Utah State University

DigitalCommons@USU

Library Faculty & Staff Presentations   Libraries

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

Part of the Library and Information Science Commons

UtahStateUniversity
MERRILL-CAZIER LIBRARY

MARC-y MARC and the Coding Bunch

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

Andrea Payant
Metadata Librarian
andrea.payant@usu.edu

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

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

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

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

Utah Library Association Annual Conference
May 21, 2021

UNIVERSITY LIBRARIES
Utah State University.

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?
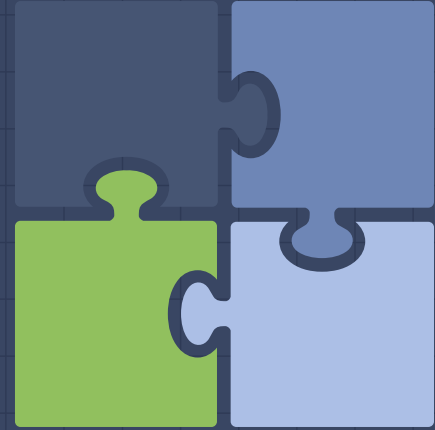
Research Questions

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

- Where are Search terms located in MARC Records?
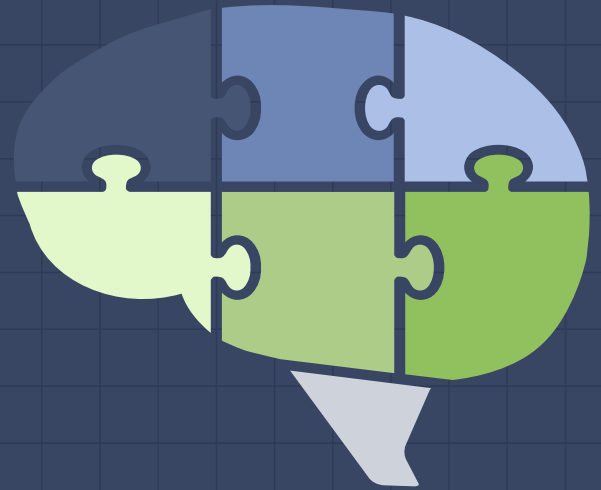
Methodology

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

- Scrapped 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?

	Batch 1	Batch 2	Batch 3	Combined
MARC–based catalog records	5264	3299	4749	13312
Records from other platforms	20326	17560	16811	54697
Total Records	**25603**	**20859**	**21560**	**68022**
Percent MARC records	*20.56%*	*15.82%*	*22.03%*	*19.57%*

Analysis 1.2:
Is there a difference between locally created records and vendor supplied records in the frequency of listing in search results?

Record Creator	# Records in results list	% Total records in results list	# Records accessed	% Total records accessed
Vendor	7,727	58.05%	163	39.00%
Cataloging and Metadata Services	5,066	38.06%	239	57.18%
Distance Campus Libraries	410	3.08%	5	1.20%
Record unavailable at time of coding	52	0.39%	2	0.48%
Patron Services, Library Media Collections, or Resource Sharing and Document Delivery	33	0.25%	8	1.91%
Acquisitions	16	0.12%	0	0.00%
Unknown	5	0.04%	1	0.24%
Natural History Library	3	0.02%	0	0.00%
Total	**13,312**		**418**	

- 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)
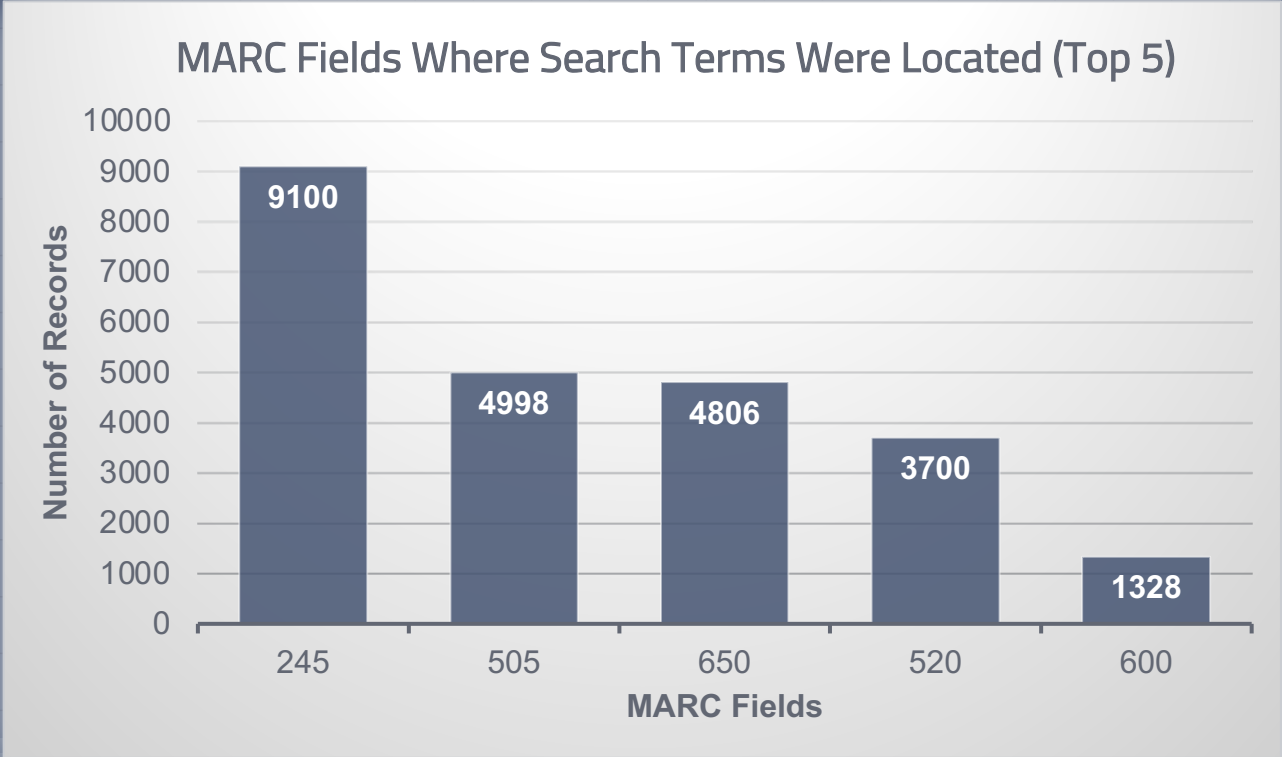
Analysis 2.2:
For records accessed by the patron, is there a difference in where search terms are located?

- 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?

Percentage of fields used in record retrieval (top 5 most frequent)			
Field	Field Description	CMS Records	Vendor Records
245	Title Statement	43.80%	51.64%
505	Formatted Contents Note	28.13%	69.65%
650	Subject Added Entry - Topical	40.89%	56.58%
520	Summary, etc.	23.41%	76.03%
600	Subject Added Entry – Personal Name	59.94%	32.68%

Analysis 2.4:
What fields are not present in the records?

	CMS		Vendor	
	Not Present	**Present**	**Not Present**	**Present**
Author (both 1xx and 7xx)	0.75%	99.25%	1.18%	98.82%
Subject (any authorized)	4.46%	95.54%	6.73%	93.27%
505 Formatted Contents Note	63.96%	36.04%	45.54%	54.46%
520 Summary Note	75.60%	24.40%	50.45%	49.55%
All Categories Present	14.86%		33.26%	

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 % Of all records in search results are Non-MARC

25 % Of MARC records place in the top 5 search results.

505/520 Occur more frequently in vendor records

1xx/6xx/7xx Occur at the same rate in Vendor and Locally created records

Analysis

Title fields are most important over all, but...

505 =

- 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

But…

.55% Would not have been displayed if field were missing

1 Instance of subject fields being "clicked on"

1xx fields were much more likely to be "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:

Liz Woolcott, Andrea Payant, Becky Skeen & Paul Daybell (2021) **Missing the MARC: Utilization of MARC Fields in the Search Process**, Cataloging & Classification Quarterly, 59:1, 28-52, DOI: 10.1080/01639374.2021.1881010

Related articles

Robert Heaton & Liz Woolcott. **Unraveling the (Search) String: Assessing Library Discovery Layers Using Patron Queries**. Library Assessment Conference, January 2021

- 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

Andrea Payant
Metadata Librarian
andrea.payant@usu.edu

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

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

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

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

Thank You!