Usage Reporting and Analysis: Current Options and Future Opportunities

Tim Lloyd
Presenter

Tricia Miller
Presenter

Katherine Swart
Presenter

Abstract

The complexity of usage reporting continues to increase, driven by the growth of Open Access publishing models and a more diverse range of stakeholders interested in understanding the usage and impact of published content. In the first part of this presentation, Katherine Swart of the Hekman Library at Calvin University reviewed and compared leading products that use Standardized Usage Statistics Harvesting Initiative (SUSHI) protocol to harvest usage reports and provide libraries with analytics on the usage of their e-resources. In the second part of this presentation, Tim Lloyd of LibLynx and Tricia Miller of Annual Reviews analyzed how evolving Open Access publishing and distribution models are changing usage reporting, how different stakeholders want to use usage data to evaluate impact and value, and how our community is grappling with and developing standardized usage data sharing practices.

Usage Statistics Harvesting and Analysis: A Comparison of Available Products

Introduction

The first presentation by Katherine Swart explored the advantages and disadvantages of four usage statistics harvesting tools: CELUS, LibInsight, EBSCO Usage Consolidation, and 360 Counter. Swart began by discussing why librarians might want to monitor usage statistics using a Standardized Usage Statistics Harvesting Initiative (SUSHI) harvesting tool. Among other reasons librarians can analyze whether subscriptions are being used, justify purchasing and licensing decisions, and gather data for surveys such as the Integrated Postsecondary Education Data System (IPEDS).

Swart created a list of basic things users expect from a SUSHI harvesting product:

- The ability to input an institution's SUSHI credentials for each provider, test those credentials before harvesting, and edit those credentials if they change.
- The ability to automatically download Counting Online Usage of Networked Electronic Resources (COUNTER) reports for each provider on a regular basis, manually upload COUNTER reports when needed, and extract already-pulled reports.
- The ability to analyze usage data at the platform level and at the individual journal level; compare platforms, journals, and e-books; and visualize that data.
- The ability to measure cost-per-use and evaluate turn-aways.
- The ability to gather data easily for annual surveys like Association of College & Research Libraries (ACRL) and IPEDS.
- The ability to manage errors with little effort.

CELUS

What makes CELUS unique is its easy-to-navigate interface and dashboard. After librarians input the SUSHI credentials for
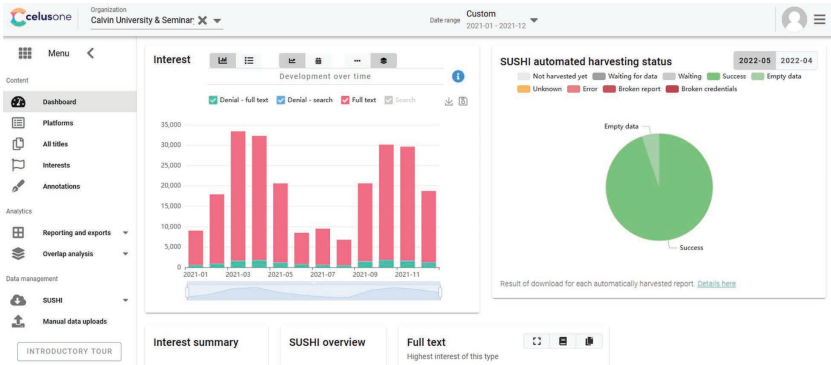
Figure 1. CELUS's main dashboard

COUNTER-compliant platforms, CELUS automatically harvests reports. Librarians keep track of harvests on a dashboard (Figure 1) and can visualize usage on charts and graphs. They can also drill down to specific platforms, e-journals, and e-books and see the usage data.

CELUS offers many advantages. Librarians can analyze subscriptions with visually appealing graphs. CELUS has intuitive design and is easy to use. Their ACRL and IPEDS reporting tool is a real time-saver. Their customer service is excellent. CELUS accommodates COUNTER 4 and COUNTER 5 data and has a way of combining the reports to show continuity of use.

One disadvantage of CELUS is that since usage is pulled monthly, librarians need to monitor the site for broken credentials or blank data sets. CELUS does make it easy to spot and fix the errors. If you want to get your data back out of CELUS, you can download machine readable data which is not like the report you would get if you downloaded a COUNTER report from a publisher's website.

LibInsight

What makes LibInsight unique is the utility and versatility of its Data-sets function. Librarians can specify which reports to harvest in the Datasets function and create high-level graphs on a dashboard like the total item requests by month graph in Figure 2. Users can also set the
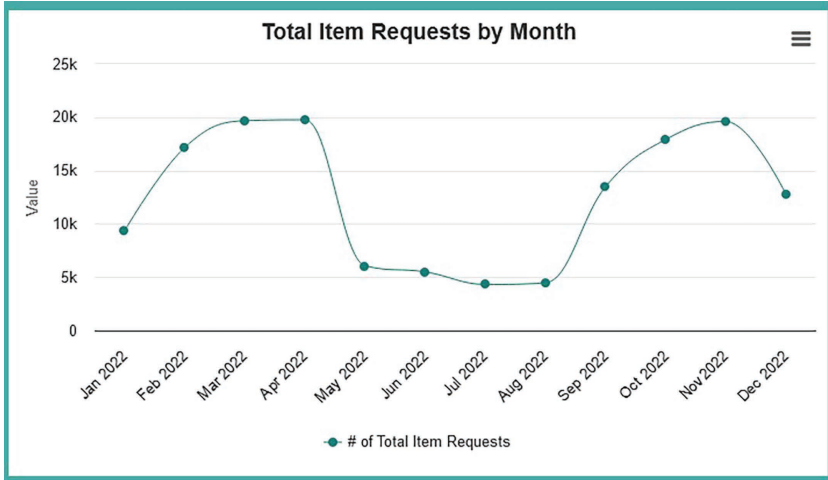
Figure 2. Graph in LibInsight measuring total item requests by month

dates they would like to view and then get a list of usage-per-platform during those dates.

An advantage to LibInsight is that cost-per-use works with journal platforms if you are willing to put in the pricing. It is fairly easy to get the number of titles and total item requests by platform for surveys like ACRL and IPEDS. The dashboard option is nice for a high-level view of your collection.

A disadvantage of LibInsight is that the design is not as slick as that of CELUS, and there was a big learning curve for Swart. LibInsight sends an email for every harvest indicating whether or not it was successful, which Swart found overwhelming. The tables and charts are downloadable, but the COUNTER reports are not.

EBSCO Usage Consolidation

What makes EBSCO Usage Consolidation unique is that it integrates with EBSCONET so that librarians can see cost-per-use without inputting the pricing data. There are many available customizations for harvesting and reports, such as a mediated option which allows librarians

to accept pending reports. Although the reporting module can create spreadsheets, there are no visuals.

The primary advantage of EBSCO Usage Consolidation is in its integration with EBSCONET. This is most useful for librarians who subscribe to several journals through EBSCO. Another advantage is that librarians can download the exact COUNTER reports that were pulled from the publisher's website. Furthermore, if you pay an additional fee, the Usage Loading Service will monitor your harvests for you.

The presenter finds the learning curve to using EBSCO Usage Consolidation a considerable disadvantage and notes there are almost too many customization options. One glaring disadvantage is that librarians must pay for a separate product called Panorama if they want visualization tools.

360 Counter is an Ex Libris product.

360 Counter is unique in that there is no limit to the number of platforms or reports it will handle. Automatic SUSHI harvests are monthly, but data from the most recent two months are unavailable. Librarians can enter prices to generate cost-per-use data. They can run custom reports and graphs in the Intota Assessment module. Like the other platforms, librarians must monitor the error log for broken reports.

One advantage is that the Intota Assessment module allows you to run analytics and do custom reports and graphs. Intota Assessment was the only part of the demonstration the presenter was not able to see, so she is unable to describe the quality or difficulty of this module.

A disadvantage is that you are only able to download JavaScript object notation (JSON) files and unable to get the original COUNTER reports. It can deal with COUNTER 4 and COUNTER 5, but there is no continuity or translation if viewing multiple years.

Conclusion

Swart concluded that CELUS is the best value. The price cannot be beat, it has a modern-looking user interface, and it is easy to use.

She likes LibInsight, but feels it is complicated compared to CELUS and the graphics are not as appealing. EBSCO Usage Consolidation is decent, but Swart got mixed reviews when she talked to librarians who use it. Acknowledging that she spent the least amount of time with 360 Counter, Swart said that it does the job, but it is a little confusing to use and may be pricey. For more information about CELUS, see Swart's review in *The Charleston Advisor*.[1]

The Increasingly Complex Nature of Open Access Usage Reporting

The second presentation in this session started by exploring how the changing environment for scholarly publishing is driving greater complexity in Open Access (OA) usage reporting.

Why Does This Matter?

Despite accounting for significantly more usage than paywalled content, reporting on usage of OA content has received relatively little attention to date. LibLynx analyzed a mixed sample of 500 million OA and paywalled usage events that they processed over the last twelve months (see Figure 3) and found that, on average, OA content received seven and a half times more requests than paywalled content.

OA content also receives usage from a much more diverse community than paywalled content. LibLynx analyzed the organizational source of the OA usage from the sample to determine that users from 189 countries had accessed content, with wealthier countries comprising the majority of usage. However, filtering the data to countries that each account for less than one percent of global usage reveals a whole host of lower middle income and low-income countries that are getting value from OA content, such as South Africa, Ethiopia, Pakistan, and Mexico (see Figure 4).

Why does this matter?

Open Access content receives significantly more usage*

Paywalled

Open Access   **7½ x**

* Comparing requests for OA vs paywalled content across a sample of 500 million events over the last 12 months

Figure 3. A Comparison of OA Versus Paywalled Content Usage
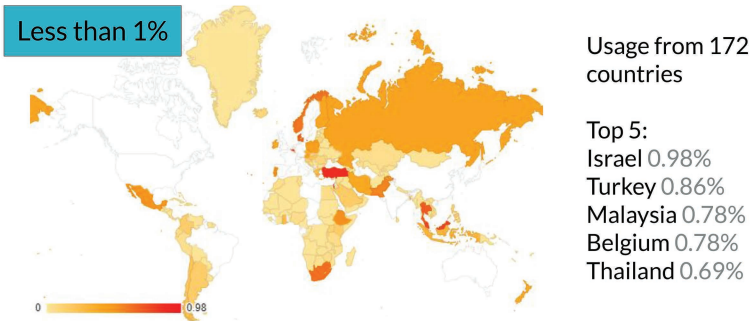

Why does this matter?

Less than 1%

Usage from 172 countries

Top 5:
Israel 0.98%
Turkey 0.86%
Malaysia 0.78%
Belgium 0.78%
Thailand 0.69%

Figure 4. Open Access Usage by Countries with Less Than 1 Percent of Global Usage

The High-Level Drivers of Complexity

Scale

Usage of OA content can be an order of magnitude greater than paywalled content. Most industry applications for usage reporting were developed for paywalled content around a traditional model of month-end batch processing. These systems struggle to transition to an environment where reporting requirements are more frequent, or even on-demand. Reprocessing is more common because data are not perfect, and some data may only become available at a later date.

Granularity

Stakeholders are increasingly interested in understanding usage at more granular levels, for example the chapter of a book, an article within a journal, or an audio or video segment. This correlates to the Item in COUNTER reporting, and it is no coincidence that COUNTER's recent 5.1 update to their Code of Practice makes the Item the default level of reporting (versus the title).[2] Item-level reporting significantly increases the volume and detail of data flowing through the system.

Stakeholders

Usage of OA content is also attracting interest from a broader set of stakeholders, outside the traditional library audience for COUNTER reports. Those managing institutional research funds want to understand the impact of their funding. Those managing and negotiating publisher relationships want to understand how their organization is generating new OA content, as well as consuming it. Authors have more choices when selecting publication venues, and usage reporting informs their understanding of these choices. These emerging stakeholders drive new use cases that add additional complexity to the process.

Data Privacy

COUNTER's Code of Practice includes a statement on data confidentiality that is based on current International Coalition of Library Consortia (ICOLC) guidelines—this statement prohibits the release of any information about identifiable users, institutions, or consortia without their permission.[3] As usage reporting grows in scale and granularity, and data is made available to a wider range of stakeholders, we need to be thoughtful about ensuring that privacy is maintained.

The Components of Reporting

Data Capture

An increasing number of publishers are syndicating their content so that usage occurs on multiple platforms, rather than just the content owner's platform. This requires the usage from these third-party platforms to be integrated into a publisher's own usage reports to provide a comprehensive view of usage—and adds more layers of complexity.

As raw usage data is sourced from a wider range of platforms, a more diverse range of inputs is to be expected. At the format level, files can be tabular (e.g. comma delimited comma separated value [CSV]) or structured (e.g. JSON). In some cases, usage data for a single platform can be split across multiple files due to the peculiarities of the database exporting the raw events. At the metadata level, examples include the use of free-text fields vs. standard identifiers, and varied conventions for time-stamping.

Data Processing

The new, emerging use cases will require additional metadata to drive new reporting. For example, adding content topics such as climate science or cancer research to enable analysis of usage against funder research priorities, or adding author identifiers to enable reporting to

be filtered for content written by a particular author. Similarly, we are seeing a need for new processing logic, such as affiliating OA usage with an organization or using third party databases to look up the funder identifiers for a journal article.

Delivery

These new use cases also reveal a need for more diverse reporting formats that support both machine-driven reporting (bulk exports; application programming interfaces [APIs]) as well as user interfaces (webpages, spreadsheets, and portable document formats [PDFs]). As OA publishing becomes increasingly global, we will also see a greater need for internationalization of reporting capabilities e.g. date formats, number formats, and multilingual support.

We are also seeing a demand for a great frequency of reporting. COUNTER reports capture a month of usage, but increasingly usage data is flowing in real-time and enabling on-demand reporting that can cover custom date ranges and be used to power reporting applications that are also working in real-time.

The Future of Open Access Usage Reporting

This future has important implications for the systems and workflows needed to support OA usage reporting. They need to be exponentially scalable and support granular, real-time (or near real-time) reporting. They need to be able to flexibly cope with a variety of input and output formats, and swap in custom processing logic for different use cases. They need to bake in standards to ensure reporting is consistent, credible, and comparable. It will also require the development of policies to underpin data collection practices and ensure they are legal and ethical. They need to be reliable and auditable, so our community can understand how they are created and rely on them for decision making.

In short, this is a big change. It will not happen overnight, and it does not need to. But this is the future we need to prepare for.

Open Access Usage and the Publisher Perspective

In the third presentation, Tricia Miller from nonprofit publisher Annual Reviews presented an analysis of how evolving OA publishing and distribution models change usage reporting and evaluation.

With the increase of OA usage, the data describing who and where scholarly resources are being used has changed. These changes, however, may not align with our original standards describing the value and use by institutions. The audiences, contexts, and purpose for the use of scientific literature are evolving and growing as more OA content is published and, therefore, the measurement and significance of OA usage data needs to be reexamined.

This reality puts publishers in an ambiguous position to balance paywalled and OA usage data to meet the needs of our users, ourselves, and other involved stakeholders. To start to respond to the changes in usage reporting, publishers must have access to and report upon how OA impacts our entire scholarly communications community.

For Annual Reviews, the challenges are amplified by their publishing model Subscribe to Open, which is a non-article processing charge (APC) OA business model that necessitates institutional, government, and corporate subscriptions to immediately convert new journal volumes to OA.[4] This means that we must go beyond traditional usage metric reporting to describe the impact of supporting OA publishing. Not only does OA usage impact more audiences but the data, including where it came from and how it's been interpreted, must also be trusted to correspond with traditional usage metrics.

How do we create a framework that correlates both OA and paywalled usage types, can satisfy equitable OA publishing, and sustainable financial support that is needed for any OA publishing model

whether it be for access to articles, supporting author manuscripts, and/or fulfilling funder and government mandates?

1. We start with establishing the reasoning for OA usage, access, and its impact on a global audience.
2. Collaborating to create a framework for shared understanding and standards.
3. And finally, the necessity to build trust in the data reported to us.

At the heart of OA is the impact on and access to a global audience. Relative to the traditional audience—those accessing articles within an institutional context—OA publishing creates a complex network of users. In 2017, Annual Reviews, opened the *Annual Review of Public Health* to help understand how OA affected usage and impact of scholarly review articles.[5] Using the LibLynx Open Analytics Platform in conjunction with COUNTER compliant usage data, we were able to find out who is using our content, in what context, and for what purpose. [6,7]

Use Case

Using data from the *Annual Review of Public Health* as a use case, we could visualize the potential and impact of OA over time.

After the first year of opening access to the journal, usage increased by just over 40 percent and in 2022, the usage increase was 130 percent higher than it was when content was behind a paywall (See Figure 5).

It should be no surprise that 90 percent of article usage comes from academic institutions. What is noteworthy is the variety of usage beyond academic institutions that continues to grow. Our data shows ninety-four different types of institutions downloading full text Hyper-Text Markup Language (HTML) and PDF articles. The variety of institution types within academic, government and corporate usage proves that there is a need and value of scholarly literature that non-OA
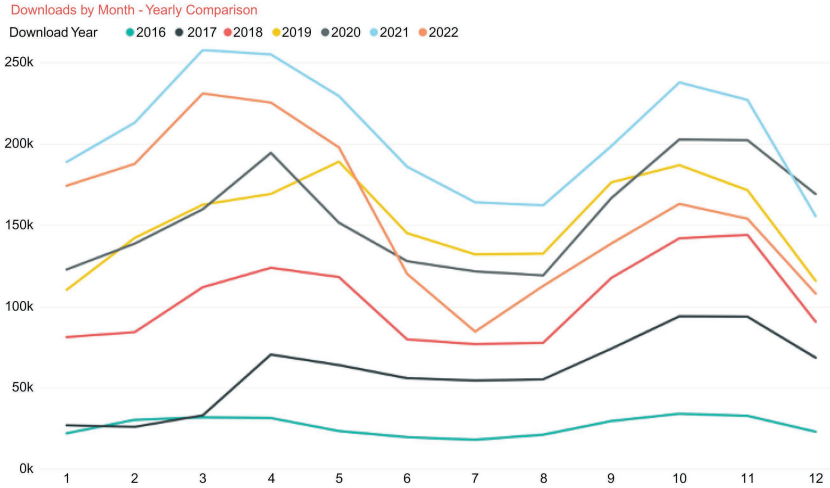
Figure 5. Annual Review of Public Health *open access usage 2016–2022*

publishing models leave out. This granularity of data is important for evaluating and supporting the needs of all users. We have found usage at places like construction companies, banks, food producers, and even prisons.

Another example of the granularity of data provided within OA usage reports is the user's areas of interest, which can reflect upon the purpose for access and impact of articles that OA publishing can support.

For Annual Reviews, their data indicates 326 different areas of interest by users (see Figure 6).

Finally, the granularity of OA data showed the impact of OA on global usage. Usage of Annual Reviews OA articles jumped from usage in fifty-five countries in 2016 to 187 countries in 2022 (see Figure 7).

While these examples are certainly not exhaustive of all the ways OA has an impact, from this use case, we can clearly see that OA has a significant impact on our usage. OA data is providing publishers with an opportunity to consider and create new products, services, and

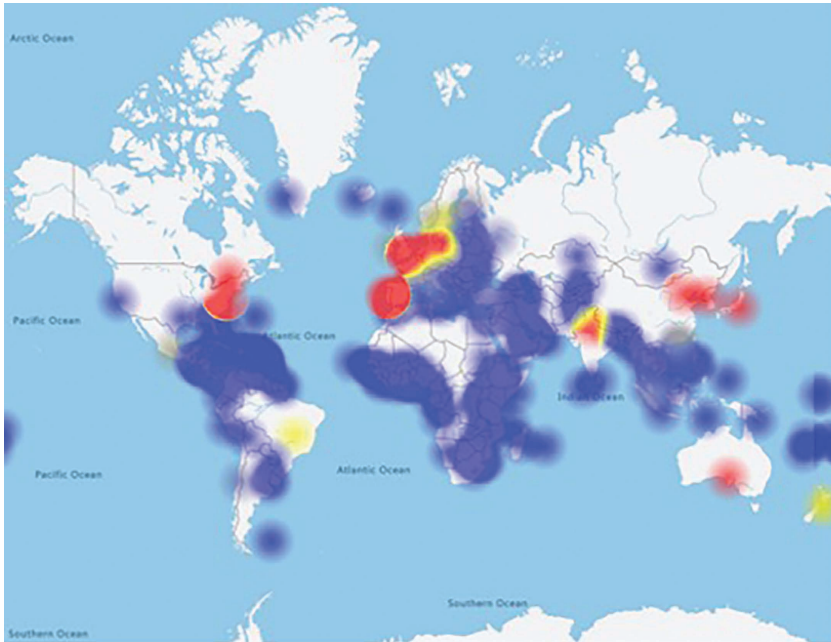*Figure 6. Annual Reviews Open Access usage by user's area of interest*

Figure 7. 2022 Annual Reviews Open Access usage by country

business models that go beyond supporting only those who can pay to participate.

Now, what we need to understand are the needs of a truly global diverse audience. How are stakeholders impacted when our audience and their needs change?

Creating a Collaborative Framework

After establishing the need for OA usage, access, and its impact on a global audience we need to develop a collaborative framework based on the integrity of data, availability of data to all stakeholders, and the reproducibility and consistency of the data. All of these things require transparency but are also subjective in interpretation without standards to establish both individual and collective goals.

The traditional usage framework, based around cost-per-use and institutional attribution, is now just part of the usage interpretation that OA reporting can provide. Many stakeholders are now also aware of and interested in a mission-driven framework where the benefit to communities, society, and global knowledge sharing exists alongside institutional benefits.

The equity in access to the global audience is one significant reason why OA publishing is accelerating so quickly. What our industry must undertake next is tackling how these individual and collective needs can co-exist and be communicated effectively.

Building Trust

Finally, trust in OA usage reporting is an important matter to publishers. Accurate OA data has the potential to build trust for a publisher and in their approach to OA publishing. We know that there are many pathways to achieving OA, but we also know that sustainability, transparency, and trust should accompany whichever model is used.

As a community, working together to achieve our collective goals, our OA data, sharing results, and interpretations helps build trust in

our relationships through these collaborations to create standards and framework.

In summary, OA data can offer evidence as to who our real audiences are, thereby allowing us to ask and understand their needs. We must also listen to institutions, libraries, funders, authors, and other publishers to help us reimagine a framework that meets the needs of all of us and is sustainable, equitable, trusted, and collaborative.

Contributor Notes

Tim Lloyd is founder and CEO of LibLynx, Washington, D.C.

Tricia Miller is Marketing Manager- Sales, Partnerships, & Initiatives at Annual Reviews, Sun Prairie, Wisconsin.

Katherine Swart is Collection Development Librarian at Calvin University, Grand Rapids, Michigan.

Notes

1 Katherine Swart, "CELUS One," *The Charleston Advisor* 24, no. 2 (October 2022): 11–14.
2 "COUNTER Code of Practice Release 5.1," Project COUNTER, accessed May 25, 2023, https://cop5.projectcounter.org/en/5.1/.
3 "ICOLC Revised Guidelines for Statistical Measures of Usage of Web-Based Information Resources," International Coalition of Library Consortia, last modified October 4, 2006, accessed May 25, 2023, https://icolc.net/content/revised-guidelines-statistical-measures-usage-web-based-information-resources-october-4.
4 "Subscribe to Open," Annual Reviews, accessed May 25, 2023, https://www.annualreviews.org/page/subscriptions/subscribe-to-open.
5 "Annual Review of Public Health," Annual Reviews, accessed May 25, 2023, https://www.annualreviews.org/journal/publhealth.
6 "Open access analytics," LibLynx, September 13, 2021, accessed May 25, 2023, https://www.liblynx.com/open-access-analytics/.
7 "Consistent, credible, comparable," Project Counter, September 2, 2022, https://www.projectcounter.org/.