



EAD3 Study Group on Discovery Report and Discussion

Society of American Archivists
Encoded Archival Standards Section

April 12, 2017

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

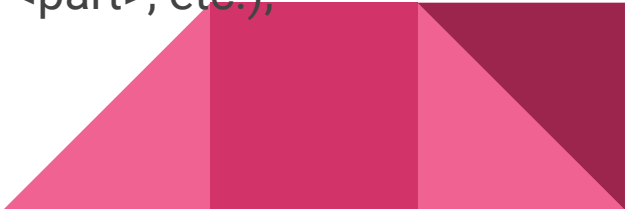


Changes in EAD3

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

THE NEW YORK PUBLIC LIBRARY
ARCHIVES & MANUSCRIPTS
PRINCETON UNIVERSITY

Princeton University Library Finding Aids Topics Names Collections Locations

Library Website | Create an Account | Site Fee

Search archivists Before or during optional date

45 Results

Only show results that contain online content

Show Advanced Search | Search Tips

Search Clear

Showing 1 - 10 of 45 Results for kw: archivists

Sort Date Created (ascending) Relevance Title (A to Z) Title (Z to A) Date Created (ascending) Date Created (descending) Creator (A to Z) Creator (Z to A)

1. A Card-Carrying ACLU Archivist

Contained in: American Civil Liberties Union Records: Subgroup 4 » Series 2007 » Publications dates not examined » Publications Box 10 dates not examined

Date range: dates not examined

Collection Creator: American Civil Liberties Union.

Extent: 1 folder

...1. A Card-Carrying ACLU Archivist ... : March 2015 Transfer. 1. A Card-Carrying ACLU Archivist...

Thomas Milton Halsey, 1903 1977, -

FILTER
1801 -
1851 -
1901 -
1951 -
2001 -

FILTER
DIVISION
Manuscripts Division
Scholarly Research Manuscripts
Rare Books
Music
Billy Rockwell
Rodger Archiver
Sound

Show collection titles only

Refine Search:

Repository
Manuscripts Division (22)
Princeton University Archives (12)
Public Policy Papers (10)

Subject
History (4)
New Jersey (3)
Princeton (3)

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

```
<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

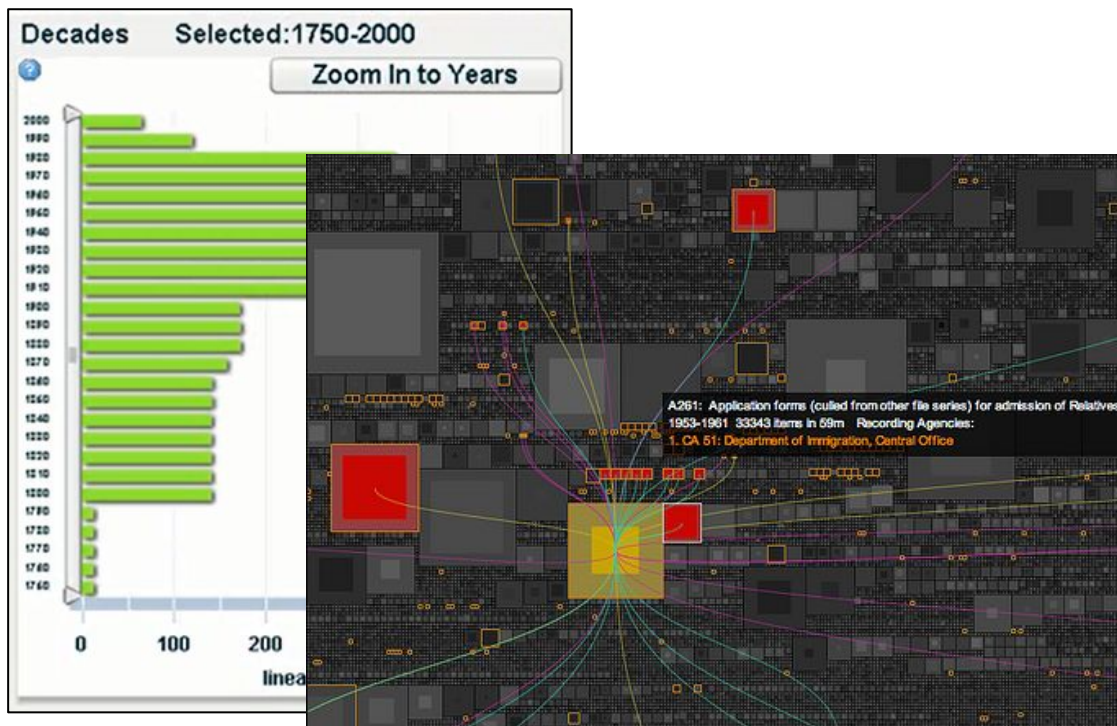


Image courtesy Mitchell Whitelaw, <http://visiblearchive.blogspot.com>

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

```
<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

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

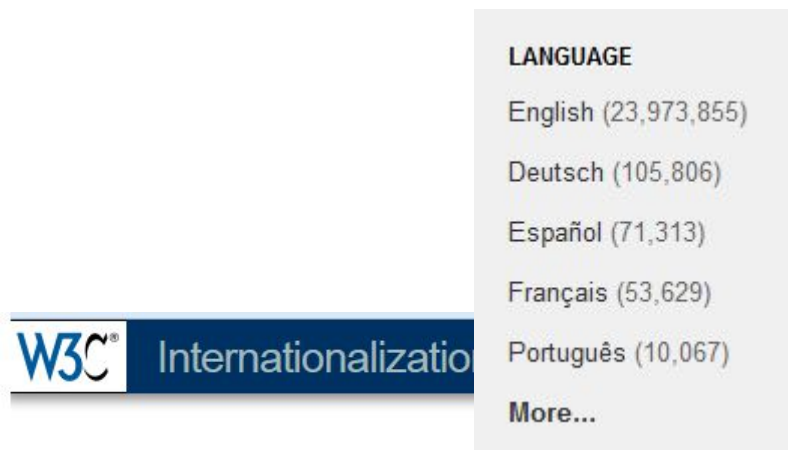
EAD3

```
<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](#)



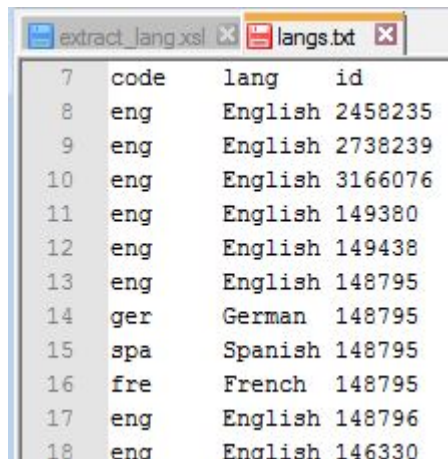
Declaring language in HTML

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



	code	lang	id
7	eng	English	2458235
8	eng	English	2738239
9	eng	English	3166076
10	eng	English	149380
11	eng	English	149438
12	eng	English	148795
13	ger	German	148795
14	spa	Spanish	148795
15	fre	French	148795
16	eng	English	148796
17	eng	English	146330
18	eng	English	146330

LANGUAGE

English (23,973,855)

Deutsch (105,806)

Español (71,313)

Français (53,629)

Português (10,067)

More...

Language elements

Recommendations:

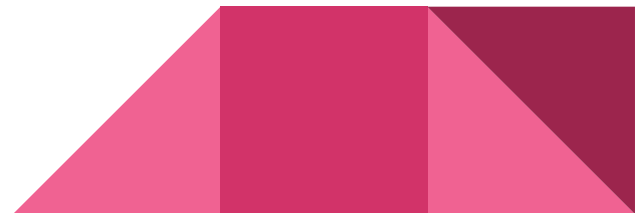
- Language of archival description

- profiledesc/language/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:

```
<languagedeclaration>
  <language langcode="eng">English</language>
  <script scriptcode="latn">Latin</script>
</languagedeclaration>
```

- Example: Finding aid in Russian, written using Cyrillic script:

```
<languagedeclaration>
  <language langcode="rus">Русская</language>
  <script scriptcode="cyrl">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:

```
<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:

```
<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:

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

Language elements

Challenges:

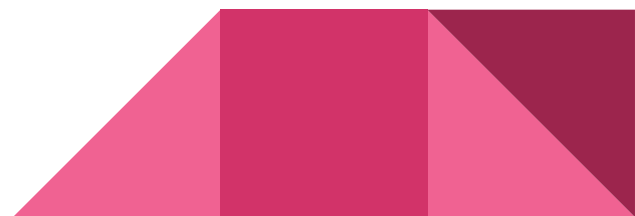
- For automated conversion, precision of existing EAD 2002 encoding will significantly impact EAD3 output

- 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)

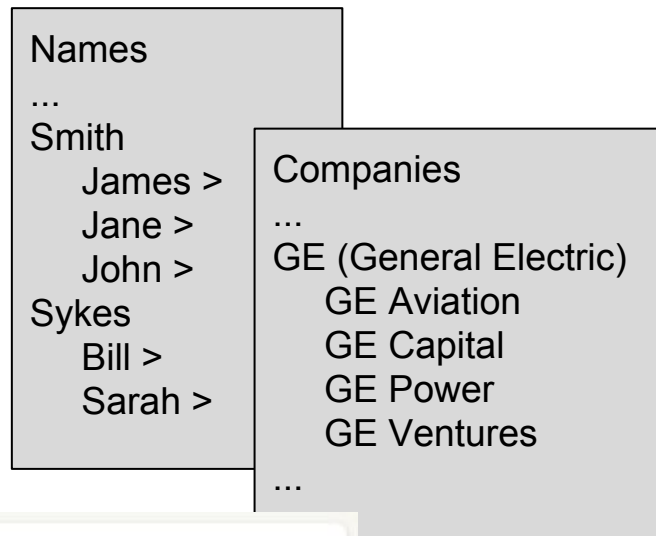


Name part element

Available in persname, corpname, famname, etc.

User tasks:

- Name-based searching
- Contextual browsing
- Bidirectional conversion (MARC <--> EAD)



[Show more options](#) ▼

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

```
<persname identifier="wizdumb001" source="wizlib" >
  <part localtype="sur">Dumbledore</part>
  <part localtype="fore">Albus Percival Wulfric Brian</part>
  <part type="daterange">1881-1997</part>
</persname>

<persname identifier="wizdumb002" source="wizlib" >
  <part localtype="sur">Dumbledore</part>
  <part localtype="fore">Aberforth</part>
  <part type="daterange">1884-</part>
</persname>
```

Names

```
...
Dumbledore
  Aberforth (1884-)
  Albus (1881-1997)
...
```

Name part element

Recommendations:

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

```
<corpname>  
  <part source="lcnaf" identifier="n79078780" localtype="corp">General Ele  
  <part source="lcnaf" identifier="n80005994" localtype="dept">Advanced Re  
</corpname>
```

```
<corpname>  
  <part source="lcnaf" identifier="n79078780" localtype="corp">General Electric</part>  
  <part source="lcnaf" identifier="no98093633" localtype="dept">Battery Business Department</part>  
</corpname>
```

```
<corpname>  
  <part source="lcnaf" identifier="n79078780" localtype="corp">General Electric</part>  
  <part source="lcnaf" identifier="n81047473" localtype="dept">Missile and Space Division</part>  
</corpname>
```

Names

...

General Electric

Advanced Reactor Systems Department

Battery Business Department

Missile and Space Division

Name part element

Recommendations:

- Include identifier and source attributes

```
<persname identifier="wizdumb001" source="wizlib" >  
  <part localtype="sur">Dumbledore</part>  
  <part localtype="fore">Albus Percival Wulfric Brian</part>  
  <part type="daterange">1881-1997</part>  
</persname>
```

```
<persname identifier="wizdumb001" source="wizlib" >  
  <part localtype="sur">Dumbledore</part>  
  <part localtype="fore">Albus Percival Wulfric Brian</part>  
  <part type="daterange">1881-1997</part>  
</persname>
```

```
<corpname>  
  <part source="lcnaf" identifier="n79078780" localtype="corp">General Electric</part>  
  <part source="lcnaf" identifier="n80005994" localtype="dept">Advanced Reactor Systems Department</part>  
</corpname>  
  
<corpname>  
  <part source="lcnaf" identifier="n79078780" localtype="corp">General Electric</part>  
  <part source="lcnaf" identifier="no98093633" localtype="dept">Battery Business Department</part>  
</corpname>  
  
<corpname>  
  <part source="lcnaf" identifier="n79078780" localtype="corp">General Electric</part>  
  <part source="lcnaf" identifier="n81047473" localtype="dept">Missile and Space Division</part>  
</corpname>
```

Name part element

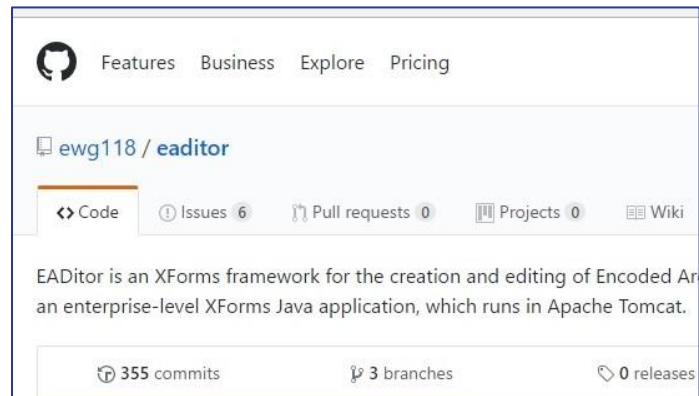
Challenges:

- 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

- 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



The GeoNames geographical database covers all countries and contains over eleven million placenames that are available for download free of charge.

Geolocation practices

Geographical Name ✕

Interface Type Geonames Local Vocabulary

Search

Location

- Charlottesville: City of Charlottesville, Virginia, United States
- City of Charlottesville: City of Charlottesville, Virginia, United States
- Charlottesville-Albemarle Airport: Albemarle County, Virginia, United States
- Charlottesville and Albemarle County Historic District: City of Charlottesville, Virginia, United States
- Wakefield: City of Charlottesville, Virginia, United States
- Albemarle Historic District: City of Charlottesville, Virginia, United States
- Azalea Park: City of Charlottesville, Virginia, United States
- Belmont Park: City of Charlottesville, Virginia, United States
- East Lawn: City of Charlottesville, Virginia, United States
- Forrest Hills Park: City of Charlottesville, Virginia, United States
- Frys Spring Park: City of Charlottesville, Virginia, United States
- George Rogers Clark Statue: City of Charlottesville, Virginia, United States
- Greenbriar Park: City of Charlottesville, Virginia, United States
- Greenleaf Park: City of Charlottesville, Virginia, United States
- Jackson Park: City of Charlottesville, Virginia, United States
- Jordan Avenue Park: City of Charlottesville, Virginia, United States
- Lamngan Field: City of Charlottesville, Virginia, United States
- Lee Park: City of Charlottesville, Virginia, United States
- Lewis and Clark Memorial: City of Charlottesville, Virginia, United States
- Madison Bowl: City of Charlottesville, Virginia, United States

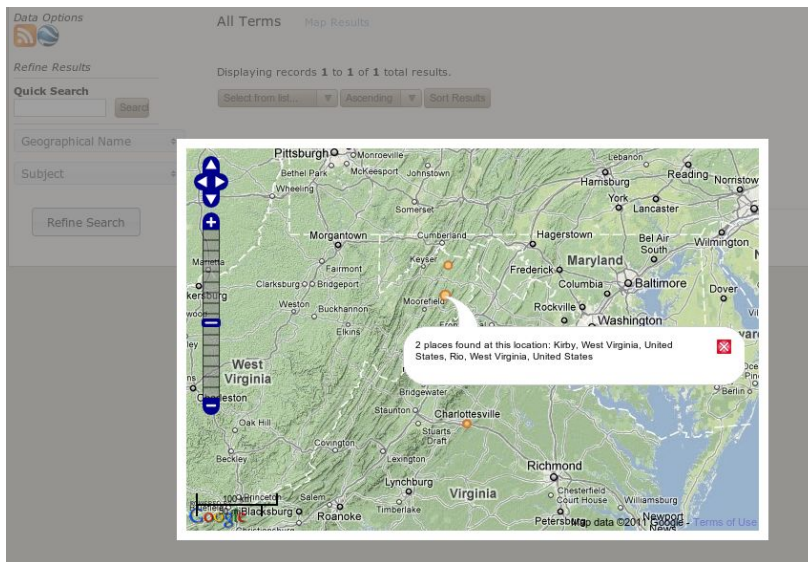
Geographical Name ✕

Interface Type Geonames Local Vocabulary

Term

Subject

- Goochland County (Va.) -- History -- 18th century ✕
- Goochland County (Va.) -- History -- 19th century



- Early implementation at EADitor

<http://eaditor.blogspot.com/2011/05/towards-georeferencing-archival.html>

Geolocation practices

Enhance Catalog Searching with Geospatial Technology | Library Lab

HOLLIS

Return to result list

Title: World's fair : Jamaica at Chicago : an account descriptive of the colony of Jamaica, with historical and other appendices / compiled under the direction of C.J. Ward, ~~Harvard~~ Commissioner for Jamaica.

Published: New York : W.J. Peil, printer, 1893.

Description: 26 cm.

Other title: Jamaica at Chicago

Cover title: Jamaica at the Columbian Exposition, 1893

Subject: World's Columbian Exposition (1893 : Chicago, Ill.)

Authors: Jamaica
Name: Charles Jamaica; non-person
Works: Jamaica. Commissioner for Jamaica.
Harvard Ernst Mayr Library preservation microfilm project : 00786.

Series titles: HOLLIS number: 005171500 MARC HOLLIS Classic

0:55 / 2:58

Home Login Register Feedback Help

Search Criteria Data Sets Additional Criteria Results

1. Enter Search Criteria

To narrow your search area, type in an address or place name, enter coordinates or click the map to define your search area (for advanced map tools, view the [help documentation](#)), and/or choose a date range.

Address/Place Path/Row Feature Circle

Show Clear

Coordinates Predefined Area Shapefile KML

Degree/Minute/Second Decimal

No coordinates selected.

Use Map Add Coordinate Clear Coordinates

Date Range Result Options

Search from: mmi/dd/yyyy to: mmi/dd/yyyy

Search months: (all)

Data Sets Additional Criteria Results

Search Criteria Summary (Show) Clear Criteria

Map Satellite (64° 28' 22" N, 029° 31' 52" W) Options Overlays

- 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 → dc:creator → Alice Walker, 1944-

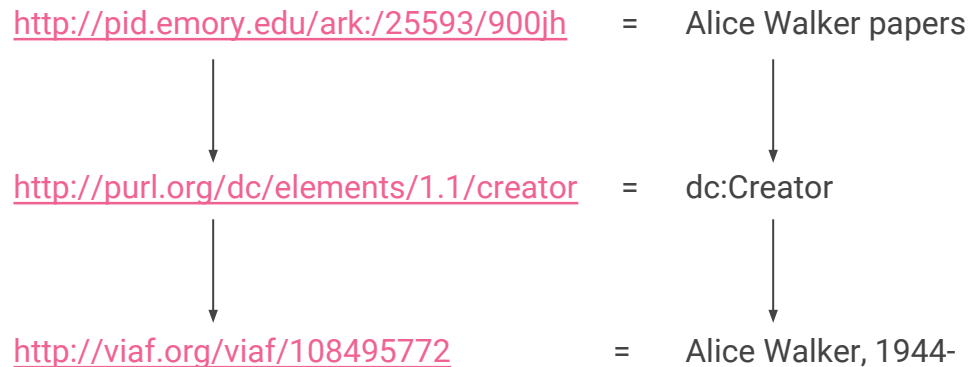
The Color Purple → dc:date → 1982

Which can then be related to another resource:

Alice Walker, 1944 → dc:creator → Alice Walker papers

A very quick overview of linked data

And uses URIs whenever possible:



...which enables semantic precision



<name>Chips</name>



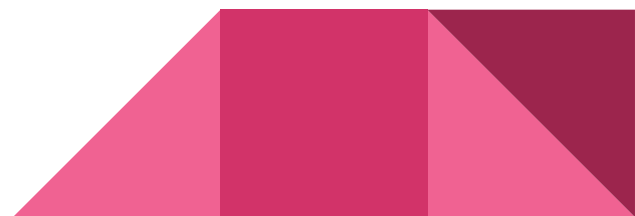
<name>Chips</name>



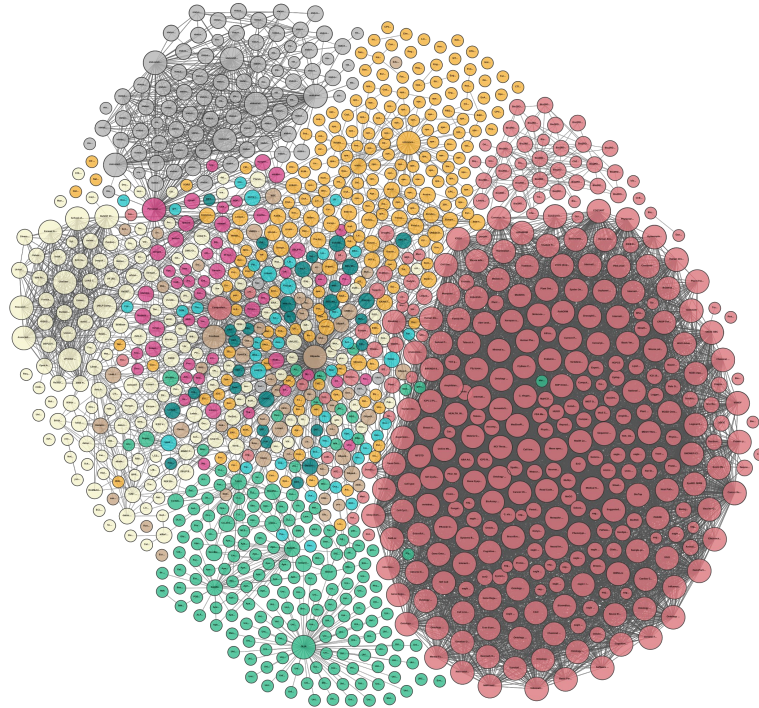
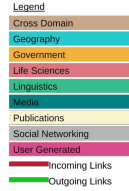
<name>French fries</name>

<https://www.wikidata.org/wiki/Q152088>

<https://www.wikidata.org/wiki/Q173265>

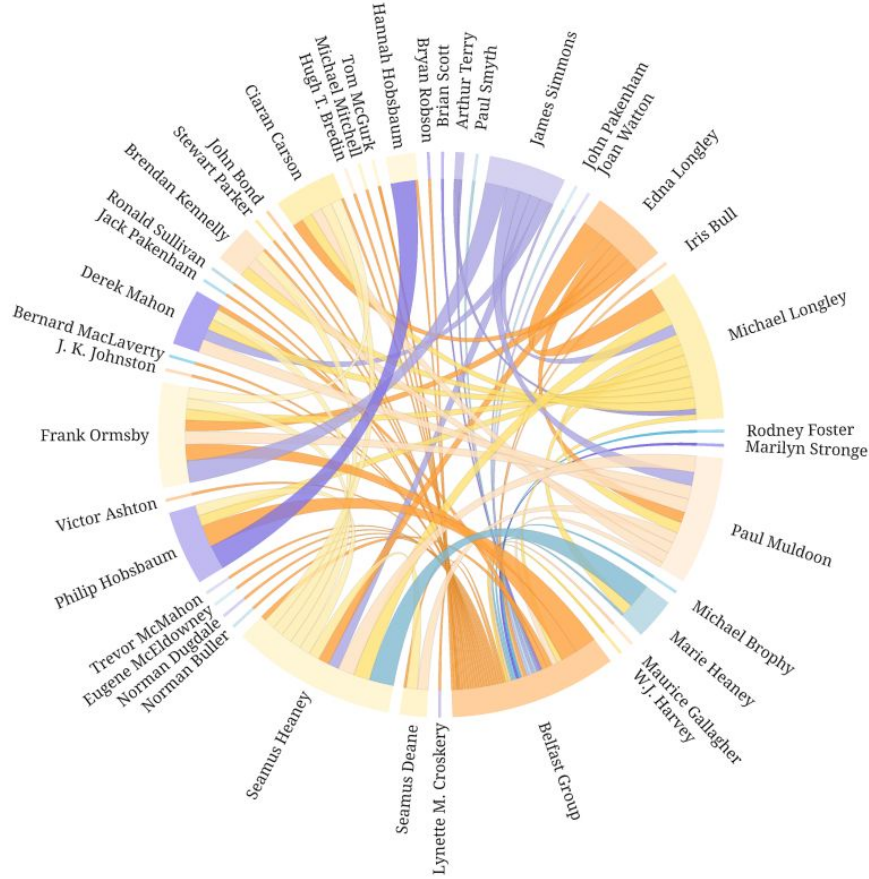


...and allows us to share information



"Linking Open Data cloud diagram 2017, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak.
<http://lod-cloud.net/>"

...in order to do interesting things



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:

```
<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.

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

Linked Data attributes: recommendations

In RDF:

@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>> .

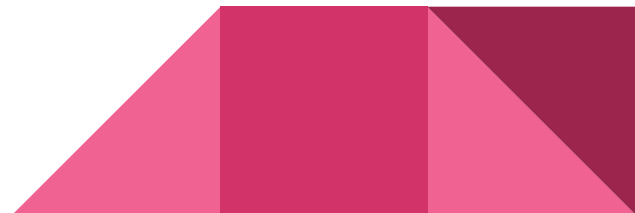
@prefix xsd: <<http://www.w3.org/2001/XMLSchema#>> .

<<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 .



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.
 - @identifer 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

