

Dear Sir/Madam,

please find enclosed a response to the "Request for Information on Public Access to Digital Data and Scientific Publications", submitted by the Wikimedia Research Committee (RCOM) on behalf of the Wikimedia Foundation. The full text of this response is archived at: [http://bit.ly/RFI\\_WH](http://bit.ly/RFI_WH)

Sincerely,

Daniel Mietchen, PhD  
*Wikimedian in Residence on Open Science and RCOM member*

Dario Taraborelli, PhD  
*Senior Research Analyst, Strategy, Wikimedia Foundation and RCOM member*

---

*This document contains a response to the "[Request for Information on Public Access to Digital Data and Scientific Publications](#)" by the Office of Science and Technology Policy of the White House, submitted on behalf of the [Wikimedia Research Committee](#) (RCOM) by Daniel Mietchen ([Wikimedian in Residence on Open Science](#) and RCOM member), with the endorsement from Dario Taraborelli (Wikimedia Foundation staff and RCOM member) on behalf of the Wikimedia Foundation.*

- Source of the RFI: <http://www.gpo.gov/fdsys/pkg/FR-2011-11-04/pdf/2011-28623.pdf> or <http://www.gpo.gov/fdsys/pkg/FR-2011-11-04/html/2011-28623.htm>
- Submission deadline: Jan 12, 2012 ([extended from Jan 2, 2012](#); [confirmation](#)).

#### Contents

[\[hide\]](#)

[1>About](#)

[2>Preface](#)

[3\\_Questions](#)

○ [3.1\\_Comment 1](#)

▪ [3.1.1\\_Comment 1a](#)

▪ [3.1.2\\_Comment 1b](#)

▪ [3.1.3\\_Comment 1c](#)

▪ [3.1.4\\_Comment 1d](#)

○ [3.2\\_Comment 2](#)

▪ [3.2.1\\_Comment 2a](#)

▪ [3.2.2\\_Comment 2b](#)

- 3.3\_Comment 3
  - 3.3.1\_Comment 3 a
  - 3.3.2\_Comment 3 b
- 3.4\_Comment 4
- 3.5\_Comment 5
- 3.6\_Comment 6
- 3.7\_Comment 7
- 3.8\_Comment 8
  - 3.8.1\_Comment 8a
  - 3.8.2\_Comment 8b

#### 4\_Additional comments

- 4.1\_Adaptation
  - 4.1.1\_Retouching
  - 4.1.2\_Translations and other modifications
  - 4.1.3\_Recontextualization
- 4.2\_Remixing
- 4.3\_Supplementary files
- 4.4\_Discoverability
- 4.5\_Wide exposure
- 4.6\_Speed

#### 5\_Conclusion

#### 6\_References

#### 7\_See also

- 7.1\_Publicly available responses by third parties
  - 7.1.1\_American Library Association and the Association of College and Research Libraries
  - 7.1.2\_Björn Brembs
  - 7.1.3\_Cameron Neylon
  - 7.1.4\_Ecological Society of America
  - 7.1.5\_Geoffrey Irving
  - 7.1.6\_Harvard University
  - 7.1.7\_Heather Etchevers
  - 7.1.8\_John Wilbanks
  - 7.1.9\_Kitware

- [7.1.10\\_National Digital Stewardship Alliance](#)
- [7.1.11\\_Stevan Harnad](#)
- [7.1.12\\_University of Oregon](#)
- [7.2\\_Similar public consultations](#)

## [\[edit\]](#) Preface

---

The [mission](#) of the Wikimedia Foundation is to empower and engage people around the world to collect and develop educational content under a free license or in the public domain, and to disseminate it effectively and globally. To this end, it operates a number of projects - most notably Wikipedia, Wikibooks, Wiktionary, Wikinews, Wikispecies, Wikiversity, Wikiquote, Wikisource and Wikimedia Commons - that provide reusable licensed free content in multiple languages.

Scholarly publications play multiple roles in these activities:

- They serve as references, providing support for a claim made in an article in one of our projects. Traditional subscription-based access modes to the scholarly literature impose barriers for such use of the literature, since most of the participants in our projects do not have subscriptions to scholarly sources, nor an affiliation with institutions that do.
- If suitably licensed, they can serve as scaffolds for or building blocks of articles in our projects. The projects run by the Wikimedia Foundation employ a [Creative Commons Attribution Share-Alike 3.0](#) license, which means that anyone can use the content for any purpose, as long as the source is acknowledged and derivative works are being made available under the same conditions. It also means that works licensed in the same way or more liberally (i.e. with a [Creative Commons Attribution](#) license, [Creative Commons Public Domain Dedication](#)) or already in the Public Domain can be incorporated into our projects, either directly or in translated form.
- Similarly, images and media originating from suitably licensed sources can be used to illustrate our educational content.

To highlight this aspect of reusing materials from Open-Access sources, our responses to the individual questions have been drafted in public and shall build and comment on publicly available responses submitted by others. A link to these responses will be provided the first time we quote them. An overview of all public responses that we consulted in drafting our own is given in the [Publicly available responses by third parties](#) section. In addition to commenting on the issues raised in the questions, we have added a [separate section](#) with examples and comments that highlight how Open Access materials can be reused on projects like those we operate.

## [\[edit\]](#) Questions

---

***"Office of Science and Technology Policy seeks further public comment on the questions listed below, on behalf of the multi-agency Task Force on Public Access to Scholarly Publications."***

[\[edit\]](#) **Comment 1**

[\[edit\]](#) **Comment 1a**

- Are there steps that agencies could take to grow existing and new markets related to the access and analysis of peer-reviewed publications that result from federally funded scientific research?

[Harvard](#) gave a comprehensive response, of which we are quoting the essence:

Yes, there are steps to grow new and existing markets arising from access to cutting-edge research. The most important step is to require public access (also called open access and free online access) to the final versions of the authors' manuscripts of peer-reviewed articles arising from publicly funded research.

and

Public access not only facilitates innovation in research-driven industries such as medicine and manufacturing. It stimulates the growth of a new industry adding value to the newly accessible research itself.

This new industry includes search, current awareness, impact measurement, data integration, citation linking, text and data mining, translation, indexing, organizing, recommending, and summarizing. These new services not only create new jobs and pay taxes, but they make the underlying research itself more useful. Translation can include translating scholarly materials into works designed for the end user, for example taking medical research literature and using it to write journal or newspaper articles or books.

Research funding agencies needn't take on the job of provide all these services themselves. As long as they ensure that the funded research is digital, online, free of charge, and free for reuse, they can rely on an after-market of motivated developers and entrepreneurs to bring it to users in the forms in which it will be most useful. Indeed, scholarly publishers are themselves in a good position to provide many of these value-added services, which could provide an additional revenue source for the industry.

[\[edit\]](#) **Comment 1b**

- How can policies for archiving publications and making them publicly accessible be used to grow the economy and improve the productivity of the scientific enterprise?

Again, the Harvard response comprehensively addresses these points:

Public access improves researcher productivity in many ways. By making published literature more visible and discoverable, public access prevents unintended duplication of effort. It prevents delays while researchers try to gain access to relevant articles they have discovered but cannot retrieve. It makes literature available to our hardware and software, not just to ourselves, and supports a fast-growing ecology of computer tools for mining and analyzing data and literature. It makes literature available to researchers outside the academy, such as

those based at hospitals, museums, non-profits organizations, and for-profit manufacturing companies. By enlarging the audience for research, public access multiplies the chances that prepared users will be able to make use of the research and translate it into clinical treatments or marketable products and services.

Some of these productivity gains can be quantified. See for example Karim R. Lakhani et al., "The Value of Openness in Scientific Problem Solving," Harvard Business School Working Paper, October 2006.

Excerpt: <http://www.hbs.edu/research/pdf/07-050.pdf>

Lack of openness and transparency means that scientific problem solving is constrained to a few scientists who work in secret and who typically fail to leverage the entire accumulation of scientific knowledge available....Our study finds that the broadcast of problem information to outside scientists results in a 29.5% resolution rate for scientific problems that had previously remained unsolved inside the R & D laboratories of wellknown science-driven firms.

#### [\[edit\]](#) **Comment 1c**

- What are the relative costs and benefits of such policies?

A [back-of-the-envelope calculation](#) already provides a flavour of the relative costs and benefits: dividing the 2009 *profits* made by Elsevier alone (ca. \$1 billion) by the total number of peer-reviewed articles published that year (ca. 1.5 million) results in about \$700 per article. Article processing fees at BioMed Central, Hindawi, Copernicus, Pensoft and other Open-Access publishers are frequently in that range or below. Bringing the profits of other publishers into the equation, the per-article rate would reach well into the range of fees charged by PLoS journals. [In other words](#),

collectively the top two or maybe three publishers take out of the academic world enough profits to pay for every research article in every discipline to be made freely available online for everyone to access using PLoS's publishing fee approach.

The matter of costs and benefits is treated in much more detail in the Harvard response, which highlights a number of studies that have been performed on the subject. They can be summed up with the following [comment](#) by SPARC, quoted in the Harvard response:

Preliminary modeling suggests that over a transitional period of 30 years from implementation, the potential incremental benefits of the proposed [FRPAA](#) [Federal Research Public Access Act] archiving mandate might be worth around 8 times the costs. Perhaps two-thirds of these benefits would accrue within the US, with the remainder spilling over to other countries. Hence, the US national benefits arising from the proposed FRPAA archiving mandate might be of the order of 5 times the costs. It is important to note that this is just the monetary cost savings. In the next few years, research to discover cheap, clean energy resources is desperately needed not only as a basic building block for the economy, but also for quality of all life on the planet. Humanities and

social research offers the potential to resolve social problems, enhance the quality of life and increase security  
- because the path to peace is through hearts and minds, not tanks.

#### [\[edit\]](#) **Comment 1d**

- What type of access to these publications is required to maximize U.S. economic growth and improve the productivity of the American scientific enterprise?

Harvard:

In practice, the way to free an article for use and reuse is to include a license or permission statement from the copyright holder explaining what the user may and may not do with it. An "open license" allows uses that would otherwise require the delay and expense of hunting down the rights-holder in order to ask permission.

and

a report evaluating the copyright licensing policies used by certain public and private funding agencies [...] recommended that research funders require the use of open licenses for funded research.

and

There are many open licenses. We recommend the Creative Commons Attribution (CC-BY) license, which permits any use provided the user makes proper attribution to the author. We recommend against licenses that bar commercial use (such as CC-BY-NC), in part because they would limit the utility of publicly funded research for businesses and industry.

This type of public-access policy would not be unprecedented for the United States. In January 2011, the Departments of Labor and Education launched the Trade Adjustment Assistance Community College and Career Training (TAACCT) program, a four-year, \$2 billion funding program for open educational resources to be released under CC-BY licenses. <http://www.whitehouse.gov/blog/2011/01/20/new-job-training-and-educationgrants-program-launched>

We also recommend against licenses that bar derivations (such as CC-BY-ND), since adapting existing resources to new needs is an important part of providing free and open educational resources of the kind the Wikimedia Foundation supports.

#### [\[edit\]](#) **Comment 2**

##### [\[edit\]](#) **Comment 2a**

- What specific steps can be taken to protect the intellectual property interests of publishers, scientists, Federal agencies, and other stakeholders involved with the publication and dissemination of peer-reviewed scholarly publications resulting from federally funded scientific research?

First, the use of the term "intellectual property" in discussions on access to the scholarly literature is misleading, since it comprises a number of legal codes that bear quite differently on research contexts. This aspect is covered in detail in the [Appendix](#) of the [Kitware](#) response.

Second, as explained in the comments by [Bremps](#) and in a related [post](#) by Cameron Neylon, publishers contribute very little to the publication of research that would fall under copyright, except some layout. The copyrights are with the authors, and for historic reasons associated with the distribution of printed copies, they are normally asked to transfer the copyright to the publishers, who then built their business models around providing toll access to the research literature. This system takes a lot of money out of research-related funds that could be used for purposes other than sustaining the profit margins of commercial publishers, so there is no scientific reason to keep supporting it. Instead, the research landscape would benefit from moves "to enshrine a simple principle in United States law: if taxpayers paid for it, they own it."<sup>[1]</sup>

[Irving](#) (emphasis added):

The easiest way to safeguard the rights of scientists is to make open access an optional but guaranteed right, as mentioned above. In the long term, mandatory open access to federally funded research may be desirable, but I believe most of the benefits can be achieved with rights alone.

The intellectual property rights of publishers could be maintained for **past** work by enforcing the right to open distribution only for **future** research. Although open access to past work is also highly desirable, a retroactive forced change to publishing agreements would constitute far more interference with private contracts, and is therefore more questionable.

#### [\[edit\]](#) **Comment 2b**

- Conversely, are there policies that should not be adopted with respect to public access to peer-reviewed scholarly publications so as not to undermine any intellectual property rights of publishers, scientists, Federal agencies, and other stakeholders?

As discussed above, the term "intellectual property" is misleading here, and in terms of copyright, it should be avoided that authors have to transfer copyrights to publishers at all. Instead, authors should retain copyright and grant publishers a non-exclusive right to publish the research under a free license - Creative Commons Attribution License (CC BY) for text, figures and media, Creative Commons Public Domain Dedication License (CC 0) for data and a free software license for code.

Consequently, if federal agencies reimburse authors for costs incurred by Open Access publication, these reimbursements "[should be limited to publication fees for true OA journals, not hybrid fees for subscription journals](#)", and they should only be partial if the publishers require transfer of copyrights from the authors.

#### [\[edit\]](#) **Comment 3**

##### [\[edit\]](#) **Comment 3 a**

- What are the pros and cons of centralized and decentralized approaches to managing public access to peer-reviewed scholarly publications that result from federally funded research in terms of interoperability, search, development of analytic tools, and other scientific and commercial opportunities?

Here, we follow Irving:

The pros of a decentralized approach are flexibility and diversity in exploring new methods of distribution, peer review, and search. The pros of a centralized approach are uniformity and simplified access, and institutional levels of security and redundancy.

The optimum, to ensure ongoing access and preservation of works, is multiple copies, so both centralized and decentralized. For practical reasons, a simple policy leaving operational details to the implementers is best. For example, PubMedCentral is a strong and now internationalized central model which works very well. Other disciplinary areas may find that a different approach works best. For example, the Research Papers in Economics service is a distributed repository.

### [\[edit\]](#) **Comment 3 b**

- Are there reasons why a Federal agency (or agencies) should maintain custody of all published content, and are there ways that the government can ensure long-term stewardship if content is distributed across multiple private sources?

There are reasons why either a Federal agency or a local university partner should maintain custody of all published content. One is that without such custody, access to the published results of U.S.-funded research could be available exclusively outside the U.S., if the publisher is owned by a private entity outside the U.S. Another is preservation. The vast majority of scholarly publishers do not attend to preservation at all. Even if a publisher does look after preservation, there is no guarantee that a private entity that changes hands will continue to provide this service, nor is there any guarantee that a private entity will choose to remain in the scholarly publisher business in perpetuity.

Full open access facilitates distributed archiving across a range of independently held repositories. In addition to that, as Irving pointed out, "a government database of open work would constitute both an alternative access point and a long term backup of published work, without interfering with nongovernmental mechanisms and institutions."

### [\[edit\]](#) **Comment 4**

- Are there models or new ideas for public-private partnerships that take advantage of existing publisher archives and encourage innovation in accessibility and interoperability, while ensuring long-term stewardship of the results of federally funded research?

#### [Cameron Neylon:](#)

There are a range of such models ranging from ArXiv through relatively traditional publishers like PLoS and BMC to new and emerging forms of low cost publication that disaggregate the traditional role of the scholarly publisher into a menu of services which can be selected from as desired. It is not the place of government, federal agencies, or even scholarly communities to attempt to pick winners at this very early stage of development. Rather the role of government and federal funding agencies is to make a clear statement of expectations as to the service level expected of the researcher and their institution as a condition of funding



and an appropriate level of resourcing the support the purchase of such services as required for effective communication of research outputs.

The role of the researcher is to select, on a best efforts basis, the appropriate services required for the effective communication of their research, consistent with the resources available. The role of the funder is to help provide a stable and viable market in the provision of such services that encourages competition, innovation, and the development of new services in response to the needs of an evolving research agenda.

The Wikimedia Foundation has also been building partnerships with academic institutions, organizations and publishers. For instance, several [scholarly societies](#) or [academic journals](#) are already partnering with Wikimedia projects or consider doing so.

### [\[edit\]](#) **Comment 5**

- What steps can be taken by Federal agencies, publishers, and/or scholarly and professional societies to encourage interoperable search, discovery, and analysis capacity across disciplines and archives?
- What are the minimum core metadata for scholarly publications that must be made available to the public to allow such capabilities?
- How should Federal agencies make certain that such minimum core metadata associated with peer-reviewed publications resulting from federally funded scientific research are publicly available to ensure that these publications can be easily found and linked to Federal science funding?

Enriching publications with metadata enables specific actions to be made to the content, rather than just labeling it. This facilitates the reuse of that content and the multiple components that it comprises (figures, tables, data, key words, semantic mark-up, etc.). It is important that the metadata model supports the appropriate context for the published works with a controlled vocabulary that permits reuse and interoperability between content platforms and databases. It is therefore also important to couple metadata with an API for standards-based data exchange.

Scholarly publishers should, at minimum therefore, support the National Library of Medicine's [Journal Article Tag Suite](#) XML standard for content and metadata as well as the [Dublin Core](#). Publishers should also deposit digital object identifiers (DOIs) into repositories such as [CrossRef](#) and, similarly, ensure they comply with requirements to allow unique author identification via platforms such as [ORCID](#), a central registry of unique identifiers for individual researchers, and an open and transparent linking mechanism between ORCID and other current author identification schemes).

To facilitate new initiatives in the semantic web, publishers should classify content with public domain taxonomies and thesauri and make these classifications available in machine-readable format in the source code. Standard taxonomies and thesauri that are created by, funded by, and recommended by scientific agencies and institutions should, whenever possible, be adopted and used by scholarly publishers to facilitate discovery across the platforms and silos that have been artificially created over the years. These efforts will also enable large-scale research projects across platforms, publishers, and resources.

Finally, there is a major need for standards to expose in a machine-readable format the provenance, attribution requirements, open access availability and reuse terms of research artifacts. The lack of a universally accepted solution allowing Web services and applications to discover and aggregate metadata about publications, datasets and other resources *based on their reusability* is hindering the innovation that the open access and linked open data initiatives represent for the scientific community. Federal agencies should invest resources to support and promote the discoverability and reuse of open licensed and openly accessible scientific contents.

#### [\[edit\]](#) **Comment 6**

- How can Federal agencies that fund science maximize the benefit of public access policies to U.S. taxpayers, and their investment in the peer-reviewed literature, while minimizing burden and costs for stakeholders, including awardee institutions, scientists, publishers, Federal agencies, and libraries?

Irving:

"much of the benefit of complete open access can be achieved simply by guaranteeing the right to freely distribute federally funded work. The costs for such a step are minimal, especially since efficient distribution and search mechanisms already exist, and will likely proliferate further if a greater fraction of publications can legally take advantage of them."

Brembs:

"The benefit is maximized by minimizing the costs associated with access. The costs are minimized by preventing third parties from adding costs to the process. One way to establish a short and thus cheap supply line is to have scholars deposit their work directly at their libraries, avoiding the costs of intermediaries such as publishers. The process of this deposition would still be identical to the current process (i.e., peer-review), albeit without intervening entities which withdraw funds but add little to no value."

#### [\[edit\]](#) **Comment 7**

- Besides scholarly journal articles, should other types of peer-reviewed publications resulting from federally funded research, such as book chapters and conference proceedings, be covered by these public access policies?

The output of all federally funded research should always be made available for reuse under a full open-access license. Although print distribution or author- and reviewer-level payments may incur additional costs, a publication fee paid to the publisher should cover these and guarantee the legal right of free online access and reuse of that material. This should include at least primary research (e.g., original research articles) and secondary research (such as systematic reviews in the medical literature) as well as conference proceedings, book chapters, and scientific protocols that are facilitated by federal funding.

#### [\[edit\]](#) **Comment 8**

#### [\[edit\]](#) **Comment 8a**

- What is the appropriate embargo period after publication before the public is granted free access to the full content of peer-reviewed scholarly publications resulting from federally funded research? Please describe the empirical basis for the recommended embargo period. Analyses that weigh public and private benefits and account for external market factors, such as competition, price changes, library budgets, and other factors, will be particularly useful.

Irving:

The embargo period should be negative: researchers should be allowed to freely distribute their work whenever they choose, including before publication or even before completion. This will minimize the delay between the completion of work and the time when other researchers can begin building on it. Producers of research that do not want their work to be immediately distributed can choose not to do so, and consumers of research that prefer to wait for the filtering and improvement effects of peer review are free to wait. Any other embargo period, even zero, would cause an unnecessary drag on the speed of dissemination and advancement of scientific knowledge."

Harvard:

If federal policy initially allows embargoes, then it should reduce their maximum permissible length over time, eventually to zero. We could support a plan to do this gradually rather than suddenly in order give publishers time to prepare.

Researchers should be encouraged to make their work openly accessible throughout the research process. If there is a flaw in the methodology, it is best to find out before starting the research project - not after results are written up, in the peer review stage. Share the data as soon as it is collected.

#### [\[edit\]](#) **Comment 8b**

- Are there evidence-based arguments that can be made that the delay period should be different for specific disciplines or types of publications?

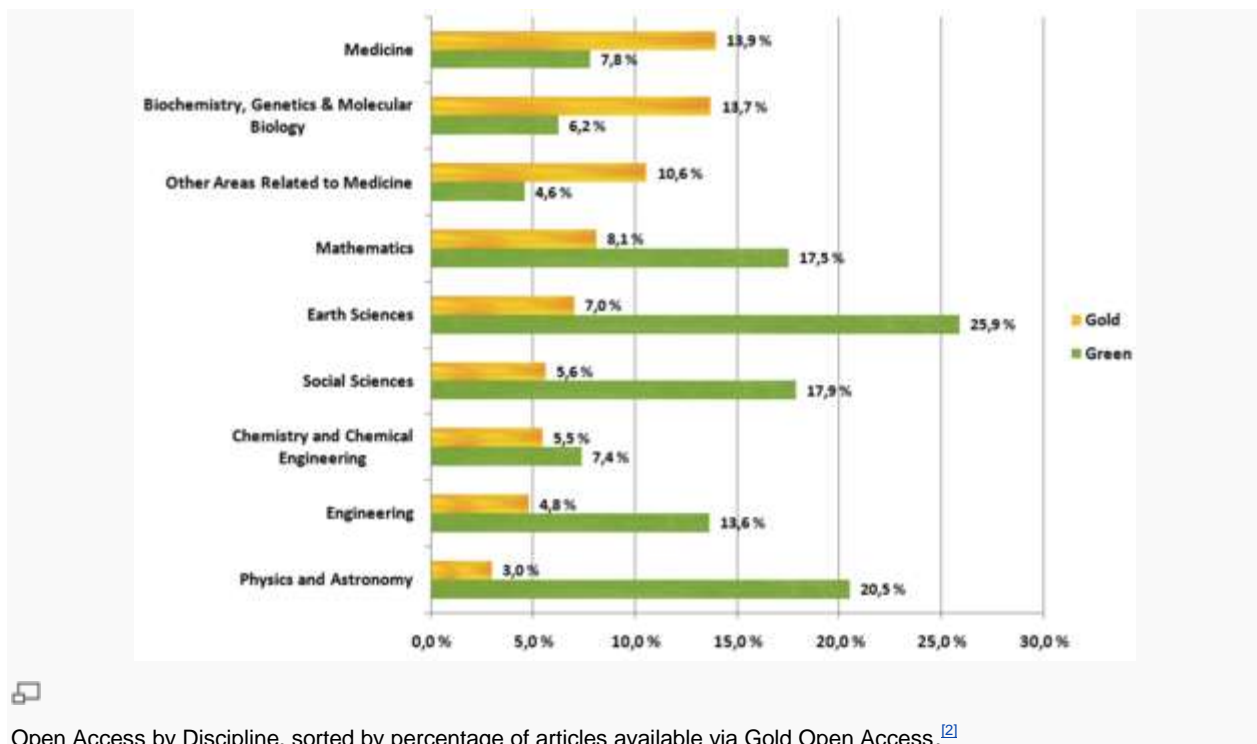
Harvard:

If government policy is to allow embargoes, even temporarily, it might allow different embargoes in different fields, on the ground that the demand for articles seems to drop off at different rates in different fields.

However, we have not seen good data on these different rates of demand decay. Publishers have this data, and if they wish to support differential embargo periods, they should provide data to justify them. In any case, variable embargo periods would burden universities by making compliance with an agency in one field different from compliance with an agency in another field (in tension with Question #6 above).

#### [\[edit\]](#) **Additional comments**

---



Open Access by Discipline, sorted by percentage of articles available via Gold Open Access.<sup>[2]</sup>

The discussions around Open Access have started out with the goal to remove barriers that prevent *researchers* from accessing the research literature. For 2009, a study found that, across disciplines, about 20% of research articles are available by way of Open Access publishing or self-archiving, with both options reaching an overall similar level of around 10%.<sup>[2]</sup>

This means that the vast majority of research articles is still hidden behind paywalls but it also means that around 20% of the contemporary scholarly literature is freely available to anyone, including researchers. Given that about half of this minority is available under free licenses (mainly CC BY, with some CC 0, and CC BY-SA), it makes sense to start exploring potential use cases for such materials beyond providing access to researchers. Wikimedia projects provide a whole battery of such use cases and can serve as a bridge between the worlds of Open Access, science outreach and Open Education.

Files listed in the Category [Open access \(publishing\)](#) on Wikimedia Commons are all imported, possibly with modifications, from Open Access sources, and most of them have been incorporated into educational content across Wikimedia projects. In November 2011, they have received a total of [over 20 million page hits](#).

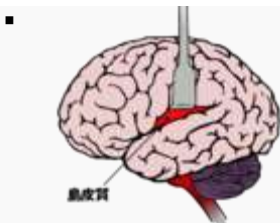
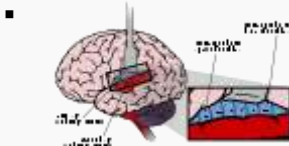
In the following, we will provide a selection of image and media files that originated from Open Access sources and have been reused in different ways on Wikimedia projects.

[\[edit\]](#) **Adaptation**

[\[edit\]](#) **Retouching**

The left image has been cropped from a figure in an [article](#) published by William M. Gray in PLoS Biology and was then retouched to remove an inset showing a pulse-chase analysis of auxin signalling in the wildtype and





The following group consists of an editable version of the top part of the first image above, along with numerous derivations thereof, especially translations. Together, the files in this section have generated [319184 page hits](#) in December 2011.



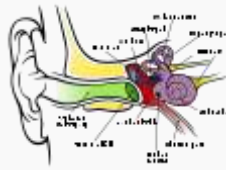
[Anatomy of the human ear – editable version](#)



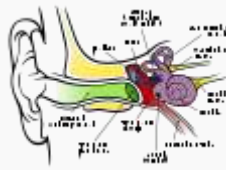
Česky



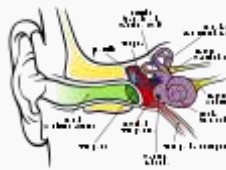
Česky



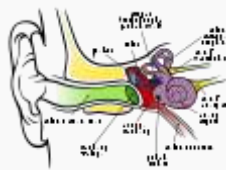
Deutsch



English



Español



Bahasa Indonesia





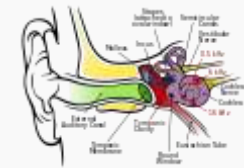
Українська



Without description



With numbers for description



English + cochlear frequency mapping

### [\[edit\]](#)Recontextualization

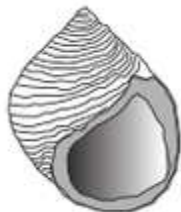
The following [two images](#) were originally published as part of two separate composite figures in an [article](#) by Banchob Sripa et al. in PLoS Medicine and now serve to illustrate pages as diverse as the entries about the dish Koi pla (ค็อยปลา) on the English, German and Thai Wikipedias and about the parasite *Clonorchis sinensis* on the English and Spanish Wikipedias. Together, these pages had [2035 hits](#) in December 2011.





In a [similar case](#), images from several composite figures from an [article](#) by Regina Cunha et al. in BMC Evolutionary Biology have been used on over 7,000 pages on the English and Vietnamese Wikipedias, getting [over 160,000 page views](#) altogether during December 2011.

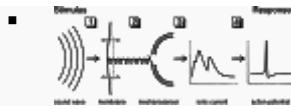






[\[edit\]](#) **Remixing**

Free licensing also allows to combine materials from different sources. The image on the left is a copy of the original file taken from an [article](#) by Tim Gollisch and Andreas Herz in PLoS Biology), converted to SVG. The image on the right has added colour, a representation of a sound source (a whistle, taken from [another freely licensed image](#)) and a mental representation of a sound.



A SVG copy of the original.



An illustration for [Qualia of sound](#)

## [\[edit\]](#) **Supplementary files**

Many scholarly articles nowadays have supplementary materials. These are rarely accessed at the publisher's sites, but those published under free licenses can be reused in other contexts, e.g. to illustrate pages on Wikimedia projects.



[The sound of a black smoker](#)





[Chlorobalius leucoviridis mimicking the song of its prey](#)



[Shimmering bees drive hornet away](#)



[Burying behaviour of an adult sandfish \(Scincus scincus\)](#)



[A lab notebook containing the record of research described in a scientific paper](#)

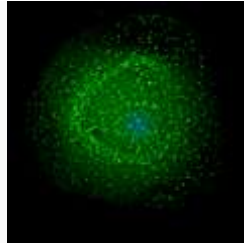
## [\[edit\]](#) Discoverability

Having an Open-Access image incorporated into a Wikipedia article may provide a considerable boost to the visibility and discoverability of the underlying research. The following image from Wikimedia Commons is the first result brought up by a (non-personalized) [image search for "Sorghum"](#). Originally published by Liza Gross [in PLoS Biology](#).



## [\[edit\]](#) Wide exposure

Wikimedia projects have various ways in which specific content can be highlighted. This way, freely licensed materials from Open Access sources can reach very large audiences.



A [Featured Picture](#). Originally published by Nolwenn Jouvenet et al. [in PLoS Biology](#)



[Picture of the Day on Wikimedia Commons](#), May 5, 2007. Originally published by Steve Bunkin [in PLoS Biology](#)



[Featured](#) on the main page of the Czech Wikipedia ([that receives ca. 150,000 hits each day](#)) in the first week of 2012. [Over 5,000 times](#), users clicked through to the image during that period. Originally published by Ewan Wolff et al. [in PLoS ONE](#).



Finalist for [Picture of the Year 2007](#) on Wikimedia Commons. Originally published by Liza Gross [in PLoS Biology](#).

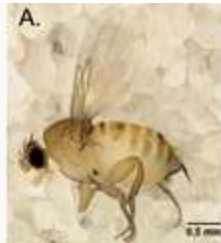
[\[edit\]](#) **Speed**



The Harvard response under [Comment 1d](#) mentions the delay normally incurred when trying to get permissions for reuse of previously published materials. Free licenses circumvent this problem, and it is thus not rare for Wikimedia projects to display content from Open Access sources within days or sometimes hours of publication.

A recent example is the case of the parasitoid fly *Apocephalus borealis* (left): last week, an [article](#) by Andrew Core et al. came out in PLoS ONE that associated the species with [Colony collapse disorder](#). The first edits on the matter were made the same day, and now, the image below is already in use at seven different pages related to the species or the disorder.

Another example is that of [Paedophryne amauensis](#), a species of frog that is now the [smallest known vertebrate](#). The description of the species was [published](#) (also in PLoS ONE) on late January 11, the respective entries in the [English](#), [Dutch](#), [Finnish](#), [Portuguese](#), [Russian](#), [French](#), [Bulgarian](#), [Hungarian](#), [Thai](#), and [Spanish](#) Wikipedias, at [Wikispecies](#) and [Wikimedia Commons](#) have all been started - with images imported from the paper - within less than a day from then on, and the article was featured "[In the news](#)" on the [Main page](#) of the English Wikipedia on January 12, 2012 ([WebCite](#)) as well as - with image - on the Thai Wikipedia ([WebCite](#)).



[An adult female fly, \*Apocephalus borealis\*.](#)



A *Paedophryne amauensis* frog on a [US dime](#).



Screenshot of *Paedophryne amauensis* as featured in the news section of the [Main page](#) of the Thai Wikipedia.

Similar [screenshot from the English Wikipedia](#).

## [\[edit\]](#) Conclusion

---

From Cameron Neylon:

To conclude, to focus on the final disposition of intellectual property arising from the authoring of research outputs relating to federally funded research is to continue a sterile and non-productive discussion. Given that the federal government funds research, and provides its agencies with a mandate to support research through direct funding to research institutions, it is incumbent upon government, federal agencies, and the recipients of that funding to ensure that research communication is carried out in such a way that it optimally supports the exploitation and the generation of outcomes from that research.

To achieve this it is necessary to purchase services that support effective communication. These services have traditionally been provided by scholarly publishers and it is right and proper that they continue to receive a fair price for those services. The productive discussion is therefore how to develop the markets in these services that means service providers are viable and sustainable, and that there is sufficient competition to prevent price inflation and encourage innovation. That such services can be economically provided through a direct publication service model where the full costs of review and publication are charged at the point of publication has been demonstrated by the success of PLoS and BioMedCentral.

However this is just a starting point. A fully functional market will encourage the development of a wide range of competitive services that will enable researchers to select the most cost effective way of communicating and disseminating their research and ensuring that it reaches the widest possible audience and in turn is exploited fully. This in turn will enable federal agencies to support research, and its communication, in a way that ensures that the public investment is exploited fully for the benefit of the U.S., its citizens, and its economy.

## [\[edit\]](#) References

---

1. ↑ Michael Eisen. "[Research Bought, Then Paid For](#)", *New York Times*, 10 January 2012.
2. ↑ <sup>a</sup> <sup>b</sup> Björk et al. (2010). "Open Access to the Scientific Journal Literature: Situation 2009". *PLoS ONE* 5 (6): e11273. doi:[10.1371/journal.pone.0011273](https://doi.org/10.1371/journal.pone.0011273). PMID [20585653](https://pubmed.ncbi.nlm.nih.gov/20585653/).

## [\[edit\]](#) See also

---

## [\[edit\]](#) **Publicly available responses by third parties**

*In this section, responses to the RFI that are publicly available and have been consulted for the drafting of our response are listed, along with some key quotes that provide an idea about the approach taken in that response. The listing of these quotes here does not imply endorsement on the part of the Wikimedia Foundation.*

### [\[edit\]](#) **American Library Association and the Association of College and Research Libraries**

#### [\[edit\]](#) **Björn Brembs**

1. "I would dispute any actual intellectual property rights of publishers over publicly funded research. I am aware that legally there are such rights as researchers most often hand over their copyrights to these publishers. However, since commercial publishers do not usually add any value to the published research results (even peer-review is performed pro-bono by researchers themselves), this practice needs to end. Publicly funded research has been bought by the public and belongs to the public. Ensuring publication licenses such as CC-BY for literature about publicly funded research is one of the possibilities to ensure the public retains its intellectual property on the research it funded."
2. "The benefit is maximized by minimizing the costs associated with access. The costs are minimized by preventing third parties from adding costs to the process. One way to establish a short and thus cheap supply line is to have scholars deposit their work directly at their libraries, avoiding the costs of intermediaries such as publishers. The process of this deposition would still be identical to the current process (i.e., peer-review), albeit without intervening entities which withdraw funds but add little to no value."

#### [\[edit\]](#) **Cameron Neylon**

#### [\[edit\]](#) **Ecological Society of America**

1. "It is not appropriate for the federal government to expropriate the additional value publishers add to research results." / "Government mandates for publishers to make their work available online without compensation will endanger the U.S. scholarly publishing system"
2. "Subscription revenue from [ESA's] journals is crucial to ESA's publishing program." / "any potential embargo period will do little to mitigate the financial losses that would result from full open access."
3. "The Society also permits liberal use of ESA publications for educational purposes."
4. "require federally-funded grantees to provide a second version of the research summaries they already prepare, specifically for the lay reader"

#### [\[edit\]](#) **Geoffrey Irving**

1. "The single and most important step to take is to explicitly grant the authors of federally funded scientific papers the right to make copies of their work freely available"
2. "The embargo period should be negative: researchers should be allowed to freely distribute their work whenever they choose, including before publication or even before completion. " / "Any other embargo period, even zero, would cause an unnecessary drag on the speed of dissemination and advancement of scientific knowledge."

[\[edit\]](#) [Harvard University](#)

1. "Yes, there are steps to grow new and existing markets arising from access to cutting-edge research. The most important step is to require public access (also called open access and free online access) to the final versions of the authors' manuscripts of peer-reviewed articles arising from publicly funded research."

[\[edit\]](#) [Heather Etchevers](#)

[\[edit\]](#) [John Wilbanks](#)

[\[edit\]](#) [Kitware](#)

[\[edit\]](#) [National Digital Stewardship Alliance](#)

[\[edit\]](#) [Stevan Harnad](#)

The minimum should be to mandate that:

1. the fundee's revised, accepted refereed final draft
2. of all refereed journal articles (including refereed conference articles) resulting from the funded research must be
3. deposited immediately upon acceptance for publication
4. in the fundee's institutional repository.
5. Access to the deposit must be made gratis OA (online access free for all) immediately (no OA embargo) wherever possible (over 60 % of journals already endorse immediate gratis OA self-archiving).

[\[edit\]](#) [University of Oregon](#)

[\[edit\]](#) **Similar public consultations**

- [Research:Committee/Areas of interest/Open-access policy/Request for Information on Public Access to Digital Data Resulting From Federally Funded Scientific Research](#) (same deadline of Jan 12, 2012)

- [US National Science Board seeks comments on a report on Data Policies, Digital Research Data Sharing and Management](#) (deadline: Jan 18, 2012)
- [Research:Committee/Areas of interest/Open-access policy/EU Consultation on scientific information in the digital age](#) (2010)
- [OSTP's Public Access Policy Forum](#) (2009)