AAT LOD Microthesauri

Marcia Lei Zeng
International Terminology Working Group (ITWG)
September 5-7, 2014
Dresden, Germany
1. Definition

Microthesaurus: designated subset of a thesaurus that is capable of functioning as a complete thesaurus.

-- ISO25964-2:2013

Microthesauri are different from:

- Derived vocabularies
  - adaptation
  - modification
  - expansion
  - partial adaptation
  - translation

Deriving new vocabularies from a source vocabulary

*New vocabularies depending on a source vocabulary*

outside
(with new categories and terms)

inside

(minor modification)

(partial adaptation)

(slight expansion)

(new encoding)

(adjusting specificity)

(translating)
2. Overview: Situations and decisions for a digital collection that wants to become LOD dataset
AAT-based Vocabularies

3. Full ATT or AAT Microthesauri

4. Sync. With AAT?
   - Yes
   - No

5. Mapped to AAT?
   - Yes
   - No
   - Mapping

6. Used AAT IDs?
   - Yes
   - No
   - Reconcile via Endpoint

Other Non-LOD Vocabs

LOD-Ready Dataset
The need to:
- use,
- create,
- derive, and
- map AAT &
- go to LOD

2. Overview: Situations and decisions for a digital collection that wants to become LOD dataset

- Digital Collection
  - Value Controlled?
    - Yes: Value Control
    - No: Use AAT?
      - Yes: AAT-based Vocabularies
        - Based on AAT?
          - Yes: Mapped to AAT?
            - Yes: Sync. With AAT?
              - Yes: LOD-Ready Dataset
              - No: Reconcile via Endpoint
            - No: Used AAT IDs?
              - Yes: LOD-Ready Dataset
              - No: Mapped to AAT?
        - No: Other Non-LOD Vocabs
  - No: Non-AAT Vocab
    - Use AAT?
      - No: Other Non-LOD Vocabs
      - Yes: Full ATT or AAT Microthesaui

- Mapping
3. Can a microthesaurus be made from an existing thesaurus?

<table>
<thead>
<tr>
<th>Structure</th>
<th>Example</th>
</tr>
</thead>
</table>
| YES Classificatory structure       | • EUROVOC  
• Chinese Classified Thesaurus  
• [English Heritage Thesauri] |
| YES Faceted structure              | • AAT  
• FAST (Faceted Application of Subject Terminology)                     |
| YES Deep hierarchies (family trees)| • AAT  
• NASA Thesaurus  
• INSPEC Thesaurus                                                         |
| NO flat structure [alphabetically organized] | • LCSH  
• hundreds of thesauri                                                  |

Microthesaurus: designated subset of a thesaurus that is capable of functioning as a complete thesaurus. -- ISO25964-2:2013
EuroVoc is split into 21 domains and 127 microthesauri. Each domain is divided into a number of microthesauri.

A microthesaurus is considered as a concept scheme with a subset of the concepts that are part of the complete EuroVoc thesaurus.

http://eurovoc.europa.eu/drupal/?q=node/555
CHIN recommends the use of the AAT for museums with broad humanities collections. The terminology found in the AAT Processes and Techniques Hierarchy is appropriate for use in the Heritage and Decorative Technique fields of the Artefacts Canada: Humanities database, as well as some Condition Fields within museum collections management systems.

CHIN has contributed approximately 2600 French terms to the AAT; these are now visible within the AAT as French language equivalents for the most common terms. This bilingual version of the AAT is used to assist with search.

Canadian Heritage Information Network (CHIN)

- Art & Architecture Thesaurus (AAT) Processes and Techniques Hierarchy
- Canadian Centre for Architecture Bilingual Term Lists
- Art & Architecture Thesaurus (AAT) Agents Facet
- Cultural Objects Name Authority (CONA)
- Art & Architecture Thesaurus (AAT) Materials Facet
- Art & Architecture Thesaurus (AAT) Physical Attributes Facet
- Art & Architecture Thesaurus (AAT)
- Art & Architecture Thesaurus (AAT) Disciplines Hierarchy
- Art & Architecture Thesaurus (AAT) Objects Facet
- Art & Architecture Thesaurus (AAT) Styles and Periods Facet
- The Info-Muse classification system for fine arts and decorative arts museums
- Testing a Vocabulary Standard Against Cataloguing Practice in Canadian Museums
- Guidelines for Forming Language Equivalents: A Model Based on the Art & Architecture Thesaurus

CHIN listed 891 recommended resources.

Only AAT has facets and hierarchies that are listed separately.
Section 2.3.4 Top Concepts

http://vocab.getty.edu/doc/#The_Getty_Vocabularies_and_LOD
The units were recommended to use by projects such as The Canadian Heritage Information Network (CHIN).
What are usually available in a flat structured LOD thesauri

<rdf:RDF>
  <rdf:Description rdf:about="http://id.loc.gov/authorities/subjects/sh85142374">
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    <skos:prefLabel xml:lang="en">Vases</skos:prefLabel>
    <skos:broader rdf:resource="http://id.loc.gov/authorities/subjects/sh85142364"/>
    <skos:narrower rdf:resource="http://id.loc.gov/authorities/subjects/sh2012001607"/>
    <skos:narrower rdf:resource="http://id.loc.gov/authorities/subjects/sh2007001063"/>
    <skos:narrower rdf:resource="http://id.loc.gov/authorities/subjects/sh2004005300"/>
    <skos:related rdf:resource="http://id.loc.gov/authorities/subjects/sh85141432"/>
    <skos:closeMatch>
      <rdf:Description rdf:about="http://d-nb.info/gnd/412636250">
        <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
        <skos:prefLabel xml:lang="DE">Vase</skos:prefLabel>
      </rdf:Description>
    </skos:closeMatch>
  </rdf:Description>
</rdf:RDF>
Easy to get an immediate family members. This is true for all thesauri, LCSH, AGROVOC, etc.
What are special in AAT & are available as LOD

Facets

Hierarchy: Furnishing and Equipment

Concept: containers (receptacles)

Guide term: <containers by form>

concept: vessels (containers)

concept: rhyta

[large] Hierarchies (full coverage, deep layer)

Sub-facets (Indicated by node labels)

concept:

concept:
4. An example
-- Use a <Guide Term> to obtain all concept URIs in a facet or hierarchy
Part 1. Get Data
After choosing a facet or a hierarchy...
1. Get the ID
2. Go to SPARQL Endpoint
   → next slide
2. Go to SPARQL Endpoint
3. Choose "Descendants of a Given Parent"
4. Replace the ID in the Query template
5. Submit
6. Get all URIs and labels under this guide term.

### SPARQL Query

<table>
<thead>
<tr>
<th></th>
<th>5.1.2 Descendants of a Given Parent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td># 5.1.2 Descendants of a Given Parent</td>
</tr>
<tr>
<td>2</td>
<td>select * {?x gvp:broaderExtended aat:300117143.</td>
</tr>
<tr>
<td>3</td>
<td>?x gvp:prefLabelGVP [xl:literalForm ?l]; skos:inScheme aat:</td>
</tr>
<tr>
<td>4</td>
<td>} order by ?l</td>
</tr>
</tbody>
</table>

Note: Here is the text of the query.
I replaced the aat ID, also inserted a line to get the labels, and sort by label:
<table>
<thead>
<tr>
<th>x</th>
<th>l</th>
</tr>
</thead>
<tbody>
<tr>
<td>aat:300391225</td>
<td>&lt;religious visual works by related event&gt;@en</td>
</tr>
<tr>
<td>aat:300391082</td>
<td>Advent candleholders@en</td>
</tr>
<tr>
<td>aat:300391224</td>
<td>Advent wreaths@en</td>
</tr>
<tr>
<td>aat:300178242</td>
<td>Andachtsbilder@en</td>
</tr>
<tr>
<td>aat:300265145</td>
<td>Bhagavad-gītās@en</td>
</tr>
<tr>
<td>aat:300263184</td>
<td>Bible stories@en</td>
</tr>
<tr>
<td>aat:300264513</td>
<td>Bibles@en</td>
</tr>
<tr>
<td>aat:300263411</td>
<td>Bibles historiales@en</td>
</tr>
</tbody>
</table>
Sparql gave me:

- beacons [N]
- buoys [N]
- lighthouses [N]
- light stations [N]
- navigational instruments [N]
- amulets
- Thor's hammers (amulets) [N]
- armrests (object genre)
- backdrops
- backrests
- ceremonial objects
- baptismal syringes
- aeronautical beacons
- airport beacons
- airway beacons
- landmark beacons
- obstruction beacons
- lighthouse lamps
- lightships [N]
7. Download JSON format data, now I have a dataset.

Download Options:
(1) JSON*
(2) XML

*JSON (JavaScript Object Notation) is a lightweight data-interchange format.
AAT URIs and preferred labels under one facet or hierarchy

AAT URIs and labels according to a Contributor

# 5.1.2 Descendants of a Given Parent

```
select * { ?x gvp:broaderExtended aat:300117143.
            ?x gvp:prefLabelGVP [xl:literalForm ?l]; skos:inScheme aat:
            } order by ?l
```
Part 2. Viewing the dataset by a non-techy person

Acknowledgement: Thanks to a Visiting Scholar En-bo Jiang for helping the testing.
How to manage by a non-techy person?

Non-techy person's wish:
I can see what are in the dataset;
I can use a spreadsheet to open and manage it.

Techy-person can prepare the file as:
1. From a JSON* file \(\rightarrow\) convert to CSV** file (can be opened as spreadsheet) using an open source converter

*JSON = (JavaScript Object Notation), a lightweight data-interchange format.

**CSV = Comma Separated Value file format
Using an online converter, turn JSON to CSV.

http://codebeautify.org/view/jsonviewer
```json
{
"root": {
"array": {
"head": {
"results": {
"bindings": [523],
0: {
"x": {
"type": "uri",
"value": "http://vocab.getty.edu/aat/300391082"
}

1: {
"xml:lang": "en",
"type": "literal",
"value": "Advent candleholders"
}
}
}
}
}
}
}
```
http://codebeautify.org/view/jsonviewer
How to manage by a non-techy person?

Non-techy person's wish:
I can see what are in the dataset;
I can use a spreadsheet to open and manage it.

Techy-person can prepare the file as:

1. From a JSON* file \(\rightarrow\) convert to CSV** file (can be opened as spreadsheet) using an open source converter, or
2. From a JSON file \(\rightarrow\) export to spreadsheet from OpenRefine
When uploaded the JSON file to OpenRefine, highlight the first enter in order for the software to tell the structure.
Establish a 'Project', then ready to edit.
To do: need to double check if all node labels and preferred terms are in.

Open the JSON file from spreadsheet on my laptop
If open the XML file from spreadsheet, it looks like:

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>result - binding</td>
<td>result - binding - name</td>
<td>result - binding - uri</td>
<td>result - binding - literal</td>
<td>result - binding - literal - xml:lang</td>
</tr>
<tr>
<td>2</td>
<td>x</td>
<td></td>
<td><a href="http://vocab.getty.edu/aat/300391225">http://vocab.getty.edu/aat/300391225</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>l</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>l</td>
<td></td>
<td></td>
<td></td>
<td>&lt;religious visual works by related event&gt; en</td>
</tr>
<tr>
<td>5</td>
<td>x</td>
<td></td>
<td><a href="http://vocab.getty.edu/aat/300391082">http://vocab.getty.edu/aat/300391082</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>l</td>
<td></td>
<td>Advent candleholders</td>
<td>en</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>x</td>
<td></td>
<td><a href="http://vocab.getty.edu/aat/300391224">http://vocab.getty.edu/aat/300391224</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>l</td>
<td></td>
<td>Advent wreaths</td>
<td>en</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>x</td>
<td></td>
<td><a href="http://vocab.getty.edu/aat/300178242">http://vocab.getty.edu/aat/300178242</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>l</td>
<td></td>
<td>Andachtsbilder</td>
<td>en</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>x</td>
<td></td>
<td><a href="http://vocab.getty.edu/aat/300265145">http://vocab.getty.edu/aat/300265145</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>l</td>
<td></td>
<td>Bhagavad-gitās</td>
<td>en</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>x</td>
<td></td>
<td><a href="http://vocab.getty.edu/aat/300263184">http://vocab.getty.edu/aat/300263184</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>l</td>
<td></td>
<td>Bible stories</td>
<td>en</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>x</td>
<td></td>
<td><a href="http://vocab.getty.edu/aat/300264513">http://vocab.getty.edu/aat/300264513</a></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Summary of the steps

1. Choose the facet or hierarchy you like to start;
2. Find the ID of that concept.
3. Use this template to get the URIs and labels:

```
# 5.1.2 Descendants of a Given Parent
select * {?x gvp:broaderExtended
    aat:300117143.
    ?x gvp:prefLabelGVP
    [xl:literalForm ?l]; skos:inScheme aat:
    } order by ?l
```

4. Use a tool that can treat JSON to view and manage.

- Replace the ID in the Query template
- Submit
- Get the URIs and labels in under this guide term.
- Sort by order (column x)
Additional: Using RelFinder to Visualize

Interactive Relationship Discovery in RDF Data

http://www.visualdataweb.org/relfinder.php
Example: Find relations between Leonardo da Vinci and Renaissance (based on DBpedia dataset)

1. Pointing to a SPARQL end point
2. Type two terms to find matching entries
3. The tool will display the triples one by one
4. Click on any concept to highlight the relations
Leonardo da Vinci and Renaissance (based on DBpedia dataset) 2
My Plan: to create a friendly SPARQL query creator for generating AAT Microthesauri
5. Conclusion

LOD AAT Microthesauri's importance in the Non-AAT World
Controlled vocabulary

Non-AAT

AAT-based vocabularies

Other

Mapped to AAT

Partially

AAT as a target

Did not map to AAT

microthesauri

makes

map to

use
AAT's importance in the Non-AAT World

THANK YOU!
Wish: Provide better SPARQL template interfaces, allowing all kinds of explorations