Scientific Publishing: Education as the Key Enabler for the Transition to Open Science

Mario Pagliaro[a]

Abstract: Originally created for facilitating scientific communication, the internet in principle makes scientific journals no longer necessary. Yet, in an almost opposite fashion to what happened to newspaper publishing, the $25 billion annual income scholarly publishing industry has further flourished following the advent of the internet. Expanding the education of today’s students and young researchers to include modern scholarly communication is the key requisite for the transition to open science.

Keywords: scientific publishing; scientific journals; scholarly publishing; open science

1. Introduction

Published since the late 1600s (the world’s oldest scientific journal appeared in France, on January 5, 1665 as a twelve page pamphlet called the Journal des scéavans [1] followed by the Philosophical Transactions of the Royal Society of London published without interruption since March 1665 [2]), today’s scholarly journals are the products of a large industry, mostly based in western Europe and North America, comprised of for-profit and not-for-profit organizations (not-for-profit publishers do make profits like any for-profit organizations, the only difference lies in the way these profits are used).

Selected figures show the relevance of this industry whose global annual turnover exceeds $25 billion ($25.2 billion in 2015) [3]. The annual revenues generated in 2017 only from English-language scientific, technical and medical (STM) journal publishing were about $10 billion [4].

In the same year, the number of peer reviewed English-language journals was about 10,000, out of 33,100 academic journals. Out of 110,000 people employed, about 40% of them were authors. The annual growth rate in the number of articles published increased to 4% per year due to the rising number of publishing researchers [4], mostly originating from China and India.

In 2017 the aforementioned journals published over 3 million articles. The annual growth rate in the number of articles published increased to 4% per year due to the rising number of publishing researchers [4], mostly originating from China and India.

In one of the first studies to include a discussion on the economics of scholarly publishing, Larivière and co-workers found in 2015 that the industry is actually an oligopoly in which the five major publishers in natural and medical sciences in 2013 accounted for 53% of all papers published [5].

In 1973, the share was slightly more than 20% [5].

Nicely explaining the unique nature of the scientific publishing market in which consumers (scholars) are isolated from the purchase because purchase and use are not directly linked (and thus price fluctuations do not influence demand), the latter study was published in a so called “open-access” (OA) journal published online by a not-for-profit STM publisher charging authors a fee (Article processing charge, APC) billed upon acceptance of the article [6].

The revenues of STM journal publishing, indeed, chiefly originate from subscriptions paid by universities and research institutions to access “pay-walled” research articles, as well as from the industry, especially pharmaceutical and chemical industry but also engineering, which usually pay much higher license fees than academia.

Developed since the mid 1990s, shortly after the introduction of the World Wide Web in 1991, the main alternative economic model for scholarly publishing is based on OA journals in which authors either publish their articles for free in journals supported by external funders (74% of OA journals listed in the Directory of Open Access Journals charge no APC [7]), or are charged with an APC which can vary from $750 for publishing in ACS Omega [8] through $3500 for publishing in PLOS Biology [6] or $5,000 in Advanced Science [9].

In principle, publishing scientific papers on the Web eliminates the need for printing journals and disseminating them via postal mail to subscribers across the world.

Today, most journals exist both in print and electronic. Articles are produced and published on the Web in different formats including hypertexts in HTML (hypertext markup language), PDF (portable document format) and ePUB (open e-book standard format). The print copies are printed according to the number of print subscriptions.

Notwithstanding the above, most journals licensed by a library are licensed as electronic copy. Customers and authors wishing to receive a printed copy of a journal’s issue are billed with its cost, and the selected journal’s issue is printed “on demand” is sent via postal mail. Finally, certain publishers earn extra revenues by selling the journal covers (front and back covers) to the authors of selected articles willing to pay the fee requested by the publisher.

“Everything points to the fact” Bartling and Friesike wrote in 2014 introducing one of the most complete (open-access)
books on open science, “that we are on the brink of a new scientific revolution” [10].

In 2007, a group of major science journal publishers had hired a “public relations” (PR) agent “to combat the open access movement” [11] that aims to make scientific articles freely accessible on the internet.

“We’re like any firm under siege” commented the manager of a publishers association organization. “It’s common to hire a PR firm when you’re under siege” [11].

Said “siege” of the open science movement apparently had little effects on publishers if a scholar based in Canada, commenting on a social network in late 2019 his refusal to write for free another book chapter, emphasized how:

«The science publishing industry which charges us to publish papers, and then charges us again for content access (through our universities or personal licenses), and not compensating us as associate editors, or being a reviewer, or to write book chapters as an expert, has to change [12]»

In an almost opposite fashion to what happened to newspapers, the scientific publishing industry not only was not been financially hurt by digitization process followed by the widespread adoption of the internet, but it actually greatly benefited from it. Both production and distribution costs, indeed, dramatically decreased.

Will this be the case also in the course of the next decade?

Expanding the education of today’s students and young researchers to include scholarly communication in the digital era so as to raise awareness, making developments visible and fostering critical thinking is the key enabler for the transition to open science. Understanding the system of scientific publishing, in other words, should be integrated in the curriculum of any PhD student.

2. Citations, Tenure and Open Access

Noting how progress to open access recently stalled, with only 20% of new papers being published as OA articles, Green has lately called for a true digital transformation of scientific publishing [13]. Rather than using information technology to simply digitize the scientific publication processes, said transformation “is about changing the way you work and designing processes using internet-era principles to deliver value” [13].

Guiding OECD Publishing, namely the publishing agency of the Organisation for Economic Cooperation and Development (OECD), Green has pioneered OECD iLibrary, a platform that disseminates OECD work to academic and research institutions across the world.

The main reason for which OA journals did not replace pay-walled journals is likely due to the fact that only few of them reached high impact factor (IF) values. Some actually did but, almost invariably, they are those with the highest article publication charges (one of the exceptions being Chemical Science, a journal “free to read and free to publish with no APCs” with an IF of 9.556 in 2018 [14]).

The impact factor, however, is a worthless criterion to forecast the impact of a specific article. For example, up to 75% of the articles in any given journal has lower citation counts than a journal’s IF [15].

Still, a recent study on the use of the journal impact factor in academic review, promotion, and tenure evaluations at North America universities found that 40% of universities granting the PhD degree explicitly mentioned the impact factor, or closely related citation-metrics, in their review, promotion, and tenure documents [16].

The picture was confirmed by the editor of Chemistry: A European Journal in a plea to authors calling for a better and fairer use of citations:

« Since the Impact Factor, and similar citation-based metrics, are likely to prevail and will continue to be used for evaluating not only journals but also researchers who publish in them, could you as an author please pay more attention to the citations, bearing in mind their implications. [17] »

As long as the academic reward system across the world will continue to evaluate scholars based on the IF of journals in which they publish, young researchers will continue to prefer publishing their work in high IF journals, even though open-access journals have significantly more citations compared to non-OA journals [18].

Studying a sample of 100 OA articles and 100 non-OA articles randomly selected among the 3,742 randomized controlled trials published in the international literature in January 2011, a team of scholars in Southeast Asia found that, whereas the IF shows moderate correlation with citations for articles published in non-OA journals, the IF does not correlate with citations for OA journals [18].

The conclusions of the study were clear: it is better to publish in an OA journal for more citations, and it is not worth paying high APCs for higher IF journals, because gain in terms of increased number of citations will be minimal.

Yet, as noted by Green only 20% of new papers are currently published as OA articles [13]. Professors guiding the work of doctoral students and post-doctoral scholars continue to prefer to submit their team’s manuscripts to high IF journals because both their own promotion and that of their tenure-seeking post-docs is driven by the journal’s impact factor.

This academic closed-loop system explains why most high IF scientific journals and their publishers not only were not impacted by the advent of the internet, but actually launched several new online-only journals which, thanks the reputation of the publisher, reached high IF valued in only a few years after the journal’s launch (four years of consecutive publishing are needed for a journal to receive an impact
factor as the journal IF is defined as the ratio between the number of citations by indexed journals in year 3, to the total number of articles published in year 1 and year 2).

All would remain unvaried for the next decades if it were not for the unexpected recent advent in other scientific disciplines of an alternative way of scientific publishing mostly used by physicists, mathematicians and computer scientists: the preprint.

3. A disruptive innovation?

Cutting time to publication, establishing priority, and eliminating subjective assessments of significance or scope, preprints allow scholars to publish the results of their scholarly work, in a fully citable form, immediately after its completion.

Routinely used and cited by physicists, astronomers, computer scientists and mathematicians since the launch of arXiv preprint server in 1991 [19], preprints are now commonly used also by biologists and life scientists [20]. Eventually, their slow uptake by chemists, forecasted to inevitably accelerate in a 2017 study [21], has lately recorded its first inflection point [22].

As mentioned above, driven by the existing recruitment and career advancement system based on citations and journal impact factor, both tenure-seeking young scholars and tenured professors (“principal investigators” in the research jargon) are chiefly interested in getting citations to their research articles.

Preprints – research articles permanently published online with a digital object identifier (DOI) – are rapidly becoming highly cited scientific documents. For example, a recent regression analysis of preprints published in bioRxiv reveals that bioRxiv preprints are directly cited in journal articles, regardless of whether the preprint has been subsequently published in a journal or not [23]. Furthermore, bioRxiv preprints are also shared online widely, particularly on Twitter and in blogs.

Publishers offer numerous services which are relevant for authors, like semantic annotation and search engine optimization for better retrieval of articles, submitting metadata to many database providers for indexing, and ensuring outreach through many technical systems. Many of these services are offered by the owners of preprint platforms.

The main service offered by scientific journals to authors submitting their work is peer review following the editorial office decision whether to send the manuscript out for review or not. Manuscript editing and formatting services offered by publishers today are far less important. Accustomed to digital technology, indeed, most today’s scholars are able to self produce easy to read versions of their studies using article templates and word processing software freely available online.

Peer review is provided for free by scholars on a voluntary basis following a request from a journal’s editor to review a manuscript. The aim is to provide authors with a critical and constructive review helping them in improving their work prior to publication.

The process, however, lacks transparency because reviews are usually not published and reviewers remain anonymous, leading numerous scholars to propose open peer review in which reviews are published online and even rated by peers [24], or double-blind peer-review in which the reviewers of the paper do not know the identity of the authors, and the authors do not know the identity of the reviewers.

Several journals today publish online the names and affiliations of the reviewers next to the published article (for example, the Frontiers OA journals) or even the reports of the anonymous reviewers (for example when authors opt for open peer review, Proceedings of the Royal Society A publishes reviewer reports, the substantive part of decision letter after review, and the associated author responses).

Similarly, authors of preprints are allowed to post online at any time a revised version of their original preprint incorporating changes suggested by other scholars usually received via e-mail or directly using the Comments form at the bottom of the web page presenting the preprint (an useful tool offered, for example, by Preprints.org).

In principle, it is already possible for a young tenure-seeking candidate to present a list of publications including highly cited preprints in place of peer reviewed articles.

If scientific quality is associated with citations from peers, why should a selection committee or a funding agency make a difference between a citation to a preprint and another to a publication in a peer reviewed journal?

Indeed, the authors of the 2018 report of a scientific publishers association emphasized in the conclusions how:

«There is some concern that preprints (which can be brought up to date) may become a go-to place for the version of record, undermining publisher business models.

«Concerns have also been raised over the loss of citations from journals to preprints servers, with well over 8,000 citations to bioRxiv reported on Web of Science [4, p.10]».

4. Educating young researchers

To enable the transition to open science, today’s undergraduate students and young researchers in all disciplines need to receive updated education on scholarly communication and scientific publishing in the digital era.

First, young researchers need to understand the new and central relevance of preprints both for their own career, and for society, as preprints dramatically speed up the time to
publication, and allow scholars to regain control of their own research work [19,20,21,23].

Studies first published as preprints are received and evaluated by the broad scientific community based on their own intrinsic quality, independent of the hosting publication platform (i.e., scientific journal), and without the need to pay any charge, neither to publish the preprint nor to access it.

Numerous university hiring and promotion committees and the most important national and international funding agencies now regularly require candidates and applicants to include preprints. It is enough to insert the preprint publication on a preprint server with a DOI in a separate list from peer reviewed articles.

Second, students and junior researchers need to learn how to copy-editing their own work, namely acting as journal editors from starting from answering the main question lately suggested by the editor of a prestigious catalysis journal: is this research meaningful? [25].

“One of the most frequent responses from reviewers is that they can’t see why the work is important” Rowan continued calling authors to answer the question: why did you perform this research? “One of the easiest ways to do this is to pose a question to answer in the introduction” [25].

Learning, for example, how to write “a descriptive and specific title, followed by a concise abstract making the results of the study understandable for a wide audience” [25], and by an informative, interesting, updated and succinct introduction requires the direct involvement of the senior scholars supervising their students and post-docs. Training their PhD students how to write a paper, in other words, should be a core task of every supervisor.

Many universities provide training courses for scientific writing. Suffice it to mention here Rothenberg’s and Lowe’s ‘Write it Right’ workshop on writing research articles, funding proposals and technical reports offered at the University of Amsterdam since 2002 [26]. Attended by more than 1600 people, the course has even become part of the educational program of the Netherlands Organization for Scientific Research (NWO).

Third, young scholars need to know more closely science evaluation practices and citation-based metrics such as the $h$-index [27] or the age-normalised $m$ quotient. By learning how to effectively use statistical data concerning one’s own research, a young researcher will better understand the impact of her/his research, and how it is being used by peers. Eventually, rather than getting rid of bibliometric indicators, such as the $h$-index or the impact factor, she/he will learn how to expand and improve their use in a useful and critical fashion [28].

Fourth, students need to be trained on how to effectively use social media, such as for example Twitter, ResearchGate, Linkedin, Vk, Instagram to share the outcomes of their research with the public and with researchers within and outside their research field [29]. The aim is not “to engage with the social media” but rather to use social media to engage with the public “in a world that increasingly values public engagement and impact” [30]. By doing so, researchers will also expand their own professional network.

Using professionally a news/social medium like Twitter, for example, scholars in science, technology and engineering will be surprised by the number of peers with a Twitter account.

Fifth, scholars usually unfamiliar with copyright legal aspects, need to regain control on their work using tools such as the Scholarly Publishing and Academic Rights Coalition Author Addendum [31]. The latter is a legal instrument that modifies the publisher’s agreement and allows authors to keep key rights to their research, “instead of blindly giving it away to publishers” [32].

The “networked system, governed by researchers themselves, designed for effective, rapid, low-cost communication and research collaboration” [32] called for by Tennant already exists.

By publishing the outcomes of their own work as well written scientific articles (including pictures and videos at virtually no cost) first as fully citable preprints and then in free OA journals, researchers wisely using social media can get all the visibility (citations) needed for their career and funding, aiding to progressively free the large amount of money (the $25 billion a year scientific publishing industry [3]) currently spent for scientific publishing in journals storing the outcomes of their work behind paywalls.

5. Outlook and Conclusions

Trying to answer the question why the Web, created by Berners-Lee in 1991 “to disrupt scientific publishing” [33] actually did not radically change it in the course of the subsequent two decades, Clarke in late 2016 concluded that this was mostly due to the journal’s role of “designation”, namely a cultural function (“the hardest to replicate through other means”) for which, based on a scientists’ publication record in existing scientific journals of high reputation (i.e., impact factor), academic institutions and funding agencies base career advancement and award decisions [33].

Thanks to the advent of preprints in all main scientific fields beyond physics, the number of citations and thus the impact of a scientist’s work is now independent of the publishing platform.

For the very first time, in principle a scientist can become a highly cited scholar without having her/his work published in conventional scientific journals following submission and peer review.

Having received no formal training on scientific publication, most today’s scholars are unaware that “the practice of sending manuscripts to experts outside of the journal’s editorial offices for review was not routine until the last half of the 20th century” [33].
Publishers often invoke their key "quality control" role, objecting that preprints would allow everyone to upload any contents. Yet, the seminal article in which Mullis published the discovery of the polymerase chain reaction (PCR) was rejected by Science [34] and eventually was published in Methods in Enzymology [35]. Three years later, emphasized Mullis in his Nobel Lecture [34], Science proclaimed PCR “Molecule of the Year”.

As to the value of the impact factor, it is instructive to learn that the IF of Methods in Enzymology in 2018 was 1.984, while Mullis’ article reporting the PCR discovery up to September 2019 had been cited 7876 times (Google Scholar).

The four revolutionary scientific papers on the theory of the photoelectric effect, Brownian motion, special relativity, and mass-energy equivalence published by Einstein in 1905, the so called annus mirabilis (extraordinary year, [36]) appeared in Annalen der Physik, namely a journal whose impact factor in 2018 was slightly more than 3 (3.276), and in 1905 was probably not higher than 1.

«We now receive many more interesting papers than we can publish...” typically reads the e-mail (a letter when Mullis in 1986-1987 repeatedly submitted his revolutionary molecular biology work) with which an editor communicates her/his decision to not send for review a manuscript.

«Hence, we regret that we are unable to process your manuscript further and suggest that you consider submission to a more specialized journal».

The time and money-wasting publication cycle of conventional scientific publication begins, with the manuscript going from a journal to another, and peer review starting each time afresh. Eventually, after months or even more than a year, the manuscript is accepted for publication.

Whether or not manuscripts are “selected on the basis of methodological rigor, novelty, quality of writing, and general significance”, until the advent of preprints there was little or nothing scholars could do to object the subjective opinion of journal’s editors and reviewers.

Now, posting online her/his preprint, an author or a team of authors is granted priority on the discoveries and the novel ideas reported therein. Novelty is established, and any media embargo until the date of online publication for which prior to embargo lifting there can be no public mention of the upcoming paper becomes unnecessary.

The infrastructure for preprints and its relatively modest costs will be faced, as it happens today, by public and private universities, foundations, scholarly societies, and even by publishers.

As mentioned in the introduction, the global scientific output is expected to double approximately every nine years [37]. This huge scientific output - most of which will shortly originate from China and India due to their huge population (amounting to 37% of world’s current population) and excellent scientific schools - will appear first and foremost as preprints.

Today’s academics, however, “work to the system in which they find themselves” [38]. Hence, further argued Laurillard calling for a better academic system facilitating and rewarding excellence in teaching and not only in research, promotion of excellence in teaching requires to change the rewarding system. Said reformed academic system will comprise the ability to evaluate scholars for hiring and career advancements based on true scientific merit, and thus also on citation-based metrics, independent of the publishing platform in which the candidate's papers have been published.

I agree with Clarke [33]: “designation” of a scientist by scientific journals is a cultural trait. The etymology of a word is most often revealing. The word “submission” originates from the Latin verb submittere and was first recorded in English around mid 15th century with a clear meaning: "humble obedience” [39].

As we enter the third decade of the 21st century, the time has come for world's scholars to replace journal article “submissions” with a free and open system widely based on freely accessible and freely reproducible preprints in which the value (i.e., quality) of scientific articles no longer needs peer review but is open to the evaluation and use of the whole scientific community which has all the tools and an obvious incentive to separate “wheat from chaff”, identifying quality work for further valorization via subsequent utilization (and citation).

The time foreseen by Évariste Galois in 1831 during which «on s'associera pour étudier, au lieu d'envoyer aux académies des plis cachetés, on s'empressera de publier ses moindres observations pour peu qu'elles soient nouvelles, et on ajoutera: 'Je ne sais pas le reste’ [40]» («scientists will team up to study, instead of sending sealed envelopes to the academies, hastening to publish their slightest observations as long as they are new, adding: 'I do not know the rest’) has finally come.

Author information

ORCID

0000-0002-5096-329X

Notes

The Author declares no competing financial interest.

Acknowledgements

This article is dedicated to Marta Bertini, schoolmate, mother and now grandmother, on the occasion of 38 years of friendship. I thank one anonymous Reviewer for selected suggestions that improved the original version of this Essay.
List of abbreviations

APC = Article Processing Charge  
DOI = Digital Object Identifier  
ePub = Electronic Publication  
HTML = Hyper Text Markup Language  
IF = Impact Factor  
OA = Open-Access  
OECD = Organisation for Economic Cooperation and Development  
PCR = Polymerase Chain Reaction  
PDF = Portable Document Format  
PR = Public Relations  
STM = Science, Technical and Medical

References

[8] ACS Publications, ACS Omega, Overview, 2019. See at the URL: https://pubs.acs.org/page/acsof/submission/authors.html  
[14] RSC Publishing, Chemical Science. See at the URL: www.rsc.org/journals-books-databases/about-journals/chemical-science  
[22] For example, ChemRxiv hosted 1,067 preprints by the end of January 2018, 17 months after its launch. By December 4, 2019 it hosted 3,064 highly read preprints. On December 3, 2019 indeed the platform had reached 4 million downloads.  
[26] Aimed at teaching how to write clear technical English, structure a document effectively, enhance titles and graphics, and make an impact, the workshop includes lectures, various exercises, and writing a mini-version of a manuscript, which is then peer reviewed. See at the URL: http://write-it-right.org.  
[27] H-Index in A. W. Harzing, Publish or Perish, Tama Software Research, Melbourne: 2010. See at the URL: https://harzing.com/poppub/ch1_2.htm  

Scientific Publishing: Education as the Key Enabler for the Transition to Open Science

M. Pagliaro

Originally created for facilitating scientific communication, the internet in principle makes scientific journals no longer necessary. Yet, in an almost opposite fashion to what happened to newspaper publishing, the $25 billion annual income scholarly publishing industry has further flourished following the advent of the internet. Expanding the education of today’s students and young researchers to include modern scholarly communication is the key requisite for the transition to open science.