Reinventing the Cataloging Hold
Smarter Processing Through Data Integration

Nina Servizzi
Associate Dean, Knowledge Access & Resource Management Services

Potomac Technical Processing Librarians 94th Annual Meeting
Washington D.C., October 12, 2018
Introduction

I. Assessment in Technical Services
II. Data Integration
III. Cataloging Hold: Backlog
IV. Cataloging Hold: Workflow Redesign
V. Cataloging Hold: Implementation
VI. Next Steps
Assessment in Technical Services

“Become a data-aware organization that promotes open Exploration of data and grounds our decisions in evidence.”

Assessment in Technical Services

What did this mean for Technical Services?

Challenge to move beyond simply reporting statistics and begin (re)designing workflows and processes that are truly data-driven.
Data Integration

Data needed for workflow development comes from many different sources and is not readily available to staff:

- ILS (Aleph Oracle tables)
- Discovery System (Primo)
- Open Link Resolver (SFX)
- Accounting System (FAME)
- University Registration (Student/Employee data)
- OCLC (MARC, Kbart)
- Vendor Systems (GOBI, Oasis)
- Database/Publisher Platforms (ProQuest, Springer, EBC)
- COUNTER statistics
Data Integration

Heterogeneous data must be normalized and pulled together in a centralized repository

Data within the repository must then be easily accessible to staff with varying levels of querying skills
Data Integration

Solution: Build a data warehouse for the Libraries

Data Integration

Libraries Data Warehouse (LDW)

Locally hosted MS SQL Server
• Developed using SQL Server Integration Services (SSIS)
  • Integrated development environment
  • Robust suite of ETL (Extract, Transform, and Load) tools
  • Cost effective (free with MS SQL Server license)
Data Integration

Data access premises:

1. End users must be able to access and manipulate data themselves
2. Reporting tool must be able to analyze and visually present data in a wide variety of ways

Primary front end tool:
Tableau, Excel

Backend tools:
SQL Server Management Studio, ODBC, Toad
Data Integration

Data access premises:

1. End users must be able to access and manipulate data themselves
2. Reporting tool must be able to analyze and visually present data in a wide variety of ways

Primary front end tool:
Tableau, Excel

Backend tools:
SQL Server Management Studio, ODBC, Toad
Cataloging Hold: Backlog

2014/2015 Cataloging backlog was over 32,000 physical volumes

Quantitative analysis of receiving and cataloging statistics revealed:

- 87,785 total receipts for New York collections
- 30.35% (n=26,803) cataloged on receipt in Resource Management Department (RM)
- 59.12 % (n=51,900) cataloged in Metadata Production & Management Department (MP&M)
- 10.53% (n=9,082) remained un-cataloged
Cataloging Hold: Backlog

Current workflows not keeping up with rate of collection growth

Qualitative analysis revealed other inefficiencies:

• FIFO workflows resulted in materials being handled by multiple staff before ending in backlog
• Original catalogers processing copy ‘on encounter’ resulted in low rate of original cataloging
• Maintaining hold in system number order for RUSH processing required 1 FTE; large increased RUSH requests
• Uncertainty about cataloging policy caused staff to set materials ‘aside’
Cataloging Hold: Backlog

Administrative numbers revealed deeper systemic problems:

- There was a $300,000 over expenditure in adjunct cataloging budget (20 adjuncts on staff = 10 FTE)
- E-Resource purchases made up 77.64% of collections budget
- Less than 1.5 FTE working on e-resources resulted in backlog of estimated 1,000,000 un-cataloged e-books
Cataloging Hold: Workflow Redesign

Redesign processing workflows between RM and MP&M to address inefficiencies particularly:

- Increase cataloging capacity to meet collection growth
- Separate copy/original cataloging streams, eliminate parallel processing streams
- Clarify cataloging policy
- Streamline hold management and reassign ownership from MP&M to RM
- Reduce hold and minimize RUSH processing requests
- Increase staff available for e-resource work and reduce reliance on adjunct catalogers
Cataloging Hold: Workflow Redesign

Spring 2015 – Fall 2016 successful proof of concept which reduced backlog from over 32,000 to under 13,000

- Pilot workflow used an internally-developed algorithm to check volumes in the hold against OCLC WorldCat
- If a full bibliographic record was found for the title, the record was automatically downloaded and overlaid in our ILS
- Staff were then provided with pull lists and the matched bibliographic records checked against physical item.
- Correctly matched items were sent to shelf
- False matches were analyzed and match algorithm updated
Cataloging Hold: Workflow Redesign

To move from proof of concept to a fully automated and well articulated workflow took about two years to fully implement workflow required changes in two areas:

1. Resource Allocation
   - Created a Metadata Policy Committee (MPC)
   - Establish new policy on rapid cataloging based on analysis of match algorithm results
   - Write new documentation and train staff
   - Reassign staff to well defined functional areas
   - Reduce adjunct catalogers through natural attrition
Cataloging Hold: Workflow Redesign

2. Automation Capacity

• Build tables in the LDW that track data for un-cataloged materials that can be used for processing triage (updated hourly)
  • Material type (DVD, CD, book, etc.)
  • Holding location (Library/sublibrary)
  • Bibliographic quality (cataloging entity, encoding level, call number type, subject headings, etc.)
  • Shelving location in processing department

• Refine local match algorithm for each type of resource

• Build interactive dashboards for staff to access data in LDW
Cataloging Hold: Implementation

Implementation of new workflows has been gradual process to allow staff to understand the context of the changes and the role they play in aligning processing with the Libraries strategic goals.

• Change in metadata policy announced
• New staff assignments introduced
• Introduction of new workflows
• Training becoming comfortable with new tools
  • GitHub pull requests, batch record enhancements, LDW dashboards, inventory systems, Jira ticketing, etc.

Throughout implementation staff feedback and close monitoring has been used to drive iterative development cycles for or LDW and automation scripts
Cataloging Hold: Implementation

Automated hold maintenance in full production July 2018 has yielded a 107% increase in volumes identified as having full-copy available and moved out of hold.

<table>
<thead>
<tr>
<th>By Unit Monthly Feb-Oct 18</th>
<th>Month of Z30 Update Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z30 Cataloger (group)</td>
<td></td>
</tr>
<tr>
<td>KADD</td>
<td>98</td>
</tr>
<tr>
<td>MP&amp;M Admin/Faculty</td>
<td>1,230</td>
</tr>
<tr>
<td>MP&amp;M Staff</td>
<td>1,872</td>
</tr>
<tr>
<td>MP&amp;M Students</td>
<td>557</td>
</tr>
<tr>
<td>RM Monographs Unit Staff</td>
<td>342</td>
</tr>
<tr>
<td>RM Monographs Unit Students</td>
<td>891</td>
</tr>
<tr>
<td>RM MP&amp;GL Staff</td>
<td>745</td>
</tr>
<tr>
<td>RM MP&amp;GL Students</td>
<td>137</td>
</tr>
<tr>
<td>Grand Total</td>
<td>5,735</td>
</tr>
</tbody>
</table>
Cataloging Hold: Implementation

Our numbers for 2017/2018:

• 76,564 total receipt of physical volumes
• 98.99% (n=76,564) of total receipts cataloged
• 38.94% ( n=30,218) cataloged on receipt
• 111.15 % (n=7042) increase in Original cataloging compared to ( n-3335) in 2015/2016
Cataloging Hold: Implementation

Other benefits:

- Collections growth stabilized around 75,000 volumes per year
- No over expenditure in adjunct cataloging budget; adjuncts have been reduced to 9
- Increase in number of staff working on e-resources to 5 FTE
- Over 1,200,000 e-books cataloged; backlog eliminated
Cataloging Hold: Next Steps

Use hold dashboard to create identify specific original cataloging projects and identify gaps in our cataloging capacity:
Questions?

Contact:

Nina Servizzi
Nina.Servizzi@nyu.edu