Charleston Library Conference

Awash in E-Journal Data: What It Is, Where It Is, and What Can Be Done with It (Is It "Too Much" or "Not Enough?")

David P. Brennan
Penn State University, dbrennan@hmc.psu.edu

Nancy J. Butkovich
Penn State University, njb2@psu.edu

David P. Brennan and Nancy J. Butkovich, "Awash in E-Journal Data: What It Is, Where It Is, and What Can Be Done with It (Is It "Too Much" or "Not Enough?")" (2013). *Proceedings of the Charleston Library Conference.*
http://dx.doi.org/10.5703/1288284315323

Awash in E-Journal Data: What It Is, Where It Is, and What Can Be Done with It (Is It "Too Much" or "Not Enough?")

David P. Brennan, Assistant Librarian, Collection Development/Digital Resources Management, George T. Harrell Health Sciences Library, Penn State University College of Medicine

Nancy J. Butkovich, Associate Librarian and Head of Physical and Mathematical Sciences Library, Penn State University Libraries

Abstract

Libraries have been collecting e-journal data for some time. With the variety of data sources available, it is often difficult to determine their utility. This session explores the Penn State experience with navigating a number of data sources and their limitations and usefulness to advance library management objectives, as well as other institutional objectives. We will look at the COUNTER-compliant JR1 and JR1a data sets and standards (how they are derived and what issues there are with the data), as well as publishing and citation data (e.g., Web of Science) that shows faculty activity in publishing and their participation in editorial activities. What can we learn from these data? The objective of the session is to stimulate ideas and discussion for applications of the data we collect, ways of manipulating data sets, and including this process in overall analysis workflows.

Introduction

The collection of e-journal data has been ongoing for a number of years. However, the experience in many libraries has been mixed with regard to the actual utility of these data beyond basic measures. This presentation will focus primarily on usage reports (JR1 files) and citation data (including impact factors) to provide some additional measures of use that will inform collection development decisions.

The impetus for the current study was our impending negotiations with Elsevier for a new ScienceDirect license. A team was formed in January 2013, and was charged with gathering data to inform decisions related to Elsevier products. The final report was submitted in June 2013 and included contributions from many individuals and departments within the University Libraries.

Although the purpose of the original task force was to look specifically at everything Elsevier, two of us decided to expand the study to four publishers with the goal of examining the level of impact these publishers and Penn State (PSU) have on each other. We focused our efforts on two major areas. David Brennan conducted an analysis of usage data utilizing COUNTER JR1/1a reports and impact factor data, while Nancy Butkovich examined source and citation data obtained from the Web of Science (WOS) database. These discussions comprise the next two sections.

Inputs: The Universe of Usage Data and Some Uses for It

Source Data

The Elsevier Study Team report outlined usage data collected in a number of areas including a discussion of the limitations of some data sources. These areas included:

- Usage data—derived from COUNTER JR1 and 1a reports. Five million-plus hits over 5 years, 80% use from 20–23% of titles. Issues with these data include: JR1/1a—subscribed versus backfile titles, defining use, vague COUNTER standards, and potential inflation of numbers depending on platform design.
- Cost-per-use data (CPU)—These numbers are based on a simple calculation derived from the contract cost in a given year

http://dx.doi.org/10.5703/1288284315323

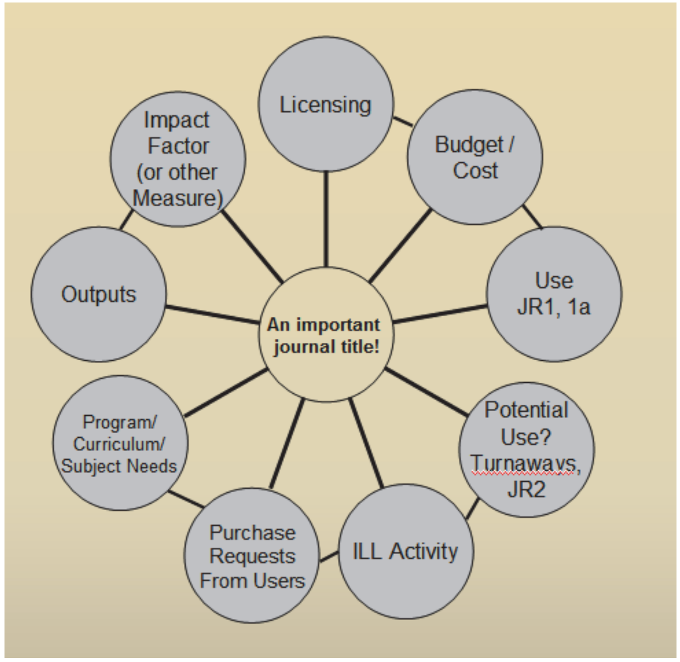

Figure 1.

against the aggregate use from the JR1. Issues include a lack of clear itemized title-by-title costs in some packages, as well as the issues above in relation to defining use.

- ILL data—Obtaining the raw data on borrowing requests was straightforward; however, determining which requests were for Elsevier titles required manual title matching. The intent of our analysis was to not only identify the extent of borrowing from Elsevier titles overall but to identify any titles for which we were exceeding the number of free requests allowed under CONTU Guidelines. Issues with ILL data on its own are that it is a prospective measure of potential use should a title be subscribed, rather than a measure, however vague, of actual use, and therefore only of value in adding titles, not removing them, and requiring a longer range of data for meaningful results.

Biases and Challenges

These data are part of the variable set for making collection development decisions. Other potential variables are shown in Figure 1.

Of these data points, a number of questions arise, particularly as to the utility of the information. Each data set has its limitations, and clearly there is no "unified" model that can take into account all of these variables. Thus, in the context of seeking funding for collection development, we chose two variables that can be easily explained to administrators: COUNTER JR1/1a usage data and Impact Factor (IF) data. Given the lack of good itemized CPU data for some packages and the complexities of the ILL data, these variables were not good examples for further study at this point (Although any of the variables in Figure 1 are targets for further analysis and investigation as to how they might be applied to collection development decisions.). Two examples can illustrate both the promises and pitfalls of this glut of data. The promise is to use analysis of publishing and citation data to show library value. The pitfall is not recognizing the issues involved with each data set.

Impact Factors and Subject Coverage

Publishing and citation data and, by extension, the use of IF (Haddow, 2007) can be used to influence collection decisions, inasmuch as there is the ability to swap titles in and out of packages, and given the limitations of IF (EASE, 2012). IF is a known quantity that is more familiar to library users than other measures and is commonly touted by vendors and publishers. Many journal landing pages prominently show their IF. Libraries do have a role in showing how to appropriately use the IF and other bibliometric data (Emory, 2013).

Even with large packages, there are still high impact titles that are outliers—recognizing these gaps and demonstrating current coverage is part of showing library value and meeting the needs of end users. The Harrell Library is using IF in conjunction with its library liaison program to demonstrate value (in relation to access to high IF titles) and responsiveness (in recognizing and responding to gaps in coverage). An example analysis is shown in Figure 2—the JCR report sorted by 5-year IF in psychiatry, and the coverage held by PSU (85%). As this analysis extends to all of the liaison areas, a clearer picture will emerge of collection strengths and needs. Preliminary data from this study indicate a high percentage of holdings of high-IF titles across all of the liaison areas (on the order of 78%).

Use—What Is It Really?

Returning to the Elsevier study report, a review of the usage data shows the dilemma of too much information. Use metrics have value, but again only inasmuch as there is the ability to swap titles in and out of a package when use data dictates. There is also the issue of the long tail (20% of titles accounting for 80% of use leaves 80% of titles with diminishing returns, even if none of them are truly "zero use"). However, use data is still an easily demonstrable measure to use, with some manipulation.

This necessary manipulation is not limited or specific to Elsevier—all vendors implement the COUNTER standards in different ways, and their platform design can influence the data. The COUNTER standards themselves can lead to confusion, particularly the JR1 versus JR1a backfile report. Figure 3 is an illustration of a title in which it is difficult to determine where use is occurring along the spectrum of holdings. COUNTER Journal Report 5 does help to alleviate this confusion, but

Journals from: subject categories PSYCHIATRY; PSYCHOLOGY
Sorted by: 5 Year Impact Factor

Penn State Analysis (as of 10/24/13) | **N=20 current= 17 (85%)** | Not current: 3 (15%) | No access: 0 (0%)

Rank	Abbreviated Journal Title (linked to journal information)	ISSN	JCR Data: Total Cites	Impact Factor	5-Year Impact Factor	Immediacy Index	Articles	Cited Half-life	Eigenfactor® Metrics: Eigenfactor® Score	Article Influence® Score	Search SS	Psu Access?
1	ANNU REV PSYCHOL	0066-4308	10635	15.265	26.624	4.818	22	9.5	0.0281	12.87	http://mk3kg9qj3t.search.:	y
2	PSYCHOL BULL	0033-2909	30814	15.575	19.676	2.316	38	>10.0	0.03238	9.087	http://mk3kg9qj3t.search.:	y
3	ARCH GEN PSYCHIAT	0003-990X	37412	13.772	14.466	2.488	125	>10.0	0.0613	5.625	http://mk3kg9qj3t.search.:	1960-2012
4	AM J PSYCHIAT	0002-953X	42730	14.721	14.396	2.574	108	>10.0	0.06219	5.504	http://mk3kg9qj3t.search.:	1 yr embargo
5	ANNU REV CLIN PSYCHO	1548-5943	2043	12.422	14.073	1.444	18	4.7	0.00962	5.738	http://mk3kg9qj3t.search.:	y
6	MOL PSYCHIATR	1359-4184	12686	14.897	13.985	3.879	116	5.3	0.04215	5.06	http://mk3kg9qj3t.search.:	y
7	PSYCHOL REV	0033-295X	21374	9.797	11.342	1.222	36	>10.0	0.02133	6.137	http://mk3kg9qj3t.search.:	y
8	BIOL PSYCHIAT	0006-3223	37840	9.247	9.773	2.131	251	7	0.09517	3.471	http://mk3kg9qj3t.search.:	y
9	SCHIZOPHRENIA BULL	0586-7614	11606	8.486	8.934	1.699	136	6.3	0.03152	2.988	http://mk3kg9qj3t.search.:	y
10	NEUROPSYCHOPHARMACOL	0893-133X	19882	8.678	7.796	1.802	248	6	0.05716	2.674	http://mk3kg9qj3t.search.:	y
11	J AM ACAD CHILD PSY	0890-8567	16470	6.97	7.148	2.01	96	>10.0	0.02751	2.621	http://mk3kg9qj3t.search.:	y
12	BRIT J PSYCHIAT	0007-1250	21472	6.606	7.112	1.872	117	>10.0	0.03229	2.416	http://mk3kg9qj3t.search.:	y
13	SOC COGN AFFECT NEUR	1749-5016	1980	5.042	6.78	1.347	101	3.7	0.01057	2.565	http://mk3kg9qj3t.search.:	y
14	J PSYCHIATR NEUROSCI	1180-4882	2253	6.242	6.473	1.647	34	5.2	0.0069	2.148	http://mk3kg9qj3t.search.:	y
15	WORLD PSYCHIATRY	1723-8617	1193	8.974	6.413	2.429	21	5	0.00373	1.962	http://mk3kg9qj3t.search.:	free only
16	J CHILD PSYCHOL PSYC	0021-9630	12841	5.422	6.235	1.025	121	8.5	0.02604	2.29	http://mk3kg9qj3t.search.:	y
17	PSYCHOL MED	0033-2917	16680	5.587	6.148	0.913	230	8.6	0.03551	2.22	http://mk3kg9qj3t.search.:	y
18	PSYCHONEUROENDOCRINO	0306-4530	9603	5.137	5.926	1.091	198	5.9	0.02404	1.758	http://mk3kg9qj3t.search.:	y
19	PSYCHOTHER PSYCHOSOM	0033-3190	2761	7.23	5.825	1.269	26	7.4	0.00517	1.567	http://mk3kg9qj3t.search.:	y
20	J CLIN PSYCHIAT	0160-6689	17639	5.812	5.639	0.726	157	8.1	0.03305	1.772	http://mk3kg9qj3t.search.:	y

Figure 2.

Biochemical systematics and ecology (0305-1978)

-from 1973 to 1994 in ScienceDirect Agricultural & Biological Sciences Backfile and ScienceDirect Environmental Science Backfile

-from 1974 to 1994 in ScienceDirect Biochemistry, Genetics & Molecular Biology Backfile

-from 06/28/1974 to 2009 in ScienceDirect Journals

Aggregate use (JR1):

1	Title	ISSN	2008 Total	2009 Total	2010 Total	2011 Total	2012 total	5 year total
267	Biochemical and molecular medicine	1077-3150					4	4
268	Biochemical education	0307-4412	23	24	19	26	72	164
269	Biochemical engineering journal	1369-703X	662	467	516	452	451	2548
270	Biochemical medicine	0006-2944					16	16
271	Biochemical medicine and metabolic biology	0885-4505					6	6
272	Biochemical pharmacology	0006-2952	1391	1495	755	1841	1762	7244
273	Biochemical systematics and ecology	0305-1978	174	164	113	107	122	680

Journal	Online ISSN	Month	Mar-12	Apr-12	May-12	Jun-12	Jul-12	Aug-12	Sep-12	Oct-12	Nov-12	Dec-12	Jan-13	Feb-13	Mar-13
Biochemical Engineering Journal	1369-703X	0	24	45	58	44	35	31	28	33	59	38	41	41	32
Biochemical Medicine	0006-2944	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Biochemical Medicine and Metabolic Biology	0885-4505	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Biochemical Pharmacology	0006-2952	0	180	142	187	146	119	131	118	145	156	105	168	263	138
Biochemical Systematics and Ecology	0305-1978	0	6	13	19	5	11	2	6	3	7	2	2	10	5

Backfile use (JR1a):

Journal	Online ISSN	Month	Mar-12	Apr-12	May-12	Jun-12	Jul-12	Aug-12	Sep-12	Oct-12	Nov-12	Dec-12	Jan-13	Feb-13	Mar-13
Biochemical Engineering Journal	1369-703X	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Biochemical Medicine	0006-2944	0	0	5	1	0	0	1	2	1	2	2	2	1	1
Biochemical Medicine and Metabolic Biology	0885-4505	0	0	0	1	0	0	2	0	0	0	0	0	0	0
Biochemical Pharmacology	0006-2952	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Biochemical Systematics and Ecology	0305-1978	0	0	3	4	4	1	1	5	4	9	3	0	2	2

Figure 3.

not all vendors supply this report, and it requires significant further data analysis to parse these data in relation to the JR1 and 1a reports to distill a useful measure.

Including or excluding these data can be debated (Bucknell, 2012). The proper analysis of "use" is of greater concern, with the well-known issues of the COUNTER standards and their implementation having an impact in how useful these data can be (Welker, 2012). Even the question of what constitutes "use" is a subject for discussion (Nicolson-Guest & Macdonald, 2013). Yet this measure with all of its limitations does present a relative picture of use, and, like IF, one that is easily explained to administrators and library users as a justification for collection decisions.

Conclusions

Obviously, there can be no model that cleanly and easily takes all of the variables into account—there are simply too many variations in each institution as how the data are collected as well as that which is vendor-supplied. Each library therefore must choose variables for which their data are useful, clearly articulate the limitations of the data, and be consistent in how it is applied to collection decisions.

Outputs: Publishers By the Numbers

The second phase of this study examines source and citation data obtained from the WOS database in order to identify key publishers for PSU in terms of where PSU authors publish and what they cite. Examining the results from four publishers allowed us to anonymize our data so that actual values involving cost and use could be presented. Other questions will also be asked; however, the key thing to keep in mind is that what is important is the method, not the actual results of this survey. All of the input numbers—subscription costs, views/downloads, citation data, etc.—will vary for each institution. While results from another large Land-Grant institution may be similar, those from a small liberal arts college will almost certainly be radically different.

Source Data

The first question involved source data—specifically, with what publishers did PSU authors publish their research in 2011? In order to answer

Publisher	Number of articles per publisher	Percentage of total articles (n=6,928)
Elsevier (and subsidiaries)	1,290	18.6%
Wiley (and subsidiaries)	774	11.2%
Springer (and subsidiaries)	453	6.5%
American Chemical Society	401	5.8%

Table 1. Top Four Publishers of PSU-Authored Papers in 2011

this, the WOS database from Thompson-Reuters was used. It is a multidisciplinary indexing and abstracting platform that also includes cited reference data back to 1900. This latter feature was very helpful when answering some of the questions posed in this paper.

Within WOS, the three classic databases: Science Citation Index (SCI), Social Science Citation Index (SSCI), and Arts and Humanities Citation Index (AHCI), were of most interest. Because our original focus was to examine a specific e-journal package, the suite of conference proceedings and book citation indexes that are also available through WOS were excluded. This decision will have an impact later in this analysis and will be discussed in the biases section. Within these three citation indexes, the phrase "Penn State" was searched in the address field, and the search was not limited by PSU campus.

Only articles published in 2011 were included. This was a matter of practicality; at the time the data were collected (early 2013) this was the most recent complete year available. There was a second practical consideration—the 2011 search results produced approximately 750,000 lines of Excel output that had to be manually cleaned, and there was insufficient time available to clean a second similarly sized data set.

To determine the publisher distribution, the 6,928 records retrieved in this search were sorted and the data from the publisher (PU) field were extracted. After studying the resulting list and matching subsidiaries to parent companies, the top publishers of PSU-authored papers in 2011 became obvious. The top four are shown in Table 1.

PSU is a Land-Grant institution with a heavy science, technology, engineering, and mathematics (STEM) focus. Because PSU also has a College of Medicine and a College of Agriculture, these areas are also included in the definition of STEM used in this paper, so these results are not particularly surprising. That the American Chemical Society (ACS) was in the fourth slot was interesting, since it publishes a very limited number of journals in a very narrow range of subjects when compared to the other three. However, chemistry is a fundamental discipline for all of STEM, as well as medicine, agriculture, and their allied fields, and PSU has a very active Chemistry Department, so it makes sense that ACS would be near the top of the publisher list.

Still, there is no denying that Elsevier and its subsidiaries published almost as many articles by PSU authors in 2011 as the next three combined. Looked at another way, nearly one paper in five was published in an Elsevier journal, and when their totals are combined, these four publishers accounted over 42% of all PSU authored papers. These publishers will be the focus of the remainder of this section.

Biases and Challenges

It would be foolish to think that the results of this section are unbiased when there is bias built into every step of the analysis, and all of those biases are in favor of STEM. These are outlined below.

- We were initially interested in a STEM-heavy multidisciplinary publisher (Elsevier), so the publisher and citation data were obtained from a STEM-heavy multidisciplinary database—Web of Science.
- Our particular area of interest with regard to Elsevier revolved around e-journals, so we focused on the three original citation indexes in Web of Science that indexed only journals; the books and conference

Publisher	Number of titles	Percentage for each publisher
Elsevier (and subsidiaries)	1,777	36.8%
Springer (and subsidiaries)	1,534	31.7%
Wiley (and subsidiaries)	1,471	30.4%
American Chemical Society	51	1.1%
Total	4,833	100.0%

Table 2. Numbers of Journal Titles on the JR1 Lists for the Four Publishers

proceedings citation indexes were deliberately excluded.

- Only PSU data were included, and PSU is a large Land-Grant institution with a heavy STEM focus.
- Only data for one year were examined. Although this probably did not add a significant bias to the results, it is possible that it could have. In other words, it is possible that 2011 could have been an anomalous year.

In addition to the aforementioned time-consuming cleanup of the Web of Science data, there were other challenges that had to be addressed. With Web of Science, there is much inconsistency in the journal abbreviations used in the cited references. In a few cases it was impossible to determine which of two journals the abbreviation represented; these references were not counted.

The JR1 files also presented challenges. The raw files required massive amounts of cleanup; fortunately, when we expanded this study, we were able to take advantage of files that had been cleaned by the members of our Serials Department. An additional hurdle that had to be overcome had to manually compare the full titles used in the JR1 files with the journal abbreviations in the WOS citation data. Finally, some of the data used in the various analyses were covered by confidentiality clauses. In order to be able to present actual numbers, some of the values in this study are based on numbers aggregated from all four target publishers.

Citation Analysis

Citation analysis has a long and rich history in collection development and management; the first true citation study was published in 1927 and consisted of an analysis of the references cited in selected chemistry journals (Gross and Gross, 1927). The principle data source for the rest of this section will be the cited references to the 6,928 PSU-authored papers retrieved in the search of the three citation indexes mentioned earlier. The JR1 title lists were used to identify the titles in the packages for the four publishers.

The first part of this citation study looks at the cited titles. These are the specific titles that were cited by PSU authors, and each journal was counted only once. What was of interest was not the journals themselves but rather the comparison of the lists of cited titles with the list of journals from the JR1 files for each publisher. The numbers of titles on the JR1 lists are shown in Table 2. Each publisher has roughly a third of the titles, except ACS, which contributes just over one percent. From this point forward, the results will be aggregated because of potential confidentiality clause concerns. This combined publisher group will be referred to as the “Big Four” in the rest of this paper.

A question that comes immediately to mind is “how many of these 4,833 titles were cited?” The answer varies by citation index:

- Science Citation Index—3,169 titles were cited (65.6%)
- Social Science Citation Index—1,430 titles were cited (29.6%)
- Arts and Humanities Citation Index—205 titles were cited (4.2%)

However, some titles were cited in more than one citation index, and 1,286 (26.6%) of the titles were not cited in any of the three indexes. Looking at the number of titles that are unique to just one citation index,

Citation Index	Total references cited by PSU authors	Big Four journal titles cited by PSU authors (percentage of total for this index)
Science	183,393	63,572 (34.7%)
Social Science	58,802	14,364 (24.4%)
Arts and Humanities	6,992	568 (8.1%)
Total	249,187	78,504 (31.5%)

Table 3. Distribution of Cited References by Citation Index

- 1,789 titles were cited only in Science Citation Index (37.0%)
- 298 titles were cited only in Social Science Citation Index (6.2%)
- 20 titles were cited only in Arts and Humanities Citation Index (0.4%)

The remaining titles were either cited in multiple databases or were not cited at all. Looking at these data from a different perspective, 73.4% of the titles were cited in at least one of the citation indexes, and roughly two-thirds of the cited titles were STEM titles.

Another question asks "what is the frequency with which the journals published by the Big Four are cited by PSU authors?" To answer this question, the data set from the previous question was used; however, the data set included all citations, not just the journals that were of interest. Table 3 shows the distribution of all citations by citation index. It also shows the results for the journals published by the Big Four.

Nearly three quarters of all the references cited by PSU authors in 2011 were in articles indexed in Science Citation Index. And of all the references cited, nearly a third was to journals published by the Big Four.

Other Calculations

To this point these data establish the importance of these four publishers to PSU authors, particularly related to where they publish and who publishes the journals that they cite. However, the citation and JR1 data allows other questions to be asked:

- What is the cost per use of these publications?
- How many articles are viewed for every article cited?
- What is the cost per citation?

Librarians have calculated cost per use since long before journals went electronic. Now the calculation is usually cost per view, but the goal is the same: determining how much the institution pays each time someone opens a document. There are potential problems with what constitutes a "view" in the JR1 data, which David discussed earlier. For the purpose of this example, the assumption is that there is no overlap between the HTML views and the PDF views for each publisher. This assumption is necessary because even if there is overlap, there is no way of determining how much there is.

In this example, the cost of the e-journal packages from the Big Four in 2011 was $4,749,866.67, and there 1,790,333 views of articles in Big Four journals. Simple division produces an average cost per view of $2.65 for the Big Four, which is actually a very good value compared to document delivery or interlibrary loan costs. Unfortunately, this is an aggregated number; the range of values for each publisher varied.

Citations are the end product; how many articles are viewed to get one citation? There are many articles that have examined the number of articles read by researchers, such as the one by King et al. (2006). However, information on the number of read (views) per citation is sparse. For example, Kurtz et al. (2005) used data from the Astrophysical Data System (ADS) and determined that a given paper in ADS was read about 20 times for each time that it was cited.

It is important to emphasize that this calculation by Kurtz et al. (2005) was based on data for the same article on both sides of the equation. Our data did not allow for that level of specificity. Furthermore, the PSUJR1 data included "views" for all kinds of uses and from all levels of users, including students, while the WOS citation data came from published output by researchers, and there was no way to correlate the data for the articles that were viewed with those of the articles that were cited. However, as long as the limitations of the data are understood, a rough value can be calculated. In this case, 1,790,333 items published in Big Four journals were viewed and Big Four journals were cited 78,504 times. Again, simple division produces a value of roughly 22.8 views for each citation.

The last value of interest was cost per citation. This value is subject to all the biases and limitations mentioned elsewhere in this paper, but in aggregate, the Big Four produced a value of $60.50 per citation. As with all the other results, this number varied widely from one publisher to another.

Final Thoughts

What is, perhaps, the most useful "take-away" from this is the method used rather than the actual results. Those numbers will vary between institutions because the variables will be different. The results can also vary depending on the sources of the data used. For example, if data are included from other products such as Scopus, Google Scholar, or SciFinder Scholar, then the results could change. However, as more data sources are incorporated into the study, more effort will be needed to remove duplicates. There is also the question of whether (and how) to include mentions in blogs, Twitter, and other social media that are measured by companies, such as Altmetrics. Where does one stop?

Looking at the publisher-level citation data can be useful, particularly since so many publishers are bundling journals into packages. Most packages have some percentage of titles that are of low importance to our institution. Therefore, if the values for the titles that are of high interest can be compared against the numbers for the whole package, then those results could become part of a decision process to determine if it is worth retaining the e-journal package or if it is more cost-effective to switch to licensing the journals on a title-by-title basis.

In an analysis like this, it is easier to identify the top level publishers than those that are in the bottom tier, particularly if citations are the basis for identification; bottom tier in this context does not necessarily mean bottom quality, particularly in a study with the biases that this one has. For example, the arts and humanities are going to be disproportionally represented in the bottom tier based on citation counts. Adding additional values, such as cost per view, views per citation, and cost per citation, to an evaluation may, in some situations, help ameliorate this imbalance. So would including data from sources that are not dependent on citations, such as interlibrary loan and document delivery data.

Acknowledgments

This was truly a team effort, and many people made valuable contributions at different stages of the project. These include the following people at Penn State University Park: Lisa German, Associate Dean for Collections, Information and Access Services; Linda Musser, Head of the Earth and Mineral Sciences Library; Barbara Coopey, Assistant Head of Access Services; Ann Thompson, Information Resources and Services Supervisor-Manager; Robert Alan, Head of Serials and Acquisitions Services; Jaime Jamison, Electronic Resources Specialist; and Alan Shay, Data Analyst Assessment. N.B. also wishes to thank Dana Roth (Chemistry Librarian, Caltech Library, California Institute of Technology), who provided a crucial literature reference. And finally, we also want to give a special thanks to all the members of the Serials Department for their efforts at cleaning up the raw JR1 data files; that was a massive undertaking, and we really appreciate your efforts!

References

Bucknell, T. (2012, August 02). Garbage in, gospel out: Twelve reasons why librarians should not accept cost per download figures at face value. *The Serials Librarian, 63*(2), 192–212. http://dx.doi.org/10.1080/0361526x.2012.680687

European Association of Science Editors. (2007, November). *The EASE statement on inappropriate use of impact factors*. Retrieved from http://www.ease.org.uk/publications/impact-factor-statement

Gross, P. L. K., & Gross, E. M. (1927). College libraries and chemical education. *Science*, *66*(1713), 385–389. Retrieved from http://www.jstor.org/stable/1651803

Haddow, G. (2013). Evidence summary: Level 1 COUNTER compliant vendor statistics are a reliable measure of for cost benefit and reporting. *ALIA Information Online 2013: Conference Proceedings.*

King, D. W., Tenopir, C., & Clarke M. (2006). Measuring total reading of journal articles. *D-Lib Magazine*, *12*(10). Retrieved from http://www.dlib.org/dlib/october06/king/10king.html

Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Murray, S. S., Martinmbeau, N., & Elwell, B. (2005). The bibliometric properties of article readership information. *Journal of the American Society for Information Science and Technology*, *56*(2), 111–128. http://dx.doi.org/10.1002/asi.20096

Robert W. Woodruff Library. (2014, January 10). *Impact factors and citation analysis.* Retrieved from http://guides.main.library.emory.edu/citationanalysis

ScienceDirect. (2009, December). Description of the ScienceDirect customer usage reports. *Elsevier*. Retrieved from http://usagereports.elsevier.com/Report_Descriptions/sd_report_description.pdf

Walker, J. (2012).Counting on COUNTER: The current state of e-resource usage data in libraries. *Computers in Libraries, 32*(9). Retrieved from http://infotoday.com/cilmag/nov12/Welker--Counting-on-Counter.shtml