

The Costs of Open and Closed Access : Using the Finnish Research Output as an Example

Jyrki Ilva^{1,2}, Markku Laitinen¹ and Jarmo Saarti²

¹ National Library, Finland. E-mail: firstname.lastname@helsinki.fi ² University of Eastern Finland Library, Finland. E-mail: jarmo.saarti@uef.fi

Abstract: The Open Access movement in scientific publishing has been gathering momentum in the European Union and its member states, partly due to the policies of some of its main research funders. Already we have seen encouraging research results on the effects of openness on the dissemination of scientific outputs. As business models of Open Access publishing are still under development, the aim of our paper is to compare the costs associated with different modes of disseminating scientific publications based on the current data available. We will also analyse the potential costs associated with Open Access publishing models and compare them with the current cost structure of – mostly – paywalled access. The discussion will include a description of current Finnish Open Access policies and their funding models. The financial analysis will be based on the statistical data found in the national Research Library Statistics database and the Finnish National Research Publications database, Juuli. We will assess the alternatives on how best to develop statistical tools to estimate the true costs of scientific publishing.

Keywords: document dissemination, open access, closed access, monetary value, Finland

1. Open Access as a paradigmatic change in scientific document dissemination

The promise of free and efficient information transfer is a cornerstone of the ideology underpinning the information society. In 1994, the mathematician, Andrew Ozlydzko, predicted a brave new world of scientific e-publishing that would be ‘dramatically cheaper’ than the traditional paper journal based model. After almost twenty years of the publishing of Odlyzko's paper, the present seems to be rather different from the future he predicted. For example, there has

been no dramatic decrease in the purchasing costs of e-journals. In fact, scientific publishing is considered to be one of the most profitable and successful business practices. The profit margins of scientific publishers have been estimated to be between 20 and 30 cents in the dollar (van Noorden 2013, 427; Monbiot 2011).

One could even argue that libraries are an integral part of the present publishing model, which is based on outsourcing the dissemination of scientific results - and the organization of peer review - to publishing companies. This model has been shown to be extremely challenging in economically turbulent times because it is based on the presumption that there will be continuous growth of the subscription costs of scientific journals. (Odlyzko 2013.)

What we have actually witnessed since the 1990s is an ever-increasing prevalence of paywalls, i.e., technical mechanisms separating the digital content for which one has to pay from the open Internet content. The majority of new scientific information lies behind these paywalls. While it is true that the traditional paper based publishing model has experienced a major transformation, many aspects of the old model still exist, only transformed into a new form. There has been an evolution but not the kind of complete digital revolution that many commentators of scientific publishing had prophesied. This development has been described as a digital divide: although the digital technology enables free dissemination of scientific information, there are many other economic, social and political barriers stopping this development (Ragnedda and Muschert, 2013).

The Open Access (OA) movement has definitely helped the scientific community in disseminating journals to a broader audience and at the same time it has led to a discussion about profits and costs involved in scientific publishing. Van Noorden (2013, 427; see also Laakso & Björk 2012 and Schimmer, Geschuhn & Vogler 2015) has argued that the OA publishing seems to be more cost effective than traditional publishing. Based on his analysis, the average cost of publishing an article in the traditional non-OA journals is \$3500 – 4000; to which one needs to add on the 20 to 30 % profit expected by the publishers so the total cost is about \$5000. The costs for publishing one article in BioMed Central or PLOS ONE are between \$1350 – \$2250. In addition, there is some evidence that the average impact of OA articles is greater than the impact of those published behind a pay wall (Antelman 2004, 379).

The turbulence within the scientific publishing is evident also in the different forms of OA dissemination; these have been characterized as Green and Gold OA. In the Green OA model, several research organizations have built institutional repositories into which researchers can self-archive their publications, even if they have been published in paywalled journals. There are also a number of subject-based repositories which have the same role. Green OA is costless to the authors, but Gold OA usually - but not always - incurs

costs for the authors. The publishing fees are called Article Processing Charges (APC). In addition, the traditional publishers have started to provide Open Access to articles published in their paywalled journals (Hybrid OA): by paying a set fee, the author(s) can provide free access to their papers.

In addition to many legitimate OA publication channels, the academic community has also seen the rise of so called “predatory OA publishers“. These companies are more interested in obtaining financial resources from academics rather than making sure that the quality of their publications meets the standards of the scientific community, either ethically or scientifically.

The provision of OA to the results of publicly funded research has been placed onto the political agenda by the research funders. For example, the European Union has decided to encourage OA publishing (see European Commission 2013). The Finnish government has lately stressed the importance of the openness of the science in its strategies and policies (see more <http://openscience.fi/>), although the actual policies and infrastructure are to some extent still under debate.

The emerging open science requires services and infrastructures, which need to be funded. Thus the aim of this paper is to investigate the different business models of scientific publishing, what effects these have on the libraries and to analyze the present library statistics: how do they manage their collections and manage statistical data about open science.

2. Open Access as a challenge for the library statistics collection

The role of the research library is likely to change substantially as Open Access becomes more ubiquitous. Instead of acquiring materials produced elsewhere for the consumption of local patrons, the library will have a larger role, in making sure that locally produced scholarly products are disseminated effectively world-wide via Open Access publication channels. This will also represent a challenge for the collection of relevant library statistics.

Many research organizations utilize a Current Research Information System (CRIS) for the collection of publication data (see De Castro, Shearer & Summann 2014 and Ilva 2014). It is very reasonable to adopt this system also for the collection of data on the Open Access status of publications. There are two general approaches to the collection of this data - either it is reported by the researcher him/herself, or the data is gathered and generated from other sources, including the local Open Access repository and lists of OA publication channels such as the Directory of Open Access Journals (DOAJ, <https://doaj.org>). Both of these approaches have their limitations, and the collection and verification of the data is not always straightforward. There may well be also other complicating factors e.g. publisher-defined publication embargos.

In many European countries, the progress on the Open Access availability of research publications is monitored at a national level and there are also a number of international services (including the EU-funded OpenAire portal) harvesting information from the local and national systems. For example, the Nordic countries, Norway, Denmark, Sweden and Finland, have their own national data collection systems, although there are significant differences between the approaches these countries have adopted (for an overview of the Nordic situation see Ilva 2014). In Norway, both the collection of publication data and the upload of self-archived publications are being integrated into the national CRISin system. In addition to local university-level repositories, the metadata of Open Access publications is available for searching, browsing and analysis in a separate NORA interface (<http://nora.openaccess.no/>), which contains a subset of the CRISin data. Denmark has its own national research publications portal (<http://forskningsdatabasen.dk/>), the content of which is harvested from the local CRIS of each Danish university. A national project, Open Access Barometer (see Price 2014), has been working on the quality of the data. In Sweden, the publication data is harvested from local repositories into a national portal, SwePub (<http://swepub.kb.se>), maintained by the National Library of Sweden.

In Finland, the Ministry of Education and Culture collects the publication data from universities as part of the annual data collection. The collection of accurate data on research publications is of paramount importance, as the number and quality of these publications is one of the main criteria used in the current funding model for the Finnish universities. At the moment, more than 200 million euros a year (13 % of total state funding to the universities) is distributed on the basis of this data. As there are plans to make the Open Access availability of research publications an additional criterion in the funding model starting from 2019, the collection of data on the prevalence of Open Access will become even more important.

The collection of publication data is handled by CSC - IT Center for Science, and the data is used both in the Vipunen statistics portal (<http://vipunen.fi>) and in the Juuli Research Publications portal (<http://www.juuli.fi>, maintained by the National Library of Finland). Unfortunately, the quality of data on Open Access is currently relatively poor, partly due to motivational reasons at the local level, partly because of problematic categories and instructions in the data collection itself. There are plans to clarify the categories and improve the methods of data collection in the near future.

As far as the Article Processing Charges of Gold OA and Hybrid OA journals are concerned, currently much of the money used for this purpose comes directly out of research funding, and in many cases, the library is not even aware of the flow of money associated with OA publications. This is by no means an optimal situation, as the libraries are paying a significant amount of money for site licenses of digital content, and there is a real danger that the publishers are

charging both licensing fees and Open Access APCs for the same content (this is called "double dipping"). (See Björk & Solomon 2014a and Björk & Solomon 2014b).

From this point of view, it would make sense to create university-level OA funds, which would collect and administer all of the money used for OA costs, including both APCs, membership fees and also voluntary subsidies collected by some of the OA publishers using alternative business models. A centralized fund would make it much easier to monitor both the prevalence of Gold OA and the associated flow of money. There would be potential savings in the transaction costs, and more important, this would make it possible to combine the data on both licensing costs and OA charges at an organizational and (with some extra effort) also at a national level. In some European countries including the United Kingdom, Norway and Sweden (see for example Eriksson 2013), university-level OA funds are already fairly common; in Finland they are currently still in early planning stages.

The development of OA business and funding models is currently undergoing a rapid spurt. While this is a major transition envisaged to exert a significant potential impact on research library budgets and workflows, the collection of library statistics does not currently provide fully adequate means for taking into account this change.

The libraries are advised to compile their annual statistics following the International Standard ISO 2789 – International Library Statistics (ISO 2789:2013(E)).

According to this protocol, electronic journals in free Internet resources (ISO 2789:2013(E):2.3.22-23) which have been catalogued by the library in its online catalogue or a database should be counted and reported separately (ISO 2789:2013(E):6.3.14.2) by counting the number of links to individual free Internet resources (electronic journals, etc.) which have been catalogued by the library in its online catalogue or a database (ISO 2789:2013(E):6.3.15).

In addition, with respect to the data collection part of the standard, there is a reference to counting the costs of institutional or single author fees for open access publishing paid by the library (ISO 2789:2013(E):6.6.1.7).

The annual statistics of Finnish libraries of higher education (HE) are compiled in the Finnish Research Library Statistics Database. The statistics are mainly collected according to the standard ISO 2789, but there are no statistics covering the use of the OA publications, nor information about their costs. Instead, the usage statistics and economy of electronic journals cover the use of both OA and PW together, with no possibility to separate them from each other.

3. Comparing costs associated with open and paywalled access in Finnish context

The aim of our study was to identify the costs associated with the OA and paywalled publications produced in Finland. The data for this study was compiled from the Finnish National Research Publications database Juuli (<http://www.juuli.fi>), the National Higher Education Statistics Portal Vipunen (<http://vipunen.fi>) and the Finnish National Research Library Database Kitt (<https://yhteistilasto.lib.helsinki.fi>). As a comparison price for OA, we used the average amount of one OA publication fee €816, this estimation was based on the work of Laakso and Björk (2012).

The study concentrated on peer-reviewed publications in journals, conference proceedings and books, using the 2011-2013 publication data collected from the Finnish universities and universities of applied sciences. The publications were categorized by their publishing channel, OA indicating open access and PW paywalled, i.e. publications in non-OA publication channels. The quality of data had some issues, as some of the organizations had reported the OA status of a large number of their publications as “unknown”. However, we assumed that all of the publications with this status were non-OA publications, although this may not be completely correct in some of these cases.

Nonetheless, conducting a comparison between the OA costs of locally produced research outputs and the licensing costs of globally produced scientific content is obviously fraught with problems. However, we have assumed that at some point in the future, Open Access publishing will become the global norm that will be adopted in all countries so that it will no longer be necessary to pay for new licensed content.

In addition, to simplify our calculations we have assumed that the volume of Finnish research output and the amount of money spent on current licensing deals are both at the average international level, which is actually not quite the case. In reality, those institutions and countries producing a higher-than-average number of publications would also pay a larger amount of money for the APCs. On the other hand, as many of the scientific publications are produced in co-operation with researchers from other organizations, it is worth noting that in these cases, the APCs are likely to be paid only by the organization of the corresponding author. There is some discussion on the effects of these issues later in this paper.

Fig. 1 illustrates both the numbers of peer-reviewed publications produced at Finnish institutions of higher education and the costs associated with the purchase of paywalled scholarly materials for these organizations. The first columns in the time series indicate the number of OA publications and the second column paywalled publications (PW). Based on this data, the share of OA publications can be estimated to be between 15-20 % of the total number of peer-reviewed research publications.

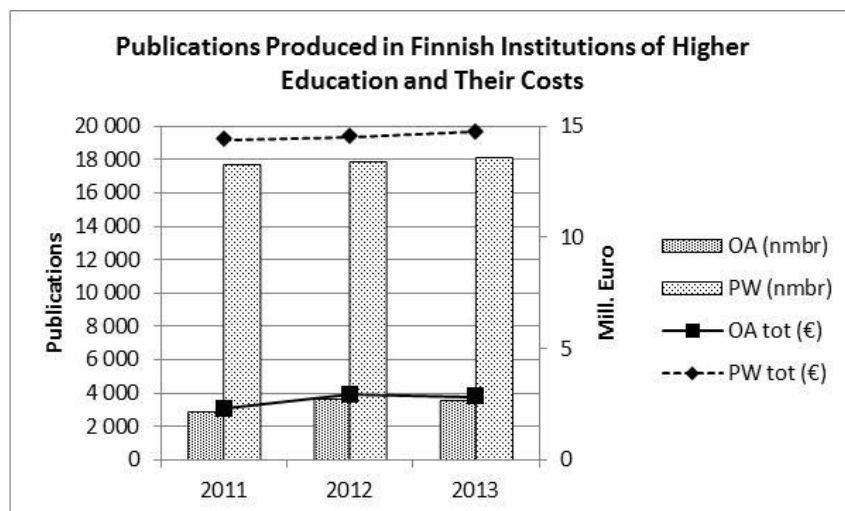


Fig. 1. The numbers of peer-reviewed OA and PW publications at Finnish HEIs and the hypothetical APC costs of the current OA and PW publications based on the international average price of OA author’s fee.

In addition, we counted the hypothetical total APC cost for the current OA and PW publications in the ideal case that they had all been published in OA publication channels. Fig. 2 shows comparison where the first column in each time series represents the costs currently paid by the HE libraries for the access to the paywalled journals. The column also includes the estimated amount of the OA costs (in lighter grey), based on the average cost of OA APCs in Gold OA journals.

The second column in each of the time series is an estimate that calculates the costs of all Finnish publications if they were to be published totally in OA. Again we have used the average price and disregarded some of the complicating factors which will be discussed below.

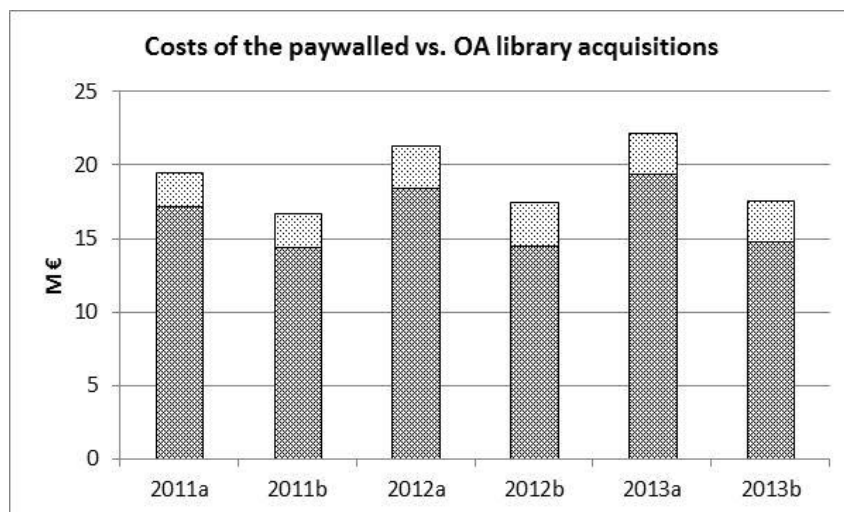


Fig. 2. Comparing the current costs associated with scientific publications (a columns) vs. the hypothetical costs associated with complete open access publishing (b columns).

As can be seen from Fig 2, the hypothetical cost of the all-OA model appears to be somewhat lower than the cost of our current model with its mostly licensed/paywalled content. However, there are two other complicating factors which mean that the actual cost savings associated with OA publishing would be somewhat higher, especially on a global level.

First, only the home organization of the contributing author of each publication will pay the APCs for publications which have co-authors from many organizations. As there are articles with tens, hundreds, or in some cases even thousands of authors, this will mean that each organization will pay APCs for only some of the articles produced by its scholars or researchers. According to Schimmer, Geschuhn & Vogler 2015, the quantity of APC-relevant publications for an institution generally lies between 40-60% of its total research output, depending on a number of factors including the research intensity of the organization. Although the difference is not quite as large at the national level, it is still significant especially for a small country like Finland, as international co-publications are very common in many fields.

Secondly, both the publication volume and the amount of money spent on licensing deals differ from country to country. Finland is currently producing a fairly high number of research publications, while compared to the examples cited by Schimmer, Geschuhn & Vogler 2015, our per capita licensing costs seem likely be rather close to the European average. This means that the potential savings for Finland may be somewhat lower than for countries with either a smaller volume of publications or higher than average licensing costs.

Of course, it should be noted that in an all-OA model, all countries and all institutions would have access to all scientific publications, which would expand the benefits of the transition far beyond the level of whatever cost savings were achieved.

One of the critical questions for the funding of Open Access in the future is whether the Article Processing Charges will stay at their current level. The leading journals with established brand names and high Impact Factors may charge higher than average sums, and there are clear incentives for researchers to try to publish in these journals. In addition, the hybrid journals tend to charge higher APCs than Gold OA journals (Björk & Solomon 2014b). On the other hand, there are also new emerging OA business models which may lead to diminished costs. These include both membership-based models like that used by PeerJ (<https://peerj.com/>) and models which are based on library subsidies, including the one adopted by the recently founded Open Library of Humanities (<https://www.openlibhums.org/>). To make sure that the publication funds are used optimally, it is very important that the libraries are able to collect information on the flow of money from the research organizations to publishers.

Another key issue for libraries, as far as licensing deals are concerned, is legacy content - the non-OA back issues of scholarly journals, which in some cases go back more than a hundred years. While they may no longer be critical for cutting-edge research in some research areas, in fields like the humanities and social sciences, they may well be very relevant. If the current licensing deals for new publications are cancelled at some point when Open Access becomes ubiquitous, will there be some other way to provide access to this content?

4. Conclusions: Open Access and its challenges for the library statistics and funding models

The on-going movement from paywalled publishing to Open Access is a major transition, which will have an impact on the role that research libraries will play within their organizations. There is also a huge demand for reliable statistical information both on the prevalence of OA and on the flow on money associated with different OA business models.

From the point of view of the research libraries, the rise of OA publishing holds a promise of significant cost savings, although it is still unclear how soon (if ever) we will reach a point at which it becomes feasible to start cancelling the major licensing deals we currently need to sign in order to obtain access to paywalled scholarly content. If the libraries wish to speed up this development and to make sure that they have a key role in the future, it is essential that they take an active role in monitoring the OA costs, making sure that the costs are administered and negotiated in an optimal way.

Although much of the statistical information on OA will be collected at an international, national or organizational level, it is also important that the collection of library statistics will be able to deal with these new demands adequately. The statistics must provide the library leadership with reliable information and functioning tools that help it to navigate the stormy waters they are experiencing today.

At the present, the number of electronic journal titles which a library has licensed to provide in its list of electronic periodicals does not differentiate their status as either OA or PW publication channels. The same applies to economy statistics and usage statistics: it is not possible to identify either of the categories from the statistics since they are often both included under the same category.

However, as shown in Fig 2 (and elaborated in the following discussion), we have to some extent confirmed our hypothesis that OA publishing would be more affordable than the current license-based model – there only are no reliable and exact statistics to explicitly prove this hypothesis. Nonetheless, in the present economic climate where libraries and their funding organizations need to be able to show that they are operating cost-effectively, it is clearly important to be able to organize data collection to support this target.

The International Standard ISO 2789 gives a clear framework for compiling the statistics of OA vs. PW. From this point of view, it would be feasible to include this data in the library statistics, especially if the library starts to administer the payments of APCs, membership fees and subsidies to the scholarly OA publishers.

The final success of Open Access depends on finding solutions for both the longtime preservation and knowledge organization of documents and data. These are two major challenges that need to be solved by the digital academic libraries and archives.

Acknowledgements

The authors are grateful to Dr. Ewen MacDonald for linguistic advice.

References

- Antelman, K. (2004). Do Open-Access articles have a greater research impact? *College & Research Libraries* 65(5):372-382.
- Björk, B-C & Solomon, D (2014a). Developing an Effective Market for Open Access Article Processing Charges. *Report for Wellcome Trust, March 2014*. <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Open-access/Guides/WTP054773.htm>. [Accessed 30 April 2014]
- Björk, B-C & Solomon, D (2014b). How research funders can finance APCs in full OA and hybrid journals. *Learned Publishing*, 27: 93-103. <http://dx.doi.org/10.1087/20140203>. [Accessed April 30 2015]
- De Castro, P., Shearer, K. & Summann, F. (2014). The gradual merging of repository and CRIS solutions to meet institutional research information management re-

quirements. *Proceedings of the 12th International Conference on Current Re-search Information Systems*. <http://hdl.handle.net/11366/197>. [Accessed 30 April 2015.]

Eriksson, J. (2013). A university fund for Article Processing Charges - experiences from Lund University. *Presentation at the Nordic Perspectives on Open Access and Open Science seminar, Helsinki, October 15, 2013*. <http://urn.fi/URN:NBN:fi-fe201310166689>. [Accessed 30 April 2015]

European Commission (2013). *Horizon 2020: the EU framework programme for research and innovation: Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*. Brussels: European Commission. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf. [Accessed 14 April 2015.]

Ilva, J. (2014). Integrating CRIS and repository – an overview of the situation in Finland and in three other Nordic countries. *Presentation at Open Repositories 2014*. <http://urn.fi/URN:NBN:fi-fe2014070432242>. [Accessed 30 April 2015]

ISO 2789:2013(E):2013. *Information and documentation - International library statistics*. International standard. 5th ed. 2013. International Organization for Standardization (ISO). 71 p.

Laakso, M. and Björk, B-C. (2012). Anatomy of open access publishing, a study of longitudinal development and internal structure. *BMC Medicine* 10(124). <http://dx.doi.org/10.1186/1741-7015-10-124>. [Accessed 18 November 2014.]

Monbiot, G. (2011). Academic publishers make Murdoch look like a socialist. *Guardian* 29 August 2011. <http://www.theguardian.com/commentisfree/2011/aug/29/academic-publishers-murdoch-socialist>. [Accessed 18 November 2014.]

Odlyzko, A. (2013). *Open access, library and publisher competition and the evolution of general commerce*. <http://arxiv.org/abs/1302.1105>. [Accessed 14 April 2015.]

Price, Adrian (2014). Open Access in Denmark. *ScieCom Info Vol 10, No 2 (2014)*. <http://journals.lub.lu.se/index.php/sciecominfo/article/view/11641/10315>. [Accessed April 30 2014.]

Ragnedda, M. and Muschert, G.W. (Eds.) (2013). *The Digital Divide, The Internet and Social Inequality in International Perspective*. Oxon: Routledge.

Schimmer, R., Geschuhn, K. K., & Vogler, A. (2015). Disrupting the subscription journals' business model for the necessary large-scale transformation to open access. *A Max Planck Digital Library Open Access Policy White Paper, 28 April 2015*. <http://dx.doi.org/doi:10.17617/1.3> [Accessed April 30 2015]

Van Noorden, R. (2013). Open access, the true cost of science publishing. *Nature* 495(7442):426–429. <http://www.nature.com/news/open-access-the-true-cost-of-science-publishing-1.12676>. [Accessed 18 November 2014.]