E-BOOK STATISTICS: HOW BIG IS THE DATA

Vicki McGarvey
vicki.mcgarvey@staffs.ac.uk
Learning and Information Services Manager
Staffordshire University
E-book big data, there appears to be so much, are we battling against an endless tide
Seeing the light, it's not about the answers it's about the questions
"Great teaching cannot be captured by dangerously soulless algorithms" Clare Taylor THES May 31st 2016

...Data has the potential to help us identify patterns of engagement and intervene for the benefit of student learning

"Across business, industry, government and other areas of human endeavor, vast amounts of data are being accumulated and processed to develop understanding of people's activities, and to optimise organisational processes and outputs" JISC

....Universities are however only beginning to understand how to exploit this to enhance the educational experience for students.
THE CHALLENGES

- Data: not always cleanly presented; understanding the terminology; aggregating from multiple places
- Data Skills: accessing; understanding; presenting to different audiences; being analytical

What we do have: reports
What we can do: create the right questions
Counter Reports

<table>
<thead>
<tr>
<th>Counter Report</th>
<th>Description use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Book Report 1: Number of successful title requests by month &amp; title</td>
<td>Tracking individual and aggregate e-book use</td>
</tr>
<tr>
<td>Book Report 2: Number of successful section requests month &amp; title</td>
<td>Tracking section/chapter use e.g. reference edited work</td>
</tr>
<tr>
<td>Book Report 3: Access denied to content items by month, title &amp; category - turnaway at title level</td>
<td>Helpful in assessing acquisition model/simultaneous users limits; needed copies</td>
</tr>
<tr>
<td>Book Report 5: Total searches by Month &amp; Title</td>
<td>Total search &amp; session at the title level</td>
</tr>
<tr>
<td>Provider</td>
<td>Report</td>
</tr>
<tr>
<td>----------</td>
<td>--------</td>
</tr>
<tr>
<td>Ebrary Academic Source Complete – search date period, current &amp; last month YTD, last 12 months Download HTML, Excel, XML</td>
<td>Category summary</td>
</tr>
<tr>
<td></td>
<td>Site Activity</td>
</tr>
<tr>
<td></td>
<td>Title Report</td>
</tr>
<tr>
<td>Dawsonera – Aggregator single title purchase: date range</td>
<td>Popular book reports</td>
</tr>
<tr>
<td></td>
<td>Usage report</td>
</tr>
<tr>
<td>Additional reports</td>
<td>MUA</td>
</tr>
</tbody>
</table>
DATA ANALYSIS IT IS ALL IN THE QUESTION

When you have enough

What to collect

Your question will inform

Where to look

How often
## The research questions

<table>
<thead>
<tr>
<th>Who &amp; where</th>
<th>What</th>
<th>Data collected</th>
<th>Data findings</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laurentian University Canada (2013)</td>
<td>Compared usage of individual selected titles with titles in bundles</td>
<td>Comparison usage data of large consortia packages with individual titles</td>
<td>Higher usage of individually purchased titles than aggregators</td>
<td>Less restrictions the higher the usage</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Comparison usage data of aggregated collections of individually selected titles with packages acquired from publisher</td>
<td>High usage packages selected from publishers with limited restrictions</td>
<td>Large packages have irrelevant titles</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Buying ebook packages may help build critical mass then later can be more selective</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Currency may impact on usage</td>
</tr>
<tr>
<td>Liverpool University (2010)</td>
<td>Influence of usage of a large collection by factors such as subject, year of publication</td>
<td>Longitudinal counter usage reports augmented with contextual title-level information from vendor</td>
<td>Number of unused titles diminished each year</td>
<td>Big deal approach still worked well</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Cost per use low</td>
<td>Older titles are significant</td>
</tr>
<tr>
<td>Who &amp; where</td>
<td>What</td>
<td>Data collected</td>
<td>Data findings</td>
<td>Analysis</td>
</tr>
<tr>
<td>-------------</td>
<td>------</td>
<td>----------------</td>
<td>---------------</td>
<td>----------</td>
</tr>
<tr>
<td>University of Westminster (2009)</td>
<td>Value for money: 1. Compared a range a subscription options to see if could reduce titles &amp; which 2. Compared aggregators a. annual credits b. restriction on simultaneous users</td>
<td>1. Vendor monthly counter reports: sessions, searches, full-text section access; total usage over lifetime of collection; frequency ratio – number of months title had shown usage divided by months it had been part of collection 2. a. When credits were used &amp; turnaways b. Turnaway reports &amp; number of times accessed</td>
<td>1. Title ranking based on usage &amp; frequency, found all collection was being. 2. a. High use titles via turnaways &amp; access reports b. High use titles numbers of credits &amp; turnaways</td>
<td>1: Comparison: Could reduce collection without having effect on access, difficult to identify which titles could drop 2:Comparison: Decided best to avoid turnaway model, go for credits As patrons get more familiar with ebooks usage may increase Low use titles may be used for background reading</td>
</tr>
<tr>
<td>University of Nevada (2012)</td>
<td>Cost effectiveness of collections from 2 aggregators to make sure most appropriate collection looking at subject and publishers</td>
<td>Longitudinal usage report over 3 years</td>
<td>Found increase in usage of 1 aggregator &amp; a decrease in the other</td>
<td>High usage if collection is larger &amp; more current; Interface may have an impact; Popularity of ebooks by publisher may help with purchasing; Low popularity may help with marketing</td>
</tr>
</tbody>
</table>
• **Annual resources review**: for budget savings using percentage threshold based on the reduction of students; do we want to cancel, or replace with something else?

• **Review composition of packages**: change titles? e.g. Safari; more UK publishers in Ebrary

• **Publisher no longer makes title available**: Do we want to purchase the title individually?

• **Low usage**: Market better? Not appearing in discovery service? Older edition ranking higher than a newer? Students accessing information elsewhere e.g. computing & arts

• **Increase in number if pages printed (eBrary)**: Change in usage behaviour; impact of discovery; dipping in rather than reading all
DATA WHAT CAN WE DEDUCT & DO

High usage title, publisher & subjects

- **Deduct:** popular titles; which disciplines are using the collection; publishers that are popular for disciplines

Low usage

- **Deduct:** background reading; change in behaviour; not being discovered; user interface; permalink problem
- **Do:** improve marketing; check discovery; feedback on user interface; check reading list links

High turnaways

- **Do:** need to buy more copies of titles; change business models e.g. MUA to unrestricted
BEYOND THE MISSING PIECES

What data will not tell you yet!
• Which student
• Which course
• Length of session
• What a user does with the information

Working around this
• Reading list stats
• VLE stats
• Interpretation of class numbers/subject areas
Where is all this data going?

- SCONUL
- Librarians
- Senior Management
- Patrons
- University Committees

Etc....
NATIONAL INITIATIVES

JUSP:  http://jusp.mimas.ac.uk/news/JUSP-ebooks-survey.docx

What do you use e-Book usage statistics for top results:
•  Value for money
•  Compiling reports for library management
•  Determining which titles are being used
•  Budget decision
•  Package comparisons

JUSP still has a way to go: standardisation; variety of reports; aggregator representation; financial information; subject areas; top books; downloads; turnaways

Lamp Project JISC Library Analytics and Metrics Project
http://jiscllamp.mimas.ac.uk/about-lamp/

E-book analytics not as developed as e-journal analytics
YOUR TURN...

What type of analysis/data do you think the following would require and how would you present the data?

- Director of service
- Librarian
- Learning and Teaching Committee
REFERENCES

- Taylor, C., Great Teaching Cannot Be Captured by Dangerously Soulless Algorithms THES May 31st 2016