

Against the Grain

Volume 30 | Issue 4

Article 81

2018

Library Analytics: Shaping the Future – Consortial Usage Statistics Analytics and the CC-PLUS Project

John McDonald
EBSCO Information Services, johnmcdonald@ebSCO.com

Kathleen McEvoy
EBSCO Information Services, k.mcevoy@ebSCO.com

Anne C. Osterman
VIVA, aelguind@gmu.edu

Jill Morris
PALCI, jill@palci.org

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



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

Recommended Citation

McDonald, John; McEvoy, Kathleen; Osterman, Anne C.; and Morris, Jill (2018) "Library Analytics: Shaping the Future – Consortial Usage Statistics Analytics and the CC-PLUS Project," *Against the Grain*: Vol. 30: Iss. 4, Article 81.

DOI: <https://doi.org/10.7771/2380-176X.8310>

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.

Library Analytics: Shaping the Future — Consortial Usage Statistics Analytics and the CC-PLUS Project

by **Anne C. Osterman** (VIVA) <aelguind@gmu.edu>

and **Jill Morris** (PALCI) <jill@palci.org>

and **Jason Price** (SCELC) <jason@scelc.org>

Column Editors: **John McDonald** (EBSCO Information Services) <johnmcdonald@ebSCO.com>

and **Kathleen McEvoy** (EBSCO Information Services) <kmcevoy@ebSCO.com>



Introduction

Library consortia serve libraries by leveraging the efforts of the few to save the time and resources of the many. They add value by capitalizing on the context, perspective, and influence that can only emerge when libraries collaborate. As libraries face the on-going mandates from their institutions to do more with less, many are investing deeply in consortial collaborations. These critical partnerships require shared, well-structured data and analytics that support effective decision-making at both the individual library and consortial levels.

Although there are many different consortial approaches to licensing resources, the need for more effective management of underlying usage data is universal. These needs can be divided into two primary areas: the collection, storage, and retrieval of usage statistics, and resource evaluation and negotiation support. This article highlights recent efforts to address these core needs through an IMLS-funded project, **Consortia Collaborating on a Platform for Library Usage Statistics (CC-PLUS)**, and a separate, but complementary, consortial data visualization project being developed in tandem.

Identifying Consortial Usage Data Needs

Beginning in 2014, a small group of leaders within the **International Coalition of Library Consortia (ICOLC)** spearheaded an initiative to analyze current consortial needs, spur discussion, and recommend shared action on community identified gaps. The group conducted an environmental scan and administered a survey of library consortia, nationally and internationally. The survey, with its 40+ consortial responses, represented thousands of libraries and revealed deep issues with the management and utilization of available e-resource usage data.

Particularly striking was the finding that many consortia had been dealing with usage data using manual processes and time-consuming methods, e.g., emailing individual vendors to obtain reports and downloading individual reports for each member institution. Only 20 percent of respondents were able to make use of automated services, such as the **SUSHI** automated retrieval standard.

The survey results also identified specific needed functionalities, including streamlined processing capabilities, the ability to combine key data points for improved contextual analysis, streamlined vendor password management, and visualization tools that would facilitate analysis. The results demonstrated a tremendous and urgent shared need for usage data and usage data system solutions and became the foundation for collective action which spurred the development of the **CC-PLUS** project.

Despite the clearly identified needs of consortia and their members in this arena, few options existed then or now to address them. A few consortia have developed and managed locally-built systems. Since the **CC-PLUS** project's inception, commercial products that meet some of the articulated needs have emerged, such as RedLink Consortia Dashboard and MPS Insights, but the need for a strong community-driven effort with the flexibility, tools, and functionality built by and for consortia, remains. It is not just a matter of putting multiple libraries together in the system, which is often the focus of commercial products; rather, it is about enabling high-level, broad product comparison, while also being able to isolate and contextualize varied consortial collections. No widely-affordable and available, or open source solution developed in direct response to consortial needs, exists.

In 2017, the **CC-PLUS** project was granted **IMLS** funding through a National Leadership Grant for Libraries (LG-72-17-0053-17) in support of **IMLS'** national digital platform strategy. This planning grant allowed the consortial community to work deliberately to develop an open source prototype to harvest, ingest, manage, and build basic journal reports from **COUNTER** data, initially using the **Jisc** Journal Usage Statistics Portal as a model. This successfully developed prototype is now available on GitHub under an open source license. **CC-PLUS** is examining future funding support options and plans to further develop this shared tool with community input. More information on the **CC-PLUS** project is available here: <http://www.palci.org/cc-plus-overview>.

To date, **CC-PLUS** has focused on the nuts and bolts of harvesting, processing, and storage — addressing the major prerequisites to the ability to analyze and visualize the data. A companion project to **CC-PLUS**, locally funded by the **Virtual Library of Virginia (VIVA)**, is the creation of an open source tool that will serve as an optional front end to the **CC-PLUS** data source, providing a simple and powerful interface for data access and visualization. A version that shows the tool with sample data is available here: <http://sampledata.vivalib.org>. It is constructed with the consortium's needs at the forefront, but also provides individual institution data and on-demand benchmarking — an expressed member need.

Meeting Consortial Usage Data Challenges

Goal 1: To simplify and automate collection, storage, and retrieval of usage statistics

Managing usage statistics for multiple libraries is complex, and providing appropriate access is complicated, but there are significant economies of scale that are gained by applying a consortial lens to these issues. The time savings on harvesting, troubleshooting, and cleaning usage statistics can be significant if the work is centralized and multiple libraries are processed at once. Economies of scale extend to automation, software, and services, where shared investment can make resources go further. Similarly, many libraries cannot afford to hire their own experts in assessment or data visualization, but a consortium can coordinate the skills and expertise of a few people to benefit a large group of libraries.

What makes the collection and storage of consortial usage statistics different than conducting these operations for individual libraries? One seemingly straightforward issue is ensuring that data for all relevant libraries has been received. It is often the case for consortia that a given content provider or product has only a subset of member libraries as subscribers, so regular quality control checks, to ensure all subscribers have data for listed products, is key. This is often difficult, due to variations of institutional naming within all usage reports, **COUNTER** and otherwise. These variations create the need for a trusted mapping of all variant names to the standardized, single name for a given institution.

The nature of shared consortial purchasing (acquiring products for a group of libraries with a central payment) is also in direct conflict with standard usage reporting. If a library subscribes to a consortial package of journals and also buys titles outside this package from the same publisher, the usage report doesn't distinguish between the two sets of titles, frustrating any type of analysis. It simply represents all the usage this library's users generated on the platform and product. Isolating the data related to the consortial acquisition is critical to understanding

continued on page 61

its value, but this can require extensive processing. This complexity deepens with eBook collections, since the sets of titles are often much larger and may change daily, weekly, or monthly, rather than annually.

If a consortium wants to extend data access to its member institutions, issues related to the level of release arise. For some consortia, there is an understanding that members can see each other's data, but for others, privacy concerns dominate. Similarly, local management of credentials might ease the work of a consortium's central office, but if that is enabled, appropriate levels of administrative access to consortial data are necessary.

Finally, basic storage needs scale up quickly for consortia. Consortia range in size, but it is not unusual to have 50, 100, or even hundreds of member libraries. If a consortium with 100 members subscribes to a journal package with 300 journals, through that one agreement, there is the need to store 360,000 rows of data in a single year for just a simple full-text total. Additional factors, such as Gold Open Access, html, pdf, and backfile counts, not to mention alternate data views, quickly multiply this number. Harvesting, checking, reconciling, and processing all of this data in a way that accounts for standardization of institutions and consortial acquisitions is a monumental task.

Although the scope is daunting, the **CC-PLUS** prototype already addresses many of these challenges. The prototype is currently capable of importing and storing consortial/library **SUSHI** credentials; harvesting, validating, and storing usage data for **COUNTER** Journal Reports 1 and 5 for multiple consortia, each with multiple libraries, from many major scholarly publishers; providing system alerts for problems with data harvests or other consortium-defined criteria, which significantly reduces the need for staff intervention; enabling administrative and viewing access at a number of levels; and reporting usage data in a dynamic interface, which responds to consortial/library operational needs.

Goal 2: To evaluate ongoing subscriptions and support renewal negotiation

The benefits of consortial data services go beyond cost and time savings, as there are also opportunities for added value. When the data for multiple libraries is available in concert, each library can deepen its understanding of its own situation through contextual comparison to, and benchmarking against, other libraries. Libraries can detect information resources that seem particularly well- or under-used much more effectively when they can compare their usage to that of their peer libraries on the same platform. Such comparisons also allow libraries to distinguish between usage pattern changes that are caused by local changes (such as discovery system configuration) and those caused by global changes (such as a publisher improving its search engine optimization or adding a whole book download button). Rich data is also a tremendous asset in analyzing the value and performance of shared information resources. Whether or not a shared acquisition is a good investment for a group of libraries is a complex question, but understanding how it is used by the various populations it is provided to is critical.

A shared acquisition evaluation has two major requirements: the ability to view all the institution-level data for a given product in concert, and the ability to view summary, comparable data for the entire consortium. Depending on the structure of the consortium, it is also often helpful to see levels of usage among institution types, such as public and private institutions, or community college and research institutions. The key metrics used for these kinds of comparisons include ones relevant to individual institutions, such as cost per use or percentage change in usage from the previous year, but there are also opportunities for new metrics within this context. The range of full-text uses per FTE across member libraries can be informative, for example, as can applying the context of **Carnegie** classifications. There are also consortium-wide key metrics that can be used to illuminate the differences among shared products, such as the proportion of usage by the highest using institution. If 80 percent of the usage is at a single institution, for example, a product might not be broadly useful, unless that institution is covering the majority of the costs.

Consider, for example, a shared purchase of a large publisher eBook or e-journal backfile collection followed by small annual content addi-

tions to keep the collection up to date. Assessment of content acquired in this way needs to take into account cumulative cost per use for the shared content, because the initial and ongoing cost of the collection differ greatly and the benefits of the expenditures are gained over a period of many years. While participants' cumulative cost per use is bound to vary greatly among libraries, even when scaled by FTE, institution level data collected year after year provides ballpark expectations that identify a reasonable range for cumulative cost per use as it changes over time. It also enables identification of outlier libraries that are paying too much or too little, allowing adjustment of the pricing model or underlying content to address discrepancies. To take this example a step further, one can imagine a few smaller, special focus libraries that cluster on the low end of breadth of use and the high end of cumulative cost per use; comprehensive rich data can help determine whether these differences are entirely appropriate or require a remedy.

Assessment of e-resources also has a direct and useful application in decision-making about acquisition, renewal, and cancellation, and increased availability of data and data analysis tools has enriched and expanded the negotiation conversation with publishers dramatically. Key metrics for shared resources can demonstrate expected levels of performance on a large scale, and these can translate directly into doing more with less through renewal negotiations.

Consortia, for example, regularly face annual percentage increases on subscriptions that are much higher than their median library budget increase. Content providers often justify this increase based on an even greater percentage growth in the volume of content or total usage, banking on the expectation that when libraries are forced to cancel content, it won't be theirs. For instance, one might argue that the participating institutions usage has grown by six percent, justifying a price increase of four percent. Leaving aside all the underlying variability and external factors that suggest that a six percent increase in usage does not indicate a six percent increase in value overall or across the board, the knowledge that the median usage increase of similar content from other providers is ten percent (for instance), can help consortia and their libraries distribute these increases more fairly, and dispassionately separate the publishers with aggressive pricing jumps from those that were previously undervalued. Consortia that can put content provider price, usage, and content changes in quantitative context go a long way toward having a strong hand in renewal negotiations, or at least the opportunity to advise their libraries on the most rational places to make cuts.

CC-PLUS began with a vision as a consortial tool, not simply a place to collate data for multiple institutions. As noted, and illuminated here, it is not enough to show only 5-10 institutions' data together; the entire scope of available, relevant data must be accessible to create a solid understanding of a resource's value, particularly for shared resources. Similarly, for resource evaluation and negotiation support, all products, or at least those with similar format types, must be viewed and analyzed together. For both of these needs, the abilities to adjust dates, see trends over time, and drill down to actual numbers are critical to creating a useful assessment environment. The **VIVA** visualization tool is one way to approach this, but it is envisioned that there will be a community of support for many routes of visualized assessment appropriate to consortia. Open source or commercial tools, such as Tableau, could be deployed with **CC-PLUS** to examine this data in new ways.

Moving Forward

E-resource usage statistics exist in a dynamic environment with large changes on the horizon. **COUNTER** Release 5 and a new **SUSHI** standard will create significant adjustments not only in how reports are structured and distributed, particularly for consortia, but new metrics that require new approaches and interpretation.

As the **CC-PLUS** project team moves its tool from a limited-use prototype to a production-ready platform over the coming year, it will continue to engage with the wider consortial and library communities. **CC-PLUS** is a tool, but it is also the foundation of a community. This project will continue to build the partnerships required both within and outside our communities to ensure that consortia can successfully address usage data issues now and into the future. If you would like to participate in the conversation, please write to **Jill Morris** at <jill@palci.org>.