Amelia Earhart's Missing Electra (Turboprop transports): controlled vocabulary in Contentdm

Janet V. Addison

Purdue University, jaddison@purdue.edu

Follow this and additional works at: http://docs.lib.purdue.edu/lib_research
Amelia Earhart’s missing Electra (Turboprop transports)

Controlled vocabulary in Contentdm

Janet V. Addison
Society of American Archivists meeting
August 30, 2007
jaddison@purdue.edu
Outline

- Background to the digital collection
- Statement of immediate issue; overview of surrounding issues of controlled vocabulary
- Some foundations of controlled vocabulary
- Desirables
- Wrap-up and conclusions
George Palmer Putnam Collection of Amelia Earhart Papers

- Gifts from George Palmer Putnam (Amelia’s husband), 1940; and Sally Putnam Chapman (George Putnam’s granddaughter), 2002
- 16.5 cubic feet
- Correspondence, photographs, creative writings, printed material, scrapbooks, personal belongings, and ephemera relating to the life and career of aviator Amelia Earhart
- Bulk 1928-1937
George Palmer Putnam Collection of Amelia Earhart Papers: digital

- 2 digitization projects 1998; 2005
- 1998 project using home-grown system
- 2005 project using Contentdm
- Thesaurus used: Thesaurus for Graphic Materials (TGM)
Choosing Contentdm: cataloger’s viewpoint

- Comes pre-loaded with Thesaurus for Graphic Materials (TGM) which we had already used in previous project
- LCSH and MeSH can be downloaded through Contentdm support page
- Ability to create cross-references
- OAI compliant: Metadata fields map to simple Dublin Core
- Words in item-level metadata are hyperlinked for additional searching
Electra (Turboprop transports)

- Library of Congress Subject Heading (LCSH)
- LCSH terms mixed in with TGM terms
Airplanes

--[nationality]--[country or state]--[city]

Public Note

Includes commercial, private, and military airplanes.

Broader Term

Aircraft

Narrower Term

Airtankers

Biplanes

Bombers

Convertiplanes

Fighter planes

Giders (Aircraft)

Monoplanes

Seaplanes

Supersonic planes

Transport planes

Related Term

Air mail service

Airline industry

Airplane cockpits

Airplane equipment
Electra (Turboprop transports)

Consistent complaint from users

... it seems your website almost always refers to the Lockheed Electra turboprop that was developed much later in the 50’s, rather than the original Lockheed Electra that Amelia flew in the 1930’s which was not a turboprop.
Electra (Greek mythology) and Electra (Carmen)
Lockheed Electra became new authorized subject heading

Subsequent effects

- Need cross reference from Electra (Turboprop transports)
- Cross-reference and heading conflicts with LCSH and with our OPAC entry
- Mixing TGM and LCSH with no differentiating labels
Final chart of decision points for AE controlled vocabulary

- Persons: Library of Congress Name Authority File (LCNAF)
- Geographic Names: Library of Congress Subject Headings (LCSH)
- Topics: (TGM I) supplemented by LCSH
- Genres: Thesaurus of Graphic Materials II (TGM II)
- Local terms
<table>
<thead>
<tr>
<th>Old ID Number</th>
<th>XV.C.5.a</th>
</tr>
</thead>
<tbody>
<tr>
<td>New ID Number</td>
<td>b13f929</td>
</tr>
<tr>
<td>Title</td>
<td>Amelia Earhart walking in front of her Lockheed Electra</td>
</tr>
<tr>
<td>Description</td>
<td>Amelia Earhart walking in front of her Lockheed Electra airplane, ca. 1930s [One original and two poor duplicate photos]</td>
</tr>
<tr>
<td>Repository</td>
<td>Purdue University Libraries, Archives and Special Collections</td>
</tr>
<tr>
<td>Date of Original</td>
<td>1930-1937</td>
</tr>
<tr>
<td>Type</td>
<td>image</td>
</tr>
<tr>
<td>Format</td>
<td>JP2</td>
</tr>
<tr>
<td>Extent of Original</td>
<td>1 photograph b&amp;w 20 x 25 cm.</td>
</tr>
<tr>
<td>Orientation</td>
<td>Landscape</td>
</tr>
<tr>
<td>Collection</td>
<td>George Palmer Putnam Collection of Amelia Earhart Papers</td>
</tr>
<tr>
<td>Subjects</td>
<td>Earhart, Amelia, 1897-1937--Photographs</td>
</tr>
<tr>
<td></td>
<td>Lockheed Electra</td>
</tr>
<tr>
<td></td>
<td>Walking</td>
</tr>
<tr>
<td></td>
<td>Photographic prints</td>
</tr>
<tr>
<td>Collection Web Site</td>
<td><a href="http://www.lib.purdue.edu/earhart">http://www.lib.purdue.edu/earhart</a></td>
</tr>
<tr>
<td>Capture Device</td>
<td>Epson Expression 10000XL Photo Scanner</td>
</tr>
<tr>
<td>Rights</td>
<td><a href="http://www.lib.purdue.edu/earhart/collection-rights.html">http://www.lib.purdue.edu/earhart/collection-rights.html</a></td>
</tr>
</tbody>
</table>
Controlled vocabulary: initial decision points

- Is a controlled vocabulary needed
- Which thesaurus will be used
- One thesaurus or multiple thesauri
- Not just topical terms, but names, places
- Consider form/genre headings also
A controlled vocabulary: For all intents and purposes

- Increase number of access points
- Allow consistency
- Gather like items
- Promote precision of searching
Controlled vocabulary: quote

“The best practice for subjects in shareable metadata is to use relevant controlled vocabularies consistently and explicitly. The controlled vocabularies chosen (there may be more than one used) should be relevant to the resource and known to the community to which the resources described would hold the most interest.”

Source: Best Practices for OAI data provider implementations and shareable metadata (wiki)
Especially helpful for visual materials

- Subject searching involves searching a standardized set of vocabulary terms.
- Using the same terminology throughout a database creates **consistency** and **precision**, thus helping the user find relevant information no matter what terminology is used elsewhere in the record.
Thesaurus choice

- Specificity and stability of the terms
- Content of the collection
- Homogeneous collection (e.g., all photographs or heterogeneous (e.g., photographs, realia, letters, video, etc.)
- Online search functionality
- Chief implementers/maintainers of the vocabulary
- Users performing the searching
The ‘other’ vocabulary

The user’s vocabulary

Did you mean: “Bob”
Cross references

Controlled vocabulary for **Subjects** field

- Electioneering USE Political campaigns
- Electioneering USE Political elections
- Elections USE Political elections
- Electoral college
- Electoral votes USE Electoral college
- Electra (Turboprop transports) USE Lockheed
- Electra
- Electric automobiles
- Electric batteries
- Electric can openers
- Electric chairs
- Electric companies USE Public utility companies
<table>
<thead>
<tr>
<th>Field name</th>
<th>DC map</th>
<th>Data type</th>
<th>Large</th>
<th>Search</th>
<th>Hide</th>
<th>Vocab</th>
<th>add field</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old ID Number</td>
<td>Identifier</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>move to</td>
</tr>
<tr>
<td>New ID Number</td>
<td>Identifier</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>move to</td>
</tr>
<tr>
<td>Title</td>
<td>Title</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>move to</td>
</tr>
<tr>
<td>Description</td>
<td>Description</td>
<td>Text</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>move to</td>
</tr>
<tr>
<td>Creator</td>
<td>Creator</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>move to</td>
</tr>
<tr>
<td>Repository</td>
<td>Publisher</td>
<td>Text</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>move to</td>
</tr>
<tr>
<td>Contributor</td>
<td>Contributors</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>move to</td>
</tr>
<tr>
<td>Date of Original</td>
<td>Date</td>
<td>Date</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>move to</td>
</tr>
<tr>
<td>Type</td>
<td>Type</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>move to</td>
</tr>
<tr>
<td>Format</td>
<td>Format</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>move to</td>
</tr>
<tr>
<td>Extent of Original</td>
<td>Source</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>move to</td>
</tr>
<tr>
<td>Orientation</td>
<td>None</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>move to</td>
</tr>
<tr>
<td>Language</td>
<td>Language</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>move to</td>
</tr>
<tr>
<td>Collection</td>
<td>Relation</td>
<td>Text</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>move to</td>
</tr>
<tr>
<td>Collection Web Site</td>
<td>Relation</td>
<td>Text</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>move to</td>
</tr>
<tr>
<td>Rights</td>
<td>Rights</td>
<td>Text</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>move to</td>
</tr>
<tr>
<td>Subjects</td>
<td>Subject</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>move to</td>
</tr>
<tr>
<td>Date Scanned</td>
<td>None</td>
<td>Date</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>move to</td>
</tr>
<tr>
<td>Capture Device</td>
<td>None</td>
<td>Text</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>move to</td>
</tr>
</tbody>
</table>
### Decisions on labels: This?

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Dublin Core Map</th>
</tr>
</thead>
<tbody>
<tr>
<td>LC Subject</td>
<td>Subject</td>
</tr>
<tr>
<td>TGM Subject</td>
<td>Subject</td>
</tr>
<tr>
<td>Local Subject</td>
<td>Subject</td>
</tr>
<tr>
<td>TGM Form or Genre</td>
<td>Subject</td>
</tr>
</tbody>
</table>
Or This?

- Field Name
- Topics
- Forms/Genres

- Dublin Core Map
- Subject
Or This?

- Field Name
  - Persons
  - Corporate Bodies
  - Topics
  - Local Headings
  - Forms and Genres

- Dublin Core Map
  - Subject
Decision points on desirables

- What metadata format will allow the library to specify which vocabulary any term came from
- What digital asset management system will allow the library to specify which vocabulary any term came from
- What user needs for best search and retrieval
- The “i” word: interoperability
Standards

- Dublin Core schema ➔ The structure
- TGM, LCSH, Local subjects ➔ The content
- End result when cross-searching ➔ Bibliographic Babel?
Controlled vocabulary: what we lacked

- Cross-references don’t show in the public display
- Unable to search single subject fields across collections
- Simple Dublin Core designed to help resource discovery and to be a mapping tool, but doesn’t provide sufficient information for other purposes
LCNAF official name:
Fasher (Sudan)
Fort-Lamy (Chad) or N’Djamena (Chad)
Documentation

- Data dictionary useful as a place to document local decisions
- Good enough? For the machine as well as the human indexer?
Dublin Core Usage in Contentdm
Purdue University Libraries

The official Dublin Core page is the Dublin Core Metadata Initiative: [http://www.dublincore.org](http://www.dublincore.org)


<table>
<thead>
<tr>
<th>Field Name</th>
<th>Dublin Core</th>
<th>Description</th>
</tr>
</thead>
</table>
| Title      | Title       | Two types: (1) transcribed and (2) cataloger-supplied.  
(1) Transcribed: Type as-is into Title field.  
(2) Cataloger-supplied: Formulate a succinct title based on the prominent aspects of the resource.  
The title should strive to uniquely identify the image. For photographs, prefer any written information from the back of the photo. For telegrams, letters, etc. follow formulation as prescribed in AACR2.  
Omit initial articles.  
Brackets, quotation marks, and ending punctuation are not used. |
| Creator    | Creator     | Person/institution chiefly responsible for the creation of the resource. Photographer of a photograph is the creator. Writer of a letter is the creator. |
Desirable

- Rich set of descriptive terms in a controlled vocabulary
- Fewer instances of mixed vocabularies without explanation
- If vocabularies are mixed, make that clear via documentation, field labels, etc.
- Cross-reference display for users
- Even better: more flexible metadata models
- Even better: user access to thesaurus for interaction with hierarchy, related terms, broader terms, narrower terms, etc.
Do your homework
References


References, continued


- Dalmau, Michelle et al., “Integrating thesaurus relationships into search and browse in an online photograph collection,” *Library hi tech*, v. 23, no. 3 (2005)