
Emily Porter, Metadata Analyst  Emory University, Atlanta, GA, USA

DIAGNOSTICS

Environmental Scan; Survey

60+ digital projects/collections  13+ schemas/standards  16+ metadata entry tools  22+ metadata storage environments  15+ non-MARC pipes into Primo

Most engagement with:  Dublin Core (Simple) (75%)  MODS (50%)  Custom/local schemas (42%)

Group Task Analysis

Created simple Dublin Core records for digital scholarship project  Test of minimal/core metadata with varying experience levels:  Inconsistent interpretation of element meanings  Data encoding and formatting issues  Need for local content standards (not just a set of elements)

SYSTEM UI INVENTORY


ANALYTICS REVIEW

Reviewed 3 months’ data (Feb-May) as available for systems:

- LUNA Digital Gallery
- Primo Discovery
- ETDs Repository
- Open Emory Repository

Investigated metadata patterns:

- Top 25 pages by URL
- Top 25 search terms

- Top Metadata Engagement:
  - Creators/Personal Names
  - Collection Names
  - Titles
  - Subjects
  - Topics/keywords
  - Time Period/Culture
  - Geographical Names
  - Identifiers/PIDs
  - School/Program Names
  - Contributors (thesis committees; donors)

BENCHMARKING STANDARDS

Descriptive Elements by Schema/Standard: Quantity and Requirements

<table>
<thead>
<tr>
<th>Rank</th>
<th>Element/Concept</th>
<th>Occurrence in % of Systems</th>
<th>UI System Interfacing Utilization</th>
<th>Analytics Activity – in % of Systems</th>
<th>Weighted Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Title</td>
<td>35%</td>
<td>10%</td>
<td>0%</td>
<td>57%</td>
</tr>
<tr>
<td>2</td>
<td>Identifier</td>
<td>50%</td>
<td>10%</td>
<td>50%</td>
<td>40%</td>
</tr>
<tr>
<td>3</td>
<td>Subject (aggregated)</td>
<td>21%</td>
<td>55%</td>
<td>75%</td>
<td>36%</td>
</tr>
<tr>
<td>4</td>
<td>Creator</td>
<td>24%</td>
<td>61%</td>
<td>58%</td>
<td>38%</td>
</tr>
<tr>
<td>5</td>
<td>Location (aggregated)</td>
<td>29%</td>
<td>32%</td>
<td>75%</td>
<td>36%</td>
</tr>
<tr>
<td>6</td>
<td>Date (aggregated)</td>
<td>35%</td>
<td>50%</td>
<td>0%</td>
<td>34%</td>
</tr>
<tr>
<td>7</td>
<td>Subject/Topic/Keyword</td>
<td>15%</td>
<td>48%</td>
<td>75%</td>
<td>31%</td>
</tr>
<tr>
<td>8</td>
<td>Collection</td>
<td>15%</td>
<td>36%</td>
<td>59%</td>
<td>36%</td>
</tr>
<tr>
<td>9</td>
<td>Type</td>
<td>32%</td>
<td>13%</td>
<td>25%</td>
<td>25%</td>
</tr>
<tr>
<td>10</td>
<td>Rights/Access</td>
<td>41%</td>
<td>0%</td>
<td>0%</td>
<td>23%</td>
</tr>
<tr>
<td>11</td>
<td>Location/Institution</td>
<td>3%</td>
<td>26%</td>
<td>75%</td>
<td>20%</td>
</tr>
<tr>
<td>12</td>
<td>Location/Institution/Repository</td>
<td>2%</td>
<td>24%</td>
<td>0%</td>
<td>19%</td>
</tr>
<tr>
<td>13</td>
<td>Language</td>
<td>26%</td>
<td>8%</td>
<td>0%</td>
<td>15%</td>
</tr>
<tr>
<td>14</td>
<td>Description</td>
<td>26%</td>
<td>8%</td>
<td>0%</td>
<td>17%</td>
</tr>
</tbody>
</table>

THEMATIC ANALYSIS

Qualitative analysis of element meanings (concept crosswalk)  Critical for enabling comparisons, quantitative analysis  Created set of tags to relate elements by concept

Based on simple DC, MODS + custom concepts (data, geospatial, other)

Account for variations in naming/semantics

Applied to benchmarking, system UI, analytics data

Tagged all elements with concept terms

Broad concepts and narrower qualifiers/facets

Analyzed tag instances’ frequency, requirements

Enabled querying for both aggregated concepts and facets e.g. Date and Date Created

SCORING

Ranking Methodology:

Pulled most-required elements across all standards (> 15%)

Averaged system UI utilization (structured search/browse/facet/sort)

Averaged analytics activity

Created weighted scoring for criteria:

- 2x for standards’ requiredness
- 1x for system UI utilization
- .5x for analytics

TOP ELEMENT CONCEPTS

<table>
<thead>
<tr>
<th>Rank</th>
<th>Element/Concept</th>
<th>Occurrence in % of Systems</th>
<th>UI System Interfacing Utilization</th>
<th>Analytics Activity – in % of Systems</th>
<th>Weighted Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Title</td>
<td>35%</td>
<td>10%</td>
<td>0%</td>
<td>57%</td>
</tr>
<tr>
<td>2</td>
<td>Identifier</td>
<td>50%</td>
<td>10%</td>
<td>50%</td>
<td>40%</td>
</tr>
<tr>
<td>3</td>
<td>Subject (aggregated)</td>
<td>21%</td>
<td>55%</td>
<td>75%</td>
<td>36%</td>
</tr>
<tr>
<td>4</td>
<td>Creator</td>
<td>24%</td>
<td>61%</td>
<td>58%</td>
<td>38%</td>
</tr>
<tr>
<td>5</td>
<td>Location (aggregated)</td>
<td>29%</td>
<td>32%</td>
<td>75%</td>
<td>36%</td>
</tr>
<tr>
<td>6</td>
<td>Date (aggregated)</td>
<td>35%</td>
<td>50%</td>
<td>0%</td>
<td>34%</td>
</tr>
<tr>
<td>7</td>
<td>Subject/Topic/Keyword</td>
<td>15%</td>
<td>48%</td>
<td>75%</td>
<td>31%</td>
</tr>
<tr>
<td>8</td>
<td>Collection</td>
<td>15%</td>
<td>36%</td>
<td>59%</td>
<td>36%</td>
</tr>
<tr>
<td>9</td>
<td>Type</td>
<td>32%</td>
<td>13%</td>
<td>25%</td>
<td>25%</td>
</tr>
<tr>
<td>10</td>
<td>Rights/Access</td>
<td>41%</td>
<td>0%</td>
<td>0%</td>
<td>23%</td>
</tr>
<tr>
<td>11</td>
<td>Location/Institution</td>
<td>3%</td>
<td>26%</td>
<td>75%</td>
<td>20%</td>
</tr>
<tr>
<td>12</td>
<td>Location/Institution/Repository</td>
<td>2%</td>
<td>24%</td>
<td>0%</td>
<td>19%</td>
</tr>
<tr>
<td>13</td>
<td>Language</td>
<td>26%</td>
<td>8%</td>
<td>0%</td>
<td>15%</td>
</tr>
</tbody>
</table>

Methodology

Apples to kiwis: structured comparison of 28+ schemas/standards reflecting Emory content types:

- Local (Emory application profiles/systems)
- Technical (XML/XSD/ISO standards)
- National/international/consortial standards

Major content standards

For each standard, reviewed and logged:

Descriptive element names, meanings

Required-ness

Repeatability

Of broader interest:

Variability of documentation practices

Lack of required-ness

Semantics/linguistics of element names

Quantitative Analysis (Selected Data)

Element Sets/Standards Reviewed  28  Emory Instances  7 (25%)

Total Elements  1166  Emory: Elements  225 (19%)

Total Required Elements  220 (19%)

Total Required if Applicable Elements  229 (20%)

Total Recommended Elements  57 (5%)

Optional/Unspecified Elements  660 (57%)

Recommended/Optional Elements  717 (61%)