Analyzing Collections for Shared Print

October 6, 2021
Susan Stearns & Sara Amato
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
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.
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
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
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.
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
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
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.
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?)

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

https://www.worldcat.org/title/peste/oclc/4383621 (1)
Series: Collection Folio, 42.

https://www.worldcat.org/title/peste/oclc/31529267 (1)
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 dates
s - Single known date/probable date
t - Publication date and copyright date
Those 3 steps give us a list of “Unique” and Overlap, opportunities to ‘top up’

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td>23,127</td>
</tr>
<tr>
<td><strong>Unique</strong></td>
<td>8,870</td>
</tr>
<tr>
<td><strong>Common</strong></td>
<td>14,257</td>
</tr>
<tr>
<td>Held at 2 libraries</td>
<td>6,947</td>
</tr>
<tr>
<td>Held at 3 libraries</td>
<td>1,128</td>
</tr>
<tr>
<td>Held at 4 or more libraries</td>
<td>6,182</td>
</tr>
</tbody>
</table>
EAST - Projects and Process

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

![EAST logo](https://eastlibraries.org/sites/default/files/BLC_Uploads/Updated%20Gold%20Rush%20vs.%20GreenGlass%20Comparison%20Chart_0.pdf)

<table>
<thead>
<tr>
<th>Function</th>
<th>GreenGlass</th>
<th>Gold Rush</th>
<th>Notes</th>
</tr>
</thead>
</table>

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