

# Wikimedia and open research: a Primer from UKRN



Authors: [Martin Poulter](#), [M.L.Poulter@bristol.ac.uk](mailto:M.L.Poulter@bristol.ac.uk), [Nick Sheppard](#),  
[N.Sheppard@leeds.ac.uk](mailto:N.Sheppard@leeds.ac.uk), <https://www.ukrn.org/primers/>  
Reviewer: [Daniel Mietchen](#), [daniel.mietchen@roninstitute.org](mailto:daniel.mietchen@roninstitute.org)

## What is Wikimedia?

Wikimedia is a group of interconnected free knowledge projects hosted by the Wikimedia Foundation, a US charity. These projects are supported by extensive, interconnected communities of affiliates and volunteers who edit, translate, and otherwise maintain the sites. The best known of these projects is Wikipedia, the internet's most popular informational site and [one of the major referrers to DOIs \(scholarly publications\)](#). Wikipedia occupies a distinctive place in the information ecosystem, linking informal discussion to scholarly publications.

Two Wikimedia projects have a special role, providing digital media or data to the other projects (as well as to sites and applications external to Wikimedia).

- [Wikimedia Commons](#) is a repository of openly licensed media files including photographs, diagrams, video and audio. It acts as a common repository for the various projects of the Wikimedia Foundation and makes it easy to embed media into a relevant article on Wikipedia for example.
- [Wikidata](#) is a store of structured open data that can be read and edited by humans or software. As well as storing basic facts about millions of topics, it points to other databases with information about each topic.

Other Wikimedia platforms where research outputs might be reused include:

- [Wikisource](#), a free library of out-of-copyright texts,
- [Wikibooks](#) and [Wikiversity](#), platforms for collaborative creation of open educational resources (OERs),
- [Wikispecies](#), a species database,
- [Wiktionary](#), a multilingual dictionary



Figure 1: some of the most relevant Wikimedia projects

Remixing existing work – whether text, media, or data – is key to how each community works. So all content on Wikimedia is freely licensed or in the public domain. “Non-

commercial” or “No Derivatives” licences (which restrict reuse) are not allowed; most content is Attribution only or Attribution ShareAlike.

### ***Why would I use Wikimedia in my research?***

The Wikimedia projects are the most popular informational sites on the internet, with about 25 billion page views each month. They are thus a way to make individual research outputs visible to a vast audience, citing and linking the source publication. As research outputs are increasingly freely licensed by default, their reuse on Wikimedia platforms is a way to maximise their reach. Wikipedia and Wikidata are used by the other major platforms through which people find information, including search engines and AI chatbots.

These sites are also platforms for contextualising and remixing. Once figures, code, and data sets are shared, the online communities can translate, combine with other research outputs, or reuse them in Wikipedia articles and other educational materials. Sharing these materials is a way to open up the research process, making the creation of a specific output such as a figure publicly verifiable. The wiki platforms take care of the long-term preservation and versioning of the materials.

In this way Wikimedia extends the principles and practices of openness described in other UKRN Primers on [Data sharing](#), [Open Access](#), and [Open code & software](#).

### **How can I use Wikimedia in my research?**

#### ***Reusing digital media and code***

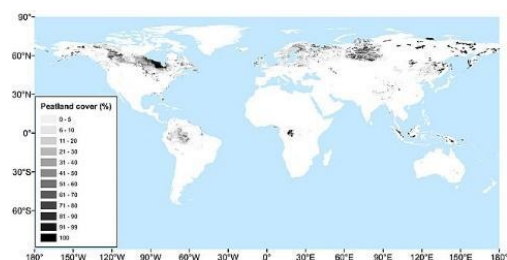
Wikimedia Commons is the digital media repository hosting images, video clips, and audio for the Wikimedia projects. Figures and other files from suitably licensed research outputs can be uploaded here and be available for use across the web or offline.

Files on Commons can have rich descriptions, including metadata about the file itself and full bibliographic data including DOIs for the source paper. Media already published on Flickr.com or Archive.org can be copied over, preserving their metadata and licence.

When media has been generated with code, [the source code can be included](#). For example, [this time series graph](#) comes with the Python code used to generate it, including a link to the source data set. As with Wikipedia, every page edit is recorded. The history of changes (viewed via the “History” tab) provides version control for updated or corrected code.

Wikipedia has more than 300 language versions, so media files will get more reuse if they are in a format that can be easily adapted. For images, this can mean Scalable Vector Graphics (SVG) format so text labels are easy to edit, or using numbered labels so that other language communities can add their own legend.

[This visualisation](#) of the [PEATMAP dataset](#) has been extracted from a research paper and uploaded to Wikimedia Commons, from where it illustrates articles about peat in eight languages, getting more than half a million views each year.



## Reusing text from papers

Text from suitably licensed papers can be added to Wikipedia articles. A template can be used to fully cite the source paper and indicate that it is being copied as well as cited.

 This article incorporates text from a scholarly publication published under a copyright license that allows anyone to reuse, revise, remix and redistribute the materials in any form for any purpose: Arnan, X.; Molowny-Horas, R.; Rodrigo, A.; Retana, J. (2012), Adler, Frederick R (ed.), "Uncoupling the Effects of Seed Predation and Seed Dispersal by Granivorous Ants on Plant Population Dynamics", *PLoS ONE*, **7** (8): 1–11, [Bibcode:2012PLoS...742869A](#), [doi:10.1371/journal.pone.0042869](#), [PMC 3413678](#), [PMID 22880125](#) Please check the source for the exact licensing terms.

For examples of Wikipedia articles improved with text from open access research papers, see [Genome diversity and karyotype evolution of mammals](#), [Golden snub-nosed monkey](#) (the "Social structures" section is from a research paper) or [Temnothorax](#).

## Reusing entire papers

If a paper's licence and style are compatible with Wikipedia, and it covers a topic for which there is no existing article, it can be pasted in to create a new article. An attribution template should be added to make clear the source of the text and give its full citation.

Since Wikipedia aims to summarise established mainstream knowledge rather than present new findings, most research papers are not suitable. Wikipedia is aimed at the general public, so highly technical papers need an accessible introduction. The ideal papers are those that survey a topic or that have an extensive summary of background literature. If the paper expresses recommendations or authors' opinions, these should be edited out. Even in an ideal case, some stylistic changes will be necessary to adapt the paper to Wikipedia, for example adding wiki-links to other articles.

For examples of Wikipedia articles that are adapted from research papers, see [Dicke model](#) or [Cameroon Press Photo Archive](#).

## Sharing data

Wikidata is a free knowledge base containing secondary data about more than a hundred million entities. It acts as a hub for the web of Linked Open Data, connecting thousands of different information sources. It fulfils the FAIR data principles (i.e. Findable, Accessible, Interoperable, and Reusable) and all its data must be free of licensing restrictions. Data from Wikidata are used across the Wikimedia projects and in many other apps, websites, and visualisations. As with Wikipedia, all factual statements should already have been published by a reliable source, for example a data repository.

Research outputs of any size can potentially be visible on Wikidata. Individual conclusions can be added to Wikidata's knowledge graph, with a citation to the research paper.

Both Wikidata and Wikipedia have been [used as crowdsourcing platforms by research databases](#). This involves sharing some items or identifiers on the wiki, monitoring changes, and re-importing statements or identifiers back into the source database. Vandalism is technically possible but rare, and can be detected by custom queries and manual checking. Examples include [GeneWiki's use of Wikidata](#) and the [protein families database Pfam's use of Wikipedia](#).

## **Adding links on Wikipedia**

Wikipedia does not aim to be a guide to the web, so “External links” sections of articles should only include very few links. Adding multiple external links is a characteristic activity of spammers, so can attract suspicion even if linking to legitimate research resources.

External links are welcome when they make it easier to verify an article’s citations. Where an open-access version of a paywalled research output is available — for example through an institutional repository — adding that link to an existing citation is helpful.

## **Managing conflicts of interest**

Wikipedia does not mind contributions from paid experts but is very resistant to paid *advocacy*, i.e. attempts to use the platform for marketing. The community is always alert to this behaviour, protective of the site’s goal to be purely descriptive and neutral. When editing Wikipedia, it is important to be aware of potential conflicts of interest. Contributors should be transparent about who they work for and avoid editing about themselves or their employers.

Contributors should declare on their user profile if they are editing on paid time. This is required by the terms of use of English Wikipedia. When there is an actual or potential conflict of interest, it is recommended to use Talk pages to request changes to articles rather than editing directly. If you only cite or link your own publications, the Wikipedia editor community may question if you are really there to improve the encyclopaedia.

## **Where can I find out more?**

Usability and documentation of the Wikimedia platforms can be intimidating to newcomers but some helpful introductions are available. Like any publisher, Wikipedia has a scope (what it will or will not publish), a house style, and standard ways to resolve disagreement; it is important to learn about these before contributing. An experienced trainer can help with this and also help to identify articles to improve. Wikimedia UK, the national charity supporting the Wikimedia projects, has a roster of trainers.

- Guidance for researchers about editing Wikipedia is found in Logan DW, Sandal M, Gardner PP, Manske M, Bateman A (2010) Ten Simple Rules for Editing Wikipedia. *PLoS Comput Biol* 6(9): e1000941. <https://doi.org/10.1371/journal.pcbi.1000941>
- A similar introduction to editing Wikidata is Shafee T, Mietchen D, Lubiana T, Jemielniak D, Waagmeester A (2023) Ten quick tips for editing Wikidata. *PLoS Comput Biol* 19(7): e1011235. <https://doi.org/10.1371/journal.pcbi.1011235>
- Guidance for adding text to Wikipedia is found at [https://en.wikipedia.org/wiki/Help:Adding\\_open\\_license\\_text\\_to\\_Wikipedia](https://en.wikipedia.org/wiki/Help:Adding_open_license_text_to_Wikipedia)
- For an introduction to sharing digital media on Wikimedia Commons, start with [https://commons.wikimedia.org/wiki/Commons:Project\\_scope](https://commons.wikimedia.org/wiki/Commons:Project_scope)
- Guidance on Conflict Of Interest (for example, editing about your own institution) is at [https://en.wikipedia.org/wiki/Wikipedia:Conflict\\_of\\_interest](https://en.wikipedia.org/wiki/Wikipedia:Conflict_of_interest)
- Expert outreach resources from Wikimedia UK can be found at [https://wikimedia.org.uk/wiki/Expert\\_outreach](https://wikimedia.org.uk/wiki/Expert_outreach)



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).