Data in Brief 19 (2018) 198–213

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Citation analytics: Data exploration and comparative analyses of *CiteScores* of Open Access and Subscription-Based publications indexed in *Scopus* (2014–2016)

Aderemi A. Atayero [a], Segun I. Popoola [a,*], Jesse Egeonu [b], Olumuyiwa Oludayo [c]

[a] *Department of Electrical and Information Engineering, Covenant University, Ota, Nigeria*
[b] *Department of Economics and Development Studies, Covenant University, Ota, Nigeria*
[c] *Department of Business Management, Covenant University, Ota, Nigeria*

ARTICLE INFO

Article history:
Received 6 January 2018
Received in revised form
3 February 2018
Accepted 2 May 2018
Available online 9 May 2018

Keywords:
Citation impact
Open Access
CiteScore
Analytics
Citation analytics
Smart campus
Data mining

ABSTRACT

Citation is one of the important metrics that are used in measuring the relevance and the impact of research publications. The potentials of citation analytics may be exploited to understand the gains of publishing scholarly peer-reviewed research outputs in either Open Access (OA) sources or Subscription-Based (SB) sources in the bid to increase citation impact. However, relevant data required for such comparative analysis must be freely accessible for evidence-based findings and conclusions. In this data article, citation scores (*CiteScores*) of 2542 OA sources and 15,040 SB sources indexed in *Scopus* from 2014 to 2016 were presented and analyzed based on a set of five inclusion criteria. A robust dataset, which contains the *CiteScores* of OA and SB publication sources included, is attached as supplementary material to this data article to facilitate further reuse. Descriptive statistics and frequency distributions of OA *CiteScores* and SB *CiteScores* are presented in tables. Boxplot representations and scatter plots are provided to show the statistical distributions of OA *CiteScores* and SB *CiteScores* across the three sub-categories (Book Series, Journal, and Trade Journal). Correlation coefficient and p-value matrices are made available within the data article. In addition, Probability Density Functions (PDFs) and Cumulative Distribution Functions (CDFs) of OA *CiteScores* and SB *CiteScores* are computed and the results are

* Corresponding author.
E-mail addresses: segunpopoola15@gmail.com, segun.popoola@covenantuniversity.edu.ng (S.I. Popoola).

https://doi.org/10.1016/j.dib.2018.05.005

presented using tables and graphs. Furthermore, Analysis of Variance (ANOVA) and multiple comparison post-hoc tests are conducted to understand the statistical difference (and its significance, if any) in the citation impact of OA publication sources and SB publication source based on *CiteScore*. In the long run, the data provided in this article will help policy makers and researchers in Higher Education Institutions (HEIs) to identify the appropriate publication source type and category for dissemination of scholarly research findings with maximum citation impact.

Specifications Table

Subject area	*Data Analytics*
More specific subject area	*Citation Analytics*
Type of data	*Tables, graphs, figures, and spreadsheet file*
How data was acquired	*Data was acquired from publication source list available in Scopus online database [1]. A set of five inclusion criteria was established namely: publication source must be indexed in the Scopus database; publication source must be active as at 28th December 2017; publication must be written in English language; publication source type must either be Book Series, Journal or Trade Journal; and publication source must have CiteScores in 2014, 2015, and 2016.*
Data format	*Secondary, analyzed*
Experimental factors	*Publication sources that did not meet any of the five criteria for inclusion in the period under consideration were excluded.*
Experimental features	*Descriptive statistics, boxplot representations, scatter plots, frequency distributions, correlation and regression analyses, Probability Density Functions (PDFs), Cumulative Distribution Functions (CDFs), Analysis of Variance (ANOVA) test, and multiple post-hoc test are performed to explore the dataset provided in this data article. All statistical computations were done using the Machine Learning and Statistics toolbox in MATLAB 2016a software.*
Data source location	*Data is available as supplementary material to this data article*
Data accessibility	*In a bid to facilitate further works on citation analytics, detailed datasets are made publicly available in a Microsoft Excel spreadsheet file.*

Value of the data

- The dataset generated and made publicly available based on the stipulated criteria will help foster further investigation into the importance of *Elsevier CiteScore* and other source ranking methods [2–4].
- Presenting this data in open access format will help researchers identify relevant sources as veritable outlets for dissemination of their research findings [5,6].
- Quite a lot of research findings often end up in subscription-only sources. This invariably limits access to such works and reduces their impact on future research significantly. This shortfall is mitigated by isolating and analyzing the OA sources of the largest global indexing body for scientific research [7–9].