IPLC’s Collection Analysis with Gold Rush

Galadriel Chilton
Ivy Plus Libraries Confederation

Jeanette Norris
Yale University Library

Sarah Tudesco
Yale University Library
Before proceeding further, we would like to pause for a moment to acknowledge and honor the Native Americans and all of the Indigenous peoples of the land upon which we are all gathered today.
Brown University
Columbia University
Cornell University
Dartmouth College
Duke University
Harvard University
Massachusetts Institute of Technology (MIT)
University of Pennsylvania
Princeton University
Stanford University
The University of Chicago
The Johns Hopkins University
Yale University
Part 1.
Two Projects

Library of Congress
1. Collection Analysis Dataset Feasibility Study
(October 2018 - April 2019)

Project Goals:
- Test methods of collaboration
- Document time and resources needed
- Explore:
  - How messy is the data?
  - Is Gold Rush’s Library Content Comparison Tool a viable option for IPLC’s collection analysis goals?
Outcome:
Collaboration requires work, synchronous communication, and learning together.
Outcome: Collaboration requires time and project management.
2. Analysis for Collaborative Collection Development
5-Part Project, 26 People (October 2019 - June 2020)

- Data Analysis
  4 + Project Manager

- Use Cases / Data Analysis Approach
  3 + Project Manager

- Defining Strategic Duplication
  5 + Project Manager

- Use Data Feasibility
  2 + Project Manager

- Data
  11 Colleagues
Part 2.
Gold Rush & Data Analysis
Data Analysis Deliverables

Project 1: Feasibility Study
1. Evaluation of data analysis methods internal & external to Gold Rush
2. Analysis of Gold Rush Match Key effectiveness for identifying duplicates
3. Recommendations for future analysis

Project 2: Collaborative Collection Development
1. Datasets - normalized and enhanced
2. Tableau Dashboard
3. R Scripts
4. Advice on when to go back to GoldRush
5. Recommendations for future analysis
Our group was charged to develop a methodology to analyze the data and answer the questions outlined by the use cases.

**Goals:**
- Evaluate GoldRush Library Content Comparison Tool
- Develop methodology for analyzing holdings vs. usage
- Provide information to inform a collaborative collection development pilot

**Process:**
- Meet regularly
- Work collaboratively
- Utilize diverse tools: R, Tableau, MarcEdit, and OpenRefine
- Normalize and enhance data to make them more useful
Data Overview

Library Holdings
Gold Rush Content Comparison Tool
- Single part print monographs
- Publication date between 2013-2017
- Holdings data from 13 institutions
- Numbers...
  - 3,921,524 records
  - ~1,678,579 distinct titles

Use Transactions
BorrowDirect (Metridoc System)
- Filled and unfilled requests
- Transactions from 2012-2019
- Requests from 13 institutions*
- Numbers...
  - 1,872,741 records
  - ~844,506 distinct titles

*Data from when institutions joined BD program.
Comparison Dataset

We could not compare at title level, so we looked at data points that are common in both datasets.

1. Top Publishers
2. LC Call Number (Class & Subclass)
3. Publication Date 2013-2017

What do they have in common? How do they differ?

Demonstration dashboard: Top publishers with over 10,000 holdings → 18 publishers in the dashboard.
Interactive Data Dashboard

The dataset focuses on top publishers.

**Use** = Borrow Direct transactions (usage data)

**Hold** = GoldRush holdings data (single-part, monographs, published between 2012-2017)

**Use%**: percentage of transactions in Borrow Direct. (e.g. Princeton Univ Press represents 7.91% of the Borrow Direct transactions in this dataset).

**Hold%**: percentage of holdings in GoldRush. (Princeton Univ Press represents 2.07% of records in this dataset).

**Diff%**: Formula \([\text{Use\%}] - [\text{Hold\%}]\). If the value is positive, demonstrates more usage relative to holdings. If the value is negative, demonstrates more holdings relative to usage.
Duplication of Holdings vs. Copies Needed

**Duplication (Holdings)**
chart shows holdings overlap between 13 Ivy Plus libraries

**Copies Needed (Usage)**
chart shows how many copies were needed to fill demand from our resource sharing transactions dataset
Holdings (%) v. Use (%)

Key:
1. Color = LC Class
2. Size of Bubble = #Holdings (larger bubbles = more titles in dataset)
# Data Analysis: Defining and Evaluating Duplicates

<table>
<thead>
<tr>
<th>Title</th>
<th>ISBN</th>
<th>Publisher</th>
<th>Library Call Number</th>
<th>Year</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDR on his houseboat: ^ the Laroco log, 1924-1926 /</td>
<td>9781438462271</td>
<td>State University of New York Press, (MH)014830896HVD01-Aleph</td>
<td>E807</td>
<td>2016</td>
<td>English</td>
</tr>
<tr>
<td>FDR on his houseboat, 1924-1926 : ^ the Laroco log /</td>
<td>9781438462271</td>
<td>State University of New York Press, (OCoLC)0cr939911243</td>
<td>E807</td>
<td>2016</td>
<td>English</td>
</tr>
<tr>
<td>Failed state : ^ dysfunction and corruption in an American statehouse /</td>
<td>Lachman, Seymour, 9781438465739</td>
<td>State University of New York Press, (OCoLC)53823604</td>
<td>K3474.7</td>
<td>2017</td>
<td>English</td>
</tr>
<tr>
<td>Failed state : ^ dysfunction and corruption in an American statehouse /</td>
<td>Lachman, Seymour, 9781438465739</td>
<td>Excelsior Editions, an Imprint of State University of New York Press, (OCoLC)53823604</td>
<td>K3474.7</td>
<td>2017</td>
<td>English</td>
</tr>
<tr>
<td>Ernest L. Boyer : ^ hope for today's university /</td>
<td>9781438455655</td>
<td>Hardcover, State University of New York Press, (OCoLC)006760525</td>
<td>LA227.4</td>
<td>2015</td>
<td>English</td>
</tr>
<tr>
<td>Ernest L. Boyer : ^ hope for today's universities /</td>
<td>9781438455655</td>
<td>Hardcover, State University of New York Press, (OCoLC-M)906760525</td>
<td>LA227.4</td>
<td>2015</td>
<td>English</td>
</tr>
</tbody>
</table>
# Data Analysis: Normalizing Publishers

<table>
<thead>
<tr>
<th>Original Publisher Name</th>
<th>Normalized Publisher</th>
</tr>
</thead>
<tbody>
<tr>
<td>routledge</td>
<td>routledge, taylor &amp; francis</td>
</tr>
<tr>
<td>routledge, taylor &amp; franc group</td>
<td>routledge, taylor &amp; francis</td>
</tr>
<tr>
<td>crc press, taylor &amp; franc group</td>
<td>routledge, taylor &amp; francis</td>
</tr>
<tr>
<td>routledge, an imprint of the taylor &amp; francis group</td>
<td>routledge, taylor &amp; francis</td>
</tr>
<tr>
<td>taylor &amp; francis</td>
<td>routledge, taylor &amp; francis</td>
</tr>
<tr>
<td>routledge is an imprint of the taylor &amp; francis group, an informa business</td>
<td>routledge, taylor &amp; francis</td>
</tr>
<tr>
<td>crc press/taylor &amp; francis</td>
<td>routledge, taylor &amp; francis</td>
</tr>
<tr>
<td>routledge, taylor &amp; francis</td>
<td>routledge, taylor &amp; francis</td>
</tr>
<tr>
<td>sage</td>
<td>sage</td>
</tr>
<tr>
<td>sage publications</td>
<td>sage</td>
</tr>
<tr>
<td>corwin, a sage company</td>
<td>sage</td>
</tr>
<tr>
<td>sage publications india pvt ltd</td>
<td>sage</td>
</tr>
<tr>
<td>sage editori</td>
<td>sage</td>
</tr>
<tr>
<td>sage/qq press</td>
<td>sage</td>
</tr>
<tr>
<td>springer</td>
<td>springer</td>
</tr>
<tr>
<td>springer vs</td>
<td>springer</td>
</tr>
<tr>
<td>springer verlag</td>
<td>springer</td>
</tr>
<tr>
<td>this palgrave macmillan imprint is published by springer nature</td>
<td>springer</td>
</tr>
<tr>
<td>springer international publishing</td>
<td>springer</td>
</tr>
<tr>
<td>springer publishing company</td>
<td>springer</td>
</tr>
<tr>
<td>palgrave macmillan imprint is published by springer nature</td>
<td>springer</td>
</tr>
<tr>
<td>springer gabler</td>
<td>springer</td>
</tr>
<tr>
<td>springer nature</td>
<td>springer</td>
</tr>
<tr>
<td>university of nebraska press</td>
<td>university of nebraska press</td>
</tr>
<tr>
<td>potomac books, an imprint of the university of nebraska press</td>
<td>university of nebraska press</td>
</tr>
<tr>
<td>john wiley &amp; sons, inc.</td>
<td>university of nebraska press</td>
</tr>
<tr>
<td>wiley</td>
<td>wiley</td>
</tr>
<tr>
<td>wiley blackwell</td>
<td>wiley</td>
</tr>
<tr>
<td>wiley-vch</td>
<td>wiley</td>
</tr>
<tr>
<td>wiley-vch verlag gmbh &amp; co. kgaa</td>
<td>wiley</td>
</tr>
<tr>
<td>jossey-bass, a wiley brand</td>
<td>wiley</td>
</tr>
</tbody>
</table>
Data Analysis: Current Tradeoffs

Control over definitions & questions

Ease of analysis & software use
Data Analysis: Takeaways

- Data normalization and analysis are tricky
  - Develop methods to keep track of your work
  - Document your code
  - Determine when it is ‘good enough’
- Diverse skill sets and tools needed to interrogate this data effectively
- Copy analysis exercise demonstrated that 4 or fewer copies meets 97.8% of the distinct titles needed for Borrow Direct.
Part 3. Application

Library of Congress, Prints & Photographs Division, Detroit Publishing Company Collection
In Progress...

DEI Definitions for Collaborative Collections

Coordinated Print Approval Plan
Potential...

- Expand the data in Gold Rush to enable:
  - Shared Print Retention Commitment Management & Analysis
  - Data source to identify commonly held and rare holdings within IPLC
  - Identifying digitized collections across IPLC to inform digitization workflows
- Use the data to help inform strategic acquisition of e-books at local institutions.
Thanks!

Any questions?

Galadriel Chilton
galadriel.chilton@yale.edu

Jeanette Norris
jeanette.norris@yale.edu

Sarah Tudesco
sarah.tudesco@yale.edu