

# Initiative aims to break science's citation payroll

Publishers agree to release proprietary data on references in millions of papers.

Quirin Schiermeier

06 April 2017 | Corrected: [07 April 2017](#)

Want to find out whether your articles are more highly cited than others? To get at the underlying data, [you'll have to pay](#). For decades, reliable, structured records of papers' authors and reference lists have been kept proprietary in two subscription databases, the Web of Science and Scopus.

Only researchers with access to these databases have been able to confidently trace citation patterns in the scholarly record and analyse the impact of particular research fields or institutions.

But citation data could soon emerge from behind their paywalls. The [Initiative for Open Citations \(I4OC\)](#) aims to allow anyone to access science papers' reference lists and to build analytical services on top of those raw data. The venture, started last year by the Wikimedia Foundation in San Francisco, California and five other partners, announced at its official launch on 6 April that 29 organizations, including some of the world's largest scientific publishers, have now agreed to openly release citation data.

"For the first time in history, swathes of scholarly citation data from the largest publishers — data that constitutes the very fabric of scientific knowledge — become available to the public with no copyright restrictions whatsoever," says Dario Taraborelli, head of research at the Wikimedia Foundation.

All science publishers already deposit their citation data at a non-profit organization, [Crossref](#), which the industry established in 2000. But until recently, only around 1% of that data had been freely available, Taraborelli says. Now, as a result of I4OC's efforts, some 40% of the data are free. "Our aim is to reach 100% coverage soon and to see more publishers and open-data organizations join the initiative," he says.

## Spot and fix errors

Publishers committed to the project include two of the initiative's co-founders: eLife and the Public Library of Science (PLOS). Other large publishers involved include EMBO Press, Wiley, Taylor & Francis and Springer Nature (the publishers of *Nature*). But Dutch publishing giant Elsevier, which contributes an estimated 30% of citation data on Crossref, is not yet on board. Elsevier also owns Scopus. (Web of Science is owned by Clarivate Analytics, which bought it from Thomson Reuters last year.)

"We are aware of the initiative but want to learn more before making a decision on whether to participate," says Tom Reller, vice-president of corporate relations with Elsevier in New York.

Making citation data open should have many advantages, says Catriona MacCallum, advocacy director with PLOS in Oxford, UK. In particular, she says, it should be easier to spot and fix errors in openly accessible citation records than it is to correct inaccuracies in closed commercial databases. The launch of I4OC means that any publisher, funder or researcher will be able to calculate the impact of their papers free of charge, potentially using new kinds of citation-based indicators that commercial firms don't yet provide.

But the drive for open records has a long way to go, MacCallum says. The records on CrossRef are raw data, not organized or structured so that non-experts can query them in useful ways (such as asking for the highest-cited paper published by a particular university in a particular year). "Building a structured database for users to be able to query and make full benefit of the data might take a few years," she adds.

Still, [open-knowledge projects such as Scholia](#) are already starting to integrate information to provide on-the-fly profiles of researchers, research topics, scholarly works and journals. And the open citation data from CrossRef is a critical resource for these services, says Finn Årup Nielsen, a data scientist at the Technical University of Denmark at Kongens Lyngby.

I4OC can't yet compete, in terms of coverage, with subscription databases that sell curated bibliographic records. "These databases are orders of magnitude larger than what we can provide today," says Taraborelli.

But he hopes that, in the long run, open-data management by a large set of players will outperform services by commercial providers. [Wikipedia ultimately overtook the popularity of services offered by Encyclopaedia Britannica](#), he notes. “I think the Wikipedia story teaches us something compelling about what commons-based communities can achieve,” he says.

*Nature* | doi:10.1038/nature.2017.21800

## Corrections

---

**Corrected:**An earlier version of this article erroneously stated that Elsevier owned 30% of the citation data on Crossref. In fact, it contributes this amount.