EZproxy and Analytics

DON HAMPARIAN, OCLC
Analytics and the Library
Data Increasingly Drives Library Decisions

- Collection management (select, deselect)
- Facility use and configuration
- Outreach programs (student, faculty, funder, community)
- Library contribution to student success
- Library contribution to institution success
More Data Needed from More Sources

- **Learning Management Systems:** Schedule, Performance
- **Library:** gate counts, circulation, ILL, Wi-Fi @ library, e-resources, facility use
- **Publishers:** COUNTER stats
- **Student Demographics & Goals**
- **Budget & Financial**
Both Evolutionary and Revolutionary

• Slow, steady process on data discovery, transformation, and curation
• Slow process on analytics management systems (Big Data, Warehouses, ETL)
• More rapid demands from institution leadership
• More data from different sources helps us understand and imagine opportunities
• Still pioneering work
Our Focus Today: E-Content Usage
E-Content Usage Statistics: Sources

Major sources of e-content usage statistics

• COUNTER statistics
• ILL e-resource statistics
• EZproxy log data
COUNTER Statistics

Strengths
• Standards based
• Specific articles/journals sited
• Good support from publishers
• WorldShare License Manager manages COUNTER stats

Challenges
• Standards based but different interpretations
• Combining from many publishers into a usage analytics data set
ILL e-Resource Statistics

Reflects increasing use of ILL for article fulfillment:

Strengths

• About 40% of OCLC resource sharing traffic is not “books” so a rich set of data exists

Challenges

• Data formats different so work involved in “mashing-up” with other e-resource data
EZproxy Log Data

Strengths
• Complete coverage of all EZproxy users and resources
• One set of data to manage vs publisher-by-publisher
• Consistent definitions of use (retrieve a PDF or TOC for example)
• Institution selects the privacy level (username, anonymized, group, department)
• Richer than SAML system statistics

Challenges
• URLs not e-resource names
• A significant amount of “noise” in the log files
Enriching the EZproxy log data

Goals

• Get identifiers/names for resources, not URLs
• Eliminate the “noise” records in the logs
• Define an e-resource access event

Challenges

• Translating identifiers/names from URLs varies with each publisher
• Almost like configuring EZproxy “in reverse”
EZproxy Analytics Pilot
EZproxy Analytics Pilot
Six Institution Pilot – 3 in EMEA, 3 in Americas
Couperin.org – consortium of HE, Research Organizations, Hospitals, large schools, national library, and museums

ezPAARSE – available for 6 years, with a large open source support community and over 60 institutions using it today
Pilot Goals

- Enrich EZproxy log data to make it more useful in library data mashups
- Provide a query and export UI
- Collaborate with libraries to learn:
  - Library & institution goals for usage data and analytics
  - How can this data be further enriched for increased usefulness
  - How can OCLC help libraries manage this data
- Launch service in 2019
EZproxy Analytics Pilot Components

- Hosted by OCLC at our Amsterdam Data Center
- All components on the institution’s EZproxy hosted server
- Hosting VMs are Linux
Sample Dashboard
Sample Queries
Sample Query

Top 10 DOIs

Count

DOI

10.2307/20096779
10.2307/2005106
10.2307/2069673
10.2307/12112483
10.2307/172351
10.2307/172351
10.2307/2069942
10.2307/2069942
10.2307/2069942
10.2307/2069942
10.2307/2069942
Sample Event and Audit Data
EZproxy Analytics Service
EZproxy Analytics: an add-on to EZproxy hosted

• All Components on dedicated institution server
• Your data stays on that server
• You access data through Kibana UI
• OCLC processes EZproxy usage and audit data for you
EZproxy Analytics Service

Available later in 2019
For More Information Contact
Your Local Account Representative
THANK YOU

Don Hamparian
OCLC
hamparid@oclc.org