Implementing EAD3: Search and Exploration

Prepared by the EAD3 Study Group on Discovery
Encoded Archival Description Roundtable
August 2016

Sarah Buchanan, University of Texas
Michele Combs, Syracuse University
J. Gordon Daines III, Brigham Young University
Cory Nimer, Brigham Young University
Elizabeth Russey Roke, Emory University

Context

This report summarizes the results of the Study Group's work to review the impact of newly introduced EAD3 elements on user discovery tasks. For each element, the group has provided a discussion of the desired functionality and possible discovery interface implementations, as well as a series of recommendations on encoding. An analysis of the costs and benefits of using these elements is also provided to assist community members in their decision-making.

Additional reports on EAD3 implementation were prepared by the Study Group on Systems and Infrastructure, and the Study Group on Conversion and Migration. While some reference is made to migration concerns, this report will not address either the technical aspects of the conversion process or specific systems issues.

Process

The group's work centered on the identification of elements and attributes introduced in EAD3 that would add functionality to user discovery interfaces. These user needs were identified based on a review of the developing tag library and associated literature, as well as through a call for comments to the archival community for what they wanted from an archival discovery system. The information gathered by this review was synthesized into user stories, which focused on five areas of interest:

1. Search and sort by date
2. Search and sort by extent
3. Search by geographic location
4. Search by language

5. Improve compatibility of name/subject entries

The first of these areas (date and extent) have been included in this report under the heading Structured Data Elements. However, we decided that the question of name/subjects needed to be addressed in two different ways--through document-centric text encoding, and through the application of linked data approaches--and is addressed in the sections on Name Part Encoding as well as Linked Data.

Considerations and recommendations

The following section includes an element analysis and specific implementation recommendations for the following five groups of elements:

- Structural data elements,
- Language encoding,
- Name part encoding,
- Geolocation metadata, and
- Linked data.

While the Study Group generally recommended the use of these elements despite their added complexity, decisions on implementing these changes also depend on migration paths and system support.

Structured data elements

One of the most significant changes included in the EAD3 release was the introduction of parallel, structured versions of the EAD 2002 <unitdate> and <physdesc> elements. These structured fields were primarily designed to improve systems compatibility and to make the field content machine-actionable. By providing greater granularity to the information in these fields, the standard provides improved support for user tasks, as outlined below.

User tasks:

For date information, the following user tasks have been identified by the community:

- Date-based searching
  Researchers need to locate source materials based on a specific timeframe, either across a repository's holdings or within a given collection.

- Faceting search results
  Researchers need to narrow results based on date information when they begin a search using keywords or other field-specific search strategies.

- Collection analysis

Researchers need repository-level information on holdings pertaining to a particular time period. Institutional staff also need to review their holdings based on the dates of collection materials, including the timespan of the content.

With physical description, and especially extent information, the following user tasks have been identified by the community:

- Searching on physical characteristics
  Researchers and institutional staff need to select materials based on physical aspects such as color content (e. g., "black & white" or "color" photographs), base materials, or dimensions.

- Space management
  Institutional staff need to understand the extent of existing holdings and the space requirements of incoming collections.

- Collection analysis
  Researchers need repository-level information on the extent of holdings associated with a topical area, as well as a means of assessing the size of individual collections while planning their research. Institutional staff also need to review the extent of their holdings based on the materials' content or other characteristics (e.g., subjects, creators).

Potential use in display systems:

Dates are used in existing finding aid display interfaces, as well as in other discovery systems (retail, travel, inventory management) in the following ways:

- Search
  Many finding aid display systems allow researchers and institutional staff to find relevant materials based on the date of materials, either in their basic or advanced search interfaces. Some examples of date search capabilities include the finding aids sites at Princeton, Rockefeller Archives (under "more search options"), Brigham Young University (advanced search), and CONTENTdm (advanced search). Similar functionality is also common in library catalogs and discovery layers, such as SirsiDynix, Innovative Interfaces, and Ex Libris Primo. Outside the archives-library sector, date-based searching functionality may also be found in specialized book sites like Alibris (advanced search), and appear prominently in travel website search interfaces. Collection analysis tools also provide date-based search tools for reviewing and assessing repository content. Some of these visualization tools include the prototype system ArchivesZ, the Library of Congress's Viewshare platform, and the Visible Archives Series Browser.

- Faceting

Another common feature of finding aid display systems that allows researchers and institutional staff to find materials according to date is faceting of initial search results. Available systems generally support this functionality using either lists of set date ranges (e.g., New York Public Library, Online Archive of California, CONTENTdm) or an adjustable slider (e.g., Rockefeller Archives). Similar result faceting is also seen in library discovery layers such as Ex Libris Primo and Innovative Interfaces Encore, and in the collection management prototype of ArchivesZ.

- Sort
  Some finding aid display systems also support sorting search results by date to allow researchers and institutional staff to more easily locate relevant materials. The finding aid sites at Princeton and Brigham Young University allow resorting results in either ascending or descending order by date. This functionality is also provided in library catalogs, such as Innovative Interfaces Encore and Ex Libris Primo.

Physical descriptions, including extent information, are not accessed in finding aid discovery systems to the same extent as date information. However, some examples of available extent-based functionality include:

- Sort
  Some experimental systems support sorting search results by the size of collections, allowing institutional staff to more easily identify groups of relevant materials with a similar extent. This includes the ArchivesZ prototype, which allows researchers to order results by physical extent. Visible Archives' Series Browser also supports visualizations based in part on the size of an archival unit as part of the application's navigation.

- Faceting
  The ArchivesZ interface also allowed faceting of search results, eliminating collections larger or smaller than a desired size.

However, if one accepts (and reverses) the adage that time is money, then the value of extent-based sorting and faceting may be reflected in price-based faceting and sorting seen on retail and travel websites (e.g., Amazon, Kayak, Airbnb). Systems might also allow searching on physical characteristics data included as part of the physical description of materials.

Connections to DACS:

Dates (DACS 2.4)

The instructions for recording date information focus on which values to record and their presentation. They do not include provisions for the normalization or use of this data. In most cases, either the <unitdate> or <unitdatestructured> options provided in EAD3 would adequately accommodate the encoding of this information in both human-readable and machine-readable forms.

One area where problems may occur is in recording estimated dates, which at times may lead to ambiguity in encoding practice. The guidelines for estimated date ranges (DACS 2.4.12) and single dates (DACS 2.4.15) provide a range of examples for indicating supplied or approximate dates in a human-readable date expression. While both <unitdate> and <unitdatestructured> include the @certainty attribute value for indicating uncertainty, the <fromdate> and <todate> elements and their attributes (e.g., @notbefore, @notafter) included in the <unitdatestructured> model permit greater specificity for some types of estimated dates.

Extent (DACS 2.5)

The guidelines for extent provided in DACS may be interpreted as supporting both unstructured and structured descriptions. The general rules given in DACS 2.5.3 and 2.5.4 are compatible with the quantity-unit model used for the EAD3 <physdescstructured> element; however, the optional instructions in 2.5.6 may only be able to be supported using a free-text field as in the revised <physdesc>. The only principle enunciated in the standard is that repositories should be consistent in recording extent statements.

As noted in the Purpose and Scope section of DACS 2.5, physical descriptions may include other components as required by companion standards. With the revisions in EAD3, the specific encoding of these elements is now only supported in the <physdescstructured> model.

Relationship with EAC-CPF:

While there are no equivalent elements in EAC-CPF for extent information, the structured date model used in EAD3 was modeled after the <existDates> structure used in the authority standard. Due to the compatibility between the EAC-CPF and EAD3 date models, it would be particularly beneficial for integrated archival management systems to implement the EAD3 <unitdatestructured> element to allow reuse of database tables.

Encoding recommendations:

Based on the user tasks and encoding considerations above, the Study Group recommends the following elements:

Element and Attributes	Status	Comments / Application Notes	References	
			EAD Tag Library	DACS
<unitdatestructured>	**Req**, Rep	May be a single date or a date range. Optionally, repeat the <unitdatestructured> element for date ranges with significant gaps (DACS 2.4.11), or for specifying different type of dates (e.g., those with differing @unitdatetype or @datechar values).	p. 381	2.4
unitdatetype="inclusive | bulk"	**Req**		p. 24	
datechar="broadcast | copyright | creation | | digitized | issued | publication | recordkeeping"	Rec	Terms given not specified in EAD3 standard, but controlled list would be useful for cross-repository search. List given taken from ArchivesSpace.	p. 13	
era="ce"	Rec		p. 14	
calendar="gregorian | julian"	Rec	Listed terms suggested in EAD3 standard, while other values allowed. DACS only allows Gregorian or Julian dates (DACS 2.4.5).	p. 10	
certainty="approximate | inferred | questionable"	Rec	Terms not specified in EAD3 standard, but controlled list would be useful for cross-repository search. List given taken from ArchivesSpace.	p. 11	
<daterange>	MA	Normalized date attributes should be	p. 137	

		recorded using ISO 8601 format.		
<fromdate>	**Req**		p. 189	
standarddate=	MA		p. 23	
notbefore=	MA		p. 19	
notafter=	MA		p. 19	
<todate>	**Req**		p. 376	
standarddate=	MA		p. 23	
notbefore=	MA		p. 19	
notafter=	MA		p. 19	
<datesingle>	MA		p. 141	
standarddate=	MA		p. 23	
notbefore=	MA		p. 19	
notafter=	MA		p. 19	
<dateset>	Opt	Wrapper element for complex date spans that apply to the described unit as a whole.	p. 139	
<physdescstructured>	**Req**; Rep	For parallel statements of extent, the <physdescset> element should be used. For multiple statements of extent (i.e., entries with @coverage of "part"), <physdescstructured> should be repeated.	p. 286	
physdescstructuredtype="carrier | materialtype | spaceoccupied"	**Req**	It is recommended that values of "carrier", "materialtype", and "spaceoccupied" be used. "Carrier" refers to the	p. 20	

		number of containers (e.g., boxes); "materialtype" indicates the type and/or number of the material types (e.g., audiocassettes); "spaceoccupied" denotes the two- or three-dimensional volume of the materials (e.g., linear ft.). If none of these values are appropriate, the value "otherphysdescstructuredtype" may be chosen and some other value specified in @otherphysdescstructuredtype.		
coverage="whole | part"	**Req**		p. 13	
<quantity>	**Req**	Content should be a number--no qualifiers (e.g., "approximately") should not be included in element).	p. 311	
approximate="false | true"	Rec		p. 9	
<unittype>	**Req**		p. 390	
identifier=	Rec	URI for the unit type should be included, if available. Vocabulary for standard containers, linear/cubic feet measures should be established.	p. 15	
source=	Opt	Terms should be taken from a list of available content standards.	p. 23	
<physfacet>	MA	Terms should be entered according to	p. 290	

		conventions of companion standard.		
<dimensions>	MA	Dimensions should be entered according to conventions of companion standard.	p. 152	
<physdescset>	MA; Rep	Element should be used as a wrapper only for parallel statements of extent (i.e., DACS 2.5.7).	p. 283	
coverage="whole | part"	MA		p. 13	
parallel="false | true"	MA		p. 20	

Migration considerations:

The following challenges should be considered before converting finding aids to EAD3:

1. Existing EAD 2002 encoding conventions will impact migration. For several elements there may not be a clean migration pathway. This is particularly true with regards to the new <unitdatestructured> and <physdescstructured> elements, which includes added granularity beyond that found in EAD 2002. This may complicate efforts to migrate in an automated way from the current encoding for dates and physical descriptions to the structured version of encoding possible in EAD3.

2. Finding aid creation and delivery system support for EAD3. Institutions should be sure that publication platforms and archival management tools used locally will support the rendering of structured elements, and/or the recording of dates and extents in a compatible fashion.

3. Availability of migration stylesheets to automate as much as possible the migration from EAD 2002 to EAD3. The currently available migration stylesheet from the Technical Subcommittee on Encoded Archival Standards does not support <unitdatestructured> or <physdescstructured>, and in the latter case removes subelement encoding from EAD 2002 <physdesc> elements when present.

Economic impact:

Conversion to EAD3 comes with several potential costs. These costs are all related directly to the increased complexity of encoding possible in EAD3 and include:

1. Increased training for hand-encoding.
2. The need to switch to systems that support the creation and/or publishing of finding aids in EAD3, or the need to upgrade systems to support the creation and/or publishing of finding aids in EAD3 (e.g., archival management systems).
3. Clean-up necessitated by migrating from a less complex version of EAD (EAD 2002) to EAD3.

Benefits:

While the use of the EAD3 structured data elements may cause some complications, encoding dates and physical descriptions in this way provides better support for identified user needs. Additionally, these elements promote interoperability with EAC-CPF and other bibliographic encoding standards.

In the case of dates, the greater specificity of approximate dates in <unitdatestructured> will improve search results, while the use of machine-readable date entries will enhance search and browse functionality. This latter function, supported in structured entries through the @standarddate, @notbefore, and @notafter attribute values, is also available in <unitdate> using the @normal attribute--which should remain Required or Mandatory where Applicable if a repository decides to use the unstructured form. However, the greatest benefit for using the structured date form is likely for archival management systems, as a consistent data model for describing dates for archival resources and archival creators may simplify implementation.

For physical description statements, the use of the <physdescstructured> is essential for supporting identified user tasks of search, space management, and collection analysis. Without the quantity-unit model (or any constraints on content), <physdesc> does not provide enough semantic meaning to support adequate indexing or machine manipulation. The ability to independently mark up the different parts of a physical description also maintains the level of compatibility with MARC Bibliographic Format available in EAD 2002.

Language encoding

Two types of language encoding are supported by EAD3: multilingual description (that is, where the finding aid itself contains multiple languages or scripts) and describing multilingual collection content (that is, identifying when the collection contains archival material in different languages or scripts). This section addresses both cases.

User tasks:

Encoding information about the language of descriptive content of individual elements in a multilingual finding aid makes possible the following user tasks:

- Language of display selection
  Researchers need archival descriptions to display in a language that they can understand in order to access the materials.

- Language of input selection
  Researchers need to interact with search interfaces in the language and script of their keyboard or other input devices.

Encoding information about the language of collection content makes possible the following tasks:

- Language-based searching
  Researchers need to locate source materials in a specific language, either across a repository's holdings or within a given collection.

- Faceting search results
  Researchers need to narrow search results based on language characteristics when they begin a search using keywords or other field-specific search strategies.

- Collection analysis
  Researchers need repository-level information on holdings by language content. Institutional staff also need to review their holdings based on language characteristics.

Potential use in display systems:

Information on the language of description is not directly used in existing finding aid display interfaces, although it may be used by Web browsers in rendering or interacting with online finding aids. Encoding information about the language of a document or element within a description supports the following functionality:

- Language display/input
  For online finding aids, display and input functionality is an integral component of the Web browsers used to access them. Encoding document language attributes may assist browsers such as Google Chrome and Mozilla Firefox to correctly select and render document components according to the language used. Currently, finding aids prepared in multiple languages must be coded separately and linked (e.g., Arizona State University's Alianza Hispano Americana records (MSS 322), available in English and Spanish). Similarly, browsers may use the language attributes of the document being displayed in the selection of keyboards for inputting data in search boxes within that page. In multilingual finding aids,

users might also be provided the option of selecting the display language for a given document or element.

- Translation services
  Modern Web browsers also provide users with translation services for online content. Google Chrome includes integrated translation services based on browser language settings. Mozilla Firefox and Microsoft Edge provide similar functionality through add-ons or plug-ins. Consistent encoding of document or element language attributes supports these and related services.

- Search and faceting
  Finding aid systems may wish to allow users to limit searches or facet search results by the language of the description. For example, in a multinational system that aggregates collection descriptions from many countries, an English-speaking user may only wish to display records written in English. Administrative users may also want to use faceting to determine the total number of records available in a given language. While not quite the same, searching by language content is also possible in Web search engines such as Google Web Search and Bing.

Encoding information about the language of collection content makes possible the following:

- Search
  While language-based searching is not currently a common feature of finding aid portals, other library-based systems include this functionality. Some examples of language-based search include Brigham Young University's ScholarSearch discovery layer (advanced search), CONTENTdm (advanced search), SirsiDynix, and Ex Libris Primo. In single component display systems, components with differing language content found lower in a finding aid hierarchy might also appear if searched for independently of the larger document (e.g., a single Latin text found in a largely English-language collection). Including consistent information about the language of collection materials would also improve searching in a shared discovery system, as this is a common component of library catalog records.

- Faceting
  Another common feature in discovery systems is post-search faceting by language content. In the archival community, the Princeton finding aid site includes content language as a default facet in their interface, but it is also found in library catalogs and discovery layers such as Innovative Interfaces Sierra, Ex Libris Primo, and Brigham Young University's ScholarSearch discovery layer. In single component display systems, components with differing language content found lower in a finding aid hierarchy might also appear in faceted results independently of the larger document. For administrative users, faceting may be

useful in helping determine the total number of collections in a repository with a given language content.

Connections to DACS:

For a single-level description or for the top level of a multi-level description, DACS requires information about the language(s) and script(s) of the archival material being described. This information is recorded in EAD3 in the <langmaterial> element in the <dsc>. An example would be:

> "The bulk of the collection material is in English, with some clippings in French, Spanish and Italian."

For subsequent lower levels of description, language information is assumed to be inherited and does not need to be recorded unless it differs from that of its parent level. In an ideal world, then, if a subset of the archival material (e.g., at the <c03> level) contains different language(s) and/or script(s), that information should be provided at that level, in addition to the more general statement at the top level.

While DACS includes a chapter on Description Control, it does not include guidelines for providing information about the language in which the description is written (in EAD, this is done using the <languagedeclaration> element in the <control> section, or by using @lang and @script attributes for individual elements).

Relationship with EAC-CPF:

EAC-CPF has equivalent elements for encoding the language in which the description is written (<languageDeclaration>) and the language and script in which the CPF entity being described was creative or productive (<languageUsed>). It also provides for the use of the <script> element within either of these.

Encoding recommendations:

Based on the user needs identified above, the Study Group recommends the following elements and attributes in each of the sections below, including components of the <control> section as well as other elements within an EAD3 instance.

In the <control> section of the finding aid, the Study Group recommends:

Element and Attributes	Status	Comments / Application Notes	References	
			EAD Tag Library	DACS

<languagedeclaration>	**Req**	Use to indicate the language of the *descriptive information* of the document as a whole, not the language of the material.	p. 221	
<language>	**Req**	Use to indicate the language of the *descriptive information*. Contains text. At least one is required at the top level of description. At lower levels, recommended when different from the <control> section value.	p. 218	
langcode=	**Req**	Use to record the language of the *descriptive information*. Provide an identifying code for the language according to the authoritative source identified in @langencoding. In most cases this will be a three-letter ISO 639-2b code. Highly recommended for interoperability and machine processing.	p. 16	
<script>	Rec	Identifies the writing script of the *descriptive information.*	p. 341	
scriptcode=	Rec	Use to record the script of the *material being described*. Provide an identifying code for the language according to the authoritative source identified in	p. 22	

		@scriptencoding. In most cases this will be a three-letter ISO 639-2b code. Highly recommended for interoperability and machine processing.		

Support for multilingual description was improved in EAD3 by adding @lang and @script attributes to all non-empty elements in EAD3, making it possible to explicitly state what language or script is used therein. Additionally, some elements were modified to allow them to repeat where previously they did not, thus enabling the inclusion of the same data in multiple languages.

Within individual descriptive elements of the EAD3 document, the Study Group recommends:

Element and Attributes	Status	Comments / Application Notes	References	
			EAD Tag Library	DACS
lang=	Opt	Use to indicate the language of the *descriptive information*, not the material. Recommended only if different from the language of the parent element.	p. 16	
script=	Opt	Use to indicate the written script of the *descriptive information*, not the material. Recommended only if different from the language of the parent element.	p. 22	

Finally, within <archdesc> and the <dsc> the language content of archival materials is described in the <langmaterial> element.

For <langmaterial> the Study Group recommends:

Element and Attributes	Status	Comments / Application Notes	References	
			EAD Tag Library	DACS
<langmaterial>	**Req**	Use to identify languages and scripts present in the *materials being described*. Required by DACS at top level. Must contain one or more <language> or <languageset> elements. At lower levels, recommended if applicable.	p. 225-226	4.5
<languageset>	Opt	Use to associate a language appearing *in the material being described*, with the script(s) in which it is written. Must contain at least one <language> and one <script>. Recommended if a language can have more than one script (e.g., Japanese can be written in hiragana or katana).	p. 223	
<language>	**Req**	Use to indicate the language of the *material being described*. Contains text. At least one is required at the top level of description. At lower	p. 218	

		levels, recommended if applicable.		
langcode=	**Req**	Use to record the language of the *material being described*. Provide an identifying code for the language according to the authoritative source identified in @langencoding. In most cases this will be a three-letter ISO 639-2b code. Highly recommended for interoperability and machine processing.	p. 16	
<script>	Rec	Identifies the writing script of the *materials being described.*	p. 341	
scriptcode=	Rec	Use to record the script of the *material being described*. Provide an identifying code for the language according to the authoritative source identified in @scriptencoding. In most cases this will be a three-letter ISO 639-2b code. Highly recommended for interoperability and machine processing.	p. 22	
<descriptivenote>	Opt	Use to provide general descriptive information about parent element, in this case <langmaterial>. Recommended in <control> section if more than one language is	p. 155-156	

		present in the material, to provide the user with some idea of what portion of the collection is in what language(s).		

Migration considerations:

The following challenges should be considered before converting finding aids to EAD3:

1. Existing EAD 2002 encoding conventions will impact migration. For institutions that have not used inline <language> tags in both their <langusage> and <langmaterial> elements, significant clean-up may be required. For example, if <language> tags are present in EAD 2002, as below:

    <langmaterial>The collection contains mostly <language langcode="eng">English</language> with some clippings in <language langcode="fre">French</language> and <language langcode="ita">Italian</language>.

    the migration style sheet will convert it to EAD3 as follows:

    ```
    <languagedeclaration>
        <language langcode="eng">English</language>
        <language langcode="fre">French</language>
        <language langcode="ita">Italian</language>
        <descriptivenote>
            <p>The collection contains mostly English, with some clippings in French and Italian.</p>
        </descriptivenote>
    </languagedeclaration>
    ```

    However, if the inline <language> elements are not present, the migration stylesheet transformation will appear as:

    ```
    <languagedeclaration>
        <language><!--LANGUAGE NAME NEEDED--></language>
        <descriptivenote>
            <p>The collection contains mostly English, with some clippings in French and Italian.</p>
        </descriptivenote>
    </languagedeclaration>
    ```

2. Finding aid creation and delivery system support for EAD3. Repositories should verify that publication platforms and archival management tools used locally will support the rendering of language encoding and multilingual displays, and/or the recording of language information according to the EAD3 model.

Language and script coding at an element level for multilingual finding aids is a new addition to EAD, so there is no pre-existing data in EAD 2002 in need of migration. For institutions that have previously maintained different copies of finding aids in each language, these documents would need to be merged manually.

Economic impact:

As suggested above, the costs involved in converting to EAD3 may vary widely depending on the existing EAD 2002 encoding conventions. These costs may include:

1. Updating training for hand-encoding.

2. The need to switch to systems that support the creation and/or publishing of finding aids in EAD3, or the need to upgrade systems to support the creation and/or publishing finding aids in EAD3 (e.g., archival management systems).

3. Clean-up of existing data, both for institutions needing to insert inline <language> tagging in EAD 2002 prior to migration, and for institutions merging finding aids maintained in separate files.

Benefits:

The primary goal of the changes in the <languagedeclaration> and <langmaterial> element in EAD3 is to reduce mixed content and improve machine-actionability. By encoding information about language in a more consistent fashion, discovery systems will be able to index the language of both the finding aid and the archival content to support search and faceting. Language coding will also improve users' Web experience, supporting browser language selection and auto-translation functions. While these changes may necessitate some clean-up work, there are few options available in the established migration path.

For institutions maintaining finding aids in multiple languages, the introduction of multilingual support in EAD3 should simplify the creation and maintenance of these documents. For users, multilingual documents may improve indexing and online findability of non-English finding aid content. Element-based language coding may also assist with auto-translation functionality, as well as manual language selection within the interface.

Name part encoding

The <part> element introduced in EAD3 may be used within any of the five name elements (<name>, <persname>, <famname>, <corpname>, <geogname>), and is used to identify specific components or access points within these elements. The primary purpose of <part> is to improve systems interoperability with MARC Bibliographic data. For example, consistent use of the @encodinganalog attribute as in the following example would simplify and improve the accuracy of data transformations:

```
<persname identifier="http://id.loc.gov/authorities/names/n81026857"
encodinganalog="600_1">
        <part encodinganalog="a" localtype="surname">Byron</part>
        <part encodinganalog="a" localtype="forename">George Gordon Byron</part>
        <part encodinganalog="c" localtype="title">Baron</part>
        <part encodinganalog="d" localtype="dates">1788-1824</part>
</persname>
```

In addition, the greater granularity of the information in these fields may also provide opportunities for additional user tasks, as outlined below. While many of these tasks are better supported using linked data strategies (see Linked data section below), alternative approaches may be used to expand the functionality to meet user needs.

User tasks:

The Study Group has identified the following user tasks associated with <part> element tagging:

- Name-based searching
  Researchers need to locate archival materials based on the names (or portions of names) of their creators, either across a repository's holdings or within a given collection.

- Contextual browsing
  Researchers need to understand the connections between archival creators and other corporate bodies, persons, and families.

- Record conversion
  Institutional staff may need to convert EAD3 records to the MARC Bibliographic Format.

Potential use in display systems:

As a new feature in EAD3, name segmentation functions are not commonly found in finding aid display interfaces. However, discovery systems for genealogical sites use separate name parts in the following ways:

- Search
  Genealogical websites have a strong focus on identifying records based on names appearing in those records, allowing researchers to search by forename, surname, and associated life dates. These can be searched separately, or in combination. Examples of this include comprehensive sites such as Ancestry.com, FamilySearch, and Archives.com, as well as sites like Find-A-Grave and BillionGraves. On some sites, configurable fuzzy string searching within individual name part fields is provided.

- Browse
  On genealogical websites with user-generated family trees, such as Ancestry.com or FamilySearch, researchers are able to click on linked individual records (e.g., parents, siblings) to access contextual information and linked genealogical resources.

Connections to DACS:

The guidelines provided in DACS suggest that there are two methods by which name-related information might be included as part of the description of archival materials: as an integrated part of a narrative description, or as associated access points.

At the top level of a description, DACS requires recording information about the creator of the materials (DACS 2.6), and encourages the inclusion of names as access points (see Overview of Archival Description section). Optionally, name information may also be included at lower levels of description as appropriate. The DACS standard recommends the use of controlled vocabularies, many of which may be maintained in more granular metadata standards than EAD. The appropriate use of the <part> element may ensure compatibility with these formats, such as the MARC Authority Format.

Relationship with EAC-CPF:

EAC-CPF includes a corresponding <part> element for use in the <nameEntry> element. Due to the compatibility between the EAD3 and EAC-CPF name part models, it may be beneficial for archival management systems that have authority control modules to reuse existing database tables.

Encoding recommendations:

Based on the user needs identified above, the Study Group recommends the <part> element be used primarily for MARC-based records conversion, if needed. To assist with

discovery, the group recognizes that this may be accomplished in two ways (listed in order of preference):

1. Using the <part> element in the <name>, <persname>, <famname>, <corpname>, or <geogname> element, and placing the link to the referenced entity in the @identifier attribute of the parent element (e.g., <persname>); or
2. Using the <part> element in the <name>, <persname>, <famname>, <corpname>, or <geogname> elements, and placing the link to associated names in the @identifier attributes of the individual <part> elements.

Both recommended approaches implement the <part> element in the same fashion, but the first option has the benefit of simplicity: it points to a single explicitly-related entity, and therefore can draw on the relationships defined in that linked record for additional information (e.g., other family members) which can be pulled in for display to the end user. While this potentially reduces the available specificity in searching, it also reduces the duplication of metadata between related EAD3 and EAC-CPF documents. The second option provides increased flexibility by offering links to multiple potentially-related entities (i.e., forename, surname, etc.) but since these are not exact matches but rather possible/partial matches (e.g., not all people with the same surname are in fact related), it may reduce the utility of data held in the linked records and may lead to confusion for the end user.

Specific recommendations for these elements and attributes are as follows:

Element and Attributes	Status	Comments / Application Notes	References	
			EAD Tag Library	DACS
<name>, **<persname>**, **<famname>**, **<corpname>**, or **<geogname>**	MA	Name-based elements should be used as needed as either creators or access points.	p. 249-250; p. 278-280; p. 178-180; p. 125-127; p. 197-199	2.6; Overview of Archival Description
identifier=	Rec	URI for the complete name in a controlled vocabulary or authority system should be included, if available.	p. 15	

		For more information on linking to remote resources, see the Linked data section below.		
encodinganalog=	Opt	Numeric code for equivalent MARC Bibliographic Format field may be included (i..e., 100, 600, 110, 610, 111, 611, 151, or 651), or a corresponding value from another data structure standard.	p. 13-14	
source=	Opt	Use to specify the source of the @identifier. Recommended if @identifier attribute value is not a URI. Providing the source of a name will enhance interoperability and make it easier for linked data systems to make correct connections.	p. 23	Introduction to Describing Archival Material
<part>	**Req**	At least one <part> element is required. More are optional.	p. 276-277	
identifier=	Opt	URI for a name <part> value in a controlled vocabulary or authority system should be included, if available. Providing a link to a name part (e.g., family name) will enhance browsability and make it easier for systems to make correct connections. See Linked Data section following for list of some well-known namespaces that could provide URIs for this value.	p. 15	
encodinganalog=	Opt	Alphanumeric code for	p. 13-14	

		equivalent MARC Bibliographic Format subfield may be included (e.g., a, b, c, or d), or a corresponding value from another data structure standard. Recommended if more than one <part> is used, and if @localtype is not used.		
localtype=	Opt	Use to specify the type of name part (e.g. surname, forename). Recommended if more than one <part> is used and @encodinganalog is not used. A controlled list of @localtype attribute values should be developed for consistent use within community.	p. 17	
source=	Opt	Use to specify the source of the @identifier. Recommended if @identifier attribute value is not a URI. Providing the source of a name will enhance interoperability and make it easier for linked data systems to make correct connections.	p. 23	Introduction to Describing Archival Material

Migration considerations:

The following challenges should be considered before converting existing finding aids to EAD3:

1. Existing EAD 2002 encoding conventions will impact migration. Since the previous version of the standard had no <part> element, in most cases granular name and access point segmentation will be lacking in existing finding aids. The migration stylesheet provided by the Technical Subcommittee on Encoded Archival Description (TS-EAD) does not attempt to break out names into separate <part>

elements. It simply wraps the entire contents of <#name> into a single new <part> element. It may be possible to use scripting to segment names that follow a consistent structure (such as Western European personal names' pattern of surname, given name, middle initial). Names that do not have a consistent formulation would be more difficult.

2. Finding aid creation and delivery system support for EAD3. Repositories should consider whether their local systems will allow the presentation and searching of <part> element content, and/or the recording of such elements. For institutions that use an archival management system for authority control, this data may already be in a format that includes name segmentation (e.g., ArchivesSpace). However, export routines and user interface updates may need to be updated to use the <part> element or other associated content.

Economic impact:

Costs associated with the introduction of <part> will vary depending on the approach taken to coding this information. Archival repositories should consider the level of specificity desired to support their discovery interfaces, and the extent to which existing controlled terms can be used. These costs may include:

1. Updating training for hand-encoding.

2. The need to switch to systems that support the creation and/or publishing of finding aids in EAD3, or the need to upgrade systems to support the creation and/or publishing of finding aids in EAD3 (e.g., archival management systems).

3. Clean-up of existing data, including name segmentation and adding @identifier attribute values (either at the name or the <part> element level).

Institutions should review their documented user needs, as well as their discovery system requirements and constraints, in determining the level of coding and granularity needed. Repositories desiring to segment name entries in EAD3 may incur significantly higher costs in remediating their data, either through manual review or automation. Still, providing segmented name entries is important for improved support for searching and browsing--particularly in a flat, document-based online discovery system.

Consideration should also be given as to whether the investment of time required will reap benefits later, as systems become smarter and better able to exploit more detailed underlying encoding. For <persname>, <famname>, and <corpname>, this investment may be more appropriate for large corporate or organizational archives, for repositories with sizable genealogical holdings, or for repositories whose collections are strongly interconnected (e.g., the papers of many members of several related families).

Benefits:

The introduction of the <part> element allows the accurate representation of multi-part names, and enhances interoperability with MARC-based systems (and thus, by extension, with other systems with which MARC is interoperable) by enabling accurate transformation to/from MARC. Name-based access point segmentation allows direct alignment with or mapping to LCNAF terms, although it may present some difficulties for linking in pre-coordinated subject strings based on these names.

While the study group recommends that name-based searching and browsing functions be enabled primarily through integration with more granular and detailed linked data records, it may be possible for these to be supported directly through EAD3 document encoding. By segmenting names and recording the semantic meaning of the component <part> elements, researchers will be able to perform more specific name-based searches. For example, using @localtype attributes to distinguish forenames and surnames would keep "Jefferson, Thomas, 1743-1826" and "Thomas, Dylan, 1914-1953" out of the same "thomas" search result set. Similarly, in a geographic name search a researcher could use @localtype attributes for primary and secondary administrative divisions (i.e., cities, states) to specify search criteria.

The use of name segmentation in finding aid displays may also be used for faceting and browsing through graphic or hierarchical display of entities that share part of a name. Some examples of this functionality include:

- For personal names, a researcher could use faceted browsing with the top-level facet being surname, and the second level facet being given name.
- For corporate names, a researcher could engage in faceted browsing by company name, then division, then department or unit. This approach is a form of navigation, but could also provide a visual representation of the corporate structure implicit in the archival material.
- For geographic names, a researcher could facet search results or browse by country, then state, then city.

The use of @identifier attributes in individual <part> values for different parts could further disambiguate indexed values. For example, in a personal name search a linked URI could be used to identify and disambiguate a given "Smith" surname value as being associated with a specific Smith family (e.g., "Smith (Family : Smith, Upton Treat, 1843-1925)", or http://id.loc.gov/authorities/names/no2014070762). As noted, however, relationships of this type may be best described independently of the finding aid in linked authority records (see Linked data section below).

Geolocation metadata

Another important addition to the EAD3 schema was the ability to consistently encode geographic coordinates within geographic elements, improving machine use of this data within finding aid displays. Geocoding in EAD3 is an emerging area of growth for the professional practice of archival description. However, while this functionality may be desirable, institutions should consider whether discovery systems might use linked data sources to implement similar interfaces rather than direct coding of geographic coordinates.

User tasks:

The community has identified the following user tasks associated with geolocation metadata:

- Location-based searching
  Researchers need to locate source materials based on a geographic location or region, either across an repository's holdings or within a given collection.

- Faceting or contextualizing search results
  Researchers need to narrow search results based on a geographic location or region, as well as using visualizations to contextualize both search results and individual records based on associated geographic locations.

- Collection analysis
  Researchers need repository-level information on holdings associated with a geographic location or region. Institutional staff also need to review the extent of their holdings associated with a geographic location.

Potential use in display systems:

Geolocation metadata is not widely used in existing finding aid display interfaces, though some examples are available in discovery systems in libraries. Other online interfaces (e.g., travel websites, Google Maps search) also employ geographic coordinates in ways that may be instructive for archives. Map-based functionality can be seen in the following functions:

- Browse
  A number of institutions have developed map-based interfaces to allow researchers and institutional staff to browse holdings based on geographic locations. In archives, a geographic visualization has been included in the interface of the EADitor XForm application. A common application has been visualizations of library catalog holdings, as seen in the Biodiversity Heritage Library discovery interface, the Kingston Public Library, and the geoHOLLIS Project at Harvard University. Similar interfaces have also been produced for digital libraries, such as the Digital Commonwealth's map to repository content

using Project Blacklight Maps, or the Digital Public Library of America (DPLA) map of its holdings.

- Faceting
  Narrowing search results based on geographic location can be done either through map browsing or by defining an arbitrary polygon or radius to limit results. Browsing is often used in commercial applications such as Google Maps local search or hotel searches (e.g., Hotels.com, Kayak) to allow users to explore search results geographically, zooming and panning to limit results and identify nearby items. Alternatively, some systems allow users to draw boxes or radii around an area on a map interface to limit search results. In a library setting, the geoHOLLIS Prototype website included this functionality. It can also seen in map-based resource websites such as the USGS EarthExplorer, or on realtor websites, such as Zillow.

- Contextualization
  Including geolocation metadata also allows immediate contextualization of archival collections or components based on associated places. Similar functionality can be seen in Google Web search results for organizations or locations, which display a map location as part of the Google Knowledge Graph display (e.g., National Archives Building results).

Connections to DACS:

The guidelines provided in DACS suggest that there are two methods by which geographic references might be included as part of the description of archival materials: as an integrated part of a narrative description, or as associated access points. The Overview of Archival Description section notes that geographic place names may be found in the title (DACS 2.3), name of creator (DACS 2.6), administrative/biographical history (DACS 2.7), and scope and content elements (DACS 3.1). Additional references to locations are also found in instructions for custodial history (DACS 5.1), immediate source of acquisition (DACS 5.2), existence and location of originals (DACS 6.1), existence and location of copies (DACS 6.2), and related archival materials (DACS 6.3).

According to the Overview section, it is recommended that place names included in these elements also be recorded in as controlled terms for indexing purposes. However, it may also be possible to obtain similar results by marking up textual elements with inline indexing of geographic names. In either case, DACS provides no guidance on the use of geolocation coding in conjunction with this indexing.

Relationship with EAC-CPF:

While EAC-CPF's <placeEntry> element served as an early model for EAD3 revisions, it is currently able to the only reference a single coordinate pair. Due to this limitation the

Technical Subcommittee for Encoded Archival Description (TS-EAD) determined to strengthen the <geogname> element in EAD3 rather than implementing the EAC-CPF <placeEntry> model. It might be expected that the integration of these standards under the new Technical Subcommittee on Encoded Archival Standards (TS-EAS) will lead to improved compatibility between the two standards.

Encoding recommendations:

Based on the user needs identified above, the Study Group recommends that <geogname> elements with associated geolocation metadata be recorded in EAD3 records. The group recognizes that this may be accomplished in two ways (listed in order of preference):

1. Relying on geolocation metadata found in linked records referenced in the <geogname> @identifier attribute; or
2. Using the <geographiccoordinates> element as a subelement of <geogname>, either within the <controlaccess> and <relation> wrappers, or as mixed content, inline tags in other EAD3 elements.

Specific recommendations for these elements and attributes are as follows:

Element and Attributes	Status	Comments / Application Notes	References	
			EAD Tag Library	DACS
<geogname>	Within <chronitem>, <indexentry>, and <relation>: Optional, not repeatable. Within all other elements: Optional, repeatable	For recording the name of a place, natural feature, or political jurisdiction. String values should be taken from controlled vocabularies such as Library of Congress Name Authority File (LCNAF) or Thesaurus of Geographic Names (TGN). The <geogname> element must contain at least one <part> element for encoding one or more parts of a	p. 197	Overview of Archival Description, 2.3.22, 2.6, 2.7.11, 3.1

		geographic name. Additional guidelines on the use of geographic name <part> elements is provided in the Name part encoding section above.		
identifier=	Rec	URI for the geographic name should be included, if available. Additional guidelines on the use of linked data identifiers for geographic names is provided below.	p. 15	
normal=	Opt	Used primarily for mixed content, inline tagging. Provides standardized form of the geographical name when not provided in the element itself.	p. 18	
relator=	Rec	URI or string specifying the relationship of the geographical name to the described materials. The schema does not limit possible values of @relator, but an institution could define and enforce values elsewhere if desired. Additional guidelines on the use of linked data identifiers for geographic names is provided below.	p. 21	
source=	Rec	Code values should be taken from a list of available content	p. 23	

		standards (e.g., https://www.loc.gov/standards/sourcelist/)		
<geographiccoordinates>	Opt; Rep	For recording a set of geographic coordinates such as latitude, longitude, and altitude representing a point, line, or area on the surface of the earth. Optional addition when @identifier is unavailable, or not sufficiently specific. When a URI-based identifier is present, do not use.	p. 200	
coordinatesystem=	Opt	Use @coordinatesystem to provide a commonly used code for the system used to express the coordinates. The World Geodetic System 1984 (WGS84) coordinate system in decimal notation is preferred. Required when <geographic coordinates> is used.	p. 12	

Migration considerations:

As geolocation metadata is a new addition to EAD, there is no pre-existing data in EAD 2002 in need of migration. In some cases, however, it may be possible to enhance existing <geogname> elements that lack URI-based @identifier attribute values through programmatic matching on either existing identifiers or controlled forms of names.

Economic impact:

Costs associated with the introduction of geolocation data will vary depending on the approach taken to coding this information. Archival repositories should consider the level

of specificity desired to support their discovery interfaces, and the extent to which existing controlled geographic terms can be used. In order to reduce implementation costs, the Study Group then recommends that archival repositories:

1. Add @identifier attributes to link <geogname> entries to vocabularies that already record geolocation data; and
2. In cases where a location does not exist in a controlled vocabulary for linking, or is not recorded with adequate specificity, use <geographiccoordinates> element instead.

Benefits:

By providing geolocation data for places associated with archival materials, repositories will be able to support user needs with map-based search and browsing of their holdings for external and internal users. This benefit can be gained either indirectly through linked data or directly through encoding of geographic coordinates, and institutions should consider both the encoding and systems support costs of each approach.

Linked data

Linked data, put simply, is a method for organizing and publishing data in a way that it can be queried and reused by a variety of systems and applications. Unlike XML, which is focused on publishing entire documents (such as a finding aid or a MODS record), linked data connects disparate pieces of data (perhaps not even stored in the same system) and links it together through semantic relationships. Most importantly, linked data is designed to be read and understood by both people and machines.

The richness of archival data is a strong argument for the necessity of exposing and sharing archival data as linked data where other communities can make use of it. Most archival data currently encoded in EAD cannot be easily read by linked data tools and applications. There are numerous efforts currently working towards mapping archival data to RDF and even expressing archival metadata natively in RDF. Prominent among them is the ICA Experts Group on Archival Description. Until an ontology for archival metadata is published, however, making EAD documents RDF-ready is a forward-thinking, simple, and sustainable approach to linked data. It enables generation of RDF formats through the transformation of EAD data when useful, but relies upon the EAD standard for encoding and storing the data.

The EAD3 revision explicitly anticipates linked data applications for archival data by providing new elements and expanded attribute sets, particularly identifiers and relator attributes, to store data needed to produce linked data. These new elements and attributes allow resources to be described using linked data principles while still conforming to the EAD standard. Applied thoughtfully, these new elements and attributes will potentially make the creation of RDF from

EAD documents possible. In essence, EAD3 allows archives to encode an "optimized" version of EAD with encoding that will enable automated translation processes from XML to RDF when needed or desired.

User tasks:

The Study Group has identified the following user tasks associated with linked data elements and attributes:

- Searching for controlled terms
  Researchers and institutional staff need to locate materials using authorized forms of terms, whether within a given collection, across a repository's holdings, or across multiple repositories.

- Browsing
  Researchers and institutional staff need to explore and discover relationships between entities, including related creators, subjects, functions, and resources.

- Contextualizing records
  Researchers and institutional staff need to have record displays augmented with contextual information on related creators, resources, and other entities.

Potential use in display systems:

Since Tim Berners-Lee coined the term in 2006, linked data and the related RDF data format has slowly been gaining a foothold in domains from commercial websites to libraries. While it has not been widely implemented in finding aid displays, some examples of of linked data functionality are available in the following functional areas:

- Search
  While the integration of linked record content in discovery layer search indexes has the potential to enhance retrieval, due to the lack of archival linked data implementations there are few available archival examples. The Jisc LOCAH and Linking Lives projects provided a proof-of-concept for how this data might be used. In the library sector, the open source VuFind project has explored how to integrate linked data indexing, while SirsiDynix has recently partnered with Zepheira to develop its BlueCloud Visibility service to improve searching and discoverability through linked data.

- Browsing and Contextualization
  A more visible feature of linked data interfaces is the display of data from linked records, and the surfacing of associated relationships between records. Archival interfaces such as OCLC's EntityJS interface to ArchiveGrid data and the Jisc Linking Lives Project include biographical information and photographs from

external records. Links to associated records (e.g., creators, topics, places, and resources) are also available for researchers to browse, allowing them to explore related materials. Similar functionality (though not based specifically on linked data) is also available in the SNAC prototype. Outside of the archives and library space, the Google Knowledge Graph is a prominent linked data application which uses data from external sources to create information boxes of information related to a search result.

Connections to DACS:

Although DACS does not address linked data specifically, it notes the importance of recording information about related entities, including names (corporate bodies, persons, and families), functions, and related archival material and other resources. The following sections of the standard are of particular applicability:

- Creator (DACS 2.6)
  The guidelines for recording creators in DACS focus on the use of access points, which are meant to assist in searching and collocation of related materials. While the Purpose and Scope portion of the element instructions focuses on consistency in form and the use of controlled vocabularies to support text string indexing, linked data approaches could provide equivalent functionality. The guidelines also include optional provisions for indicating relationships between creators and archival materials using controlled terms.

- Biographical/Administrative History (DACS 2.7)
  DACS provides two options for recording and maintaining information about the creator: incorporation into the description using a biographical/historical note, or in a separate system of authority files linked to the archival description and displayed together. A portion of the guidelines for this element includes recording information about the relationships of the creator to other people (DACS 2.7.19, 2.7.20, 2.7.32), organizations (DACS 2.7.17, 2.7.18, 2.7.19, 2.7.28, 2.7.29, 2.7.30), or geographical locations (DACS 2.7.16, 2.7.25), as well as functional relationships (DACS 2.7.27). As suggested in the element commentary, it may be more efficient to record some of the components of the description within a linked authority record connected to the resource, rather than within the EAD file itself..

- Related Materials Elements (DACS 6)
  DACS also encourages adding to finding aids information about the existence and location of materials closely related to the materials being described in the finding aid, whether by provenance, sphere of activity, or subject matter. This also encompasses the existence and location of copies, originals, and published versions of materials in the collection. While the examples provided in Chapter 6 of DACS only include unstructured or structured descriptions of these related

materials, these might be expressed in a linked data-friendly form using the <relations> element.

- Access Points
  In the Overview of Archival Description section, DACS discusses the creation of access points, which may take several forms: names, places, subjects, documentary forms, occupations, and functions. These access points are designed to facilitate searching by providing standard fields for indexing as well as the form in which they appear. These access points have typically been recorded in EAD 2002 in the <origination> and <controlaccess> portions of the record, though in EAD3 these could also be placed in the <relations> area.

Relationship with EAC-CPF:

The <relations> element set in EAD3 duplicates the <relations> element guidelines laid out in EAC-CPF, and its addition to EAD3 is intended to ensure compatibility between the two standards. This similarity in design should allow archival management system developers to use a common approach for linking records internally, as well as reducing the need for and training in hand-encoding of these types of records.

Encoding recommendations:

Based on the user needs identified above, the Study Group recommends that repositories implement linked data principles in their EAD3 encoding. The group suggests that this may be done in two ways:

1. Referencing linked data URIs in access points using the @identifier and @relator attributes; and/or
2. Using the <relations> element to record linked data relationships.

Specific recommendations for these these approaches are below:

Identifiers and Relators

A major principle of linked data is to use dereferenceable (i.e., interpretable, or actionable) URIs when possible. This is usually done using @identifier and/or @relator attributes of the relevant element. The encoding of identifiers and relationships for entities and resources within EAD3 is key to producing a linked-data-friendly document. The two work in tandem to define the relationships necessary for generating linked data. Although populating a finding aid with identifiers and relationships will NOT produce linked data without some additional work, it is a good head start and makes data easily harvestable and reusable.

Even in repositories that do not intend to use linked data, the addition of identifiers to EAD3 data should be standard practice. It is akin to validating subject and creator access points in an OCLC record. The addition of identifiers to EAD3 documents is an important first step in optimizing it for transformation to linked data structures, but it is not the only benefit. Search and discovery/display systems for finding aids can make use of identifiers for improved search and browsing capabilities, much as an authority database does within an integrated library system.

It is recommended that repositories add @identifier and @relator attributes to EAD3 data using widely used, linked data-friendly namespaces whenever possible. Repositories that intend to produce linked data from EAD3 must use linked-data friendly namespaces, but even repositories that simply want to encode identifiers should do so as a best practice to optimize their data and to reduce the need to re-encode data in the future.

Below is a list of vocabularies that are widely used/supported both in the library and outside communities, and are available in RDF. The dereferenceable HTTP prefix form of the term rather than an internal identifier string should be used when possible. If a URI is not available, it is recommended that the @identifier attribute be used in conjunction with the @source attribute.

@identifer Attribute Value Vocabularies Value records URI location of related record.	
EAD3 Element	**Controlled Vocabulary Sources**
<persname>	Library of Congress Name Authority File (http://id.loc.gov/authorities/names.html); Virtual International Authority File (http://www.viaf.org/); DBpedia (http://wiki.dbpedia.org/)
<famname>	Library of Congress Name Authority File (http://id.loc.gov/authorities/names.html); Virtual International Authority File (http://www.viaf.org/); DBpedia (http://wiki.dbpedia.org/)
<corpname>	Library of Congress Name Authority File (http://id.loc.gov/authorities/names.html); Virtual International Authority File (http://www.viaf.org/); DBpedia (http://wiki.dbpedia.org/)
<name>	Library of Congress Name Authority File (http://id.loc.gov/authorities/names.html); Virtual

	International Authority File (http://www.viaf.org/); DBpedia (http://wiki.dbpedia.org/)
<subject>	Library of Congress Subject Headings (http://id.loc.gov/authorities/subjects.html) Recommended practice is to record multiple simplified Library of Congress Headings or validation records with a single, defined URI, or to use the (experimental) OCLC FAST ontology (http://id.worldcat.org/fast/)
<title>	Library of Congress Name Authority File (http://id.loc.gov/authorities/names.html); DBpedia (http://wiki.dbpedia.org/)
<geogname>	Geonames (http://www.geonames.org); Library of Congress Name Authority File (http://id.loc.gov/authorities/names.html); Thesaurus of Geographic Names (http://www.getty.edu/research/tools/vocabularies/tgn/); DBpedia (http://wiki.dbpedia.org/)
<genreform>	Library of Congress Genre/Form Terms (http://id.loc.gov/authorities/genreForms.html); Library of Congress Subject Headings (http://id.loc.gov/authorities/subjects.html); Art & Architecture Thesaurus (AAT) (http://www.getty.edu/research/tools/vocabularies/aat/); pbcoreGenre (http://metadataregistry.org/concept/list/page/1/vocabulary_id/148.html)
<physfacet>	Art & Architecture Thesaurus (AAT) (http://www.getty.edu/research/tools/vocabularies/aat/); pbcorePhysicalFormat (http://metadataregistry.org/concept/list/vocabulary_id/145.html)
<function>	Art & Architecture Thesaurus (AAT) (http://www.getty.edu/research/tools/vocabularies/aat/)
<occupation>	Library of Congress Subject Headings (http://id.loc.gov/authorities/subjects.html); Art & Architecture Thesaurus (AAT) (http://www.getty.edu/research/tools/vocabularies/aat/); DBpedia (http://wiki.dbpedia.org/)
<term>	Recommended practice is to establish a local

	vocabulary of authorized terms and publish as URIs

@relator Attribute Value Vocabularies Value records URI of description of relationship.
RDA Registry (http://www.rdaregistry.info)
MARC Relators Code List (http://id.loc.gov/vocabulary/relators)
Schema.org (http://schema.org)
pbcoreCreatorRole (http://metadataregistry.org/concept/list/page/1/vocabulary_id/128.html)
pbcoreContributorRole (http://metadataregistry.org/concept/list/page/1/vocabulary_id/127.html)

<relations> Element Set

Once identifiers for resources have been added to the finding aid, to make an EAD description linked data-ready institutions should define the relationships between the collection and its related entities. Precisely defining these relationships semantically is a fundamental principle of linked data. This is done in EAD3 using the <relations> element set.

The <relations> element set encodes semantic relationships between resources, both internal and external, and can be a powerful way to describe these relationships precisely. It is a data-centric approach to description that assumes the EAD3 finding aid encodes data as well as the finding aid document itself. The <relations> element set in combination with the standardized attribute set for names, functions, titles, and subjects can be used to optimize EAD3 encoding for translating EAD3 data to RDF.

The Study Group recommends that the <relations> element be used to ensure compatibility with linked data, rather than as a replacement for simple hyperlinking. Organizations looking to create simple links between documents/resources without encoding specific semantic relationships (e.g., connecting a folder in one EAD3 document to another collection) should use other linking elements within EAD3 instead, such as <dao>, <ptr>, or <ref>. Unlike the recommendations for @identifier and @relator that encourage broad use, the

<relator> element set requires an understanding of linked data constructions and ontologies and may not be appropriate in all cases.

The following guidelines suggest some strategies and best practices for using <relations> and its associated elements and are focused on the question of how to generate linked data from an EAD document. The data model considered here assumes that information about resources encoded within a finding aid are linked to external authority records and other relevant data sources whenever possible.

Element and Attributes	Status	Comments / Application Notes	References	
			EAD Tag Library	DACS
<relations>	Rec (as appropriate)	Required wrapper element for using the relations element set. Can occur at <archdesc> and all <c0x> levels.	p. 325-330	
<relation>	**Req**	Required child element of <relations>. Describes the relationship between the resource and the related entity named in <relationentry>.	p. 321-322	
relationtype="cpfrelation | functionrelation | resourcerelation | otherrelation"	**Req**	Fixed value set that encodes the type of relationship: cpfrelation (for names), functionrelation (for functions), resourcerelation (for related collections, bibliographic resources), and otherrelationtype (places, events, topics)	p. 21	
arcrole=	Req	URI that defines the specific relationship between the EAD resource and the resource defined in @href. Values may be taken from vocabularies listed in @relator table above.	p. 10	

href=	Req	URI that links to the related resource. Values may be taken from the vocabularies listed in the @identifier table above.	p. 15	
linkrole=	Opt	URI that defines the nature of the related resource (person, collection, etc.). Do not use if @href points to a linked data resource.	p. 17	
<relationentry>	Opt	Human-readable description of the related resource. This should generally be in the form of a preferred access point.	p. 323-324	
<descriptivenote>	Opt	Extended description of relationship, local display purposes only.	p. 145-146	

Other attributes not listed above are optional. Elements such as <relationentry> and <descriptivenote> may be helpful for local systems. Both are designed to provide human-readable text, but are not needed for the production of linked data. <relationentry> is redundant if URIs are encoded in the <relation> attributes and point to a resource encoded in linked data formats such as RDF or json-ld because this information can be harvested from the data source. However, it may be useful if the linked resource does not contain linked data.

General Principles and Best Practices for Applying <relations>

- *Use URIs whenever possible.* All recommended attributes within the <relations> element should contain URIs. This guideline fulfills a fundamental linked data principle to use URIs as names for things. When recording relationships in other elements, please note that although the @relator attribute does not specifically allow a URI and they are not actionable, well-constructed URIs would be valid. Future revisions of EAD3 should consider allowing URIs in all appropriate attributes, especially the @relator attribute.

  If a URI does not exist, consider minting one. Repositories should follow best practices, publishing terms using either a local vocabulary server (e.g., TemaTres) or an online service (e.g., Open Metadata Registry). For cross-institutional needs, the archival community should consider establishing common vocabularies.

- *Define the semantic relationship as specifically as possible.* Take the time to explicitly define the relationship between the entity and the related resource each time this element is used rather than relying on default values in the @relationtype attribute. At a minimum it is recommended to use the @arcrole, @href, and @relationtype attributes. Defining a more specific relationship takes minimal effort but results in vastly more useful data.

  Consider this example in EAD3:

```
<archdesc>
        <did>
                <unittitle>Southern Christian Leadership Conference
                records</unittitle>
                [....]
        </did>
        <relations>
                <relation relationtype="cpfrelation"
                href="http://viaf.org/viaf/100170140">
                        <relationentry>King, Martin Luther,
                        1929-1968</relationentry>
                 </relation>
        </relations>
</archdesc>
```

  Which through an EAD3 to RDF transformation results in (in Turtle serialization):

```
@prefix ns1: <https://schema.org/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix schema: <http://schema.org/> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

<http://pid.emory.edu/ark:/25593/90wfs> schema:mentions
<http://viaf.org/viaf/100170140>
```

  This transformation generates only a single relationship and that relationship, which has to be assumed by the transformation code (schema:mentions), is fairly general. In cases where @relationtype="otherrelationtype", no relationship could be assumed at all, rendering the <relations> encoding useless for RDF serialization.

  Now consider the same example with the addition of the @arcrole attribute in the EAD3 encoding:

```
<archdesc>
        <did>
```

```
                <unittitle>Southern Christian Leadership Conference
                records</unittitle>
                [....]
        </did>
        <relations>
                <relation relationtype="cpfrelation"
                arcrole="http://http://schema.org/Creator"
                href="http://viaf.org/viaf/100170140"
                linkrole="http://schema.org/Person">
                        <relationentry>King, Martin Luther,
                        1929-1968</relationentry>
                </relation>
        </relations>
</archdesc>
```

These additions provide the more accurate RDF transformation below (in Turtle serialization):

```
@prefix ns1: <https://schema.org/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix schema: <http://schema.org/> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

<http://pid.emory.edu/ark:/25593/90wfs> schema:mentions
<http://viaf.org/viaf/100170140> ;
 schema:creator <http://pid.emory.edu/ark:/25593/900g7> .

<http://viaf.org/viaf/100170140> a schema:Person .
```

This encoding has produced three sets of triples, defining the type of related resource, the specific relationship, and the general relationship defined in the first example. Most importantly, because all of these relationships were encoded in the EAD, the transformation code does not need to make any assumptions that might be incorrect or too general.

It is worth mentioning here that there is a great need for the archives community to develop ontologies and vocabularies specifically suited to archival relationships. Although existing vocabularies should be utilized when possible, there is room for the archives community to define our domain more specifically and accurately.

- *Do not duplicate information in <relations> that can be found in an EAC-CPF record.* Rather, reference the EAC-CPF record using @identifier attributes within names, topics, and functions elements instead of redefining the relationship directly in the EAD3 record.

- *Consider local impacts when using <relations>.* As the Technical Subcommittee on Encoded Archival Description notes in the introduction to EAD3, <relations> is an experimental element and best practices for this element set are still being determined. It was included in EAD3 to accommodate data-based approaches to description while still supporting document-based description. Institutions that choose to use <relations> may wish to also record some descriptive elements that serve as access points (names, places, subjects) in the standard <controlaccess> location to accommodate these differing approaches to description and to ensure backwards-compatibility with EAD 2002 data structures and systems.

  The following examples suggest some ways <relations> can be used, each satisfying different use cases. All of these examples should be able to support linked data serialization.

  Mixed attribute reuse and Relations
  This example uses <relations> at the container level, but not at the collection level where some relationships (subject, origination) might be assumed by the structure of the EAD3 document. This example encodes linked data relationships while retaining traditional document-based finding aid elements for front-end display.

```
<archdesc level="collection">
    <did>
        <unittitle>Michael Longley papers</unittitle>
        <origination>
            <persname relator="http://schema.org/creator"
            identifier="http://id.loc.gov/authorities/names/n50044
            363" encodinganalog="100_1">
                <part encodinganalog="a"
                localtype="surname">Longley</part>
                <part encodinganalog="a"
                localtype="forename">Michael</part>
                <part encodinganalog="d">1939- </part>
            </persname>
        </origination>
        [...]
    </did>
    <controlaccess>
        <corpname relator="http://schema.org/about" identifier="
        http://id.loc.gov/authorities/names/n84236929"
        encodinganalog="610_2">
            <part encodinganalog="a">Arts Council of Northern
            Ireland</part>
        </corpname>
        <genreform relator="http://schema.org/genre"
        identifier="http://id.loc.gov/authorities/subjects/sh85080672"
        >
```

```
                    <part>Manuscripts</part>
                </genreform>
            </controlaccess>
            [...]
                <c02 level="file">
                    <did>
                        <container type="box">1</container>
                        <container type="folder">22</container>
                        <unittitle>Edna Longley manuscript of The
                        Living Stream</unittitle>
                    </did>
                    <relations>
                        <relation relationtype="cpfrelation"
                        arcrole="http://schema.org/creator"
                        href="http://id.loc.gov/authorities/names/n81
                        147007">
                            <relationentry>Longley,
                            Edna</relationentry>
                        </relation>
                    </relations>
                </c02>
            [...]
        </archdesc>
```

Relations only

This example takes a data-centric approach to EAD description. It may not be compatible with some EAD 2002 documents and systems because origination and subject access points are encoded within <relations> and will require some transformation to produce a traditional finding aid document for display.

```
        <archdesc level="collection">
            <did>
                <unittitle>Michael Longley papers</unittitle>
            </did>
            <relations>
                <relation relationtype="cpfrelation"
                arcrole="http://schema.org/creator"
                href="http://id.loc.gov/authorities/names/n50044363">
                    <relationentry>Longley, Michael, 1939-
                    </relationentry>
                </relation>
                <relation relationtype="cpfrelation"
                arcrole="http://schema.org/about"
                href="http://id.loc.gov/authorities/names/n84236929">
                    <relationentry>Arts Council of Northern
                    Ireland</relationentry>
                </relation>
```

```
                <relation relationtype="otherrelationtype"
                arcrole="http://schema.org/genre"
                href="http://id.loc.gov/authorities/subjects/sh85080672">
                    <relationentry>Manuscripts</relationentry>
                </relation>
            </relations>
            [...]
                <c02 level="file">
                    <did>
                        <container type="box">1</container>
                        <container type="folder">22</container>
                        <unittitle>Edna Longley manuscript of The
                        Living Stream</unittitle>
                    </did>
                    <relations>
                        <relation relationtype="cpfrelation"
                        arcrole="http://schema.org/creator"
                        href="http://id.loc.gov/authorities/names/n81
                        147007">
                            <relationentry>Longley,
                            Edna</relationentry>
                        </relation>
                    </relations>
                </c02>
            [...]
        </archdesc>
```

Attributes only

This example encodes relationships information in EAD3 without using <relations> at all. This encoding is not as precise as <relations>, but may be easier to implement in environments that need to support both EAD 2002 and EAD3 documents.

```
        <archdesc level="collection">
            <did>
                <unittitle>Michael Longley papers</unittitle>
                <origination>
                    <persname relator="http://schema.org/creator"
                    identifier="http://id.loc.gov/authorities/names/n50044
                    363" encodinganalog="100_1">
                        <part encodinganalog="a"
                        localtype="surname">Longley</part>
                        <part encodinganalog="a"
                        localtype="forename">Michael</part>
                        <part encodinganalog="d">1939- </part>
                    </persname>
                </origination>
                [...]
```

```
                </did>
                <controlaccess>
                        <corpname relator="http://schema.org/about" identifier="
                        http://id.loc.gov/authorities/names/n84236929"
                        encodinganalog="610_2">
                                <part encodinganalog="a">Arts Council of Northern
                                Ireland</part>
                        </corpname>
                        <genreform relator="http://schema.org/genre"
                        identifier="http://id.loc.gov/authorities/subjects/sh85080672"
                        >
                                <part>Manuscripts</part>
                        </genreform>
                </controlaccess>
                [...]
                        <c02 level="file">
                                <did>
                                        <container type="box">1</container>
                                        <container type="folder">22</container>
                                        <origination>
                                                <persname
                                                relator="http://schema.org/creator"
                                                identifier="http://id.loc.gov/authorities/
                                                names/n81147007"
                                                encodinganalog="100_1">
                                                        <part encodinganalog="a"
                                                        localtype="surname">Longley
                                                        </part>
                                                        <part encodinganalog="a">
                                                        localtype="forename">Edna</
                                                        part>
                                                </persname>
                                        </origination>
                                        <unittitle>Edna Longley manuscript of The
                                        Living Stream</unittitle>
                                </did>
                        </c02>
                [...]
        </archdesc>
```

Migration considerations:

Institutions should consider the following challenges related to linked data before pursuing EAD3 conversion:

1. Stylesheet migration of existing linked data-friendly finding aids. For institutions already using EAD 2002 element attributes to record URIs, the existing migration stylesheet provided by TS-EAD should be adequate to address relevant changes in EAD3. These include small changes to the attribute set, replacing the @role attribute with @relator and the @authfilenumber attribute with @identifier. For institutions interested in migrating existing attribute-based linked data entries to the <relations> element, additional stylesheet development and data remediation may be needed.

2. Finding aid creation and delivery system support for EAD3. Content management systems should consider data structures that enable archivists to describe information once and then serialize to EAD3 or RDF. Ideally, the data elements recorded within the <relations> element should be available throughout the description to ensure the most specific data possible can be recorded. For instance, a subject heading for a finding aid may be described as:

```
<archdesc>
        <did>
                <unittitle>Frank Ormsby papers</unittitle>
        </did>
        <controlaccess>
                <persname identifier="http://viaf.org/viaf/109956735">
                        <part>Carson, Ciaran, 1948-</part>
                </persname>
        </controlaccess>
</archdesc>
```

Or using <relations>:

```
<archdesc>
        <did>
                <unittitle>Frank Ormsby papers</unittitle>
        </did>
        <relations>
                <relation relationtype="cpfrelation"
                arcrole="http://schema.org/about"
                href="http://viaf.org/viaf/109956735"
                linkrole="http://schema.org/Person">
                        <relationentry>Carson, Ciaran,
                        1948-</relationentry>
                </relation>
        </relations>
</archdesc>
```

Encoding within the <relations> element expresses the same relationship and gives content creators much more semantic control. Ideally, content management systems should follow the <relations> model for descriptive best practice, allowing users to define relationships to the creator or collection (@arcrole attribute), the URI for the resource (@href attribute), and the type of resource (@linkrole attribute) rather than traditional descriptive elements such as subject headings, creator, and related resources. Using the <relations> element--an element set meant to implement linked data--should make RDF serialization simple.

For EAD serialization, EAD documents will probably need to continue to contain both <relations> and the older EAD 2002-style elements. Even though the information is semantically the same, both types of encoding are necessary to accommodate differing approaches to description and to ensure compatibility with EAD 2002 data structures.

Economic impact:

Implementing linked data approaches following conversion to EAD3 comes with several potential costs. These include:

1. Increased training for hand-encoding.
2. The need to switch to systems that support recording linked data attributes and/or integrating linked data content in EAD3 finding aid displays, or the need to upgrade existing systems to support recording and/or publishing linked data content in EAD3 finding aids (e.g., archival management systems).
3. Data clean-up following migration to EAD3 for insertion of linked data values.

The <relations> element offers better description and optimizes finding aids for future RDF transformation. However, in a hand-coded environment they are labor-intensive to record. Institutions will have to consider whether the benefits outweigh the cost, which will depend on whether they intend to create/use linked data or develop robust systems that pull data from different sources together within a single user interface.

Benefits:

Archival information systems would benefit from implementing fields to record the information needed for the <relations> element as recommended above, as well as the ability to use URIs throughout the system where appropriate. These enhancements would enable archival institutions to participate in and contribute to the linked data environment, connect to other archival institutions, and improve metadata accuracy.

Using linked data would support discovery and use of archival materials in a number of ways. As noted above, for existing discovery systems the benefits might include improved search and browse functionality based on linked data content due to the reduced

dependence on text string matching. Linked data-based applications would also be able to integrate external content into displays, providing better contextualization and browsing of records. For archival management systems, using linked data services would also provide an opportunity for supporting automated updating and enhancement of locally cached records.

In addition to these discovery-focused benefits, integrating linked data principles into EAD3 encoding could simplify descriptive practice and improve compatibility of archival metadata. Using the <relations> element in particular would provide greater descriptive precision for archival description, while referencing standard vocabularies encourages greater connectivity among archival and library resources. This may be particularly true for archival management systems, where a consistent data model and standard values would simplify interfaces and unify system functionality. Employing a linked data model will also allow for easier transformation of EAD3 records into RDF or other formats, allowing archives to share their finding aid content more widely--improving both access and discoverability.