Everything That's Wrong with E-Book Statistics: A Comparison of E-Book Packages

Karin Byström
Uppsala University Library, karin.bystrom@ub.uu.se

Follow this and additional works at: <https://docs.lib.purdue.edu/charleston>

Part of the [Library and Information Science Commons](#)

An indexed, print copy of the Proceedings is also available for purchase at:

<http://www.thepress.purdue.edu/series/charleston>.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information Sciences. Find out more at: <http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences>.

Karin Byström, "Everything That's Wrong with E-Book Statistics: A Comparison of E-Book Packages" (2012). *Proceedings of the Charleston Library Conference*.
<http://dx.doi.org/10.5703/1288284315105>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Everything That's Wrong with E-Book Statistics: A Comparison of E-Book Packages

Karin Byström, Head, Section for E-resources, Uppsala University Library

Abstract

This poster presentation highlights the problems that exist in defining “a download” for e-books. Even though there is a COUNTER code of practice, a download can still be defined as either a page, chapter or title use. Many e-book publishers don't follow COUNTER at all, and then the differences are even bigger. Libraries face many problems because of this, and this poster aims to raise awareness on the problems concerning analyzing e-book usage statistics.

Introduction

Digital e-book collections are an important part of a modern academic library collection, and it is important to analyze the use. However, there are many difficulties regarding the usage statistics, for example, analyzing, comparing, and gathering (Cox, 2008).

Each year libraries gather usage statistics from publishers and vendors, but it's difficult to do something relevant with it. Libraries want to use usage statistics in the renewal process. They want to establish value for money by comparing the number of downloads and price per download.

Usage statistics for e-books is much more complicated than e-journal statistics, and that brings great implications when comparing different publishers. Libraries need to understand how usage statistics can differ between publishers and the underlying mechanisms on how usage statistics are gathered.

Method

An investigation was done on the type of usage statistic delivered for 13 e-book collections at Uppsala University Library. In many cases it was difficult to find detailed information, and vendor customer services had to be contacted.

Usage statistics were gathered from the vendor platforms and financial information was received from the library's economic department.

Background

There is an existing COUNTER Code of Practice for e-books, and compliant vendors are listed on the COUNTER website (COUNTER, 2012a). The current standard includes two different types of book reports, Book Report (BR) 1 and Book Report 2. BR1 shall deliver statistics on the “title” level, and should only be given if the e-book is constituted of one file and the usage cannot be analyzed in more detail. Title-level statistics means that if a user reads from many chapters, it still is only registered as one download.

BR2 shall report use of *sections* of e-books, on “the first level of subdivision” of a book, which in most cases means chapter level. But this wording also leaves room for different interpretations on what constitutes “the first level of subdivision,” and consequently, what constitutes a download. Some vendors count use by page, and others also include printing and copying of pages. Using page level as “the first level of subdivision,” of course gives publishers a higher number of downloads than chapter or title level.

In April 2012 the fourth release COUNTER Code of Practice was published, and it included a few changes for the book reports (COUNTER 2012b). The most important change is for BR2, and means that vendors must state *which kind* of “first level of subdivision” they use, that is, chapter or page. This might sound like a small change, but this means that vendors will have to be clearer about what kind of statistics they supply. The deadline date for implementation of Release 4 is 31 December 2013.

	A	B	C	D	E	F	G	H
1	Book Report 2 (R4)	Number of Successful Section Requests by Month and Title						
2	<Customer>	Section Type:						
3	<Institutional Identifier>	<Chapter, etc.>						
4	Period covered by Report:							
5	yyyy-mm-dd to yyyy-mm-dd							
6	Date run:							
7	yyyy-mm-dd							
8		Publisher	Platform	Book DOI	Proprietary Identifier	ISBN	ISSN	Reporting Period Total
9	Total for all titles		Platform Z					998
10	Title AA	Publisher X	Platform Z			9-787543-690548		134
11	Title BB	Publisher X	Platform Z			9-783924-691539		124
12	Title CC	Publisher Y	Platform Z			9-786432-692557	0154-1521	454
13	Title DD	Publisher Y	Platform Z			9-781945-693446		286
14								
15								
16								
17								

Figure 1. The Section Type (Chapter, Encyclopedia Entry, Etc.) Used in this Report Must Be Indicated in the Report Itself as Shown. Where More Than One Type of Section Is Used, Simply List the Predominant Type Covered in this Report.

Figure 2. Usage Statistics for E-Book Collections at Uppsala University 2006-2011

Results

Total Use

The total number of downloads for 13 e-book collections can be seen in Table 1. This is how librarians usually look at their statistics. Just looking at the graph, it's easy to think that ebrary is the absolutely best collection, and the ones in the bottom should be cancelled. However, a

deeper analysis of the different levels of usage statistics reveals a different picture.

Type of Download

In this analysis 9 of 13 packages follow the COUNTER standard. But even though they are COUNTER compliant there are still differences. Books@Ovid is the only vendor that gives BR1 with title level statistics. However, MedicinesComplete also gives title level statistics,

but they use BR2. ebrary and Thieme give statistics at page level, and Ebrary also includes print and copying of pages, since ebrary considers each access a “successful section request”. This of course means a lot more “downloads” and for Uppsala printing and copying stands for over 10% of the total use. Five collections use BR2 with statistics on chapter level; Cambridge Histories Online, ECCO, Emerald, Sage, and Springer.

Four collections don’t follow the COUNTER standard. Two of them use the International Coalition of Library Consortia (ICOLC) “Revised Guidelines for Statistical Measures of Usage of Web-Based Information Resources” (ICOLC, 2006),

namely Oxford Scholarship Online (OSO) and Early English Books Online (EEBO). It is positive that they follow a standard, but since it isn’t similar to the others it doesn’t help libraries much.

Two collections, AccessMedicine and Knovel, don’t follow any standard. AccessMedicine reports nearly all use of their web content as “downloads”, and the statistics not only include book chapters, but also video lectures and self-assessment tests found on their web site. This of course gives AccessMedicine a very high usage. Knovel includes interactive tables in their usage statistics that wouldn’t be counted if they used the COUNTER standard.

Package name	Status	Level	Description
AccessMedicine	Not COUNTER		Chapter + self-assessments + updates
EEBO	"Spirit of ICOLC"		pdf, ASCII, document/page
Knovel	Not COUNTER		Chapter + "data"
Oxford scholarship online	ICOLC		5 pages = one "full content unit"
Books@Ovid	COUNTER	BR1	Title
MedicinesComplete	COUNTER	BR2	Title
Ebrary	COUNTER	BR2	Page + print + copy
Thieme	COUNTER	BR2	Page
Cambridge Histories online	COUNTER	BR2	Chapter
ECCO	COUNTER	BR2	Chapter
Emerald	COUNTER	BR2	Chapter
Sage	COUNTER	BR2	Chapter
Springer	COUNTER	BR2	Chapter + entry in dictionary

Figure 3. Type of Download

Figure 4. COUNTER Compliant Collections with Chapter-Level Usage Statistics

The most common type of report and level of statistics was BR2 and chapter level. However, among the 13, only 5 delivered this type of report. The following graph shows the five collections with comparable statistics.

Discussion

The comparison of the type of usage statistics of many e-book packages clearly showed the lack of a common standard for e-book statistics. This means that a relevant comparison cannot be done for the majority of the library's collections. They are not all COUNTER compliant and even those who are cannot be accurately compared, because they all have a different definition of "a download". Therefore, although libraries might think otherwise, the COUNTER standard is, in fact, not a real standard.

This means libraries have no real use of the usage statistics collected every year. The only thing that can be done is comparing each package from year

to year to see the trends, provided that the package doesn't differ a lot in content from year to year.

The experience from our library is that librarians know there are problems with comparing usage statistics, but they do the comparisons anyway. These usage statistics are all we have, so we use them even if they don't give us the answers we need.

The lack of comparable statistics leaves Acquisition Librarians confused, with no support in the renewal process, because there is no way of knowing how to compare "title" use and "page" use. It also makes it hard to establish value for money and to justify the purchases to our stakeholders. There is no way to establish a reasonable cost for an e-book download if there is no way of comparing. Without the detailed knowledge of the differences in definitions of "downloads", it would be an easy answer to cancel a package because of a high cost per

download, when there, in fact, might not be a real difference.

It was also surprising to notice how difficult it was to get information on the type of statistics delivered. This made it hard to get all the facts we needed to make an analysis. It would facilitate if vendors would clearly describe what kind of statistics they collect on their website. The new COUNTER Code of Practice will mean a small but important improvement. It will also be an eye-opener for librarians, and it will hopefully put more focus on this. Unfortunately, the problem of comparing statistics remains.

It is also important to remember that usage statistics is only one of many things to consider; qualitative aspects such as content, platform interface, and administrative and technical issues are also important in the renewal process. There are also new types of metrics, like altmetrics *Altmetrics: a manifesto*, n.d.) and the forthcoming COUNTER “usage factor” for e-journals (“Usage Factor,” n.d.).

This presentation has no easy solution to the problem, but it highlights the importance of a common standard for e-book statistics and higher awareness among librarians.

References

- Altmetrics: A manifesto. (n.d.). *altmetrics.org*. Retrieved from <http://altmetrics.org/manifesto/>
- COUNTER. (2012, January). Code of practice for books and reference works: Release 1. Register of vendors providing usage reports compliant with release 1 of the code of practice for books and reference works. Retrieved from http://www.projectcounter.org/R1/R1CV_Jan2012.pdf
- COUNTER. (2012, April). The COUNTER code of practice for e-resources: Release 4. Retrieved from <http://www.projectcounter.org/r4/COPR4.pdf>
- COUNTER. (n.d.). Code of practice. Retrieved from http://www.projectcounter.org/code_practice.html
- COUNTER. (n.d.). Usage factor. Retrieved from http://www.projectcounter.org/usage_factor.html
- Cox, J. (2008). Making sense of e-book usage data. *Journal of Electronic Resources Librarianship*, 19(3), 193–212. <http://dx.doi.org/10.1080/08963570802026278>
- ICOLC. (2006, October 4). Revised guidelines for statistical measures of usage of web-based information resources. Retrieved from <http://icolc.net/statement/revised-guidelines-statistical-measures-usage-web-based-information-resources>