# Open access and Wikipedia

*Stuart Lawson[1]*

## Abstract

The Wikipedia community has been involved in open access for a long time so it is fitting that the Wikimedia Foundation has now implemented an open access policy. This talk will discuss the links between open access and Wikipedia, and discuss the history of the interaction between the two communities. Projects in the area range from tools to get more open access content into Wikimedia, such as the Open Access Media Importer, to infrastructure projects designed to improve citation data. Open access shares many of the goals and ideals that drive Wikipedians to share their knowledge with the world, and having access to open access resources is vital for those who contribute to the encyclopedia. This talk will examine in detail some of the specific projects going on at the moment and show how they support the long-term goals of creating a knowledge commons.

### Open access and Wikipedia

I've been asked to talk about open access and Wikipedia. My day job is at Jisc Collections, the UK's library consortium, where I'm a Research Analyst trying to understand the economics of the scholarly communication system, including open access. Before that I worked in an academic library. I've also been a Wikipedian for some time, with my primary off-Wiki contribution being helping to organise the open access strand of the Wikimania conference in London last summer.

Open access is when scholarly publications are made available for anyone to read at no cost and to re-use with minimal restrictions. That description could also apply to Wikipedia, although of course Wikipedia is not restricted to only scholarly knowledge. Open access shares many of the goals and ideals that drive Wikipedians to share their knowledge with the world. This similarity of purpose has resulted in various projects and initiatives linking open access and Wikipedia, the most notable recent one being the implementation of an open access policy by the Wikimedia Foundation.

### Open access

To give a very brief overview of open access, for those who are less familiar with it; it's about expanding access to the published results of research beyond those working in

---

[1] Research Analyst, Jisc, Brettenham House, 5 Lancaster Place, London WC2E 7EN

academic institutions with well-funded libraries – not just to all researchers, but to everyone with the ability to access and use the internet. So far the two ways this has been achieved are by publishing in open access journals – open access to books, and theses, and other research outputs is also happening, and that's important, but most of the work around open access has focused on journal articles so far – publishing in open access journals, so the official version of record of the journal article is made openly available from the moment of publication; and also through depositing work in open access repositories, which are online archives that are either run by higher education institutions or are sometimes subject-specific archives.

Open access has grown significantly over the last 20 years. According to one EU-funded study last year (Science-Metrix 2014), over 50% of recently-published articles are now available as open access, either through journals or repositories. So a large proportion of research is now openly available, and this is having a lot of repercussions for all stakeholders in the scholarly communication system: for publishers who are having to re-think their business models; for researchers, who have to think differently about where they publish; and for academic libraries, who often now manage open access for their institutions, in addition to their other roles.

This all relates to Wikipedia in a number of ways. The first, and most obvious, is that in order for Wikipedia to be a comprehensive and reliable source it needs to be able to cite the most relevant and up-to-date research literature, and if contributors – the majority of whom are outside of the university system – now have access to more research themselves, then they can incorporate the knowledge from those articles into the encyclopedia. The more research is open access, the easier this can be.

A key facet of open access is open licensing. Open licenses – which in this context, for all practical purposes means Creative Commons licenses, if we're talking about text rather than software – these are legal mechanisms used by the creators and publishers of works to explicitly allow their work to be used and re-used in certain ways.

Some definitions of open access, such as the Berlin Declaration (often referred to as a founding document of the open access movement, although it was published in 2003 which is actually over a decade after the first open access journals and archives appeared; open access is almost as old as me, but that term wasn't always used), are very close in wording to the Creative Commons Attribution license (CC BY) when they describe the permissions that should be allowed for using and re-using open access research.

Wikipedia, of course, is itself licensed under a [Creative Commons Attribution-ShareAlike license](#) (CC BY-SA). This means that if open access research is released under an open license – CC BY, CC BY-SA, or the CC0 public domain waiver – then its text can be directly incorporated into the encyclopedia. The most extreme example of this are the [topic pages published by PLOS Computational Biology](#). First published in 2012, these are short, Wikipedia-style articles published both in PLOS Computational Biology, and on Wikipedia itself. So the version on PLOS acts as the version of record, and the Wikipedia version was then subject to the normal revisions and additions as any other Wikipedia article. So that's a very different afterlife for an academic text than usual, with anyone able to come along and directly amend it. Wikipedia has incorporated out-of-copyright and openly-licensed texts from its early days, but this was the first time that a scientific article was written directly for both academic and Wikipedia audiences. For people who are interested in peer review reform and open peer review – you can't get much more open than allowing anyone, including anonymous authors, to change what you've written. So it places a great deal of trust in the Wikipedia community; trust that's been earned by the high quality of much of the encyclopedia and the ongoing commitment of the volunteer contributors.

### History of Wikimedia and open access

I'm going to give you a little history about the interaction with the Wikimedia and open access movements.[2]

The Wikimedia movement has a long history of engagement with research ([https://blog.wikimedia.org/2014/01/15/wikimedia-and-open-access/](https://blog.wikimedia.org/2014/01/15/wikimedia-and-open-access/)), as I'm sure this audience is aware. Through the [Research Committee](#) (which oversees the monthly [Research Newsletter](#) published as part of the Signpost), through support of the [WikiSym (now OpenSym)](#) annual conference series, or through other forms of [significant support](#) such as funding or endorsement.

The Foundation has also publicly supported open access. The research committee responded to public consultations such as the [EU public consultation on the nature of scientific information in the digital age](#) in 2011, and the request from the White House [for Information on Public Access to Peer-Reviewed Scholarly Publications Resulting From Federally Funded Research](#) in 2012. Following that, the White House released a [directive](#) that requires all large public research funders in the US to develop public access policies.

---

[2] The following two paragraphs draw heavily from [https://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Signpost/2015-03-25/News_and_notes](https://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Signpost/2015-03-25/News_and_notes) by [User:Resident Mario](#) and [User:Tony1](#).

Away from the Foundation and at the grassroots level; WikiProject Open Access was started in 2012 as a central place to work on both Wikipedia articles about open access – the main article for open access actually goes back to 2003 – and projects which involve open access within Wikipedia. The level of activity varies over time, but there are over 100 Wikipedians who have signed up as members.

Every year Wikimedia holds an annual conference for its contributors, called Wikimania, now in its 11th year. All sessions are proposed and run by the Wikimedia community, plus a few invited keynotes. The last few years have all had some open access-related sessions, with a whole strand dedicated to open scholarship at last year's London event, and as invited speakers we had the open science advocate Peter Murray-Rust and the CEO of PLOS, Elizabeth Marincola. Sessions included a panel featuring representatives from SPARC, Open Knowledge Foundation, Open Society Foundations, and Jimmy Wales, co-founder of Wikipedia. So that was a strong line-up of organisations who are pushing the boundaries of open access showing their support for the importance of the interaction between the Wikimedia and open research communities.

In explaining our decision to include open scholarship so prominently in the event we proposed that Wikipedia and the other Wikimedia projects don't just benefit from the results of open access research, but can be seen as examples of open research themselves, because the process of their creation is transparent and designed to encourage wide participation, and results in a body of work that contributes to global knowledge. One statement from the website ran: 'At Wikimania, we invite you to imagine what it would be like if the process of scholarly research and other areas of knowledge creation were as open as this.' So we were suggesting that if writing the summary of current scholarly knowledge, in the form of encyclopedia articles, can be an open participatory process, why not extend that ethos towards earlier stages in the research process, as we see within the movement towards open science.

In this year's Wikimania one of the sessions is on the Wikimedia Foundation's new open access policy.

Policies and mandates have been one of the instruments that has helped to accelerate the adoption of open access worldwide. They have been widely used by research funders and research institutions. Typically, a mandate will require that any research that is funded by a particular funder, or undertaken by a researcher at a particular institution, must be made open access. Most policies ask authors to deposit work in open access repositories, but we're now seeing an increase in policies which support publishing in open access journals as well.

The Wikimedia Foundation's policy, which was introduced in March this year, requires the results of research to be made open access not only if it's directly funded by the Foundation, but also if the researchers receive support in any other way, such as a letter of endorsement or use of office space or equipment. This policy is a logical step to take given the interest that Wikimedia has taken in open access over the years, and the benefits it's derived from it. It's also in keeping with open access policies implemented by other charitable foundations, notably the Wellcome Trust and as of this year the Bill and Melinda Gates Foundation. Initial work on a Wikimedia open access policy actually began in 2010 and was presented at Wikimania the following year[3] but it was never formalized into an actual policy until now.

### Projects

There are a few specific projects I'll now mention to give you an idea of some of the work done by Wikimedians.

The OA media importer is designed to increase the amount of multimedia scholarly research content in Wikipedia. It's a bot which crawls PubMed Central, which is an archive of journal articles run by the US National Institutes of Health – it has over three million articles in it; automatically harvests any media files associated with open access articles and published under a Wikimedia-compatible license; and then adds a copy to Wikimedia Commons. To date it has uploaded over 19,000 media files to Commons, significantly enriching the audiovisual resources available to Wikipedians for inclusion in articles, and raising the visibility of research-based video and audio more generally; Wikimedia Commons files tend to appear highly on Google search results.

Throughout this talk I've mostly been concentrating on Wikipedia rather than the other Wikimedia projects, but open access does touch on some of the others too. Wikimedia Commons, as an archive of openly-licensed media files, is beginning to play an important role in aggregating open access content that is not collected anywhere else in the same way. The other project it's important to mention here is Wikisource. This is an openly-licensed archive of textual resources, and since the vast majority of scholarly outputs are text, there is a lot of potential to include open access research in Wikisource. As opposed to Wikipedia, Wikisource is focused on preserving the text exactly as it is in the original, so it performs a much more traditional archival role, which might make it easier for more traditionally-minded organisations like publishers to engage with.

---

[3] https://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Signpost/2015-03-25/News_and_notes

The OA signalling project is a project led by the researcher and Wikimedian Daniel Mietchen to include information in each citation in Wikipedia to show whether or not the citation is to an open access work, thus allowing readers to see whether they can follow the link and read the underlying source. The exciting part of this project is that if the article is open access, tools will automatically find the text, media files, and metadata, and add them to Wikisource, Wikimedia Commons, and Wikidata. Further progress for the project now depends on third-party activities, including development of Wikidata, although a lot of the basic technical functionality has been built. Full-text imports from PubMed Central into Wikisource use JATS as the XML format via a JATS-to-Mediawiki tool. Along with the Open Access Media Importer, this tool has discovered many issues with JATS that are now being worked on by the JATS for Reuse working group at http://jats4r.github.io/. So work within Wikimedia can feed back into other communities.

I should mention that these activities could potentially have important implications for accessing scholarly journal articles in the Global South through Wikipedia Zero. Wikipedia Zero is the service whereby people in participating countries – 59 so far – can access Wikipedia on mobile devices free of data charges. Openly licensed scholarly articles which are mirrored in Wikisource could be accessed through Wikipedia Zero. Research that gets cited in a Wikipedia article we'd hope would be among the most important and relevant work on that topic, so if the OA signalling project achieves its aims, that would mean that the most important research, *if it is open access*, would be made available to the widest possible audience.

In order to ensure that the most important open access research is cited in Wikipedia, a project was started called Open access reader. The aim of this is to find the most important and highly cited open access research on any given topic, and then try to make sure it's cited in Wikipedia by bringing it to the attention of editors who are members of WikiProjects related to that article. The story of the Open access reader project highlights something typical of work within Wikimedia which is that if projects are volunteer-run, they tend to stop-and-start unless there are lot of people working on them; and in the case of the Open access reader, this actually had some funding from the Foundation, but only for a short time, after which not much progress has been made. But we're still hoping it will be developed further.

### Wikidata

The launch of Wikidata, the Wikimedia project which stores structured data for Wikipedia and the other projects, has led to some interesting possibilities. I've already mentioned how the OA signalling project will use Wikidata to host bibliographic data, and a lot of the other developments that are currently underway to do with citation infrastructure are happening within Wikidata. Just two weeks ago a new Wikidata

[WikiProject Open Access](#) was started to co-ordinate this. Personally I'd like to see Wikidata used to host journal- and article-level bibliographic and citation data, subverting the need for closed commercial databases.

A related WikiProject, [Wikidata for research](#), has produced a project proposal for EU funding under the Horizon 2020 program. This proposal was publicly drafted on the wiki and is itself openly licensed.[4] This project aims to integrate Wikidata into other research infrastructure, positioning it as an open data platform to facilitate open science. There was an interesting [blog post](#) recently by employees of CrossRef and PLOS – writing in a personal capacity – about the need for scholarly communications infrastructure to be under community control rather than subject to commercial interests. Perhaps Wikidata will prove to be one such piece of infrastructure.

### Scholarly communication system

I've diverged slightly here from talking strictly about open access, and that is because the scholarly communication system is a complex one which relies on many different pieces of infrastructure and communities of practice, and there is a possibility of increasing the levels of openness throughout the system; from open research data and open citation data, through to open peer review systems and open source publishing platforms. And I think the current interests among those Wikipedians who advocate for open access is very much focused on open data and infrastructure, and on how the Wikimedia projects can play a role in the wider research environment.

But open access to research outputs remains of central importance to improving the quality of Wikipedia and the relationships between the Wikipedia and research communities. When researchers publish work about Wikipedia which is closed access, most of the community won't be able to read it; so for those of you who are Wikipedia researchers and don't currently publish all of your work as open access I would strongly encourage you to do so, whether it's in an open access journal or in a pre-print archive or repository, which most closed access journals allow authors to use. This would allow Wikipedians to learn more about other perspectives on the projects they are involved in, as well as offer you the usual advantages of open access – increased exposure, higher citations etc.

### Wikipedia as the expanded library

Now (finally) I'd like to refer to the theme of the conference, 'Wikipedia as the expanded library'. I hope it's clear why open access fits with this theme but I think it's worth

---

[4] [https://www.wikidata.org/wiki/Wikidata:WikiProject_Wikidata_for_research/EINFRA-9-2015](https://www.wikidata.org/wiki/Wikidata:WikiProject_Wikidata_for_research/EINFRA-9-2015)

discussing this in more depth. Libraries have always been central to scholarship and the spread of knowledge by providing access to texts; preserving the scholarly record, and enabling access to it. Since the establishment of public libraries in the nineteenth century, they've also played a key role in making knowledge available to the general public. This knowledge dissemination role is one that Wikipedia complements really well. One of the sessions from last year's Wikimania was entitled 'Wikipedia as the front matter to all research', which I think is a nice analogy; like with a traditional print journal when you have the front matter that explains newly published research and puts it in context. Wikipedia has the potential to act in a similar role but for the whole of research. Wikipedia articles are often a great way to be introduced to a topic, and then libraries provide both the sources that articles are based on and also further reading when you need to know more detail.

Open access is important to libraries because it unlocks a wealth of knowledge which becomes available to library users, without directly increasing libraries' acquisitions budgets, but with librarians still playing a key role in mediating access to these resources and helping with discovery. In this way, open access and Wikipedia can both be seen as supporting the mission of libraries. They both make huge amounts of knowledge openly available, but the library's purpose of providing information to people by collecting, preserving, and facilitating access to information resources is still as vital as ever. Even if physical collections decline in importance, the librarians' role as a guide to finding information still remains. Perhaps this is where librarians in research libraries can help to promote Wikipedia. A number of universities, sometimes through the Wikimedian-in-residence program, now engage with it by teaching students how to edit, and using the experience as an information literacy tool to show the importance of evaluating information sources.

So to conclude: we can envisage Wikipedia as the gateway to knowledge, open access research as the deep well of scholarship lying behind it, and librarians and educators as the guides to understanding.