Observing changes in EAD Tag Usage to Support Discovery, 2013-2021

Merrilee Proffitt
Senior Manager
OCLC Research Library Partnership

Bruce Washburn
Principal Architect
OCLC Research
• Partnership with California Digital Library, OCLC, University of Virginia Library, and Chain Bridge Group

• Explore and create the blueprint for implementing a national finding aid network that is community-driven, -sustained, and –governed

• Find out more at https://confluence.ucop.edu/display/NAFAN/Building+a+National+Finding+Aid+Network
NAFAN Activity Streams

- Research
- Technical assessment and prototyping
- Community engagement
- Sustainability and governance planning
OCLC’s Primary Role: Research

- Research
- Technical assessment and prototyping
- Community engagement
- Sustainability and governance planning
Qualitative: Gain an updated understanding of the needs of both researchers and cultural heritage institutions regarding finding aid aggregation.
Research Summary

**Qualitative**: Gain an updated understanding of the needs of both researchers and cultural heritage institutions regarding finding aid aggregation.

**Quantitative**: Evaluate an aggregation of finding aids in order to:

- scope the network's functionality to the finding aid data
- lay the groundwork for data remediation and expanded network features.
Research Summary

**Qualitative:** Gain an updated understanding of the needs of both researchers and cultural heritage institutions regarding finding aid aggregation.

**Quantitative:** Evaluate an aggregation of finding aids in order to:
- scope the network's functionality to the finding aid data
- lay the groundwork for data remediation and expanded network features.
145,671 Aggregated Finding Aids
Dimensions for analysis

**Search**: all discovery systems have a keyword search function; many also include the ability to search by a particular field or element.

**Browse**: many discovery systems include the ability to browse finding aids by title, subject, dates, or other facets.

**Results display**: once a user has done a search, the results display will return portions of the finding aid to help with further evaluation.

**Sort**: once a user has done a search, they may have the option to reorder the results.

**Facet**: once a user has done a search, they may have the option to narrow the results to only include results that fall within certain facets.
EAD elements that impact discovery

- Dates
  - unitdate

- Extent data
  - extent

- Collection title sources
  - unittitle, titleproper/@type=filing

- Content tags in dsc
  - corpname, famname, function, genreform, geogname, name, occupation, persname, subject

- Content tags in origination
  - corpname, famname, name, persname

- Content tags in controlaccess
  - corpname, famname, function, geogname, name, occupation, persname, subject

- Material type
  - controlaccess/genreform

- Repository
  - repository

- Notes
  - abstract, bioghist, scopecontent

Derived from "Thresholds for Discovery: EAD Tag Analysis in ArchiveGrid, and Implications for Discovery Systems" published in code4lib Journal issue 22.
## Defining Thresholds for Discovery

<table>
<thead>
<tr>
<th>Usage</th>
<th>Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-50%</td>
<td>Low</td>
</tr>
<tr>
<td>51-80%</td>
<td>Medium</td>
</tr>
<tr>
<td>81%-95%</td>
<td>High</td>
</tr>
<tr>
<td>96%-100%</td>
<td>Complete</td>
</tr>
<tr>
<td>Discovery Category</td>
<td>EAD Element</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Dates</td>
<td>unitdate</td>
</tr>
<tr>
<td>Extent data</td>
<td>extent</td>
</tr>
<tr>
<td>Collection Title sources</td>
<td>unittitle</td>
</tr>
<tr>
<td></td>
<td>titleproper&amp;type=filing</td>
</tr>
<tr>
<td>Content tags in dsc</td>
<td>corpname</td>
</tr>
<tr>
<td></td>
<td>famname</td>
</tr>
<tr>
<td></td>
<td>function</td>
</tr>
<tr>
<td></td>
<td>genreform</td>
</tr>
<tr>
<td></td>
<td>geogname</td>
</tr>
<tr>
<td></td>
<td>name</td>
</tr>
<tr>
<td></td>
<td>occupation</td>
</tr>
<tr>
<td></td>
<td>persname</td>
</tr>
<tr>
<td></td>
<td>subject</td>
</tr>
<tr>
<td>origination and content tags</td>
<td>origination</td>
</tr>
<tr>
<td></td>
<td>corpname</td>
</tr>
<tr>
<td></td>
<td>famname</td>
</tr>
<tr>
<td></td>
<td>name</td>
</tr>
<tr>
<td></td>
<td>persname</td>
</tr>
<tr>
<td>archdesc/controlaccess and content tags</td>
<td>controlaccess</td>
</tr>
<tr>
<td></td>
<td>corpname</td>
</tr>
<tr>
<td></td>
<td>famname</td>
</tr>
<tr>
<td></td>
<td>function</td>
</tr>
<tr>
<td></td>
<td>geogname</td>
</tr>
<tr>
<td></td>
<td>name</td>
</tr>
<tr>
<td></td>
<td>occupation</td>
</tr>
<tr>
<td></td>
<td>persname</td>
</tr>
<tr>
<td></td>
<td>subject</td>
</tr>
<tr>
<td>Material type</td>
<td>genreform</td>
</tr>
<tr>
<td>Repository</td>
<td>repository</td>
</tr>
<tr>
<td>Notes</td>
<td>scopecontent</td>
</tr>
<tr>
<td></td>
<td>abstract</td>
</tr>
<tr>
<td></td>
<td>bioghist</td>
</tr>
</tbody>
</table>
“If the archival community continues on its current path, then the potential of the EAD format to support researchers or the public in discovery of material will remain underutilized. Minimally, collection descriptions that are below the thresholds for discovery will hinder their discovery efforts and maximally will remain hidden from view.

Perhaps with emerging evidence about the corpus of EAD, continued discussion of practice, recognition of a need for greater functionality, and shared tools both to create new EAD documents and improve existing encoding, we can look forward to further increasing the effectiveness and efficiency of EAD encoding and develop a practice of EAD encoding that pushes collection descriptions across the threshold of discovery.”

https://journal.code4lib.org/articles/8956
More Research Opportunities

- Finding aid links to digital content
- Specifying physical characteristics and genre
- Content element value connections to controlled vocabularies
- Consistency of institutional contact information
- Informing researchers about access to, use of, and reuse of materials
More Research Opportunities

🔗 Finding aid links to digital content

🔍 Specifying physical characteristics and genre

🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗🔗˥
Finding aid links to digital content

- How many links are there, per finding aid?
- Which EAD elements and attributes include links?
- What types of digital objects are linked?
- Are links a complete URL or are they relative?
- How many links resolve?
Topics Covered:

• The NAFAN “Building a National Finding Aid Network project”
• NAFAN activity streams and OCLC’s role
• Qualitative and Quantitative research objectives
• The NAFAN EAD aggregation
• A Discovery-oriented EAD tag analysis, past and present
• More research opportunities

Contact Us

Merrilee Proffitt
Senior Manager
OCLC Research Library Partnership
merrilee_proffitt@oclc.org

Bruce Washburn
Principal Software Engineer
OCLC Research
bruce_washburn@oclc.org