Assessing Subject Metadata for Images

Hannah Marie Marshall, hmm88@cornell.edu
Metadata Librarian for Image Collections
Cornell University Library
ARLIS/NA+VRA 2016
March 11, 2016
Seattle, Washington
Background
Assessment Goals

• Determine retrieval rates
• Determine the search utility

• Primary Terms
  • “What is the image of?”

• Secondary Terms
  • “What is the image about?”

• Tertiary Terms
  • “How does the image communicate to the viewer?”
Challenges of subject analysis for images

• "Image indexing is a complex socio-cognitive process that involves **processing sensory input** through classifying, abstracting, and mapping sensory data into concepts and entities often expressed through **socially-defined and culturally-justified linguistic labels and identifiers**" (Heidorn, 1999)

• "Concept-based indexing has the advantage of providing higher-level analysis of the image content but is **expensive** to implement and suffers from a **lack of inter-indexer consistency** due to the **subjective nature of image interpretation**" (Chen, Rasmussen, 1999)
Findings – types of terms

Search Utility

• **Primary Terms**
  • “What is the image *of*?”

• **Secondary Terms**
  • “What is the image *about*?”

• **Tertiary Terms**
  • “How does the image communicate to the viewer?”

• **Non-subject Terms**
  • *Descriptive terms that don’t address the subject matter of the work* (i.e. worktype, materials/techniques, style/period)

### TYPES OF TERMS

<table>
<thead>
<tr>
<th></th>
<th>EXISTING DATA</th>
<th>USERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary Terms</td>
<td>64%</td>
<td>37%</td>
</tr>
<tr>
<td>Secondary Terms</td>
<td>12%</td>
<td>16%</td>
</tr>
<tr>
<td>Tertiary Terms</td>
<td>19%</td>
<td>34%</td>
</tr>
<tr>
<td>Non-Subject Terms</td>
<td>5%</td>
<td>13%</td>
</tr>
</tbody>
</table>
Findings – types of terms

Search Utility

- Higher levels of correspondence for images of two-dimensional works
  - Higher retrieval rates
  - Higher search utility
- Users were 2.5 times more likely to use non-subject terms to describe and search for images of three-dimensional works (and non-representational/abstract works)
  - Pottery, jewelry, sculpture
Findings – types of terms

Search Utility

Most common types of non-subject access points

- Culture
- Materials/Techniques
- Style/Period
- Worktype

• Users were 2.5 times more likely to use non-subject terms to describe and search for images of three-dimensional works (and non-representational/abstract works)
  • Pottery, jewelry, sculpture
Findings – literal terms

Retrieval Rates

- Literal matches = successful image retrieval
- Non-matches = unsuccessful image retrieval
- Successful retrieval = 8.5%
- Unsuccessful retrieval = 91.5%
Findings – literal terms
Retrieval Rates

- Of that 8.5%...
  - Primary Terms (75%)
    - “What is the image of?”
  - Secondary Terms (3%)
    - “What is the image about?”
  - Tertiary Terms (16%)
    - “How does the image communicate to the viewer?”
  - Non-subject Terms (6%)
    - Other descriptive metadata that does not address subject meaning (i.e. materials and techniques)
Conclusions

- Primary terms yield the greatest search utility and higher levels of successful image retrieval.
- High numbers of non-subject terms applied to images of three-dimensional and non-representational works suggest that subject metadata is a weak access point for them.
Thank you!