The Web of Science database as a source of information on Open Access publications from the Czech Republic and Poland

Małgorzata RYCHLIK
Poznań University Library, Poznań, Poland
rychlik@amu.edu.pl

INFORUM 2018: 24th Annual Conference on Professional Information Resources
Prague, May 29-30, 2018

Abstract

This paper aims to present the Web of Science database as an important source of information on Open Access publications. The indirect goal of the study was to examine whether there were any differences between scholars and researchers from the Czech Republic and Poland in open publishing and, if so, to what degree?
Methods: the first stage of the study involved data extraction from the WoS CC database of publications that dealt primarily with OA publishing. The data included 9,244 publications from Poland and 3,870 publications from the Czech Republic published in 2013. For both countries under scrutiny, the following comparisons were made: type of documents, language of publication, the most popular research areas in OA and bibliometric indicators (citations and H-index). In addition, the productivity of scholars from both countries was juxtaposed and compared (the supporting statistical data was provided by OECD). The data covering the period 2010-2016 was also analysed to examine the number of works published in OA in Poland and the Czech Republic. The second stage involved extraction of the altmetric indicators for most highly cited works from both countries, and their further comparison. For this particular study, the Altmetric Explorer was used.
Results: it has been proved that the resources of the WoS CC database provide convenient data environment for different analyses, both qualitative and quantitative, and, additionally, provide a strong predictive value in defining future trends in open access publishing.

Introduction

Recent studies show that almost a half of the scholarly literature is available OA (Archambault et al., 2014; Piwowar et al., 2018). Numerous new tools, services, platforms, etc. have mushroomed to provide Open Access (OA) to all research literature. Web of Science (WoS), one of the most widely used and comprehensive scientific citation index databases, found itself among those that provide researchers with OA content. ImpactStory, a non-for-profit organization, has obtained a grant from Clarivate Analytics and designed an open online service – oaDOI (oaDOI is called now Unpaywall)[1]. As a result of this partnership, users of WoS are provided with trusted OA materials. The prerequisite for using oaDOI is that the article has a DOI (Digital Object Identifier) available. The difference between DOI and oaDOI is that DOI redirects users to a paywall, whereas oaDOI points to an OA version of the paper in the pdf format.

[1] http://unpaywall.org/data

The service searches from within a multitude of resources, such as DOAJ, DataCite, CrossRef's database, the BASE OA Search Engine and PubMed Central and, as the creators of oaDOI maintain, "the majority of our OA content comes from independently monitoring over 50,000 unique online content hosting locations, including Gold OA journals, Hybrid journals, institutional repositories, and disciplinary repositories".
The main aim of this study is to show that WoS is a valuable source of trusted OA. The user can encounter the following types of OA: Gold OA, Hybrid OA and Green OA. In the case of Green OA, WoS links only to peer-reviewed versions from open repositories.
The indirect aim of this paper is to determine the differences between Czech and Polish researchers that emerge in the field of OA publishing. On the basis of the obtained data, the present author will try to answer the following questions:

1. Which research areas have the highest percentage of available OA literature?
2. In what languages do researchers write their OA articles?
3. What are the most common types of OA literature?
4. What are the differences between the bibliometric indicators for both countries?
5. What is the productivity of Czech and Polish researchers and what is the average number of publications per researcher, and citations per article?
6. Is the OA scholarly publishing trend ascending among Polish and Czech scientists?
7. What is the average number of the most common altmetric indicators for most highly cited OA articles written by Polish and Czech researchers?
8. What is the average number of altmetric indicators and citations for most highly cited papers published by Polish and Czech scholars?
9. Are citation counts influenced by Altmetric Attention Score for most highly cited OA articles of Polish and Czech researchers?

Materials and methods

The study was divided into two stages. The first stage was to collect and filter the data from the WoS Core Collection. The metrics data collected from WoS CC are related to the authors affiliated in Polish and Czech scientific institutions. The chronological scope of the study covered the year 2013. Since the intention was to compare relevant citation counts and other bibliometric indicators and measures, the guiding principle in the study was to select those papers that potentially had enough time to be cited. Another significant factor determining the selection of this particular year of publication was the total number of publications - WoS has its own limitations and makes it possible to process citation reports for up to 10,000 units only. Due to the fact that OA literature was the core of the analysis, the search results were refined by using open access search filter. Eventually, 9,244 papers from Poland and 3,870 from the Czech Republic were obtained. The supporting statistical data was provided by OECD to investigate the productivity of scholars from both countries. The number of researchers (headcounts) in higher education amounts to 69,027 (Poland) and 22,957 (the Czech Republic) in 2013[2]. In addition, the data covering the period 2010-2016 were also analyzed to examine the trends in OA in both respective countries.

[2] http://stats.oecd.org/viewhtml.aspx?datasetcode=PERS_OCCUP&lang=en

The second stage involved the use of the Altmetric Explorer (http://www.altmeric.com) to check whether the citations counts of most highly cited articles were influenced by Altmetric Attention Score (AAS)[3]. There are 3 important factors concerning AAS (Mukherjee, Subotić, & Chaubey, 2018):

1. volume: the more people mention a paper, the higher AAS (only one mention from each person per source is taken into consideration)
2. sources: the score is the weighted count of the amount of attention per a research output
3. authors: it is essential who writes the mentions and to whom, e.g. an automated share from a journal account weights less than if a neurobiologist shares a link with other neurobiologists.

Altmetric indicators were collected for 1% of most highly cited OA papers by Czech researchers (*n=39*) and Polish scholars (*n=91*).
The data was collected on 14th April 2018.

Results

RQ1. **Which research areas have the highest percentage of available OA literature?**
Figure 1 presents 10 most popular research areas that Czech and Polish scholars publish in OA. For both countries, the highest percentage of OA papers was represented by physics. The lowest for Czech researchers was cardiovascular system cardiology (2%) and for Polish scholars research experimental medicine (less than 4%).

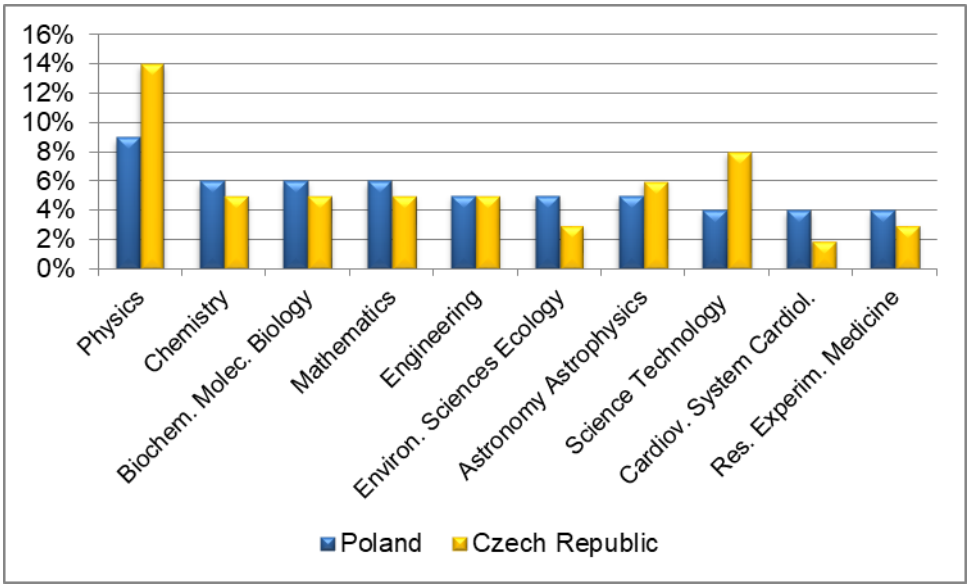

Figure 1. Percentage of 10 most prolific areas which provide OA content for Polish and Czech researchers.

RQ2. **In what languages do researchers write their OA articles?**
Figure 2 shows that only 4% of papers whose authors are affiliated to Poland were written in languages other than English. In the case of Czech scientists, this percentage amounted to 3%.

[3] https://help.altmetric.com/support/solutions/folders/6000216249

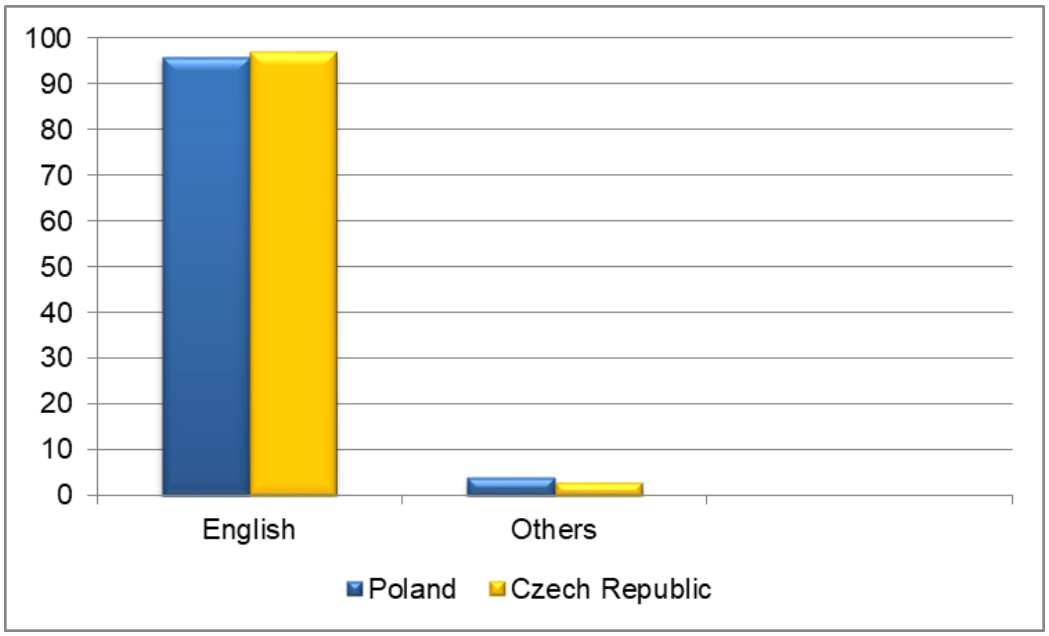

Figure 2. Percentage of papers written in individual languages.

RQ3. **What are the most common types of OA literature?**
Figure 3 shows that the most common type of a document, both in the group of Polish (N=9244) and Czech scientists (N=3870), was an article. In the case of Polish authors, non-article content accounted for 15% of all types of publications, while in the case of Czech authors this percentage was slightly higher and amounted to 21%.

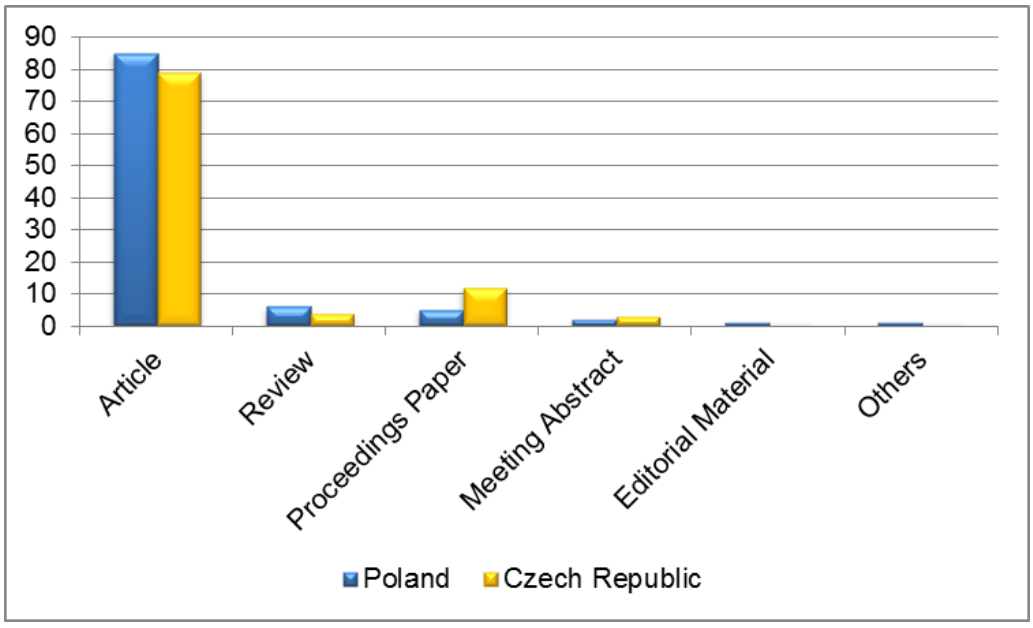

Figure 3. Percentage of publication types.

RQ4. **What are the differences between the bibliometric indicators for both countries?**
It turned out that the average number of citation per paper was over 5% higher for Czech scholars, whereas H-index was higher for Polish researchers.

Tabela 1. Bibliometric indicators for Polish and Czech OA papers

Bibliometric indicators	Poland	Czech Republic
Average number of citation per paper	12,38	17,81
H-index	114	102

RQ5. **What is the productivity of Czech and Polish researchers and what is the average number of publications and citations per researcher?**
Figure 4 shows that both the average number of citations per researcher as well as the average number of papers per 100 researchers was higher for Czech scientists. The Czechs had 1.25 more citations per researcher than the Poles and 3 more citations per 100 scientists. Here, the supporting statistical data provided by OECD was used - the number of researchers N=69,027 (Poland) and N=22,957 (the Czech Republic).

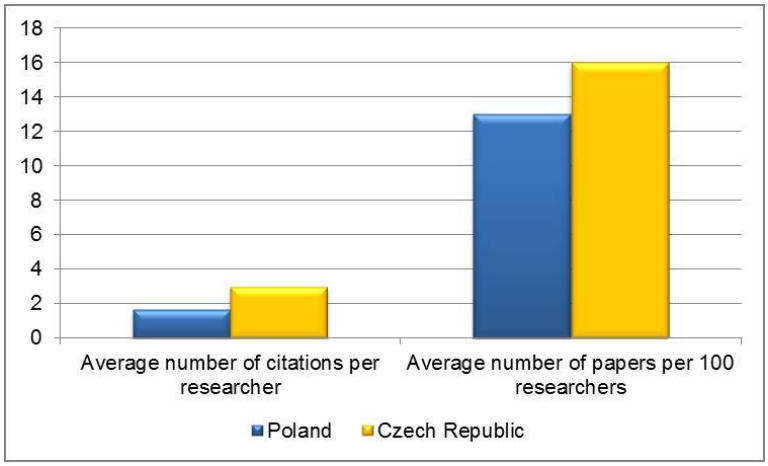

Figure 4. Average number of papers per 100 researchers and average number of citations per researcher.

RQ6. **Is the OA scholarly publishing trend ascending among Czech and Polish scientists?**
Figure 5 presents an ascending trend in OA scholarly publishing**.** In the case of Polish researchers, the percentage of available OA papers increased as much as three times, while in the case of Czech scientists it increased slightly more than two times for papers published between 2010 and 2016.

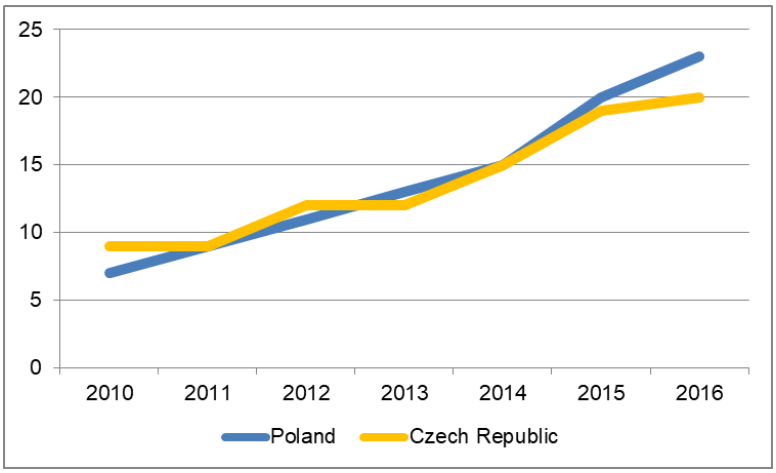

Figure 5. Ascending trend in Open Access publishing.

RQ7. **What is the average number of the most common altmetric indicators for most highly cited OA articles written by Czech and Polish researchers?**
Figure 6 shows that the average number of altmetric indicators for most highly cited papers by Polish scholars (*n=91*) was 3 times lower than that for Czech scientists (*n=39*). It should be noted at this point, however, that one of the Czech works under scrutiny had an exceptional number of Mendeley readers (6,812) and this fact affected the final result. If this particular result were reduced (omitted), the number of Mendeley readers would be only 1.6 times higher as compared to the Poles, instead of nearly 3 times higher. In the case of tweets, the situation was similar - one of the papers of Czech scientists had 641 tweets, which affected the result. If this single result were deleted, the average number of tweets for most highly cited papers would be 49, instead of 65.
The median value of Mendeley readers was 192 and that of tweets 26 for Czech scholars, whereas for Polish researchers it amounted to 107 for Mendeley readers and 13 for Twitter mentions.

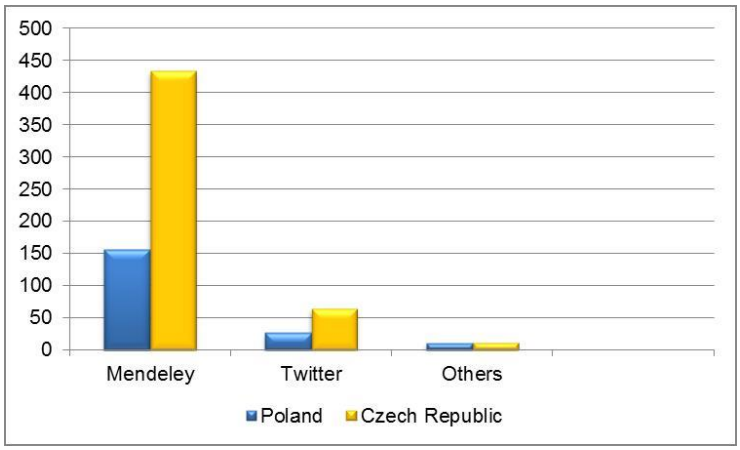

Figure 6. Average number of altmetrics for highly cited papers.

RQ8. **What is the average number of altmetric indicators and citations for most highly cited papers published by Czech and Polish scholars?**
Figure 7 shows that both the average number of altmertic indicators as well as the average number of citation counts for most highly cited papers are higher for Czech researchers.

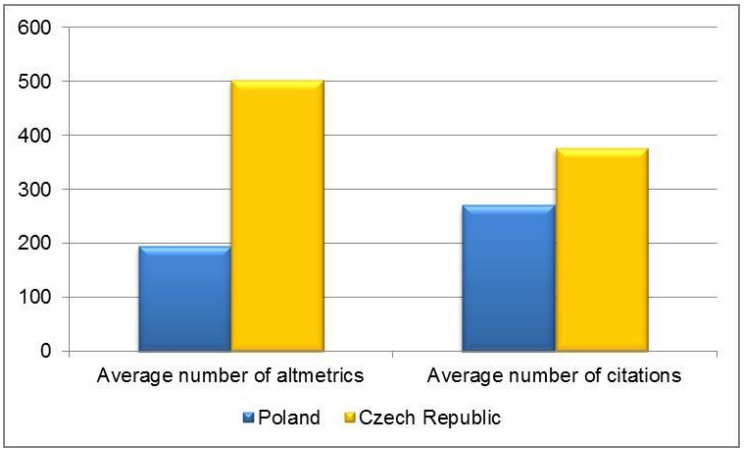

Figure 7. Comparison of average number of altmetrics and citations.

RQ9. **Are citation counts influenced by Altmetric Attention Score for most highly cited OA articles by Czech and Polish researchers?**

The number of citations of Czech open-access papers were positively dependent on the AAS (Fig. 8). This relationship was highly significant (Linear regression: R^2 = 17.85%, p = 0.0074, n = 39, y = 280.55 + 1.275x). Similarly, the number of citations of Polish open-access papers were significantly and positively dependent on the AAS (Fig. 9; Linear regression: R^2 = 9.67%, p = 0.0027, n = 91, y = 231.31 + 0.790).

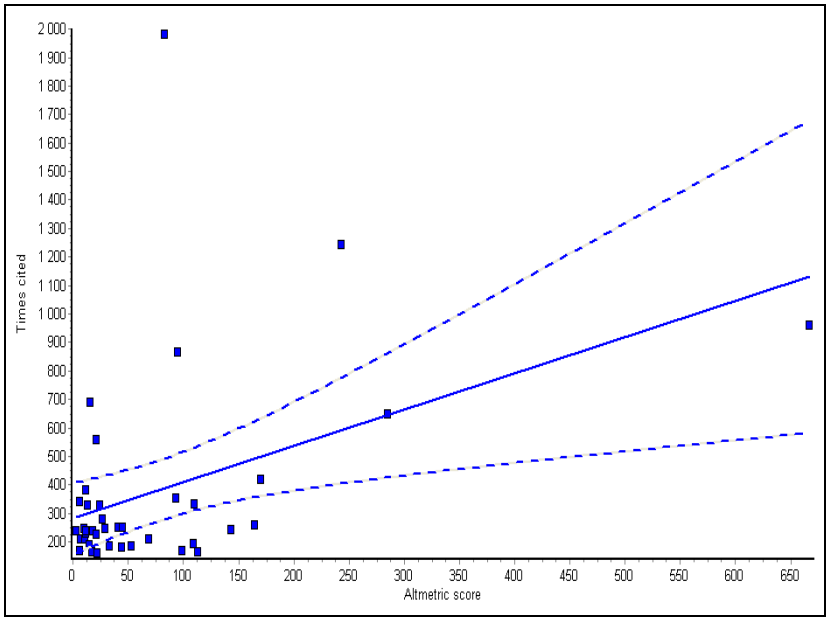

Figure 8. Linear regression for Czech most highly cited papers.

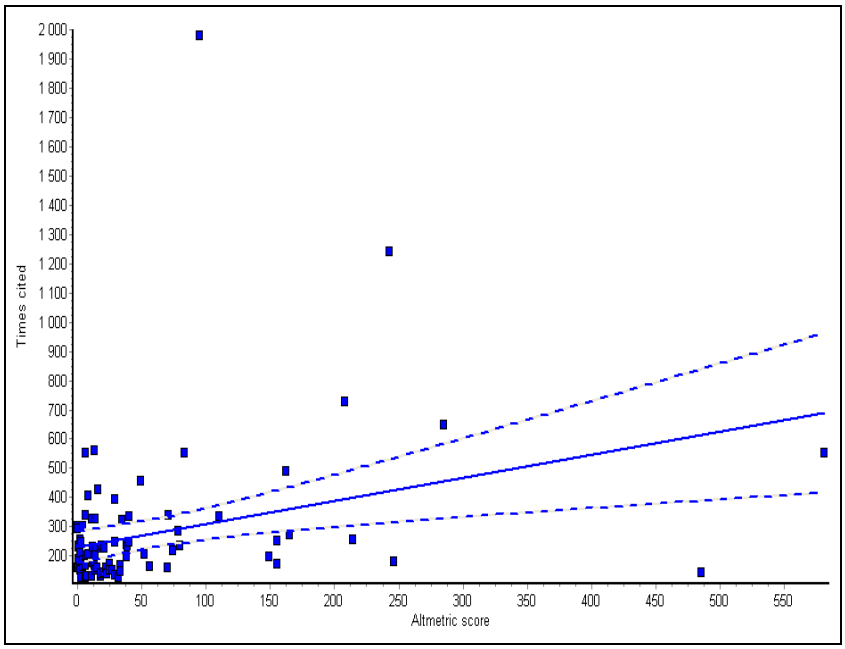

Figure 9. Linear regression for Polish most highly cited papers.

Conclusions

It is worth noticing that the obtained WoS datasets are dynamic as the content of the database changes on a daily basis. Authors constantly self-archive their research output in repositories, embargoes expire and deposited articles become OA, so there are more and more OA papers available. But there is also another possibility that articles that are free-to-read but without an open license are moved from OA to toll-access. This is the case of the so-called Bronze OA (Piwowar et al., 2018).

The author compared research areas to check in which areas the percentage of OA was the highest. There were two schemes to choose from: the research area scheme and the WoS categories scheme. The research area scheme was chosen deliberately because analyzing such data requires a broader subject classification. The WoS categories scheme is recommended rather for analyses of data concerning, for instance, a single author or department. For Poland and the Czech Republic 139 and 125 research areas were analyzed, respectively.

Most of the collected data covered only two indexes: Science Citation Index Expanded and Conference Proceedings Citation Index-Science. Data on the Humanities and social sciences were scarce. The one reason for this is that WoS covers far more materials in hard sciences than soft. Mongeon and Paul-Hus compared the scope of journals in WoS and Scopus databases with periodicals indexed in the Ulrichsweb catalogue. It turned out that in social sciences and the humanities WoS covered only about 15% of journals included in the Ulrichsweb catalogue (Mongeon & Paul-Hus, 2015). The other reason is that oaDOI applies only to papers that are assigned by DOIs. Recent studies conducted on Polish humanities journals showed that the percentage of articles with DOIs was low (Rychlik, 2017). The lack of relevant results in the humanities and social sciences may also have an impact on the language of a publication. Papers in the humanities and social sciences are much more often published in national languages than those published in hard sciences, where English is a pervasive language.

Deliberately, no analysis of Gold vs. Green OA was undertaken, as for all OA articles, WoS preferred links to the publisher's Gold version.

The analysis showed that the productivity of Czech scholars was higher both in terms of bibliometric and altmetric indicators. The Poles have only a slightly higher H-index. One of the reasons can be that gross domestic spending on Research and Development is twice as high in the Czech Republic (1.900% of GDP) as compared to Poland (0.871% of GDP)[4].

Bibliography

Archambault, É., Amyot, D., Deschamps, P., Nicol, A., Provencher, F., Rebout, L., & Roberge, G. (2014). Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels—1996–2013. Retrieved from www.science-metrix.com

[4] https://data.oecd.org/rd/gross-domestic-spending-on-r-d.htm

Mongeon, P., & Paul-Hus, A. (2015). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, *106*(1), 213–228. https://doi.org/10.1007/s11192-015-1765-5

Mukherjee, B., Subotić, S., & Chaubey, A. K. (2018). And now for something completely different: the congruence of the Altmetric Attention Score's structure between different article groups. *Scientometrics*, *114*(1), 253–275. https://doi.org/10.1007/s11192-017-2559-8

Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., … Haustein, S. (2018). The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*, *6*, e4375. https://doi.org/10.7717/peerj.4375

Rychlik, M. (2017). Perspektywy stosowania wskaźników altmetrycznych w ocenie dorobku polskiej humanistyki. In *Komunikacja naukowa w humanistyce, red. Kulczycki, E., Wydawnictwo Naukowe Instytutu Filozofii UAM, Poznań.* Retrieved from https://repozytorium.amu.edu.pl/handle/10593/17600