Prologue to Perfectly Parsing Proxy Patterns

Jeremy Brown
Gretchen Smith
Scott Gillies
Introduction

- Why? The question that inspired us
- Literature Review
- How we gathered/analyzed data
- What we found
- Future Directions
  - Problems we discovered
  - Enhancements
Why are we here?
Why are we here?

- We know they use e-books
- Online degree program
- Hybrid degree program
- Loans constitute only 2% of our collection use
- Reserves transactions have outpaced loans
Who is using what?

- Use proxy logs to match resources and program of study
- Justify the purchase of eResources over print
- Allocation decisions

But e-resources are still a hard sell for some....
Distance Learning
Changes in the market

- Improved quality
- Liberalized digital rights rules
- More high quality content available
- More concurrent users
- New acquisition models
- Memories of poor usability
- No longer an “additional resource”
More pushback
“I find everything online....”
Why are you making the collection smaller?
### Extent of electronic resources

<table>
<thead>
<tr>
<th>Type</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>eJournals</td>
<td>185,541</td>
</tr>
<tr>
<td>eBooks</td>
<td>593,527</td>
</tr>
<tr>
<td><strong>Total eResources</strong></td>
<td><strong>947,628</strong></td>
</tr>
<tr>
<td>Physical Resources</td>
<td>370,572</td>
</tr>
</tbody>
</table>
Why buy a one user license?

VS.
We only use these!
The Question

• 200+ databases
• Access to hundreds of thousands of eJournal articles
• “My students do not use e-resources”
• Can we (dis)prove that?
Literature Review – Institutional Level Studies


Literature Review –
Library Studies Focused on Electronic Resource Usage


Literature Review –
Studies That Directly Impacted Our Approach


How we did it
The Pieces

• Patron database
• Proxy usage logs
• Usage database
Patron Database

- Used for proxy authorization
- Contains rich demographic data
Proxy Logs

- LogFormat %h %{ezproxy-session}i %u %t %r %s %b %{referer}
- Helpful Documentation: https://www.oclc.org/support/services/ezproxy/documentation/cfg/logformat.en.html
- %h — IP address of remote user
- %{ezproxy-session}i — session ID
- %u — username (our MUID)
- %t — date/time request was made
- %r — complete URL sent through
- We don’t really care about the remaining fields.

Proxy Usage Database Design

- Track all the useful data
- Discard all the personally identifiable data
- One set of data for students and employees
Mapping Data

- operational Title
- division
- campus
- URL
- count
- version-title
- date inserted
- IP address
- program
- program version (degree)
- grade level
- level
- program-division
- session
- campus
- Employee
- Student Enrollment
- MUID
- MUID
Capturing Data

• We rotate the EZproxy logs daily.
• After we rotate, go through each line of the log.
  • The first instance of an IP/MUID/session tells us everything we need, including the initial database URL
  • We increment the counter for subsequent occurrences of the session
• We add a date stamp for each session so we can capture trends by date
Data Analysis
Tools

- SQL queries to generate data
- Excel
  - Simple tables and charts for summary analysis
  - Pivot tables to help spot trends over time
Usage by program over time

Program Daily Activity

- Spring Break

Legend:
- Adjunct
- English Language Institute
- Georgia Baptist College of Nursing
- School of Engineering
- unknown
- College of Health Professions
- Eugene W. Stetson School of Business & Economics
- James & Carolyn McAfee School of Theology
- School of Medicine
- Tift College of Education
- Walter F. George School of Law
- College of Liberal Arts
- General University
- General University Employee
- MERC
- Penfield College
- Townsend School of Music
- College of Pharmacy
Usage by program over time
Usage by program over time
Which programs use our resources?
Usage grade level

- 4th Year Graduate/Professional: 15%
- 3rd Year Graduate/Professional: 7%
- 2nd Year Graduate/Professional: 4%
- 1st Year Graduate/Professional: 9%
- Graduate - Master's: 13%
- Doctoral Candidate: 8%
- Employee: 6%
- Other: 3%
- Sophomore: 2%
- Freshman: 1%
- Unknown: 22%
- Graduate Specialist - Post Master's: 0%
- Special Non-Degree Undergrad: 0%
- Special Non-Degree Graduate: 0%
- International Exchange Student: 0%
Campus usage summary

Percentage

- Atlanta: 56.02%
- Unknown: 24.47%
- Macon Main: 8.33%
- Distance Learning: 4.73%
- Macon Center: 2.59%
- Henry County: 2.25%
- Douglas County: 1.27%
- Atlanta Center: 1.12%
- Navicent Health: 0.72%
- Macon: 0.16%
- MDC: 0.11%
- Eastman: 0.05%
- Newman Center: 0.02%
- Medical School Macon: 0.01%
- Atlanta Non-Credit: 0.01%
- Savannah: 0.01%
- Newman Center: 0.00%
Campus usage over time

That's a lot of unknowns!
What we know about “unknown”

• Generally on-campus users
• We have an IP address
• We know where the IP addresses are
  • Atlanta — graduate/professional
  • Macon — traditional undergraduate
  • Henry/Douglas/Newnan — non-traditional undergraduate
Our unknowns are traditional undergraduates
Our “unknowns” & undergraduates

• Macon is a residential campus.
• Most students live on campus.
• Most of them use the campus LAN to access resources.
• They made up a very small percentage of the previous reports.
Resource Usage
Problems We Discovered

• Our URLs aren’t perfect. Platforms appear together.
• Our “Unknowns” — there are ways of getting user information about these folks.
• Only collecting data for first resource used in a session.
Future Directions

• Track other Library Services.
• Correlate library use with student achievement.
• Track effectiveness of database marketing efforts.
• Utilize data for collection development.
Conclusions

• We can effectively track our database usage by demographic group.
• This proves who our strongest customers are.
Questions?

Slide background: By Max Ronnersjö (Own work) [CC BY-SA 3.0], via Wikimedia Commons: https://commons.wikimedia.org/wiki/File%3AAeonium_tabuliforme.jpg