Study Group members

Sarah Buchanan, University of Missouri
Michele Combs, Syracuse University
J. Gordon Daines III, Brigham Young University
Cory Nimer, Brigham Young University
Elizabeth Russey Roke, Emory University
Changes in EAD3

Technical Subcommittee on Encoded Archival Description revision goals:
- Greater conceptual and semantic consistency;
- Connection with, exchange, or incorporation of data from other protocols;
- Multilingual support;
- Usability.

Significant areas of revision:
- `<control>` element;
- Structured data elements (e.g., `<unitdatestructured>`, `<part>`, etc.);
- Language elements and attributes;
- Simplified linking elements;
- `<relations>` element
Study Group process

Focus on user discovery tasks:
1. Search and sort by date
2. Search and sort by extent
3. Search by language
4. Improve compatibility of name/subject entries
5. Search by geographic location
Date elements

User tasks:
- Searching by date
- Faceting by date
- Sorting by date
Date elements

Recommendation:
● Use `<unitdatestructured>` and associated elements

Justification:
● Improved correspondence between normalized and display values
● Increased specificity of approximated dates
● Data model consistency with EAC-CPF

```xml
<unitdatestructured unitdatetype="inclusive" datechar="creation" era="ce" calendar="gregorian" certainty="approximate">
  <daterange>
    <fromdate notbefore="1917-04">approximately 1917 April</fromdate>
    <todate notafter="1918-11">approximately 1918 November</todate>
  </daterange>
</unitdatestructured>
```
Date elements

Challenges:

- Existing data
- System support
- Migration tools

EAD 2002

<unitdate type="inclusive" datechar="creation" era="ce" calendar="gregorian" certainty="approximate" normal="1917-04/1918-11"> approximately 1917 April-approximately 1918 November</unitdate>

EAD3

<unitdatestructured unitdatetype="inclusive" datechar="creation" era="ce" calendar="gregorian" certainty="approximate"> <daterange> <fromdate standarddate="1917-04"> approximately 1917 April</fromdate> <todate standarddate="1918-11"> approximately 1918 November</todate> </daterange> </unitdatestructured>
Extent elements

User tasks:

- Searching by physical characteristics
- Faceting on physical characteristics
- Collection analysis

Extent elements

Recommendation:
- Use `<physdescstructured>` and associated elements

Justification:
- Improved correspondence between normalized and display values
- Increased ability to reuse physical extent information
- Data model compatibility with MARC Bibliographic Format

```xml
<physdescset parallel="true" coverage="whole">
  <physdescstructured physdescstructuredtype="carrier">
    <quantity approximate="false">1</quantity>
    <unittype>box</unittype>
  </physdescstructured>
  <physdescstructured physdescstructuredtype="spaceoccupied">
    <quantity approximate="false">0.5</quantity>
    <unittype>linear ft.</unittype>
  </physdescstructured>
</physdescset>
```
Extent elements

Challenges:
- Existing data
- System support
- Migration tools

EAD 2002

```xml
<physdesc>
  <extent>1 box</extent>
  <extent>(0.5 linear ft.)</extent>
</physdesc>
```

EAD3

```xml
<physdescset parallel="true" coverage="whole">
  <physdescstructured physdescstructuredtype="carrier">
    <quantity approximate="false">1</quantity>
    <unittype>box</unittype>
  </physdescstructured>
  <physdescstructured physdescstructuredtype="spaceoccupied">
    <quantity approximate="false">0.5</quantity>
    <unittype>linear ft.</unittype>
  </physdescstructured>
</physdescset>
```
Language elements

User tasks:

- Language of archival description
  - Browser recognition and adjustment of display or input
  - Translation services
  - Searching and faceting in aggregate systems such as ArchiveGrid, ArchivesHub, Archives Portal Europe

Declaring language in HTML

Intended audience: HTML coders (using editors or scripting), script developers (PHP, JSP, etc.), Web project managers, and anyone who needs to better understand how to declare the language of text on a Web page.
Language elements

User tasks:

- Language of archival material
  - Targeted location of source materials in a given language
  - Faceting search results by language of content
  - Collection analysis
Language elements

Recommendations:

- Language of archival description
  - profiledesc/ langusage/language no longer used
  - Encode at top level in the <control> section using <languagedeclaration>, <language> (ISO 639-2b), and <script> (ISO 15924) elements
  - Example: Finding aid is in English, written using standard Latin script:
    ```xml
    <languagedeclaration>
      <language langcode="eng">English</language>
      <script scriptcode="latn">Latin</script>
    </languagedeclaration>
    ```
  - Example: Finding aid in Russian, written using Cyrillic script:
    ```xml
    <languagedeclaration>
      <language langcode="rus">Русская</language>
      <script scriptcode="cyril">Кириллица</script>
    </languagedeclaration>
    ```
Language elements

Recommendations:

● Language of archival description (cont.)
  ○ If lower-level sections of finding aid are in another language or script, use @lang and @script attributes on the appropriate elements as necessary.
  ○ Example: bioghist provided in English and French:

```
<bioghist lang="eng" script="latn">
  <head>Biographical History</head>
  <p>George Washington was born...</p>
</bioghist>

<bioghist lang="fra" script="latn">
  <head>Biographical History</head>
  <p>George Washington est né...</p>
</bioghist>
```
Language elements

Recommendations:

● Language of archival materials
  ○ Encode at top level (in archdesc/did) using <langmaterial>
  ○ Note that <langmaterial> can no longer contain narrative text; if necessary, place in <descriptivenote>
  ○ Example:

  ```xml
  <langmaterial>
    <language langcode="eng">English</language>
    <language langcode="fra">French</language>
    <language langcode="deu">German</language>
    <descriptivenote>Bulk of collection is in English, with some clippings in French and German.</descriptivenote>
  </langmaterial>
  ```
Language elements

Recommendations:

- Language of archival materials (cont.)
  - Use languageset as appropriate
  - languageset pairs a language and a script
  - Example:

```xml
<langmaterial>
  <language langcode="eng">English</language>
  <languageset>
    <language langcode="jap">Japanese</language>
    <script scriptcode="Hira">Hiragana</script>
  </languageset>
  <descriptivenote>The bulk of the collection is English, but the "Reports" series contains substantial material in Japanese, in the Hiragana script.</descriptivenote>
</langmaterial>
```
Language elements

Recommendations:

- **Language of archival materials**
  - Encode at lower levels when the language differs from that given in the `<control>` section
  - Example:

```xml
<did>
  <unittitle>Reports</unittitle>
  <langmaterial>
    <languageset>
      <language langcode="jap">Japanese</language>
      <script scriptcode="Hira">Hiragana</script>
    </languageset>
  </langmaterial>
</did>
</c01>
```
Language elements

Challenges:

● For automated conversion, precision of existing EAD 2002 encoding will significantly impact EAD3 output
  ○ EAD 2002:
    
    `<langmaterial>The collection contains mostly English, with some clippings in French.</langmaterial>`
  ○ EAD3:
    
    `<langmaterial>
    <language> <!-- LANGUAGE NAME NEEDED --> </language>
    <descriptivenote>The collection contains mostly English, with some clippings in French.</descriptivenote>
    </langmaterial>`
Language elements

Challenges (cont):

- Encoding of language (@lang) and script (@script) at individual element level is a new capability; multiple versions of a finding aid may need to be merged.
- Addition of <languageset> may require manual correction.
- Some indexing and display systems may not be able to exploit language encoding at this level of granularity (e.g., they may not offer language as a facet).
Name part element

Available in persname, corpname, famname, etc.

User tasks:
- Name-based searching
- Contextual browsing
- Bidirectional conversion (MARC <--> EAD)
Name part element

Recommendations:

- Explore `<part>` encoding for *name elements if:
  - Your records need regular interchange with MARC
  - Your repository is heavily genealogy-based
  - Your repository has collections that are extensively interconnected by family
Name part element

Recommendations:

● Explore <part> encoding for *name elements if:
  ○ Your repository has collections that are extensively interconnected by department

```xml
<corpname>
  <part source="lcnaf" identifier="n79078780" localtype="corp">General Electric</part>
  <part source="lcnaf" identifier="n80005994" localtype="dept">Advanced Reactor Systems Department</part>
  <part source="lcnaf" identifier="n80005994" localtype="dept">Battery Business Department</part>
  <part source="lcnaf" identifier="n98093633" localtype="dept">Missile and Space Division</part>
</corpname>
```
Name part element

Recommendations:

- Include identifier and source attributes
Name part element

Challenges:
● Previous version of EAD had no <part> element within *name elements. Automated conversion to <part> encoding is therefore difficult or impossible
● Current discovery and display systems may not support part-based browsing or searching
Geolocation in archival data

• Let’s consider geocoding in archival description alongside linked data use practice

• Possible uses? User tasks identified:
  1. location-based searching
  2. faceting or visualizing archival record hits by location
  3. collection analysis (for researchers, archivists)

• Early implementations:
  EADitor: search querying of GeoNames, map rendering with OpenLayers
Geolocation metadata

- first, the context:
  HTML version of EAD3 Tag Library (4/2016):

  noting: attribute @identifier
  element <geogname>
  element <geogcoordinates>
  element <part>

Which one to use?
Geolocation metadata

• What can EAD3 do for places?

• EAD3 modifies the <did> (Descriptive ID) elements to require one or more <part> elements. Common attributes facilitate interoperability with external vocabularies:
  @identifier (the URI)
  @source (the vocabulary)
  @rules (formulation of terms)

• The <geogname> element remains, but gained some optional attributes and at least one **required** child element, <part>,
  followed by optional <geographiccoordinates> for lat, long, and altitude text

• <geographic coordinates> requires
  @coordinatesystem (what code is used? WGS84...)
Studying purpose & use of Geolocation elements

• Alongside such facets as date : <unitdatestructured>
  extent : <physdescstructured>
we explored EAD3 search by geographic location : <geographiccoordinates>

• Is there overlap with LD entries in <geogname> in (finding aid, EAD) records
  with unstructured geo-coordinates?

• Some evidence that <geogname> and its @identifier makes
  <geographiccoordinates> redundant.

• AND, why not keep using GeoNames ...
  it can be transformed into RDF
Geolocation practices

- Early implementation at EADitor
  http://eaditor.blogspot.com/2011/05/towards-georeferencing-archival.html
Geolocation practices

- Catalog visualization: GeoHOLLIS (Harvard)
- Faceting: USGS, Digital Commonwealth
Geolocation archival metadata recommendations

- EAD3 elements with DACS: (p.28 of report)
  integrated in a narrative description
  as access points (title, creator, admin/bio history, scope and content etc.)

- Acknowledged but grew apart from EAC-CPF’s <placeEntry>

- Recommendation:
  use <geogname> @identifier (geolocation metadata)
  do also use <geographiccoordinates> @coordinatesystem

- Migration or application considerations:
  No legacy data in EAD 2002!
  Have <geogname>? Add @identifier
  Don’t have <geogname>? Add <geographiccoordinates>
Linked Data attributes

EAD3 provides new elements and an expanded attributes set to provide support for linked data applications.

- @identifier/@relator: for referencing linked data URIs in access points
  - For use in persname, famname, corpname, name, subject, title, geogname, genreform, physfact, function, occupation, term

- <relations> element set to record semantic relationships
A very quick overview of linked data

Linked data/graph data is comprised of a series of triples, statements that establish a semantic relationship between two entities:

The Color Purple \(\rightarrow\) dc:creator \(\rightarrow\) Alice Walker, 1944-

The Color Purple \(\rightarrow\) dc:date \(\rightarrow\) 1982

Which can then be related to another resource:

Alice Walker, 1944 \(\rightarrow\) dc:creator \(\rightarrow\) Alice Walker papers
A very quick overview of linked data

And uses URIs whenever possible:

http://pid.emory.edu/ark:/25593/900jh = Alice Walker papers

http://purl.org/dc/elements/1.1/creator = dc:Creator

http://viaf.org/viaf/108495772 = Alice Walker, 1944-
...which enables semantic precision
...and allows us to share information

Network Graph of relationships in the Belfast Group participants based on data from finding aids.
Linked Data attributes: recommendations

For everyone:

Use URIs in access points using the @identifier and @relator attributes:

```xml
<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>
```
Linked Data attributes: recommendations

For advanced users wanting to encode linked data relationships within EAD:

Use the `<relations>` element set to record semantic relationships.
Linked Data attributes: recommendations

In RDF:

```rml
@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> .

```
Linked Data attributes: Best practices

- Use URIs (rather than identifier numbers) whenever possible.
- Define semantic relationships as specifically as possible.
- Do not duplicate information in <relations> that can be found in an EAC-CPF record.
- Consider local impacts and use cases.
I don’t have a triplestore. Is this for me?

<relations> and semantic attributes:

• Improve search and discovery
• Generate metadata that is human AND machine-readable
• Store linked data-ready metadata within existing XML structures

EAD archival metadata optimized for linked data!
Linked data attributes: Takeaways

• EAD remains an XML-based standard.

• @identifier and @relator are strongly recommended; a list of recommended vocabularies is available in the report

• <relations> and its accompanying tags are optional

• Use of <relations> will not automagically create linked data, but is a first step for those who want to implement linked data principles.
Questions?

Report available at
http://www2.archivists.org/sites/all/files/EAD3_Study_Group_on_Discovery_Recommendations_20160719.pdf

E-mail questions about the report may be directed to Cory Nimer (cory_nimer@byu.edu) or other members of the Study Group