

Half of papers searched for online are free to read

Large study of open research analysed reader data from Unpaywall tool, which finds freely available versions of articles.

Dalmeet Singh Chawla

07 August 2017 Corrected: 09 August 2017

Almost half of the scholarly papers that people attempt to access online are now freely and legally available, according to a huge study that tracked 100,000 online requests for journal papers in June.

The work, published on 2 August in *PeerJ Preprints*¹, examined reader data from a web-browser extension called Unpaywall, which trawls the Internet to find free-to-read versions of paywalled papers.

The tool, which launched in April, was developed by two authors of the study, Jason Priem and Heather Piwowar, who co-founded the non-profit company Impactstory in Vancouver, Canada. It has been installed by more than 80,000 people worldwide and is used around 50,000 times a day, says Priem.

When Unpaywall users land on a journal paper, the tool queries a database called oaDOI — also developed by the pair — that contains records of all 67 million journal articles with digital object identifiers (DOIs), an identifier code widely used for academic publications. The widget then signals to the user whether a free-to-read version of the article is available.

Related stories

- Unpaywall finds free versions of paywalled papers
- Open-access website gets tough
- Half of 2011 papers now free to read

More related stories

Related stories

- Unpaywall finds free versions of paywalled papers
- Open-access website gets tough
- Half of 2011 papers now free to read

More related stories

The study authors analysed server logs of 100,000 papers that Unpaywall users tried to access during one week in June, and found that 47% of accessed studies were legally available to read for free

somewhere on the web. Around half the content being accessed was published in the past two years, says Priem.

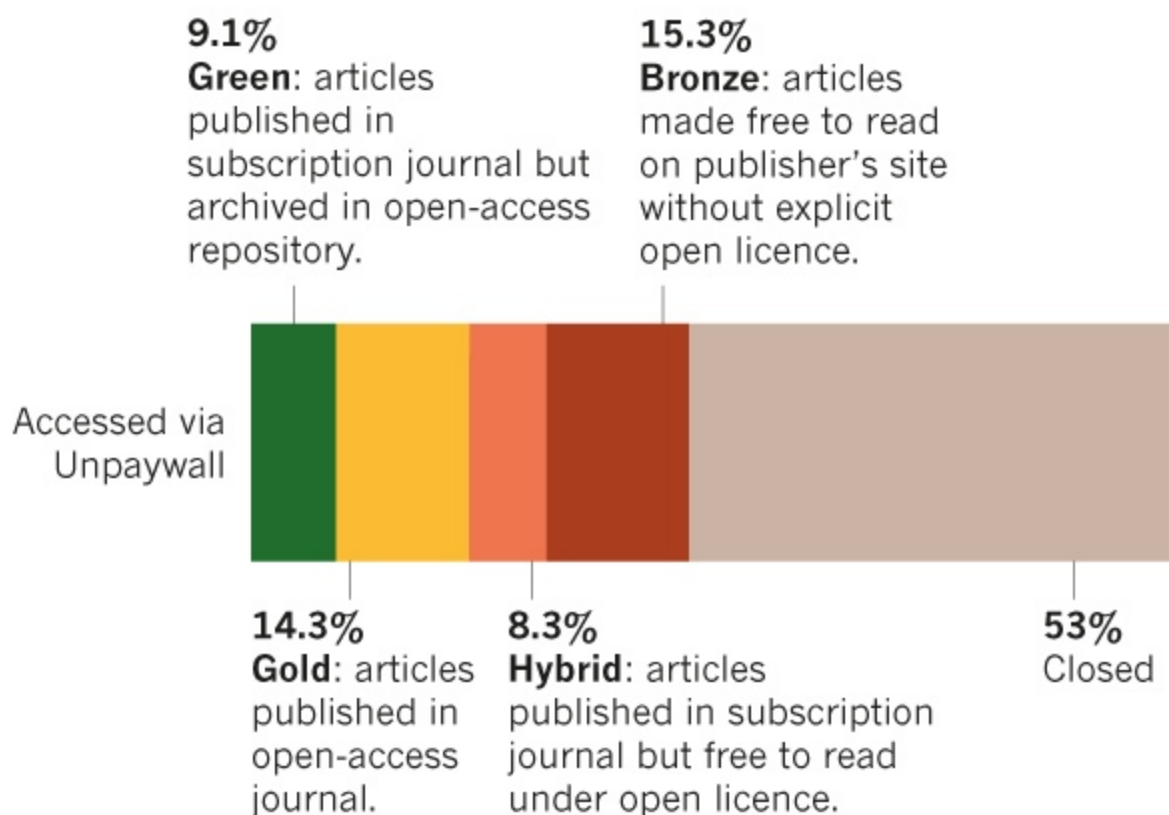
The study, which hasn't yet been peer-reviewed, is "careful and extensive", says Ludo Waltman, deputy director of the Centre for Science and Technology Studies at Leiden University in the Netherlands who edits the *Journal of Informetrics*.

The study authors say theirs is the first broad analysis of the state of open research since a 2014 report produced for the European Commission. But the two analyses employed different methods: the earlier one used automated software to search online for papers drawn at random from the Scopus database. It also scoured social scholarly networks such as Academia.edu and ResearchGate — which the Unpaywall study does not examine — and estimated that, at the time, more than half of peer-reviewed research articles published from 2007–12 were free to read online. Given the methodological differences, that's roughly comparable to the finding in the new work, Piwowar says.

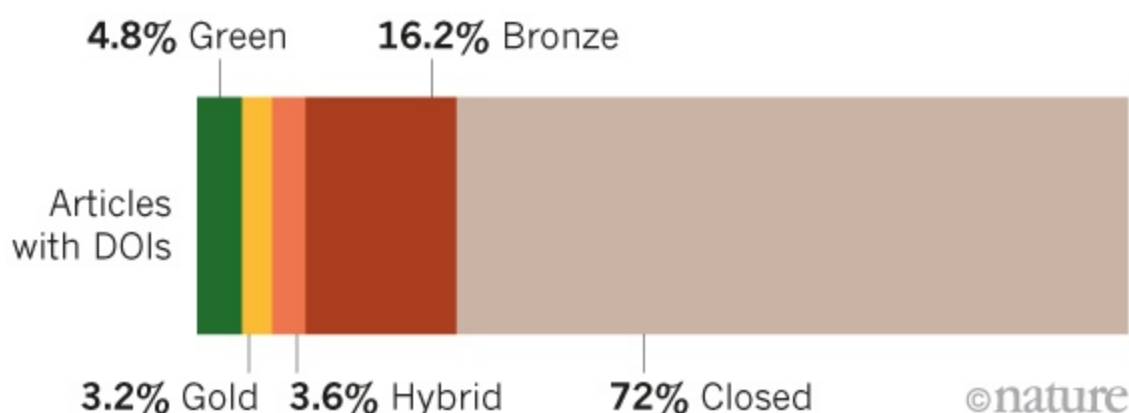
The latest work also delves into how papers become free to read. More than 20% of scholarly articles searched for through Unpaywall were available directly from journals, with clear licences describing whether the papers were free not just to read, but also to download or redistribute. Another 9% of the studies were still published behind a paywall, but authors later uploaded their paper — or some version of it, such as a peer-reviewed manuscript — to an online repository (see 'The state of open research').

THE STATE OF OPEN RESEARCH

Nearly half of scholarly articles that researchers are trying to access using the browser extension Unpaywall are freely readable somewhere on the web, finds an analysis. Around half were published in the past two years.



In the scholarly literature as a whole — approximated as the 67 million journal articles with DOIs — 28% are free to read online.



The most intriguing category of papers were the 15% that were posted on a publisher's site as free to read, but without any explicit open licence. The authors say this type of open-access — which they call 'bronze', in contrast to the widely used 'gold' and 'green' definitions — has been scarcely discussed.

Citation complications

To measure the prevalence of free-to-read papers in the scholarly literature as a whole, the authors used oaDOI to identify the publication statuses of 100,000 articles chosen randomly from the 67 million journal articles available on the DOI registry Crossref. In this sample, 28% of articles were free-to-read, predicting a total of 19 million such articles in the literature. Of papers published in 2015 — the most recent year examined — 45% were freely available, which suggests that newer articles are more likely to be open.

The study also investigated the claim that open-access articles are more cited than paywalled studies. It analysed another random set of 100,000 papers from the 8 million indexed in the Web of Science database between 2009 and 2015, found that, for a given subject area and publication year, free-to-read articles are cited 18% more than the average.

The trend is supported by several previous studies², but some have questioned whether the effect exists. Waltman says that it's difficult to know for sure whether these studies are being cited more frequently specifically because they are open. To be certain, he says, one would need to check whether researchers citing the studies have access to paywalled content.

Priem says that one limitation of the study is that its samples included only articles with DOIs, which aren't always used by publishers in the arts and humanities disciplines and in the developing world.

Still, "the percentage of literature that is OA continues to grow quite steadily", he says. And that could have implications for academic libraries. As tensions over the costs of institutional subscription packages grow between universities and publishers, the finding that roughly half of recently published research may be available to read for free could "tip the scales toward cancellation for some institutions", the study says.

To Priem, the future looks open. "In the next few decades, we're going to be seeing nearly all the literature available freely."

Corrections

Corrected: An earlier version of this story stated that free-to-read articles are cited 18% more than paywalled articles; in fact, the comparison is with the average for all articles for a given subject area and publication year.

References

1. Piwowar, H. *et al.* Preprint at <http://dx.doi.org/10.7287/peerj.preprints.3119v1> (2017).
2. Gargouri, Y. *et al.* *PLoS ONE* **5**, e1363 (2010).