Using Encoded Archival Description with Manuscript Collections: The Guide to Australian Literary Manuscripts

Toby Burrows

Dr Toby Burrows is Principal Librarian of the Scholars' Centre in the University of Western Australia Library, where he is responsible for a range of services which include rare book and manuscript collections, and document delivery and inter-library loans. He has published widely in various fields of librarianship, as well as in medieval history. He is a member of the National Resource Sharing Policy Committee, and Reviews Editor of *Parergon*, the journal of Australian & New Zealand Association of Medieval and Early Modern Studies. He was the director of the national project which developed the *Guide to Australian Literary Manuscripts*.

The Guide to Australian Literary Manuscripts is a web service which provides access to the contents of literary manuscript collections in Australian libraries. This article looks at the methodologies employed in developing the Guide, especially in relation to the use of Encoded Archival Description (EAD) as its metadata standard. The future of EAD, and of the Guide, in the Australian information infrastructure is also discussed.

The *Guide to Australian Literary Manuscripts* (*findaid.library.uwa.edu.au*) is a web service containing finding aids for more than 80 collections of Australian literary manuscripts. Launched in May 2001, it was developed with funding from the Australian Research Council through its Research Infrastructure (Equipment and Facilities) Scheme.

The project involved a collaborative effort between six major academic and research libraries: the National Library of Australia, the State Library of New South Wales, the University of Queensland, the University of Sydney, the University of Western Australia, and the Australian Defence Force Academy Library at the University of New South Wales. The University of Western Australia managed the project, and the Guide is housed on one of its servers.

Background

The Guide was designed to fill a significant gap in the provision of information about collections of literary manuscripts and archival material in Australia. Prior to the development of the Guide, there were three main sources of such information:

•   the catalogues of individual libraries and the National Bibliographic Database (NBD);

•   entries in the Register of Australian Archives and Manuscripts (RAAM), maintained by the National Library of Australia; and

•   detailed individual finding aids produced by the specific repositories.

The first two sources provide only a summary collection-level record of manuscript material. While most library catalogues, together with the NBD and RAAM, are accessible via the Web, it is not possible to obtain an in-depth understanding of a collection's contents from these sources.

Only the individual finding aids produced by specific repositories provide this level of detail, and these have traditionally taken the form of in-house documents. In the past, researchers either had to order a copy of these guides, or had to consult them personally at the repository itself.

While some guides have now been published electronically, and may even be linked to the collection-level catalogue records, they still suffer from significant limitations. Their format is inconsistent, they are not searchable except in a simplistic way, and they are dispersed across the websites of the individual institutions. There is no integrative mechanism to bring together collections relating to particular subjects or creators. The project aimed to address these difficulties in relation to a specific subject area, Australian literature.

A virtual manuscript collection

The Guide brings together detailed information about 80 different collections of manuscript materials relating to Australian literary authors – assembling them, in

effect, into a virtual manuscript collection. This collection can be browsed alphabetically by author's name, and each finding aid can be browsed through its table of contents.

The complete collection, as well as individual finding aids, can also be searched by keyword. This makes it possible to look for all occurrences of a particular person or theme anywhere in any of the collections. Several of the finding aids relate to collections of manuscripts originating from a single author, but held in different repositories.

The poet John Kinsella, for instance, has collections of material in the National Library of Australia, the Australian Defence Force Academy Library and the University of Western Australia Library. These collections can be searched as a specific sub-group within the overall Guide. As a result, the physically dispersed material from a particular author is reunited in a virtual sense.

The finding aids contained in the Guide are presented in a consistent, standardised format (Figure 1, below). The metadata standard used is Encoded Archival Description (EAD), and in fact this project was the first significant collaborative application of EAD in Australia. EAD was chosen because of its widespread international acceptance and use as the standard for electronic finding aids in a networked environment.[1]
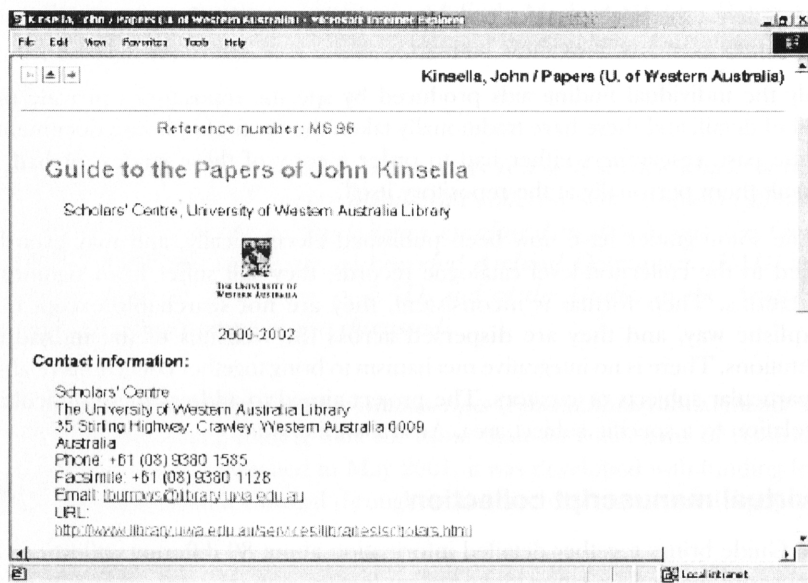

Figure 1. Title page, EAD finding aid

EAD is the result of extensive discussion in the archival and manuscript communities about collection description, and has been used in a range of similar collaborative projects in North America and Europe. It was designed to be compatible with ISAD(G), the General International Standard Archival Description. As a standard based on Standard Generalised Markup Language (SGML) and Extensible Markup Language (XML), it is also independent of any particular software platform.

Using the EAD standard

The project partners worked mainly with existing finding aids for established collections though some new finding aids for new collections were also prepared during the project. In a few cases, the legacy documents were considered to be of insufficient quality and their content was extensively revised and improved. For the most part, though, the main aim of the project was to shape into a consistent and standard format a set of finding aids which varied considerably in structure, arrangement and level of detail. This required two concurrent steps:

· normalising and standardising the structure of each finding aid to conform with the structure prescribed by EAD (Figure 2, previous page);

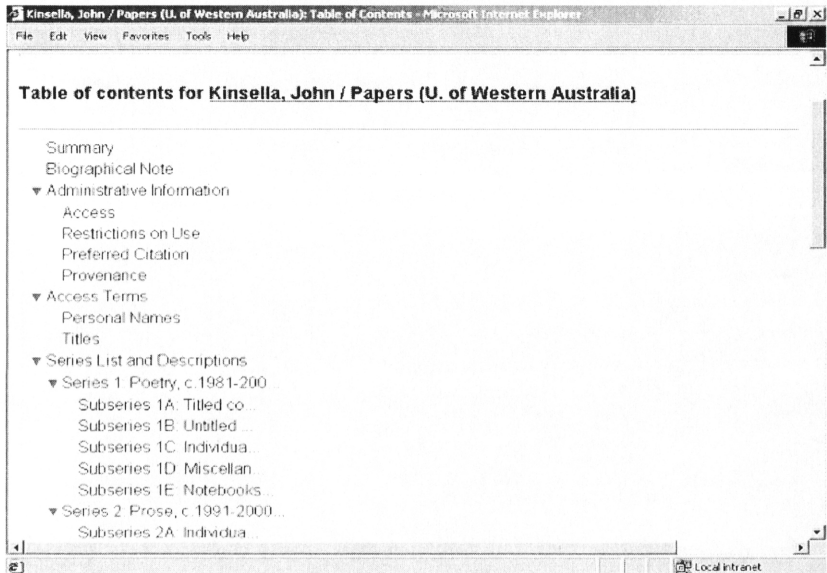· identifying areas where EAD was flexible and non-prescriptive, and

Figure 2. Table of contents, EAD finding aid

determining the extent to which the project should prescribe its own approach in these areas.

The project partners endeavoured to find an acceptable balance between the need to ensure consistency in a consortial environment and the desire to allow some room for individual variation, especially in terms of the amount of detail provided. These issues were much the same as those experienced in similar projects in North America.[2]

The EAD standard itself is very flexible. Only eight of its more than 100 elements are mandatory. It is supplemented by the EAD Application Guidelines, but these only contain recommended approaches rather than mandatory ones.[3] Each institution – and each consortial project – which uses EAD must decide whether to develop its own specific rules for applying EAD or to use rules produced elsewhere. In this situation, the EAD Cookbook produced by Michael Fox has become quite widely used as a de facto set of rules for applying EAD in practice, though this was not its original purpose.[4]

To standardise its use of EAD, the Australian project began with a set of retrospective conversion guidelines developed by Daniel Pitti (the main developer of EAD) for three North American consortia: the Online Archive of California, Virginia Heritage and the Online Archive of New Mexico. These guidelines were
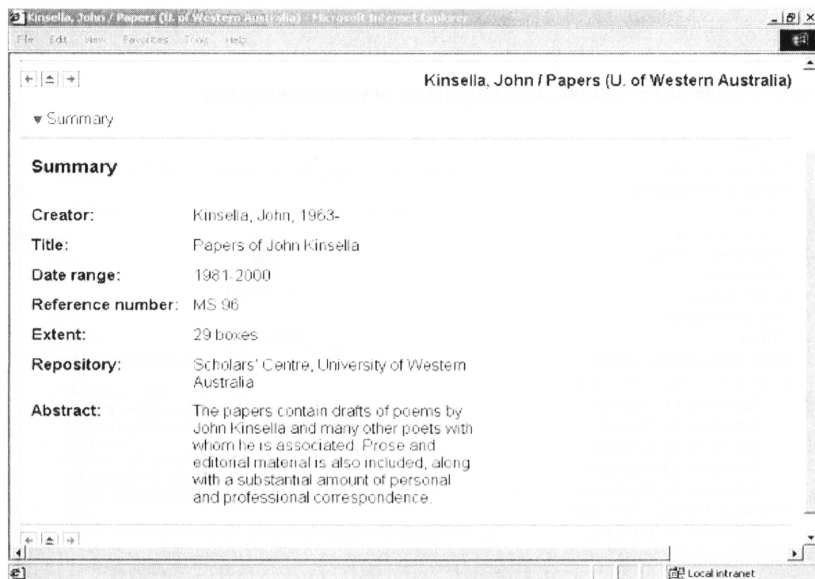

Figure 3. Descriptive identification <did>

modified gradually during the project, to reflect the consensus reached by the partners. Areas which were of particular interest and importance included:

- which elements in the Descriptive Identification (<did>) were mandatory, and the order in which they should appear (see Figure 3 and Appendix 1);

- the extent to which access points such as names and subject headings were required in the Controlled Access (<controlaccess>) area, and the controlled vocabulary which should be used for each kind of access point (see Appendix 2); and

- whether to use the enumerative <c01> ... <c12> pattern to describe subordinate components like series and subseries, or the recursive <c> pattern (the decision was to use the former rather than the latter).

These Australian guidelines were further revised and consolidated for an EAD workshop held in Sydney in September 2001, in association with the *Computing Arts: Digital Research Resources for the Humanities Conference*.[5]

Conversion and publication of finding aids

Each partner library was responsible for encoding its own finding aids, which existed in a mixture of formats: Word, Hypertext Markup Language (HTML), or typescript. In most cases, staff in the participating libraries did this conversion themselves, but some work was outsourced to a commercial firm which specialises in SGML encoding.

Some libraries experimented with specialised SGML/XML authoring software like XMetaL, but most preferred to use a general text editor. NoteTab, in particular, was widely used and several versions of a NoteTab 'clipbook library' for EAD were developed from an original provided by Daniel Pitti (see Figure 4, following page). This included templates for EAD elements as well as links to parser software (nsgmls), an SGML to XML converter (SX), an XML to HTML converter (Saxon), and an Extensible Stylesheet Language (XSL) stylesheet.

The project did not attempt to carry out a systematic evaluation of other methods of encoding finding aids. Other projects have used one or more of the following approaches:[6]

- macros for converting files into EAD format from proprietary word-processing software;

- web-based templates for creating EAD components;

- scripts (usually written in perl) for creating finding aids from generalised databases which use Microsoft Access or Oracle; and

- output of EAD files from data in a specialised archival management system.

For the Guide, a centralised publication process was developed, partly to ensure consistent quality but also to ensure that the finding aids could be searched as a coherent collection and presented as more than static webpages. After encoding their finding aids, the participating libraries sent the files to the University of Western Australia Library, where they were initially published on a restricted-access test website.

The DynaWeb and DynaText suite of software, which indexes and formats SGML files so they can be displayed and searched over the Web, was used for the publication process. After this initial publication, each finding aid was checked, and details of any required revisions were sent to the contributing library, where staff made the suggested changes and resubmitted the corrected files to the University of Western Australia Library. Only after this quality assurance process was complete did the finding aid appear on the public website.

DynaWeb was used for the publication process for three main reasons. Firstly, the University of Western Australia Library already had considerable experience
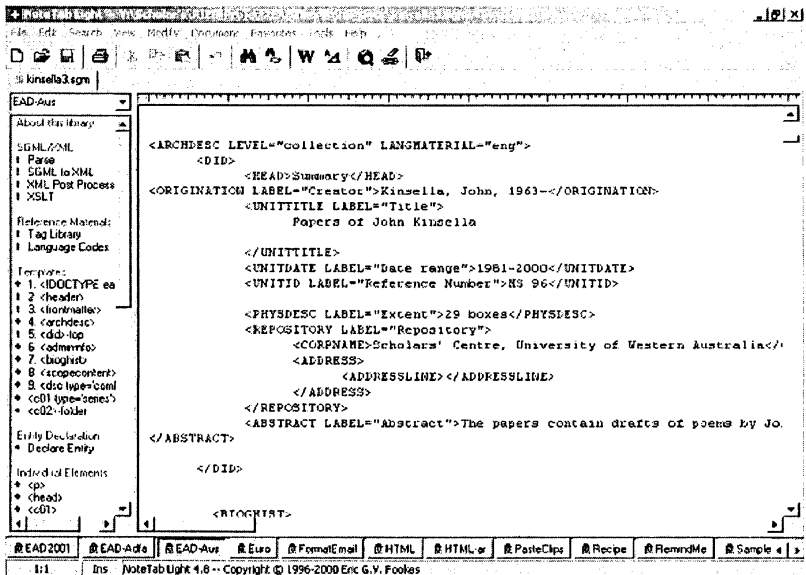

Figure 4. NoteTab view of EAD file

in publishing SGML files with DynaWeb for its Electronic Text Service, though these were large full-text collections distributed by Chadwyck-Healey Ltd. Secondly, DynaWeb has been widely used in North America for publishing significant collections of EAD finding aids, notably at the Online Archive of California and the American Heritage Virtual Archive.[7]

Thirdly, it was a key goal of the project that the finding aids be searchable as well as browsable, and DynaWeb includes this kind of functionality. The project did not explore other methods of publishing EAD files to the Web, though various alternatives have been employed by other projects:

- using other brands of text retrieval and web publication software, like OpenText and Verity;

- producing static HTML files for viewing by a standard web browser; and

- producing static SGML or XML files, which require suitable viewing software at the user's site.

EAD training

Among the project partners, none had any substantial experience in the use of EAD, though a few had already looked at the possibility of using it for their finding aids. Suitable training was therefore a major requirement. The project partners sponsored a visit to Australia in March 2000 by Daniel Pitti, who provided in-depth training for project participants as well as giving public seminars on EAD in Canberra, Sydney, Brisbane and Perth. A return visit by Pitti in September 2001, which included a public EAD training workshop, was co-sponsored by the project. The areas covered by this training included:

- an introduction to SGML;

- a history and overview of EAD;

- encoding a sample finding aid in EAD;

- converting it to XML and HTML; and

- an overview of software tools for parsing, converting and publishing EAD finding aids.

Expertise is now available in the School of Information Systems, Technology and Management at the University of New South Wales to carry out further training workshops of this kind, but there is as yet no coordinated approach to EAD training in Australia.

A national model?

One of the key goals of the project was to link the Guide into the existing information infrastructure, and to investigate future national models for resource discovery through finding aids. Initially, this has involved incorporating a search of the finding aids in the Guide into the AustLit database,[8] as well as building links to the Guide from collection-level records in library catalogues and the National Bibliographic Database and from several web gateway sites. Links from RAAM are also under investigation.

The Guide is a member of the Australian Subject Gateway Federation, though at present this is mainly concerned with sharing information and establishing best practice among subject gateways.

The broader structural issue, still under investigation by the project, is the kind of national model which might be followed in the future for the publication of finding aids using the EAD format. The project database is an example of a centralised model, which focuses on a specific subject area. It remains to be seen whether expanding this database into other subject areas is an appropriate and sustainable approach.

In the longer term, the project database might become one node in a partly decentralised structure, or might be migrated to a central hub or dispersed to its contributing institutions. Another possible option is a linkage with one of the North American or British finding aid services. Some Australian finding aids have been contributed to the Research Libraries Group's Archival Resources service, though this is only available to subscribers.[9] The Archives Hub in the United Kingdom, in contrast, is a free web service.[10]

A centralised approach has the advantage of removing the need for individual repositories to find ways of converting their EAD files to HTML for viewing and searching on the Web. It also ensures a consistency of presentation as well as a uniform level of access and searching functionality. It provides a single starting-point for researchers interested in this kind of data, giving them the ability to do a single search across finding aids from a variety of institutions. But it also raises numerous issues about mechanisms for contributing files to the central site, retention of intellectual property rights, and governance of the whole process.

An additional difficulty might be to identify a single institution as the central repository covering both the library and archival sectors. It could also be argued that this model does not take full advantage of the distributed architecture of the Web. This will become increasingly important as institutions populate their finding

aids with links to images, full text and other media, reflecting the actual content of their collections.

Using a decentralised model, in contrast, would mean that individual repositories retain and publish their own finding aids. This approach avoids difficulties with intellectual property and uses the distributed architecture of the Web. It offers, however, much less likelihood of consistency or of a uniform level of access and presentation. Smaller repositories would probably be discouraged from publishing their finding aids in EAD format, and would find it difficult to acquire sufficient expertise and to provide the infrastructure for publishing SGML-based files, though XML may help to reduce these barriers. There would undoubtedly be problems for researchers in locating relevant finding aids in a fully decentralised structure, as well as an inability to search across files held in a variety of locations.

Perhaps the most effective approach might be one which combines centralisation and decentralisation. Within a distributed network, it would still be possible to have a central site which acted as a directory service. This facility could be constructed by harvesting summary data from sites on the network, in order to create a central database which pointed to finding aids held on remote servers. It could also include a single search interface which worked across this distributed network. A decentralised structure of this kind could also contain various nodes to which smaller repositories without the resources to publish their own finding aids would be able to contribute raw files for encoding and publication.

A key element in this kind of approach might be the Z39.50 standard.[11] The development of EAD attribute sets for Z39.50 will open up the possibility of searches of distributed collections of finding aids, in a manner similar to existing searches of multiple library catalogues. With Z39.50 as the foundation, it should be possible to build portal-type services which search across collections of finding aids in specific subject areas. The Archives Hub is developing an architecture that includes Z39.50-based searching and support for data encoded in SGML or XML.[12] Individual repositories can house and make available their own EAD files for searching by a central harvesting service. This may well serve as a suitable model for a future national approach in Australia.

A different kind of national architecture might also be developed at a more summary level. This could be done by extracting information from the header of each EAD file (the <eadheader> element) and converting it to the Dublin Core metadata standard.[13] This could be harvested, in its turn, by the kind of high-level resource discovery services currently being discussed by the National Library of Australia and the National Archives of Australia.[14]

The future

A future national approach to the publication of EAD-based finding aids is one of the issues which should be addressed in the future management of EAD as an Australian metadata standard. The establishment of a group to take on this role is highly desirable. It would need to combine representation from the library and archives sectors, as well as from relevant professional bodies. It could also feed Australian advice into the North American working group which considers amendments to EAD, on which Australia is already represented. Promotion of EAD, and coordination of EAD training, would also be important activities.

The Round Table on 'Archives in the National Research Infrastructure', held by the National Scholarly Communications Forum in November 1999, endorsed a 'vision for the creation of a web-based distributed search/access infrastructure for archives, based on common descriptive and technical standards' and urged the exploration of EAD for this purpose.[15] The *Guide to Australian Literary Manuscripts* project explored the possibility of using the EAD format as the basis for this kind of infrastructure, and demonstrated the value and suitability of this approach. The project also identified a range of issues which need to be addressed in the development of a national model for the use and exchange of EAD files. Some of these issues are structural and technical, while others relate to disseminating more widely knowledge of EAD and expertise in its use. Tackling these issues and adopting EAD as an Australian standard will ensure that information about our archival and manuscript collections can become part of a rapidly growing international resource discovery framework.

Appendix 1: Descriptive Identification <did>: An Extract from the Encoded Archival Description Draft Australian Guidelines

C. Descriptive Identification <did>

1. Template.

```
<did>
  <head>Summary</head>
  <origination label="[Creator or Collector]">
  </origination>
  <unittitle label="Title"></unittitle>
  <unitdate label="Date Range"> </unitdate>
  <unitid label="Reference number"></unitid>
  <physdesc label="Extent"></physdesc>
  <repository label="Repository">
    <corpname>Repository name</corpname>
```

```
<address>
  <addressline>Repository address</addressline>
</address> </repository>
<physloc label="Location"></physloc>
<abstract label="Abstract"></abstract>
</did>
```

The following table gives the <did> elements in prescribed order; whether they are required (R), not required (NR), or required if available (RA); and the recommended LABEL attribute value or values.

Element	Required?	LABEL	Comments
<head>	R		Summary
<origination>	R	Creator *or* Collector	Use AACR2 for form of name
<unittitle>	R	Title	
<unitdate>	RA	Date Range	
<unitid>	R	Reference number	
<physdesc>	R	Extent	
<extent>	NR		
<repository>	R	Repository	Use AACR2 for form of name
<physloc>	NR	Location	
	NR	Abstract	Recommended
<note>	NR	[as appropriate to descriptive content]	

Appendix 2: Controlled Access <controlaccess>: An Extract from the Encoded Archival Description Draft Australian Guidelines

Names and titles

For <corpname>, <famname>, <geogname>, <persname> and <title> relating to Australian authors and works, the form of name or title should match that used in the AUSTLIT database. Otherwise, use the verified form found in the National Bibliographic Database (Kinetica). Date ranges should be included after each personal name where possible.

<subject>

EAD allows for any number of different subject thesauri to be used, indicated by the attribute "source", eg <subject source="lcsh">. Thesauri not specifically listed in the Tag Library[16] can be included by using:

   <subject source="othersource" othersource="mythesaurus">

For <subject source="lcsh"> the standard will be the National Bibliographic Database (Kinetica).

<function>

No standard specified as yet.

<genreform>

EAD allows for any number of different genre/form lists to be used, indicated by the attribute "source", eg <genreform source="aat">

Genre/form sources not specifically listed in the Tag Library can be included by using:

   <genreform source="othersource" othersource="mythesaurus">

For <genreform source="lcsh"> the standard will be the National Bibliographic Database (Kinetica).

<occupation>

The Register of Australian Archives and Manuscripts is the best source. See: *www.nla.gov.au/raam/browseoccupation.html.*

Awards (Literary prizes)

Use <name> and follow AUSTLIT practice.

ENDNOTES

[1] The best introduction to EAD and its use is Jackie M Dooley (ed.), *Encoded Archival Description: Context, Theory and Case Studies,* Society of American Archivists, Chicago, 1998 (also published as the *American Archivist,* vol. 60, nos. 3-4). The official EAD website is at *lcweb.loc.gov/ead.*

2 Elizabeth J Shaw, 'Rethinking EAD: Balancing Flexibility and Interoperability', *New Review of Information Networking*, vol. 7, 2001, pp. 117–32.

3 Society of American Archivists, Encoded Archival Description Working Group, *Encoded Archival Description Application Guidelines*, Version 1.0, Society of American Archivists, Chicago, 1999.

4 Christopher Prom, 'Extending the Capabilities of the EAD Cookbook', *OCLC Systems & Services*, vol. 17, no. 2, 2001, pp. 89–95. The EAD Cookbook itself is at *jefferson.village.virginia.edu/ead/cookbookhelp.html.*

5 The Guidelines are available at *findaid.library.uwa.edu.au/RIEF_findaid_standards.html.*

6 For more information, see: Michael Fox, 'Implementing Encoded Archival Description: An Overview of Administrative and Technical Considerations' in Dooley, pp. 67–79, and several of the papers in Daniel V Pitti and Wendy M Duff (eds), *Encoded Archival Description on the Internet*, Haworth Press, New York, 2001 (also published as the *Journal of Internet Cataloging*, vol. 4, nos. 3–4).

7 Timothy P Hoyer, Stephen Miller and Alvin Pollock, 'Consortial Approaches to the Implementation of Encoded Archival Description (EAD): The American Heritage Virtual Archive Project and the Online Archive of California (OAC)', in Pitti and Duff, pp. 113–36 (esp. 127–9). The OAC recently migrated from DynaWeb to software produced by the University of Michigan Library's Digital Library eXtension Service.

8 AustLit aims to be the definitive source of information on creative and critical works on Australian literature, as well as on authors and literary organisations. See *www.austlit.edu.au.*

9 See *www.rlg.org/arr.*

10 See *www.archiveshub.ac.uk.*

11 Z39.50 refers to International Standard, ISO 23950: 'Information Retrieval (Z39.50): Application Service Definition and Protocol Specification'. The official website for Z39.50 is at *www.loc.gov/z3950/agency.*

12 Paul Watry, 'Archive Hub Specifications and Architecture', at *sca.lib.liv.ac.uk/cheshire/hubarchitecture3.html.*

13 The Dublin Core metadata standard is a simple set of fifteen elements for describing networked resources. The official website is at *au.dublincore.org.*

14 National Library of Australia, 'National Resource Discovery Metadata Initiative', at *www.nla.gov.au/initiatives/resdisc.html.* The National Archives of Australia is developing a proposal for a national online archival network.

15 National Scholarly Communications Forum, Round Table No. 10 (November 1999): 'Archives in the National Research Infrastructure: Communique and Resolutions', at *www.asap.unimelb.edu.au/nscf/roundtables/r10/r10_resolutions.html.*

16 See *lcweb.loc.gov/ead/tglib/tlhome.html.*