

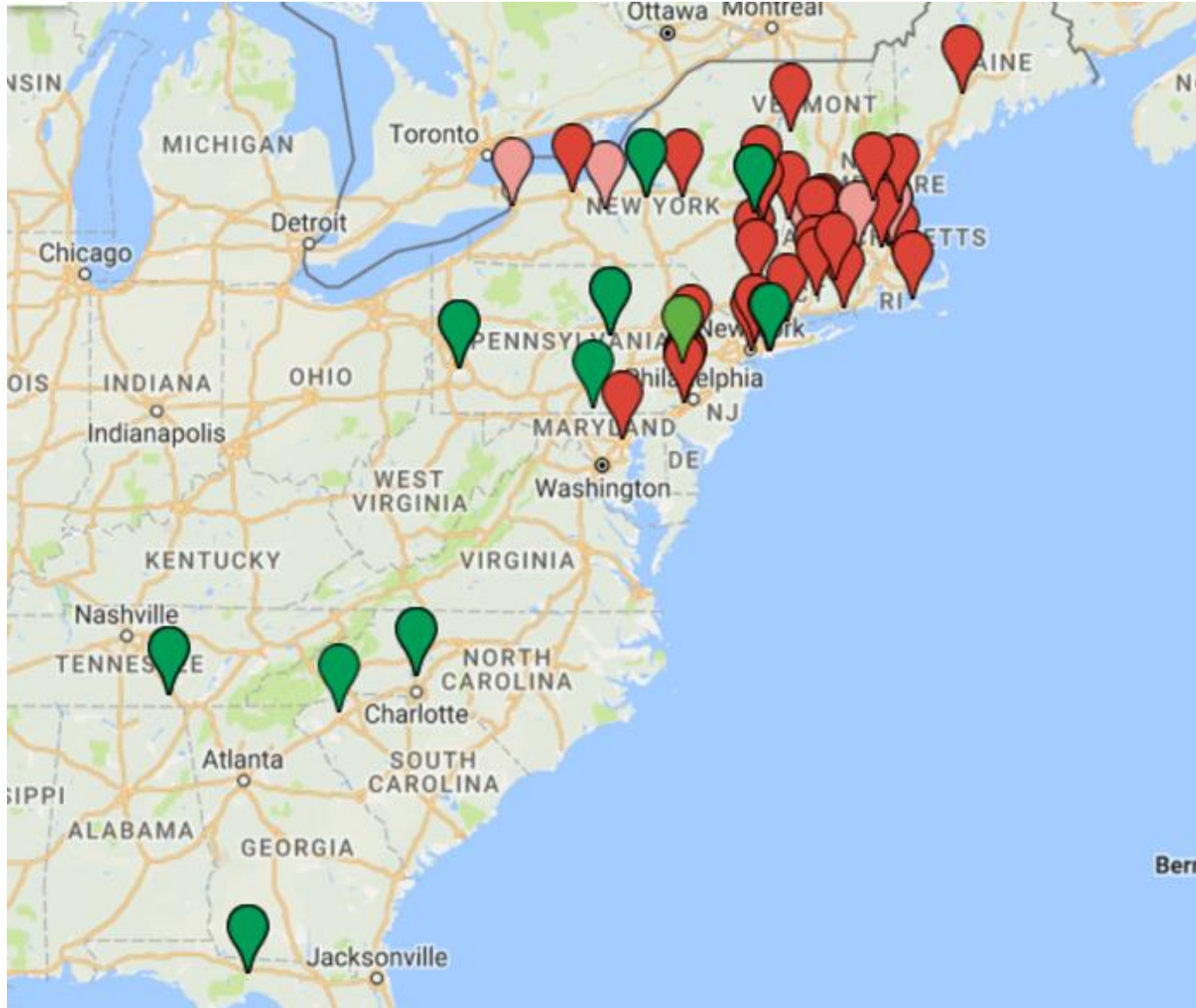
# From Descriptive Analysis to Prescriptive Analytics: Visualization and Validation in Shared Print

Charleston Conference  
November 8, 2017

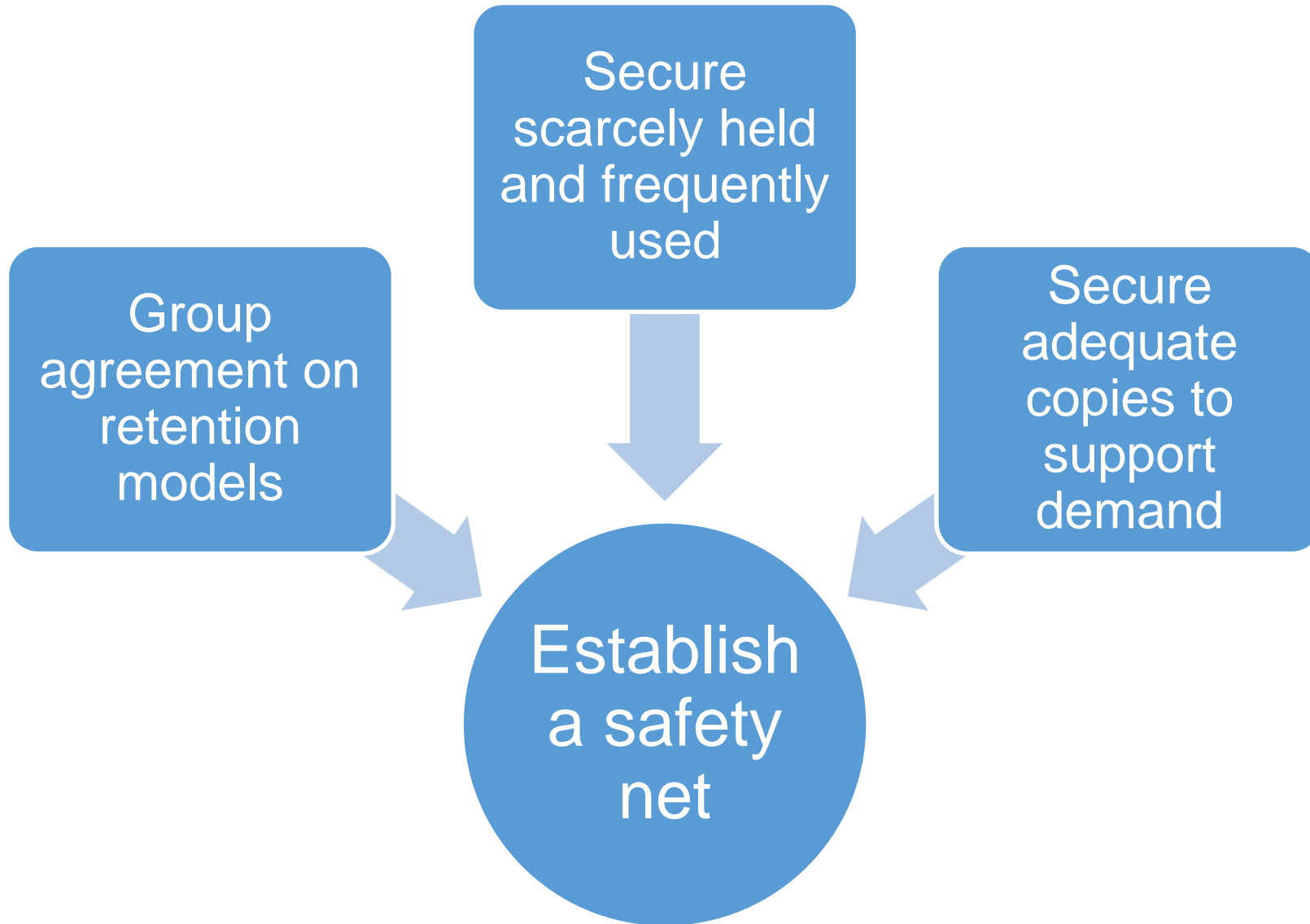
Susan Stearns, Project Director  
Mei Mendez, Project Manager

# The EAST mission

Securing the scholarly  
print record in support  
of teaching, learning and  
research



- 60 members from Maine to Florida
- 52 Retention Partners
- 8 Supporting Partners
- Diversity in collection size, research vs. teaching focus and collection philosophy



**EAST**

EASTERN ACADEMIC  
SCHOLARS' TRUST

# Collection Analysis is Key

- For EAST: to identify titles to be committed to retention
- For individual libraries: to provide insight into de-selection and overall collection management
- Partnering with OCLC Sustainable Collection Services (SCS)<sup>®</sup>
- Analysis for original 40 Cohort 1 libraries completed in 2016
- Now undertaking analysis with 12 Cohort 2 libraries
- Scope of analysis is circulating print monographs only

EAST 2

EAST 1

12 libraries

8.97M holdings

62% held by 1  
library

40 libraries

16.5M holdings

20% with >20  
aggregate uses

39% in HathiTrust

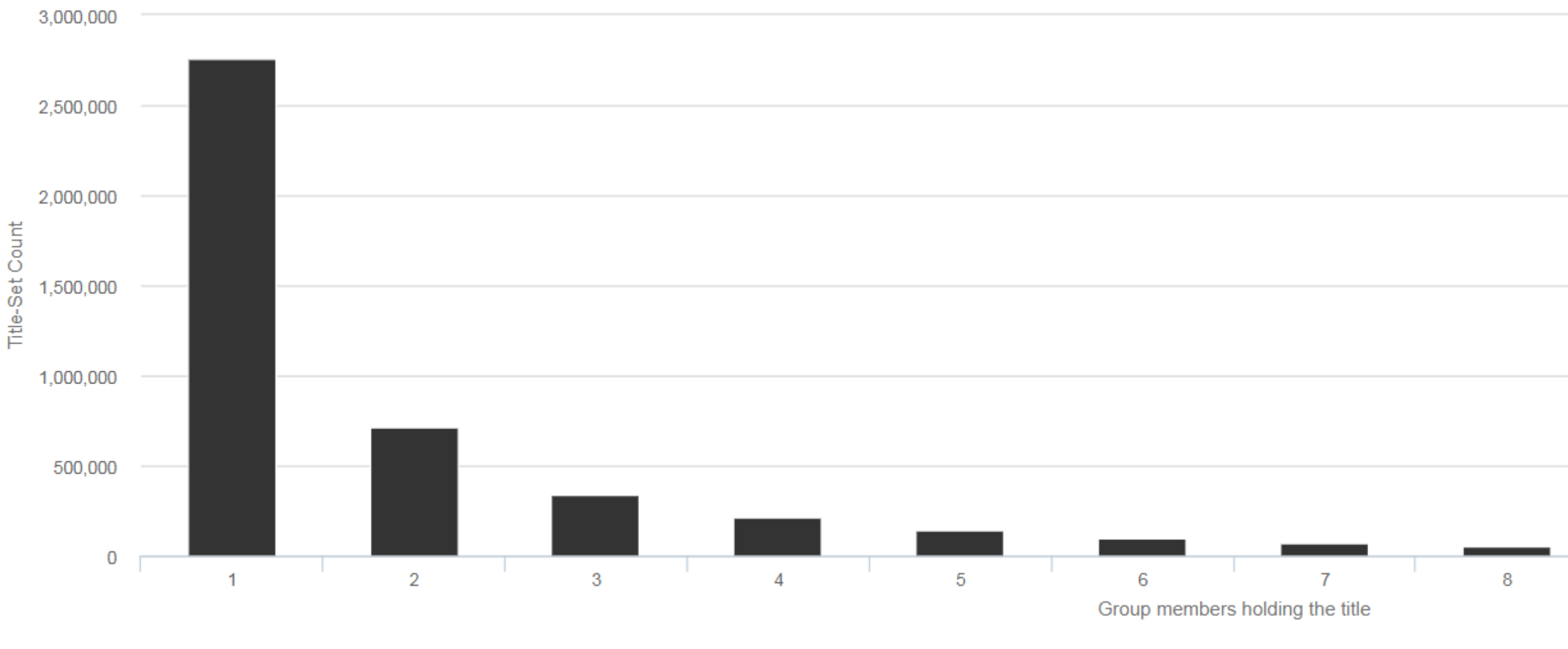
Over 2.2M  
holdings  
unique to  
EAST 2

A Comparison of the  
EAST Cohorts 1 and 2  
Collective Collections



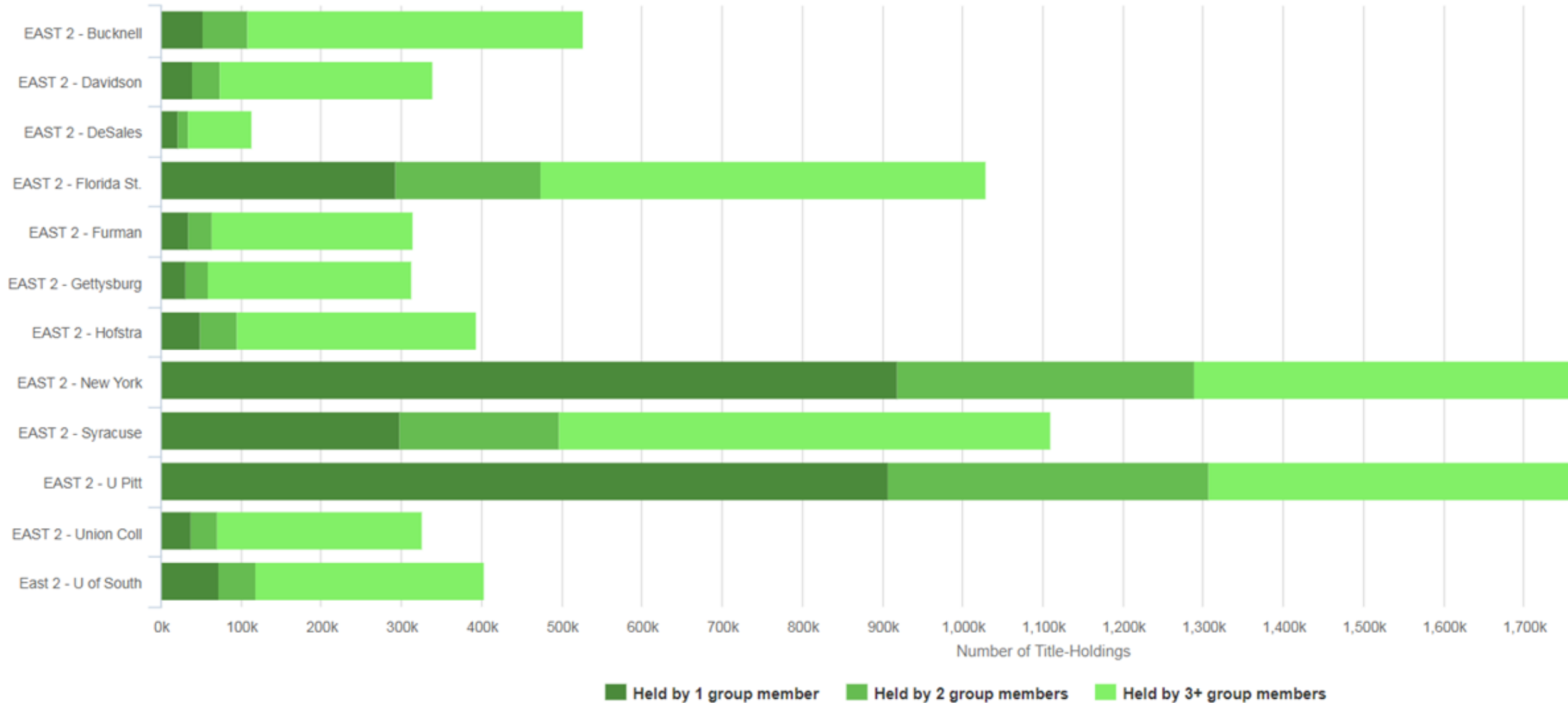
EAST

EASTERN ACADEMIC  
SCHOLARS' TRUST

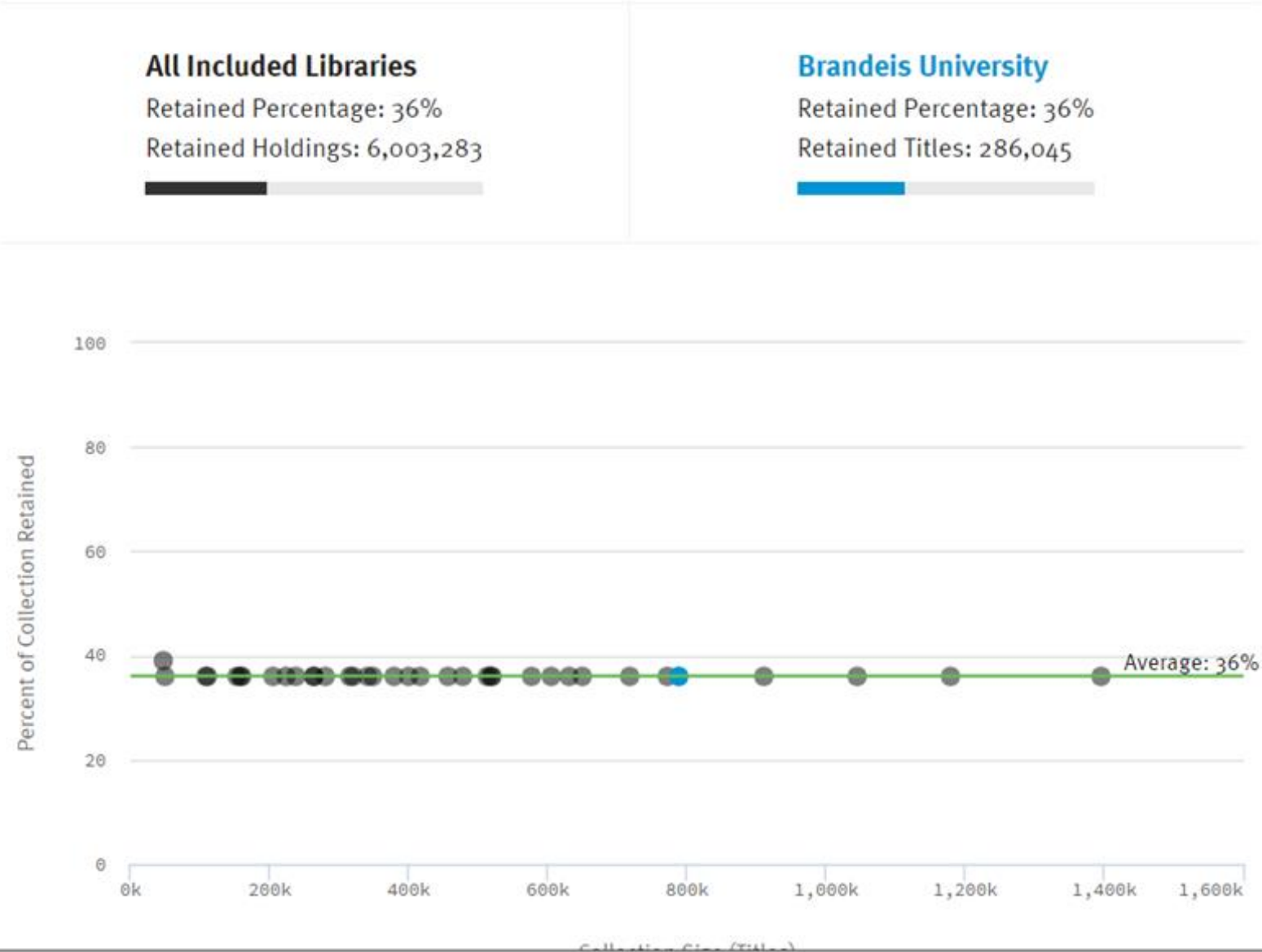


## Uniqueness and overlap in the EAST collective collection – Cohort 1

# Overlap by library – Cohort 2







Criteria
LIBRARIES
Includes 36 of 36 Libraries
RETAIN ALL
EAST holdings FEWER THAN 5
US holdings FEWER THAN 40 (ANY EDITION)
Regional Large Academic Libraries holdings FEWER THAN 5 (ANY EDITION)

# Validation Sample Study

## What is the EAST Validation Sample Study?

A small sample study to determine an “availability metric” that provides a rough estimate of the percentage of library holdings that are likely to be available (are accounted for).

## Why validate?

Trust.

# Validation Sample Study

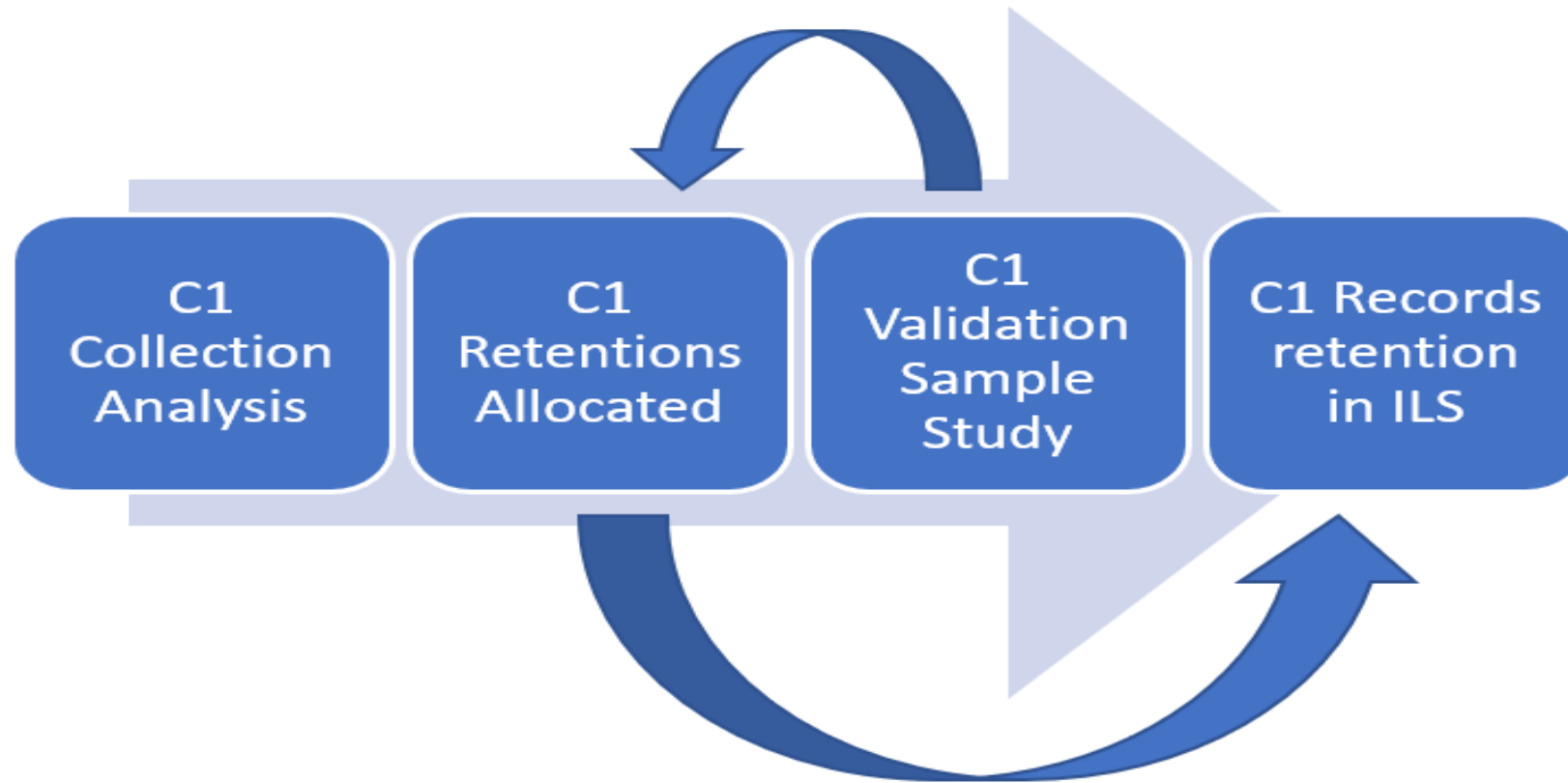
## What does the validation look like?

Worked with statistical consultant Grant Ritter and EAST Data Librarian Sara Amato to create an availability and condition matrix.

Determined that 6,000 items per library would be enough for a statistically valid sample.

SCS generated lists of 6,000 titles randomly selected from the in-scope records that were sent to them.

# Validation Sample Study



**EAST**

EASTERN ACADEMIC  
SCHOLARS' TRUST

# Validation Sample Study Results

240,000

- Titles checked in the Validation Sample Study

92,575

- Of titles checked received retention commitments

70,000

- Of the full set of retained titles were potentially at risk of being in poor condition or not being on the shelf

Statistician reviewed titles that had a greater than 7.5% chance of being missing or in poor condition AND that EAST had surplus copies that were not already under retention

Libraries with additional copies of these items not already under retention received requests to retain them.

# Validation Sample Study Results: Missing Items

- Calculation of percentage of monographs missing for each participating EAST library
- Summary Statistics and Distribution on Missing Monographs at 40 participating EAST libraries (all estimates accurate to 1%)

	mean	std	Min	5 <sup>th</sup> pctl	25 <sup>th</sup> pctl	50 <sup>th</sup> pctl	75 <sup>th</sup> pctl	95 <sup>th</sup> pctl	max
Rate	3.0%	2.1%	.3%	.6%	1.3%	2.8%	4.2%	7.4%	9.7%



**EAST**

EASTERN ACADEMIC  
SCHOLARS' TRUST

# Validation Sample Study Results: Condition

- Calculation of percentage of monographs in poor, acceptable, or excellent condition for each EAST library
- Summary Statistics and Distribution on Condition of Monographs at 40 participating EAST libraries (all estimates accurate to 2%):

	mean	std	Min	5 <sup>th</sup> pctl	25 <sup>th</sup> pctl	50 <sup>th</sup> pctl	75 <sup>th</sup> pctl	95 <sup>th</sup> pctl	max
Poor	10.7%	9.8%	1.1%	1.3%	3.0%	8.2%	15.2%	29.3%	44.2%
Acceptable	55.7%	17.3%	34.8%	34.8%	45.5%	54.1%	65.9%	88.1%	91.1%
Excellent	36.6%	17.7%	6.8%	7.8%	20.7%	33.6%	43.3%	70.5%	78.8%



**EAST**

EASTERN ACADEMIC  
SCHOLARS' TRUST

# Validation Sample Study Follow Up: Missing Items

- Missing rates at most participating libraries were very low. Only two libraries had rates greater than 7.4%, and no library had a rate greater than 10%
- Factors significantly affecting likelihood of being missing were small
- No factor appeared important enough to suggest a modification to the current EAST retention plan



# Validation Sample Study Follow up: Condition

- Poor condition rates at some participating libraries were large enough to merit attention. A quarter of the libraries had rates greater than 15% and two libraries had rates greater than 29%
- Certain factors affecting likelihood of being in poor condition are large enough to recommend changes to retention plans
- Retention plan might keep extra copies of older monographs, monographs frequently circulated, and monographs in subject area of painting

# Validation Sample Study, Deeper Dive into Data

## **Missingness:**

- If the missing volume is one of several retained copies, the problem is likely to be minimal. On average, the mean probability of two volumes of the same title being missing would be negligible.
- If the missing copy is the only copy of the monograph within EAST at allocation, then there is nothing to be done to mitigate the issue.
- However, 7,882 titles had unallocated copies at other EAST libraries. It is requested that an additional copy of these titles be retained.

## **Condition:**

72,757 titles are more than 50% likely to be in poor condition and have unallocated copies at other EAST libraries. It is requested that an additional copy of these titles be retained.



**EAST**

EASTERN ACADEMIC  
SCHOLARS' TRUST

# Validation Sample Study, Cohort 2

- Professor Ritter determined that the same methodology and sample size could be applied to the Cohort 2 collections.
- EAST Cohort 2 libraries are conducting the validation sample study and expect to be completed in mid-December, when the data will be provided to Professor Ritter for analysis.
- The results of his analysis will be incorporated in the Cohort 2 retention modeling and used to select the final retention model.

# Retention of Serials and Journals

- 29 of the 60 EAST libraries are Serials/Journals Retention Partners
- In late 2016 began to investigate options for collection analysis
- No decision support tools similar to GreenGlass® available today
- Contracted with the Center for Research Libraries for a series of reports on overlap and uniqueness as well as depth of holdings based on library contributed metadata
- No ability to compare outside of the EAST data set

# Totals by Library

*EAST Holdings (total holdings for these titles in EAST)*

*Library Title Sets (how many titles in my library)*

	Library Title Sets	EAST Holdings
Amherst	6,551	21,188
Bard	902	8,581
Boston College	28,307	84,231
Boston University	27,463	78,495
Brandeis	13,980	50,834
Connecticut College	9,501	37,214
Elms	141	1,038
Fairfield	1,285	9,569
Five Colleges	17,349	50,229
Hamilton	5,592	30,191
Lafayette	5,729	32,248
Loyola	2,139	14,633
Mount Holyoke	6,379	31,386
Siena	1,729	13,247
Smith	7,087	27,660
Swarthmore	2,733	19,623
Trinity	4,684	28,228
UMass Amherst	33,596	80,846
Union	4,227	26,460
Williams	10,258	45,810
Woods Hole	6,482	17,325

EAST Holdings



Jstor

- ☐ (All)
- ☒ Not in JSTOR
- ☐ In JSTOR

Cohort 1 serials/journals titles as presented in Tableau for analysis: 108,863 title sets representing 195,793 holdings

# Totals by Library

*EAST Holdings (total holdings for these titles in EAST)*

*Library Title Sets (how many titles in my library)*

	Library Title Sets	EAST Holdings
Amherst	1,187	5,693
Bard	118	589
Boston College	4,535	21,559
Boston University	4,423	20,972
Brandeis	2,755	13,216
Connecticut College	1,592	7,724
Elms	44	216
Fairfield	372	1,795
Five Colleges	2,503	11,888
Hamilton	1,253	6,171
Lafayette	1,564	7,705
Loyola	598	2,883
Mount Holyoke	1,406	6,879
Siena	374	1,847
Smith	1,247	6,044
Swarthmore	832	4,145
Trinity	1,217	6,012
UMass Amherst	3,993	18,886
Union	1,219	5,998
Williams	2,406	11,606
Woods Hole	767	3,681

EAST Holdings








Jstor

- ☐ (All)
- ☒ Not in JSTOR
- ☐ In JSTOR

Totals, by library, of those titles held by 4-6 of the retention partners: 7,348 titles representing 34,405 holdings

# Understanding your local collection – an EAST example

SCS has been delivering a set of Key Metrics to every client library since the launch of GreenGlass in 2012. For each measure, we offer a green hyperlink to the corresponding percentage of your overall collection. The accumulation of these measures provides an interesting opportunity for you to compare your own key measures with most of the libraries we have worked. These numbers are based on more than 220 US academic libraries and over 98 million holdings. They include research libraries, state universities, 4-year colleges, and community colleges.

METRIC	FILTERED ITEMS	% OF YOUR FILTERED ITEMS   SCS AVERAGE	COMPARED TO THE RANGES AND
Zero recorded uses	<a href="#">511,471</a>	42% •   42% ▣	
Greater than 3 recorded uses	<a href="#">305,404</a>	25% •   25% ▣	
Publications more than 10 years old	<a href="#">1,049,533</a>	86% •   88% ▣	
More than 100 US holdings - same edition	<a href="#">677,230</a>	56% •   74% ▣	
Fewer than five US holdings - any edition	<a href="#">18,031</a>	1% •   2% ▣	
Unique in New York - any edition	<a href="#">52,782</a>	4%   NA	NA



### EAST overlap

no restriction ▼

### Aggregate uses

no restriction ▼

### Ephemera (?)

No restriction ▼

### Allocated for retention

No restriction ▼

### HATHITRUST

Include items that meet the following criteria:

No restriction ▼

### COMBINE QUERIES

Combine criteria above with the following queries:

- ☒ Don't combine with other queries
- ☐ Combine results with other queries
- ☐ Intersect results with other queries



MATCHED  
26,071

CO  
863

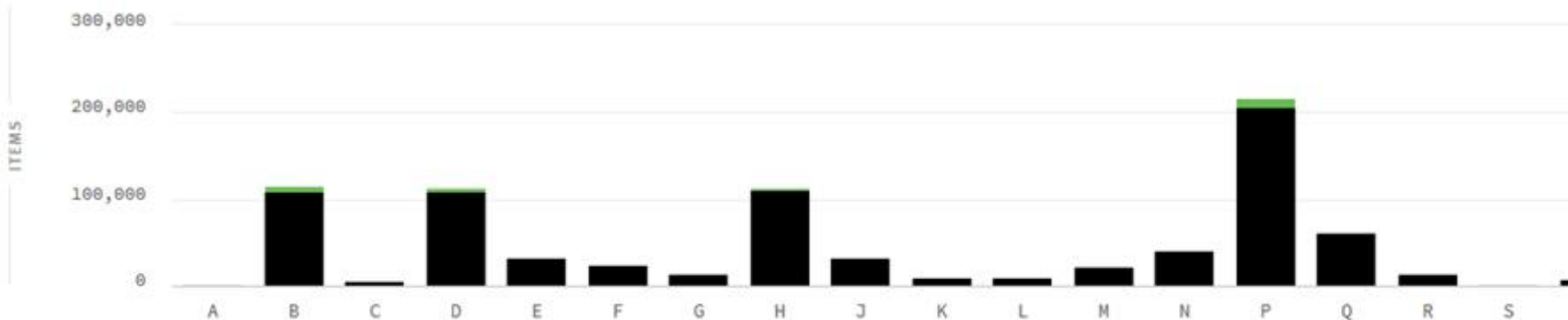
Recorded Uses

< 5

Pub Year

before 1900

UNMATCHED 96.98% | 837,691    MATCHED 3.01% | 26,071



SUBJECT

PERCENT MATCHED

MATCHED

Total

3.02 %

26

A General Works

3.79 %

Titles published before 1900 with minimal usage by LC class



[New Query](#)[← QUERY LIST](#)[EXPORT XLSX](#)

## EAST overlap

no restriction ▾

## Aggregate uses

no restriction ▾

## Ephemera (?)

No restriction ▾

## Allocated for retention

No restriction ▾

## HATHITRUST

Include items that meet the following criteria:

No restriction ▾

HIDE EDITOR

■ MATCHED  
307,380■ COLLECTION  
863,762

Edition Matching

same edition

Boston Library Consortium (BLC) Holdings

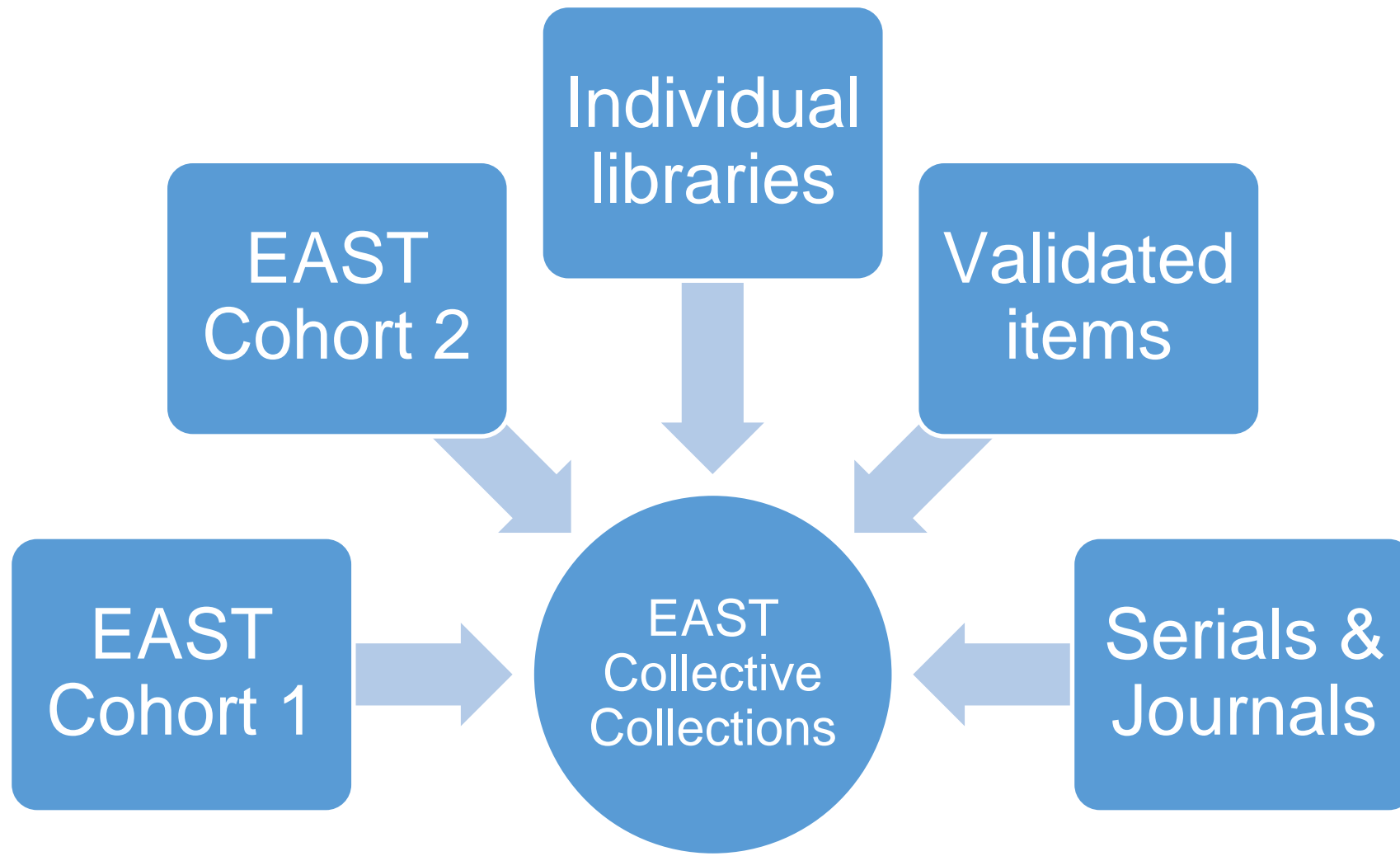
&lt; 2

■ UNMATCHED 64.41% | 556,382    ■ MATCHED 35.58% | 307,380



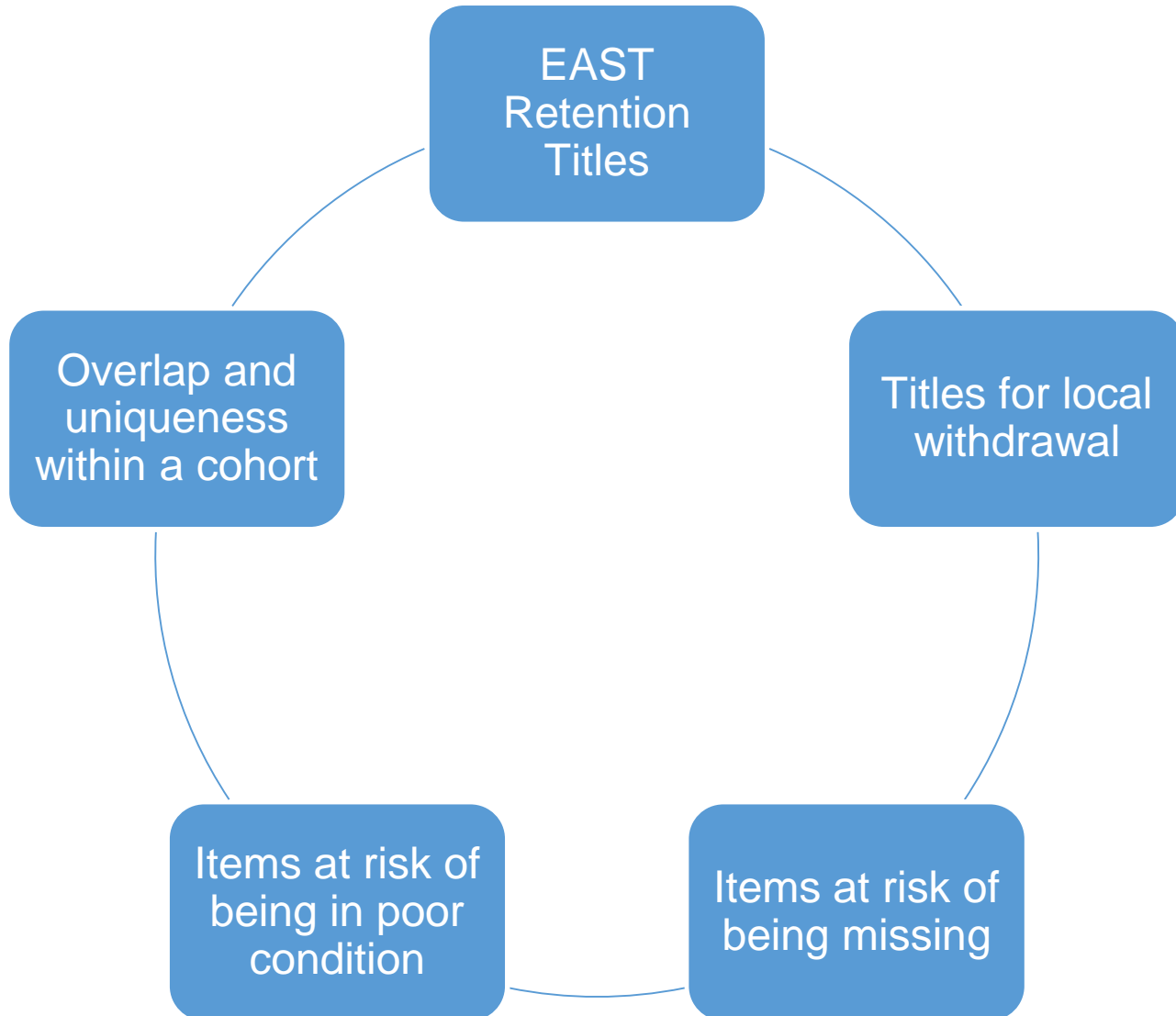
Comparing the local collection to the consortial collective data set

# Descriptive Analysis and Visualizations



Descriptive analyses provide measures of composition and distribution of data

# Prescriptive Analytics



Prescriptive analytics assist you in finding the best course of action in a given situation.



**EAST**

EASTERN ACADEMIC  
SCHOLARS' TRUST

# Thank you.

[ssstearns@blc.org](mailto:sstearns@blc.org)

[mmendez@blc.org](mailto:mmendez@blc.org)



Education, Owl  
graphics by Freepik,  
Flaticon, CC BY 3.0 (via  
Logo Maker)



**EAST**

EASTERN ACADEMIC  
SCHOLARS' TRUST