRBR: Enriching and integrating digital libraries. In: Proc. ACM/IEEE Joint Conference on Digital Libraries, pp. 160--269. ACM, Chapel Hill, NC (2006)
13. Buchanan, G., Hinze, A.: A Generic Alerting Service for Digital Libraries. In: Proc. ACM/IEEE Joint Conference on Digital Libraries, pp. 131--140. ACM, Colorado (2005)
14. FORSITE, http://foresite.cheshire3.org/
15. Van Noorden. R.: Microsoft Ventures into Open Access Chemistry. In: Chemistry World. (Jan 2008) http://www.rsc.org/chemistryworld/News/2008/January/29010803.asp
16. Cole, T.W.: OAI-ORE experiments at the University of Illinois Library at Urbana-Champaign, (Apr 2008) http://www.openarchives.org/ore/meetings/Soton/Cole-OAI-ORE-Roll-Out-OR08.pdf
17. Davis, I., Newman, R.: Expression of Core FRBR Concepts in RDF, (2005) http://vocab.org/frbr/core
18. A Selection of Bibliographic Ontologies - OAI-ORE Wiki, http://foresite.cheshire3.org/wiki/CommunitiesHome/ScholarlyWorksHome/BibliographicOntologies
19. Dublin Core Metadata Element Set, Version 1.1, http://dublincore.org/documents/dces/
20. DCMI Metadata Terms, http://dublincore.org/documents/dcmi-terms/
21. FOAF Vocabulary Specification 0.91, http://xmlns.com/foaf/spec/
22. Lagoze, C., et al.: Fedora: An architecture for complex objects and their relationships. In: International Journal on Digital Libraries, vol. 6, no. 2, pp. 124--138. Springer, Heidelberg (2006)

Appendix: AustLit Ontology

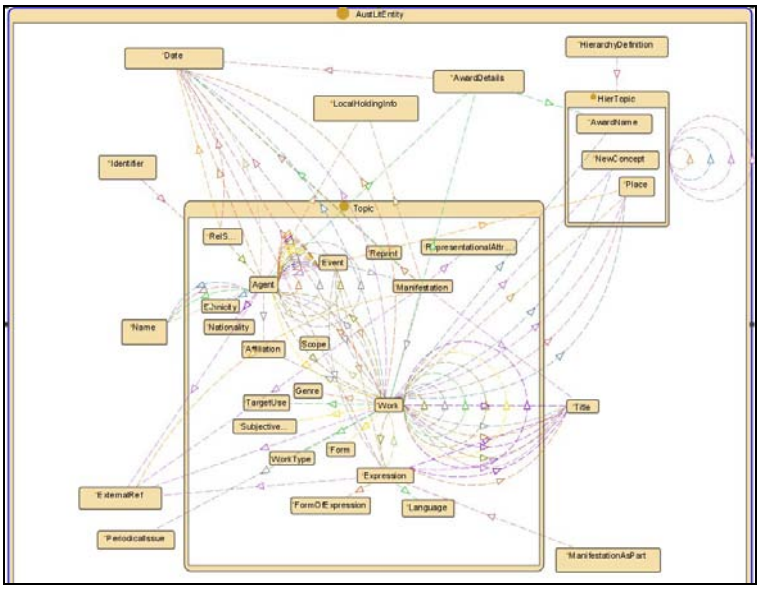

Fig. 5. A subset of the OWL object properties from the AustLit Ontology