Funders should mandate open citations

All publishers must make bibliographic references free to access, analyse and reuse, argues **David Shotton**.

Over the past two decades, open access to journal articles, software and research data has changed from aspirational to commonplace. However, truly open scholarship also requires that bibliographic references be freely available for analysis and reuse.

Citations — the links created when a published work acknowledges other works in its bibliographic references — knit together independent works of scholarship into a global endeavour, and they are important for assigning credit to other researchers.

Analyses of citations can reveal how scientific knowledge develops over time and illuminate patterns of authorship. Such information is essential for assessing scholars' influence and making wise decisions about research investment. Bibliographic databases and citation indices are also crucial to individual researchers: they enable automated tools to hunt for relevant papers throughout the literature.

Making reference lists from articles free to view is insufficient for these purposes; to be useful, open references must be stored in a machine-readable format in a centralized repository. Crossref, the DOI-registration agency used by most academic publications, has provided such a repository since 2000, but its references are freely available only if publishers explicitly specify that they be made open. Funders and the scientific community must push harder for this.

Last year was eventful for open references. In April, more than 60 publishers (including Springer Nature) responded to a call from the Initiative for Open Citations (I4OC) — an effort that I co-founded — to unlock the reference lists of their scientific articles. By September, more than half of the nearly one billion journal-article references deposited at Crossref had been made open, up from 1% before I4OC launched. Bibliometric visualizations using this open data set have already appeared. They reveal, for instance, how co-authorship maps within particular disciplines and, at a larger scale, links between disciplines. In December, an open letter signed by more than 250 scientometricians called for publishers to open up their references (see go.nature.com/2crblo9). For reasons of both international equity and methodological integrity, scholars need access to comprehensive open reference data, and they need to be able to show the raw data behind their analyses.

That is presently not the case. The two most authoritative sources of citation data are Clarivate Analytics' Web of Science, which grew from the Science Citation Index created by Eugene Garfield in 1964, and Elsevier's Scopus, launched in 2004. Neither is open or comprehensive. Most research universities pay tens of thousands of dollars annually to access one or both of them, whereas institutions and independent scholars that cannot afford such a cost have no access.

However, the idea that references are proprietary data is fading. In addition to the half-billion references already made open by Crossref,

the OpenCitations Corpus, the repository I run with computer scientist Silvio Peroni, has already published 12.8 million citation links from PubMed Central under a Creative Commons waiver that puts them in the public domain. These are fully curated and semantically enhanced in Linked Open Data format to assist automated analysis.

Two significant barriers prevent comprehensive reference availability through Crossref. First, although it is easy to do so, two-thirds of Crossref's publisher-members, in particular the smaller ones, do not submit references along with the other details of their publications.

The second obstacle is created by publishers that submit references to Crossref, but do not make them open. Elsevier is by far the largest member of this group, which also includes the American Chemical Society, IEEE and Wolters Kluwer Health. Elsevier deposits about one-third of all journal-article references stored by Crossref, these constitute nearly two-thirds of those that are not presently open.

> **WHERE POLITE ENCOURAGEMENT FALLS ON DEAF EARS, STERNER MEASURES ARE REQUIRED.**

The rationale for Elsevier not opening up its references is financial: free availability of its numerous bibliographic references would undermine Elsevier's ability to sell access to such data.

Companies such as Elsevier have invested considerable resources over many years into creating databases that can be used for bibliometric analyses. Elsevier argues that it is reasonable to charge for high-quality citation analysis, that curating citation data entails costs, including licensing fees, and that it cannot make reference lists from its journals freely available because it could not then afford to add value to these data.

However, I believe that Elsevier's decision not to open up its raw reference data is misguided. Because it is bad for scholarship, it cannot be good in the long term for a business that seeks to serve scholars. In an increasingly open world, Elsevier's reputation will suffer, and its publications will become less visible. Instead, Elsevier executives should have more confidence in the advantage their analytical services give them in the citations market.

I call on all parties who could potentially benefit — including researchers, librarians, bibliometricians, funders, academic and research administrators, governmental agencies, members of the general public, and other stakeholders committed to open scholarship — to campaign for comprehensive open access to bibliographic references, and to actively develop, support and use services providing such access. However, where polite encouragement falls on deaf ears, sterner measures are required. Specifically, major funders should extend their open-access mandates and require grant recipients to publish only in journals whose publishers ensure their references are open. ∎

David Shotton *is co-director of OpenCitations, and a senior researcher at the Oxford e-Research Centre, University of Oxford, UK. e-mail: david.shotton@opencitations.net*