Usage Data for Collection Development and Management: Changes, Issues, Local Utilization, and Shared Possibilities

Deborah Blecic
6/29/2013
OCLC Library Analytics: Data to Inform Decision-making and Measure Impact
The “Old” Days – gather use data by hand

- Counting each time a print journal volume or issue was reshelved
- Citation studies – random sample of faculty articles
- Counting date due stamps in a print book

Results shared via publication, but only small samples of the data
Today – more quantitative data

- Publisher/vendor-generated use statistics for databases, books and journals, much of it standardized
- Turnaways for owned, subscribed, and unowned content
- ERMS statistics
- Better OPAC’s with more data available, often consortially
- Online citation indexes
- DDA/PDA data
International Coalition of Library Consortia (ICOLC) – early standardization for publisher/vendor data


http://legacy.icolc.net/webstats.html
Project COUNTER – 11 years and continuously evolving

April 2012- Release 4 covers all types of e-resources in one single, integrated Code of Practice, including journals, databases, books, reference works and multimedia (implementation date Dec 31, 2013) – 23 reports, 13 standard and 10 optional

New reports include Multimedia Report 1, report on usage on mobile devices, etc…

Changes: Book report 2- section must be defined; for databases session counts no longer required, but record views and result clicks are, and will give more meaningful information
Project COUNTER

Librarian research and recommendations have played a key role in changes over the years – we must continue.

Delving into the data not only helps us to evaluate use at our libraries, but allows us to contribute at the broader level to the evolution of usage statistics standards as we share our findings.
Charleston Conference 2012

Everything That’s Wrong with E-book Statistics – a comparison of E-book Packages

Harvesting the Crop: Implementing a Usage Statistics Management System at Georgia State University

Does Format Matter? Comparing Usage of E-books and P-books

Collaborating to Analyze E-Journal Use Data: A Discussion of Cross-Institutional Cost-Per-Use Analysis Projects within the UNC System
Database Use

Searches and Sessions were standard measures for the 15 + years,

At UIC, 2 studies focused on the evolution of search and session data from ICOLC to COUNTER, looking at ratios of searches to sessions, variability over time, etc…and reported how automated searches and federated search engines were affecting search and session counts
Database Use Statistics Progression

Since 2008 automated and federated searches broken out in COUNTER-compliant statistics

But the ability to search multiple databases on a platform still resulted in high search and session counts for many resources
Database use statistics with COUNTER4

Sessions have been retired

Result clicks and record views are the new measures

New types of data will tell us new things about how our users interact with databases
E-journals

COUNTER standard is the Successful Full-Text Article Requests (SFTAR)

Often combined with cost to get a cost-per-SFTAR figure for both single titles and packages

At UIC, recently published a study about metrics developed to compare “Big Deal” journal packages:

## Deal-Level Metrics and Rankings for Three Big Deals

<table>
<thead>
<tr>
<th>Deal-Level Metric</th>
<th>Big Deal 1</th>
<th>Metric Rank</th>
<th>Big Deal 2</th>
<th>Metric Rank</th>
<th>Big Deal 3</th>
<th>Metric Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All Journal Titles</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost Per SFTAR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>$2.62</td>
<td>1*</td>
<td>$8.51</td>
<td>3</td>
<td>$4.89</td>
<td>2</td>
</tr>
<tr>
<td>2007</td>
<td>$2.14</td>
<td>1</td>
<td>$6.59</td>
<td>3</td>
<td>$4.48</td>
<td>2</td>
</tr>
<tr>
<td>2008</td>
<td>$2.53</td>
<td>1</td>
<td>$5.82</td>
<td>3</td>
<td>$5.36</td>
<td>2</td>
</tr>
<tr>
<td><strong>Subscribed Journal Titles</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average of SFTARs **</td>
<td>213.1</td>
<td>2</td>
<td>100.1</td>
<td>3</td>
<td>335.7</td>
<td>1</td>
</tr>
<tr>
<td>Median of SFTARs</td>
<td>84.7</td>
<td>2</td>
<td>38.7</td>
<td>3</td>
<td>147</td>
<td>1</td>
</tr>
<tr>
<td>% with Average SFTARs per Year Equal to 25 or Higher</td>
<td>75.1</td>
<td>2</td>
<td>58.5</td>
<td>3</td>
<td>77.2</td>
<td>1</td>
</tr>
<tr>
<td><strong>Add-on Journal Titles</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average of SFTARs</td>
<td>79.9</td>
<td>1</td>
<td>30.9</td>
<td>3</td>
<td>77.6</td>
<td>2</td>
</tr>
<tr>
<td>Median of SFTARs</td>
<td>26.5</td>
<td>1</td>
<td>9.3</td>
<td>3</td>
<td>25.3</td>
<td>2</td>
</tr>
<tr>
<td>% with Average SFTARs per Year Equal to 25 or Higher</td>
<td>52.1</td>
<td>1</td>
<td>28.7</td>
<td>3</td>
<td>50.2</td>
<td>2</td>
</tr>
<tr>
<td><strong>Subscribed and Add-on Titles</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage that Accounts for 80% of Subscribed and Add-on SFTARs</td>
<td>26.7</td>
<td>1</td>
<td>23.4</td>
<td>3</td>
<td>23.6</td>
<td>2</td>
</tr>
<tr>
<td>Deal Composite Rank Score</td>
<td>13</td>
<td>30</td>
<td>17</td>
<td>17</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* For all Deal-Level Metrics and Deal Composite Rank Score, the lowest number is best and the highest number is the worst.

** All averages and medians are based on the three year average of SFTARs for each journal title.
Metrics for Journals within Big Deals

3-year average of SFTARs and cost-per-SFTAR both ranked then ordered by adding the ranks together

Why? If two journals have the same cost-per SFTAR, the journal with the higher average SFTARs will rank ahead of the other. If two journals have the same average SFTARs, the one with the lower cost per SFTAR will rank ahead of the other.
Metrics for Journals within Big Deals

<table>
<thead>
<tr>
<th>Journal Title</th>
<th>Journal List Price 2009</th>
<th>Cumulative Journal List Price</th>
<th>3-year Average of SFTARS per year</th>
<th>Rank of Average SFTARS</th>
<th>Cost per Average SFTARS</th>
<th>Rank of Cost per SFTAR</th>
<th>Journal Combined Rank Score</th>
<th>Combined Rank Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>BJOG: An International Journal of Obstetrics and Gynaecology</td>
<td>$567</td>
<td>$567</td>
<td>2,627.3</td>
<td>1</td>
<td>$0.22</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Child Development</td>
<td>$190</td>
<td>$757</td>
<td>1,577.7</td>
<td>7</td>
<td>$0.12</td>
<td>1</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Journal of the American Geriatrics Society</td>
<td>$748</td>
<td>$1,505</td>
<td>1,535.0</td>
<td>8</td>
<td>$0.49</td>
<td>6</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>Journal of Obstetric, Gynecologic, and Neonatal Nursing</td>
<td>$426</td>
<td>$1,931</td>
<td>1,057.3</td>
<td>14</td>
<td>$0.40</td>
<td>4</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>Journal of Personality</td>
<td>$969</td>
<td>$2,900</td>
<td>1,480.0</td>
<td>9</td>
<td>$0.65</td>
<td>9</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>Journal of Child Psychology and Psychiatry and Allied Disciplines</td>
<td>$773</td>
<td>$3,673</td>
<td>925.0</td>
<td>17</td>
<td>$0.84</td>
<td>13</td>
<td>30</td>
<td>6</td>
</tr>
<tr>
<td>Public Administration Review</td>
<td>$347</td>
<td>$4,020</td>
<td>651.7</td>
<td>29</td>
<td>$0.53</td>
<td>7</td>
<td>36</td>
<td>7</td>
</tr>
</tbody>
</table>
Other Big Deal Factors Studied

Characteristics of use for Humanities, Social Sciences, STEMath, and Health Sciences and implications for the collection

How many SFTARs occur in full-text databases rather than on the publishers platforms – overlap analysis
## Overlap Analysis Summary for Overlap Providers with 30 or More Titles That Overlap Publishers’ Titles, Subscribed and Add-on Titles Only

<table>
<thead>
<tr>
<th></th>
<th>Big Deal 1</th>
<th>Big Deal 2</th>
<th>Big Deal 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average SFTARs per Year on Publisher’s Platform</td>
<td>106,620</td>
<td>61,537</td>
<td>73,980</td>
</tr>
<tr>
<td>Average SFTARs per Year on Overlap Providers’ Platforms</td>
<td>70,552</td>
<td>7,377</td>
<td>5,390</td>
</tr>
<tr>
<td>Total of Average SFTARs on All Platforms Studied</td>
<td>177,172</td>
<td>68,914</td>
<td>79,370</td>
</tr>
<tr>
<td>% of Average SFTARs on Publisher’s Platform</td>
<td>60.2%</td>
<td>89.3%</td>
<td>93.2%</td>
</tr>
<tr>
<td>% of Average SFTARs on Overlap Providers’ Platforms</td>
<td>39.8%</td>
<td>10.7%</td>
<td>6.8%</td>
</tr>
</tbody>
</table>
Questions for Shared Data

Are the deal-level metrics consistent across libraries? Can we share cost data or is it impacted by non-disclosure agreements?

The journal analysis uses list price, not actual cost. Are rankings consistent across similar libraries, or very different?

The percentage of title overlap would be consistent across libraries, but are the SFTAR patterns similar?

Subject data – are patterns consistent across similar types of libraries?
Another Recent UIC Study


Demonstrated a strong Spearman Rank-Order Correlation among the three measures
## Results – correlation coefficients

<table>
<thead>
<tr>
<th>Data Subsets</th>
<th>Link Resolver/Vendor</th>
<th>Link Resolver/Citation</th>
<th>Vendor/Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of Titles</td>
<td>Spearman’s Correlation</td>
<td>Spearman’s Correlation</td>
</tr>
<tr>
<td>All Journals</td>
<td>2619</td>
<td>.843</td>
<td>.752</td>
</tr>
<tr>
<td>1-25th Percentile of titles (Link-resolver use ≤26)</td>
<td>663</td>
<td>.454</td>
<td>.392</td>
</tr>
<tr>
<td>25th-75th Percentile of titles (Link-resolver use ≥27-≤214)</td>
<td>1302</td>
<td>.563</td>
<td>.408</td>
</tr>
<tr>
<td>75th -100th Percentile of titles (Link-resolver use ≥215)</td>
<td>654</td>
<td>.586</td>
<td>.511</td>
</tr>
<tr>
<td>25th -100th Percentile of titles (Link-resolver use ≥27)</td>
<td>1956</td>
<td>.703</td>
<td>.581</td>
</tr>
</tbody>
</table>
Sample of titles with anomalous results

<table>
<thead>
<tr>
<th>Title</th>
<th>Link-Resolver</th>
<th>Vendor</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amer. Journal of Clinical Nutrition</td>
<td>1,295</td>
<td>0</td>
<td>140</td>
</tr>
<tr>
<td>Biochemistry (Easton)</td>
<td>935</td>
<td>0</td>
<td>364</td>
</tr>
<tr>
<td>Infection and Immunity</td>
<td>142</td>
<td>0</td>
<td>122</td>
</tr>
<tr>
<td>Internet J. of Peds. &amp; Neonatology</td>
<td>0</td>
<td>301</td>
<td>0</td>
</tr>
<tr>
<td>J. For Specialists in Ped. Nursing</td>
<td>0</td>
<td>799</td>
<td>0</td>
</tr>
<tr>
<td>Critical Care</td>
<td>665</td>
<td>19</td>
<td>21</td>
</tr>
<tr>
<td>Diabetes Care</td>
<td>1,905</td>
<td>730</td>
<td>190</td>
</tr>
<tr>
<td>Journal of Studies on Alcohol</td>
<td>520</td>
<td>496</td>
<td>38</td>
</tr>
<tr>
<td>British Journal of Ophthalmology</td>
<td>508</td>
<td>512</td>
<td>35</td>
</tr>
<tr>
<td>Gerontologist</td>
<td>590</td>
<td>608</td>
<td>26</td>
</tr>
</tbody>
</table>
Conclusions

• Link-resolver data from our ERMS correlates with vendor use statistics, and thus can aid with collection development decisions, especially if vendor data cannot be obtained.

• If a journal has high Link-resolver counts, one can be fairly confident that vendor stats will be high. Link-resolver data can be used to identify journals that need further evaluation for retention decisions.
Questions for Shared Data

Are the correlations similar for other subject areas?

Are the correlations similar at other libraries?

Do they vary based on size of user population or other factors?

Are they similar for other ERMS systems?
Concerns about COUNTER journal data

Overview:

Issues include variability in the number of years and articles, news item counts the same as a research article, usage spikes, journal transfers and title changes, open access articles, and the influence of the platform design.
Platform Effects on E-journal statistics

• First reported by Davis and Price in 2006, study of shared data from 32 research libraries in US, UK, and Sweden

• Some publishers connect a user initially to an abstract, others to the html version of the article. In the latter case, if user then clicks on pdf, one use generates two SFTAR counts. Other variables affect ratio also.

• At UIC, one vendor was 3.91 pdfs per html one year and 6.41 two years later, another vendor was 14.47 one year and 18.87 the next. A third vendor more consistent at 11.22, 12.61, and 11.73.

• Not fixed in COUNTER4 reporting, seems unlikely it can be

• We can’t know for sure if the user was satisfied with the html
Other factors

Mentioned in Big Deals study – overlap with full-text databases

Also overlap with subject and institutional repositories –
ArXiv
PubMedCentral
Social Sciences Research Network
RePEc: Research Papers in Economics
Etc…. 
COUNTER4 Changes for E-journals

- New Required Report: Number of Gold OA SFTARs
- Optional reports include reporting of usage on mobile devices
- New multimedia reports may impact reporting for journals such as JOVE
E-books: New Types of Data

- Print books: one user, many weeks or months

- E-books: many users in one day
  - New way to meet some user needs
  - New types of use patterns and use
  - May vary by field

- COUNTER statistics allow two types of reporting – not yet standardized
  - Book Report 1: Title Requests
  - Book Report 2: Section Requests
    - New in COUNTER 4: The Section must be defined
E-books are used – some UIC data

• Ovid E-books (COUNTER Book Report 1)
  – 2010 600 E-books; 4,545 title requests
  – 2011  799 E-books; 9,774 title requests

• Springer E-books (COUNTER Book Report 2)
  – 2010: ~45,000 e-books; 32,641 section requests
  – 2011:  ~50,000 e-books; 56,104 section requests

• ebrary (COUNTER Book Report 2)
  – 2010  7,733 E-books used; 486,231 section requests
  – 2011  10,968 E-books used; 822,547 section requests
E-book Advantage


1 print copy – currently checked out

ebrary e-book: 23,261 section requests in 2011
Conclusions

• We have more quantitative data than ever before

• No type of data is perfect, but all give us information that is helpful for making decisions

• Use many types of data in collection development and management decisions

• Reporting and sharing results and problems will help the evolution of codes of practice

• Sharing data for cross-institutional analysis is possible and will yield new types of research
Shared Data Possibilities

- Standardization such as Project COUNTER makes it easier to share – data is consistent across institutions.

- Cost per use data - impacted by non-disclosure agreements

- Time and effort – currently time consuming to collect data locally let alone contribute to larger repositories, but harvesting initiatives and products will help

- Consortia already doing internal analysis
More Shared Data Possibilities

- Consortial catalogs already have data from more than one institution

- Citation databases offer data on multiple institutions

- Libraries that have the same ERMS can share the same types of data

- DDA/PDA projects, especially consortial, may yield shared data
Questions?

dblecic@uic.edu
Bibliography

