

Applications of the *QA Catalogue for Analysing Library Data*

David Floyd, Chief Cataloging Librarian
Sasha Frizzell, Catalog/Metadata Management Librarian
Binghamton University
CITSIG, Core Interest Group Week, March 8, 2024





Agenda

- Introduction to the QA catalogue
- Projects we've completed based on QA findings
- Future project ideas

Web application, developed by Péter Király, that provides a visual interface for the analysis of large sets of MARC metadata.

Outputs include:



Record completeness



MARC coding issues



Quantitative access point review



Several unique analyses designed for serials metadata

Shelf-Ready completeness

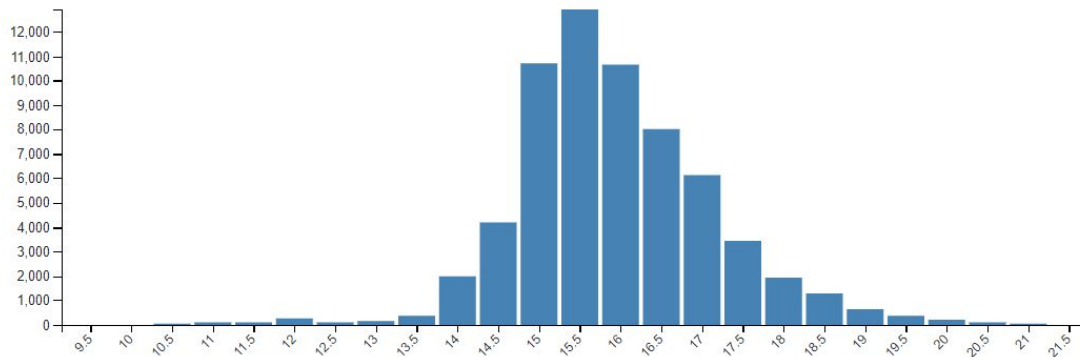
These scores are the implementation of the following paper:

Emma Booth (2020) *Quality of Shelf-Ready Metadata. Analysis of survey responses and recommendations for suppliers*
Pontefract (UK): National Acquisitions Group, 2020. p 31. https://nag.org.uk/wp-content/uploads/2020/06/NAG-Quality-Shelf-Ready-Metadata-Survey-Analysis-and-Recommendations_FINAL_June2020.pdf

The main purpose of the report is to highlight which fields of the printed and electronic book records are important when records are coming from different suppliers. 50 libraries participated in the survey, each selected which fields are important. The report listed those fields which gets the highest scores.

The current calculation based on this list of essential fields. If all data elements specified are available in the record it gets the full score, if only some of them, it gets a proportional score. E.g. under 250 (edition statement) there are two subfields. If both are available, it gets score 44. If only one of them, it gets the half of it, 22, and if none, it gets 0. For 1XX, 6XX, 7XX and 8XX the same logic applies. A record gets the full scores if at least one of those fields (with subfield \$a) is available. The total score became the average. The maximum score would be 28.44, which could be accessed if all the data elements are available in the record.

histogram



Can't you get this information from your ILS?

QA tool is better at:

- 🔍 Comprehensive overviews
- 🔍 Addressing records and fields not indexed by the ILS

QA tool is also:

- 🔍 Open-source
- 🔍 Not ILS dependent

How it works

1

Collect MARC
files in .mrc
format

2

QA tool
transforms them
into CSV files

3

Web interface
provides
visualizations of
data

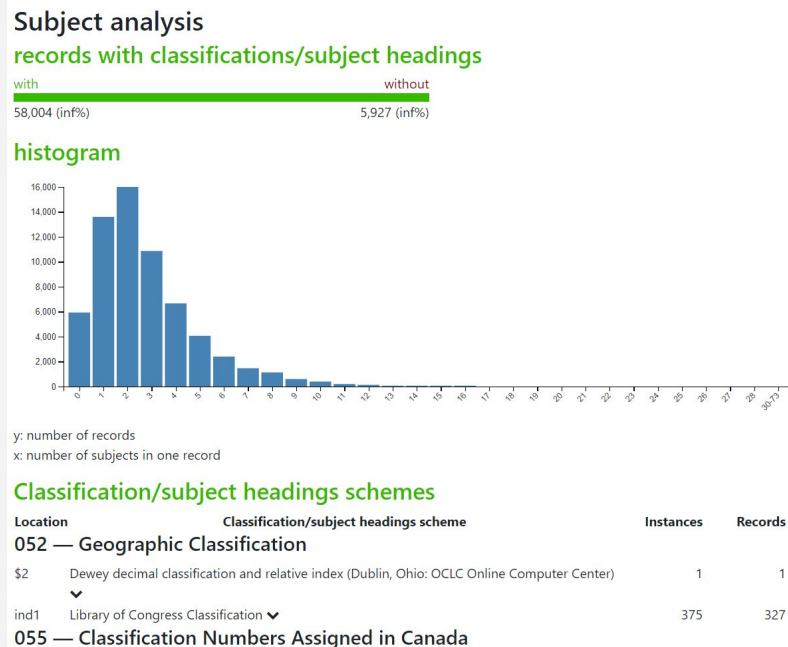


Projects stemming from QA Analysis

1. Vendor metadata quality assessment

Comparison of vendor metadata in the Alma Community zone, Binghamton's Institution Zone, and Cornell University's catalog.

Focus on if the records included classification. Additional review of subject heading quantity, quality, and source vocabularies.



Projects stemming from QA Analysis

2. Metadata enhancement project for Pleasure Reading

Completeness of MARC21 fields

path	label		terms	records		occurrences				
				count	%	count	min	max	mean	stddev
Subject Access										
655 — Index Term - Genre/Form										
\$ind1	Type of heading	<div></div>	<div></div>	12,533	19.60%	20,609	1	13	1.64	0.00
\$ind2	Thesaurus	<div></div>	<div></div>	12,533	19.60%	20,609	1	13	1.64	0.00
\$a	Genre/form data or focus term	<div></div>	<div></div>	12,533	19.60%	20,610	1	13	1.64	0.00
\$b	Non-focus term		<div></div>	2	0.00%	2	1	1	1.00	0.00
\$c	Facet/hierarchy designation		<div></div>	21	0.03%	27	1	3	1.29	0.09
\$v	Form subdivision		<div></div>	6	0.01%	7	1	2	1.17	0.14
\$x	General subdivision		<div></div>	23	0.04%	33	1	6	1.43	0.23
\$y	Chronological subdivision		<div></div>	216	0.34%	291	1	6	1.35	0.03
\$z	Geographic subdivision		<div></div>	198	0.31%	403	1	7	2.04	0.04
\$0	Authority record control number		<div></div>	281	0.44%	553	1	6	1.97	0.02
\$2	Source of term	<div></div>	<div></div>	11,097	17.36%	15,451	1	13	1.39	0.00
\$3	Materials specified		<div></div>	1	0.00%	4	4	4	4.00	0.00
\$5	Institution to which field applies		<div></div>	50	0.08%	77	1	7	1.54	0.12

Projects stemming from QA Analysis

3. Enhancement of records created by Binghamton,

First subset: language of cataloging empty

Numbers and Code

040 — Cataloging Source

\$a	Original cataloging agency			63,931	100.00%
\$b	Language of cataloging			63,516	99.35%
\$c	Transcribing agency			63,927	99.99%
\$d	Modifying agency			62,587	97.90%
\$e	Description conventions			8,668	13.56%



Future projects

- 🔍 Enhancement of local origin records
- 🔍 Evaluation of Special Collections cataloging
- 🔍 Targeted authority control for faculty
- 🔍 Implement into decision making workflows for large sets of vendor metadata

We have accomplished a lot already with this tool, and full disclosure, we are still at about 40% functionality. The British Library (<http://141.5.102.114/bl/?tab=completeness>) has a successful full implementation that shows signs of where things can go from here.



Thank you



QA Catalogue on GitHub
<https://github.com/pkiralv/qa-catalogue>

Questions? Contact

David - DFloyd@Binghamton.edu

Sasha - SFrizzell@Binghamton.edu