

Analyzing Collections for Shared Print

October 6, 2021
Susan Stearns & Sara Amato



PARTNERSHIP FOR SHARED BOOK COLLECTIONS

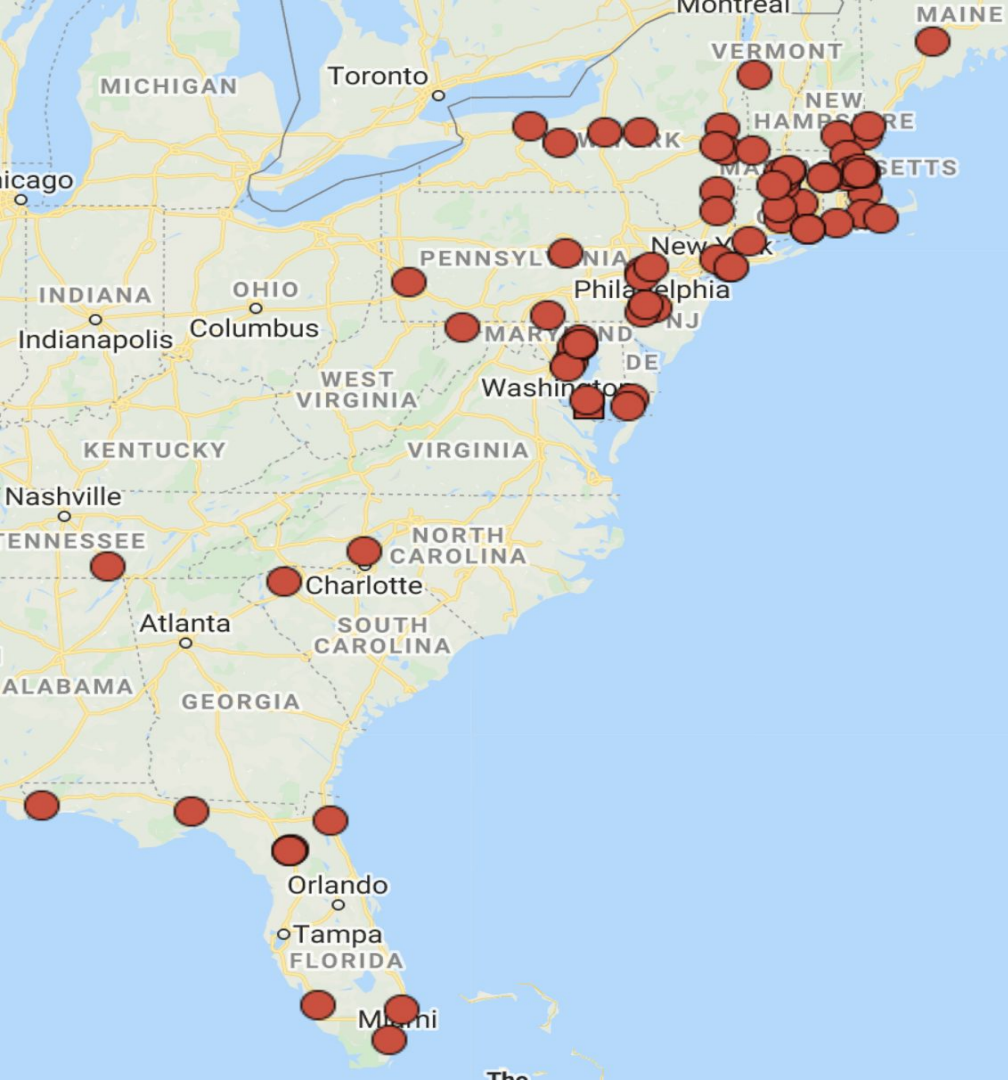


EAST

EASTERN ACADEMIC
SCHOLARS' TRUST

A Brief History of EAST

- Grew out of a Mellon-funded planning grant in 2013/4
- Formally launched in July, 2015 with grant funding
- 40 original members in Cohort 1, 12 added in Cohort 2 in 2016/7, Florida Cohort in 2019/20, & USMAI libraries in 2020/21
- Almost 10 million monograph titles committed to retention
- Launched journals work in late 2017 and have committed over 28,000 titles
- Member of the Rosemont Shared Print Alliance (Journals)
- Founding member of the Partnership for Shared Book Collections



- 82 members from Maine to Florida
- 77 Retention Partners
- 5 Supporting Partners



EAST

EASTERN ACADEMIC
SCHOLARS' TRUST

Land and Labor Acknowledgement

EAST wishes to acknowledge that our member libraries occupy ancestral land from many of the First American Nations including those listed [here](#).

We also acknowledge the hundreds of thousands of enslaved people from Africa and the Americas who preceded us and whose labor was taken to build and sustain many of the academic and research institutions which are members of EAST.

This acknowledgement commits us to work to dismantle these systems of oppression and to work for peace and reconciliation as caretakers of the lands we occupy.

EAST - History with Gold Rush

- Need for a Retentions Database in 2016/7 as WorldCat registration service not yet complete
- Initial consideration of Gold Rush rejected, but provided feedback
- 2019 began using Gold Rush both as a retentions database and to offer lightweight collection analysis



PARTNERSHIP FOR SHARED BOOK COLLECTIONS



EAST

EASTERN ACADEMIC
SCHOLARS' TRUST

EAST Data and Uses

Gold Rush as a repository for EAST retention commitments

- All EAST member libraries deposit MARC records of their retained material
- Data is updated yearly

This allows us to:

- Offer a Retentions Database that is an alternative to WorldCat
- Do Lightweight Collection Analysis for New Members



PARTNERSHIP FOR SHARED BOOK COLLECTIONS



EAST

EASTERN ACADEMIC
SCHOLARS' TRUST

The EAST Retention Model - Cohorts 1 and 2

Retain ALL unique or scarcely held titles: Scarcely was defined as fewer than 40 U.S. holdings (any edition) in WorldCat, publication year before 2011 and not published by a publisher on the 'ephemera' list.

Retain sufficient copies of widely used titles: Retain UP TO 5 copies (same edition) of works that had more than 20 aggregate uses across the libraries, publication year before 2011 and not published by a publisher on the 'ephemera' list.

Retain extra copies of materials likely to be in poor condition: Retain UP TO 3 copies of monographs published before 1900.

Retain ONE copy of all other monographs, publication year before 2011 and not published by a publisher on the 'ephemera' list.

The EAST Model Light - Developed for Gold Rush

Retain ALL uniquely held titles: Titles *unique to EAST*, publication year before 2011 and not published by a publisher on the 'ephemera' list.

Retain sufficient copies of widely used titles: Retain UP TO 5 copies of titles with *EAST retentions of more than 2 but fewer than 5*, publication year before 2011 and not published by a publisher on the 'ephemera' list. Also consider *High Circ Low EAST*

Retain extra copies of materials likely to be in poor condition: Retain UP TO 3 copies of monographs published before 1900.

EAST - Gold Rush Analyses

To date have worked with:

Lehigh University - academic library

USMAI - 13 libraries allied to the University of Maryland system

Tufts University (new museum school added)

Soon to begin work with

State Library of Massachusetts - EAST's first state library

Temple University - academic library



PARTNERSHIP FOR SHARED BOOK COLLECTIONS



EAST

EASTERN ACADEMIC
SCHOLARS' TRUST

EAST - Projects and Process

Search interface for EAST (Eastern Academic Scholars' Trust) database.

Search filters: EAST (dropdown), All Fields (dropdown); Search... (text input), Search (button), Select One (dropdown), Select Multiple (button), Site Compare (button).

Location lookup (link)

Load saved search (button)

This collection contains 6,922,344 unique Match Keys. [View Indexing History](#) (link)

*Tufts University SMFA (text input)


Step 1 - Compare library holdings to all EAST - based on matchkey


ballads_and_poems_relating_to_the_burgoyne_campaign_____1893_____j
munsa_____stone__william_l_____p


EAST - Projects and Process

Step 2 - Check FORMAT - exclude all formats except Book

Note that Gov docs identified only by 086

Limit your search... **Apply** 

Format 

<input type="checkbox"/>		
<input type="checkbox"/>	Book X	14496
<input type="checkbox"/>	Electronic X	15
<input type="checkbox"/>	Gov. Doc. X	15
<input type="checkbox"/>	Video X	14
<input type="checkbox"/>	SoundRecording X	3
<input type="checkbox"/>	Unknown X	3
<input type="checkbox"/>	Journal/Serial X	2



PARTNERSHIP FOR SHARED BOOK COLLECTIONS



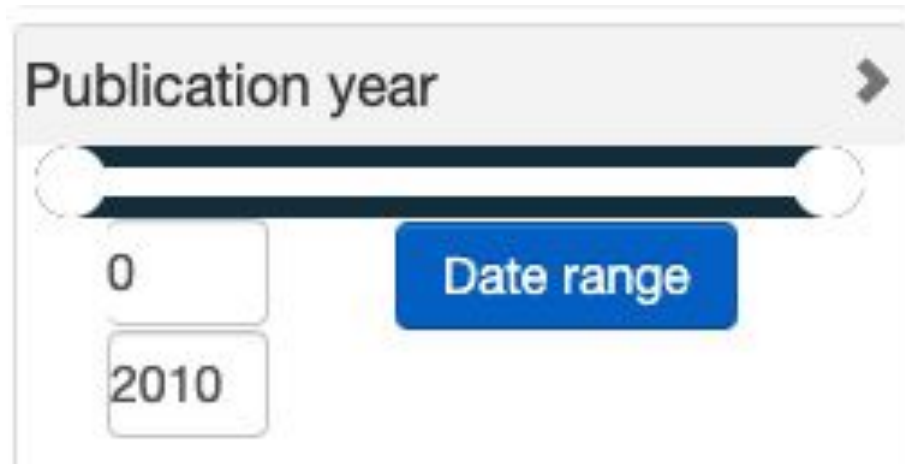
EAST

EASTERN ACADEMIC
SCHOLARS' TRUST

EAST - Projects and Process

Step 3 - Limit dates pre 2011 (pub year 0-2010)

Note that “date” comes from 008 2nd date, 264c, 260c in that order.*



The image shows a web interface for selecting a date range. At the top, there is a label "Publication year" with a right-pointing arrow. Below this is a horizontal slider with two white circular handles. Under the left handle is a text input field containing the number "0". Under the right handle is a text input field containing the year "2010". To the right of these input fields is a blue rectangular button with the white text "Date range".

Overlap in Gold Rush - Problem Example

These 3 were identified as the same work in Gold Rush, though they are distinct OCNs and thus not the same work in OCLC's GreenGlass (would be grouped together at the 'work' level?)

la_pest_____1947_____gallia_____camus__albert_____
_____p

<https://www.worldcat.org/title/peste/oclc/342736> (9)

Publisher: [Paris] : Gallimard, ©1947.

<https://www.worldcat.org/title/peste/oclc/4383621> (1)

Publisher: [Paris] : Gallimard, [1978], ©1947.

Series: Collection Folio, 42.

<https://www.worldcat.org/title/peste/oclc/31529267> (1)

Publisher: [Paris] : Gallimard, 1992, ©1947.

Series: Collection Folio, 42.

Gold Rush looks at 008 first to determine date, and will take the 2nd date if valid. It does NOT look for the value in the [008/6](#) to determine if that second date is a copyright date (or a reprint date)

r - Reprint/reissue date and original date
s - Single known date/probable date
t - Publication date and copyright date

EAST - Projects and Process

Those 3 steps give us a list of “Unique” and Overlap, opportunities to ‘top up’

Total	23,127
Unique	8,870
Common	14,257
Held at 2 libraries	6,947
Held at 3 libraries	1,128
Held at 4 or more libraries	6,182



PARTNERSHIP FOR SHARED BOOK COLLECTIONS

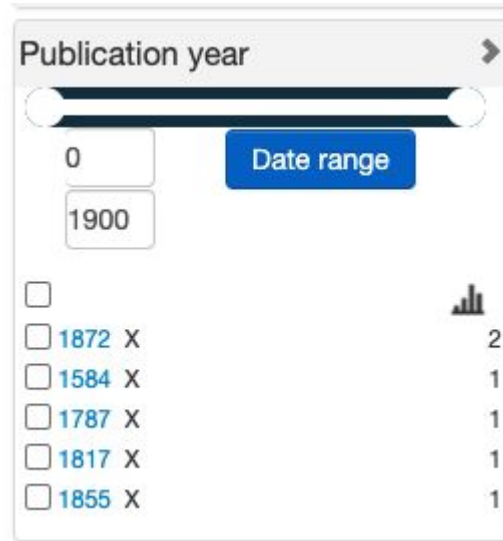


EAST

EASTERN ACADEMIC
SCHOLARS' TRUST

EAST - Projects and Process

We do the same process, but limit by year for pre 1900 materials



The screenshot shows a search interface for the EAST project. At the top, there is a 'Publication year' filter section. It features a horizontal slider bar with a right-pointing arrow. Below the slider, there are two input boxes: the first contains '0' and the second contains '1900'. A blue button labeled 'Date range' is positioned to the right of these boxes. Below the date range section, there is a list of specific years with checkboxes and counts. The years listed are 1872, 1584, 1787, 1817, and 1855, each followed by an 'X' and a count. A small bar chart icon is visible to the right of the list. The counts for each year are: 1872 (2), 1584 (1), 1787 (1), 1817 (1), and 1855 (1).

Year	Count
0	
1900	
1872	2
1584	1
1787	1
1817	1
1855	1



PARTNERSHIP FOR SHARED BOOK COLLECTIONS



EAST

EASTERN ACADEMIC
SCHOLARS' TRUST

EAST - Projects and Process

Sometimes also look at High Circ / Fewer than 5 in EAST

Using Circ from local catalog, which can be loaded into Gold Rush

EAST - Projects and Process

Post Processing:

- ~~Checking pre 1900 in case are really reprints~~ - to be fixed this calendar year
- Remove “ephemera” (Script that runs a regex match, always growing!)
- Compare to WorldCat using API to assure no more than 5 EAST retentions already exist (catches many of the false Uniques)
- Prepare list (or lists for a consortium) of proposed retention titles for review by the library(ies)
- Library finalizes commitments which are registered in the local catalog, WorldCat
- Gold Rush updated to reflect new commitments with original full dataset discarded

EAST - Promises and Pitfalls

Good (or good enough) at overlap

Okay at uniqueness, we double check OCN

Iterative improvements - Edition matching (1st vs First),
Some publisher normalization,
Overlap facet



PARTNERSHIP FOR SHARED BOOK COLLECTIONS



EAST

EASTERN ACADEMIC
SCHOLARS' TRUST

EAST - Promises and Pitfalls

Does catch and pull together some variation on metadata, e.g. publisher “Dover” vs. “Dover Publications” do match, but some variation still problematic, e.g.

Publisher: H.N. Abrams

architecture_from_prehistory_to_postmodernity_____2002____2__hnabra_____
__trachtenberg_marvin_____

Publisher: Harry N. Abrams

architecture_from_prehistory_to_postmodernity_____2002____2__harrya_____
__trachtenberg_marvin_____p

Publisher: Harry N. Abrams but title from a 245 \$b

architecturefrom_prehistory_to_postmodernity_____2002____2__harrya_____
__trachtenberg_marvin_____p

EAST - Promises and Pitfalls

Garbage in,

Can't load interleaved LHRs - data is BIB LEVEL data (e.g. for us 583 has to be in bib not lhr)

Ephemera matching done outside of Gold Rush - makes final lists a bit of a pain to produce (and can't just reproduce on the fly)

Mismatches in matchkey with already retained items - False Uniques
This problem of false uniques is by no means unique to Gold Rush

Gold Rush geared toward individual library analysis
not great for a group project, e.g. several libraries coming in as a consortium,
No mechanism to load balance retention assignments
EAST did work outside of Gold Rush to accommodate a small consortium joining, but that work not currently scalable.

Comparing GreenGlass and Gold Rush Functionality

We get asked this question enough that we made a chart:



https://eastlibraries.org/sites/default/files/BLC_Uploads/Updated%20Gold%20Rush%20vs.%20GreenGlass%20Comparison%20Chart_0.pdf

Comparison of GreenGlass and Gold Rush

Prepared by Matthew Revitt (EAST), March 2020 & updated January 2021

Function	GreenGlass	Gold Rush	Notes
----------	------------	-----------	-------

Strengths/Weaknesses in each - fitting the tool to your project needs and budget.



Thank you.

Susan Stearns
Project Director, EAST
sstearns@blc.org

Sara Amato
Data Librarian, EAST
samato@blc.org