

# Is the System of Scientific Publications on the Eve of a Revolution? And if so, Toward What?

Marie Farge (École Normale Supérieure, Paris) and Frédéric Hélein (Université Paris Diderot - Paris 7, France)

Today, researchers benefit from an extremely powerful electronic distribution system for scientific articles, which allows any researcher (whose institution has paid the subscription fees) to access, via the internet, most of the articles they need, instantaneously and regardless of one's location. Electronic publishing has indeed enabled libraries to subscribe electronically to a very large number of journals, through so-called big deals, at rates that were initially affordable, especially in cases where these libraries partnered in consortia to negotiate with publishers.

However, behind this simplicity for the researcher lies a digital infrastructure much more complex than that of traditional libraries acquiring and preserving printed articles on paper and, behind the apparent impression of gratuity that the researcher may have, there are exorbitant and constantly increasing costs and contracts that have become opaque. The creation of journal bundles has resulted in pricing that no longer has any relation to the real cost of production for the publisher, and the dematerialisation of articles has allowed commercial publishers to multiply tolls according to usage: pay to read the papers to which your library has subscribed this year, pay to read the archives of journals to which it subscribed in previous years, pay to publish in Open Access, etc. At the same time, electronics has allowed large publishers to make enormous reductions in the costs of producing an electronic paper (and any researcher knows that the raw material and its evaluation are provided free of charge).

In economic terms, this development has mainly benefited big publishers, whose profit margins far exceed those of pharmaceutical companies or banks, but it is increasingly weighing on public finances. Indirectly, it is also the small publishing houses, private or academic, that bear the costs because it is by cancelling subscriptions with these small publishers that libraries manage to balance their budgets. This maintains a vicious circle, which results in a growing concentration of scientific journals in the hands of the largest publishers. These economic mechanisms have had an impact on the quality of scientific publications, as major publishers have a commercial interest in encouraging the proliferation of journals and articles. A visible effect of this is that many scientific communities are saturated by the number of papers and the quality of peer review has fallen. Another effect is the multiplication of predatory journals (there are nearly 10,000 today, all disciplines combined).

## The Open Access project

Since the early 2000s, and particularly since an initiative [1] launched in Budapest in 2002, researchers have dreamed of building a model for distributing articles free of charge for all (thanks to the internet) and thus freeing themselves from the big publishers. Although at first glance this dream seemed within reach, it is far from being realised as, even today, more than 80% of articles are not published in Open Access. The first reason is that any system allowing rapid dissemination (e.g., the arXiv) has a cost, even if it is small in comparison to the one charged by publishers.

Publishers, who initially feared the Open Access movement, have managed to turn the Open Access project to their advantage by proposing a model in which the author pays publication fees, often called "APC" (for *Article Processing Charges*), so that their article is immediately Open Access. However, it would be more correct to call APC a *licence fee* (which we will do in the following) because the amount usually has no relation to the real cost of the service.

There is therefore still a long way to go to build an Open Access publishing model, in which researchers and their institutions regain full possession and free use of the data they themselves have produced, and in which the costs associated with their publication and dissemination are charged at the right price. Indeed, the transition is tremendously complex to set up between, on the one hand, extremely well organised multinationals and, on the other hand, scientists and institutions, divided according to their disciplines, their institutions and their countries and where dialogue is not necessarily well organised neither information does not circulate very well.

Moreover, a fundamental obstacle is that scientific journals combine several essential functions, inherited from the age of printing, namely:

- Evaluation process. This task is mainly carried out by researchers, as members of editorial boards or as referees. It should, however, be noted that this task involves secretarial work, which is usually carried out by a secretary, but also sometimes by researchers, who may be assisted by software. Secretariat funding is provided by the publisher or a public institution, in varying proportions depending on the journals.
- Label provided by the reputation of the journal.
- Referencing and notification of an article. Several actors contribute to this function: the journal itself, data-

bases, such as *Web of Science* and *Scopus* for most disciplines and *MathSciNet* and *Zentralblatt* for mathematics, and libraries, as well as institutional reference systems that are developing.

- Diffusion and promotion.
- Archiving.

The latter two functions can be performed by journals or by archive platforms, such as *arXiv*, but also, as far as archiving is concerned, by public institutions and non-profit organisations, such as *CLOCKSS* [2] and *PKP-LOCKSS* [3], not forgetting libraries for printed versions.

It can be seen that almost all of these functions are services that can be the subject of a call for competition on a market, or be taken over by public institutions or non-profit organisations. All of them are services except ... the label of a journal, which is attached to recognition by the community and which is thus unique. This is the reason why the market for scientific journals is essentially without competition, and this is the main explanation for prices unrelated to the cost of services. As long as journals combine these functions, it will always be difficult to bring subscription prices down to reasonable levels.

### Where do we stand?

After two rather calm decades, it seems that we have reached an unstable situation, and are probably on the eve of great upheaval. Indeed, tough negotiations with publishers are multiplying (as we will see later), institutions are unsubscribing from big deals (see a list of institutions that have cancelled their subscription to big deals in [4]) and the Sci-Hub pirate platform is enjoying worldwide success, with hundreds of thousands of illegal downloads of articles per day. In Europe, the pressure is increasing because the European Union has set, under Horizon 2020, the objective of free dissemination of European scientific production. Finally, the development of alternative solutions is accelerating, as we shall see next.

### An overview of countless innovations

Open Archive platforms like *arXiv* [6], which was created in 1991 by Paul Ginsparg, were long confined to certain science disciplines, such as high-energy physics,

computer science, mathematics and economy.<sup>1</sup> Today, the situation is changing rapidly as new platforms are being created, such as *bioRxiv* [12] in biology in 2013, *ChemRxiv* [13] in chemistry in 2016, and *EarthArXiv* [14] and *ESSOAr* [15] in geosciences in 2018. In addition, recent agreements between institutions to accelerate the development of these archives and to federate them, together with initiatives such as *ASAPbio* [16], are aimed at encouraging researchers in biology and medicine to deposit their preprints on public archives. These initiatives have an international dimension.

This rise in the power of Open Archives can be largely, but not only, explained as a reaction to abusive tariffs by major publishers or to *licence fees* for publishing in Open Access. It also appears that this flexible and commercially unfettered method of dissemination is incomparably faster than the journal system and better protects researchers from plagiarism.

Open Access journals without licence fee, often referred to as *Diamond Open Access journals* [17], continue to grow. The existence of such journals is most often based on projects, such as *Episciences* [18] or *Cedram* [19] (whose activities are expanding with the creation of the *Centre Mersenne* [20]) for mathematics. Similar projects exist in all disciplines (especially in the humanities, which is at the forefront of this movement), such as: the *Public Knowledge Project* [21], which develops *Open Journal Systems*, an open source software to publish journals; organisations such as *Knowledge Unlatched* [22] and *Open Library of Humanities* [23], which propose crowd-funding for the publication of Open Access books and journals; *OpenEdition* [24], which provides platforms for books, journals and blogs; and the *OA Cooperative Project* [25]. In biology, the mega journal *PLOS One* [26], supported by a non-profit organisation, is an Open Access journal that was free of publication fees in its early days (but unfortunately now charges publication fees of the order of \$1,500 per article). In Latin America, the *SciELO* [27] platform, founded by Brazil, includes 1285 journals, most of which are Open Access. These projects are supported by various foundations and organisations, such as *LingOA* [28], *MathOA* [29], the *Foundation Compositio Mathematica* [30], *SPARC* [31] and *SCOSS* [32].

These Open Access journals can either be a new journal, or an already existing title whose editorial board resigns or leaves a commercial publisher so that the journal can continue in an open framework. This is called *emancipation* of a journal [33]. The most notable example of emancipation of a mathematical journal was *Topology*, which became in 2006 the *Journal of Topology*, but there have been other journals before and after, including *Acta Mathematica* (since 2017). A list of emancipated Open Access journals can be found on the *Journal Declaration of Independence* [34]. A list of Open Access mathematical journals without publication fees can be found on the *Cimpa website* [35]. (See also, for example, the *Directory of Open Access Journals* [36] for all fields.)

---

<sup>1</sup> Open archive platforms were developed in the last century by researchers to share their preprints and reprints (note that those years publishers were providing reprints to authors for free and asked them to send them to colleagues for advertising). The *SPIRES High Energy Physics database*, developed at *SLAC-Stanford University* in the '70s, was the first made accessible via the Web in 1991 and replaced by *Inspire* [10] in 2012. It was followed by the database *ADS* [5] developed for astrophysics in 1988 and transferred to the Web in 1994. In 1991 *arXiv* [6] was created by Paul Ginsparg at Los Alamos National Laboratory for physics. In 1994 Michael Jensen founded *SSRN* [7] for social sciences and humanities (in 2013 it was the largest open repository in the world and in 2016 it was sold to Elsevier). In 1997 *PubMed* [8] was designed for medicine and *RePEc* [9] for economy.

Lastly, there are some very interesting innovations in terms of evaluation, including Open Peer Review [17, 37]. The principle is to organise the evaluation of an article in an open way, i.e. by making public the reports of the referees, the authors' responses to the referee and even the contributions of other researchers. Such a system may give rise to fears, but many variants exist, and the result will depend on the quality of the editorial board that oversees this process and on the adjustment of its details (in particular, one can choose to keep references anonymous and to make public only the positive reports, in which case one should really speak of recommendations).

Among the first experiments on Open Peer Review are Copernicus in 2001 [38] and F1000Research [39] (a non-profit organisation currently with low publication fees) and SciPost [40] (without publication fees); they seem to lead to a significant improvement in the quality of reviews. A similar experience, perhaps even more innovative, is "Peer Community in..." [41], a recommendation platform that is not a journal, offering positive evaluations of preprints or articles. These projects take up an idea already proposed in the Self Journal of Science project [42], which has unfortunately had difficulties getting off the ground. The interest in these latter projects is to decouple evaluation (and its associated label) from the rest of the services.

We could also mention the European Digital Mathematical Library [43], and the platform Dissemin [44] which detects papers behind pay-walls and invites their authors to upload them in one click to an open repository, in order to boost open access while respecting legal constraints [44]. Another interesting innovation is the DOI (Digital Object Identifier) resolver that points to a free version of the requested article, when available, instead of its version under pay-wall [45].

The previous list of innovations, institutions and tools is by no means exhaustive since this would be impossible, as new innovations are emerging every month. The term *bibliodiversity* has been proposed to refer to this proliferation and led to the "The Jussieu Call" [46], an initiative aimed at supporting these alternative solutions, which has been signed by more than 100 institutions from 13 countries.

### **The temptation to contract with major publishers for Open Access publishing**

*(This section reproduces a large portion of the article by the authors, "Transition vers l'Accès Libre: le piège des accords globaux avec les éditeurs", which was published in the French newspaper Mediapart on 14 April 2018 [47].)*

A worrying policy in some countries is to conclude national agreements with large multinational publishing companies to pre-pay licence fees for Open Access publishing. As it happens, the countries at the forefront of this movement are essentially those in which these global publishing giants are mostly established, namely the United Kingdom, the Netherlands and Germany. We are therefore entitled to ask ourselves whether this is a

coincidence and we propose here hypotheses to interpret this "coincidence".

### **Let us put ourselves in the shoes of a multinational scientific publishing company**

To understand this situation, it is useful to adopt the point of view of the major commercial publishing groups and ask what would be the most effective strategy for these groups to adapt to these changes whilst maintaining or increasing their profits:

- Firstly, *preserve current income*, without rushing but simply by taking advantage of the inertia of the system that exists today, with more than 80% of articles published in journals with subscription, whilst multiplying the tolls to access articles on the electronic platforms of the publishers.
- Then, invest for the future by developing, just like Google, new "services" based on algorithms to mine the large amount of data accumulated. This is the direction taken by Elsevier, the largest multinational publishing company, which buys back several start-ups that have developed such services each year. The amount of data processed can come both from the vast scientific corpus contained in papers and from the data concerning the researchers themselves (as authors or experts). In this "market", Elsevier competes with Clarivate (formerly Thomson–Reuters).
- Finally, build an Open Access model that is stable and that guarantees a firm's profits to be as large as those they currently make with subscriptions. We will see how delicate this operation is.

Indeed, for the major publishing groups, Open Access has remained, since 2000, both a *source of additional income* and a *source of concern*.

- It is a *source of additional income* thanks to the payment of *licence fees* by researchers (or their institutions) for each article published in Open Access (noting, in passing, that these fees may correspond to articles that are either published in journals that are totally Open Access or *hybrid journals*, i.e. for which libraries already pay for subscriptions!).
- It is a *source of concern* because of the risk that researchers and institutions might emancipate themselves from the current system of journals by building their own Open Access system (for example, this concern can be seen directly through Elsevier's share price, which fell at the time of the *Budapest Initiative* in 2001 and then on other similar occasions). To avoid this risk, these multinationals must build a stable Open Access publishing model that allows them to collect fees as cost-effectively as the current subscription model but, at the same time, does not provoke a hostile and destabilising reaction from researchers and their institutions.

With the current subscription system, a publisher receives, on average, more than 5000 euros per article

published, this revenue being even higher in the case of *Elsevier*. Therefore, to maintain comparable profit margins with the Open Access publishing system, such a publisher would have to charge royalties at an average price of the same order of magnitude. On a laboratory scale, such rates would result in an extremely high bill (and in mathematics, even charging 2000 or 2500 euros, the fees for all items in some laboratories, would absorb the entire budget!). Such a solution is therefore unrealistic. Moreover, even if the budgetary limits of the laboratories were disregarded, this would have a devastating effect on the publisher's image because it would reveal to researchers, in a concrete way, the scandalous level of fees charged, of which, in the majority of cases, they are unaware. Moreover, the fees give those who control the credits excessive power, which is likely to degrade relations between colleagues. Finally, the development of such a model risks, even before it has come to an end, provoking an acute awareness and reaction from the scientific community, which would compromise the commercial publisher's projects. Indeed, there are now more and more initiatives from the scientific community that could well be the premise for such a reaction.

The best solution for these publishers therefore remains to obtain payment of fees from their traditional interlocutors, libraries, which already have large budgets capable of supporting expensive subscriptions, and national agencies or institutions. Thus, the fees will not weigh directly on laboratories' budgets, will remain invisible to researchers and, even better, will "smoothly" replace the astronomical sums that libraries are used to paying. Publishers will thus be able to continue their "business as usual" in complete security.

The first country where such a model was tested was the United Kingdom. The British have been engaged since 2013 in a transition programme toward Open Access, combining the use of Open Archives and the payment of fees to publishers. This required the establishing of a complex protocol for the institutions and the creation of a special fund to finance the additional costs involved. Despite a political will to control overall costs (subscriptions and fees), it is clear that no expenditure could be contained. Worse still, the system of reimbursement of fees to universities by the special fund has created a bureaucracy whose cost has added to the bill. There is therefore quite a bit of discontent with this system in the United Kingdom. This experience has encouraged the major publishing groups and countries tempted by this direction to move towards global agreements on a country scale, the bill of which would be paid by libraries.

This is the path taken by the Netherlands, by concluding a first agreement at the end of 2014 with Springer that integrates a subscription to a bundle of journals with the right for Dutch researchers to publish Open Access at no additional cost (these are therefore included in the subscription invoice). Similar agreements have since been concluded with other publishers and in other countries: in Austria and Germany at the end of 2015, in Sweden in 2016, in Finland in 2018, etc. The type of contract varies but there is a shift toward contracts in which the propor-

tion of fees for Open Access publications is becoming increasingly important. Thus, from the end of 2016, Germany embarked on an even more radical path: instead of wanting to conclude mixed agreements, concerning both subscriptions to read and fees to publish, Germany believes that it is no longer a question of paying to read but just paying to publish.

In any event, it is striking to observe that this movement toward global contracts including Open Access fees (which, as we have seen, is certainly the most satisfactory solution for multinational publishing companies) is developing mainly from the United Kingdom, Germany and the Netherlands and in the geographical area around these countries. But it is also striking to note that, with the exception of the American Chemical Society, the main publishing companies (*Elsevier*, Springer Nature, Wiley and Taylor & Francis) are precisely located in these three countries. It should also be noted that small scientific, commercial or academic publishers are not involved in these agreements and risk, once again, paying the price for these developments.

There may be several explanations for this coincidence: the result of lobbying by these publishing firms cannot of course be excluded, but the most plausible explanation is the conflict of interest situation in which these three countries naturally find themselves. Indeed, public institutions in these countries must certainly try to reduce, if not contain, the bill they pay to publishers but, at the same time, we can assume that it is difficult for them to make choices that would harm the multinationals based in these countries, not so much for fear of weakening them as for fear that these companies would threaten to relocate their activities to other countries.

In Germany, the Max Planck Society (*Max Planck Gesellschaft*), the main German research institution, which has played an active role in promoting Open Access since 2003, is also in an ambiguous situation. Stefan von Holtzbrinck, who owns more than half of Springer Nature's capital, is a member of its Board of Directors. Stefan von Holtzbrinck is also the president and co-founder of the *Max Planck Förderstiftung*, a foundation that financially supports the Max Planck Society.

With regard to Germany, it is important to distinguish two trends that are developing in independent directions:

- On the one hand, the German institutions have joined forces within the DEAL consortium [48] to negotiate hard with Elsevier, Springer Nature and Wiley, in order to obtain significant price reductions and to obtain transparent contracts (since, until now, the details of these contracts have remained confidential, an opacity that benefits publishers). Having failed to reach an agreement with Elsevier at the end of 2016, the consortium decided that, as of January 2017, all contracts with Elsevier that were due to expire would be terminated, resulting in a standoff with this publisher. Of course, we can only applaud this exemplary determination.
- On the other hand, as we have seen, the Germans demand contracts that guarantee them free access to journals but they agree to pay to publish in Open Ac-

cess. This radical position is inspired by a strategy developed by the Max Planck Society, which is the subject of a promotional campaign inviting institutions around the world to rapidly switch to an Open Access publication system, even if this means paying licence fees in advance. The Max Planck Society thus proposes to institutions around the world to commit to this path by signing the Initiative *OA2020* for *Open Access 2020* [49] – not to be confused with the *H2020* programme of the European Union for financing research! Note that this initiative also calls for transparency of costs – currently refused by publishers – and for a certain diversity of models.

For the moment, no final agreement with the publishers has been concluded. It seems that the reason for this is, not so much the reverse model (paying to publish instead of paying to read), but the price to pay and the complexity of the deals. As long as publishers are not asked to significantly reduce their turnover, they have every reason to be satisfied to the extent that they are offered a stable solution. Indeed, Springer and Wiley affirm their will to bring these negotiations to a successful conclusion and, even if relations with them are much more tense, Elsevier clearly indicated in a note [50], made public in September 2017, that they do not disagree with the principle but on the price and details of the implementation of the changeover.

On the other hand, the selling point [51] underlying the *OA2020* initiative, promising a significant reduction in tariffs, unfortunately seems simplistic and it is hard to see what serious element could support it once one market without competition is replaced by another market without competition. On the contrary, the risk seems great that the result will be the creation of a new Open Access publishing model in which large companies will be able to continue to dictate their financial terms (see [52]).

### As a conclusion

As we have seen, Germany has been engaged since January 2017 in a tough negotiation with Elsevier. Likewise, the French institutions, grouped within the Couperin consortium, have recently followed a similar approach with Springer (but for a big deal subscription contract that does not include Open Access). Failing to obtain a significant reduction in Springer's rates, these institutions terminated their contracts in January 2018 ... until this publisher become reasonable. Finally, many institutions [53] around the world have simply cancelled their subscriptions to big deals. All these steps may signal a change in the power balance. It is interesting to note that in Germany, after cutting access to institutions that had terminated their contracts for a few weeks, Elsevier finally reinstated them in February 2017. Similarly, in France, Springer has not cut off access to French institutions since January 2018.

On the other hand, there is a risk that in many countries, decision-makers and library managers, eager to make a transition to Open Access, will give in to temptation to follow the examples of northern European coun-

tries by signing agreements with major publishing groups that would include the advance payment of licence fees for Open Access publishing. We must avoid this scenario, which would further strengthen the hold of these large groups.

In addition, the transition to Open Access must not replace commercial barriers to read with barriers to publish. Research results and, more generally, data produced by public institutions must stop being privatised, which does not exclude using private providers to disseminate and make them visible.

### References

- [1] Budapest Open Access Initiative: <http://www.budapestopenaccessinitiative.org/read>.
- [2] CLOCKSS: <https://www.clockss.org/>.
- [3] PKP-LOCKSS: <https://pkp.sfu.ca/pkp-lockss/>.
- [4] SPARC Big Deal Cancellation Tracking: <https://sparcopen.org/our-work/big-deal-cancellation-tracking/>.
- [5] Astrophysics Data System: <http://adsabs.harvard.edu/>.
- [6] ArXiv: <https://arxiv.org/>.
- [7] SSRN: <http://www.ssrn.com/en/>.
- [8] PubMed: <https://www.ncbi.nlm.nih.gov/pubmed/>.
- [9] Research Papers in Economics (RePEc): <http://repec.org/>.
- [10] Inspire: <http://inspirehep.net/?ln=en>.
- [11] PubMed: <https://www.ncbi.nlm.nih.gov/pubmed/>.
- [12] bioRxiv: <https://www.biorxiv.org/>.
- [13] ChemRxiv: <https://chemrxiv.org/>.
- [14] EarthArXiv: <https://eartharxiv.org/>.
- [15] ESSOAr: <https://www.essoar.org/>.
- [16] ASAPbio: <http://asapbio.org/about-2>.
- [17] M. Farge, Scholarly publishing and peer-reviewing in open access, <https://zenodo.org/record/1227093>.
- [18] Episciences: <https://www.episciences.org/>.
- [19] Cedram: <http://www.cedram.org/>.
- [20] Centre Mersenne: <https://www.centre-mersenne.org>.
- [21] Public Knowledge Project: <https://pkp.sfu.ca/>.
- [22] Knowledge Unlatched: <http://www.knowledgeunlatched.org/>.
- [23] Open Library of Humanities: <https://www.openlibhums.org/>.
- [24] OpenEdition: <https://www.openedition.org/>.
- [25] OA Cooperative Project: <http://oa-cooperative.org/about.html>.
- [26] PLOS ONE: <http://journals.plos.org/plosone/>.
- [27] SciELO: <http://www.scielo.org/php/index.php?lang=en>.
- [28] LingOA: <https://www.lingoa.eu/>.
- [29] MathOA: <http://www.mathoa.org/>.
- [30] Foundation Compositio Mathematica: <https://compositio.nl/stichting2.html>.
- [31] SPARC: <https://sparcopen.org/>.
- [32] SCOSS: <http://scoss.org/>.
- [33] I. Daubechies, IMU blog: <https://blog.wias-berlin.de/imu-journals/>.
- [34] Journal Declarations of Independence: [http://oad.simmons.edu/oadwiki/Journal\\_declarations\\_of\\_independence](http://oad.simmons.edu/oadwiki/Journal_declarations_of_independence).
- [35] CIMPA: <https://www.cimpa.info/en/node/62>.
- [36] DOAJ (Directory of Open Access Journals): <https://doaj.org/>.
- [37] J. Bordier, OpenEdition experiments open peer review protocols for scientific articles, <https://oep.hypotheses.org/1661>.
- [38] Copernicus: <https://publications.copernicus.org/>.
- [39] F1000Research: <https://f1000research.com/>.
- [40] SciPost: <https://scipost.org/about>.
- [41] Peer Community in...: <https://peercommunityin.org/>.
- [42] Self Journal of Science: <http://www.sjscience.org/>.
- [43] EuDML: <https://eudml.org/>.
- [44] Dissem.in: <https://dissem.in/>.
- [45] Digital Open Access Identifier: <http://doai.io/>.

- [46] Jussieu Call: <http://jussieucall.org/index.html>.
- [47] <https://blogs.mediapart.fr/edition/au-coeur-de-la-recherche/article/130418/transition-vers-l-acces-libre-le-piege-des-accords-globaux-avec-les-editeur>.
- [48] DEAL: <https://www.projekt-deal.de/about-deal/>.
- [49] OA2020: <https://oa2020.org/>.
- [50] Elsevier: <https://www.elsevier.com/connect/working-towards-a-transition-to-open-access>.
- [51] R. Schimmer, K.K. Geschuhn, A. Vogler, Disrupting the subscription journals' business model for the necessary large-scale transformation to open access, <http://pubman.mpdl.mpg.de/pubman/faces/viewItemOverviewPage.jsp?itemId=escidoc:2148961>.
- [52] <https://poynder.blogspot.fr/2018/03/the-open-access-big-deal-back-to-future.html>.
- [53] SPARC: Big Deal Cancellation Tracking, <https://sparcopen.org/our-work/big-deal-cancellation-tracking/>.



Marie Farge [[marie.farge@ens.fr](mailto:marie.farge@ens.fr)] is a research director at CNRS-INSMI and has worked at LMD-ENS Paris since 1981. She has a Master of Science from Stanford University, a PhD in physics from Université Paris VII and a PhD in mathematics from Université Paris VI. She studies turbulent flows of incompressible fluids and hot plasmas by direct

numerical simulation and their applications in geophysics, nuclear fusion and aerodynamics of insect flight. In 1988, she pioneered the use of wavelets in fluid mechanics and participated in the creation of the journal *ACHA* (Applied and Computational Harmonic Analysis) in 1992. She has been member of the Ethics Committee of CNRS, member of RISE (Research Innovation and Science Experts) of the European Commission, and is a member of the Council of the *Academia Europaea*.



Frédéric Hélein [[helein@math.univ-paris-diderot.fr](mailto:helein@math.univ-paris-diderot.fr)] is a professor at the University Paris Diderot. He first worked on the analysis of harmonic maps between Riemannian manifolds, the asymptotic analysis of the Ginzburg–Landau equation and the study of some completely integrable systems in differential geometry. His recent work concerns mathematical physics, especially the study of the Hamiltonian structure of field theories and gauge theories within a differential geometric viewpoint. He is the scientific chair of the French Réseau National des Bibliothèques de Mathématiques and participates in negotiations with publishers and in various discussion groups working on scientific publications policy.

---

# Can Statistics Predict the Fields Medal Winners?

Adam Bannister and Olaf Teschke (FIZ Karlsruhe, Berlin)

---

With the upcoming ICM in Rio de Janeiro, the seasonal speculation about who will receive the 2018 Fields Medals at the opening ceremony is once more in full swing. With the big data industry measuring us in all possible ways, a natural question might be whether statistical approaches could possibly predict the committee's choice of the Fields Medallists. We undertake some experiments here to see which predictions are provided by standard approaches based on data from zbMATH and linked data sources.

## Fields Medal is small data

One obvious obstacle, however, is that the set of Fields Medallists is small by its very nature and may easily defy statistics with all kinds of outliers. As a nice example, one could recommend reading Borjas-Doran's study on a statistical decline of Fields Medallists' productivity [BD14] and Kollár's amusing review [K15]. Without further discussing the fundamental problem of measuring a mathematician's productivity by publication and citation numbers – the fallacies of this approach have been frequently discussed in the newsletter, for example in [BT17] – we just note here the very last observation in [K15]:

*“The limits of statistics are illustrated by the numbers contained in the penultimate line of [BD14, Table 1]. (It is not commented on in the paper.) While most of the Fields Medallists and contenders are happily alive, Figure 3 shows a disturbing pattern about those who have passed away [...Namely, an average age of death of 74.0 for Fields Medallists compared to only 66.3 for contenders...] Thus, if you got a Fields Medal, you can expect to enjoy your extra US\$120,000 per year for almost eight more years.”*

Firstly, we may take this as an illustration of how seemingly exact science is often perturbed by possibly unreliable data. It was, for us, impossible to reproduce the average age of death of 74 from [BK14] ([K15] does not comment on this figure). Submitted in 2014 before the death of Grothendieck, the nine Fields Medallists deceased at that time reached an average age of 78.5. (A closer look at the appendix of [BK14] reveals that the 1936 medallists Ahlfors and Douglas seem to have been excluded from the study but that has almost no effect on this average). Secondly, as we are all sadly aware, this figure has been significantly affected since then by the