5-21-2021

MARC-y MARC and the Coding Bunch

Anna-Maria Arnljots  
*Utah State University*

Paul Daybell  
*Utah State University*

Kurt Meyer  
*Utah State University*

Andrea Payant  
*Utah State University*

Becky Skeen  
*Utah State University*

Liz Woolcott  
*Utah State University*

Follow this and additional works at: [https://digitalcommons.usu.edu/lib_present](https://digitalcommons.usu.edu/lib_present)

Part of the Library and Information Science Commons

**Recommended Citation**

MARC-y MARC and the Coding Bunch

Anna-Maria Arnljots
Metadata Assistant
anna-maria.arnljots@usu.edu

Paul Daybell
Archival Cataloging Librarian
paul.daybell@usu.edu

Kurt Meyer
Government Information and E-Resource Cataloger
kurt.meyer@usu.edu

Andrea Payant
Metadata Librarian
andrea.payant@usu.edu

Becky Skeen
Special Collection Cataloging Librarian
becky.skeen@usu.edu

Liz Woolcott
Cataloging and Metadata Services Unit Head
liz.woolcott@usu.edu

Utah Library Association Annual Conference
May 21, 2021
Background

• Multi-year research into user search behavior for all metadata standards employed by the unit
  ▪ First phase: MARC
  ▪ Next phases: EAD, Dublin Core

• Project started just as the library moved everyone to work from home

• Whole unit was able to participate in the coding project
Problem Statement

What is the correlation between user search terms, the placement of MARC records in search results lists, and the performance of individual MARC fields in a search process?
Research Questions

• What is the frequency and placement of MARC records in search results list?

• Where are Search terms located in MARC Records?
Methodology
Web Log Analysis

• Focused on the Discovery Layer (Encore) because it was the primary search portal used by patrons

• Pulled list of all URLs accessed on three days

• Put into Airtable and coded
Web Scraping

- Filtered for URLs that lead to search results pages
- Fed URLs into Octoparse, a web-scrapping tool
- Scraped the list of search results, URLs, pagination, and results
- Numbered the results and put into Airtable, linked to originating URL
Airtable

- Search Results List and URLs
  - Extracted bib #
  - Created formula to link to MARC view of bib
  - Unit members pulled up Bib record and copy/pasted it into Airtable
  - Assigned codes for:
    - Creator of record
    - Material type
    - MARC fields where term was found
    - Fields that were not present
  - Automated formula examined wordcount of record
Airtable (continued)

- Web Log URLs
  - Coded for basic search features:
    - Page Types
    - Advanced Search fields used
    - Facets used
    - Page Number
  - Coded the queries (search terms) for:
    - Search term construction
    - Search categories (known item, topical)
    - User Path
    - Known Item Titles
Airtable (continued)

- Known Items pulled out specifically and coded (most for a separate project looking at the discovery layer)
  - Format/Genre
  - Availability
  - Physical or Electronic
  - Location
  - Steps to access
  - Listed by
  - Final Content Provider
  - Checkouts
  - Discoverability in Google Scholar
    - Steps to Access
Results

Research Question #1

What is the frequency and placement of MARC records in search results lists?
Analysis 1.1: How frequently are MARC records showing up in search results?

<table>
<thead>
<tr>
<th></th>
<th>Batch 1</th>
<th>Batch 2</th>
<th>Batch 3</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>MARC-based catalog records</td>
<td>5264</td>
<td>3299</td>
<td>4749</td>
<td>13312</td>
</tr>
<tr>
<td>Records from other platforms</td>
<td>20326</td>
<td>17560</td>
<td>16811</td>
<td>54697</td>
</tr>
<tr>
<td><strong>Total Records</strong></td>
<td><strong>25603</strong></td>
<td><strong>20859</strong></td>
<td><strong>21560</strong></td>
<td><strong>68022</strong></td>
</tr>
<tr>
<td><strong>Percent MARC records</strong></td>
<td><strong>20.56%</strong></td>
<td><strong>15.82%</strong></td>
<td><strong>22.03%</strong></td>
<td><strong>19.57%</strong></td>
</tr>
</tbody>
</table>
Analysis 1.2: Is there a difference between locally created records and vendor supplied records in the frequency of listing in search results?

<table>
<thead>
<tr>
<th>Record Creator</th>
<th># Records in results list</th>
<th>% Total records in results list</th>
<th># Records accessed</th>
<th>% Total records accessed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vendor</td>
<td>7,727</td>
<td>58.05%</td>
<td>163</td>
<td>39.00%</td>
</tr>
<tr>
<td>Cataloging and Metadata Services</td>
<td>5,066</td>
<td>38.06%</td>
<td>239</td>
<td>57.18%</td>
</tr>
<tr>
<td>Distance Campus Libraries</td>
<td>410</td>
<td>3.08%</td>
<td>5</td>
<td>1.20%</td>
</tr>
<tr>
<td>Record unavailable at time of coding</td>
<td>52</td>
<td>0.39%</td>
<td>2</td>
<td>0.48%</td>
</tr>
<tr>
<td>Patron Services, Library Media Collections, or Resource Sharing and Document Delivery</td>
<td>33</td>
<td>0.25%</td>
<td>8</td>
<td>1.91%</td>
</tr>
<tr>
<td>Acquisitions</td>
<td>16</td>
<td>0.12%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>Unknown</td>
<td>5</td>
<td>0.04%</td>
<td>1</td>
<td>0.24%</td>
</tr>
<tr>
<td>Natural History Library</td>
<td>3</td>
<td>0.02%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>13,312</strong></td>
<td><strong>418</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Analysis 1.3: How are MARC records ranked in the search results list?

- Most common position for MARC records in a search result set of 25 items, is position 4
- MARC records appear in the top five search results 25.35% of the time
Analysis 1.4:
Where do MARC records for known items rank in the search results list?

| Percentage of Times Available Whole Object Appeared in Search Results by Position Number |
|-----------------------------------------------|----------------|----------------|----------------|----------------|-----------------|----------------|----------------|----------------|
|                                               | Result 1 | Result 2 | Result 3 | Result 4 | Result 5 | Results 6-10 | Results 11-15 | Results 16-20 | Results 21-25 |
| **Total #**                                   | 125      | 107      | 61       | 49       | 37       | 104           | 67             | 56             | 35             |
| **% in results**                              | 18.7%    | 16.0%    | 9.1%     | 7.3%     | 5.5%     | 15.6%         | 10.0%          | 8.4%           | 5.2%           |
Results

Research Question #2

Where are search terms located in MARC records?
Analysis 2.1:
What fields are used most in retrieving records?

MARC Fields Where Search Terms Were Located (Top 5)
Analysis 2.2:
For records accessed by the patron, is there a difference in where search terms are located?

- The 245 Title statement remained highest, appearing 64% more often than the next most utilized field.
- Instead of the 505 Formatted Contents Note being in second place, the 650 Subject Added Entry is the next most used field.
- The 505 Formatted Contents Note and 520 Summary fields retained a spot in the top four fields.
Analysis 2.3: For locally created records and vendor-supplied records, is there a difference in where search terms are located?

<table>
<thead>
<tr>
<th>Field</th>
<th>Field Description</th>
<th>CMS Records</th>
<th>Vendor Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>245</td>
<td>Title Statement</td>
<td>43.80%</td>
<td>51.64%</td>
</tr>
<tr>
<td>505</td>
<td>Formatted Contents Note</td>
<td>28.13%</td>
<td>69.65%</td>
</tr>
<tr>
<td>650</td>
<td>Subject Added Entry - Topical</td>
<td>40.89%</td>
<td>56.58%</td>
</tr>
<tr>
<td>520</td>
<td>Summary, etc.</td>
<td>23.41%</td>
<td>76.03%</td>
</tr>
<tr>
<td>600</td>
<td>Subject Added Entry – Personal Name</td>
<td>59.94%</td>
<td>32.68%</td>
</tr>
</tbody>
</table>
## Analysis 2.4:
What fields are not present in the records?

<table>
<thead>
<tr>
<th></th>
<th>CMS</th>
<th>Vendor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not Present</td>
<td>Present</td>
</tr>
<tr>
<td>Author (both 1xx and 7xx)</td>
<td>0.75%</td>
<td>99.25%</td>
</tr>
<tr>
<td>Subject (any authorized)</td>
<td>4.46%</td>
<td>95.54%</td>
</tr>
<tr>
<td>505 Formatted Contents Note</td>
<td>63.96%</td>
<td>36.04%</td>
</tr>
<tr>
<td>520 Summary Note</td>
<td>75.60%</td>
<td>24.40%</td>
</tr>
<tr>
<td>All Categories Present</td>
<td>14.86%</td>
<td></td>
</tr>
</tbody>
</table>
Analysis 2.5:
Which fields would make the greatest impact if not included in the record?

- The top four fields with the greatest impact on retrieval, if not found in a record: 505, 245, 520, and 650

- Without the 505 or 520, 16.86% of all records appearing in results would not have shown up

- In contrast, without 650 and 600 fields, only 0.66% of records would not have appeared in the search results
Analysis
Analysis

• Non-MARC records have advantage over MARC

• MARC vendor records appear more often than locally created MARC records

80 %

25 %

505/520

1xx/6xx/7xx

Of all records in search results are Non-MARC

Of MARC records place in the top 5 search results.

Occur more frequently in vendor records

Occur at the same rate in Vendor and Locally created records
Title fields are most important over all, but...

- Ranked higher than 245 for records where search terms matched only one field
- Consistently in the top 4 fields that retrieved a record (along with 520)
- If missing, 12% of all MARC results would not have been displayed
Analysis

Subject fields are important

3rd Most important field for matching search terms

2nd Most important field for records viewed by patrons

1xx fields were much more likely to be “clicked on”

But...

Would not have been displayed if field were missing

.55% Instance of subject fields being “clicked on”
Take-Aways

Cataloger will retain ability to make best judgment for each record, but will be asked to consider the following guidelines:

- More emphasis on creating 505 and 520 notes in local records
- Less emphasis on 6xx fields as an entry point
- More emphasis on 1xx fields as an entry point
MARC-y MARC's Coding Bunch

- Anna-Maria Arnljots
- Josee Butler
- Ryan Bushman (Stats)
- Paul Daybell
- Barbara Fleming
- Maddie Gardner
- Alisha Grant
- Bryn Larsen
- Sabrina Leatham
- Rachel Olsen
- Andrea Payant
- Kurt Meyer
- Jessica Mills
- Abby Rodabough
- MaKayla Roundy
- Melanie Shaw
- Becky Skeen
- Sara Skindelien
- Seth Westenburg
- Liz Woolcott
Resources
Full Procedures: https://usulibrary.atlassian.net/l/c/8H7jgU98

Article with final results:

Related articles
• Presentation: https://www.libraryassessment.org/program/2020-schedule/#jan21
• Paper: https://www.libraryassessment.org/2020-proceedings/
Questions?

Anna-Maria Arnljots  
Metadata Assistant  
anna-maria.arnljots@usu.edu

Paul Daybell  
Archival Cataloging Librarian  
paul.daybell@usu.edu

Kurt Meyer  
Government Information and E-Resource Cataloger  
kurt.meyer@usu.edu

Andrea Payant  
Metadata Librarian  
andrea.payant@usu.edu

Becky Skeen  
Special Collection Cataloging Librarian  
becky.skeen@usu.edu

Liz Woolcott  
Cataloging and Metadata Services Unit Head  
liz.woolcott@usu.edu
Thank You!