Researchers and Scholarly Communications - An Evolving Interdependency

David C Prosser
Executive Director, RLUK

We tend to think of the history of scholarly journal publishing as an unbroken thread of consistent activity since the founding of the first journals. In reality, journals have changed in many ways over the past 400 years, and so has the research environment. It may be obvious that publishing activity will alter to reflect changes in the research process, but it may also be argued that the research process itself has changed, at least in part, as a result of developments in scholarly communication. This paper will investigate the two-way flow of influence between research and publishing, both in terms of a 400 year history and looking forward to new developments.

The Rise of Journals

The birth of the modern system of scholarly communications is generally taken to be the the launch of the first two scientific journals, the *Journal des Savants* in 1665 and the *Philosophical Transactions of the Royal Society* in 1666. These journals reflected the growing interest in natural philosophy at the time, but also, through their very existence, began to initiate changes in scholarly practice. Before the mid-seventeenth century the norm was not for 'scientists' (to use an anachronism: they would not have recognised the term) to share their findings. While as concerned with issues of priority as modern researchers - witness the famously bitter battle between Newton and Liebnitz regarding the invention of calculus in mathematics - they were hoarders of knowledge, reluctant to give what we would describe as a competitive advantage to their rivals. This led to some wondrous contortions, not the least of which was Galileo's 'announcement' of his discovery of the rings of Saturn in 1610. Galileo wanted to be able to prove priority, but did not want to actually release the information, so his announcement took the form of a 37-letter anagram in Latin, which even if solved would prove to be unhelpfully cryptic.

It was within this environment of secrecy that Henry Oldenburg, the first publisher of the *Philosophical Transactions*, had to operate. To overcome the natural reluctance of researchers Oldenburg wrote to all the leading scientists of the day stressing the advantages of publication as a means of establishing priority. If the discovery was described in black and white there could be no quibbles over who the discoverer was. So, from the very start, a major incentive of publication was

not communication to one's peers in the hope that they may be able to build upon the work described, but rather a desire to prove one's priority and intellectual worth.

As time passed, research (particularly in the sciences) became industrialized and professionalized. By the mid-nineteenth century 'amateur' scholars such as Darwin, who lived as a country gentlemen with inherited wealth, worked in parallel with a growing number of professionals, such as Faraday, who was employed by the Royal Institution. Darwin was notoriously reluctant to publish his findings and theories. He was compelled to publish his first paper on natural selection only because he wanted to establish his priority over (or at least independence from) Wallace who was about to publish a remarkably similar theory. Darwin's general reluctance to publish is often attributed to both his perfectionist nature and his concern at the reaction to the perceived blasphemous nature of his theories. But perhaps it was also that Darwin did not *need* to publish. He was seeking no grants, nor did he have employers to appease or REF returns to consider. As an independent researcher he was free from the compulsion of 'publish or perish'.

Faraday on the other hand published regularly, with over 45 papers under his name in the *Philosophical Transactions*. True, he was more of an experimentalist than Darwin, but the key difference was that Faraday was an employee, a professional scientist, whose publication record formed evidence of skill as a researcher. Although not formally judged on the number of papers he published, it is clear that publication was a means of establishing both priority and prestige for Faraday (and, by extension, the Royal Institution).

So, in less than 200 years we can witness a shift from researchers as secretive hoarders of knowledge to researchers as published self-publicists.

The Role of Journals

As we have seen, communication is not necessarily the only, or even the most important, part of scholarly communication. In fact, journal publishing has traditionally been described as fulfilling four functions: Registration, Certification, Awareness, and Archiving.[i] That is:

> Registration - the author wishes to ensure that she is acknowledged as the person who carried out a specific piece of research and made a specific discovery.
>
> Certification - through the process of independent peer-review it is determined that the author's claims are reasonable.

Awareness - the research is communicated to the author's peer group.

Archiving - the research is retained for posterity.

Obviously, the weight assigned to each of these functions by the researcher will vary depending on their role as either author or reader. Registration, as described above, is most important for authors as it is the means by which they can stake a claim to the research. Certification benefits both authors and readers. For the author it improves the quality of their work (by providing independent feedback) and allows it to enter the scientific record as a valid piece of work. For the readers, certification guarantees a certain level of quality and relevance to what they are to read.

Awareness ensures that the author's work is widely known (increasing the chance that their work will be read and cited) and it ensures that the reader will be able to find the work they need. The guarantee of long-term archiving gives authors comfort that they will forever be associated with a particular piece of work and readers the reassurance that they will be able to find historical research in the scientific record.

An emergent function of the scholarly communication system is that it provides input into the scholarly reward structure. There is increasing competition for academic posts and for research grants (for example, the UK's Medical Research Council currently funds only 15% of grant requests). Administrators (both within institutions and at funding agencies) need means by which they can rank researchers and separate the funded from the unfunded. Increasingly they have turned to researchers' publication histories for this, working on the basis that past results are an indicator of future success. Ideally the quality of individual papers should be assessed. But this is a huge task, and administrators look to the standing of the journal to act as a proxy for the quality of the research and researcher.

The most famous proxy of quality is the Impact Factor, published by Thomson Reuters. The Impact Factor (IF) is a measure of the average number of citations papers published in the journal in a given year received in a defined time-window. The problems with using IF as a proxy for the quality of individual papers are well-documented (and have been highlighted even by their creator, Eugene Garfield[ii]). Not least is the fact that in even the most highly cited journals a significant proportion of papers receive no citations. At best, the IF tells you that your work is in good company, not that it necessarily has merit in itself. But for many the convenience of the IF and

their reassuringly scientific-looking accuracy to three decimal places outweigh the problems, and some funders have gone so far as to preferentially reward researchers who have managed to get their papers in high-impact journals, irrespective of the intrinsic quality of research. Fundamentally, researchers do not just want to publish good research, they want to publish in the highest impact journals they can, or at the least in a journal covered by Thomson Reuters. There is a perceived hierarchy of journals in each field (based mainly on impact factor rating) and authors want to publish as far up this hierarchy as possible. The greater the kudos of the journal, the greater the kudos of the researcher, and the greater the chance of a successful future promotion or research grant.

This use of journal quality to make funding and promotion decisions is a clear example of the way in which the scholarly communications system has had a direct effect on the way in which research is done.

Electronic Futures

It is remarkable that despite the ubiquity of the internet over the past 20 years it has had little effect on the fundamentals underpinning journal publishing. This may appear surprising - what about online access, citation links between papers, big deals, etc? Certainly, the delivery mechanism has obviously changed beyond recognition. No longer is a publisher required to package papers into bundles to form print issues, to be posted around the world (taking weeks or months to reach the customer), and placed on library shelves to be perused by researchers (or, more likely, graduate students on their behalf). Today, access is 24/7, at the desktop - for those affiliated with a subscribing institution. And new business models such as big deals (where libraries purchase electronic access to a publisher's entire output rather than selected individual titles) mean access to more content - again for those affiliated with a subscribing institution. New thinking around business models has also led to the open access movement and this will, I believe, soon become the dominant business model for scholarly journals in the internet age.

But while electronic journals, big deals and open access are important in themselves, they are 'only' variations on the access mechanisms and business models. They have had almost no effect on the content of journals. It is true that citations can now link straight through to papers, protein names can be linked to databases and suchlike, but the underlying nature of the journal article is the same today as it was 50 years ago. They retain their decades-old structure and are still written in a rather

stilted, international style that many claim is not conducive to communication or the reproducibility of results.

To look at what scholarly communication can be, we need to separate the functions of a journal as described above. We need to accept that a document that allows for a means of conferring reputation on a researcher may not be the same as a document that transmits the maximum amount of information.

One factor that will affect the way in which papers are written is the changing way in which papers are read. There have been complaints of information overload since the beginning of written communications, so we must take today's complaints with a small pinch of salt. But one thing is indisputable: the literature is growing, and the extent of discipline range in which any given researcher can hope to retain mastery is ever narrowing. As an example, a search of *PubMed* for papers published in 2011 with 'HIV' in the title returns 8,500 hits - a publication rate of approximately one paper an hour over the year. If we extend the search to both title and abstract we get over 13,500; and a search in the same period for 'HIV' or 'AIDS' gives almost 17,000 returns - one article published every 30 minutes. In many fields the literature is growing more quickly than practitioners can hope to read it. And this assumes that researchers will read only in their own field. A scientist who wishes to gain insights from similar research in other fields will have an even greater mass of literature to read.

Data Mining

Thankfully, technology gives us one way of tackling the ever growing literature. The emerging fields of text and data mining use computers to scour the literature, looking for connections between papers that human readers may not have made. A researcher studying the causes of epilepsy may not realise that their work has implications for those studying the causes of migraines (and *vice versa*). These researchers would have no cause to read the literature in the other's field. But text mining could discover the links, to the potential benefit of both fields.

There are a number of barriers to the successful uptake of text and data mining. Not least is the fact that most publishers do not automatically allow text mining of their corpus even if the researcher's institution is paying for access. The researcher needs to contact the publisher and ask for permission. An example of the problems this causes was highlighted by the Wellcome Trust as part of a case study on data mining in the field of malaria. The Wellcome began by compiling a corpus

of potentially interesting papers related to the study of malaria. What they discovered was that these papers represented output from almost 100 different publishers. Somebody wishing to data mine these papers would need to gain permission from each of these publishers - a massive administrative burden.[iii]

In the UK, the recent *Hargreaves Review of Intellectual Property* has recommended that a copyright exception be introduced to allow text and data mining of content to which customers had legal access (for example through subscriptions)[iv]. This is being fiercely contested by publishers, presumably as they view text and data mining services as a future revenue-generating opportunity .

But practical considerations aside, a bigger question relates to the nature of information. If the predictions that papers will increasingly be read by machines rather than humans proves to be true then surely these papers should be written in a style that makes them easier for machines to read? Why surround them with the trapping that humans need to render them readable? Publication could then be the expression of facts in a machine-readable way. The facts would not necessarily have to possess the 'novelty' or import that is traditionally expected of a paper-worthy result; and negative results could be included more easily than they are at the moment. (There is a perceived bias against publishing null and negative results even though they can provide interesting and useful data points.) This would give a much richer database of information out of which new connections could be discovered.

Of course, for this to work then new reward and citation structures would be required. An emergent discovery that comes about from data-mining 10,000 genetics papers cannot possibly cite all the papers and credit all the authors. The power of this method will come from the aggregation of massive datasets, not necessarily from individual genius and insight. And if we move away from citation then we potentially move from the impact factor which currently serves as an easy means of assessment for administrators.

In fact, it could be argued that the reason scholarly communications in the internet age looks so similar to scholarly communications in the print age is precisely because the needs of promotion, tenure, and granting committees outweigh the needs of communication and instil an inertia in the system which prevents a move to new systems of greater utility. There are few ways in which this inertia could be overcome. One would be for far-sighted funders to revise completely the ways they assess researchers and their worth. In the UK perhaps the greatest hope would be for the Wellcome Trust to make such a move.

The other way in which a change in the system might happen would be if we found that we could no longer afford the current means of scholarly communication. The serials crises is well known and although the move online and big deals (which gave access to increased numbers of titles for much the same cost) provided temporary respite we now see a number of libraries having to cancel titles. But there are also less well-known hidden costs, such as the that of peer review. It has been estimated that peer review costs the UK approximately £200 million per year in opportunity costs.[v] Many journal editors complain that it is becoming increasingly difficult to find reviewers (although this may be a perennial complaint of journal editors). For their part, reviewers complain that they are asked to review papers for one journal that they have already rejected for another. It may prove to be that the entire peer-review, subscription journal model becomes a luxury that we can no longer afford

A new type of journal that begins to address this issue of multiple review has appeared over the past few years. *PLoS One* and its growing number of clones judge papers on their technical correctness, rather than any attempt to evaluate appropriateness for the journal's audience or impact.[vi] The job of evaluating the importance of the research is seen as being that of the readers rather than the reviewers. Rather than asking 'is this paper right for the journal?' the referees are essentially asked 'is this paper right?'. This strips out the multiple submit-review-reject-resubmit loops and so may go part of the way to reducing the costs of publication.

Cost-reduction may also result following a move to open access as we have an opportunity to introduce the market forces that are currently lacking in subscription models where readers are totally divorced from the costs of subscriptions. (Although it should be noted that if we are not careful in how we structure the payment of publication fees there is a risk that we will replace a failed subscription market with a failed open access market.)

Interdisciplinary and International Research

One of the current fashions amongst research funding bodies is to encourage inter-disciplinary research. The belief is that major new breakthroughs will take place not necessarily by digging deeper and deeper into more and more narrow sub-disciplines, but by bringing together insights from different disciplines and encouraging cross-pollination of ideas and techniques. The traditional print journal does not encourage inter-disciplinary research - it bundles together a group of similar papers into issues and volumes. As these bundles become larger and larger the journals

may split into daughter journals - Parts A, B, C,... - each daughter part more narrowly focused than the parent. In the days of journal browsing this gave readers the opportunity to gain a deep knowledge of their sub-field, but it does not encourage the acquisition of knowledge from beyond that field.

Today, browsing through journals is rarely the researchers' primary means of identifying articles for researchers. Most will start with a search of abstracting and indexing (A& I)services (perhaps the ultimate abstracting and indexing service, *Google*). These services begin to break down the physical and virtual boundaries of the journal and bring all papers of potential interest to the reader (assuming all the relevant journals are covered - gaps in coverage of A&I services being a topic for another paper).

This breaking-down of barriers has implication for the library and its notion of 'the collection'. It has been noted that after moving to big deals institutions see a great deal of usage of journals not previously subscribed to. There are a number of reasons for this. First, no library could afford to subscribe to all of the titles it needed to satisfy the information needs of its users. There was always bound to be unmet demand. Second, the 'subscribed' titles today often reflect the holdings of ten years ago or more. The list of individual titles that a library would have subscribed to in the absence of the big deal would have changed over time. But is also the case that even the most focused 'core collection' would always miss certain journals (and articles) likely to be of interest to their researchers. This is especially true in inter-disciplinary research. A library may have an excellent collection of surgery journals, but the key research insight may come from a paper published in a materials science journal.

PLoS One described above is beginning to move us away from the idea that papers need to be sorted into journals and the idea of a journal as a discrete unit may be on the wane as *PLoS One* and its imitators continue to thrive and grow.

Large publishers offering big deals would argue that by moving from subscribed titles to the big deal database sales models they are empowering inter-disciplinary research. This is true, but the fear is that as more libraries move away from big deals (as they slowly, but surely are doing) we shall reduce the ability to carry out inter-disciplinary research at the very moment that we wish to expand it.

There are similar arguments concerning collaborative research across institutional and national borders, again an area that receives attention from research funders and is heavily promoted. To be successful researchers must be able to share research infrastructure. This may be elaborate such as vastly expensive instrumentation (especially in high energy physics and astronomy), through to the more prosaic internet and email. One part of the infrastructure is the research literature. At the moment, distribution of access to the research literature is highly skewed to the richer (mainly Western) countries, with good access in some poorly developed countries where publishers offer discretionary charity access (e.g. *HINARI, AGORA*, etc). This highly skewed pattern is not conducive to good collaboration. It is harder for two people to work together as equals if one has access to a much greater range of the scholarly literature than the other.

As we continue to see greater emphasis on cross-disciplinary and collaborative research then the failings of the current communications systems will become more apparent and there will be greater pressure to revise how we communicate. Open access will provide part of a solution and it is interesting that the fields dominated by the largest shared instrumentation (CERN for high-energy physics, telescopes for astronomy, etc) we already have very popular and highly used open access solutions (*arXiv, ADS*, etc.). We also see this in biomedicine, where genetic resources (the human genome, protein structure databases, etc.) are made freely available.

Conclusion

Publication of research results has been an integral part of the research process for over 400 years. Journals have adapted to fulfil the evolving requirements placed upon them by the scholarly community and they continue to adapt as they find their place online. But journals have also, by their very nature, affected the research process and moulded the behaviour of researchers. This two-way interaction is set to continue as new technology and the shifting priorities of research funders allow new iterations of a centuries-old tradition.

[i] Roosendaal, Hans E. and Peter A. Th. M. Geurts (1997). 'Forces and functions in scientific communication: an analysis of their interplay.' *Cooperative Research Information Systems in Physics*, August 31—September 4 1997, Oldenburg, Germany. <http://www.physik.uni-oldenburg.de/conferences/crisp97/roosendaal.html>.

[ii] The Agony and the Ecstasy— The History and Meaning of the Journal Impact Factor, Eugene Garfield http://www.psych.utoronto.ca/users/psy3001/files/JCR.pdf

[iii] http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtvm054838.pdf, p10

[iv] http://www.ipo.gov.uk/ipreview-finalreport.pdf

[v] http://www.rin.ac.uk/our-work/communicating-and-disseminating-research/activities-costs-and-funding-flows-scholarly-commu

[vi] http://www.plosone.org/static/information.action