

Rowland Lorimer
Simon Fraser University

Abstract

Background This technical report contains written versions of the spoken narration that accompanies five slide-based videos providing instruction on and reviewing the analysis generated by software developed by the Canadian Association of Learned Journals Readership Analytics Project.

Analysis First come usage instructions. The second and third videos offer a case study-based summary of the Standard and Premium reports. Fourth is a multi-year analysis of the case-study data. Fifth are some observations and insights.

Observations and insights The data provide a foundation for a detailed understanding of journal usage. The specific case-study data (of the Canadian Journal of Communication) brings forward an extensive set of findings, including ongoing growth in usage in an environment of declining library subscriptions, the widespread use of articles published throughout the 40-plus years of operation, and a clear predominance of HTML usage over PDF usage.

Keywords Journal metrics; Online journal usage; Journal publishing; Open access; Data visualization; Scholarly Communication

Résumé

Contexte Ce rapport technique contient des transpositions écrites de la narration accompagnant cinq diaporamas. Ces derniers portent sur une analyse réalisée grâce à un logiciel développé dans le cadre du Projet d'analyse de lectorat de l'Association canadienne des revues savantes.

Rowland Lorimer is the Founding Editor of *Scholarly and Research Communication* and Professor Emeritus at the CISP Journal Services, Simon Fraser University, Vancouver, BC. Email: lorimer@sfu.ca .

CISP Journal Services
Scholarly and Research Communication
Volume 10, Issue 3, Article ID 0301343, 10 pages
Journal URL: www.src-online.ca <http://doi.org/10.22230/src.2019v10n3a343>
Received June 17, 2019, Accepted June 17, 2019, Published July 26, 2019

Lorimer, Rowland. (2019). An Introduction to Canadian Scholarly Journal Online Usage-Analytics Software, *Scholarly and Research Communication* 10(3): 0301343, 10 pp.

© 2019 Rowland Lorimer. This Open Access article is distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc-nd/2.5/ca>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Analyse Le premier diaporama offre un mode d'emploi. Les deuxième et troisième se fondent sur une étude de cas pour résumer les rapports standard et détaillé du Projet. Le quatrième présente une analyse pluriannuelle des données provenant de l'étude de cas. Le cinquième comporte certaines observations et certains constats pertinents.

Observations et constats Les données recueillies permettent d'atteindre une compréhension approfondie de l'utilisation des revues. Quant aux données spécifiques provenant d'une étude de cas sur le Journal canadien de la communication, elles sont très révélatrices, démontrant : une utilisation croissante de la revue dans un contexte où de moins en moins de bibliothèques s'y abonnent; une consultation importante de la part du public d'articles parus tout au long des quarante années d'existence de la revue; et un recours à des documents HTML plutôt que PDF.

Mots Clés Mesures de revues; Utilisation de revues en ligne; Édition de revues; Libre accès; Visualisation de données; Communication érudite

Introduction

In a digital environment, and especially with the growth of open access, article and journal usage metrics take on increased importance. Journals associated with large commercial enterprises tend to appear in one location on the internet and, in pursuit of assessing journal performance and strategic planning, their publishers collect and analyze proprietary usage data. In contrast, smaller journals tend to appear in a variety of locations, including their own website and the websites of various aggregators or platforms. Customarily, online usage in each location is tracked in log files, and it is most often made available to journals in spreadsheet form. In some cases, accessing valid and reliable data for, say, annual usage, is difficult. Also, smaller journals often lack the staff time and expertise to pay much attention to usage data, let alone determine its validity and reliability. It is even further beyond the resources of small journals to combine the various data sets from each internet location into a meaningful whole.

The Canadian Association of Learned Journals Readership Analytics Project (CALJ-RAP) was designed to assist its members, most of which are small, independent titles, in addressing these realities. By addressing the needs of the group, it was designed to take advantage of the economies of scale in accomplishing three tasks. First was to ensure the data were valid and reliable. Second was to convert the data from spreadsheet form to meaningful figures and tables. Third was to present the data to users in both combined and separated formats, that is to say, from each internet location and from all locations combined.

The beta format of CALJ-RAP is now available to journals and includes data from Open Journal Systems (OJS) software, EBSCO, ProQuest, Project Muse, and JSTOR. It is capable of handling data provided by Érudit, but Érudit is currently unable to provide usage data to its clients. Plans are afoot to add at least one more data source in the near future.

As a means of circulating instructions to potential participants and to allow journals to see the results the software generates, five short slide-based videos were made and loaded onto YouTube. This report contains the lightly edited written scripts for the narration of the slide presentations in video format.

The videos these scripts reference can be found in the following locations:

- https://www.youtube.com/watch?v=ifHVz_-okYE&feature=youtu.be

- https://www.youtube.com/watch?v=xo_mB8QBL-c&feature=youtu.be
- <https://www.youtube.com/watch?v=Guytts7Ofw&feature=youtu.be>
- <https://www.youtube.com/watch?v=r64HH6WL3Tg&feature=youtu.be>
- <https://www.youtube.com/watch?v=aSfRXCalbjY&feature=youtu.be>

Usage instructions

These instructions are meant for journals that have already established contact with CALJ-RAP and have received an agreed-upon user name and password for accessing the CALJ-RAP site. The sign-in page is publicly accessible and by clicking on the Data Disclaimer and User Guidelines, a journal can access instructions for making contact with the project. What follows is the script for this first video.

Welcome to the Canadian Association of Learned Journals, Readership Analytics Project (CALJ-RAP).

On the screen is the sign-in page for CALJ-RAP.

The important first step is that you read the Data Disclaimer and User Guidelines document in either French or English.

Then you can enter your agreed upon or suggested username and password with a CALJ-RAP administrator (currently, lorimer@sfu.ca) and move toward inputting data.

Once that is entered, a click on the Sign-in button for first-time users will take you to the Participant Form. In subsequent visits, you will bypass the form.

The Participant Form asks for five types of content:

1. The full formal name of your journal.
2. The language or languages of your published content.
3. The primary discipline of your journal.
4. Your use of open access (OA) and subscriptions.
5. The approximate percentage of income that you derive from a number of possible categories of revenue. This is requested but not required. It will provide a foundation for comparing the performance of similar journals and would not be carried out without the permissions of all involved.

When the form is complete, please click on Submit. This will take you to the Data Upload Tool.

The Data Upload Tool can be accessed by filling out the Upload a New Data File form at the bottom of the page—in this example, you can see an indication that the data that have been uploaded for the Canadian Journal of Communication for various years from various sources. To proceed, you enter the year of the data that you are uploading. You then select the data source from the drop-down menu. If you select OJS, for example, you then click on Submit and the software takes care of everything else. (It accesses the data from your journal's website and inputs the data into CALJ-RAP). If you select a secondary aggregator such as EBSCO, you then search on your hard drive for the CSV file it has provided to you, select and upload that file, and click on Submit. The software then uploads the file to the system. You will see, on the right of the screen, that the word queued appears.

After a while the word processing appears, and after a longer period processed appears, and the data are ready for you to view by clicking on Standard Reports (in the upper left of the screen).

That is it for data uploading. The next two videos examine the Standard and Premium reports.

Thank you.

CALJ-RAP's Standard Report

This (video) presentation introduces the structure and content of the Standard Report of the Canadian Association of Learned Journals Readership Analytics Project. It does so through a case study of the Canadian Journal of Communication (CJC). The CJC is a 12-month delayed open access journal oriented primarily to Canadian communication scholarship. Prior to the emergence of such scholarly social networking sites as ResearchGate and Academia.edu, most researchers in the social sciences and humanities and many society-run journals lived in relative ignorance of the amount of attention anyone was paying to their articles. The world was different for outstanding authors who were often cited, and in certain disciplines, the exhaustive and complete citation of previously relevant research was *de rigueur*. But for many authors and disciplines, the evidence of usage was paltry.

There now exists, and CALJ-RAP provides, an integrated presentation of data from a number of well-used data sources on:

- usage levels;
- the location of views;
- preferred formats (HTML, PDF, title and abstracts only);
- the various data sources that display journal content;
- the frequency of access;
- individual article performance; and
- multi-year data access.

But before examining the data, here is a basic orientation. In 2018, the CJC's website generated 80 percent of all article views. The secondary aggregator, ProQuest, added a further 14 percent. And EBSCO added on the last six percent. This gives a total number of article views of 569,623.

Does this mean that someone went to an article nearly 570,000 times and looked at the full text? No, not really. Each time that happens, it counts as at least four article views. But everyone quotes the 570,000 "article view" as the view number. And this conforms to a publisher/library coordination organization called COUNTER. Added to the 570,000, if views of titles and abstracts are included, total "content views" for the CJC in 2018 rises to more than 780,000 views.

1. At the top of the Standard Report is the year being viewed. Users can select any year for which data have been uploaded. The first table presents the total global views of the content and the sources from which they come: the journal website, EBSCO, and ProQuest. Note the sec-

- ond line is the views of titles and abstracts.
2. Next you see the Canadian views. For CJC in 2018, they only accounted for 16 percent of all views. Total Canadian views of “article content” on all sites are just under 104,000.
 3. The third item shows the views of HTMLs and PDFs. Note EBSCO does not differentiate between HTML and PDF views. Note also the predominance of HTML views.
 4. Item four is really an overview that gives a visual and quantitative sense of where the views come from. Note also the reach of CJC for 2018, i.e., the total number of countries in the upper left.
 5. Item five compares the viewership of the website with all secondary sources for the top-20 articles. It gives a sense of which countries use secondary sources and to what degree. Some countries do, some not so much.
 6. Item six examines viewership within Canadian provinces. As well as giving a sense of relative usage, what is interesting here is that the relative preference for HTML and PDF varies by province, as does the relative use of abstracts.
 7. Item seven is a first look at individual articles and their readership from all data sources for 2018, that is, the year being viewed. Note the outstanding performance of one article and the more usual levels for well-performing individual articles.
 - a. The following chart shows article performance on the journal website alone. There are only minor differences because the website is so predominant. However, there is a short list of articles for which that is not the case.
 8. Item eight shows the performance of individual articles over the past four years (the time-span of this item will expand to a maximum of five years). Note the number-one article slides into the number-two spot, bumped by the quite dramatic use of a commentary about the media treatment of Stéphane Dion, seven years after the article was published. Again, as with item seven, it is followed by an analysis of the data from the website alone.
 9. Item nine is very important as it shows the overall performance of content by its publication year. Note the darkness levels in the tabular presentation. In the “Multi-Year Analysis” video, there is a summary table and a chart that reveals more about performance by publication year.
 10. The last item presents data on frequently used words in titles by all articles and by the top-50 articles. This item is meant to give a sense of often-discussed elements and any differences between all articles compared with the top-50 articles.

CALJ-RAP’s Premium Report

The Premium Report delves into article performance over the four years that CJC has submitted data. The intent here is not to focus on high-perfor-

mance articles but to find a way of exploring the dynamics of usage using a subset of often-accessed articles.

1. First is a look at the top-20 articles from the four years of collected data, their viewership, and the source of that viewership with five article-identity elements. Note the hover function and the two views: chart and table. Note also the articles with substantial views from secondary sources.
2. Item two does the same thing for titles and abstracts. It turns out that certain articles attract only title and abstract views and few full-text views. But this is quite unusual.
3. Item three is a “where” analysis of the viewership of the top-50 articles, including the number of countries, the number of E.U. countries, the number of Canadian provinces, and the number of U.S. states that CJC articles reached over the four years. The point is to look at which articles appeal to viewers in which locations. A follow-up question is, of course, why?
4. Item four reports on a stratified viewership of the full texts of the 10 most-viewed articles, the next 10, the following three tens, and then all others. Note particularly the high usage level of the top 10 (nearly 30% on the journal’s website), then the relatively low percentages for the “other tens” (in the low single-digit percentages for the website), and then all articles other than the top 50 (57% for the website). This indicates the importance of the entire collection of articles that the journal has assembled over the years.
5. Item five does the same for titles and abstracts and, interestingly, shows about the same patterns of widespread use of the collection as a whole.
6. Item six is a kind of “hit parade” analysis that will grow in importance over the years. It includes the year of publication, the rank on the website for the current year, the number of years in the top 20, and the highest rank obtained. This provides insight into the nature of the viewership that individual articles generate.
7. Item seven does the same for titles and abstracts. There are notable differences between the two tables, that is, full-text views and title and abstract views.

That is it for the annual reports produced by CALJ-RAP. Next is a multi-year analysis, i.e., the nature of viewership in 2015 through to 2018.

The multi-year analysis

At this time, CALJ-RAP software does not generate a separate multi-year report. Certain items within the Standard and Premium reports do address up to five years of data. This analysis was undertaken separately by downloading data from the Standard and Premium reports for multiple years.

1. The previous presentations focused on the two annual reports generated by CALJ-RAP. This presentation looks at CJC’s data for the initial four years of the project.

2. Recall the predominance in HTML viewership. This slide shows that predominance and also the apparent growth of HTML viewing in recent years. Both the high figure and the upward trend should give pause to those journals publishing articles in PDF alone. It also emphasizes the value of well-produced HTML files.
3. The CALJ-RAP reports provide data on usage by the 20 highest-using countries. This slide shows the top 10. It would be easy to assume that the top 10 countries would always be approximately the same from year to year. Turns out that is wrong. Note, for example, the sudden appearance in 2018 (in green) of Sweden and Belgium in the top 10 and the substantially increased use in Germany (50,000) and the Netherlands (8,000). Note also the high usage in the Philippines and the off-the-chart notable use in 2018 in the Seychelles. The presence of the journal publisher Hindawi in the Seychelles alongside a very interesting university may have something to do with this surprising finding. By the way, it was mainly the performance of one highly used French-language article that accounted for the changes noted. Usage may have been stimulated by an EU proposal or policy on wireless communication, the subject of the article.
4. This slide gives a different look at usage by country. Colours designate one particular year. It is gratifying to see that Canadian use predominates, except that the reason it does is based on one article viewed over 300,000 times. Note both the stable markets (all colours approximately equal) and the variable markets (one or a few colours).
5. This table illustrates that use across many countries does not correlate perfectly (by any means) with overall rank. Note for example, the “Excellence in Journalism” article in 2018, which was used in many countries but ranked number 27. In contrast, the “Looking at Shirley” article was used in 11 fewer countries but ranked fifth in overall usage.
6. Note here the usage in Canadian provinces compared with the percentage of the population: Ontario has a higher percentage of usage than its percentage of the population. British Columbia is even, and the rest of the Canadian provinces have a smaller percentage of usage than their percentage of the population. The presence or absence of communication and journalism programs within a province has some effect on usage.
7. This slide shows a growing and significant trend. While the usage level of the top ten most accessed articles is high (and bear in mind that no one knew which articles these were prior to this study), the use of articles ranking 11th through to the 50th spot is quite low. Complementing that low usage is widespread (but low) usage across the whole of the collection of journal articles. This is reflective of a dynamic engagement with an ever-shifting field of attention and inquiry. It reflects curiosity and diversity, a positive pattern in the exploration and acquisition of knowledge. Building on the previously mentioned high percentage of the usage of articles outside the top 50

are two elements. First is that the percentage of these articles (now 57%) has been growing over the past four years. Second is that ProQuest shows an even higher percentage of usage outside the viewership of the top 50 articles: more than 90 percent.

8. As this related slide shows, there is substantial change in rank and usage level each year, even in the top articles. This table provides the article titles, classifies the articles into five major groupings, and shows the year-over-year performance.
9. This chart shows viewership by year of publication. For journals, the data indicate the value of the content curated over the years.

Observations and insights

This concluding presentation offers some observations and tentative insights.

1. Whereas on average, non-website usage for CJC articles attracts 15 to 25 percent of all usage, for a very limited number of articles, all of which are listed here, the ratio is dramatically different.
2. The most dramatic ratio is for an article on cybervictimology. It has attracted nearly 40 times the viewership from secondary aggregators compared to the average article. For these articles, 10 times the usage is not unusual. The nature of the articles and their means of access suggest they might be very popular with undergraduates, a target audience for both ProQuest and EBSCO. Of these articles, only two were in the top 20 overall: the article dealing with missing and murdered women and the article dealing with interpersonal surveillance.
3. Outside events can drive dramatically higher overall usage. Professors gain public attention, for example, and the public (or possibly many students) search for their publications. One CJC author rose from a few hundred annual article views to over 10,000 in one year and then dropped down dramatically in the following year.

On the other hand, single articles can generate prominence independent of public events. In 2016, an article published in 2009 suddenly attracted the most attention that any CJC article has ever received, over 300,000 article views in a single year.

Bringing these two dynamics together, a single event captured by a single article can send usage soaring.

4. There is no sign of above-average performance by individual authors. No author had more than one article in the top 20 in 2018. This would suggest that users are driven by subject matter not author. It also seems to indicate that there appear to be no stars in the firmament of Canadian communication research scholars (outside of Harold Innis and Marshall McLuhan, both of whom are dead).

Related to the above, it appears that CJC authors refrain from sending hoards of undergrads online to download their articles.

The limitations of these interpretations are that they do not tap into illegal or unapproved usage via course-management systems; social networking services, such as ResearchGate and Academia; and pirate sites, such as Sci-Hub.

In this regard, it is very interesting that in the call for openness by universities and their librarians, the inclusion and usage levels of articles in course-management systems are hidden from journals and authors.

5. Where does all this bring us? The publication of research in CJC and, it would appear, in social science and the humanities in general, is far from a useless activity. It takes content around the world and, as growing altmetrics data indicate, contributes to the knowledge foundation of society.

Whereas subscriptions have dwindled to really pitiful levels, usage is steadily climbing. The shame of this steadily increasing usage is that the predominant discourse regarding scholarly journals focuses on bringing free access in an environment in which, at least in Canada, nothing is in place to underwrite the costs of production. Complemented by weakened copyright laws, this sets the stage for either a collapse of production or institutional capture and control. Inevitably, institutional control leads down a conservative path, reflective of institutional interests. The user community worldwide, which accounts for over 750,000 content views, contributes less than five cents per full-text article view to the CJC.

Canadian Journal of Communication users read not just the most popular scholarly articles but a wide variety of them—the classic long tail of usage. This is a salutary finding, in terms of education.

Apparently, the main role of science articles seems to be to feed science. In that context, citation indexes tell an important story.

However, in social science and the humanities it appears that the predominant value of articles is their exposure to a wide population of users, including students who can gain a sense of both article content and research techniques. While citation takes place, article views by a wide range of users are much more prevalent than citation for journals such as the CJC.

Quantitative analysis, even of top articles, does not turn usage metrics into a horse race. Rather it establishes the significant contribution being made by researchers and their journals to ideas, opinions, knowledge, and understanding. Quantitative analysis also provides a valuable foundation to understand the nature of usage.

Usage patterns, when examined closely, suggest familiar profiles that help capture the dynamics of scholarly inquiry and its uptake in society. Here are some beginning examples of article profiles that suggested themselves:

- Icaruses: fast-rising, attention-grabbing articles that establish something new and then quickly fade to rare use.
- Methuselahs: Articles that gain in importance and are used and cited for a very long time.

- Solomons: Articles that resolve a tension in the field and provide emergent wisdom.
- Newton-Sleepers: Highly reductionistic articles that bring forward powerful new paradigms.
- Einsteins: Idea-based articles that bring forward highly significant changes in a field.
- Scheherazades: Sequences of articles that, step by step, open new knowledge fields and generate continuing interest in each successive phase.
- Pied Pipers: Articles that generate a great deal of follow-on parallel research in a particular field.
- Jodies: Articles that challenge the status quo based on fundamental principles.
- Julians: Articles that bring forward suppressed data and provide new insights.

Conclusion

This ends the initial introduction to CALJ-RAP. The value of CALJ-RAP will increase with the growing attention being paid worldwide to metrics and by its adoption and control by journals themselves rather than external agencies, which inevitably have their own interests to pursue.

Websites

Academia, <https://www.academia.edu/>

Counter, <https://www.projectcounter.org/>

EBSCO, <https://www.ebsco.com/>

Érudit, <https://www.erudit.org/en/>

JSTOR, <https://www.jstor.org/>

Open Journal Systems, <https://pkp.sfu.ca/ojs/>

ProQuest, <https://www.proquest.com/>

Project MUSE, <https://muse.jhu.edu/>

ResearchGate, <https://www.researchgate.net/>

Sci-Hub, <https://sci-hub.se>