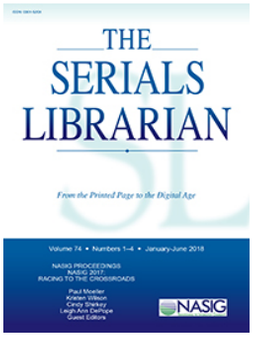


The Serials Librarian
From the Printed Page to the Digital Age

ISSN: 0361-526X (Print) 1541-1095 (Online) Journal homepage: https://www.tandfonline.com/loi/wser20

Beyond COUNTER-Compliant: Ways to Assess E-Resources Reporting Tools

Kelly Marie Blanchat

To cite this article: Kelly Marie Blanchat (2018) Beyond COUNTER-Compliant: Ways to Assess E-Resources Reporting Tools, The Serials Librarian, 74:1-4, 87-93, DOI: 10.1080/0361526X.2018.1428464

To link to this article: https://doi.org/10.1080/0361526X.2018.1428464

THE SERIALS LIBRARIAN
2018, VOL. 74, NOS. 1–4, 87–93
https://doi.org/10.1080/0361526X.2018.1428464

Routledge
Taylor & Francis Group

Beyond COUNTER-Compliant: Ways to Assess E-Resources Reporting Tools

Kelly Marie Blanchat

Presenter

ABSTRACT

This article describes the implementation of ProQuest's usage assessment tool, 360 COUNTER [Counting Online Usage of Networked Electronic Resources], at Yale University Library, and the subsequent methods used by Yale's E-Resources Group to assess the accuracy of its reporting output compared to the original data seen in COUNTER reports. The assessment of the usage reporting tool took place in multiple phases, and involved working closely with ProQuest to understand how COUNTER data were being normalized behind the scenes for consolidation to the Authority Title. The assessment method used to evaluate the reporting accuracy can be adapted and modified for other usage assessment tools. This article also discusses how Yale has begun to modify the use of 360 COUNTER over time, using it in combination with a home-grown system built within Tableau.

KEYWORDS

COUNTER; e-resources; usage statistics; assessment; collection management

In the spring of 2015, the Electronic Resources Group at Yale University Library began implementing ProQuest 360 COUNTER [Counting Online Usage of Networked Electronic Resources].[1] The 360 COUNTER system is made up of two parts: 360 COUNTER itself and Intota Assessment. 360 COUNTER harvests and archives COUNTER reports, and Intota Assessment consolidates usage across both time and providers (see Figure 1). For the E-Resources Group the primary goal for implementation was to outsource usage statistics harvesting so that valuable staff time could be allocated elsewhere. Prior to 2015, library staff had saved usage reports as Excel spreadsheets every year, twice per year, for hundreds of content providers. This manual process was extremely time consuming and also did not provide the ability to easily compare, analyze, and distribute usage reports. The 360 COUNTER system would also provide the opportunity to enhance reporting metrics with consolidated views of usage within Intota Assessment.

In September 2015, 360 COUNTER delivered the first set of usage statistics for the January to June 2015 reporting period, and the E-Resources Group began previewing the consolidated reports from Intota Assessment. Although informal, this initial assessment discovered missing titles from consolidated reports; titles that were known to have been in the raw data harvested from 360 COUNTER. Most notably, the Oxford English Dictionary, published by Oxford University Press, had disappeared from consolidated Book Report 2 (BR2) reporting. Due to the importance and value of the Oxford English Dictionary, the E-Resources Group moved from informal testing to systematic assessment of consolidated reports. The purpose of the systematic analysis was to understand how much data were missing from Oxford University Press on its own, in the hopes that a single data handling error could be discovered and corrected.

CONTACT Kelly Marie Blanchat kelly.blanchat@yale.edu 110 Wall Street, New Haven, Connecticut 06520, USA.
At the time of doing this work and presenting at NASIG, author's title was E-Resources Support Librarian at Yale University Library.

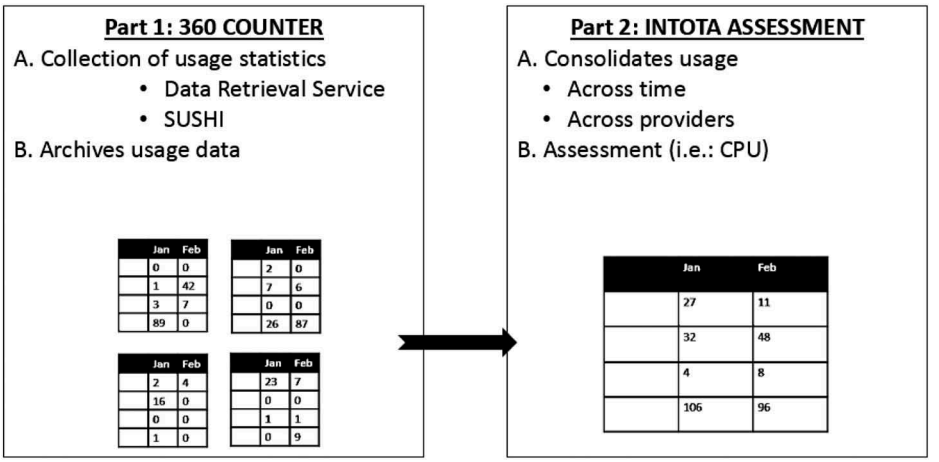


Figure 1. How 360 COUNTER and Intota Assesment work together for usage retrieval, storage, and consolidation.

Phase one

To evaluate the accuracy of usage consolidation the E-Resources Group needed to compare an original BR2 COUNTER report against the output from Intota Assessment. For this task, the group chose the Excel VLOOKUP formula for a title-level analysis. Although a manual process, VLOOKUP can isolate where data points match between two sources. The International Standard Serial Number (ISSN) and International Standard Book Number (ISBN) fields was used as the initial point of comparison; however, because the VLOOKUP formula only works where data are available, if an ISSN is missing or variant, it will return no results (or "N/A"). Although not perfect, this method was a good starting point for title-level assessment of BR2 consolidation for Oxford University Press.

The initial analysis found a total of 342 missing titles from Intota Assessment's consolidated BR2 from Oxford University Press, which represented a total loss of 29,170 uses. At the granular title level, the E-Resources Group discovered the following major categories of errors in the consolidated report from Intota Assessment:

(1) Duplicate titles with duplicate ISBNs were deleted from the Intota Assessment report; because both duplicate titles were removed from the report, the reporting period total represented a decrease from the original COUNTER report (seen in 20 unique titles).
(2) Usage from titles with the same name but different ISBNs—distinct editions—were combined into a single entry, thereby decreasing the total title count (seen in 34 titles).
(3) Intota's consolidated reports displayed variant ISBNs from the original COUNTER reports and a comparison could not be made due to a mismatch in the VLOOKUP data point (seen in 63 titles).

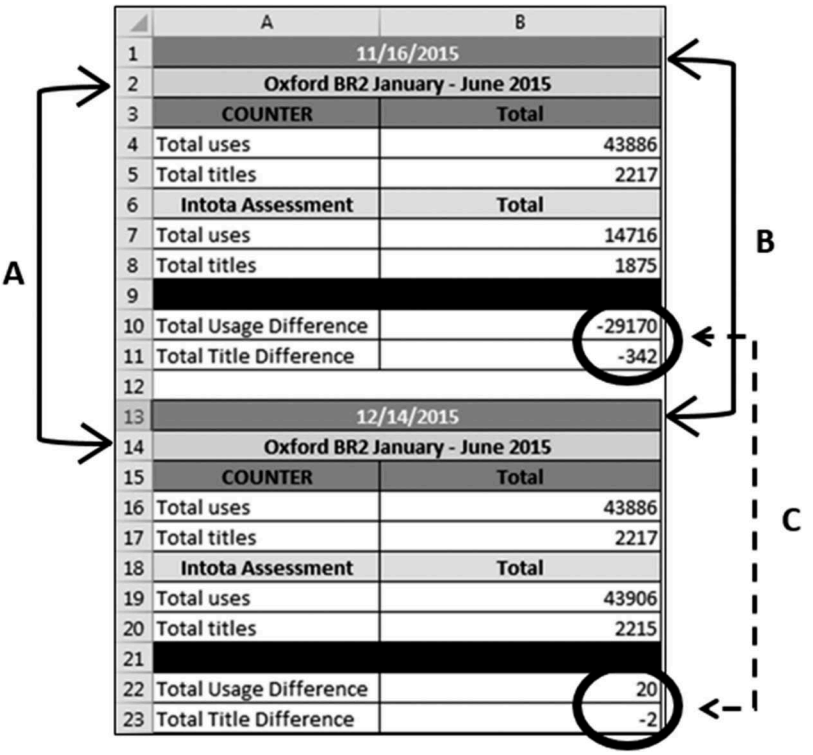
Although these categories and title counts did not address all the reasons behind the full 342 missing titles, they were a good start, and in November 2015 the E-Resources Group provided ProQuest with a title-level error ticket. In December 2015, ProQuest confirmed the findings and implemented a bug fix, after which only two titles and 20 uses remained unaccounted for in BR2 consolidated reports. These results were favorable—as two titles and 20 uses would not impact decision making—and demonstrated ProQuest's flexibility and understanding of the E-Resource Group's concerns. However, due to the volume of content providers and report types in Yale's instance of 360 COUNTER, Yale's workflow to identify system errors would not be sustainable to continue verifying the accuracy of other

consolidated reports. It was also not clear that a bug fix had been universally implemented for Intota Assessment.

Phase two

As a next step the E-Resources Group opted to move away from a detailed, title-level assessment in favor of high-level data collection, or Phase 2. By focusing on totals for title counts and usage between the original COUNTER reports and the consolidated reports—and by not collecting title-level information—the system output quality could still be assessed to ensure that consolidated reports had a direct relationship to original COUNTER data. The group created a generic assessment worksheet in Microsoft Excel to allow for more data to be collected over a shorter period of time on a variety of different report types and providers. Although an Excel worksheet was being used again for Phase 2, this worksheet had a built-in subtract formula to calculate differences in totals between COUNTER reports and consolidated reports.[2]

Figure 2 demonstrates the Phase 2 data collection in Excel for Oxford University Press BR2, where a subtract formula is built-in to cells B10, B11, B22, and B23. From the formula results seen in these cells, a positive number indicates an increase in titles and/or usage in a consolidated report compared with the original COUNTER data. A negative number indicates that the consolidated report is missing titles and/or usage contained in the original COUNTER report. An ideal calculation will return a number as close to "0" as possible (with the understanding that title history can affect consolidation), indicating no or very little change between COUNTER data and consolidation. This



	A	B
1	11/16/2015	
2	Oxford BR2 January - June 2015	
3	COUNTER	Total
4	Total uses	43886
5	Total titles	2217
6	Intota Assessment	Total
7	Total uses	14716
8	Total titles	1875
9		
10	Total Usage Difference	-29170
11	Total Title Difference	-342
12		
13	12/14/2015	
14	Oxford BR2 January - June 2015	
15	COUNTER	Total
16	Total uses	43886
17	Total titles	2217
18	Intota Assessment	Total
19	Total uses	43906
20	Total titles	2215
21		
22	Total Usage Difference	20
23	Total Title Difference	-2

Figure 2. The phase 2 collection worksheet showing the differences before and after a support ticket was submitted to ProQuest, where A is a constant (report type), B is a variable (date of data collection), and C is the result, representing system accuracy at a point in time.

spreadsheet was used to proactively identify the accuracy of the consolidation process, and was also eventually used to reactively evaluate the success of regular system updates and bug fixes.

The next support ticket that the E-Resources Group submitted to ProQuest included discrepancy data for eight content providers for both JR1 and BR2 report types, to which ProQuest responded with more granular information about *how* data are consolidated behind the scenes. Essentially, the 360 COUNTER system is connected to the ProQuest knowledgebase, Serials Solutions, and all COUNTER data harvested by the system are normalized behind the scenes to the Authority Title. For the E-Resources Group, this normalization process made sense because usage consolidation enables a uniform view of content available from multiple sources, whether from an aggregator database or from another publisher. Due to the complexity of serials—and e-books for that matter—title normalization should, in theory, remove inconsistencies. However, applying title normalization can be a tricky business especially for multiple format types and if normalization match points (such as ISSN and ISBNs) are missing.

ProQuest also walked the E-Resources Group through the problem tickets submitted from Phase 2, and explained how the underlying title normalization process could sometimes inadvertently inflate or deflate overall title counts and usage totals seen within consolidated reports. Because normalization relies on ISSN and ISBN data, the system needs to have a way to handle problems with those data—a concept similar to the problems found by the E-Resources Group during Phase 1 when ISSN and ISBN data were missing. For example, if an ISSN is duplicated in a COUNTER report, but the title name is distinct (such as when a title name includes a specific date range) the consolidated system picks a version to display and omits the rest (deflating usage). Yale University Library subscribes to the title "Annali di Mathematica Pura ed Applicata" from SpringerNature. This title has four different iterations, as follows: Annali di Mathematica Pura ed Applicata (1858–1865), Annali di Mathematica Pura ed Applicata (1867–1897), Annali di Mathematica Pura ed Applicata (1898–1922), and Annali di Mathematica Pura ed Applicata (1923-) had.

Because "Annali di Mathematica Pura ed Applicata" has distinct name variations and duplicated ISSNs over time, Intota Assessment pulled just one version for consolidation. In this case Intota Assessment happened to select the version of the title with 0 uses. As a result the consolidated report displayed all iterations of the title as having 0 uses for reporting period 2015, when in fact the version "(1923 -)" had 847 total uses (see Figure 3). Depending on the size of a library's collection, a loss of 847 uses can make a significant impact on decision making, and this type of title normalization—if repeated throughout the consolidated system for multiple reports and providers—can make a big impact on any library's understanding of its collection use overall.

Journal	Print ISSN	Online ISS	Reporting Period Total	Reporting Period HTML	Reporting Period PDF
Total for all journals			3069969	19429	2977464
ANNALI DELL'UNIVERSITA' DI FERRARA	0430-3202	1827-1510	411	0	409
Annali di Matematica Pura ed Applicata (1858-1865)		0373-3114	0	0	0
Annali di Matematica Pura ed Applicata (1867-1897)		0373-3114	0	0	0
Annali di Matematica Pura ed Applicata (1898-1922)		0373-3114	0	0	0
Annali di Matematica Pura ed Applicata (1923 -)	0373-3114	1618-1891	847	0	830

					2015		
Title	Authority Title	Provider	Print ISSN	Online ISSN	Use	Cost	Cost Per Use
Grand Total					0		
Annali di matematica pura ed applicata, Annali di Matematica Pura ed Applicata (1923 -), Annali di	Annali di matematica pura ed applicata	Springer-Verlag	0373-3114	1618-1891, 0373-3114	0		

Figure 3. The top spreadsheet shows JR1 the usage from SpringerNature COUNTER report for the journal "Annali di Mathematica Pura ed Applicata," while the bottom chart shows the usage after title normalization and consolidation within Intota Assessment, according to the rule for titles with duplicate ISSNs with variant names.

Phase 2 of the E-Resource Group's assessment of data handling between 360 COUNTER and Intota Assessment helped to reveal what was happening behind the scenes. In short, the process is complex because usage data are connected to the Serials Solutions knowledge base, and the knowledge base is already a complex place on its own. The resulting considerations from both the title-level (Phase 1) and high-level (Phase 2) assessment workflows included:

(1) Which assessment results are caused by the title normalization?
(2) Which assessment results should trigger a request for a bug fix?
(3) Which assessment results are a result of incorrect COUNTER data (and therefore are not the fault of the system)?

Phase three

The E-Resources Group immediately saw clear benefits from harvesting usage data in 360 COUNTER. However, after an extensive testing period, the impact of title normalization on reporting still contained too many unknowns for Intota Assessment to be used in making collection decisions. In summary, 360 COUNTER was kept in place to harvest and archive usage statistics and in August 2016 planning for an alternative method of usage assessment—Phase 3—began.

Working with the Library's Assessment Librarian, the E-Resources Group began piloting a method for home-grown usage assessment by transforming COUNTER reports into a data source to consolidate within Tableau, an interactive data visualization program. By transitioning usage consolidation into Tableau, the E-Resources Group could take ownership over the consolidation process by removing any guess-work regarding data handling. Tableau was also already being used at Yale University Library to report ILS financial data, and there were opportunities to match COUNTER data to integrated library system (ILS) reports behind the scenes for additional usage reporting enhancement.

By utilizing 360 COUNTER, the E-Resources Group was also taking on the burden of not using title normalization, with the understanding that there might be some data errors either due to missing ISSN and ISBNs or name changes over time. This risk was worth it because, moving forward, Tableau would offer robust data visualization opportunities which the E-Resources Group could reasonably use to create a method to identify data errors that could affect cost-per-use calculations.

The Phase 3 workflow officially moved the E-Resources Group out of assessment and into production. It involved retrieving reports harvested by 360 COUNTER and transforming them into a data source by applying the Tableau plugin for Microsoft Excel. Formatting COUNTER reports as a data source essentially moved the data from a horizontal display to a vertical display, so that each title in the report has multiple rows of data that correspond to an individual month of usage, instead of a single line entry for the full reporting period across columns (see Figure 4). Once the COUNTER data had been set up as a data source, files were ingested into an Access database for eventual staging in Tableau for visualization and analysis.

Phase 3 was piloted for the collection and analysis of the library's Association of Research Libraries (ARL) statistics, which provided a proof of concept on high-impact resources. The pilot demonstrated immediate success in its ability to quickly and easily switch Tableau data from calendar year to fiscal year (a reporting structure that ARL requires but is notoriously difficult to achieve). The pilot also resolved errors with missing titles and dropped usage seen from the consolidated reports in Intota Assessment. For instance, the previous consolidated version of the 2015 SpringerNature JR1 had the 847 missing uses (see Figure 3). In Tableau it accurately displayed multiple years of usage, in addition to the total 2015 usage for all four iterations of "Annali di Mathematica Pura ed Applicata" (see Figure 5).

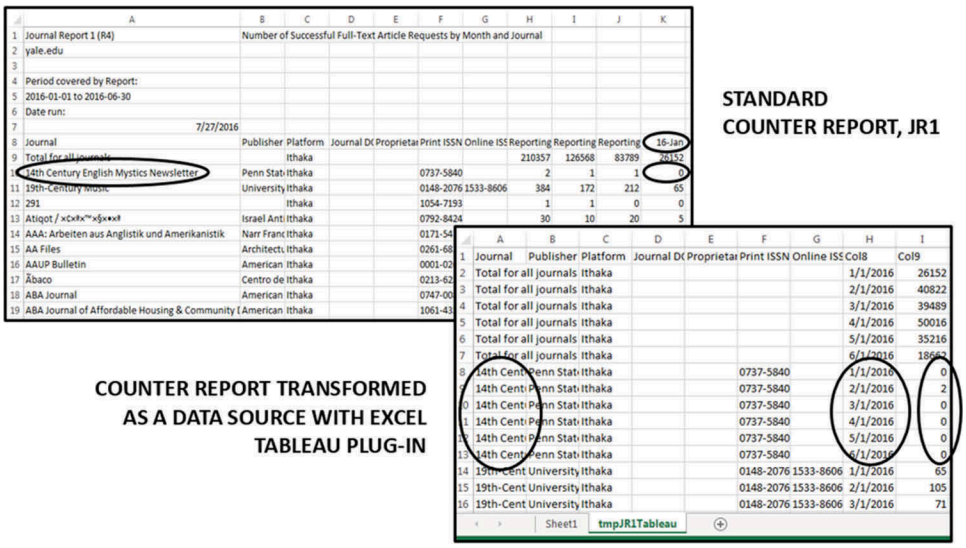


Figure 4. An original COUNTER report (top left) compared to COUNTER as a data source after applying the Tableau plug-in for Microsoft Excel (bottom right).

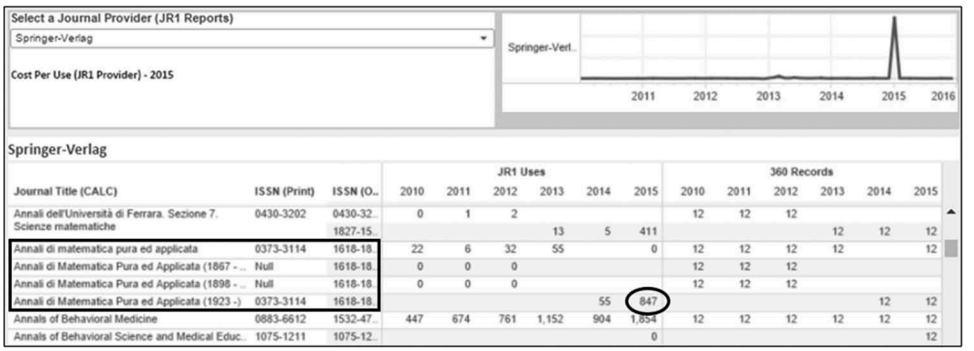


Figure 5. Proof of concept of the home-grown consolidated system within Tableau, using "Annali di Mathematica Pura ed Applicata" as seen with an error within Intota Assessment from Figure 3.

On the horizon

In Phase 3 the E-Resources Group was successfully able to pilot a workflow for transforming COUNTER data for Tableau consolidation as a direct result of leveraging the additional staff time made available due to 360 COUNTER usage statistics harvesting. Although the Phase 3 workflow was still manual for staff, it provided more opportunities to enhance reporting. Moving forward the goal will be to make these processes less manual—thereby freeing up even more staff time for other projects—and more robust so that Tableau consolidation can be utilized for more reports from more providers. Toward this goal, in February 2017 the E-Resources Group began conversations with the Assessment Librarian to explore migrating to a Python and SQL solution, where Python would be used to both retrieve 360 COUNTER reports and transform them into a data source, and an SQL database would archive the transformed data for ingest into Tableau. With a more robust collection of consolidated usage reports available in Tableau, the E-Resources Group will be able to begin to explore setting up standard reporting views so that other librarians can use the system in a self-service way.

As of June 2017, the E-Resources Group still utilized the "Data Retrieval Service" (DRS) harvesting option for 360 COUNTER under the COUNTER 4 Code of Practice. While DRS is extremely valuable to the E-Resources Group, it is a manual process for ProQuest. ProQuest might soon require that the E-Resources Group move to automated Standardized Usage Statistics Harvesting Initiative (SUSHI) harvesting, which poses a significant challenge for data maintenance and quality assurance due to the volume of monthly reports that SUSHI delivers and that will be archived in 360 COUNTER. With that said, the worksheet used in Phase 2 can be leveraged again to test and keep track of the quality of data arriving from SUSHI. Additionally, if a Python solution is successful, Python can also be leveraged to retrieve and clean-up the SUSHI reports. Now that COUNTER Release 5 is imminent, the E-Resources Group will need to carefully consider and evaluate which data points are used and how they might cause variations in display and usage consolidation. Although Phases 1–3 have been successful as ways to evaluate and understand usage data, there is still assessment work to be done to ensure reporting accuracy.

Discussion

Questions from attendees ranged from why certain assessment decisions were made to asking about staffing. One audience member asked why individual providers were evaluated for the Phase 2 worksheet, when usage consolidation can combine providers and years, and the answer related to the need to understand data handling at a granular level in order to assess the system as a whole. Another attendee asked how the workflow can be replicated for a library that has a much smaller staff. The author's answer related to moving slowly during Phase 1—which was accomplished solely by the E-Resources Support Librarian—and then acquiring buy-in from stakeholders to participate in the project.

Notes

1. There were two important decisions relating to Yale University Library's implementation of 360 COUNTER: (1) The E-Resources Group decided against the SUSHI harvesting option in favor of the "Data Retrieval Service" in which a ProQuest team manually harvests and uploads COUNTER reports into the system, and (2) The consolidated system (Intota Assessment) can be used to calculate cost-per-use metrics, but only if cost data are added to the system, which the E-Resources Group did not add due to the manual work it would require.
2. The Excel worksheet used in Phase 2 can be leveraged for any usage consolidation system assessment, and is available for use at https://tinyurl.com/y7bvlg27.

Notes on contributor

Kelly Marie Blanchat is Undergraduate Teaching and Outreach Librarian at Yale University Library, New Haven, Connecticut.

ORCID

Kelly Marie Blanchat http://orcid.org/0000-0003-2689-9758