Automated Enhancement of Legacy Metadata in the CONTENTdm Project Client

Andrew Weidner, Metadata Librarian
UNIVERSITY of HOUSTON | LIBRARIES

Upper Midwest Digital Collections Conference 2014
University of St. Thomas, Minneapolis, MN
UHDL Background

- Est. 2009
- CONTENTdm 6
- 60 collections
- 55,000+ images
- Metadata Upgrade started in 2013
Rationales for Metadata Upgrade

• Improve access of digital collection data in Primo
• Make our metadata interoperable with other systems
• Enhance discoverability of digital objects
• Improve display and retrieval of our digital metadata through our upcoming new UHDL interface
• Prepare for linked data environment

More Robust, Reliable and Retrievable Metadata!
Upgrade Strategy

- **Phase One:** Standardize the digital library's metadata schema and create metadata dictionary
Upgrade Strategy

- **Phase One**: Standardize the digital library’s metadata schema and create metadata dictionary (relatively easy)
Upgrade Strategy

- **Phase One**: Standardize the digital library’s metadata schema and create metadata dictionary (relatively easy)

- **Phase Two**: Standardize collection names and update metadata fields to conform to schema
Upgrade Strategy

- **Phase One:** Standardize the digital library’s metadata schema and create metadata dictionary (relatively easy)

- **Phase Two:** Standardize collection names and update metadata fields to conform to schema (relatively easy)
Upgrade Strategy

- **Phase One**: Standardize the digital library’s metadata schema and create metadata dictionary (relatively easy)

- **Phase Two**: Standardize collection names and update metadata fields to conform to schema (relatively easy)

- **Phase Three**: Standardize item level metadata
Upgrade Strategy

• **Phase One:** Standardize the digital library’s metadata schema and create metadata dictionary (relatively easy)

• **Phase Two:** Standardize collection names and update metadata fields to conform to schema (relatively easy)

• **Phase Three:** Standardize item level metadata (you want us to do what?!?!)
Phase Three Challenges

• Limitations of existing interfaces
  o No easy way to automate or batch edit legacy metadata in Project Client or CDM Administration interfaces (Catcher is a possibility)
  o Must edit item-by-item, which is time consuming and resource intensive
Phase Three Challenges

• Limitations of existing interfaces
  o No easy way to automate or batch edit legacy metadata in Project Client or CDM Administration interfaces (Catcher is a possibility)
  o Must edit item-by-item, which is time consuming and resource intensive

• Limitations of tab-delimited process
  o Exporting metadata, editing tab delimited files, and re-uploading content can strain server and software
  o OpenRefine can tell us where the problems are, but the work still has to be done in the Project Client or Web Admin
Phase Three Challenges

• **Limitations of existing interfaces**
  - No easy way to automate or batch edit legacy metadata in Project Client or CDM Administration interfaces (Catcher is a possibility)
  - Must edit item-by-item, which is time consuming and resource intensive

• **Limitations of tab-delimited process**
  - Exporting metadata, editing tab delimited files, and re-uploading content can strain server and software
  - OpenRefine can tell us where the problems are, but the work still has to be done in the Project Client or Web Admin

• **Limitations on staff time**
Phase Three Solutions

• Decided to work from project client
Phase Three Solutions

- Decided to work from project client
- Develop tools layered on top of the project client to automate key areas of the upgrade process
Phase Three Solutions

• Decided to work from project client

• Develop tools layered on top of the project client to automate key areas of the upgrade process
Upgrade Tools: Subject App

- Automates mapping from various vocabularies to LCSH
- Facilitates harvesting of URIs from LC linked data service (id.loc.gov)
Upgrade Tools: Subject App
Upgrade Tools: Subject App
<table>
<thead>
<tr>
<th>Thumbnail</th>
<th>Donor</th>
<th>Subject Topical (LCSH)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Melott, Scott;</td>
<td>Fire fighters; Funeral rites and ceremonies</td>
</tr>
<tr>
<td></td>
<td>Melott, Scott;</td>
<td>Firefighters; Funeral ceremonies; Tropical</td>
</tr>
<tr>
<td></td>
<td>Melott, Scott;</td>
<td>Floods (natural events); Firefighters; Fire engines; Fire stations;</td>
</tr>
</tbody>
</table>
Search Results

Do you need assistance with your search?

**Please Note: LC Classification entries are not included in general search results. You must explicitly select LC Classification in order to search the scheme. This is temporary while the impact of adding LCC to the current system is better understood.**
**Search Results**

Do you need assistance with your search?

**Search**

floods

**LC Subject Headings**

- LC Subject Headings
- LC Name Authority File
- LC Classification
- LC Children's Subject Headings

**Please Note:** LC Classification entries are not included in general search results. You must explicitly select LC Classification in order to search the scheme. This is temporary while the impact of adding LCC to the current system is better understood.

[Search]  [Reset]

---

<table>
<thead>
<tr>
<th>Label</th>
<th>Vocabulary</th>
<th>Concept Type</th>
<th>Subdivision Type</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Floods</td>
<td>LC Subject Headings</td>
<td>Topic</td>
<td></td>
<td>sh85049168</td>
</tr>
<tr>
<td>Floodings; Inundations</td>
<td>LC Subject Headings</td>
<td>Topic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Floods in literature</td>
<td>LC Subject Headings</td>
<td>Topic</td>
<td></td>
<td>sh2004004831</td>
</tr>
</tbody>
</table>
Floods

From Library of Congress Subject Headings

Details ▶ Visualization ▶ Suggest Terminology

Floods

URI(s)
- http://id.loc.gov/authorities/subjects/sh85049168
- info:lc/authorities/sh85049168
- http://id.loc.gov/authorities/sh85049168#concept

Instance Of
- MADS/RDF Topic
- MADS/RDF Authority
- SKOS Concept

Scheme Membership(s)
- Library of Congress Subject Headings

Collection Membership(s)
- LCSH Collection - Authorized Headings
- LCSH Collection - General Collection
Floods

From Library of Congress Subject Headings

Details Visualization

Floods

URI(s)
- http://id.loc.gov/authorities/subjects/sh85049168
- info:lc/authorities/sh85049168
- http://id.loc.gov/authorities/sh85049168#concept

Instance Of
- MADS/RDF Topic
- MADS/RDF Authority
- SKOS Concept

Scheme Membership(s)
- Library of Congress Subject Headings

Collection Membership(s)
- LCSH Collection - Authorized Headings
- LCSH Collection - General Collection

Authorized Term

floods (natural events)
Enter an authorized term:
Floods

OK Cancel
Floods

From Library of Congress Subject Headings

Details Visualization Suggest Terminology

Floods

URI(s)
- http://id.loc.gov/authorities/subjects/sh85049168
- info:lc/authorities/sh85049168
- http://id.loc.gov/authorities/subjects/sh85049168#concept

Instance Of
- MADS/RDF Topic
- MADS/RDF Authority
- SKOS Concept

Scheme Membership(s)
- Library of Congress Subject Headings

Collection Membership(s)
- LCSH Collection - Authorized Headings
- LCSH Collection - General Collection

Authorized Term URI

Floods

Enter the URI:

http://id.loc.gov/authorities/subjects/sh85049168

OK Cancel
<table>
<thead>
<tr>
<th>Thumbnail</th>
<th>Donor</th>
<th>Subject Topical (LCSH)</th>
<th>Subject Topical (TGM-1)</th>
<th>Subject Topical (AAT)</th>
<th>Subject Topical (SAA)</th>
<th>Subject Topical (Local)</th>
<th>Subject Topical (LCNAF)</th>
<th>Subject Topical (HOT)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Melott, Scott;</td>
<td>Fire fighters; Funeral rites and ceremonies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Melott, Scott;</td>
<td>Floods: Fire fighters; Fire engines</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Melott, Scott;</td>
<td>Firefighters; Fire; Engine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Melott, Scott;</td>
<td>Firefighters; Hose Carriages</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Melott, Scott;</td>
<td>Firefighters; Fire Stations; Fire; Engine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Upgrade Tools: Subject App
Upgrade Tools: Subject App
<table>
<thead>
<tr>
<th>Subject Topical</th>
<th>URL</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGM Schools</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85118451">http://id.loc.gov/authorities/subjects/sh85118451</a></td>
<td>20140513</td>
</tr>
<tr>
<td>TGM Cotton</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85033302">http://id.loc.gov/authorities/subjects/sh85033302</a></td>
<td>20140513</td>
</tr>
<tr>
<td>TGM Government facilities</td>
<td>Government buildings</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85108620">http://id.loc.gov/authorities/subjects/sh85108620</a></td>
</tr>
<tr>
<td>AAT buildings</td>
<td>Buildings</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85017769">http://id.loc.gov/authorities/subjects/sh85017769</a></td>
</tr>
<tr>
<td>AAT farming</td>
<td>Agricultural systems</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85002407">http://id.loc.gov/authorities/subjects/sh85002407</a></td>
</tr>
<tr>
<td>AAT hospitals</td>
<td>Hospitals</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85062285">http://id.loc.gov/authorities/subjects/sh85062285</a></td>
</tr>
<tr>
<td>TGM Jails</td>
<td>Jails</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85069253">http://id.loc.gov/authorities/subjects/sh85069253</a></td>
</tr>
<tr>
<td>AAT banks</td>
<td>Banks and banking, Central</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85011647">http://id.loc.gov/authorities/subjects/sh85011647</a></td>
</tr>
<tr>
<td>AAT City blocks</td>
<td>City blocks</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh8504924">http://id.loc.gov/authorities/subjects/sh8504924</a></td>
</tr>
<tr>
<td>TGM Waterworks</td>
<td>Waterworks</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85145757">http://id.loc.gov/authorities/subjects/sh85145757</a></td>
</tr>
<tr>
<td>TGM Buildings</td>
<td>Buildings</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85017769">http://id.loc.gov/authorities/subjects/sh85017769</a></td>
</tr>
<tr>
<td>TGM Portrait photographs</td>
<td>Portraits</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85105182">http://id.loc.gov/authorities/subjects/sh85105182</a></td>
</tr>
<tr>
<td>TGM Shipping</td>
<td>Shipping</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85121572">http://id.loc.gov/authorities/subjects/sh85121572</a></td>
</tr>
<tr>
<td>TGM Advertising</td>
<td>Advertising</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85001086">http://id.loc.gov/authorities/subjects/sh85001086</a></td>
</tr>
<tr>
<td>TGM Business &amp; finance</td>
<td>Business</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85010260">http://id.loc.gov/authorities/subjects/sh85010260</a></td>
</tr>
<tr>
<td>AAT buildings (structures)</td>
<td>Buildings</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85017769">http://id.loc.gov/authorities/subjects/sh85017769</a></td>
</tr>
<tr>
<td>AAT skyscrapers</td>
<td>Skyscrapers</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85123270">http://id.loc.gov/authorities/subjects/sh85123270</a></td>
</tr>
<tr>
<td>TGM Railstation stations</td>
<td>Railroad stations</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85111042">http://id.loc.gov/authorities/subjects/sh85111042</a></td>
</tr>
<tr>
<td>TGM Industry</td>
<td>Industry &amp; trade summary</td>
<td><a href="http://id.loc.gov/authorities/subjects/n92090429">http://id.loc.gov/authorities/subjects/n92090429</a></td>
</tr>
<tr>
<td>AAT factories</td>
<td>Factories</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85046233">http://id.loc.gov/authorities/subjects/sh85046233</a></td>
</tr>
<tr>
<td>TGM Advertisements</td>
<td>Advertising</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85001086">http://id.loc.gov/authorities/subjects/sh85001086</a></td>
</tr>
<tr>
<td>TGM Construction</td>
<td>Construction</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85005337">http://id.loc.gov/authorities/subjects/sh85005337</a></td>
</tr>
<tr>
<td>AAT architectural drawings</td>
<td>Designs and plans</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh87005580">http://id.loc.gov/authorities/subjects/sh87005580</a></td>
</tr>
<tr>
<td>AAT banks (buildings)</td>
<td>Bank buildings</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85011556">http://id.loc.gov/authorities/subjects/sh85011556</a></td>
</tr>
<tr>
<td>TGM stretcher</td>
<td>Stretchers</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85077674">http://id.loc.gov/authorities/subjects/sh85077674</a></td>
</tr>
<tr>
<td>AAT floods (natural events)</td>
<td>Floods</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85049168">http://id.loc.gov/authorities/subjects/sh85049168</a></td>
</tr>
<tr>
<td>Subject</td>
<td>Authority URL</td>
<td>Date</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>TGM Schools Schools</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85118451">http://id.loc.gov/authorities/subjects/sh85118451</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>TGM Cotton Cotton</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85033302">http://id.loc.gov/authorities/subjects/sh85033302</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>TGM Government facilities Government buildings</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85108620">http://id.loc.gov/authorities/subjects/sh85108620</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>AAT buildings Buildings</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85017769">http://id.loc.gov/authorities/subjects/sh85017769</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>AAT farming Agricultural systems</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85002407">http://id.loc.gov/authorities/subjects/sh85002407</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>AAT hospitals Hospitals</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85062285">http://id.loc.gov/authorities/subjects/sh85062285</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>TGM Jails Jails</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85069253">http://id.loc.gov/authorities/subjects/sh85069253</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>AAT banks Banks and banking, Central</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85011617">http://id.loc.gov/authorities/subjects/sh85011617</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>AAT City blocks City blocks</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85014924">http://id.loc.gov/authorities/subjects/sh85014924</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>TGM Waterworks Waterworks</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85145757">http://id.loc.gov/authorities/subjects/sh85145757</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>TGM Buildings Buildings</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85017769">http://id.loc.gov/authorities/subjects/sh85017769</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>TGM Portrait photographs Portraits</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85105182">http://id.loc.gov/authorities/subjects/sh85105182</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>TGM Shipping Shipping</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85121579">http://id.loc.gov/authorities/subjects/sh85121579</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>TGM Advertising Advertising</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85001086">http://id.loc.gov/authorities/subjects/sh85001086</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>TGM Business &amp; finance Business</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85018260">http://id.loc.gov/authorities/subjects/sh85018260</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>AAT buildings (structures) Buildings</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85017769">http://id.loc.gov/authorities/subjects/sh85017769</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>AAT skyscrapers Skyscrapers</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85123270">http://id.loc.gov/authorities/subjects/sh85123270</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>TGM skyscrapers Skyscrapers</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85123270">http://id.loc.gov/authorities/subjects/sh85123270</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>TGM Banks Banks and banking, Central</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85011617">http://id.loc.gov/authorities/subjects/sh85011617</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>TGM Railroad stations Railroad stations</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85111042">http://id.loc.gov/authorities/subjects/sh85111042</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>TGM Industry Industry &amp; trade summary</td>
<td><a href="http://id.loc.gov/authorities/subjects/nm92090429">http://id.loc.gov/authorities/subjects/nm92090429</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>AAT factories Factories</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85046623">http://id.loc.gov/authorities/subjects/sh85046623</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>TGM Advertisements Advertising</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85001086">http://id.loc.gov/authorities/subjects/sh85001086</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>TGM Construction Construction</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85005337">http://id.loc.gov/authorities/subjects/sh85005337</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>AAT architectural drawings Designs and plans</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh87005850">http://id.loc.gov/authorities/subjects/sh87005850</a></td>
<td>CPH 20140513</td>
</tr>
<tr>
<td>AAT banks (buildings) Bank buildings</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85011556">http://id.loc.gov/authorities/subjects/sh85011556</a></td>
<td>CPH 20140513</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subject</th>
<th>Authority URL</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAT floods (natural events) Floods</td>
<td><a href="http://id.loc.gov/authorities/subjects/sh85049168">http://id.loc.gov/authorities/subjects/sh85049168</a></td>
<td>AJW 20140513</td>
</tr>
</tbody>
</table>
Upgrade Tools: Subject App
Upgrade Tools: Name App

- Large collections with legacy metadata
- All names entered in LC NAF field and formatted to look like an LC NAF name
- App moves names to correct authority
<table>
<thead>
<tr>
<th>LCNAF</th>
<th>Theis, Susanne Theis, Susanne</th>
<th>AJW 20140424</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCNAF</td>
<td>Johnson, Dianna A. Johnson, Dianna A.</td>
<td>AJW 20140424</td>
</tr>
<tr>
<td>LCNAF</td>
<td>Young, Anne Young, Anne, RN</td>
<td>AJW 20140424</td>
</tr>
<tr>
<td>LCNAF</td>
<td>Sidhu, Bansi Sidhu, Bansi</td>
<td>AJW 20140424</td>
</tr>
<tr>
<td>LCNAF</td>
<td>Storrs, Landon R. Y Storrs, Landon R. Y.</td>
<td>AJW 20140424</td>
</tr>
<tr>
<td>LCNAF</td>
<td>Storrs, Landon R. Y. Storrs, Landon R. Y.</td>
<td>AJW 20140424</td>
</tr>
<tr>
<td>LCNAF</td>
<td>Sondock, Ruby K. Sondock, Ruby K.</td>
<td>AJW 20140424</td>
</tr>
<tr>
<td>LCNAF</td>
<td>Storey, Gail Donohue Storey, Gail Donohue</td>
<td>AJW 20140424</td>
</tr>
<tr>
<td>LCNAF</td>
<td>Walker, Margie Walker, Margie</td>
<td>AJW 20140424</td>
</tr>
<tr>
<td>LCNAF</td>
<td>Koen, Karleen Koen, Karleen</td>
<td>AJW 20140424</td>
</tr>
<tr>
<td>HOT</td>
<td>Allen, John M., (?-1847) Allen, John M. (?-1847)</td>
<td>AJW 20140424</td>
</tr>
<tr>
<td>HOT</td>
<td>Allen, John M. (?-1847) Allen, John M. (?-1847)</td>
<td>AJW 20140424</td>
</tr>
<tr>
<td>HOT</td>
<td>Alley, Abraham, 1803-1862 Alley, Abraham (1803-1862)</td>
<td>AJW 20140424</td>
</tr>
<tr>
<td>HOT</td>
<td>Alley, Abraham (1803-1862) Alley, Abraham (1803-1862)</td>
<td>AJW 20140424</td>
</tr>
<tr>
<td>HOT</td>
<td>Atchison, Daniel D., 1820-1867 Atchison, Daniel D. (1820-1867)</td>
<td>AJW 20140424</td>
</tr>
<tr>
<td>HOT</td>
<td>Atchison, Daniel D. (1820-1867) Atchison, Daniel D. (1820-1867)</td>
<td>AJW 20140424</td>
</tr>
<tr>
<td>HOT</td>
<td>Ayres, David, 1793-1881 Ayres, David (1793-1881)</td>
<td>AJW 20140424</td>
</tr>
<tr>
<td>HOT</td>
<td>Ayres, David (1793-1881) Ayres, David (1793-1881)</td>
<td>AJW 20140424</td>
</tr>
<tr>
<td>HOT</td>
<td>Austin, William Tennant, 1809-1874 Austin, William Tennant (1809-1874)</td>
<td>AJW 20140424</td>
</tr>
<tr>
<td>HOT</td>
<td>Austin, William Tennant (1809-1874) Austin, William Tennant (1809-1874)</td>
<td>AJW 20140424</td>
</tr>
<tr>
<td>HOT</td>
<td>Ball, George, 1817-1884 Ball, George (1817-1884)</td>
<td>AJW 20140424</td>
</tr>
<tr>
<td>HOT</td>
<td>Ball, George (1817-1884) Ball, George (1817-1884)</td>
<td>AJW 20140424</td>
</tr>
<tr>
<td>HOT</td>
<td>Battle, Mills M., 1800-1856 Battle, Mills M. (1800-1856)</td>
<td>AJW 20140424</td>
</tr>
<tr>
<td>HOT</td>
<td>Battle, Mills M. (1800-1856) Battle, Mills M. (1800-1856)</td>
<td>AJW 20140424</td>
</tr>
</tbody>
</table>
Vocabulary

LCNAF

Theis, Susanne
Theis, Susanne

Johnson, Dianna A.
Johnson, Dianna A.

Young, Anne
Young, Anne

Sidhwa, Bansi
Sidhwa, Bansi

Storrs, Landon R. Y.
Storrs, Landon R. Y.

O'Connor, Kyrie, 1954-
O'Connor, Kyrie, 1954-

Storey, Gail
Storey, Gail

Walker, Margie
Walker, Margie

Koen, Karleen
Koen, Karleen

Allen, John M., ?-1847
Allen, John M., ?-1847

Alley, Abraham, 1803-1862
Alley, Abraham (1803-1862)

Atchison, Daniel D., 1820-1867
Atchison, Daniel D. (1820-1867)

Ayres, David, 1793-1881
Ayres, David (1793-1881)

Austin, William Tennant, 1809-1874
Austin, William Tennant (1809-1874)

Ball, George, 1817-1884
Ball, George (1817-1884)

Battle, Mills M., 1800-1856
Battle, Mills M. (1800-1856)
<table>
<thead>
<tr>
<th>Match Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storrs, Landon R. Y</td>
</tr>
</tbody>
</table>
Authorized Form
Mapping Function

```c
NameMap(vocabulary, listentry, name, confirmed) {
    count = 0
    Loop, parse, listentry, 't
    {
        count++
        if (count == 1)
        {
            IfEqual, vocabulary, %A_LoopField%
            {
                Continue
            }
        }
        else
        {
            Break
        }
    }
    if (count == 2)
    {
        matchname = %A_LoopField%
        StringReplace, matchname, matchname, '\n', All
        StringReplace, matchname, matchname, 'r', All
    }
    if (count == 3)
    {
        authorizedname = %A_LoopField%
        StringReplace, authorizedname, authorizedname, '\n', All
        StringReplace, authorizedname, authorizedname, 'r', All
        Break
    }
    IfEqual, name, %matchname% ; if the current name matches
    {
        global match = 1
        IfNotInString, confirmed, %authorizedname% ; add authorized form
        {
            confirmed := confirmed . authorizedname . ";" ;
        }
    }
    Return confirmed
}
```

Mapping Function

- Check if line belongs to the right vocabulary
Mapping Function

- Check if line belongs to the right vocabulary
- If so, store the match name...
Mapping Function

- Check if line belongs to the right vocabulary
- If so, store the match name...
- ...and the authorized form
Mapping Function

- Check if line belongs to the right vocabulary
- If so, store the match name...
- ...and the authorized form
- Add to the authorized list if it's a positive match
Mapping Function

- Check if line belongs to the right vocabulary
- If so, store the match name...
- ...and the authorized form
- Add to the authorized list if it’s a positive match
- Return authorized list for pasting into CDM field
Basic AutoHotkey Automation

- Delete junk data
- Copy data
- Type repetitive text
Delete Junk Data

• Legacy data repeated at the item level in compound objects

• Manual deletion would take too much time

• Automation makes it possible
Delete Junk Data

```plaintext
^!d::
Loop
{
  Click 1870, 1000 ; metadata scroll bar (lower right)
  Click 980, 715 ; use and reproduction
  Send, ^c ; copy
  Clipwait ; wait for data on clipboard
  StringLen, len, Clipboard ; get string length
  if (len > 5) ; if data is present
  {
    Send, {Delete} ; use and reproduction
    Send, {Up 2}
    Send, {Delete} ; repository
    Send, {Up 4}
    Send, {Delete} ; original collection
    Click 660, 1000 ; page scroll bar
    Click 490, 1025 ; page navigation panel
    Send, {Down} ; next page
  }
  else ; end loop if no data present
  Break
}
```
Delete Junk Data

```java
^!d::: Loop
{
  Click 1870, 1000 ; metadata scroll bar (lower right)
  Click 980, 715 ; use and reproduction
  Send, ^c ; copy
  Clipwait ; wait for data on clipboard
  StringLen, len, Clipboard ; get string length
  if (len > 5) ; if data is present
  {
    Send, {Delete} ; use and reproduction
    Send, {Up 2}
    Send, {Delete} ; repository
    Send, {Up 4}
    Send, {Delete} ; original collection
    Click 660, 1000 ; page scroll bar
    Click 490, 1025 ; page navigation panel
    Send, {Down} ; next page
  }
  else ; end loop if no data present
  Break
}
```
Delete Junk Data

```
^d::
Loop
{
  Click 1870, 1000 ; metadata scroll bar (lower right)
  Click 980, 715 ; use and reproduction
  Send, ^c ; copy
  Clipwait ; wait for data on clipboard
  StringLen, len, Clipboard ; get string length
  if (len > 5) ; if data is present
  {
    Send, {Delete} ; use and reproduction
    Send, {Up 2}
    Send, {Delete} ; repository
    Send, {Up 4}
    Send, {Delete} ; original collection
    Click 660, 1000 ; page scroll bar
    Click 490, 1025 ; page navigation panel
    Send, {Down} ; next page
  }
  else ; end loop if no data present
    Break
}
```
Delete Junk Data

```
^!d::
Loop
{
    Click 1870, 1000 ; metadata scroll bar (lower right)
    Click 980, 715 ; use and reproduction
    Send, ^c ; copy
    Clipwait ; wait for data on clipboard
    StringLen, len, Clipboard ; get string length
    if (len > 5) ; if data is present
    {
        Send, {Delete} ; use and reproduction
        Send, {Up 2}
        Send, {Delete} ; repository
        Send, {Up 4}
        Send, {Delete} ; original collection
        Click 660, 1000 ; page scroll bar
        Click 490, 1025 ; page navigation panel
        Send, {Down} ; next page
    }
    else ; end loop if no data present
    Break
}
```
Delete Junk Data

```bash
^!d::
Loop
{
  Click 1870, 1000 ; metadata scroll bar (lower right)
  Click 980, 715 ; use and reproduction
  Send, ^c ; copy
  Clipwait ; wait for data on clipboard
  StringLen, len, Clipboard ; get string length
  if (len > 5) ; if data is present
    {
      Send, {Delete} ; use and reproduction
      Send, {Up 2}
      Send, {Delete} ; repository
      Send, {Up 4}
      Send, {Delete} ; original collection
      Click 660, 1000 ; page scroll bar
      Click 490, 1025 ; page navigation panel
      Send, {Down} ; next page
    }
  else ; end loop if no data present
    Break
}
```
Delete Junk Data

```plaintext
9  ^!d::
10  Loop
11  {
12       Click 1870, 1000 ; metadata scroll bar (lower right)
13
14       Click 980, 715 ; use and reproduction
15
16       Send, ^c ; copy
17       Clipwait ; wait for data on clipboard
18       StringLen, len, Clipboard ; get string length
19
20       if (len > 5) ; if data is present
21           {
22              Send, {Delete} ; use and reproduction
23              Send, {Up 2}
24              Send, {Delete} ; repository
25              Send, {Up 4}
26              Send, {Delete} ; original collection
27              Click 660, 1000 ; page scroll bar
28              Click 490, 1025 ; page navigation panel
29              Send, {Down} ; next page
30         }
31     else ; end loop if no data present
32        Break
33  }
```
Delete Junk Data

```sh
# !d::
Loop
{
  Click 1870, 1000 ; metadata scroll bar (lower right)
  Click 980, 715 ; use and reproduction
  Send, ^c ; copy
  Clipwait ; wait for data on clipboard
  StringLen, len, Clipboard ; get string length
  if (len > 5) ; if data is present
    Send, {Delete} ; use and reproduction
    Send, {Up 2}
    Send, {Delete} ; repository
    Send, {Up 4}
    Send, {Delete} ; original collection
    Click 660, 1000 ; page scroll bar
    Click 490, 1025 ; page navigation panel
    Send, {Down} ; next page
  }else ; end loop if no data present
    Break
}
Delete Junk Data

```
^!d::
Loop
{
  Click 1870, 1000 ; metadata scroll bar (lower right)
  Click 980, 715 ; use and reproduction
  Send, ^c ; copy
  Clipwait ; wait for data on clipboard
  StringLen, len, Clipboard ; get string length

  if (len > 5) ; if data is present
  {
    Send, {Delete} ; use and reproduction
    Send, {Up 2}
    Send, {Delete} ; repository
    Send, {Up 2}
    Send, {Delete} ; original collection
    Click 660, 1000 ; page scroll bar
    Click 490, 1025 ; page navigation panel
    Send, {Down} ; next page
  }
  else ; end loop if no data present
  Break
```
Delete Junk Data

```plaintext
^d::
Loop
{
  Click 1870, 1000 ; metadata scroll bar (lower right)
  Click 980, 715 ; use and reproduction
  Send, ^c ; copy
  Clipwait ; wait for data on clipboard
  StringLen, len, Clipboard ; get string length
  if (len > 5) ; if data is present
  {
    Send, {Delete} ; use and reproduction
    Send, {Up 2}
    Send, {Delete} ; repository
    Send, {Up 4}
    Send, {Delete} ; original collection
    Click 600, 1000 ; page scroll bar
    Click 490, 1025 ; page navigation panel
    Send, {Down} ; next page
  }
  else ; end loop if no data present
  Break
}
```
Delete Junk Data

```ruby
^!d::
Loop
{
  Click 1870, 1000 ; metadata scroll bar (lower right)
  Click 980, 715 ; use and reproduction
  Send, ^c ; copy
  Clipwait ; wait for data on clipboard
  StringLen, len, Clipboard ; get string length
  if (len > 5) ; if data is present
    {
      Send, {Delete} ; use and reproduction
      Send, {Up 2}
      Send, {Delete} ; repository
      Send, {Up 4}
      Send, {Delete} ; original collection
      Click 660, 1000 ; page scroll bar
      Click 490, 1025 ; page navigation panel
      Send, {Down} ; next page
    }
  else ; end loop if no data present
    Break
  }
```
Delete Junk Data

```plaintext
"!d::
Loop
{
  Click 1870, 1000 ; metadata scroll bar (lower right)
  Click 980, 715 ; use and reproduction
  Send, ^c ; copy
  Clipwait ; wait for data on clipboard
  StringLen, len, Clipboard ; get string length
  if (len > 5) ; if data is present
    {
      Send, {Delete} ; use and reproduction
      Send, {Up 2}
      Send, {Delete} ; repository
      Send, {Up 4}
      Send, {Delete} ; original collection
      Click 660, 1000 ; page scroll bar
      Click 490, 1025 ; page navigation panel
      Send, {Down} ; next page
    }
  else ; end loop if no data present
    Break
}
```
### Delete Junk Data

<table>
<thead>
<tr>
<th>Image</th>
<th>Description</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td>Note from Stephen F. Austin to Samuel May Williams regarding order for disbursement</td>
<td>Austin, Stephen F. (Stephen Fuller), 1793-1836</td>
</tr>
<tr>
<td><img src="image2.png" alt="Image" /></td>
<td>Ordenanza for the municipality of Austin</td>
<td></td>
</tr>
<tr>
<td><img src="image3.png" alt="Image" /></td>
<td>Order to Capt. Harris of the Sloop “Ohio”</td>
<td>Burnet, David Gouverneur (1788-1870)</td>
</tr>
<tr>
<td><img src="image4.png" alt="Image" /></td>
<td>Orders for a “faithful” discharge of John S. Black for his service to the Texas Army</td>
<td>Austin, Stephen F. (Stephen Fuller), 1793-1836</td>
</tr>
</tbody>
</table>
Delete Junk Data

- AutoIt Window Spy for Click commands
Delete Junk Data

- AutoIt Window Spy for Click commands
- Place mouse over correct position on screen
Delete Junk Data

- AutoIt Window Spy for Click commands
- Place mouse over correct position on screen
- Use “In Active Window” screen coordinates
## Copy Data

<table>
<thead>
<tr>
<th>Thumbnail</th>
<th>Publisher</th>
<th>Place of Creation (TGN)</th>
<th>Place of Creation (HOT)</th>
<th>Place of Creation (Local)</th>
<th>Date (ISO 8601)</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Thumbnail" /></td>
<td>China</td>
<td></td>
<td></td>
<td>China</td>
<td>193x</td>
<td>1930s</td>
</tr>
<tr>
<td><img src="image2.png" alt="Thumbnail" /></td>
<td>China</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image3.png" alt="Thumbnail" /></td>
<td>Hawaii</td>
<td></td>
<td></td>
<td>Hawaiian Territory</td>
<td>1939-11?</td>
<td>November 1939</td>
</tr>
<tr>
<td><img src="image4.png" alt="Thumbnail" /></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Copy Data

```
6 7 pause::pause
8 9 ^!m::
10 Sleep, 100
11 Send, ^c
12 Sleep, 100
13 Send, {Right 21}
14 Sleep, 500
15 Send, ^v
16 Sleep, 500
17 Send, {Left 21}
```
Copy Data

Pause key pauses script

```
6
7   pause::pause
8
9   ^!m::
10   Sleep, 100
11   Send, ^c
12   Sleep, 100
13   Send, {Right 21}
14   Sleep, 500
15   Send, ^v
16   Sleep, 500
17   Send, {Left 21}
18
```
Copy Data

Pause key pauses script

Ctrl + Alt + m (hotkey)
Copy Data

Pause key pauses script

Ctrl + Alt + m (hotkey)

Ctrl + c (copy)
Copy Data

Pause key pauses script

Ctrl + Alt + m (hotkey)

Ctrl + c (copy)

Right x 21 (move to Sub.Geo)
Copy Data

Pause key pauses script
Ctrl + Alt + m (hotkey)
Ctrl + c (copy)
Right x 21 (move to Sub.Geo)
Ctrl + v (paste)
**Copy Data**

Pause key pauses script

Ctrl + Alt + m (hotkey)

Ctrl + c (copy)

Right x 21 (move to Sub.Geo)

Ctrl + v (paste)

Left x 21 (move to POC)
Copy Data

Sleep commands, measured in milliseconds, give the GUI time to react
# Metadata Dictionary

## Descriptive Metadata

- Title
- Alternative Title
- Creator
- Contributor
- Publisher
- Place of Creation
- Date (ISO 8601)
- Description
- Caption
- Inscription
- Donor
- Subject: Topical
- Subject: Name
- Subject: Time Period
- Subject: Geographic
- Geographic Coordinates
- Genre
- Language
- Language Code (ISO 639.2)
- Physical Description
- Type (DCMI)
- Format (IMT)
- Original Item Extent

## Administrative Metadata

- Original Item Location
- Original Item URL
- Original Collection
- Original Collection URL
- Digital Collection
- Digital Collection URL
- Repository
- Repository URL
- Use and Reproduction
- Transcript

<table>
<thead>
<tr>
<th>Label</th>
<th>Definition</th>
<th>Input Rules</th>
<th>CONTENTdm Settings</th>
<th>Crosswalk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>&quot;The name given to the resource. Typically, a Title will be a name by which the resource is formally known.&quot; (DCMI)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Letter:** "Letter from x to y (if needed: regarding z)"
**Newsletter:** "title of newsletter and issue date"
**Book:** title taken from the catalog record | |
|         |                                                                          | DC Map: Title: Data Type: Text Search: Yes Hide: No Required: Yes Vocab: No |
|         |                                                                          | MARC: 245 Sa 245 $b                                                         |
### Descriptive Metadata

**Title**: "The name given to the resource. Typically, a Title will be a name by which the resource is formally known." (DCMI)

**Letter**: "Letter from x to y (if needed: regarding z)"

**Newsletter**: "title of newsletter and issue date"

**Book**: title taken from the catalog record

### Administrative Metadata

**Original Item Location**

**Original Item URL**

**Original Collection**

**Original Collection URL**

**Digital Collection**

**Digital Collection URL**

**Repository**

**Repository URL**

**Use and Reproduction**

**Transcript**

---

**CONTENTdm Settings**

**Crosswalk**

**MARC**: 245 Sa

**245 $b**
Metadata Dictionary: digital.lib.uh.edu/about/metadata

### Descriptive Metadata
- Title
- Alternative Title
- Creator
- Contributor
- Publisher
- Place of Creation
- Date (ISO 8601)
- Date
- Description
- Caption
- Inscription
- Donor
- Subject: Topical
- Subject: Name
- Subject: Time Period
- Subject: Geographic
- Geographic Coordinates
- Genre
- Language
- Language Code (ISO 639.2)
- Physical Description
- Type (DCMI)
- Format (IMT)
- Original Item Extent

### Administrative Metadata
- Original Item Location
- Original Item URL
- Original Collection
- Original Collection URL
- Digital Collection
- Digital Collection URL
- Repository
- Repository URL
- Use and Reproduction
- Transcript

<table>
<thead>
<tr>
<th>Label</th>
<th>Definition</th>
<th>Input Rules</th>
<th>CONTENTdm Settings</th>
<th>Crosswalk</th>
</tr>
</thead>
</table>
| Title          | "The name given to the resource. Typically, a Title will be a name by which the resource is formally known." (DCMI) | "Letter: "Letter from x to y (if needed: regarding z)"
Newsletter: "title of newsletter and issue date"
Book: title taken from the catalog record | DC Map: Title
Data Type: Text
Search: Yes
Hide: No
Required: Yes
Vocab: No | MARC: 245 Sa 245 $b |
Type Repetitive Text

Metadata Dictionary: digital.lib.uh.edu/about/metadata

- **Descriptive Metadata**
  - Title, Alternative Title, Creator, Contributor, Publisher, Place of Creation, Date (ISO 8601), Date, Description, Caption, Inscription, Donor, Subject, Temporal, Subject.Name, Subject.Time Period, Subject.Geographic, Geographic Coordinates, Genre, Language, Language Code (ISO 639.2), Physical Description, Type (DCMI), Format (IMT), Original Item Extent

- **Administrative Metadata**
  - Original Item Location, Original Item URL, Original Collection, Original Collection URL, Digital Collection, Digital Collection URL, Repository, Repository URL, Use and Reproduction, Transcript

**Title**: "The name given to the resource. Typically, a Title will be a name by which the resource is formally known." (DCMI)

**Definition**: Letter: "Letter from x to y (if needed: regarding z)"

**Input Rules**: MARC: 245 Sa 245 $b
Type Repetitive Text

```plaintext
^!p::
Send, %A_Space%style="text-align: center;"
Send, {Right}<strong>
Send, {End}{Left 4}</
```
::hdr::Title `Alternative Title `Creator (LC NAF)` Creator (HOT)` … (70+ fields)

; type “hdr” to activate
Resources

- Authority App Source Code: https://github.com/meta_weidner
- Notepad ++ (free text editor): http://notepad-plus-plus.org/
- Metadata Dictionary: http://digital.lib.uh.edu/about/metadata
Thank You

Andrew Weidner
ajweidner@uh.edu