

Notes:

- This is the wrapper element for the bulk of the description of archival material.
- The level attribute must always be set. Values are "collection", "file", "fonds", "item", "otherlevel", "recordgrp", "subgrp", "subseries".
- The language of the archival materials should be set using the three letter codes from *ISO 639-2, Code for the representation of names of languages*.
- The *ISAD(G)* requirements for multi-level description are particularly pertinent within <archdesc>. The design supports: description from the general to the specific, information relevant to the level of description, linking of descriptions and non-repetition of information.

Descriptive Identification (<did>)

Encoding 10 - DID - Case Study (XML)

```
<did>
  <head>Descriptive Summary</head>
  <unittitle label="Title">Records of the Institute of Race Relations' ANC
Collection</unittitle>
  <unitdate label="Dates">1928-1973</unitdate>
  <origination label="Creator">
    <corpname>African National Congress</corpname>
  </origination>
  <physdesc label="Extent">
    <extent>12 boxes</extent>
  </physdesc>
  <repository label="Repository">
    <corpname>University of the Witwatersrand. Library. Department of Historical
Papers</corpname>
  </repository>
  <abstract label="Abstract">Forms part of the archives of the South African
Institute of Race Relations. Constitution, minutes, reports, resolutions, Treason
Trial legal records, press statements, speeches, correspondence, press clippings,
printed items, notes and items on the ANC Youth League and ANC Women's
League together with other organisations such as the All African Convention, All
African People's Conference and the South African Congress of Democrats +
Subjects are apartheid, African Nationalism, passes, education, Freedom Charter
and the activities of the ANC in exile. + There is information on Albert John
Luthuli, AP Mola, Nelson Mandela, Silas Modiri Molema, James S Moroka, Godfrey
Mothibe Pitje, Pixley Ka Isaka Seme, Robert Mangaliso Sobukwe, Oliver R Tambo
and Alfred Bitini Xuma.</abstract>
</did>
```

Notes:

- Descriptive Identification (<did>) is the primary element for describing an archival unit. It includes Container, Origination, Physical Description, Physical Location, Repository, Date of the Unit, Identification of the Unit, and Title of the Unit. Generally, the ANC inventory provides an item level title, sometimes with a date, while a large number of items detail the physical description such as number of pages and type of document.
- Input into Abstract was sourced from the NAREM description. The attempt to format the Abstract into paragraphs using <lb>, which is called an 'empty element', was unsuccessful. The *EAD Application Guidelines* state "...XML standard declares that the empty-element tag form must be employed 'for interoperability.'" While the meaning of this statement is admittedly vague, it is easiest simply to use the empty-element-tag syntax (<lb/>) as your default in XML documents."¹⁴¹ However, this did not work, it was either ignored (which may have been a stylesheet omission) or was rejected on parsing (validation). This demonstrates a significant difference in the use of 'empty' elements between SGML and XML, the language used for the case study or an elusive solution.
- The Union template has specific requirements for <did>. These are <unittitle>, <origination> (with the attributes, Provenance, Creator, Collector), <physdesc> (with the attributes Extent, Size, Physical Characteristics), <repository>, <physloc>. The Abstract is not a required element in the Union template.

Contextual Elements in <did>

Encoding 11 - DID Contextual Elements - Case Study (XML)

```
<bioghist>
  <head>History of the African National Congress</head>
  <p>The African National Congress (ANC) was founded on 8 January 1912 ...</p>
  <p>Between 1990 and 1992 meetings took place between the government and
the liberation movements, of which the most numerous and powerful was the
```

141 *EAD Application Guidelines for Version 1.0*, 1999

ANC. The details are being worked out but the intention is that there should be universal franchise, a democratically-elected government and a Bill of Rights. After 80 years of struggle, many of the ANC's aims seem about to be realised.</p>
</bioghist>

<scopecontent encodinganalog="520" id="a3">

<head>Scope and Contents</head>

<p>It is important to note...</p>

</scopecontent>

<organization>

<head>Classification</head>

<p>The papers have been organised into the following series:</p>

<list type="ordered" numeration="upperalpha">

<item>CONSTITUTION AND HISTORY</item>

<item>MINUTES AND OTHER PAPERS</item>

<item>RESOLUTIONS AND REPORTS</item>

...

</list>

</organization>

<add>

<relatedmaterial encodinganalog="544 1" id="a6">

<head>Related Material</head>

<p>The following references may be accessed through the South African National Archives search engine NAAIRS for additional archived material relating to the ANC:</p>

<list type="simple">

<item>REFERENCE AAS46 DESCRIPTION JS Moroka Accession</item>

<item>REFERENCE AAS84 DESCRIPTION SM Molema</item>

<item>REFERENCE AL2494 DESCRIPTION South African History Archive

Periodicals Miscellaneous</item>

<item>REFERENCE AL2516 DESCRIPTION African National Congress, Commission of Inquiry</item>

<item>REFERENCE PC31/1 DESCRIPTION African National Congress collection</item>

</list>

</relatedmaterial>

</add>

<odd>

<head>Introduction</head>

<p>The records were originally sorted at the Institute of Race Relations and its broad arrangement has been retained with some reclassification in various places. A detailed inventory is available from Historical Papers at the William Cullen Library at the University of the Witwatersrand. As the correspondence seemed to be of particular importance, it has been itemised and summarised. A brief sketch of the ANC is supplied together with an index of all personal names mentioned in the inventory and of selective subject fields.</p>

- <p>

The following abbreviations have been used:

- <list type="simple">

<item>AL Letter in the handwriting of the Author but not signed</item>

```

<item>ALS Autograph letter signed: the letter is both written and signed by the
author</item>
<item>ANC African National Congress</item>
<item>Ms Manuscript</item>
<item>AL Letter in the handwriting of the Author but not signed</item>
<item>NEC National Executive Committee</item>
<item>p Page</item>
<item>Ptd Printed</item>
<item>TL Typed letter, not signed</item>
<item>TLS Typed letter signed by the author</item>
<item>Ts Typescript</item>
</list>
</p>
</odd>

```

Notes:

- These elements are available following <did>: <admininfo>, <bioghist>, <controlaccess>, <scopecontent>, <organization>, <arrangement>, <add>.
- <odd> or Other Descriptive Data is useful for encoding old finding aids such as inventories where there is a mixture of information which is not readily separated for use in other elements. It was used here to display a list of abbreviations.
- The last paragraph of the "History of the African National Congress" has a "forward looking" statement written before the first democratic elections. This concerns methods of description, rather than encoding, but also raises the question as to how description and associated encoding method can be dynamic and accommodate changes in description. Other examples would be the addition of further materials to an open collection, a change in access restrictions or special conditions associated with viewing fragile materials.
- References to other collections were added to the description illustrating how contextual description can be extended outside the limits of the collection document. Usually EAD linking elements would be used to link directly to the relevant web pages. This could not be tested as NAREM does not consist of permanent collection-level web pages, but extracts the collection description when queried through NAAIRS, formats and displays it.

- Contextual elements offer a wide scope in recording almost any aspect of the collection description. The ANC Inventory has extensive introductory notes with the history of the ANC, the administrative history of the collection, and abbreviations. Additional information includes the physical extent, the range of dates of the material, the names of those who donated material, material of particular interest and the form of arrangement of materials.

Description of Subordinate Components

Encoding 12 - Subordinate Components - Case Study (XML)

```

<dsc type="in-depth">
  <head>Container List</head>
  <c01 level="series">
    <did>
      <unittitle>CONSTITUTION AND HISTORY</unittitle>
      <unitid>A</unitid>
    </did>
    <c02>
      <did>
        <unitid>A1</unitid>
        <unittitle>Introductory note on the constitution by R.W. Msimang,
<unitdate>1919</unitdate></unittitle>
        <physdesc>2p Ts</physdesc>
      </did>
    </c02>
    <c02>
      <did>
        <unitid>A2</unitid>
        <unittitle>Constitution, <unitdate>1 Dec. 1942</unitdate></unittitle>
        <physdesc>20p Ts</physdesc>
      </did>
    </c02>
    <c02>
      <did>
        <unitid>A3</unitid>
        <unittitle> Constitution signed by A.B.Xuma, President-General and J.A.Calata,
Secretary General, Bloemfontein, <unitdate>16 Dec. 1943</unitdate>
        </unittitle>
        <physdesc>5p Ts</physdesc>
      </did>
    </c02> ... (continued)
  </c01>
</c01 level="series">
  <did>

```

```

<unitid>B</unitid>
<unittitle>MINUTES AND OTHER PAPERS</unittitle> ...(continued)

</c01>
</dsc>
</archdesc>
</ead>

```

Encoding 13 - Subordinate Components - Union Template (SGML)

```

<dsc type="analyticover">
  <head>[Series Description]</head>
  <c01 level="series">
    <did>
      <unittitle>[Title], <unitdate type="inclusive">[Date or date
range]</unitdate></unittitle>
    </did>
    <scopecontent>
      <p>...</p>
    </scopecontent>
    <c02 level="subseries">
      <did>
        <unittitle>[Title], <unitdate type="inclusive">[date or date
range]</unitdate></unittitle>
      </did>
      <scopecontent>
        <p>...</p>
      </scopecontent>
    </c02>
  </c01>
</dsc>

```

Notes:

- The inventory description was encoded exactly into the hierarchical structure. The items that needed additional attention were those in tabular format, additional notes and paragraphs.
- Description relating to physical description such as "page 5 is missing" was placed in <physdesc>. The number of pages that is a component description component may be more appropriate in <extent>, a sub-element of <physdesc>.
- Extent is a required element for compliance with the ISAD(G).

- The first line of description following the folder number was used as <unittitle> uniformly, although some may be too long, such as "A deceptive calm prevails over south of the Sahara, says the writer, who discusses the significance of anti-white secret societies in Africa, by Louis Kraft, extracted from Star '. These were at lower levels of description, usually at item level. These may be more appropriate in a paragraph with a shorter title entered into <unittitle>.
- Some of the abbreviations for the types of materials contain fullstops, and their encoding became inconsistent. I removed all the fullstops in the abbreviations using a 'Replace All' edit, as it was proving very time-consuming to identify these errors embedded in almost 100 pages of code.
- There is an inconsistent use of fullstops and abbreviations in dates.
- It was unclear in which element the description " Note: These minutes are continued at the end of Book 1 in Part 11. " and this was placed in a paragraph after <unittitle><unitdate>
- Language Material is not applied at the item level where there are references to African languages. Unit titles which are not in English such as " Na oena oetsan G? Mahlo a ho shebile" need to have a translation added and Language Material added for greater accessibility. Some items are described as "in the vernacular" or the term "Bantu languages" is used and these should be more explicit and identify the language instance.
- The Unit Identification proved problematic to encode and this was primarily because the EAD separates the physical and logical identifiers. The Unit Identification according is "primarily a logical destination, which sometimes secondarily provides location information, as in the case of a classification number. Use... <physloc> and <container>, to designate specifically the physical location of the described materials."¹⁴² The sequence number in the inventory was initially used as Unit Identification, but later removed, as it

142 *Encoded Archival Description Tag Library*, 1998, p254

should contain, when used, "an accession number, classification number, lot number, or other such unique and permanent identifier"¹⁴³. This interpretation stemmed from the understanding that the numbering did not provide a permanent identifier, an identifier that would remain unchanged if some of the materials were to be re-arranged or removed. However it was pointed out that this is the only means of identifying and accessing a unit so was re-encoded. It is necessary to interpret the use of the <unitid> within the repository's conventions for physical storage and logical identification. Also, if the materials are moved regularly to optimise storage space usage, note that this can result in having to update the EAD documentation, if a physical location identifier is used.

The Markup of NAREM to the EAD Standard

The approach taken with this part of the case study was to enter the description sourced from the NAREM document (Document 1 in Appendix). A complete encodement was not possible as NAREM has very few discrete elements and much of the information is embedded within Remarks and Summary.¹⁴⁴ As the content of these two paragraphs is entirely dependent on what the holding institution has entered, they cannot be considered as candidates for entry into EAD elements. This approach was relaxed with some elements i.e. extent and the use of the abbreviated Description as "African National Congress". A template for the minimum EAD requirements extracted from *EAD Application Guidelines* was used to enter the elements.

Encoding 14 - NAREM Markup - Case Study¹⁴⁵

```
<ead>
  <eadheader audience="external" langencoding="iso 639-2"
findaidstatus="unverified-partial-draft"
```

¹⁴³ *Encoded Archival Description Tag Library*, 1998, p7

¹⁴⁴ This encoding was not parsed and may have syntax errors.

¹⁴⁵ Bold text indicates description sourced from NAREM. Square brackets "[" and "]" contain required but not available content or explanation of content. Some values for attributes have been added.

```

    <eadid type="catalog">PUBLIC "-/Wits University::Library//TEXT
[COUNTRY CODE ISO 3166]::[NATIONAL REPOSITORY CODE]::AD2186::African
National Congress (ANC)//EN"
    </eadid>
    <filedesc>
        <titlestmt>
            <titleproper> [Inventory of the] African National Congress
(ANC)</titleproper>
            <author>[AUTHOR]</author>
        </titlestmt>
        <publicationstmt>
            <publisher>[PUBLISHER]</publisher>
            <date>[DATE]</date>
        </publicationstmt>
    </filedesc>
    <profiledesc>
        <creation>[CREATION]</creation>
        <langusage>[Finding aid is written in]
            <language>[LANGUAGE]</language>
    </profiledesc>

```

<archdesc> encompasses the text of the archival finding aid¹⁴⁶

```

    <archdesc level="[collection]" langmaterial="[eng]" legalstatus="[LEGAL
STATUS]" level="[collection]"

        <did>
            <repository>
                <corpname>Wits University Library</corpname>
                <origination label="[Creator]"><corpname>African National
Congress</corpname> OR <persname>, <corpname>, <famname> as appropriate
            </origination>
                <origination label="[Collector]"><corpname>South African
Institute of Race Relations</corpname> OR <persname>, <corpname>, <famname>
as appropriate
            </origination>
            <unittitle>[Records of the] African National Congress
(ANC)</unittitle> .
            <unitdate>1928-1990</unitdate>
            <physdesc label="[Extent]">
                <extent>12 boxes</extent>
            </physdesc>
            <unitid> [UNITID with COUNTRYCODE and REPOSITORYCODE
attributes
            <abstract>[Abstract - N.B. A brief summary]</abstract>
        </did>

```

Elements that can be used within a number of other 'wrapper' elements including <did>

```
<admininfo>
    [subelements as appropriate]</admininfo>
<bioghist>[Historical or biographical summary]</bioghist>
<scopecontent>[Describe range etc of materials]</scopecontent>
<controlaccess>
    [subelements as appropriate]</controlaccess>
```

Description at levels below collection-level

```
<dsc> with TYPE attribute
    <c0x> or <c> with LEVEL attribute in as many levels as
appropriate
    <did>
        <container>
        <unittitle>
        other subelements as appropriate
```

The template provides for description below collection-level, not applicable to NAREM. However, if a full encoding is done, a single document should be encoded, not two, one for collection-level and one for a full encoding as this becomes difficult to maintain when changes or corrections are made to the encoding.

The *EAD Application Guidelines* for the minimum of information identifies the questions that a researcher needs to answer and the corresponding element:

Which repository holds the material? - Repository <repository>
Who created the material? - Origination <origination>
What is the title of the material? - Title of the Unit <unittitle>
When was the material created? - Date of the Unit <unitdate>
How much material is there? - Physical Description <physdesc>
What is the material's subject matter? - Abstract <abstract>

These are the most basic elements needed for *display* purposes. This emphasis is important as this data on its own would serve no useful function in data exchange, union databases or inclusion in widely used catalogues such as USMARC. The lack of South African standards for elements that reference standard headings, forms of names and international codes would also exclude its use in any joint project using automation, with overseas archiving organisations.

Conclusion

NAREM and the EAD

The conclusion drawn from this study is that the NAREM description is not useful for EAD markup. The following is suggested to bring the display requirements to EAD standards:

- Element definition should be precise, removing the general narrative elements Remarks and Summary. The choice of elements would depend on the extent to which the EAD format is adopted.
- Rules for each element should be precise, with required elements made mandatory. An example of a rule is that the definition of dates should reflect the dates when the materials were created.
- The rules must make allowance for individual collection anomalies. An example is where the material dates are unknown. An alternative could be that the date associated with description or acquisition or another relevant date could be used. Under these circumstances a 'Type' attribute is needed so the researcher knows what type of date is displayed.
- As NAREM includes materials that are not described at collection level, the Level must be displayed. Even if NAREM was confined to collection-level description, the cross-database query in NAAIRS

makes viewing confusing. The Level description is regarded as one of the most important by most archival standards organisations.

- The Labels or Headings on NAREM displays need to be aligned with international best practices, using for example Title instead of Description. The EAD illustrates the use of this in its application guidelines.
- The use of international codes and the development of local codes are a longer-term goal, discussed in Chapter 5.

The ANC Inventory Markup and the EAD

The EAD DTD provides all the description elements needed to encode the ANC inventory. Further, the structure of the DTD accommodates the hierarchical structure of the inventory and maintains the contextual relationships between items. Items presented in tabular format are accommodated, areas for leader pages or printed inventories can be encoded and requirements for publishing and contact information are fulfilled.

The *ISAD(G)* lists reference code (consisting of country, repository and collection identifiers); title; creator; date(s); extent of the unit of description; and level of description, as essential elements.¹⁴⁷ The ANC inventory generally provides all of these, except for the national and international identifiers in the reference code and an explicit reference for level of description. EAD requirements stipulate that *ISAD(G)* equivalents are mandatory. 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). For practical purposes, if the EAD document is to be used as the source for a variety of presentation formats, this information is needed at every level. This would also accommodate

¹⁴⁷ *ISAD(G)*, 2000, p9

situations where a summary level instead of item level is used for display purposes.

Some elements have 'attributes' that provide specific values for the element and are mandatory, such as "record group" or "series" for the element Archival Description. Other non-mandatory attributes provide for greater detail. Language Material is not applied at the item level in the ANC Collection where there are references to material in African languages that are "in the vernacular" and "Bantu languages". Legal Status, which refers to the policy affecting access to the unit, needs to be added. 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.

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 using creator names, such as the Library of Congress Name Authority File or the Art and Architecture Thesaurus.¹⁴⁸ This kind of standardisation is not referred to in the ANC inventory. 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. The use of codes 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.

An important feature of the EAD design is that there should be no repetition of information. 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. Description relating to

¹⁴⁸ EAD Cookbook, 2000, p1

the item content (not attributes such as number of pages, medium, etc), at levels above the individual item must not be repeated at the item level. The EAD DTD handbook also stipulates this as a rule, but technically, it cannot be reinforced, as the validation process does not check content. The onus is on the archivist or encoder to enter non-repetitive description. The examination of the ANC collection inventory raised the question as to the exact meaning of repetition. The simple definition and one used by information technologists is that the same data input should not be repeated. In the ANC collection the markup takes place from the arranged and described collection and while this paper is not intended to provide a critique on methods of arrangement, because "arrangement facilitates description; description is shaped by arrangement"¹⁴⁹, it is useful to examine an example.

In the group headed "Minutes and Other Papers", generally, the items are minutes, placed under subheadings for conferences and other meetings. An initial unit description provides the conference name and date, and is followed by "Minutes". Yet further down the group, only conference names and dates are provided and it is assumed that these items are minutes. The following record groups are headed "Resolutions and Reports", and "Memoranda and Deputations", and these items could be placed under these headings and this would provide consistently ordered description with the minutes grouped together with their attached documents, and other types of documents placed in the appropriate groups.. An alternative is the arrangement of a mix of different types of documents under different conference and meeting headings. However, these items are contained in the reporter's notebooks

The inventory's levels (A, A1, A1.1, etc) were used to develop the Archival Description hierarchy, but may not always be appropriate. An example is the breakdown within a record group and with an additional level header.

A - Book 1

¹⁴⁹ Harris, 2000, p53

A1 - Part 2

A1.1 - Unit description

If not, methods of arrangement, which fall outside the scope of this project, need to be examined, as the encoder only has the hierarchical structure provided by arrangement to interpret levels of description. In examples viewed, there appears to be a meta-definition of each level, which is consistent across all the record groups. This appears to be based on a mix, i.e. type of material (financial records, letters, etc) or event (conference, meeting).

Subject areas can be defined in the EAD DTD to provide quick search facilities across collections, facilitating rapid retrieval for the researcher. Internal subject areas need to access a broader thesaurus of terms that are common to most archival repositories. 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.¹⁵⁰ A comprehensive markup has to be tested against the criteria for successful searches across collections and repositories. This can only be done where such an infrastructure exists, and was therefore, not tested. An attempt was made to link the Related Material to the NAREM display, but proved impossible. The NAREM search engine extracts the requested description, as needed from its database and then displays this on a web page; it is not a permanent web page.¹⁵¹ 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. The depth and detail or granularity of coding is affected by considerations such as these.

¹⁵⁰ EAD Application Guidelines for Version 1.0, 1999
¹⁵¹ Clifford, 2002

In summary, the ANC Collection can be marked up to the EAD for internal use, but the results would not be suitable for use across repositories and collections. In anticipation of providing a finding aid which has a broader use, standards for codes, standardised names and subject areas need to be developed. Ideally, the effect of arrangement on description in areas such as the assignment of levels and the duplication of description needs analysis and should precede encoding, as there is a cost associated with changing the EAD document later.

CHAPTER 5

The Findings of the Research

NAREM and the EAD

If NAREM is to provide the data content and data elements for the basis of an *ISAD(G)* standard, the data must be consistent in element use to convert to a new electronic standard such as the EAD. The analysis of the markup of NAREM documents revealed a mixture of description styles, which parallels the findings of studies done in the USA and Canada in the 1980s and 1990s.

The three collection-level documents that were first examined contained detailed description, which provided the researcher with appropriate information. There was no standardisation of content design across contributing institutions, which is the result of the wide parameters provided by the input guidelines and the element design of the NAREM display screens. It must be emphasised that NAREM is intended for collection-level description while some descriptive content attempted to provide a hierarchical view of the materials by encoding detail by series. Further investigation revealed that some entries referred to item level description with no formal reference to a parent or collection level and this was achieved by using similar names for entries that belonged together. This also parallels international experience where archival description was encoded into bibliographically-based finding aids which were unable to reinforce the hierarchical structure necessary to archival description practices. The loss of contextual information, which should be inherited by the lower levels of description from the higher ones, is the result. A detailed format with more exacting rules for mandatory elements and content results is needed for consistency in the description.

As the search engine, NAAIRS, permits across-database searches, descriptions from public records were also analysed to find a related standard. Because there is no clear description on the displayed document of the source

and the level of description, the result is confusing. There is encoded meaning in a number of elements that have to be 'translated' for context and content information. These need to be standardised with more precise definitions and adhere to a recognised standard of archival terminology. The use of NAREM as a holder of collection-level description is not adhered to by all participating institutions. In some cases, this may be the result of bibliographic approaches to arranging and describing archived material as self-explanatory units.

The markup of NAREM description was not possible. An important contributing factor is that much of the content needed for markup was embedded in the narrative in Remarks and Summary, and only discretely defined elements are candidates for markup into the EAD format. NAREM does, however, offer a useful code in the collection number and a central point for viewing South African collection-level description.

The conclusion drawn is that the content, in its current form, cannot provide the consistency needed for conversion, without enhancement of the content into a more rigorous definition of data elements across all the entries. In addition, the *ISAD(G)* and the EAD provide for hierarchical levels of description; NAREM is not intended for lower levels of description and it would be more appropriate to use collection-level description, when converting to the EAD standard and applying *ISAD(G)* rules.

The EAD in Context

The EAD is a modern standard, preceded by broad agreement on archival principles such as the principle of provenance, and in the application of content rules from the *ISAD(G)*. However there are differences in application as illustrated by the Commonwealth Series System. The EAD, like the *ISAD(G)*, does not cater adequately for some variances in descriptive practices.

Multi-provenance is particularly problematic, as authority records are not included. Shepherd and Smith write, regarding the unsuitability of the *ISAD(G)* in

the description of **electronic datasets** that "it does not allow for the possibility of macro-level descriptions (biographical or administrative history) being accommodated in linked authority files, an approach which grew out of the Commonwealth Series System... Increasingly organisational change and flexibility in structures has resulted in multi-provenance records".¹⁵² Other areas of concern in the area of electronic records are that they do not have the same clearly defined boundaries as paper records. There are difficulties in identifying related filing systems, and electronic records do not naturally fall in hierarchical groupings. Electronic records can be active and changing and this activity occurs in a virtual world where there is more than one location from which records originate which may not be distinctly identifiable or even apparent. The EAD has similar weaknesses as it does not include separately-defined authority records, although the EAC or Encoded Archival Context DTD may prove to fulfil this role.

The impact of these weaknesses needs to be placed in a South African context. Older collections originating from private persons and organisations are largely paper-based and the limitations in respect of electronic records are not relevant. The need for multi-provenance is also limited to those few collections where there was dynamic change in administrative structures and the archiving convention applied to multi-provenance requires more flexibility in the EAD structure. Open private collections, collections of electronic records and public records would require more dynamic archiving solutions. The future use of the EAC DTD or the use of discrete but linked record groups could also provide part of a solution to these limitations.

The question about whether South African archival description provides appropriate description for EAD markup assumes that there are conventions and rules that can be described as 'South African'. The investigation of the NAREM markup revealed varied approaches to what constitutes description, and the conclusion reached was that there is no South African standard. In order to focus

¹⁵² Shepherd and Smith, 2000, p57

on the possibilities for a South African standard, the case study of the markup of the unofficial ANC Collection inventory description is referenced.

South African Archival Description and the EAD

The ANC inventory was cohesive and reflects arrangement and description that adheres to the principles of provenance and the maintenance of context through a hierarchy of description levels. The challenges involved in the encoding related to the lack of standardised codes and the variations in the formatting of content items such as dates and names. The description contained most of the information components needed for a full encoding, but some had to be isolated from the narrative description to fit into the EAD elements. However, local standards are needed for a consistent view within and across collections. The use of the encoded document for data exchange is contingent on standardisation of international codes and the addition of system-related codes, standardised names for identification of the local repository and country, and standardised creator name.

The EAD is not a content standard but rather defines broadly what the various elements should contain, and allows considerable licence in the selection of elements. The *EAD Cookbook* recommends the use of a national standard, which has not been developed in South Africa. Without a South African standard, my analysis was based on a combination of archival theory and debate and my own intuition. The *ISAD(G)* rules were applied as a point of reference providing further measures of the element content for the EAD. While there was no critique of the *ISAD(G)* assumptions, this does not mean the *ISAD(G)* should be accepted uncritically. Chris Hurley comments, 'One doesn't even have to disagree with their (the ICA) principles and purposes to find their standards to be theory laden and very confining.'¹⁵³ The *ISAD(G)* does, however, offer a standard that has been co-operatively developed by archivists representing several

¹⁵³ Hurley, 1998, p63

countries, who tried to define a standard that is broad enough to accommodate as many approaches as possible, constructed on the foundation of archival theory.

The *EAD Cookbook* states that "EAD encoding is not a substitute for sound archival description. Before marking up a single document, consider what information you wish to record. Descriptive conventions like *Archives, Personal Papers, and Manuscripts (APPM)*, the *Rules for Archival Description (RAD)*, and the *General International Standard for Archival Description (ISAD(G))* all have very useful things to say about what information ought to be included in a good archival description and how to structure that data. Finding aids do not differ fundamentally from other tools we create to assist our users in identifying and locating relevant materials in our collections. Consider using indexing terms drawn from standardized vocabularies in controlled elements to improve retrieval across metadata systems."¹⁵⁴ These sources recognise the importance of a content standard, based on traditional archival principles and interpreted through a well-defined formatting standard in producing useful archival description.

The translation of the ANC inventory and the investigation of NAREM highlighted the need for descriptive standards for, among others, headers and index terms, codes, language usage, naming conventions of entities, dates, types of material and extent of material. Some description such as the use of lists, of single item entries and hierarchies pointed to differing arrangement practices. In my examination of NAREM and the results of NAAIRS queries, there was at times doubt that archival principles were always applied or whether the mix of item and collection-level description resulted from a bibliographic approach. The conclusion drawn is that South African archival description is varied, some reflects a strong adherence to archival principles. Individual institutions can mark up their inventories to the EAD but a South African standard is needed to support

¹⁵⁴ *EAD Cookbook*, 2000, p4

inter-operability between institutions and participation in national and international information-sharing initiatives.

The Way Forward

Content, Context and Description

In the practice of description, archivists need guidelines in the application of archival principles in practical situations. The history of other countries' standards development processes can also inform South African archivists in their development of a South African content standard. The development of national standards in Canada, the USA and the United Kingdom covers most of the early standards-development ground and are useful sources of reference for a content standard.

The Society of American Archivists' NISTF defined archival requirements for USMARC formats, while *APPM* was written to provide content for the AACR2 format. The USA has two traditions, which include records administration and manuscript collections, with emphasis on cataloguing. The Canadian *RAD* incorporates both traditions with more emphasis on records administration.¹⁵⁵ Generally, standards development projects over the last ten years have referenced the *ISAD(G)* rules and principles and more recently the *ISAAR(CPF)*. The USA took a technology driven approach, defining standards with consideration given to data exchange functions as required by MARC, while the Canadian's took a more formal approach with studies, committees and consultations preceding the development of *RAD*. Their findings illustrate the problems that are commonly found in countries that are embarking on standards development. An example is the conclusion reached by the Canadian Working Group that "inventories and lists appeared to conform, without much regard to archival theory, to the idiosyncrasies of organization imposed on their records by

¹⁵⁵ Stibbe, 1996, p259-274

their creators or archivists".¹⁵⁶ This lack of reference to archival theory is also found in South African practices and it necessitates debate and agreement when developing local standards.

The lack of electronic encoding and little or no information exchange processes between local and overseas archival institutions is reflected in the lack of use of standardised codes, lists, headings and names. In applying standards codes, one recognises that most sources for these terms are grounded in a Western socio-political terrain.¹⁵⁷ The use of other countries' thesauri must be combined with South African and African sources that contribute towards producing description that includes the 'other' and the African, and extend and enhance international searches.

Authority work for areas such as headings may not in the past have been a general practice and the adoption of such a standard places a further discipline on the archivist. In addition to the traditional goal of establishing a single unique form of a name as an authorised heading, authority work now includes biographical and historical information and other contextual notes. Personal and corporate relationships, geographical mobility and restructuring of organisations are reflected in enhanced authority records and are necessary to provide the context for a multidimensional network of associations.¹⁵⁸ It is important that the results of authority work are shared to avoid duplication of effort. Formats and content also require standardisation with co-operation between archival organisations.

A central point for the co-ordination of standards such as these is needed as they are dynamic and need to be enhanced and added to. An example of this type of standard is the Library of Congress Name Authority File. It cannot be assumed that these headings provide for the full scope of archival description requirements. Roe describes them as inadequate to the needs of archival

¹⁵⁶ Duff, 1999, p28

¹⁵⁷ Bowker and Star, 2000, p285

¹⁵⁸ Matters, "Authority Work for Transitional Catalogs", 1990, p91-113

description for form, function, occupation, etc. and suggests that they need to be used in conjunction with other vocabulary lists and thesauri.¹⁵⁹

Standards Compatible with Goals

Any initiative towards standards development has to be geared towards future goals. The EAD is not an isolated standard and operates within a global context. South Africa has experienced years of isolation due to apartheid and a standard such as the EAD extends further than the creation of inventories and on-line displays. Data exchange opens avenues for enriching overseas research and widening the audience for previously neglected areas of South African heritage and history. A further goal is to identify and participate in related data exchange initiatives such as MARC and Dublin Core, which would require the inclusion of EAD elements relating to external data mapping requirements. Best practices in EAD usage, as put forward by the RLG Advisory Group, make the following recommendation "to make the metadata in the EAD instance as robust as possible, and to allow for crosswalks to other encoding schemes, we mandate the inclusion of the related encoding and encoding analog attributes... <eadheader> elements will commonly map to a related encoding standard such as Dublin Core, while <archdesc> elements commonly map to content standards such as MARC 21 or descriptive frameworks such as *ISAD(G)*v2"¹⁶⁰. The American Heritage Virtual Archive Project provides one example of the development of a union database¹⁶¹, a goal for South African archivists which would provide impetus in the development of a South African standard. An Australian example is the Australian Research Council's has funding the development of a new model for resource discovery of Australian manuscript collections, using the EAD to publish finding aids for collections on the Web.¹⁶²

¹⁵⁹ Roe, 1990, p157-158

¹⁶⁰ RLG *Best Practice Guidelines for Encoded Archival Description*, 2002, p2-7

¹⁶¹ *The Encoded Archival Description Retrospective Conversion Guidelines*, 1997

¹⁶² Burrows et al, 2001

These initiatives provide studies and road maps that are useful in developing a South African standard in the appropriate use of the EAD.

A South African Road Map

The establishment of a South African standard for archival description is contingent on the inclusiveness of the process. Initial steps would involve achieving consensus in the means of obtaining a standard within broadly defined goals. In order to achieve agreement, the use of a common terminology and archival principles need to be formulated. It is suggested that the rules and principles associated with the *ISAD(G)* would provide a foundation for the start of the standards development process, as most commonly used technical and content description standards have been aligned with the *ISAD(G)*.

This study recognises that the adoption of an EAD DTD goes beyond the implementation of the format. Other standards are necessary to enable inter-institutional co-operation. The standards matrix used by the WGSAD and the extensive lists and contact information provided in Walsh's *Standards for Archival Description* and other application guidelines are useful references. The development of a content standard is particularly relevant and the United States, Canada and England, among others, offer content standards and the benefit of their research, which can aid a more rapid development of a South African standard.

The designers of the EAD make very few demands in the extent of the adoption of EAD elements. A decision as to the intended use of archival description in respect of data exchange, the creation of finding aids and the administration of archives would define the extent of the development of the South African standard for EAD implementation. The experience and resulting standards and guidelines from organisations involved in the development of union databases and web-based inventories and catalogues can assist in these areas of standards development. A stepped approach in reaching this would

allow for visible progress with initial steps, for example, the definition of the standard for a shorter-term initiative such as the encoding of old finding aids.

The identification and development of national standards for headings and subject terms would necessitate the investigation of standards such as those managed by the *International Standards Organization* and the Library of Congress. Other standards would also have to be researched and their compatibility with South African needs measured. Agreement on the source of a standard, additions that are needed, and the liaison with the organisations responsible for standards administration, would assist in finding compatibility with local requirements. Other related standards used in South Africa could be consulted to provide input into archival standards identification and development, such as the bibliographic standards in use. However, the process must be driven by the South African archival community, with reference to their users' needs. The *ISAAR(CPF)* standard, which is aligned to the *ISAD(G)* for the description of creators and the area of authority control for the names of those creators also needs to be referenced as its use is dependent on co-operation between organisations. The ICA Ad-Hoc Commission on Descriptive Standards was tasked with investigating standards for specific media and an international standard number for authority records, and the effect on local standards needs to be anticipated.

An area that requires detailed definition is the rules for content formatting. The RLG best practices include the normalisation of all dates according to the *ISO 8601*, the use of punctuation and white space, standardisation on character encoding, the recommendation of the UTF-8 Unicode character set and consistency in the content of headers and labels. These standards provide consistency within a union environment.¹⁶³

Ongoing consideration of South African requirements and existing standards is needed. South African sources such as thesauri and description and

¹⁶³ RLG *Best Practice Guidelines for Encoded Archival Description*, 2002, p2-7

encoding practices are essential to ensure a local 'flavour'. Other considerations would be the management of the South African standard with the maintenance of controlled vocabularies and the publication of associated standards. The process of standards development is not static. As circumstances necessitate change, the inclusive process of development must continue. At the same time, participation in international standards development would allow for the inclusion of local considerations while enabling the maintenance of an aligned South African standard.

The development of a standard should ideally be accompanied by efforts to implement an EAD-driven description project within institutions. This is the ideal, which for many is limited by funds, time and available skills. However, a medium-term goal would be to identify a rules-driven description method, which incorporates the elements of description, in the EAD document structure, which can be used as input into future EAD projects.

The use of existing South African resources should be leveraged to enable a more rapid integration of local finding aids. It is suggested that NAREM provides a base for an initial union database. This would need further investigation as to its feasibility and the existing technology infrastructure, which is outdated, would need to be enhanced. NAREM does not provide permanent web pages for external links, which is essential to place the web pages in the reach of other search engines and for links from overseas resources to NAREM. Archiving projects have historically been under-funded, but entry into the international arena on a very visible medium, provided by the Internet, could result in more importance being assigned to such a project. Archiving organisations with automated systems, web access and digitisation experience offer experience in technology and the management of encoding which would inform standards development and implementation.

Conclusion

Standards development in countries such as Canada has taken decades. South African archivists do not need to retrace all the steps taken to develop a national standard. Experience with *RAD* highlighted a need for a common understanding with the consistent use of terminology. A common understanding and acceptance of the goals and methods used for defining a South African description standard are also prerequisites. International experience has also required description standards that are used with arrangement that is done according to accepted archival principles. This has resulted in the need for training, with better documentation and the development of primers and guides as support in the description process. It has also resulted in the development, by the International Council on Archives, of a *Dictionary of Archival Terminology*.

Access to resources such as time, funds and skills are often cited as problems in archiving projects. A stepped or phased approach with the sharing of knowledge and where possible, resources, reinforces the co-operative approach which is the foundation of the development and implementation of a national standard. Duff, in a study of the implementation of *RAD* in Canadian organisations comments that there is a "need to have a holistic approach in empowering organisations to use standards."¹⁶⁴

Technology has been shown, as in the American experience, to drive standards development. This, in a country with more limited resources, can be expensive. Hasty implementation of a new electronic standard without consideration given to the development and integration of a national standard would result in islands of EAD-defined documents with description that cannot be used across institutions or for the development of a union database. The development of a national standard from a process that is not inclusive, could result in a standard that is not widely adopted. The benefit of an efficient

¹⁶⁴ Duff, 1999, p38-43

operation of electronic systems relies on good design and well-defined and adhered to standards.

However, there are dangers in relinquishing a critical approach to standards development in the quest for this 'holy grail' or in adopting a generic or 'vanilla' standard, with no recognition of uniquely South African requirements. Internationally defined standards that do not reflect South African culture and heritage will result in impoverished description and run the risk of alienating South Africans. Examples of these are the use of terms such as 'stokvel' and 'lobola'. Another example of a standard that needs assessment within the South African context is that of the classification of a corporate body in the *ISAAR(CPF)*. This has a eurocentric approach in assigning an identity to an organisation and needs to be deconstructed to ensure its usefulness for South African organisations that do not fall neatly into its description. International standards should not dictate South African standards and it is up to the South African archival community to act as facilitators and protectors of descriptive standards that relate to South African heritage and culture.

APPENDIX A

NAAIRS Documents

Document 1 - African National Congress (ANC)

National Archives of South Africa (NASA)	
Database: National Registers of Manuscripts and Photographs (NAREM and NAREF)	
National Archives' cartographic and library material, microfilms and copies	
NAAIRS - Result Details	
Document 1 of 1	
DEPT:	Nits University Library
TYPE:	Manuscript
REFERENCE:	AD2186
DESCRIPTION:	African National Congress (ANC)
STARTING:	1928
ENDING:	1990
REMARKS:	Records; 12 boxes; inventory. Forms part of the archives of the South African Institute of Race Relations.
SUMMARY:	+ Constitution, minutes, reports, resolutions, Treason Trial, legal records, press statements, speeches, correspondence, press clippings, printed items, notes and items on the ANC Youth League and ANC Women's League together with other organisations such as the All African Convention, All African People's Conference and the South African Congress of Democrats. + Subjects are: apartheid, African nationalism, passes, education, Freedom Charter and the activities of the ANC in exile. + There is information on Albert John Luthuli, AB Mola, Nelson Mandela, Silas Modiri Molema, James S. Moroka, Godfrey Mphahlele Pitje, Pixley Ka Isaka Seme, Robert Mangaliso Sobukwe, Oliver R. Tambo and Alfred Bitini Xuma.