Journal of Informetrics 12 (2018) 430–435

Contents lists available at ScienceDirect

Journal of Informetrics

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

# Dimensions: A competitor to Scopus and the Web of Science?

Mike Thelwall

*University of Wolverhampton,Wulfruna Street, Wolverhampton WV1 1LY, UK*

ARTICLE INFO

*Article history:*
Received 25 February 2018
Received in revised form 11 March 2018
Accepted 20 March 2018

*Keywords:*
Dimensions
Scopus
Web of Science
Citation analysis
Citation indexing

ABSTRACT

Dimensions is a partly free scholarly database launched by Digital Science in January 2018. Dimensions includes journal articles and citation counts, making it a potential new source of impact data. This article explores the value of Dimensions from an impact assessment perspective with an examination of Food Science research 2008–2018 and a random sample of 10,000 Scopus articles from 2012. The results include high correlations between citation counts from Scopus and Dimensions (0.96 by narrow field in 2012) as well as similar average counts. Almost all Scopus articles with DOIs were found in Dimensions (97% in 2012). Thus, the scholarly database component of Dimensions seems to be a plausible alternative to Scopus and the Web of Science for general citation analyses and for citation data in support of some types of research evaluations.



## 1. Introduction

Citation counts are used by researchers and research managers to help evaluate the quality or impact of published research, particularly when it is impractical to employ peer judgements or a second opinion is needed. In the early years of citation analysis there was a single pre-eminent data source for citation counts, Eugene Garfield's Science Citation Index (Garfield, 1964) but today Scopus has become a viable alternative (Archambault, Campbell, Gingras, & Larivière, 2009; Falagas, Pitsouni, Malietzis, & Pappas, 2008). There are also free online citation indexes, such as Google Scholar (Halevi, Moed, & Bar-Ilan, 2017; Martin-Martin, Orduna-Malea, Harzing, & López-Cózar, 2017; Prins, Costas, van Leeuwen, & Wouters, 2016) and Microsoft Academic (Harzing & Alakangas, 2017; Hug, Ochsner, & Brändle, 2017; Sinha et al., 2015; Thelwall, 2018). The existence of alternatives has three main benefits for research evaluators. First, the free alternatives may reduce the cost of evaluations and make informal impact self-evaluations possible for many researchers that would not pay to access data. Second, all citation indexes are imperfect and the availability of alternatives allows data from one to be cross-checked against the alternatives. Third, each citation index may have coverage advantages or capabilities that make it a better fit for a given impact evaluation task.

In January 2018, Digital Science launched Dimensions, a new online scholarly platform for publications, grants, clinical trials and patents, giving free partial online access (Adams et al., 2018). The platform replaced a previous grant analysis tool, also called Dimensions, to support "portfolio analysis and planning for science funders".[1] This article focuses on the publication component of Dimensions. Publications in Dimensions are categorised as articles (75,698,402 on 19 February 2018), chapters (9,525,334), proceedings (4,975,857), monographs (328,484) and preprints (19,734). The relatively small numbers of preprints and lack of other sources suggests that the Dimensions data is predominantly from publishers. The preprints originate from bioRxiv, and, according to its founder Christian Herzog (personal communication) and (private)

*E-mail address:* m.thelwall@wlv.ac.uk

[1] https://web.archive.org/web/20170820121951/http://www.uberresearch.com:80/dimensions-for-funders/.

https://doi.org/10.1016/j.joi.2018.03.006
1751-1577/

developer FAQ, it plans to index more preprint archives and some institutional repositories. In contrast, Microsoft Academic and Google Scholar also index web content from crawlers (Halevi, Moed, & Bar-Ilan, 2017; Harzing & Alakangas, 2017). This presumably means that the citation counts in Dimensions are lower than those of Microsoft Academic and Google Scholar but has the important advantage that its data is less easy to spam. Generating fake papers with self-citations and posting them to academic domains is an effective way to spam indexes that look for academic content online (Delgado López-Cózar, Robinson-García, & Torres-Salinas, 2014). Whilst there are also low quality academic journals (Gutierrez, Beall, & Forero, 2015) and unethical practices in peer reviewed publications (Chorus, 2015), these can be policed by the academic community (e.g., the Committee on Publication Ethics (COPE) claims 12,000 members: publicationethics.org/about) and are not as powerful as unlimited self-publishing to academic domains. Dimensions might also contain fewer data processing errors a result of avoiding web data. It therefore apparently fulfils a unique niche as a large scale partly free citation index that is protected against spam.

Given the potential value of Dimensions for research evaluations, it is important to assess its key properties to decide whether it contains enough data to be useful and whether its citation counts are plausible.

## 2. Research questions

The aim of this study is to give insights into Dimensions rather than to provide comprehensive information. As a young service, it may evolve soon, undermining the value of a detailed empirical analysis. Dimensions is compared to Scopus but not the Web of Science since Scopus has consistently been found to have greater overall coverage of academic journals (Mongeon & Paul-Hus, 2016; Waltman, 2016) and so represents best practice in terms of comprehensiveness. The following exploratory research questions drive the study.

1. How comprehensive is the coverage of Scopus journal articles in Dimensions?
2. Are the average citation counts for journal articles in Dimensions comparable to those of Scopus?
3. Are Dimensions citation counts for journal articles interchangeable with those of Scopus, in the sense of having a very high correlation with them?

## 3. Methods

The Scopus narrow category Food Science was chosen for an exploratory analysis. This is an average category from the perspective of online citation counts, with the median average (geometric mean) Microsoft Academic citation count of all Scopus narrow fields in a recent study (Thelwall, 2018). The years 2008–2017 were selected to allow an analysis of changes over time in the results and the data was collected in February 2018 from Scopus using its Applications Programming Interface (API). Citation counts and DOIs were extracted from the Scopus records. Articles without DOIs were discarded since these could not be easily matched accurately. Only documents recorded in Scopus as (standard) journal articles were used, excluding books, conference papers, reviews and editorials, for example.

In February 2018 Dimensions was queried by DOI for all journal articles with DOIs returned by the Scopus queries ($n = 84691$) and the Dimensions citation counts were recorded.

For RQ1, the coverage of Dimensions was compared against Scopus as a benchmark by calculating for each year the percentage of Scopus articles with DOIs that were also in Dimensions with a DOI. This is a one-way comparison since Dimensions may cover articles that are not in Scopus but should nevertheless give broad insights into whether Dimensions has substantial coverage of science.

To check whether Food Science is an unusual case, a random sample of 10000 articles with DOIs from Scopus in 2012 was also checked for matching records in Dimensions. This sample was selected using a random number generator from a list of the most recent 5000 (a system limitation) articles in all 326 non-empty Scopus narrow fields. Recycled Scopus data that had originally been collected 26 August 2017 for another paper was used for this.

For RQ2, the geometric mean citation counts for Dimensions each year were compared against those of Scopus. For this calculation, two different comparisons were made. For the first, articles not found in Dimensions were excluded. For the second, articles not found in Dimensions were included and given a citation count of 0. The geometric mean is a better measure of central tendency than the arithmetic mean because citation data is highly skewed (Fairclough & Thelwall, 2015).

For RQ3, citation counts from Scopus and Dimensions were compared for each year using Spearman correlations, as appropriate for skewed data. A high correlation suggests that the two may be interchangeable in practice for impact calculations. The correlations were calculated for data with and without articles not found in Dimensions, as for RQ2.

Although not directly addressing the research questions, Altmetric Scores and RCR (Relative Citation Ratio) values were also extracted from the data to provide additional context about Dimensions. Altmetric Scores provided by Dimensions are derived from Altmetric.com (Adie & Roe, 2013), and are weighted sums of all transparent scores collected by Altmetric, including citations from blogs, Twitter and Facebook but not Mendeley. RCR uses the co-citation network of an article to normalise its citation count (Hutchins, Yuan, Anderson, & Santangelo, 2016). This is more sensitive to the field of an article than the field of the journal publishing the article. This indicator has been criticised for a lack of transparency and technical problems with the calculation, such as with the method used to estimate the publication field of an article (Janssens, Goodman, Powell, & Gwinn, 2017). Dimensions also reports a Field Citation Ratio, which is field normalised citation score