

Toward alternative metrics of SSH research impact: A comparison of visit, download and citation data

Jadranka Stojanovski¹, Franjo Pehar², Tomislav Jagušć³ and Draženko Celjak⁴

¹ jadranka.stojanovski@irb.hr

University of Zadar, Department for Information Science, F. Tuđmana 24i, 23000 Zadar, Croatia
/ Ruđer Bošković Institute, Library, Bijenička c. 54, 10000 Zagreb, Croatia

² fpehar@unizd.hr

University of Zadar, Department for Information Science, F. Tuđmana 24i, 23000 Zadar, Croatia

³ tomlav.jagust@fer.hr

University of Zagreb, Department of Information Science, Unska bb, 10000 Zagreb, Croatia

⁴ drazenko.celjak@srce.hr

University of Zagreb, University Computing Centre, J. Marohnića bb, 10000 Zagreb, Croatia

Introduction

Readers of the journal articles can express their interest, positive attitude, acknowledgment or tribute in different ways. The most traditional way is to cite the article in the own article's reference list. Nowadays' channels of the scholarly communications are changing, and journal articles are visited, downloaded, bookmarked, mentioned, commented, shared, and cited using different platforms and tools, social media and web applications (Donato, 2014). Indicators of the usage can be also visits and downloads, and in addition to the citations, can give us the more comprehensive picture about popularity and impact of journal articles. (Brody, Harnad, & Carr, 2006) The relationship between downloads and citations from a bibliometric perspective was investigated and possible differences in obsolescence patterns for articles published in a single journal were revealed. (Moed, 2005) The same study showed that, as a cohort of documents grows older, its download distribution becomes more and more skewed, and more statistically similar to its citation distribution.

Two criteria for academic careers are prevailing in the Croatian scientific community: number of papers published in the journals, and the prestige of the source journals that is measured, according to the discipline, with the Journal Impact Factor (JIF) (Macan & Petrak, 2015) and other qualitative indicators. In the social sciences and humanities the prestige of the journal is proved by the inclusion of the journal in the commercial databases like Scopus (Elsevier) and Web of Science Core Collection – WoSCC (Clarivate Analytics, former Thomson Reuters) etc., or the specific journal is recognized as high-quality journal by the community and listed as eligible for the assessment. Academic policy decision makers neglect the existence of other bibliometric indicators and the article level metrics. Because of the present criteria, authors are under pressure to publish their papers in high JIF journals. In the fields of social sciences and humanities researchers are often facing a dilemma to publish in prestigious, frequently fee-based, international journals or in local OA journals well recognized by the national scientific and professional community. The repository of the Croatian Open Access journals Hrčak includes 429 OA journals, and the majority of them come from the social sciences, humanities, and arts (n=302). In this fields research is more regionally and nationally engaged, and biased metric as JIF is particularly problematic (Hicks, Wouters, Waltman, de Rijcke, & Rafols, 2015).

In this paper, the popularity and usage measured by a number of visits and downloads of the articles published in Croatian OA journals are compared with the number of citations from WoSCC and Scopus databases at journal and article level. The research question for this study is: Is there a correlation between metrics based on citations and metrics based on visits and downloads? According to the research question, the objective of our research is to investigate the relationship between visits and download data and traditional impact metrics such as citation analysis, in order to find out if the different methods are correlating or complementary.

Methodology

The usage data for 155.602 papers exported from 429 OA journal titles were provided by the repository of the Croatian OA journals Hrčak, and stored into a database with the following metadata: ID, article title in Croatian, article title in English, journal title in Croatian, journal title in English, ISSN, eISSN, volume, issue, publication year, date of publishing, date of the inclusion in Hrčak, number of visits, and number of downloads.

Even the Hrčak repository was launched in 2006, because of additional re-calculation of the usage data in 2013 which remove all robotic and non-human traffic, the usage data from Hrčak is available only from March 2007 to the present. Available Hrčak web usage statistics includes cover journal's visits, issue visits, article (bibliographic records) visits and full-text (PDF and HTML) downloads. For the purpose of this study article page visits and full-text downloads were considered. The article page visit with the bibliographic data includes a unique visit from the same IP address during the period of 30 minutes.

We identified 48 OA journals available in Hrčak repository and indexed by WoSCC and Scopus databases. Finally, we selected 17 journals from the field of social sciences and humanities. Citation counts for the articles published in these journals were obtained by using the print and online ISSN as the main search criteria. We exported 6.641 paper records from WoSCC and 4.482 papers from Scopus. Following fields from both databases were recorded in two separate CSV files: authors, source title, publication year, volume, issue, beginning page, ending page, and a number of total citations.

Further, we assemble a single data source based on three datasets. Since not all Croatian journals have DOI and there was not unique ID among three data sets, the data from Hrčak, Web of Science and Scopus were matched using the Levenshtein distance algorithm to determine the similarity between paper titles. In the first iteration all types of the papers were included: 'Book Review', 'Conference paper', 'Editorial', 'Essays', 'In memoriam, Obituary', 'Meeting abstract', 'News', 'Original scientific paper', 'Other', 'Preliminary communication', 'Professional paper', 'Reminiscences', 'Review article', 'Short communication, Note'. Further, only the papers with the citation potential were selected: 'Conference paper', 'Original scientific paper' and 'Review article'. This procedure in the last iteration matched 2.226 unique papers with data about visits and downloads from Hrčak and citation counts from WoSCC and Scopus.

The data was collected during the period 1 to 20 February 2017.

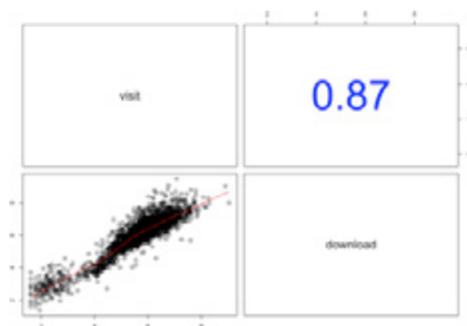


Figure 1. Hrčak total bibliographic record visits vs full-text downloads

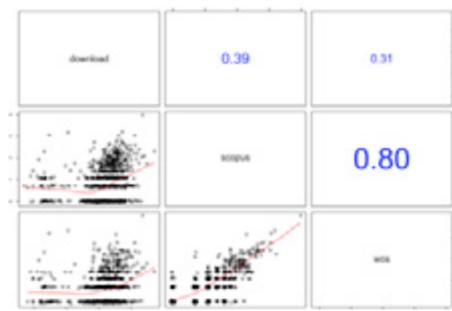


Figure 2. Hrčak full-text downloads vs Scopus and WoSCC citations

Results and discussion

The comparison between the total number of bibliographic record visits and full paper downloads is shown in Figure 1.¹ The correlation coefficient of the Hrčak bibliographic records visits and paper downloads shows high positive correlation (Spearman $r=0.87$, $p=,00$). The Spearman correlation results in Figure 2 show that there is a high positive correlation between Scopus and WoSCC citations ($r=0.80$) and medium to low correlation between full paper downloads and Scopus citations ($r=0.39$), as well as for WoSCC citations ($r=0.31$) for the same set of papers. These results show that full paper download frequency within the Hrčak repository weakly correspond to the Scopus and WoSCC citations.

1 We added 1 to all variables to enable it to be plotted on a log scale.

Conclusion

This paper presents the results of the study where different methods measuring usage (citations, visits, and downloads) has been applied. As described, there was medium to low correlation between the download data from Hrčak repository and citations from WoSCC and Scopus. This could imply that research of the regional or local relevance, published in the local journals and the local language, cannot gain a significant number of citations from the global community. Also, the usage patterns could be different between citing and downloading, related to the claim that downloads and citations measure different aspects of usage (Moed & Halevi, 2015). Citation databases as WoSCC and Scopus contain various biases such as under-representation of non-English journals (Jarvey, Usher, & McElroy, 2012), and the citations from non-indexed Croatian SSH journals are not included in WoSCC and Scopus.

The analysis of the relation between visits of the bibliographical records and the number of downloads demonstrated a high positive correlation, which could be the result of the links to the full-text article at the bibliographical level. The latter is important when discussing the advantages of open access publishing. There was also a high positive correlation between WoSCC and Scopus citations, which was expected according to the different studies (Archambault, Campbell, Gingras, & Larivière, 2009; Meho & Rogers, 2008). Since reading and citation populations are different (Moed & Halevi, 2015) the results could indicate the difference in reading and citing habits from the local community, and the international research community, where the visibility, readability, peer judgment, and citations depend on the English language, global relevance of the topic and prestige of the journal or author.

This study makes a contribution to the clarification of the question of the broad impact of altmetrics. Furthermore, these results raise questions regarding the appropriateness of the JIF and citations as a sole metrics in the assessment of article impact and suggest the possibility of devising impact metrics based on usage information in general. (Bollen, Van de Sompel, Smith, & Luce, 2005)

Limitations of the study

Hrčak serves as a national repository of Croatian OA journals, and about 70% of Hrčak journals have their own websites, with available bibliographic data and full-text articles. Because all articles are OA and licensing allow copies and redistribution of metadata and full-text, articles' PDFs are spread across various databases, repositories, and social networks. To get a complete picture about the number of downloads all these data should be aggregated.

References

- Archambault, É., Campbell, D., Gingras, Y., & Larivière, V. (2009). Comparing bibliometric statistics obtained from the web of science and Scopus. *Journal of the American Society for Information Science and Technology*, 60(7), 1320–1326. <https://doi.org/10.1002/asi.21062>
- Bollen, J., Van de Sompel, H., Smith, J. A., & Luce, R. (2005). Toward alternative metrics of journal impact: A comparison of download and citation data. *Information Processing & Management*, 41(6), 1419–1440. <https://doi.org/10.1016/j.ipm.2005.03.024>
- Brody, T., Harnad, S., & Carr, L. (2006). Earlier Web Usage Statistics as Predictors of Later Citation Impact. *Journal of the American Society for Information Science and Technology*, 57(8), 1060–1072. <https://doi.org/10.1002/asi.20373>
- Donato, H. (2014). Traditional and alternative metrics : The full story of impact. *Revista Portuguesa de Pneumologia*, 20(1), 1–2. <https://doi.org/10.1016/j.rppneu.2013.11.001>
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature*, 520(7548), 429–431.
- Jarvey, P., Usher, A., & McElroy, L. (2012). *Making Research Count : Analyzing Canadian Academic Publishing Cultures*. Toronto. Retrieved from <http://www.higheredstrategy.com>
- Macan, B., & Petrak, J. (2015). Bibliometrijski pokazatelji za procjenu kvalitete znanstvenih časopisa. In I. Hebrang-Grgić (Ed.), *Hrvatski znanstveni časopisi : iskustva, gledišta, mogućnosti* (pp. 37–53). Školska knjiga, Zagreb.
- Meho, L. I., & Rogers, Y. (2008). Citation Counting , Citation Ranking , and h -Index of Human-Computer Interaction Researchers : A Comparison between Scopus and Web of Science. *Journal of the American Society for Information Science and Technology*, 59(11), 1711–1726.
- Moed, H. F. (2005). Statistical relationships between downloads and citations at the level of individual documents within a single journal. *Journal of the American Society for Information Science and Technology*, 56(10), 1088–1097. <https://doi.org/10.1002/asi.20200>
- Moed, H. F., & Halevi, G. (2015). On Full Text Download and Citation Distributions in Scientific-Scholarly Journals. *Journal of the Association for Information Science and Technology*, 67(2), 412–431. <https://doi.org/10.1002/asi.23405>