

PUBLIC RESEARCH BENEFITS AND PUBLISHERS' PROFITS

Prepared by Claudio Aspesi for SPARC

The public research agenda is affected in several ways by the interests of the publishing industry. The issue has been recognized for quite some time: the DORA declaration of 2012 explicitly addresses this¹, and yet many people in the research community still do not understand how serious or pervasive the issue is.

One of the factors contributing to this continuing underestimation of the distortions introduced by publishers is the limited public knowledge of the costs and profitability of scholarly journals. The publishers contribute to this by withholding data that would shed a true understanding of their costs and profitability, as well as by arguing that the aggregate cost of dissemination is a small fraction of total research costs and that the research community is well served by these expenditures.

While the cost of dissemination is indeed a small percentage of federally funded research, the profit margin charged by the publishers on top of their costs is significant, and additionally less expensive alternatives exist. The argument that “the research community is well served by the publishing industry” bears close scrutiny: a system that has to be repeatedly “worked around” when society confronts a crisis is broken. Fighting COVID-19 required establishing emergency mechanisms outside of the current publishing system, and there is no logical argument why cancer or heart disease patients are any less worthy of the benefits of open access and open science. Similarly, addressing other societal crises, like climate change, the loss of biodiversity and their impacts requires similar actions to those that supported the global fight against COVID-19.

Cost of dissemination is a small percentage of total research spending

We estimate that the publishers incur costs in the region of \$2,000-3,000 in order to publish an article. There is a vast literature on this subject, with results ranging from a few hundred dollars to several thousand. Several variables affect these calculations, including the characteristics of a publisher (for example, larger publishers have more articles to amortize the cost of their fixed spending in software and technology; journals with higher rejection rates have more costs that need to be charged to the published articles; different companies have different productivity rates) and the quality of the analysis. We believe the range of \$2,000 to \$3,000 captures the vast majority of articles published, and interviews confirm this.

In 2020, US-based authors published 626,295 articles,² leading us to estimate that publishers spent in the range of \$1.252 to \$1.878 billion dollars to publish these articles. There is no exact data on how many of these articles were funded, at least in part, through federal funding, but we can assume that they are at least equal to the share of federally funded research. The latest data available for the mix of spending is from 2018³, and it indicates that the Federal Government accounts for 41.8% of total basic research spending. Applying these

¹ <https://sfdora.org/>

² Source: <https://www.oecd.org/sti/scoreboard.htm>

³ Source: <https://sgp.fas.org/crs/misc/R44307.pdf>

percentages to the articles published, we can estimate that federally funded research accounted for 261,791 articles, at a total cost of \$523.6 to 785.4 million for the publishers. In addition, non-federal government spending accounts for 2.6%, and universities and other nonprofits (which, it could be argued, are still funded by the US taxpayer through favorable tax benefits) account for an additional 26.6%. Hence, under a more expansive definition, the US taxpayer supported research that may have cost between \$889.3 and \$1,334 million to publish.

How does the \$523.6 to \$785.4 million compare to federal spending on research? Federal government spending for research has increased in recent years: it stood at \$70 billion in 2016, but grew to \$90.3 billion by 2020⁴. Since research often takes years between funding and publication, it appears prudent to use the five-year average federal research spending (\$79.9 billion). Using all these assumptions, publishing costs stand at about .65 to .98 percent of research spending.

The profit margin charged by the publishers on top of their costs is significant

There is no definitive data on the profitability of scholarly publishers because few of them are subject to public markets disclosure requirements, and they often belong to companies with other business units and some of their data is reported at a company level. Additionally, even within their scholarly publishing divisions they often have multiple business lines (textbooks, databases, digital services) with different economic profiles. However, there are financial indicators that can be used to determine a reasonable estimate of these profits.

The most widely used data points are the Adjusted Operating Margins of the STM business of RELX, colloquially referred to as Elsevier, (37.2% in 2020) and the EBITDA of Wiley (35% in 2020). Interviews suggest that operating margins in the region of 30 to 40% are not uncommon for the leading and even for middle-sized scholarly publishers. It should be noticed that these costs are not transparent, and that it is likely they include the allocation of corporate costs that are not necessary to the management of a scholarly journal business (for example, Corporate officers like Group CEO and CFO and their staff, the investor relations staff, etc.).

These operating margins mean that – for an article generating \$5,000 in revenues (either through subscriptions or through OA revenues, such as APCs) - the publisher earns anywhere between \$1,500 and \$2,000. Applying this range to the estimated 261,791 articles funded by the Federal Government, dissemination of federally-funded research generated an estimated profit (before interest and taxes) of \$393 to \$523 million for the publishers.

These figures become even larger when all research funded by the US taxpayer through federal and state taxes (including the tax-advantaged status of nonprofits) is taken into consideration. As noted earlier, taxpayer-supported research funding accounts for 71% of US basic research spending; this equates to an estimated profit of \$657 to \$877 million.

Even the estimated profits on the smaller figure (specifically, only federally funded research) are not inconsequential: they equate as much as 8% of the research budget of the National

⁴ Source: <https://www.aaas.org/programs/r-d-budget-and-policy/historical-trends-federal-rd>

Science Foundation (\$6.7 billion in 2020) or almost 18% of the research budget of the USDA (\$2.9 billion in 2020).

Less expensive alternatives exist

The shift to a “Green” Open Access policy, with the requirement to deposit a manuscript copy in a repository, would save substantial amounts of money. The estimated cost per article has been estimated in surveys to run between \$7 and \$15 per article⁵. These surveys are dated; however, in 2020 arXiv (a physics repository operated by Cornell University) incurred annual costs of \$2.4 million and received an estimated 190,000 articles⁶ (\$12.75/article – and some of the costs relate to the upkeep of submissions the previous year). At \$12.75/article, the estimated 261,791 articles published in 2020 associated with federally funded research would have generated incremental costs of about \$3.3 million.

This approach would lessen the dependency on legacy publishers. Services - including editing and peer-review management, could be provided at reasonable cost by non-profit publishers or independent groups of editors that would likely offer these services at competitive prices, potentially lowering the costs of these services. Other categories of costs (like digital rights management (DRM) systems, billing systems, software and sales and sales accounting staff) could be greatly reduced or eliminated. Conservatively, we estimate that 10% of the operating costs of a scholarly journal would go away altogether.

Under the most conservative scenario, the savings that could be obtained by assigning editorial services to independent operators and those obtained by eliminating some categories of costs would pay for the profits of these new operators. Essentially, the entire profit pool we calculated earlier (\$393 to \$523 million) could be saved and it would become available for the academic system to invest otherwise. The extension of such a policy to all taxpayer-funded or subsidized research would increase the money that could be reallocated to \$657 to \$877 million.

These figures are conservative, as they assume that editors currently employed by publishers would want to earn profits in addition to their current compensation, that all corporate costs allocated to current costs per article also become profits for the independent operators, and that innovative models (like the open peer review process used by innovative platforms like F1000) do not lead to reducing or eliminating the existing peer review administrative costs.

⁵ Source:

<http://openaccess.eprints.org/index.php?/archives/702-Annual-Costs-Per-Deposit-of-Hosting-Refereed-Research-Output-Centrally-Versus-Institutionally.html> and <https://www.dlib.org/dlib/january13/burns/01burns.html>

⁶ https://static.arxiv.org/static/arxiv.marxdn/0.1/about/reports/2020_arXiv_annual_report.pdf