EZPROXY ANALYTICS AT EMORY: DEVELOPMENTS, CHALLENGES, AND APPLICATIONS

ASERL Webinar Series
23 October 2019

Jessie Copeland, Assistant Director, Resource Management, Woodruff Library

Chris Palazzolo, Head of Collections and Social Sciences Librarian, Woodruff Library
• Impetus for the EZProxy Project at Emory
• Basic Technical Background/Introduction*
• Processes and Documentation (Prioritization, User Manual, Project Management)
• Data Applications
• Data Challenges (Vendor, Archiving, Education)
• Next Steps for Development

*For more details on the technical aspects and backend of the project, we highly suggest you review our LTDS Core Systems colleagues’ presentation at 2018 Code4Lib Southeast!
Continued need for the refinement of collection assessment methods to more effectively steward collection budgets to meet patron needs and administrative accountability (funders)

Need for cross-comparative data that goes beyond the institutional level that COUNTER provides to better understand e-resource usage across patron groups and schools for the purposes of selection, deselection, and marketing of our resources

Essential to note that at Emory users are required to authenticate on and off campus
Known project at Couperin.org (EZPaarse)

Not meant to replace COUNTER statistics for vendors; to be viewed as an alternate data point

No easy access to publisher logs to determine how resource usage is actually counted/measured

Recognition that not all e-resource URLs can be parsed, and of those that can be, level of granularity can vary widely for purposes of differentiation
BRIEF TECHNICAL BACKGROUND

Parsing is analyzing a string of text and converting it to something that is human-readable.

In this case, we are parsing the ezProxy logs.

There are limitations with some vendors.
WHAT IS BEING PARSED?

- Actions
- Data
- Unique Sessions
- Unique Title IDs
- Unique Unit IDs
- Unique User IPs
PARSING LIMITATIONS

Vendors’ URLs

Turn-aways

Retrospective Data
• Business Case developed with Core Systems
• Key set of stakeholders established (collection managers)
• User Manual drafted (more on this, later)
• **How are parsers prioritized?**
  
  • Quarterly list of domains
  
  • Collection managers prioritize based on a number of factors (spend, requested evaluation, new vendors, non-COUNTER equivalents, new URL for addition to existing parser)
  
  • Priorities sent through listserv for adding to monthly projects
  
  • GitHub site for parser status and .csv file for current status and addition of parser
  
  • Enhancements for visualization, dashboards and views etc. also managed through listserv and documented on GitHub site.
  
  • Regular stakeholder meetings established
• Project Management
  • Box
  • GitHub
  • Trello

• Documentation
  • User Manual
  • Data Dictionary
  • LibStats Tableau Dashboard Glossary
  • Service Level Agreement
User Manual
- Provide guidance
- Outline advantages
- Relationship of data sources

Data Dictionary
- Defines relationships, meanings, and data elements
- Emory Shared Data
- EZProxy logs

Tableau Glossary
- Terms used in Tableau
- Assists in interpreting data
- Does not replace Data Dictionary

SLA
- Outline relationship between responsible parties
- Shared expectations
- Provide a framework
• Purpose
• Audience
• Background
• Advantages of LibStats
• LibStats vs. Vendor Stats
• Data Definitions
• Resource Types
• Actions and Counts
• Dashboards
• Best Practices
• Appendix 1: Procedure for new parsers
LITS has begun to utilize more heavily in data visualization for library actives and services, including collections (Emory By the Numbers, ARL Benchmarking, Print Acquisition assessment).

Emory has invested in and adopted Tableau as a broad university wide tool for data visualization.
• Creation of various dashboards to meet different stakeholder needs and visualize various “slicing and dicing” of the parsed data

• Platforms, publishers, user groups, resource types serve as basis

• Different schools have their own dashboards

• Data can be filtered by elements

• Different levels of data downloading (crosstabs, .csv, image [.png])
## Platforms, Publishers, and Publication Titles by Division

<table>
<thead>
<tr>
<th>Name (Stan-orded)</th>
<th>DIT</th>
<th>ELSEVIER</th>
<th>ACAD EMERGENCY MEDICINE</th>
<th>TAYLOR &amp; FRANCIS</th>
<th>CHEMICAL ABSTRACTS</th>
<th>FORENSIC SCIENCE INTERNATIONAL</th>
<th>50K</th>
<th>100K</th>
<th>150K</th>
<th>200K</th>
<th>250K</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERICO PUBLISHING</td>
<td>16,505</td>
<td>13,943</td>
<td>1,571</td>
<td>1,970</td>
<td>1,836</td>
<td>547</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OTHER</td>
<td>16,505</td>
<td>13,943</td>
<td>1,571</td>
<td>1,970</td>
<td>1,836</td>
<td>547</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHEMICAL ABSTRACTS SERVICE</td>
<td>13,526</td>
<td>1,571</td>
<td>1,970</td>
<td>1,836</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELSEVIER</td>
<td>13,526</td>
<td>1,571</td>
<td>1,970</td>
<td>1,836</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAYLOR &amp; FRANCIS GROUP</td>
<td>1,526</td>
<td>1,571</td>
<td>1,970</td>
<td>1,836</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHEMICAL ABSTRACTS SERVICE</td>
<td>1,526</td>
<td>1,571</td>
<td>1,970</td>
<td>1,836</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Select Measure
- Total Data (Bytes)
- Unique Sessions
- Unique Title IDs
- Unique Units
- Unique User IDs

### Recorded Date
- 10/30/2019

### Active Flag of Healthcare...
- Yes
- No

### Division
- Emory Campus Life
- Emory College
- Emory Temporary S...
- Emory Children's P...
- EUS Retiring/Disability
- EUS/Resort/C Liaise
- EVP Provost Academ...
- Exec VP for Health...
- Executive Vice Pres...
- Strueta Business S...
- Graduate School
- Human Resources
- Internal Audit
- ITS Library and IT...
- OIC of Doc & Alumni...
- Office of General Co...
- Oxford College
- President’s Office
- Research Administra...
- School of Law
- School Of Medicine
- School Of Nursing
DATA CHALLENGES

Vendor Limitations

Data Limitations

Comparing Data
DATA CHALLENGES

- Data Archiving
- Anonymity
- Data Retention Policy
- Stakeholder Education
APPLICATIONS:

E-BOOK TASK FORCE

CONTEXT

• Assessment of e-book collections and acquisitions over time at Emory
• Evaluation and use patterns of e-titles over time
• Providing recommendations for future purchases and strategy

____________________

• Challenges with data:
  • Not all e-book vendors can be parsed for just books (e.g., Ebook Central, which can account for significant use)
  • Incorrect parsing (different parser writers define resource differently)
  • Recent additions of heavy-use collections (JSTOR)
E-Book Use by Patron Group, FY19

- Students: 65%
- Faculty: 27%
- Walk-Ins: 8%
- Staff: 6%
- Others: 4%

E-Book Use by School/Division

- Candler
- Healthcare
- Nursing
- Business
- Law
- Oxford
- Public Health

E-Book Use by Student Level

- Graduate: 27%
- Professional: 8%
- Undergraduate: 65%
CONTEXT

• Final year of contract
• Establishment of new, standardized for assessing Emory’s journal packages
  • Desire to compare vendor data with local data

Data Challenges

• Challenges in merging title level use data in EZProxy with title list for ROI
• Use may be inflated due to parsing
### Actions by Resource Type

<table>
<thead>
<tr>
<th>Person Type</th>
<th>ARTICLE</th>
<th>BOOK</th>
<th>TOC</th>
<th>Grand Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null</td>
<td>4,197</td>
<td>65</td>
<td>1</td>
<td>4,262</td>
</tr>
<tr>
<td>Administrative</td>
<td>1</td>
<td></td>
<td>30</td>
<td>16</td>
</tr>
<tr>
<td>Faculty</td>
<td>14,667</td>
<td>816</td>
<td>146,337</td>
<td></td>
</tr>
<tr>
<td>Pre-start</td>
<td></td>
<td></td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Retired</td>
<td>1</td>
<td></td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Sponsored</td>
<td>0</td>
<td></td>
<td>222</td>
<td></td>
</tr>
<tr>
<td>Staff</td>
<td>0</td>
<td></td>
<td>87,083</td>
<td></td>
</tr>
<tr>
<td>Staff/Student</td>
<td>27,130</td>
<td>215</td>
<td>27,345</td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>64,092</td>
<td>611</td>
<td>64,648</td>
<td></td>
</tr>
<tr>
<td>Student/Applicant</td>
<td>19</td>
<td></td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Student/Staff</td>
<td>108,520</td>
<td>1,345</td>
<td>112,200</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>1,738</td>
<td>8</td>
<td>1,746</td>
<td></td>
</tr>
<tr>
<td>Grand Total</td>
<td>441,149</td>
<td>3,698</td>
<td>444,854</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Student Level</th>
<th>ARTICLE</th>
<th>BOOK</th>
<th>TOC</th>
<th>Grand Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced Standing</td>
<td>49,391</td>
<td>397</td>
<td>50,161</td>
<td></td>
</tr>
<tr>
<td>Freshman</td>
<td>1,063</td>
<td>95</td>
<td>1,158</td>
<td></td>
</tr>
<tr>
<td>Full Standing</td>
<td>60,587</td>
<td>814</td>
<td>61,407</td>
<td></td>
</tr>
<tr>
<td>Graduate</td>
<td>775</td>
<td>2</td>
<td>777</td>
<td></td>
</tr>
<tr>
<td>Graduate Year 1</td>
<td>13,880</td>
<td>63</td>
<td>13,943</td>
<td></td>
</tr>
<tr>
<td>Graduate Year 2</td>
<td>25,489</td>
<td>207</td>
<td>25,697</td>
<td></td>
</tr>
<tr>
<td>Graduate Year 3</td>
<td>2,081</td>
<td>5</td>
<td>2,086</td>
<td></td>
</tr>
<tr>
<td>Graduate Year 4+</td>
<td>288</td>
<td>5</td>
<td>293</td>
<td></td>
</tr>
<tr>
<td>Junior</td>
<td>10,240</td>
<td>143</td>
<td>10,383</td>
<td></td>
</tr>
<tr>
<td>Non-Degree Graduate</td>
<td>9,389</td>
<td>76</td>
<td>9,465</td>
<td></td>
</tr>
<tr>
<td>Non-Degree Professional</td>
<td>847</td>
<td>4</td>
<td>851</td>
<td></td>
</tr>
<tr>
<td>Non-Degree Undergrad...</td>
<td>734</td>
<td></td>
<td>734</td>
<td></td>
</tr>
<tr>
<td>Professional Year 1</td>
<td>1,575</td>
<td>18</td>
<td>1,593</td>
<td></td>
</tr>
<tr>
<td>Professional Year 2</td>
<td>750</td>
<td>6</td>
<td>756</td>
<td></td>
</tr>
<tr>
<td>Professional Year 3</td>
<td>4,352</td>
<td>22</td>
<td>4,374</td>
<td></td>
</tr>
<tr>
<td>Professional Year 4</td>
<td>5,854</td>
<td>13</td>
<td>5,867</td>
<td></td>
</tr>
<tr>
<td>Colonel</td>
<td>9,720</td>
<td>94</td>
<td>9,814</td>
<td></td>
</tr>
</tbody>
</table>
FUNDING VARIATIONS (HEALTH DASHBOARD)

CONTEXT

- Health Sciences Library receives funding from a variety of schools
- Schools often inquire re: their return on investment
- Emory Healthcare (clinical staff) has distinct user privileges (but not access to all Emory resources)
- Large numbers of residents (classified as staff)
- Growth in hospital acquisitions by Emory
  - Need to determine potential use of this content
FUNDING VARIATIONS (HEALTH DASHBOARD)
FUNDING VARIATIONS (HEALTH DASHBOARD)
NEXT STEPS FOR DEVELOPMENT

- Enhancing Views and Filtering of Data
- Data Retention
- Incorporating acquisitions data (cost)
- Adding new parsers and refining existing parsers where possible
SPECIAL THANKS!

• Special thanks go to our Core Systems team, without whom none of this would be possible!
  • Alex Cooper, Applications Support Analyst, Lead
  • Lisa Hamlett, Applications Support Analyst II
  • Ann Hulton, Applications Support Analyst II
  • Clare Barton, Applications Support Analyst II

Link to Presentation at 2018 Code4Lib Southeast:
QUESTIONS