Open access to publications

Open access (OA) to scientific publications has become more common since the early 2000s. Today, the majority of publications annually produced by Finnish researchers are openly available either on the publisher's service or as a parallel published copy in an open access repository. Some publications are openly available prior to publishing as a preprint version, which has not yet gone through the peer review process.

The transition to open access has been gradual. Initially, the change was expected to happen almost by itself through recommendations to researchers and publishers, but since then it has had to be accelerated by a variety of mandates from research organisations, research funders and the government. In recent years, the Plan S initiative by European research funders, which aims to promote a rapid transition to the immediate open access to scientific publications, has attracted the most attention.

There is still some disagreement on the definition of open access. The definition used can have a significant impact on which publications can be interpreted as open. Several categories have been developed to describe the different types of open access (gold OA, green OA, hybrid OA, etc.), and their definitions are not entirely consistent or unambiguous, either.

Open access repositories and parallel publishing

The parallel publishing of scientific articles in open access repositories has been a key objective of the international Open Access movement since the launch of the Budapest Open Access Initiative in 2002. Open access repositories can be either organisation specific or discipline specific. Today, almost all universities and many other organisations have their own open access repositories. The most prominent of the field-specific open access repositories is ArXiv, originally created in the 1990s as the preprint archive for high-energy physics.

Features of the open access repositories include support for publication metadata harvesting to other services using the Open Archives Initiative Protocol for Metadata Harvesting. In this respect, the open access repositories differ from commercial networking services targeted at researchers, such as ResearchGate or Academia.edu, which do not share the metadata on the collected dataset with other services.

Originally, in the early 2000s, it was thought that researchers offering their articles to journals would agree with the publisher on the right to parallel publish them. In practice, this was not realistic, which is why parallel publishing has been carried out under the conditions set by the publishers.

The majority of scientific journals allow their articles to be parallel published. However, the terms of publishing vary. Many publishers will only accept parallel publishing of the author's own final version and will not allow the final publisher's version to be made available for open distribution. Some journals also require a certain embargo period (e.g. 6, 12 or 24 months) before opening a parallel published article to the public.

Some universities (e.g. Harvard) have taken the decision to withhold parallel publishing of their researchers' outputs, and research funders have also taken similar decisions regarding the research they fund (e.g. Plan S Rights Retention Strategy). In some countries, the researcher's right to parallel publishing has been set out in legislation.

Open scientific journals and books

Open scientific journals have been published since the 1990s. Some open-access journals have been new, but more and more of the longer-published traditional journals have gradually shifted to open publishing. Open publishing is not cost-free either, so it has taken time to find new funding models.

At international level, scientific publishing activities are largely under the control of large commercial publishers (Elsevier, Springer Nature, Wiley and Taylor & Francis). These publishers have been very profitable, with many of them generating the profits of more than 30% of their annual turnover. This has been based on the fact that publishers own the majority of the traditional and respected publication channels, and publishing on them is important for researchers to achieve academic merit.

In turn, libraries and library consortia have obtained this data for their own organisations by negotiating 'Big Deals' with publishers and other intermediaries of data. In Finland, for example, the total price of the data agreements negotiated by FinELib has been in the range of €30 million per year.

For a long time, major publishers were reluctant to open up access to their publications, which is why open publishing was adopted relatively slowly. However, the situation has changed since the 2010s. Alongside the traditional subscription-based business model, a new business model based on the Article Processing Charge (APC) has emerged, which is also highly scalable to widespread use. A good example of this are mega journals using the "streamlined" evaluation process (e.g. PLOS ONE and Scientific Reports), which have published up to tens of thousands of articles a year. Open-access publishers such as MDPI and Frontiers have also grown rapidly in terms of their publication volume and market share. As a separate side effect, the trend towards APCs has also led to the proliferation of predatory journals: these journals charge a publication fee but do not carry out proper peer review processes or editorial work.

The average rate of APCs for open-access journals has been in the range of €1,500–2,000, but there are significant differences between journals. For scientific monographs, the Book Processing Charge (BPC) may even be significantly higher. The charges have been paid from the budgets of both research projects and libraries, and sometimes from researchers' personal funds. To alleviate this situation, many research organisations have set up centralised funds to pay these charges.

In addition to the publication channels offered by the new open-access publishers, APCs have also become increasingly widespread in journals published by traditional publishers. Many of these journals have become hybrid OA journals, with both paid and open content. This duality has led to concerns that publishers are double dipping, i.e. charging twice for both the subscription and the APCs. APCs for publications on hybrid channels have also typically been more expensive than on fully open-access publication channels.

To address this situation, since the late 2010s, libraries have begun negotiating transformative Read-and-Publish agreements with publishers, agreeing at the same time on the access to the publisher's subscription material and the open access to publications produced by the library's own organisation. As the name suggests, these agreements are intended as a solution for the transition period to speed up the transition to full open access. At least in the short term, the agreements have led to a rapid increase in the proportion the previously shunned hybrid OA makes up of the publications produced by Finnish researchers, for example.

Although the majority of open-access articles appear in paid journals, there are also many channels where publication is available free of charge. These kinds of Fair OA publication channels often rely heavily on voluntary work, but new funding models have been developed to support them, which may be based on, for example, support from research funders or voluntary contributions from research organisations (e.g. Open Library of Humanities).

Other types of funding models have also been developed for monographs. For example, Knowledge Unlatched – a commercial entity that operates for profit – has adopted a model where libraries collectively purchase individual books for open use.

Immediate or indirect open access?

Definitions of open access have generally required that the publications appearing in open-access journals are openly available immediately and without restriction from the time of their publication. Many journals open their content for open use retrospectively (e.g. after a year), but this is not counted as actual open-access publishing.

For parallel published items, embargoes have previously been considered acceptable solutions, although the maximum acceptable length of embargoes has been debated in different fields of science. Recently, however, new policies such as Plan S have started to require immediate open access to parallel published items as well.

Publication rights and further use

Open access is closely linked to the issue of the related rights. Especially in the case of international article publications, it used to be a common practice to transfer all publication rights without compensation to the journal's publisher. This practice was justified by the fact that researchers were rewarded for their publications through the merit it provided. At the same time, however, this practice allowed publishers to control the availability of publications as they pleased.

The rights to open-access publications are often defined by Creative Commons licences. Originally launched in 2002, the use of licences in scientific publication has gradually become more widespread. Especially in article publishing, the recommended licence is usually CC BY 4.0, but for monographs, other Creative Commons licences may be recommended. In humanities and social sciences, the use of CC licences has also posed problems.

The purpose of Creative Commons licences is to indicate the right to the further use of the publication in an unambiguous and standardised way. For example, the CC BY licence allows the use (incl. commercial use) and modification of a publication, as long as the original authors of the publication are credited. The licence also allows the publication to be available not only in the publisher's service but also in other locations, e.g. as a parallel published copy in an open access repository.

In some definitions of open publishing, the use of CC licences is considered a criterion of transparency ('Libre OA') as this is considered a guarantee of permanent open access to the publication. Some definitions, on the other hand, consider that the open access to a publication is sufficient in itself ('Gratis OA'), even if its rights to the further use are not defined.

Publication versions

In many cases, there are different versions of the publications available. Especially in natural sciences, it is increasingly common for non-refereed *preprint* versions of article manuscripts to be available, even before they are published. At this stage, the bibliographic information of the publication is not yet known, not necessarily even the final publication channel. During the evaluation and editorial process, the content of the publication may undergo a variety of changes.

The open access to preprint versions is seen as an important objective for the distribution of scientific information, and in some fields, publications are already widely read and used in this form. However, most definitions of open access are based on the idea that the open access to preprint versions is not sufficient in itself, but that a peer-reviewed version of the publication must also be made available for open distribution.

After the peer-review process, the publication may be available both as the *publisher's version* published in a printed journal or on the publisher's online service and as the *author's own final version* (*postprint* or *final draft*), which may differ in appearance from the publisher's version. Publishers and journals have different policies as to which version of the publication can be parallel published, the author's final version or the final publisher's version.

Definitions and categories

The following terms are commonly used to categorise different open access types, largely based on the classifications outlined above:

- **Gold OA:** The publication has been published on a fully open-access publication channel, which may either be free or charge a fee for publication.
- **Green OA:** The publication has been parallel published in an open access repository where it is available either immediately or after an embargo period specified by the publisher.
- **Hybrid OA:** The publication is openly available for a fee on a channel where some publications have open access and some are only available to subscribers.
- **Diamond or platinum OA:** The publication is openly available in a fully open-access publication channel that does not charge a fee for publishing and finances its activities in some other way.

- **Bronze OA:** The publication is currently openly available in the publisher's service but its rights for further use are not defined by a Creative Commons licence.
- **Black OA:** The publication is available illegally without the publisher's permit, e.g. via the Sci-Hub service. Black OA is not usually considered actual open access.

The above categories overlap to some extent, and there may also be conflicting views on their definitions. In the case of Gold OA, it is sometimes assumed that article or book processing charges have been paid for the publications in the category. However, this is not a generally accepted perception – publications appearing through free channels are usually also included. Similarly, issues related to CC licences share views. For example, the Unpaywall service counts open-access articles published without a CC licence as a separate Bronze-OA category.

Open access to publications in the Ministry of Education and Culture's publication data collection

In its publication data collection, the Ministry of Education and Culture has been collecting data on the open access to publications in the VIRTA Publication Information Service since 2011. Initially, the data was not very reliable, so from 2016 onwards the data collection adopted a more precise definition for open access and a revised three-part categorisation. The open-access publications could appear either 1) in full or 2) in part in the open-access publication channel, and they could also be 3) parallel published. As many of the parallel published publications are also openly available on the publisher's service, these categories largely overlapped.

This classification did not fully meet the requirements of the ever-changing OA policies either. Since 2020, data collection has started to shift towards an attribute-based classification, which is not tied to any predefined categorisation. The attributes have been selected in such a way that they can be used to generate the current and future OA categories as comprehensively as possible. For this purpose, the collection of data into VIRTA started, concerning data such as publication licences and publication fees. During the transitional period, the data will be available in both the old and new format.

Data sources and search engines

- Sherpa/Romeo: A service that provides information on the parallel publishing policies of scientific journals.
- Journal Checker Tool: A service that can be used to find out if a scientific journal is compatible with the requirements of Plan S.
- Directory of Open Access Journals (DOAJ): An international directory of open access journals in which only scientific journals meeting certain criteria are eligible. In addition to journal-level data, DOAJ also harvests the metadata on articles published in journals.
- Directory of Open Access Books (DOAB): A sister service to DOAJ that focuses on book publications.
- OpenDOAR: The most comprehensive international directory of open access repositories.
- OpenAPC: An initiative that compiles information on APC payments made by research organisations to publishers.
- ESAC Initiative: An initiative that compiles information such as transformative agreements concluded with publishers.
- Unpaywall: A database that compiles information on the open access to publications identified by DOIs, both in the publishers' services and in the open access repositories. The data is available through a browser extension and integrated into many other services through APIs.
- OpenAIRE: An EU-funded cooperation project that has developed a search engine that aggregates the content of European open access repositories and research information systems. Also maintains the Zenodo open access repository, which is used globally.
- Bielefeld Academic Search Engine (BASE): A search engine maintained by Bielefeld University that harvests the metadata on publications from open access repositories and many other sources as well.
- CORE: A similar British search engine that also harvests and indexes full-text materials archived in open access repositories.
- Research.fi: A service developed by CSC that includes, for example, the publication data of Finnish research organisations collected in the Ministry of Education and Culture's publication data collection, including data on the open access status of publications.
- Vipunen.fi: A reporting portal of the Finnish National Agency for Education that also includes statistical and monitoring data on open-access publication volumes.
- Finna.fi: A search service that provides a common search feature for material on both the Journal.fi service and most national open access repositories.