

# Encoded Archival Description in the National Library of Australia

Diana Dack

Diana Dack has worked at the National Library for over twenty years in a variety of roles mostly related to automated systems and projects. In 1995 she was responsible for the initial planning and implementation of the National Library's web server, and as Director of Systems Projects was responsible for the development of the Register of Australian Archives and Manuscripts (RAAM). She is currently working with EAD and its application to the delivery of digitised manuscript collections.

*This article describes the implementation of EAD at the National Library. Issues such as the availability of suitable software, retrospective conversion of existing finding aids, staffing, and the use of EAD encoded finding aids to support access to digitised manuscript collections are discussed.*

The National Library is committed to the conversion of its manuscript finding aids to Encoded Archival Description (EAD)<sup>1</sup> and to the deployment of this standard as a key component of the national information infrastructure.

## Finding aids online

The Library has recognised the potential of making its finding aids widely available online since the earliest days of the Web. When the National Library website went live in 1995, a number of finding aids were marked up in Hypertext Markup Language (HTML) and made accessible. All finding aids in machine-readable

form have subsequently been added. New finding aids are added as they are created and older finding aids added as they are gradually converted to machine-readable form. Of the 1177 finding aids, 560 (almost 50%) are now available online ([www.nla.gov.au/ms/findaids](http://www.nla.gov.au/ms/findaids)).

The availability of the finding aids on the website has made the contents of the Library's collections much more accessible and has resulted in an increase in the number and complexity of enquires, including many from overseas clients. The entire text of the finding aids is indexed, and the ability to do keyword searching over the whole set of finding aids has enabled the Manuscripts staff to deliver a more comprehensive and effective reference service and researchers to access the content of some of the collections from their homes and offices. Usage of the manuscript collection has remained constant despite general reductions in the time and travel budgets of academics. Manuscripts staff attribute at least some of this use to the opening up of the collections online.

As a natural extension of this activity, the National Library became interested in the development of EAD in the mid-1990s. When the Research Libraries Group (RLG)<sup>2</sup> was invited to run an EAD workshop in Australia in 1998, National Library staff attended. However, the success of the online finding aids removed the imperative to adopt EAD quickly. It meant that the Library could afford to wait until the environment was more favourable and the standard had matured before devoting the necessary resources to implementing it.

### **Australian Literary Manuscripts Project**

The Library obtained its first practical experience of applying EAD to its own collections as a participant in the Australian Literary Manuscripts Project<sup>3</sup> which commenced in March 2000.

The Library selected 20 finding aids from its notable literary collections for conversion and inclusion in the project. Considerable work was required on the finding aids prior to coding. Of the finding aids chosen, five were paper-based and needed to be converted to machine-readable form; three needed substantial rearrangement and relisting of the collection; five needed collection summary and biographical notes; and the structure of one was found to be impossible to convert to EAD. Many finding aids also needed minor structural changes to make them compatible with the EAD format.

The encoding itself was outsourced to a company specialising in Standard Generalised Markup Language (SGML). This was not as successful as had been hoped. Considerable time was spent developing encoding instructions to deal

with the many variations amongst the finding aids and in closely proofing and editing mistakes in the returned SGML files. However it provided valuable experience concerning the effort involved in outsourcing retrospective conversion of existing finding aids to EAD. For retrospective conversion on a large scale, it may be more efficient to use an organisation such as RLG that specialises in the creation of EAD finding aids and whose staff are familiar with the nature of manuscript collections.

## **Implementation of EAD in the current operations**

The Library is committed to implementing EAD in its current operations. Since participation in the Australian Literary Manuscripts Project, the Library has focused on fully incorporating EAD into its operations. This work has two aspects:

- the need to develop tools and methods to enable the coding of newly created finding aids in EAD; and
- retrospective conversion of the 1177 finding aids currently available either in HTML or on paper.

### **Creation of new finding aids in EAD**

New finding aids have to date been produced as Word documents by the Manuscripts Section and have been subsequently marked up in HTML by Web Services staff. All new finding aids are to be created as Extensible Markup Language (XML) files in EAD format and the HTML version will be generated from the XML file using a style sheet. Conversion to HTML will be done until such time as most standard web browsers can use XML and the conversion to HTML becomes unnecessary. Finding aids currently produced by the Manuscripts Section are standardised and will readily lend themselves to EAD coding with only minor changes to descriptive practices.

In incorporating implementation of EAD in current workflows for newly created finding aids, there are, as with any implementation of this nature, two main factors that the Library has had to consider:

- software and systems to enable the task to be done; and
- staffing, training and work methods.

### *Software and systems*

A significant challenge in implementing EAD has been finding tools that will provide a user-friendly systems environment for non-Information Technology (IT) staff. HTML, which is a single, simple, commonly used implementation of SGML, has

spawned an industry in 'What You See is What You Get' (WYSWYG) editing and markup tools. Although some of the markup produced using these tools is very suspect, standard web browsers are also very forgiving and capable of ignoring bad markup.

EAD, on the other hand, is a specialised implementation of SGML/XML<sup>4</sup> and users need to rely on general XML and SGML markup tools that can accommodate a variety of XML/SGML implementations, of which EAD is only one of many. HTML markup has a single use, in web browsers, whilst EAD is multipurpose, ie the marked up documents will be used for many different purposes, such as online delivery, indexing, printing, converting to HTML or loading into databases. It is therefore necessary for EAD to conform strictly to the standard so that the documents can be processed successfully by a variety of software to produce a variety of products and derivatives.

There is, at present, no totally satisfactory XML markup tool for the non-specialist. Recent discussions on the Text Encoding Initiative (TEI) mailing list<sup>55</sup> ([listserv.brown.edu/archives/tei-l.htm](mailto:listserv.brown.edu/archives/tei-l.htm)) have bemoaned the fact that there is nothing that can fully meet the needs of the literary community for encoding documents in XML.

The Library has spent some time on the investigation of both public domain and commercially available software products for the creation of manuscript finding aids in EAD format and will continue to look for an ideal product. We have recently decided upon using NoteTab ([notetab.com](http://notetab.com)), a readily available text editor widely used for document markup. NoteTab enables the development of customised templates to be used to automatically create the required code.

The particular implementation of NoteTab is based on a set of software tools developed for use with EAD that were initially developed by Daniel Pitti, together with some modifications originating with the Australian Literary Manuscripts Project, and made available by the University of Sydney.

The implementation includes:

- templates for the creation of EAD markup using NoteTab;
- an XML parser for validating an XML document against the EAD Document Type Definitions (DTD); and
- an Extensible Stylesheet Language Transformations (XSLT) processor for processing the XML finding aid using XSL stylesheets.

The parser and the XSLT processor are invoked from within NoteTab by batch scripts. They are public domain packages. The software resides on desktop

machines and the documents, templates and stylesheets are stored and accessed on shared drives on a central server for ease of maintenance and management of the files.

Considerable modification was required to meet the Library's needs, and the templates and stylesheets will require ongoing refinement as experience is gained in applying them to different styles of finding aid.

### *Stylesheets*

An XSL stylesheet is a file that describes to a piece of software, an XSLT processor, how to manipulate and display an XML document of a given type. Use of XSL stylesheets has provided a standard means to display and make use of an EAD document. Stylesheets and the XSLT processor are used to enable National Library staff to generate a variety of HTML documents for printing and loading on the web server. A number of stylesheets have been developed that can be used to generate different products from the same EAD document, eg collection register sheets, a version of the finding aid for printing and use offline, and an HTML version of the finding aid in National Library web style for loading on the National Library web server.

The availability of software such as that used in this package reflects an environment that is increasingly favourable for the implementation of EAD. The widespread adoption of XML in a variety of application environments on the Web has already ensured that a wide range of products is available for manipulation of documents encoded using XML. The associated development of XSL stylesheets has provided a way of manipulating EAD documents with general-purpose software products. Although the creation of stylesheets requires expertise in the use of XSL, such expertise is also becoming more available for those who cannot develop the skills in-house and would prefer to buy the service. The software used by the National Library could be established and used on a single desktop PC in a small organisation with minimum of expense.

### *Application of the standard*

As far as possible, the finding aid elements used by the National Library have been based on the elements used in the Australian Literary Manuscripts Project for compatibility. There are some differences, but the flexibility of the standard means these will not adversely affect compatibility with the Guide to Australian Literary Manuscripts. Some of the differences simply involve the use of the attribute 'audience' set to the value 'internal' so that elements can be suppressed from public display and stripped if the finding aid is made available outside the Library, eg for an element containing confidential information relating to copyright. Other

differences involve the use of certain elements specifically for the digitisation process. These can also be stripped if necessary.

### *Staffing and workflows*

Many of the staff are familiar with EAD at a general conceptual level, either through involvement in the RLG training course or by involvement in the preparation and editing of finding aids for the Australian Literary Manuscripts Project. In addition to this, two of the staff are fully conversant with EAD at both a theoretical and at a practical operational level through intense involvement in the Australian Literary Manuscripts Project and involvement in the development of standards, tools and templates in the current implementation. The remaining staff are yet to be fully trained.

Considerable thought has been given to the optimum implementation and work pattern for EAD in the section. Personal preferences differ; some prefer to complete the finding aid before marking it up, others prefer to mark up as the finding aid is created. Work responsibilities in the section may also be organised in different ways; the markup task may be centralised and done by a few people who mark up all the finding aids created in the section, or distributed, with each person marking up the finding aids they create. A case could be made for both ways.

The National Library Manuscripts Section is large compared to most in-house archives, having six professional staff. The staff are engaged in a variety of tasks and this means that the sorting of a collection and the creation of a finding aid can take many months. Manuscripts Section staff at all levels, including temporary staff and volunteers, undertake this work.

This makes it difficult to predict how best to ensure that the staff trained in EAD are able to obtain sufficient practice to develop and retain EAD markup skills if they only apply them on an intermittent basis. On the other hand, centralising the skills in one or two people makes the section vulnerable to staff movements. Regardless of the arrangements for coding, all staff must be familiar enough with the standard and the EAD structures to create their finding aids in a pattern that requires coding only and not re-ordering or editing.

Initially, only the professional staff will be trained in EAD markup. Other staff may be trained at a later date, although it is unlikely that the training of temporary staff and volunteers will be considered a sound investment. Experience in using the standard will help determine the most practical level of training and involvement for the particular environment in the National Library.

### Retrospective conversion

The full benefit of EAD encoding can be achieved only if all finding aids are encoded. However the retrospective conversion of existing finding aids presents different, but significant, challenges.

Once a pattern has been determined, EAD imposes a degree of uniformity on the development of new finding aids and can help to standardise them. The retrospective conversion of existing finding aids can require considerable resources in a library such as the National Library where finding aids have been created over many years by many people and with varying degrees of skill and resources. Augmenting and converting finding aids retrospectively must be adequately resourced in order to avoid adversely affecting current processing.

### *Re-evaluating the finding aids*

As was demonstrated in the Australian Literary Manuscript Project, there is often considerable intellectual work involved in preparing retrospective finding aids for markup. Since the project was completed, Manuscripts staff have done an analysis of all the finding aids marked up in HTML. Of the 560 finding aids, only 43% could be converted to EAD without further work being undertaken on them. Of the 57% that need editing, 49% required scope and content notes, biographical notes or structural changes (eg incorporating additions into the series descriptions, placing appendixes into the main description, turning headings into series or subseries); 13% were incomplete (ie only the most recent instalments were recorded in the HTML version); 1% were too problematic to be salvaged by mere editing or were located on the Australian Science Archives Project (now the Australian Science and Technology Heritage Centre, Austehc) site; and 5% had already been converted to EAD.

It is worth noting that these guides are, on the whole, the most recently created and are therefore more likely to conform to current standards. Staff are aware that a survey of the guides that are still only in paper form would reveal that the vast majority would need substantial work.

The determination of minimum acceptable standards is an important consideration for retrospective conversion. While the HTML finding aids are on average of better quality than the paper-based ones, many lack the summary and biographical information that provide context for the collection. Although this information is not mandatory and could be omitted, it is clearly highly desirable to improve the usability of the finding aid. If resources are not available for upgrading, it may be better to convert as cleanly as possible, making only structural changes and adding the required information to those finding aids that are partially complete.

### *Conversion of finding aids currently in HTML*

As mentioned earlier, the Library has over 500 finding aids marked up in HTML and available on the website. Initially it was suggested that some form of automated conversion from HTML to EAD might be possible. However this would be very difficult, as the two standards are so different. HTML describes the layout of a document and EAD describes the content. It is quite possible that it would take as long to proof and edit the resulting files as to code the document in EAD from scratch. Since that time, the survey of existing finding aids has shown that many of these files also need substantial work before conversion.

It is now apparent that the most likely option is to encode the documents from scratch. This can either be done in-house or outsourced to an organisation such as RLG that is experienced in EAD markup and offers the service to libraries at a reasonable cost. The relative merit of these options will become clearer as the Library staff become more familiar with EAD markup and have more information to compare the time and effort involved in marking up in-house or managing an outsourcing process and checking the results. Either way, additional staff will be needed to undertake the required upgrading.

Once the upgrading is done, the decision will be whether resources for markup are directed towards the provision of staff in-house or are spent on outsourcing.

### *Retrospective conversion of paper-based finding aids*

In addition to those finding aids in machine-readable form, the Library has a further 600+ older finding aids in paper form only. These finding aids are generally more varied and of a much lower standard. The work involved in bringing them up to a more acceptable standard and encoding them in EAD makes it a significant task. Some will be straightforward to convert, but many will require extensive work and some will require reprocessing and relisting the collection.

Once again, the extent of the work required depends on the minimum standard the Library is prepared to accept for its EAD finding aids. If resources are not available for extensive upgrading, it may be better to encode the existing paper-based finding aids with minimum change simply to make them available. They could then be gradually upgraded by giving priority to upgrading significant collections.

## **EAD and digitisation of manuscripts**

Early National Library interest in EAD was fuelled by a recognition that the structured and flexible nature of the EAD encoding had the potential to provide



an ideal vehicle for the support and navigation of the objects in digitised collections. In the National Library, the digitisation of manuscript collections has a strong relationship with the implementation of EAD.

The National Library will use EAD encoded finding aids to provide the intellectual underpinning of its digitised manuscript collections. The Library's policy is to provide access to all its digital collections through the catalogue. The catalogue record alone will suffice if it represents a single image, or small group of images, but for manuscript collections where a single catalogue record can represent thousands of images, it is necessary to provide additional structure and navigation aids.

An EAD finding aid can provide that structure. Use of the inherent structure of the finding aid can enable the user to browse the collection in a logical, efficient and consistent way. Because the structure and the content of the finding aid are labelled, the various components can be identified, extracted, indexed, displayed and used in a variety of ways. The standard is also flexible enough to allow the use of certain elements in ways that were not anticipated in order to facilitate linking of the digital images with the descriptions.

There is a resource requirement associated with this use of finding aids. Not all finding aids can be used in this way without considerable modification, and it will always be necessary to evaluate the suitability of the finding aid for this purpose.

### The Barton Project

The experimental digitisation of Sir Edmund Barton's papers successfully demonstrates the use of an EAD encoded finding aid to provide the structure for the navigation and online browsing of a digitised manuscript collection.

For this project, document-level description was provided for the majority of the digital images. This required a major enhancement of the finding aid, which described the collection at a much more general level. It is unlikely that the detailed level of access provided for the Barton collection will be undertaken for current or future digitisation projects since the level of work involved in providing such fine granularity is unsustainable except for the smallest collections.

The Barton Project is well documented on the National Library website ([www.nla.gov.au/ms/findaids/ms51](http://www.nla.gov.au/ms/findaids/ms51)).

### Current digitisation of manuscripts

A mass digitisation project, commenced in 2001, is to include some key manuscript collections in addition to maps, music and pictures.

EAD encoded finding aids will provide the structure for navigating the digital manuscript collections. The mechanism by which this will be realised is not yet complete. The management and delivery of the Library's digital collections is to be undertaken using the Digital Collection Manager (DCM) currently under development. Phase one of the DCM is operational, but the enhancements necessary to support manuscript delivery will not be available until the next release of the system towards the end of the year.

The first step in the digitisation of a manuscript collection is the evaluation of the finding aid to assess its suitability for supporting navigation and presentation of the digitised collection. For this purpose, even a good informative finding aid that could be coded in EAD without modification with satisfactory results may require modification. This is demonstrated in the case of the Library's most recent manuscript digitisation project, the Alfred Deakin papers.

The Alfred Deakin collection has a reasonably detailed finding aid that could be encoded in EAD format, and provides a wealth of information about the contents of the collection for prospective researchers. However it is not sufficiently detailed to provide the necessary structure to navigate through the collection of digital files. Although there is a lot of information in the series descriptions, it is in narrative form. The information will need to be extracted, coded and linked with the relevant digital files and supplemented with additional levels of description to break up large bodies of material into groupings of a manageable size for manipulation online and downloading. This work would not have been necessary except for the digitisation of the collection, but will nevertheless result in an improved finding aid.

#### Manuscripts acquired in digital form

The Library receives a variety of electronic files amongst the papers in its manuscript collections. The Manuscripts and Digital Preservation sections are currently undertaking a project to locate computer disks in manuscript collections and to migrate the data contained within them to Rich Text Format. Work is proceeding on storing them in the Library's archival storage system where they will eventually be managed using the DCM.

The plan is to link the files to the EAD finding aids in the same way as the digitised images of the paper-based materials are linked and thus make them available online to users of the collection. However, unlike the digitised images, most 'born digital' material will need to be restricted to in-house use because of copyright and access restrictions.

## Future directions

The implementation of EAD will enable the National Library to make progress with a number of strategic initiatives. In addition to encoding its manuscript finding aids and exploiting EAD finding aids to support digitisation of manuscript collections, the Library will, in the near future, evaluate EAD as a vehicle for providing access to the contents of its extensive ephemera collections.

An added benefit of the creation of new finding aids in EAD will be the opportunity it offers to automatically generate catalogue entries in Machine-Readable Cataloguing (MARC) format from the XML files.

The replacement of the Library's existing Integrated Library Management System (ILMS) offers further opportunities to exploit the data definition provided by EAD to integrate the entries in the Manuscripts catalogue with the finding aids to provide a single entry point for users of the Library's manuscript collections. This will allow a single search across all collections, either at the collection level or at a more detailed level that includes the detailed descriptions contained in the finding aid.

Integration of catalogue entries describing the collection, or record group, with finding aids is a strategy that is also being considered for the redevelopment of the popular Register of Australian Archives and Manuscripts site. The ideal mechanism for achieving this is still under discussion. Any such project to provide access to the content of Australian archival collections will involve the gathering and indexing of finding aid data from contributors to the service, either by harvesting it or by asking participants to contribute the data.

It will therefore require collaboration with other Australian libraries and archival agencies to develop a national model. Whatever the model adopted, integration of the indexing data will be easier if data is available in a standard format. The adoption of EAD by a significant body of libraries and archives will make the development of a centralised index significantly easier.

The Library therefore has an interest in a widespread and uniform application of the standard to support such resource discovery activities. It has a role to play, in cooperation with other libraries and archives, in supporting the management of the standard in Australia and in promoting its use.

**ENDNOTES**

<sup>1</sup> EAD official website is located at the Library of Congress at *www.loc.gov/ead*.

<sup>2</sup> Research Libraries Group (RLG) at *www.rlg.org*.

<sup>3</sup> Australian Literary Manuscripts Project at *findaid.library.uwa.edu.au*.

<sup>4</sup> XML is a simplified subset of SGML optimised for the web environment. The EAD standard is designed to be compatible with both XML and SGML. Use of XML makes it possible to use standard XML and XSL tools and is compatible with the National Library's other operations.