

Wikipedia is open to all, the research underpinning it should be too.

*Often thought of as 'the last good place on the internet', Wikipedia plays a key role in the online information ecosystem by linking its entries to current and historic research papers. But, after following these links, how much of this research is openly accessible? Presenting evidence from a recent study, **Andy Tattersall**, finds that around 50% of research linked to Wikipedia from the White Rose Universities of Sheffield, York and Leeds, is openly accessible. As Wikipedia's stated aim is for its sources to be verifiable, he argues openness should be central to the use of research on Wikipedia.*

There has (too) long been a debate as to Wikipedia's relationship with academia and whether the former is a credible platform for sharing and citing research. Much of that discussion has been around how it is used by students. Whilst certain academics have had a tendency to turn their nose up at the popular encyclopaedia, not seeing it as a credible source of knowledge.

Yet, for many it is simply the first stop for authoritative information and as such it offers an opportunity for the research community to share their work with a huge, global audience. This happens when their research is cited as part of a Wikipedia entry, as the encyclopaedia is built upon evidence, not anecdotes. One of the first news features on Wikipedia in [Nature](#) in 2011 suggested that editing the platform could be an influential way of improving a researcher's visibility and communicating their work to the academic community. Bringing that forward a decade and we can see this go much further, as the [World Health Organization \(WHO\)](#) and the [Wikimedia Foundation](#) collaborated to expand the public's access to the latest and most reliable information about COVID-19.

At the White Rose universities of Sheffield, Leeds and York, we looked at how much of our research is cited by Wikipedia and more importantly, how much of that is available via Open Access. In an age where more research is being published open access, it can be very easy to assume that every link to a research paper in the world's most accessible encyclopedia is also freely accessible. Sadly much current and historical research is still behind a publisher's paywall. This of course undermines one of Wikipedia's three principal core content policies of 'verifiability' with the other two being 'neutral point of view', and 'not original research', meaning that it [does not publish original thought](#). If a piece of evidence is behind a paywall it becomes harder for someone to verify it for themselves, even though it might have been through a peer review process. It is also incoherent and ironic, like research papers on the topic of open access that are themselves behind a publisher paywall.

If a piece of evidence is behind a paywall it becomes harder for someone to verify it for themselves, even though it might have been through a peer review process

To deal with the issue of articles cited in both journal websites and repositories, Wikipedia introduced the option for dual references to be added to a Wikipedia citation, meaning that the repository version of a research paper can be included alongside one that might still be behind a subscription wall. Some institutions, such as Leeds have hosted their own Wikipedia Editathons to address a variety of issues, such as de-colonisation of Wikipedia which heavily favours [white, male content](#), in addition to linking to open access materials. Wikipedia does promote the use of the [OABOT](#) tool that facilitates making links to the OA versions of publications.

As part of our research we obtained data from Altmetric.com to explore how much research across the three universities had been cited by Wikipedia. The data showed there were 6454 Wikipedia citations across the three institutions (Sheffield 2523, Leeds 2406, York 1525). We used an Unpaywall API to check the DOIs of all articles appearing in the sample to explore which of these articles were open access via the Gold publishing model and through our institutional repository available under the 'Green' route.

The two tools we employed to explore this data, Unpaywall and Altmetric.com, are largely automated, whereas research that is cited in Wikipedia is created manually. To validate our sample we carried out a manual, random check of 100 Wikipedia citations from each of the three institutional datasets to check for accuracy and confirm that each paper was attributed to that institution correctly. We also checked that the open access status given by Unpaywall was correct.



The oldest publication that was available open access and cited in a Wikipedia entry was from 1910, whilst the oldest paywalled research article was published in 1922. The fact there is a paywalled journal article from 100 years ago is rather depressing in itself. It is noteworthy that publication data that is tracked in Altmetric.com appears to go back to as far as 1666. We also looked at which disciplines received the most citations and found Biological Sciences and Medical and Health Sciences had by far the highest number of citations for each institution. Several disciplines returned similar results across the institutions, whilst others did much better than their fellow White Rose universities. Physical Sciences research at University of Sheffield received considerably more Wikipedia citations than Leeds or York. The University of Leeds Earth Sciences and Chemical Sciences research received much higher numbers of citations than the other two. York led the way in History and Archaeology compared to Sheffield and Leeds.

Our sample indicated that around half of all academic citations on the platform are paywalled. This is a major flaw in the Wikipedia model.

All three institutions performed similarly well in terms of open access coverage in Wikipedia. York did best with 56% of their references openly available compared to Sheffield with 54% and Leeds with 52%. Even though that highlights a majority of open links, it also shows there is still some way to go for a truly open resource. The data from Altmetric.com also highlighted editing patterns with multiple Wikipedia entries edited by the same accounts. Sadly we do not know the source of these editors, but can only assume they are either academics or professionals working in that particular field or possibly citizen scientists with a keen interest in current research.

Our study reveals there is still much work to be done in opening up research citations on Wikipedia. Differences in coverage across disciplines also likely reflect wider issues around the availability of open access. However, Wikipedia's ethos of verifiability should extend to the accessibility of academic references. Our sample indicated that around half of all academic citations on the platform are paywalled. This is a major flaw in the Wikipedia model. Openly available published research helps support the development of Wikipedia. This in turn assists Wikipedia's ultimate goal of access to transparent and evidence-based knowledge. It would also lower barriers to access research, which ultimately is good for academics and society.

We appreciate that not everything is open for the rest of society and it might be some time before that happens. But, given Wikipedia's global influence and stated mission, the research that underpins each entry should be as open and accessible as possible. To take full advantage of this it requires a greater understanding amongst academics and Wikipedians as to the importance of citing open access works over those behind a paywall.

*This post draws on the authors' co-authored paper: Tattersall A, Sheppard N, Blake T, O'Neill K and Carroll C, [Exploring open access coverage of Wikipedia-cited research across the White Rose Universities](#), published in *Insights*.*

The content generated on this blog is for information purposes only. This Article gives the views and opinions of the authors and does not reflect the views and opinions of the Impact of Social Science blog (the blog), nor of the London School of Economics and Political Science. Please review our [comments policy](#) if you have any concerns on posting a comment below.

Image Credit: [Everett Bartels](#) via Unsplash.
