

Citation data are now open, but that's far from enough

The reference lists of more than 60 million papers on the linking site Crossref are now openly available. That is welcome – but further steps must follow.

A few years ago, researchers would find considerable hurdles when attempting to study citation patterns to illuminate trends in a field, identify new areas of research interest or pinpoint questionable practices such as excessive self-citation.

First, they'd need to request access to one of the large scholarly databases containing citation data, such as Web of Science or Scopus. Even if access was granted, they wouldn't be able to make public the proprietary data on which their findings were based.

That is now changing. Most online papers are identified with a unique set of characters called digital object identifiers (DOIs). This system is administered by Crossref, a non-profit association based in Lynnfield, Massachusetts, that has around 15,000 publishers, funding agencies and other institutions as members. Last month, Crossref announced that the citation data associated with the more than 60 million journal articles in its database were now openly available for downloading and use.

That's largely thanks to the Initiative for Open Citations (I4OC), a collaboration between academic publishers, researchers and other stakeholders, which since its launch in 2017 has been encouraging publishers to make citation data open. Uptake in some quarters, including among some big publishers, was initially slow. A *Nature* editorial in 2019 called for those publishers still dragging their feet to jump on board (see *Nature* 573, 163–164; 2019). (Springer Nature, publishers of *Nature*, joined the initiative in 2018. *Nature*'s news team is independent of its publisher.)

The opening up of citation data is welcome. It means greater transparency and accountability for research studies designed to inform academics, funders and governments in their decisions about areas of research they should focus energy and money on.

But more is needed. Not all publishers index papers on Crossref, and not all indexed papers have citation data associated with them. One study published in July found that about one-third of papers indexed in 2021 are lacking such data (N. J. van Eck and L. Waltman. Preprint at <https://doi.org/10.31222/osf.io/smxe5>; 2022). Some of these articles – particularly editorials, letters, corrections and book reviews – might not have any references, but this by no means applies to all of them. Uploading citation data

should not be seen as optional.

There is an important caveat to any quest for bibliometric openness. The 2012 San Francisco Declaration on Research Assessment (DORA) states that metrics should never be used out of context or in isolation to judge researchers and their work. We should be careful not to place too much reliance on citation data, especially when evaluating scientists for promotions and job applications. But if used wisely, it can only be better to have such data open to all.

And openness should not end with citation data. Crossref also allows publishers to post other types of metadata, such as author affiliations, funding information, data- and code-availability statements, and ORCID IDs, which are used to identify individual researchers. However, not all publishers do this. In an open letter in June, the Open Research Funders Group, a partnership of philanthropic organizations – including the Bill & Melinda Gates Foundation and the Chan Zuckerberg Initiative – that advocates the open sharing of science rightly argued that such metadata should be made available (see go.nature.com/3qvfp3u).

Furthermore, the Initiative for Open Abstracts (I4OA), launched in 2020, has been pushing for abstracts of studies to be openly accessible. This would make it easier for researchers to discover, read and cite studies, and would open up more possibilities for analyses using machine-learning techniques, for instance to identify trends in the use of terms. (Such methods are already providing insights in other areas of science when data are opened up, for example in assessing the quality of peer-review reports; see page 455.) According to the July study, only 39% of the articles with a Crossref DOI indexed in 2021 have open abstracts – although that proportion has almost doubled since 2018.

Depositing all relevant metadata on Crossref should become the norm in scholarly publishing, as should generating DOIs for every paper. For those publishers that don't have the time or resources to do this, I4OC, I4OA and others in the open-science community have declared themselves ready to offer assistance.

Ultimately, all these moves must be only steps towards the goal of having all research papers openly available in their entirety. But until we arrive at that point, they are key to the transparency and reproducibility of research. They should be supported by all.

“If used wisely, it can only be better to have citation data open to all.”

Her Majesty Queen Elizabeth II (1926–2022)

Nature joins with many in the United Kingdom, in the Commonwealth and around the world in mourning the death of Her Majesty Queen Elizabeth II. She was a constant figure in public life, and the life of science, for more than 70 years. We send our condolences to the royal family and all those affected by her death.