Rare Book MARC Data as University Archives: New Directions for Special Collections Collaboration

Jennifer King, Leah Richardson, Dolsy Smith

THE GEORGE WASHINGTON UNIVERSITY — WASHINGTON, DC

GW | Libraries

Project Goals

In Fall 2017, a group of four librarians embarked on a project to learn more about GW library's rare book holdings using MARC data. The goals for this project were to gain a more complete understanding of the strengths, coverage, and gaps of the rare book collection and to strengthen our nascent data analysis skills. We plan to use the results from our data analysis to guide collection development and instruction.

Research Questions

How many rare books do we have?
What is the range of publication dates?
Where are the places of publication?
Who were the printers and/or publishers?
What languages are represented?
What are the subjects covered?

Research Methods

Using a dataset that captured specific MARC fields within the bibliographic records that held the potential answers to our research questions we used various methods of data clean-up and organization to analyze and visualize the data.

Mapping Questions onto MARC

What is the range of publication dates?
Where are the places of publication?
Who were the printers?
Who were the publishers?
What languages are represented?
What are the subjects?
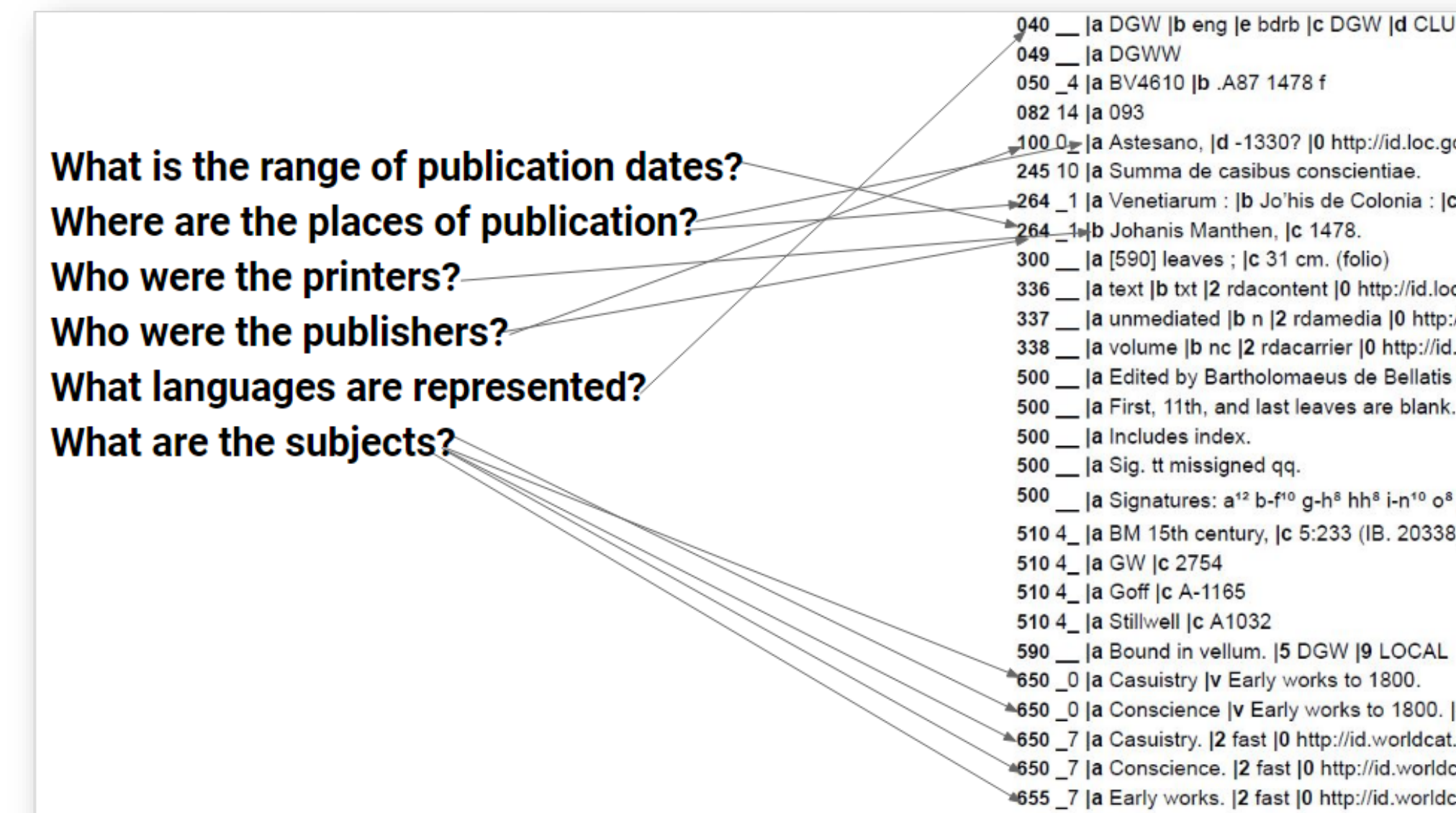
73,000+ records in initial data extractions

Issues encountered:
- **Difficult to capture relationships between fields and subfields**
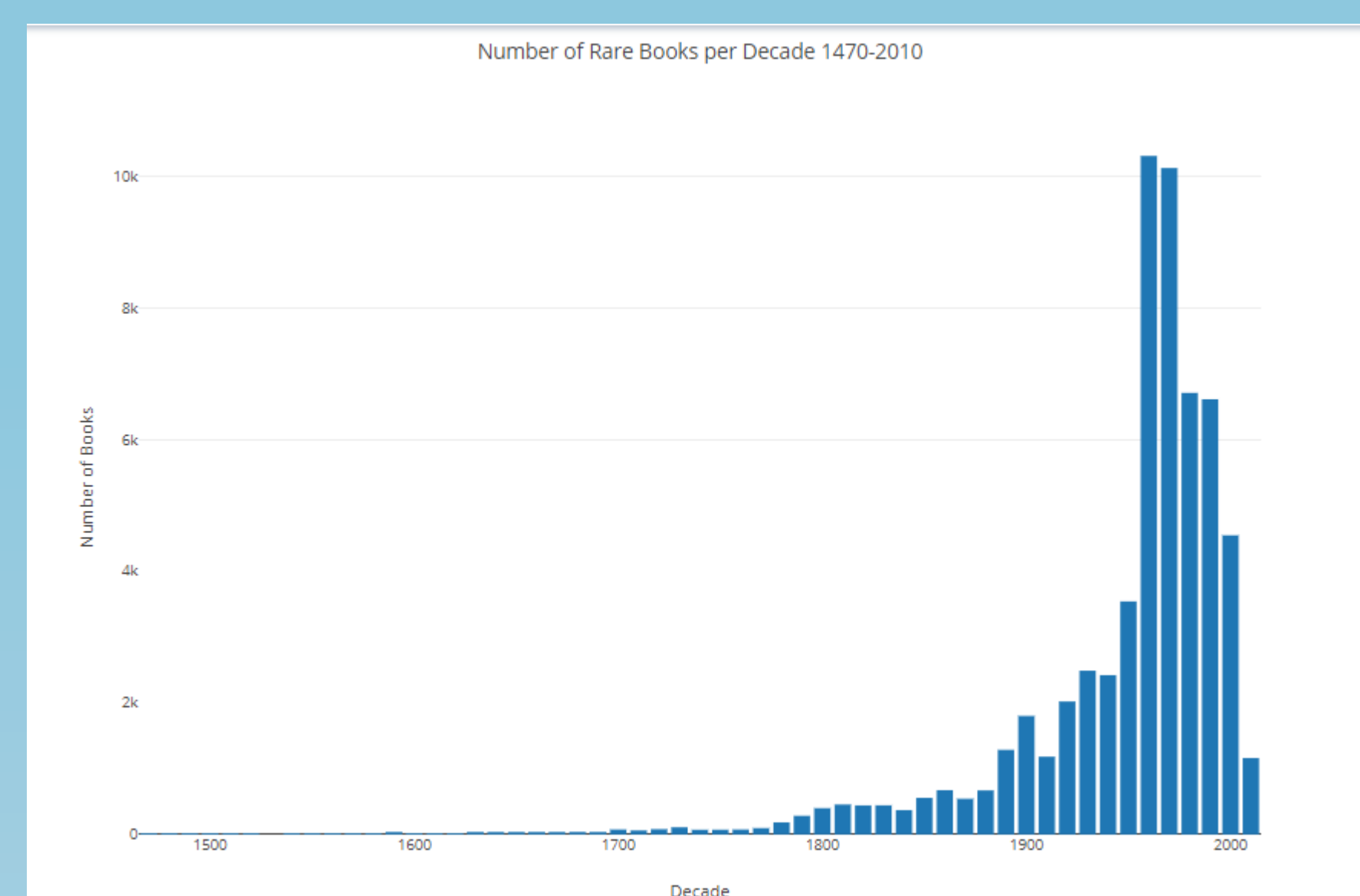- **Inconsistent cataloging practices**
- **Duplicate MARC fields**

Process

✓ **Data-clean up:**
   ✓ **Replaced MARC indicators with plain language**
   ✓ **Normalized dates**
   ✓ **Excluded non-book records, MS#**
   ✓ **Removed punctuation and Unicode characters that didn't translate**

Initial Findings

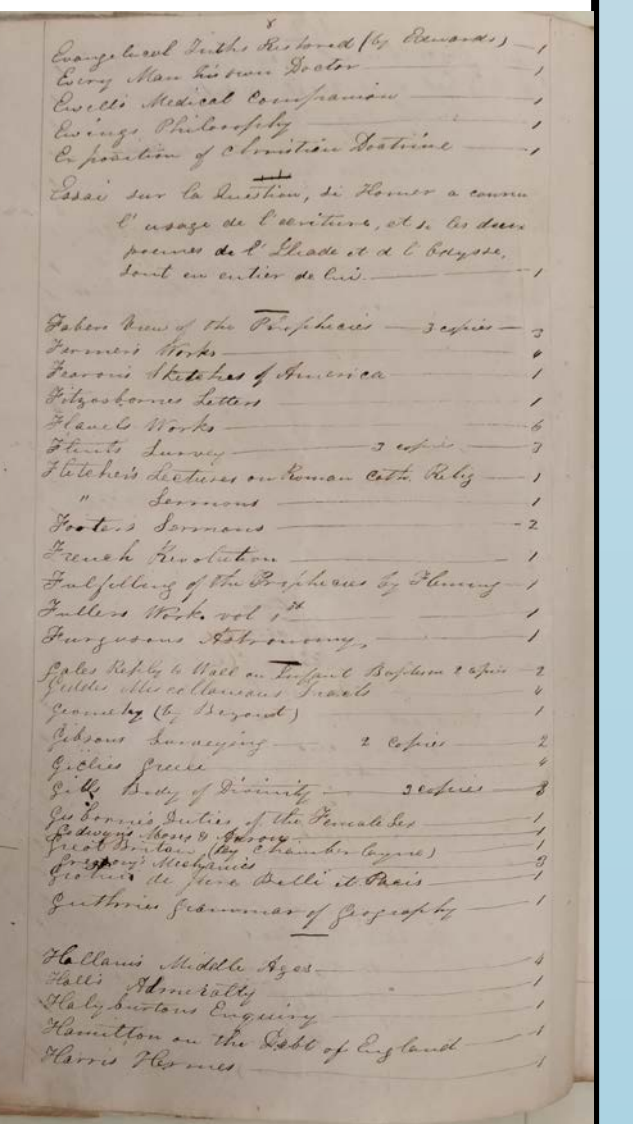Number of Rare Books per Decade 1470-2010

Number of books:
60,238

Date coverage:
1470s to 2017
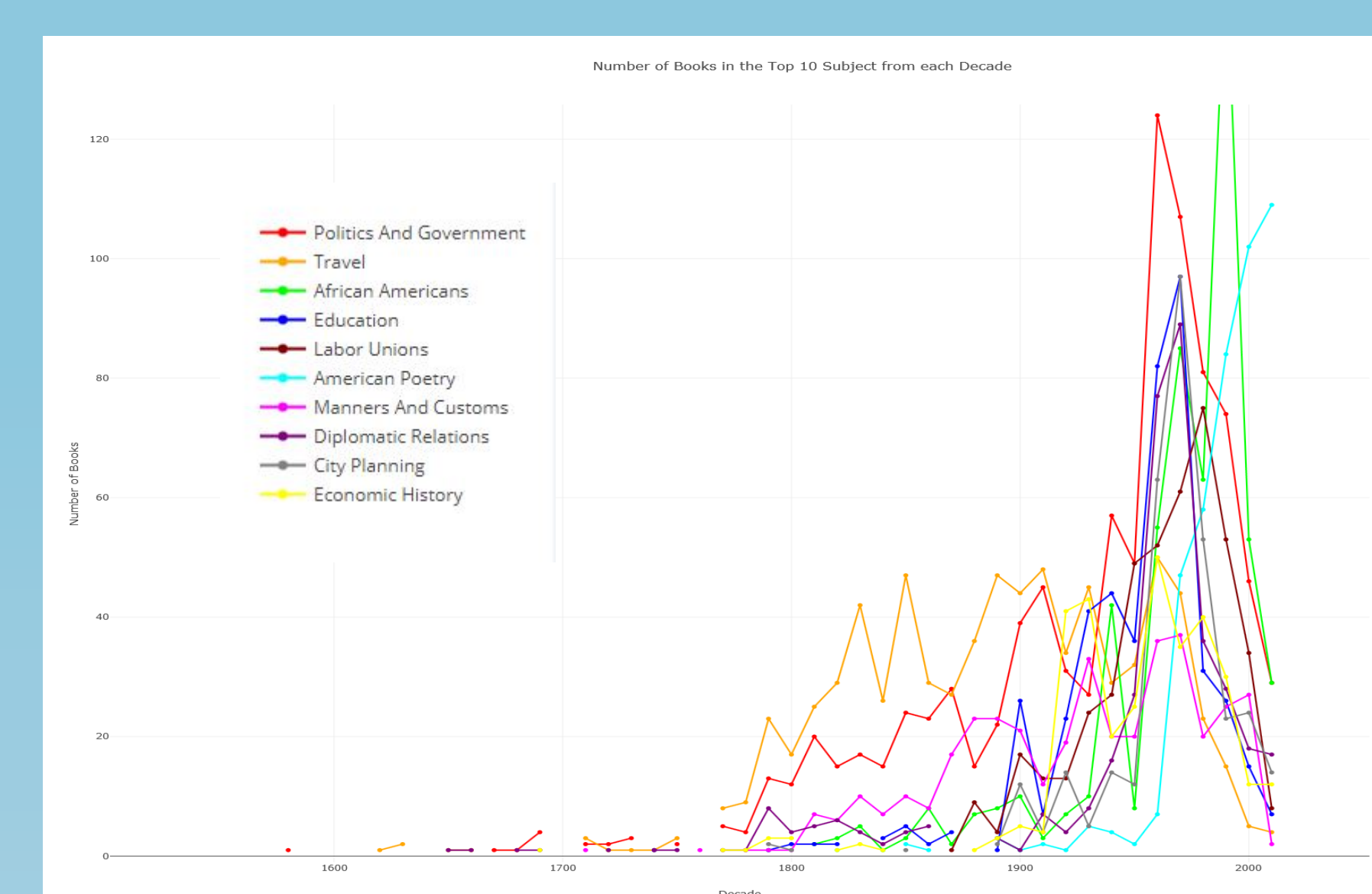
University Archives in MARC

We were able to locate titles that remain from the original library collection of the Columbian College. We also located University Archives that were cataloged as books.

A page from the original inventory of the Columbian College library's collection, written in 1825.

500 __ |a From the Columbian College Library: Schedule A, 1825: No. 216

The note in the 500 field indicating books listed in the above 1825 inventory.

Outcomes

- Greater understanding of MARC data
- Collaborative approach to learning to write code
- Traced the intellectual work of librarians, archivists, and curators since the university was founded

Next Steps

- Explore data in 5xx and 7xx fields
- Use data to support strategic collection development
- Disseminate data to support discovery and collections as data research