

lack of knowledge of the existence of collections, such as the ANC's official collection, will impede the researcher.

The Canadian Working Group also found that different finding aids described materials at different levels, *fonds*, series, and items. Initially it appeared as if collection-level description alone was provided on NAAIRS, with some detailing a series or, for smaller collections, items. However, some collections had individual entries for each item, others had a mix of entries for individual items, and sequences of items were described with their individual reference numbers on a single entry. Others had references attached to the items described in Remarks. This mix is understandable in a query facility that accesses items at different levels, but the lack of information providing the level of description, the form of arrangement and cross references to related levels of description prevent the researcher from identifying contextual information.

CHAPTER 3

The EAD and its Technical Application

The Origins of the EAD Document Type Definition

The EAD DTD originated at the University of California, Berkeley in 1993. The aim was to "investigate the desirability and feasibility of developing a non-proprietary encoding standard for machine-readable finding aids..."⁹⁷. An important consideration for the development of the EAD DTD was the conversion of existing finding aids to the new format.⁹⁸ When applying the EAD, the technical 'package' consists of the DTD, which enforces the rules of the format, and a Tag Library, which defines the permissible data elements and their use. The Library of Congress Network Development/MARC Standards Office acts as the maintenance agency for the EAD.⁹⁹

There has been increasing acceptance of the EAD DTD as a standard for encoding. It is in the utility and adaptability of a standard that it is often measured. In the professional milieu, it is also important to examine critically the purposes and principles that a standard does or does not support. The application of standards is predicated on a clear view of the position of the standard within the operational and theoretical environment that it serves. The EAD is not an isolated standard. It contains standards defined by external sources in its structure and it has to conform to the needs of archiving as a discipline, not only as a competent electronic standard. It also has to interface with other electronic standards, for presentational requirements such as printed inventories and formatted Internet web pages. Exchange requirements dictate their standards and it must be accessible for administrative and control functions. In terms of user access, it has to provide the means for interfacing with search

⁹⁷ EAD *Application Guideleines for Version 1.0*, 1999

⁹⁸ EAD *Application Guideleines for Version 1.0*, 1999

⁹⁹ EAD *Application Guideleines for Version 1.0*, 1999

and enquiry technologies, and include features that enable access points and subject indexing.

A Multi-Dimensional View of Standards

The EAD makes use of standards that are defined externally. An example of this is the data value standard for the encoding of type of language, which is recommended by the developers of the EAD but defined and managed by the International Standards Organization.

In order to implement a 'standardised' use of the EAD it is necessary to situate the EAD and related standards. Standards consist of various types, with varying degrees of flexibility and applicable to different elements of description and systems components. It is important to identify the role that each type of standard plays as this aids in including the practices of archivists, information technologists and related practitioners. The SAA's Working Group on Standards for Archival Description used David Bearman's framework to create a three-dimensional matrix, which defines and positions the standards that operate within the encoding of archival description.¹⁰⁰

100 Klaassen (ed.), 1989, p452-455

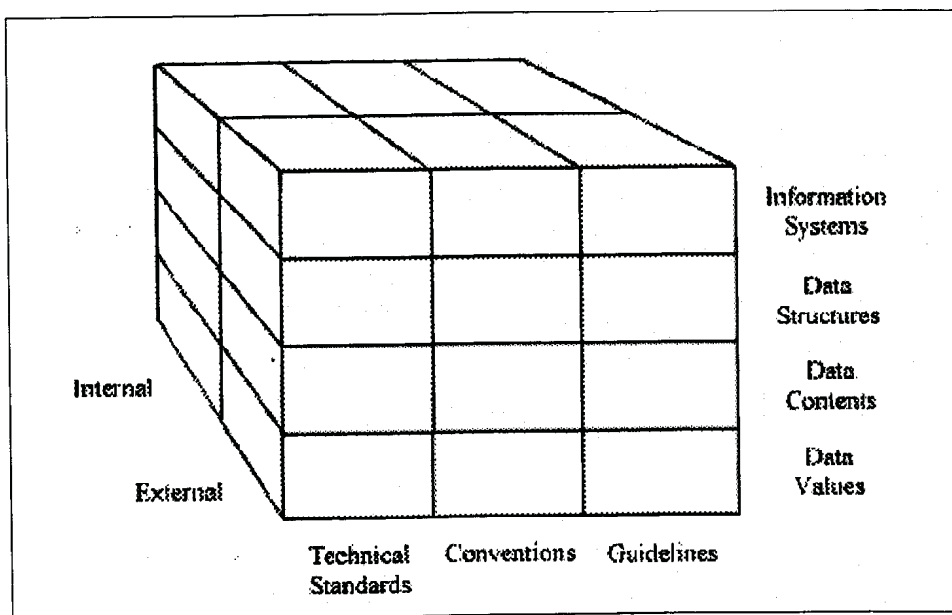


Figure 1 - Three-dimensional Matrix of Standards Applicable to Archival Description¹⁰¹

This matrix is useful in situating the standards applied in archiving and encoding practices. Once these are situated, candidate standards and guidelines can be identified in the appropriate domain and investigated in terms of their usefulness, without 'reinventing the wheel'. Internal standards relate to the professional, in this case archival standards, while external standards reference those from related professions or generally used standards. Walsh points out that there is not always a clear distinction between standards defined outside the archiving profession and internal standards. External standards can be adopted to such an extent that they are assumed to be internal and the "motivation behind trying to draw the external/internal distinction lies largely with establishing how much influence archivists can have in future modifications or development of new standards in these areas."¹⁰²

Examples of standards are:

- Technical standards are often inflexible and their use results in predictable and consistent outcomes. Often their use is unconscious

101 *Standards for Archival Description: A Handbook*, 1994

102 *Standards for Archival Description: A Handbook*, 1994

and tied to prescribed systems. EAD is based on Standard Generalized Markup Language (ISO 8879) in the form of a Document Type Definition (DTD).¹⁰³ SGML has technical standards while the EAD DTD includes a number of internal standards including archival conventions and guidelines. In 1986, Standard Generalized Markup Language (SGML) was first codified by the International Standards Organization. It is a descriptive method of encoding textual information in a form that can be interpreted and used by computers. It is a markup language meta-standard or a standard for constructing markup languages. The use of the EAD DTD is an internal standard while the use of SGML or XML draws on standards defined by external practices in the information technology sector.

- Conventions consist of common practices. Examples of these are:
 - Hensen's Archives, Personal Papers, and Manuscripts (1989) provides an internal data content standard.
 - An internal data structure standard, the USMARC Format for Archival and Manuscripts Control was endorsed by the SAA. The EAD DTD provides an internal data structure standard in its hierarchical design.
 - Data content for some elements of the EAD DTD are prescribed. An example is the inclusion of a collection identifier.
 - Data values for some elements of the EAD DTD are prescribed, such as the level of description.
 - Data values such as the /SO codes, mostly external standards, are conventionally used in the USA.

103 *The Encoded Archival Description Retrospective Conversion Guidelines*, 1997

- Guidelines offer advice on practices, and are the most flexible in the matrix. Examples of these are found in the manuals that the SAA provides, which are used to train archivists.¹⁰⁴

Walsh notes that as standards development matures across organisations, the flexible 'guidelines' become 'conventions' and may become technical standards. In environments that involve data exchange and the development of centrally managed databases from different organisations, technical standards become increasingly important for a composite view of the archival description. This is made possible with pre-determined formats, content, lists or values, standardised names and common usage of alphanumeric codes as some of the components that facilitate inter-institutional co-operation. The benefits of these standards are compatibility between different systems and greater data integrity with a common understanding and use of data elements.¹⁰⁵ These enable relatively seamless interfaces between systems and integration of data between different systems without manual intervention. This automation increases the volumes of information that can be handled, gives the researcher an integrated and coherent view of collections across repositories and enables the archivist to perform administrative and archival functions in collaboration with others. The guidelines provided with the EAD set very few technical standards in content and the mandatory use of elements. This allows considerable licence, but it is also apparent that while the resulting encoding may satisfy the individual institution's EAD encoding requirement, a minimalist and idiosyncratic approach does not lend itself to inter-institutional co-operation.

Because of SGML's complexity, in 1996 the World Wide Web Consortium founded the Extensible Markup Language (XML) Working Group, which developed a simplified subset of SGML, XML. XML has been used for the case study. It provides a markup language but does not provide presentational markup

¹⁰⁴ Rodrigues, 2002

¹⁰⁵ *Standards for Archival Description: A Handbook*, 1994

for printing inventories or displaying archival description on web pages.¹⁰⁶ Coombs writes that a Generalized Markup Language (which includes XML) allows the encoder to concentrate on *descriptive* markup as opposed to presentational markup.¹⁰⁷ Technologies that support descriptive markup have several advantages:¹⁰⁸

- The author's focus is on the content, and not on the control of a computer program to produce a semblance of the desired display.
- The original markup document is not subject to changes for purely presentational requirements and remains a reliable source.
- The development of a standard for description relies on a defined format, which can and should be applied to the original description, and not to input forms or presentational format.
- Standardisation enables portability, and in the case of archival requirements, document exchange.
- Different presentational requirements are based on the description in the source document, ensuring consistent content across all presentations.

The maintenance of professional standards or archiving conventions has had varied successes and uses in the past. These need to be compared against a new standard such as the EAD. The creation of a hierarchy of description levels provides context and enforces the archival principle of provenance. The widely used USMARC Format for Archival and Manuscripts Control was developed to overcome this limitation in the USMARC, which did not encode the multi-level archival hierarchy of arranged archival materials. This is a convention (and an internal standard) that needs to be reflected in an encoding format.

Successful use of MARC AMC, as with the EAD, is based on the application of a data content standard. The 1978 *Anglo-American Cataloguing*

¹⁰⁶ Pitti, 2001

¹⁰⁷ Coombs, 1987

¹⁰⁸ Coombs, 1987

Rules Second Edition or *AACR2* was based on bibliographic conventions needed for the description of books and other library materials and did not reinforce archival principles. In contrast, Hensen's *APPM*, focussed on archival requirements and has become an established standard. It attaches importance to foundational archival concepts such as the principle of provenance, that archival materials occur in collectivities, that the use of archival material is not in its importance as an item but in its relationship to the creator's activities and the importance of multi-level description. However, Hensen says, "It is important to understand that this is not a manual of general archival description, nor is it a guide for the development and construction of archival finding aids".¹⁰⁹ Rather, its use is aimed at the creation of catalogue items in USMARC. However, the widespread use of MARC AMC results in a need for the EAD to include considerations of compatibility with the MARC AMC element requirements.

An increasing recognition in the archival community that an international standard will further co-operative goals and make archival holdings more accessible, dictates the EAD's inclusion of the *ISAD(G)*. In contrast to the Americans where the MARC AMC format found wide acceptance, European archivists focussed on the developing *ISAD(G)*. *ISAD(G)* defines twenty-six elements for the description of an archival entity at any level. The ICA has developed the *ISAD(G)* as a general guidance for the preparation of archival descriptions, to be used with existing national standards or as the basis for the development of national standards. The purpose of archival description is "to identify and explain the context and content of archival material in order to promote its accessibility."¹¹⁰ The *ISAD(G)* rules for multi-level description are designed to ensure that an archival standard is maintained with: description from the general to the specific, information relevant to the level of description, linking of descriptions and, non-repetition of information.

¹⁰⁹ Hensen, 1995

¹¹⁰ *ISAD(G)*, 2000, p7

The EAD DTD

Encoded Archival Description (EAD) is a data structure standard that is designed for the storage of data relating to archival descriptive guides, inventories and catalogues. It is intended to enable delivery of archival data on the Internet and across different systems. The EAD Tag Library identifies the names and definitions of EAD data elements for the application of the structure and content rules. The EAD is intended to accommodate the creation of new finding aids and the conversion of old finding aids including paper-based inventories. Its designers also intended to accommodate description, control, navigation, indexing, and online or print presentation, but not necessarily data that addresses collection management needs specific to local requirements.¹¹¹

The EAD design is possible with the development of open source standards and object-oriented design principles. Open source standards, unlike those found with proprietary software, give users access to the code of a program. Perens of the Open Source Initiative describes the Initiative's *Open Source Definition* as a bill of rights for the user, providing for the right to copy a program and having access to the source code which allows for changes and improvements of the code for one's own use without having to pay licence fees.¹¹² Object-oriented design results in the XML document that describes the archival entity (a collection, in this case) as a complete information object. The rules are embedded in the document and it can be used across a number of systems without any changes having to be made to it.¹¹³ Open source standards and object-oriented design provide flexibility, which underpins the ability to adapt software to requirements and to develop a standard which is usable across multiple systems and for different archival processes.

Users of older electronic systems such as bibliographically based finding aids have found the limitations imposed by the technology inhibit the

¹¹¹ *EAD Application Guidelines for Version 1.0*, 1999

¹¹² Perens, undated

¹¹³ Holman, 2001

development of description. While some of these limitations may, for the outsider, appear minor, within archival description, the impact was severe. Archivists had to wrestle with too small fields, a lack of appropriate fields, and the failure of the finding aid data structure design to support archival practices such as the maintenance of the principle of provenance and other contextual needs. As the archivist entered limited descriptions and attempted to manipulate the use of the data entry fields to accommodate a more complete display, the results were unsatisfactory. The usefulness of the finding aid deteriorated with the lack of overall consistency in the descriptions as archival description was subordinated to bibliographic description requirements. The EAD design was informed by this experience, among other design considerations.

The high level design of the EAD DTD consists of three basic elements:¹¹⁴

- EAD Header, a mandatory element, which relates information about the finding aid.
- Front Matter, an optional element, for additional description for publications.
- Archival Description is the mandatory element for the description of the body of archival material, in this case, a collection. This segment is further divided into
 - hierarchically organised information describing a unit of documents and the
 - component parts, from the general to the specific, reflecting archival principles of arrangement providing 'contextual views, and
 - adjunct information that relates to the collection and provides contextual and other information assisting in the use of the collection.

The EAD DTD design arranges elements hierarchically. High-level elements contain or 'wrap' lower levels. EAD is the highest wrapper element, with EAD Header, Front Matter and Archival Description contained within EAD. These three elements in turn wrap lower level elements.

114 *Encoded Archival Description Tag Library*, 1998, p5-9

EAD Header Element

The description for the EAD Header was modelled on the work done by the Text Encoding Initiative, which is an international humanities-based effort “to develop a suite of DTDs for encoding literary texts or other objects of study”.¹¹⁵ The data in the EAD Header goes further than identifying the repository, collection or finding aid. It is expected to provide the data needed for searching across repositories based on date, repository, language and other important aspects. It can also be used for version control, if a document has to be altered or added to at a later stage.

The EAD Header consists of four main elements:

- the Identifier (<eadid>)¹¹⁶ provides unique identification for the finding aid. While a textual description of the holding institution followed by the name of the collection can be used, encoding examples that were referenced provided standards-based codes. The International Council on Archives’ *ISAD(G)* recommends a unique identification using a country code in accordance with the latest version of ISO 3166, a repository code in accordance with the national repository code standard or other unique location identifier, and a specific local reference code, control number, or other unique identifier.¹¹⁷
- File Description (<filedesc>) contains bibliographic information about the inventory or collection.
- Profile Description (<profiledesc>) describes the electronic finding aid..
- Revision Description summarises information about revisions of the finding aid.

Front Matter Element

The Front Matter is an optional element used to generate title pages to local requirements, encode prefaces, dedications or other text not catered for in

¹¹⁵ *Encoded Archival Description Tag Library*, 1998, p5

¹¹⁶ The XML notation is provided in brackets, while element names are capitalised.

¹¹⁷ *ISAD(G)*, 2000, p13

the EAD Header. The rationale behind this optional element is to allow for local variations, whereas the EAD Header has to follow a prescribed order to ensure uniformity across finding aids.¹¹⁸ Front Matter, when used on projects directed at creating 'union' databases, i.e. integrated databases of collection descriptions from different repositories, regard this as a mandatory element.¹¹⁹

Archival Description Element

The Archival Description element is an important wrapper element containing all the information relating to the material in a collection and contains elements that are important for the primary level of description. The first is the collection-level Descriptive Identifier or <did> which is a wrapper element. The Administrative Information or <admininfo> element provides administrative information about access to the materials and informs archivists managing the collection. The Biography History or <bioghist> element contains biographical and/or historical information about the creator of the materials, while the Scope and Contents (<scopecontent>) element contains the scope and content information for the collection. Depending upon the finding aid being encoded, different options are available for the use of <scopecontent> <arrangement> and <organization>.

The EAD DTD takes advantage of the versatility of modern technology through cross-referencing and pointers. The use of the controlled access element may offer the 'poly-centric' view that the Bearman describes as necessary to researcher needs. The controlled access element "designates key access points for the described materials and enables authority-controlled searching across finding aids on a computer network."¹²⁰ The elements that can be included, <corpname>, <famname>, <function>, <genreform>, <geogname>,

118 *Encoded Archival Description Tag Library*, 1998, p6

119 *The Encoded Archival Description Retrospective Conversion Guidelines*, 1997, p8

120 *Encoded Archival Description Tag Library*, 1998, p94

<occupation>, <persname>, <subject> and <title> should conform to national or international controlled vocabularies.

The <did> is one of the most important elements in EAD because it contains all the information relating to the collection or unit. This is done within an hierarchical structure containing different levels. This emphasis is particularly important as South African archivists have generally not had the exposure of their Western counterparts to the extensive use of finding aids based on bibliographic requirements, which could not encode the complexities of hierarchical structures and enforce the relationships between units of description. Description from higher levels provides context for the lower levels. The lower levels of information inherit information from higher levels, so there is a contextually based structure. Metadata is generic and the same elements are repeated, on the different levels, but content is appropriate to the level described. The *Tag Library* emphasises that description must not be repeated at the item level. Technically, this cannot be reinforced, and the onus is on the encoder to adhere to this standard.

Each level is numbered sequentially as <c01>, <c02>, etc. The XML syntax of the EAD DTD requires a descriptive identification (<did>) followed by the unit title (<unittitle>). The <did> can contain a number of elements for the description of the level and these elements can be repeated where the content is pertinent to that level of description. At the highest levels a level value must be provided from a predefined list that includes 'series', 'collection', 'file', 'fonds', 'item', etc. The designers have limited the mandatory elements in order to accommodate the conversion of old finding aids. Some elements may not be relevant at all levels. If all the items have the same source, repository and physical location, description for these is applied at the collection level, or other appropriate level.

The <extent> tag may be altered reflecting local use and unit of measure. The <physloc> element is intended for the shelf location but repositories may find that the maintenance of this element is complicated by the movement of material.

Access restrictions aid the user in accessing material, and should be completed, either at the highest level where there are none, or a default can be embedded in the template at lower levels. Copyright information or special conditions for using the material are also appropriate in this element. <processinfo> provides the opportunity for describing the archival processes that the collection has undergone.

An important feature of the EAD design is that there should be no repetition of information. The *ISAD(G)* lists reference code; title; creator; date(s); extent of the unit of description; and level of description, as essential elements.¹²¹ The *ISAD(G)* reference code, at collection level, is enforced in the EAD DTD with the Unit Identification (<unitid>) sub-element within Archival Description.¹²² EAD requirements stipulate that *ISAD(G)* equivalents are mandatory and Country Code and Repository Code are needed. An interpretation of the rules associated with these elements is that date(s) and extent should be described at every level, with the number of physical units, the unit of measurement and the specific medium (media).

Mandatory attributes for EAD DTD elements, which parallel the *ISAD(G)* requirements include Level, such as 'series' or 'record group' for the element, Archival Description. For purposes of best practices, the authors of the EAD recommend the use of encodinganalog, parent, and id attributes, which are primarily for use across different electronic systems and finding aids.

Standardised codes, particularly ISO codes are used in the EAD and standards manuals. The advantages of these are uniformity for interpretation and enabling automated access and data exchange, and these codes need to be considered for adoption. National codes for the identification of repositories and institutions are used by overseas archivists, while internally, institutions issue codes for the unique identification of collections and systems. The use of codes

¹²¹ *ISAD(G)*, 2000, p9

¹²² See Figure 2 in Appendix for mandatory elements and attributes

within electronic systems has a number of aims including reducing the chance of error where there are differences in spelling or punctuation of names and providing a broadly understood standard for archivists across national and regional boundaries.

The EAD includes pointer and linking references that are available for simple (one-way) links, and multi-directional links, allowing the user to navigate more rapidly through the display and to access related description.¹²³ In designating links to described items, the use of index terms etc, should also be informed by researcher needs. A further consideration is that extensive linking at item level may result in an unmanageable number of query responses, resembling a search on the public domain search engines such as Google. EAD is also hospitable to embedded links to external files and entities, such as related guides and records as well as images and text files. External links to other objects such as web pages for other collections and images, video and sound material is enabled with the Digital Archival Object (<dao>).

Representational markup is not completely excluded from the XML version of the EAD DTD, which includes tags for formatting description such as <drow> and <dentry> which define a tabular layout. This removes the markup from purely descriptive markup as described by Coombs and adds to the complexity of the encoding. Where possible, stylesheets that are designed to provide this type of format for predetermined tags such as container or history description are a more efficient method of representing formatting, which may be specific to a particular end-product.

The <odd> or Other Descriptive Data element is used for information that does not fit in with the definition of other elements. The designers warn that this should be used very selectively. It is useful for the encoding of legacy or old finding aids, where methods of description may not fall easily within the EAD

123 Encoded Archival Description Official Web Site

elements. Resource constraints may dictate rapid encoding with little time for separating the description into constituent parts.

The <add> Adjunct Descriptive Data element provides supplemental information for the use of materials. This includes additional finding aids and lists of materials. The *Tag Library* emphasises that the supplemental information must be distinguished from the core of the materials. This element is also useful for the encoding of old inventories that do not follow current practices in description. An <add> entry can be contained in any level of the description hierarchy, so the information remains within the core description.

XML is case sensitive so encoding standards such as the use of lower-case for attribute names and tags. This standard results from the compatibility of XML with Unicode character set specifications. Special characters are a concern in EAD-encoded documents. While the EAD and its associated technologies and processes define a number of standards which provide the guidance for encoding, the content of description has to be standardised through standards defined at the encoding location. Headings, abbreviations, date formats and abbreviations should not become idiosyncratic, dependent on individual preferences. The layouts of sections, the ordering of the data elements and the where and if summary information such as date ranges and extents are encoded need a consistent approach. Inconsistency can result in confusion for the researcher. Some elements may need default values set up such as an Access Restriction default such as "Available for research" or the Preferred Citation and some publication and copyright information.

Intersection of the EAD with other standards

The EAD has been developed in a world where standards have to be compatible with those of other practices and technologies. Standards may not be identical, reducing everyone to the lowest common denominator, but they do have to be able to interact. The EAD presents a relatively new standard and the

designers have taken into account the requirements of related standards. These include technical standards for other electronic formats such as MARC AMC and content standards such as *ISAD(G)* and *ISAAR(CPF)*, and the conventions embedded in archival arrangement and description.

THE EAD generally does not dictate content standards, but provides the designation for encoded archival descriptions, i.e. it provides the elements and the overall design. Content of the elements comes from the description, and standards such as *APPM*, *RAD* and *MAD* are some of the sources for standards for description. These national standards also include core archival principles such as the principle of provenance.

The widespread use of MARC AMC especially in the USA, is acknowledged and supported by the EAD. The attribute "encodinganalog" gives the encoder the means of documenting the data needed for a catalogue reference for MARC AMC. Examples of these are the entry of the name of the creator entered in authority form as specified in existing MARC records for the collection. The datamapping or "crosswalks" of USMARC to EAD provide the relationship between the EAD and USMARC.¹²⁴ This provides a reference for entering content into the EAD element which has a MARC equivalent. A cross-reference is entered into encodinganalog.¹²⁵ Where a new finding aid is created, the *APPM* and *AACR2* rules can be referenced to describe authority records to fulfil MARC requirements. The EAD includes a data element to provide the equivalent field number for MARC or any other finding aid for conversion of EAD data to another format. The description can then be available to other finding aids and extracted programmatically. International data exchange is supported by the Dublin Core elements. As in the identification of equivalent MARC records, the Dublin Core's crosswalk of the EAD to the Dublin Core elements can be

¹²⁴ *Encodes Archival Description Tag Library*, 1990, Appendix A

¹²⁵ *EAD Application Guidelines for Version 1.0*, 1998 beginning at p. 240

referenced to ensure compatibility. The EAD Header was purposely designed to ensure compatibility with the TEI standard.

The EAD presents a document that is a record of a collection and its contents. This has to be processed into presentational formats and electronic exchange formats before it has any utility. XML is not accessible as it is an encoding 'language', it is SGML made simpler by stripping much of the representational language from it. This necessitates an external conversion process, using Extensible Stylesheet Language or XSL, to apply representational requirements. In addition, Linking Language or XLink is a standard for hypertext and hypermedia linking. XML Query (XQuery) supports the standardisation for searching texts.¹²⁶ These have their own technical standards and within the conversions of the DTD to HTML or other readable format, other standards need to be defined. The manner in which the display is designed and presented is dependent on local repository requirements. Once displayed, a lack of standards in the encoding of sections, names, headings and other description may be identified, necessitating a rapid definition of internal encoding standards.

The ICA's *ISAD(G)* and *ISAAR(CPF)* represent international standards that have increasingly informed the development of description standards worldwide. The developers of EAD have accommodated the *ISAD(G)* mandatory elements while maintaining compatibility with the *ISAD(G)*'s principles of multilevel description.¹²⁷ It prescribes essential data elements and establishes codes and conventions necessary for capturing and distinguishing information within those elements for future use or manipulation.¹²⁸ Use of the *ISAAR(CPF)*, which provides a standard for authority control records for the provision of access points to documents relating to the same creator, across repositories, is referenced in both the *EAD Guidelines* and the *ISAD(G)*. Standardised forms of EAD elements are needed to allow for access points for successful searches

126 Pitti, 2001

127 *Encoded Archival Description Tag Library*, 1998, p1

128 *Encoded Archival description Tag Library*, 1998, p1

using creator names, such as the Library of Congress Name Authority File or the Art and Architecture Thesaurus.¹²⁹ The codes used access international coding standards such as the *ISO* and suggest the use of national and local coding standards. Controlled vocabularies, thesauri and headings provide other instances of standardisation. The use of terminology specific to the archival profession is also brought under the spotlight, and this necessitates a common understanding of these terms.

Interjections to the EAD design

The debates taking place around description and associated standards have not all been laid to rest. The EAD has the marks of a 'traditional' custodial approach in some areas. Although the focus is no longer on the safekeeping and control of holdings, but rather on the researcher, the methods and type of data used for description and encoding are based on traditional practices. Recent standards developments are not predicated on a particular product or outcome. Bearman earlier took issue with this approach, saying that content standards have to be based on the perceived use of that content.¹³⁰

The Australian Series System separates the authority control or administrative records, for record creating entities described by functions and activities, from the records description. This enables a dynamic linking of records to a new creator when the activity generating those records is moved from one entity to another. The EAD finding aid does not have complete flexibility for this. However, a Working Group has been formed to develop an Encoded Archival Context or EAC standard. In *Toronto Tenets: Principles and Criteria for a Model for Archival Context Information*, It recognises "the existence of other information, such as functions and business processes, geographic places, events, concepts, and topics, that are crucial to archival description...its primary purpose is to

¹²⁹ *EAD Cookbook*, 2000, p14

¹³⁰ Bearman, 1992

standardize descriptions about records creators so that they can be discovered and displayed in an electronic environment, linked to each other to show/discover the relationships amongst record-creating entities, and linked to descriptions of records".¹³¹ The utility of the EAC DTD in providing context about functions, activities and processes separated from, but linked to, the EAD and to what extent it supports the Australian Series System will be seen in the future.

The Australians have also found a lack of elements designed specifically for record management functions. A recent report stated that, "The need for linkages between collection description and collection management is being addressed by the Australian Science and Technology Heritage Centre (Austehc), which is currently working to enable the production of EAD-encoded finding aids from its collection management database application, Online Heritage Resource Manager".¹³²

An EAD document should provide summary information such as extent and date ranges at the header levels of the hierarchy. Summarised description can be encoded but ideally this information can be programmatically calculated. This is useful when documents with contents summarised at header level, rather than item level, are published. This reduces the need for extensive revision should items be added or removed but also assumes the availability of technical skills for programmatic summary. The design of the EAD was particularly concerned with the encoding of inventories and the provision of exchange data. Administrative and the more dynamic requirements of the records-continuum were not a focus of the design.

The Australian's recordkeeping approach to archival description accentuates the need for a dynamic document, one that is not created after a collection or record group has been closed. The maintenance of an EAD document is time-consuming when entries for elements change, such as access

¹³¹ *Report from Toronto Archival Context Meeting, 2001*

¹³² Burrows et al, 2001

restrictions and changes in the physical location. The addition of description of large amounts of new material is laborious. Collections may have living donors who make several donations over a period of time. The practical use of the EAD document under these circumstances was questioned on the international EAD on-line forum. Responses varied but indicated that most encoders do not have any experience of this kind of situation; some assert that one can only encode a closed collection. Some solutions are: to add each accession at the end of the document or, to create a new document (i.e. a new finding aid) for each accession. The arrangement then appears to be assumed to relate to the sequence of accession. Another response referred to the description of a large and regularly added-to record group. A 'main' document was created for the record group with other documents created for 'sub-groups', which are linked from the main document to the sub-groups. The Arrangement tag is used to list and link the components. This maintains the 'hierarchical' approach to arrangement and description, by using main and sub-documents as upper levels of description. Another response from a repository dealing with collections was, "the existing EAD file should be amended/expanded to include the accrual; our finding aids aim to present the collections in original and/or classified order, regardless of whether it arrived in one or several accessions."¹³³

The development of the EAD continues as the international archival community responds to their experiences in using the EAD DTD. The designers have asserted that the development of the EAD DTD is an ongoing process and not confined to the development of inventories for printed and on-line reference and document exchange. It is also intended, in the future, to provide data for the management and organisation of archival systems and takes into consideration standards and products that have international standing.

133 *EADListserv*, 2002

CHAPTER 4

A Case Study in EAD Markup

The Format and Assumptions for the Markup Specifications

A case study involving the markup of a South African collection to the EAD standard provides insight into the adaptability of an example of local description to the EAD format. In Chapter 2, the analysis of samples of NAREM description identified a mix of standards application from different organisations. The ANC Collection from Historical Papers at the University of the Witwatersrand was referenced on NAREM and in this chapter, extracts of the results of an EAD markup of the inventory¹³⁴ are used to illustrate the use of the EAD. The description in the inventory is detailed, adheres to the principle of provenance and uses the hierarchical design for different levels of description. It also contains detailed description at every level and wide-ranging contextual description. These factors combine to provide a comprehensive markup. This is followed by a markup of the NAREM description of the same collection to check its appropriateness for conversion to the EAD format.

The EAD DTD offers considerable flexibility in its application and the result is that no single formula can be prescribed for encoding. The encoded description relies on the description produced by the archivist. The EAD DTD maintains the hierarchy of description levels, and I did not find any description component in the African National Congress collection, which could not be encoded. The challenges arose with formatting, deciding on the appropriate level of element detail and occasionally having to decide which EAD element should be used. The list of elements is extensive, and while this study was not intended to include any changes to the collection's description, it gave rise to some critical thought on what could be done to enhance the given description to make greater use of the tagging, and the opportunity to create one comprehensive document

¹³⁴ *Records of the African National Congress*, 1992

which would be the sole source for data exchange, on-line display and printed inventories and guides, produced at any desired level of description. The content of this chapter illustrates the working code produced for the case study, in the order of writing, followed by notes on what was encoded from the hardcopy inventory and NAREM. The encodement was intended for an on-line finding aid and little of the description was changed. I have also added template extracts and information from *The Encoded Archival Description Retrospective Conversion Guidelines, A Supplement to the EAD Tag Library and EAD Guidelines*, as this is particularly concerned with the creation of a 'union' database, a union of description from many institutions' collections, and the content of this publication provides further insight into the possibilities for an encodement.¹³⁵ This is followed by an analysis of NAREM and the requirements for a NAREM entry to be encoded according to the EAD specification. Conclusions are then drawn from both areas of study.

EAD Header (<eadheader>)

EAD Header Attributes

Encoding 1- Header - Case Study (XML)

```
<eadheader audience="internal" langencoding="iso 639-2"
findaidstatus="unverified-full-draft">
```

Notes:

The values associated with the eadheader attributes¹³⁶ are:

- Audience can be "internal" or "external". This description is intended for the internal use of the repository staff and not for general display.
- Langencoding refers to the use of the International Standards Organization's ISO 639-2 language standard.

¹³⁵ I refer to these standards as the 'Union template'

¹³⁶ An attribute provides specific information from a list of values about an element. I use the term 'element' which is the more general term than 'tag'.

- Findaidstatus has four values, i.e. "unverified-partial-draft", "unverified-full-draft", "edited-partial-draft", "edited-full-draft". The "full-draft" and "partial-draft" refer to either a complete encoding of a collection or an encoding with a subset of the series in a collection. An "edited" document is one that has been checked and approved as opposed to an "unverified" document.

EAD Identification (<eadid>)

Encoding 2 - EADID - Case Study (XML)

```
<eadid type="File">PUBLIC systemid="WITSLIB">ANC</eadid>
```

Encoding 3 - EADID - Union Template (SGML)

```
<eadid type="SGML catalog">PUBLIC "-//Name of owner::subordinate named  
division of owner//TEXT (Country code (ISO 3166)::National repository code::local  
repository reference code::Title of archival unit)//EN" "filename.sgm"</eadid>
```

Notes:

- The <eadid> must provide a unique reference.
- The Union template provides an example of the encoding of the full names.
- One of the reasons for the different encoding is because the union standard for the <eadid> is set according to an SGML Open Catalog specification, and "can be used directly in an SGML catalog file as an entry."¹³⁷ In this case, Type must be set to "SGML catalog".
- The file name at the end of the Union template links to an external file within the computer storage. This is transferred by an institution to be used in the union database. A central means of managing collection files would also use this facility. This works with the SGML version, but I found that I didn't need this in the case study.
- A full encoding of all the attributes would look like this:¹³⁸

¹³⁷ *The Encoded Archival Description Retrospective Conversion Guidelines*, 1997
¹³⁸ *EAD Application Guidelines for Version 1.0*, 1998

```
<archdesc audience="external" relatedencoding="marc"
langmaterial="eng" legalstatus="public" level="fonds"
type="register">
```

- The Relatedencoding attribute identifies MARC AMC as another 'target' system for some of the data. The EAD includes a data element to provide the equivalent field number for MARC or any other finding aid for conversion of EAD data to another format. The description can then be available to other finding aids and extracted programmatically. As there is no local equivalent of a common electronic finding aid, this data element was not used. At the more detailed level of encoding, the attribute Encodinganalog is used to provide a reference to the related system's field or element, so the EAD element can be recognised for extraction.
- The Legalstatus attribute is useful for repositories with materials that are restricted.
- The Type provides for "register" or "inventory".
- The Union template uses AACR2 references for names of repositories, and owning entities which must also be described in the encoding. In the case study, I experimented with codes for the repository and collection, but the names are usually described in full. Country and repository codes are not used at Historical papers, and at the moment there are no national standards applied, while the local unique identifier could be the number '20' which is printed on the hardcopy inventory, or the number which appears on the National Register of Manuscripts database 'AD2186'.

File Description (<filedesc>)

Encoding 4 - FILEDESC - Case Study (XML)

```
<filedesc>
  <titlestmt>
    <titleproper> Inventory of the Institute of Race Relations' ANC Collection,
    <date>1992</date>
  </titleproper>
  <author>Edited by Anna M. Cunningham; processed by Anna Cunningham;
```

```

machine-readable finding aid created by Deborah Wilson</author>
</titlestmt>

<publicationstmt>
  <publisher>University of the Witwatersrand. Department of Historical
Papers</publisher>
  <date>©2000</date>
  <p>University of the Witwatersrand. All rights reserved.</p>
</publicationstmt>
</filedesc>

```

Notes:

- The Union template requires the <filedesc>, <titlestmt>, <titleproper>, and <publicationstmt><publisher>.
- The Title Proper, which refers to the finding aid title, proved problematic. The *ISAD(G)* states that the name of the creator should be included in the title, and the definition of creator(s) provides for 'the organization(s) or the individual(s) responsible for the creation, accumulation and maintenance of the records in the unit of description. The name should be given in the standardised form as prescribed by international or national conventions in accordance with the principles of *ISAAR(CPF)*."¹³⁹ As the ANC created the records, but the collection is part of the collection of the Institute of Race Relations, both names were included in the title, as it was not clear how the connection between the two should be encoded. The source for a standardised form of the names was not known and so the full names were used. The recommendation for the Union template is that a "prefatory phrase" such as 'Inventory of ...' is used.
- The Union Template recommends a separate file be kept for all contact information include addresses, phone and fax numbers, and e-mail addresses. In SGML the file can be declared as an external entity in the entity declaration which is coded before the EAD Header. This external entity can be then referenced in the Publication Statement element. This means that if

139 *ISAD(G)*, 2000, p18

any of this information changed, only the file would have to change and the EAD document would not have to be edited.

- The Union template also has copyright information with '©' , the ISO character entity for the copyright symbol in *ISO 8879-1986 Numeric and Special Graphic* character entity set. The addition of this special symbol in an XML-encoding was not tested.
- The Case Study derived File Description content from the paper-based inventory, with the addition of my name as the creator of the electronic version of the inventory.

Profile Description (<profiledesc>)

Encoding 5 - PROFILEDESC - Case Study (XML)

```
<profiledesc>
  <creation>Machine-readable finding aid derived from hardcopy Inventory 20
  </creation>
  <language>Finding aid is written in <language>English.</language>
  </language>
</profiledesc>
```

Encoding 6 - FILEDESC - Union Template (SGML)

```
<profiledesc>
  <creation>Machine-readable finding aid derived from [paper by means of
  scanning and OCR; OCR file edited for typographical errors before
  encoding][word processing program. [version number if known]] [typescript by
  rekeying] [[database name] database [version number if known] by means of
  machine processing]. [Text converted and initial EAD tagging provided by Apex
  Data Services, <date>[Date received from Apex]</date>]
  [Machine-readable finding aid created by [staff/name]</creation>]
  [Date of source: <date>.</date>]
  </creation>
  <language>Description is in <language>English.</language></language>
</profiledesc>
```

Notes:

- The inclusion of "machine-readable finding aid created by..." in the Union Template's <profiledesc> differs from the Case Study which included this

statement in <titlestmt> in the File Description, and demonstrates how elements do not necessarily have one fixed position in the EAD DTD. The difference here is explained in the output product. The Case Study is intended to provide an on-line inventory, where <titlestmt> will be printed, while the Union Template provides for a union database and this statement is embedded in the <creation> narrative.

- The Union Template includes detailed technical information with related dates about software, databases, version numbers and information sources, which is useful in the management of electronic resources.
- The use of a number of elements in a sentence has to be coded with care to ensure there are spaces in the appropriate spaces. An example is "<language> Finding aid is written in <language>English.</language>". A space is left after "written in", anticipating the next element's content. Minor errors can result in poor encoding such as "Finding aid is written inEnglish" and an internal standard is needed to avoid this.

Revision Description <revisiondesc>

Encoding 7 - REVISIONDESC - Union Template (SGML)

```
<revisiondesc>
  <change>
    <date>[date of change]</date>
    <item>[brief textual description of significant change, name of editor]</item>
  </change>
</revisiondesc>
```

Notes:

- Revision Description is usually used for significant changes only, not minor errors such as typographical errors. It is useful for change control, where a record of the change can be kept.
- As the Case Study was the first EAD document for the ANC collection, this element was not relevant.

Front Matter (<frontmatter>)

Encoding 8 - FRONTMATTER - Case Study (XML)

```
<frontmatter>
  <titlepage>
    <bibseries>
      Historical and Literary Papers: Inventories of Collections
    </bibseries>
    <num>20</num>
    <titleproper>Inventory of the Institute of Race Relations' ANC
Collection</titleproper>
    <date>1992</date>
    <publisher>Department of Historical Papers, The Library, University of the
Witwatersrand, Johannesburg</publisher>
    <date>1992</date>
    <list type="simple">
      <head>Contact Information:</head>
      <item>Department of Historical Papers</item>
      <item>The William Cullen Library</item>
      <item>Private Bag X1</item>
      <item>P.O.Wits 2050</item>
      <item>University of Witwatersrand</item>
      <item>Johannesburg, South Africa</item>
      <item>Telephone: (011) 717 1940</item>
      <item>Fax: (011) 339 4137</item>
      <item>
        E-mail: <extref href="michelep@library.wits.ac.za" role="email"
inline="true">Michele Pickover</extref>
      </item>
      <item>
        E-mail: <extref href="carola@library.wits.ac.za" role="email"
inline="true">Carol Archibald</extref>
      </item>
    </list>
    <list type="deflist">
      <defitem><label>Processed by:</label>
        <item>Anna M. Cunningham</item>
      </defitem>
      <defitem><label>Edited by:</label>
        <item>Anna M. Cunningham</item>
      </defitem>
      <defitem><label>Encoded by:</label>
        <item>Deborah Wilson</item>
      </defitem>
    </list>
    <p>© 2002 The Library, University of the Witwatersrand, Johannesburg. All
rights reserved.</p>
  </titlepage>
</frontmatter>
```

Notes:

- The rationale behind this optional element is to allow for local variations, whereas the EAD Header has to follow a prescribed order to ensure uniformity across finding aids.¹⁴⁰ Publication details were supplemented with address and contact information for Historical Papers.
- The Union template requires <titlepage>, <titleproper>, <publisher>, <date> (of publication), contact information (in <list>), and <p> including copyright statement, in that order.
- The Case Study used the symbol '©' for copyright, rather than the ISO standard described in the Union Template.
- This encoding demonstrates examples of the use of formatting elements for the definition of lists. <label> provides the label for the printed element. The paragraph <p> can be used within a number of elements. I found that <p> could be used to overcome the problem associated with finding the right element to encode some description. As <p> has no archival definition associated with it, it should only be used within another element. The element <lb> is also a formatting element and describes a new line for the enclosed description.
- The Union Template also provides an example of an external pointer or <extptr> accessing a graphics file containing the repository's seal. This has to be declared in the declarations prior to the EAD Header.
- Other Union Template requirements are: <titleproper> should match exactly the <titleproper> given in the <eadheader>, the name of the repository must be used for <publisher>.

Archival Description (<archdesc>)**Encoding 9 - ARCHDESC - Case Study (XML)**

```
<archdesc level="collection" langmaterial="eng">
```

140 *Encoded Archival Description Tag Library*, 1998, p6