Publishing perspectives from a poacher-turned-gamekeeper

Based on a paper presented at the UKSG One-Day Conference, ‘Untying the knots and joining the dots: researchers’ needs from funding to outputs and beyond’ London, November 2014

The open access (OA) agenda has gained traction in the UK, and the academic community is embracing it. Despite this, compliance rates of OA deposition into journals and repositories are below the levels expected and hoped for. This is despite the growth of OA and hybrid journals and the OA mandates from funders. In this article I will attempt to address this from the perspective of an academic researcher, to shed some light on the OA compliance gap. What are the pressures on academics and what is limiting the successful transition to full compliance? What are the pinch points in the process from an academic viewpoint? What are academics looking for from publishers and repositories to help them in their research? I will also attempt to outline four key areas in this publication and reporting process that we, as an academic publishing community, should aim to improve.

Introduction

The growing pressures on academics to publish their research in high impact, open access (OA) journals show no signs of decreasing. Everyone wants ‘bang for their buck’ across a broad spectrum of stakeholders, including governments (via the Research Excellence Framework [REF] 2014), funders, deans and heads of schools, institutional repositories (IRs), academic colleagues, students, and the general tax-paying public, as well as the publishers themselves. How have these various pressures, many of which are in part conflicting, impacted on the academic? I hope to offer some observations in the following article from a variety of perspectives, some of which may not be obvious to those working outside the academic sector or who have not had direct experience of managing a research group in a UK higher education institution (HEI) in the last ten years or so.

Firstly, who am I, and why do I feel qualified to spout forth? I would describe myself as a jobbing academic working in the higher education sector, having published around 100 papers, with broad experience of publishing and funding. I run a modest-sized lab with six to nine postdocs and students whose careers and aspirations I am partly responsible for, as they also need to publish. My research is also often collaborative and multidisciplinary and, like me, my colleagues are naturally keen to publish their work. Hence, my natural curiosity to conduct biological research and native instinct to publish and see my work disseminated throughout the research community drives me on as a ‘poacher’. In this regard, doing science unfettered by bureaucracy and associated irritations (like writing it up and publishing it, writing/reviewing more grants, etc.) is part of what attracts people like me and my group to research. However, there is also (and always will be) a bit of a ‘gamekeeper’ in most academics. We recognize the requirement to publish our research in OA journals, the need to seek funding in a competitive system, and the value of peer review to both these processes. Similarly, there is a wider recognition of the impact agenda in academia. For my part, I am a grant reviewer, as a chair of one of the Biotechnology and Biological Sciences Research Council’s (BBSRC) responsive mode committees, so I see hundreds of grants a year. I sit on the Project Board of Europe PubMed Central, a project actively seeking to promote open access publishing.

‘pressures on academics to publish their research in high impact, OA journals show no signs of decreasing’
I will cover some of the challenges encountered by scientists when reporting research findings and depositing our outputs into journals, repositories and databases. Additionally, I will give a brief flavour of my personal preferences for searching the literature and associated repositories, which indirectly informs the publication process. Finally, I am going to try and propose some solutions via a four-point plan which addresses some of the problems associated with OA publishing.

Although I believe I am fairly representative of the busy, time-poor academic with a full research, teaching and administration portfolio, I should just add this disclaimer: much of what I write here is personal to my own discipline and me. I am a bioscientist specializing in computational biology and proteomics, with training in biology and physics, and you should take all of what I write, therefore, with the appropriate pinch of salt.

**360° publishing pressure**

As already noted, there are a large number of somewhat conflicting pressures on modern academics, just like everyone else in the scholarly publishing ecosystem. And just like everyone else, academics are easily confused by the large number of requests for information, requirements to deposit articles and metrics, fill out forms and collect all the associated data so it can be reported in a variety of guises. There are seemingly so many things academics are supposed to do, with requests from all directions. For example, if you are lucky, funders give you money. Naturally, they want to see published outputs and want associated impact to be generated, so they can report back to governmental bodies like the Department for Business, Innovation and Skills (BIS) on just how wonderfully successful it has all been and therefore, please can we have some more money for science (cf. the current Comprehensive Spending Review [CSR]). Academics also have line managers, which might come as a bit of a shock to some, perhaps. In my case, it is a supportive dean of my faculty, who is also an OA evangelist. Regardless, the bosses want their academics to publish papers in high impact journals for the REF. And the dean has a boss too, the vice chancellor, who wants the faculties to make a high quality REF submission since university funding, league tables and reputations also depend on this and, in some cases, maybe even livelihoods.

In tandem with school/faculty pressures to publish, the university library is also interested. In my case, at the University of Manchester: a very useful and enlightened Library. They want me to put all my outputs in the institutional repository (IR) so they can track the trajectory of the next REF submission. It is also self-evidently good practice to internally monitor and store the collective digital outputs of our research, as well as provide a unified point from which to disseminate it.

The Library also liaises with the publishing community on our behalf, since this is the traditional publication route for many in academia, whether it be for a journal article, conference proceeding or book/chapter. Academic publishers therefore also want us to interact with them, submit our papers and pay for our work to be published in their journals. Naturally, they also want their journals to be read, so I have to tell my Librarian the ones I want to read, so they can then negotiate the right price, and so on.

Institutional repositories are not the only ones we need to deal with. In many subject areas, notably so in biomedical science, there are other repositories that also want my outputs (e.g. EuropePMC, Dryad, ResearchFish). This list seems to grow on a regular basis without any apparent pruning.

Then there is everybody else; a long list of other stakeholders, arguably equally important, who also want to engage with me with respect to publications and outputs: undergraduate and postgraduate students, funding agencies, national and international colleagues, branches of the media, science networking tools, the rest of
science, etc. The modern academic is supposed to do all of these things, generate impact, publish papers for the REF, and even possibly spare five minutes to worry about enhancing their own CV for promotion prior to asking the dean for a pay rise – and all this on top of the OA agenda.

In summary, the above represent most of the perceived pressures and concerns that are impacting on modern researchers today. These need to be taken into consideration when analysing the attitudes and behaviour of academics to the publication of their research.

The funding-publishing cycle

The pressures to publish are matched by the pressures to obtain funding. This is an area that has received much recent media attention, with reports suggesting that income target-setting is becoming widespread in the UK university sector. This is therefore relevant to the current topic as the funding-publishing cycle sustains most academic careers and, naturally, the grant-funding process is informed by outputs, papers, journal articles and the like. These are the ‘inputs’ to this process, which are vital if you want to stay abreast of the literature and to write a well-informed, competitive grant. Idealistically, one can consider the process like this: building on the literature and your own prior work, you have a great idea, you write it up, you get some money, you publish your results in the form of more papers, which feeds back into the cycle and you then go back around the loop. Although this is one paradigm for research funding, publication, literature searching and scholarly communication, it is not the only model. A lot of academic research is not directly funded by grants, but through other routes, such as self-funded or charity-funded students, the occasional piece of scholarly endeavour, and serendipitous outputs from informal collaborations. Indeed, one can argue that this melting pot of ideas and creativity is a real strength of the UK research sector and it should not be assumed that all research outputs of worth or value originate from a research grant with a funding code.

Regardless of the origin of the funding, searching the literature remains key to successful research, and the OA movement has helped the community in this respect. However, my personal experience in the biomedical literature is that most researchers are still somewhat naive about the repository landscape – hardly a surprise considering the speed at which it evolves. So where do you go to find the key papers and outputs to inform your grants and papers given the wide choice of tools available? This ‘confusion’ of options can lead to a certain natural conservatism, which results in researchers sticking with their tried and tested choices. For example, in biomedical science, PubMed has been around a long time and is familiar to most biologists. It contains the vast majority of all abstracts and metadata that have been published, currently representing over 24 million articles. However, linked to this, researchers at the National Institutes of Health (NIH) in the US have generated the companion full-text version of this, PubMed Central (PMC). This in turn has led to a number of federated versions, via PMC International, including EuropePMC. Outside these, bioscientists frequently use Google and, increasingly, Google Scholar, whilst some prefer Web of Knowledge/Science from Thomson Reuters. Other commercial tools exist that build largely on the same content, such as Quertle, adding to the mix. In fact, this list is just the tip of the iceberg, as highlighted recently by Lu.

So what do researchers prefer to use? Although many people will use several of these, my strong suspicion is that most rely on just one: PubMed. There are several reasons for this, in my opinion. Firstly, habit. Even though I am a EuropePMC advocate, I frequently find myself going back to PubMed as my first port of call, particularly for simple searches such as for a single author. Secondly, I think many academics are probably ignorant of the huge array of alternatives, including PMC and its cousins, equating to a general lack of awareness. Thirdly, mastering a new tool along with its idiosyncrasies requires an investment of time and effort, or at least this is the perception from non-adopters. There is always an inertia to overcome to persuade people to switch. The fourth issue relates to content. Many colleagues know their favourite journals are indexed by PubMed and trust they can therefore be ‘found’, whereas some residual uncertainty may exist with other tools. Finally, the look and feel of a trusted browser-driven tool offers security; once people get used to a search tool and a way of working, they will stay loyal.
Full text: you can have your cake and eat it

The PubMed Central repositories have built on the success of PubMed and also support open access. The crucial difference, however, perhaps underappreciated by many in the biomedical community, is that they also support full-text versions of many of the articles in PubMed. For example, EuropePMC (formerly UKPMC, and built on PMC as part of a network of PMC International [PMCI]11), contains over 30 million abstracts (including the almost 25 million from PubMed) of which 3.3 million are also available as full-text articles. It also supports additional tools that for more advanced text-mining approaches, such as EvidenceFinder12, offer novel ways of interacting with the literature.

Somewhat surprisingly, there does not yet appear to be much published research on the advantages of searching full-text articles, although the research I have found is generally very positive13. Unsurprisingly, the research suggests that searching full text is more likely to find relevant articles than searching only abstracts. Clearly, this is an area that will grow in significance with the pervasive nature of open access and a community that by and large supports it14. Researchers can see the benefits of making all their papers available via gold OA, so not only can they be read for free but, importantly, they can also be computationally indexed and searched. In theory, this ought to herald a new age of discovery, and support advanced research and science generally.

The reporting burden and competition

Obviously, researchers recognize the need to publish free-to-access outputs, a view shared by all our stakeholders, including funders and governments, who are keen to monitor their investments. This therefore requires a system that captures and collates the data. Ideally, this would be a simple, one-stop shop that automatically captures outputs. Alas, this is not currently the case, for a variety of reasons. So what is out there at present for reporting outputs and linking them together?

Recent developments include ResearchGate, offering a scientific social-networking site, encouraging users to validate the papers it spots and upload OA versions. It supports lots of networking tools and discussion and has become quite popular. In parallel, a service that is gaining quite a lot of currency is ORCID15, which allows researchers to distinguish themselves via a unique, digital identifier and then populate their record with outputs and information. This is clearly a really useful advance and also links out to other resources including ResearcherID, Scopus and EuropePMC (to name a few), which makes it easy to link all your articles to your ORCID identifier (ID) in just a few clicks.

Researchers also have to link their papers to their funding sources. Traditionally, this is by virtue of the acknowledgements section of an article, but we are also asked to do a little more these days. EuropePMC contains one such tool to do this via EuropePMC Plus, though ResearchFish is the favoured site for RCUK-funded scientists, who are now obliged to use this as the central point to report on all of their outputs. It supports the alignment between grants and all forms of outputs including impact, as well as information on staff, their movements, whereabouts and more. This is a serious, well-intentioned audit of activities, but feels quite painful to complete and, one suspects, fails to capture all the information it truly seeks. So, although the phasing out of more formal grant reporting appears to be dead16, this replacement still has work to do.

The reporting demands of academics are also internal. Our IRs want the same information, as well as commercial and national/international repositories, and this has come to feel like a real chore for academics. We are asking a lot of people in the research community for the same information again and again; it feels like there are too many players in the reporting space. Although we are often ably supported by colleagues in our research offices and libraries, in my experience, it is more often than not the academic who ends up doing this – sadly, with tools that are not actually currently fit for purpose.

Open access alchemy: green into gold

From an academic perspective, the OA space is still quite confusing. Academics do not fully understand the difference between green and gold, and they probably do not understand the difference between all the Creative Commons licence codes17. (Do you know the difference between

‘Ideally, this would be a simple, one-stop shop that automatically captures outputs’

‘From an academic perspective, the OA space is still quite confusing’
CC BY and CC BY-NC-ND?) Some still suspect that open access poses a threat to journals that subsidize learned societies. Many are also confused about hybrid journals and whether they are a good or a bad thing, who pays, and the usual canards to do with double dipping. Funders also request their fundees publish OA, though policies are generally not uniform and hence require a detailed knowledge of funders’ mandates.

Most academics will also recognize another pressure, from governmental bodies, in the form of the REF. For outputs to be considered in the next REF they must be OA, at least green, from April 2016 onwards. Incentives to publish OA are therefore very strong for the academic community and I sense that in the sciences, the moral argument is mostly won. Like everyone else, academics are tax payers and want to be morally right, so we are very happy to see our tax pounds, dollars and euros spent to enable our articles to be freely accessed. Interestingly, hot on the heels of OA publications is OA data; and although I am personally a big fan of this as a computational biologist, this is arguably technologically an even greater challenge that we need to simplify for the academic community.

So taking as read the ‘Eleventh Commandment’: thou shalt publish in OA journals, the community is pretty sympathetic to this and happy to sign up. So why is there an apparent gap? Ben Johnson from the Higher Education Funding Council for England (HEFCE) reported in his talk at the UKSG 2014 One-Day Conference that although theory tells us that 96 per cent of outputs could be published OA, the actual level achieved might be closer to 12 per cent. So why is there such a big gap? To paraphrase a Shakespearean idiom, I characterize this as ‘many a slip ‘twixt cup and lip’. One of the key stages in the academic publishing pipeline is when the paper is accepted by the journal and the academics are informed. Hooray! Triples all round! But wait a minute. There is still some bureaucracy. Click on to the publisher’s website and fill out a few forms, make sure the appropriate copyright transfer is done, and … what’s this? Oh yes, you have to think about OA as well and make a decision about it now. So to OA or not to OA, go green or gold? For me this is a potential challenge. Some publishers are making this easier and I welcome this, as it is now much easier to capture an OA paper at this stage if the academic (and it is often the academic driving this process) can select their institution or funder from a drop-down list. This can work well, as most academics are well aware of the RCUK block payment for OA made to individual institutions that will cover the costs. I suspect some publishers have not yet had the time or resources to grasp this nettle. And this is a key point. If, as a community, you want us to be OA compliant, make it easier, please, and make it easier to be ‘gold’.

There are still further challenges for the green route submitters, aiming to meet the regulatory requirements but perhaps unable to be fully gold compliant. I have personally fallen foul here and failed to be fully OA compliant, particularly with older manuscripts. It may seem obvious, but for self-archiving through the green route, you need to know where your manuscript is. In principle, this should be easy. But there are complications. Which version should you submit? Is it the preprint, the postprint, the pre-postprint, the post-preprint, etc? Do I have to do this now or later, whilst recognizing the publisher’s embargo periods? Costs can also be an issue. Who pays? As noted above, this problem is starting to be solved. However, university OA funds are not bottomless pits and the key decision-making stage for an academic might involve being faced with a web page which requests credit card details on it against which article processing charges (APCs) can be levied. This might persuade someone to opt for green rather than gold if there is no simple option short of putting their own personal details in.

A further compliance complication that I have personal experience of relates to the nature of modern science; it is often highly multidisciplinary, which implies there are also multi-author papers. My CV has many of these. As a co-author, but perhaps the principal one at my home institution, I am asked to deposit the paper in one of the various OA repositories. However, I do not actually possess the accepted version of that manuscript, or a green-compatible pre/postprint. In some cases, this is because during the revision process small edits were made and not communicated to all authors, which in some cases is entirely reasonable. I can always pick up the final PDF off the publisher’s website, right? Yes, of course, but I cannot submit that to the repository owing to publisher copyright restrictions. Similarly, sometimes I do not have the figures to hand. In principle, these ought not to be problems either since the lead author can be e-mailed and can provide the files, or indeed they might also have submitted the article to an OA repository themselves, from which I can download the material. However, although all this is true, it does place additional and tedious barriers to the busy academic at the time of action when deposition is requested. So although the goodwill is there, the mechanisms to facilitate compliance are
arguably just not yet slick and easy enough, some crucial files might be missing, and jobs like this can slip off the ‘to do’ list fairly quickly. Capturing the articles at the point of publication is therefore rather important.

**Promoting your wares**

Another new frontier facing academics is the task of generating and quantifying the impact of our research. Notwithstanding the debate about what impact actually is (nor how we measure it), this is clearly still a new area for many academics who are learning about the joys of ‘pathways to impact’ statements, along with the additional publication and funding challenges. Though most in academia recognize that we are judged by our outputs and metrics such as impact factors, citations and h-indices, most are sceptical of their value. However, we are being asked as a community to do better in generating impact from our publications. What does this imply? Well, one avenue is epitomized by the growing altmetrics sector, involving social media, tweeting and alternative sources of reporting external to the journal articles themselves. This should increase the exposure of our science, boost citations and drive our outputs further up the impact altmetric ‘charts’. Although academics are not set against this (indeed many embrace it wholeheartedly), it is yet another job to add to the list.

Some recent studies by altmetric evangelist Jason Priem, founder of ImpactStory, have attempted to quantify the academic and scholarly use of Twitter. Around one in 40 scholars are currently active on Twitter, which I suspect is a modest underestimate, though use is clearly growing. However, as a community, academics appear to be slow to buy into the new social media as a medium for scholarly communication. It is fair to say that the jury is clearly still out on altmetrics as a truly valid means to judge impact and more work is obviously needed to convince academics of its true value.

**Searching the literature: how and why**

Returning to the theme of literature searching, I offer a few observations on the general habits of an academic’s search strategy and the motivation behind it. Why do academics search the literature? Often to advance their own personal research, or when reviewing a paper or a grant, when the author’s previous work is of relevance. Another common use case is the student writing a review or literature survey as part of a thesis or dissertation. One of the most common searches I conduct on EuropePMC or PubMed is to search for myself: ‘Hubbard SJ’. I suspect many other academics are similar, and certainly a basic author search is the default on most literature repositories. Frequently, it is to generate a publication list for a grant or CV, to recall details about their own papers, or check their own citations. This may seem like vanity, but it is actually rather useful because citation data is a handy way to track who is reading and citing your own work (such as via the convenient click-through links provided by Google Scholar on your profile page), and it is also a really nice way to browse the literature. If somebody has cited my work, there is a fair bet it is linked to my own research interests and is therefore a very neat way of keeping up with what my peers are doing. I call this ‘citation serendipity’; it is a great way to find out what is going on, who has built on what you have done, and how your field is developing. Some tools, such as ResearchGate, offer updates every time a paper is cited.

**A four-point plan for improving OA compliance?**

So, as promised, here is a four-point plan for discussion. This is not meant to be a complete plan, rather four observations from an academic perspective that need to be factored into any concerted efforts to improve the OA space.

1. **Take academics out of the loop.** This might sound odd, but what I suggest here is that if you want to get open access working and get scholarly communication streamlined, you need to take us out of the loop, i.e. prevent academics from inhibiting the smooth progress of accepted papers from journals into repositories. Once a paper is accepted for publication, if certain criteria are met, it ought to go straight into an OA repository and become immediately available. This might mean resolving financial issues at the submission stage and removing the bureaucracy, but this seems to be manifestly possible for some publishers. In this regard, the gold route seems cleaner and simpler. Green implies an embargo and the possibility that the academic can slow it down, mess it up, or otherwise impede the process. In short, improve the communications between journals, institutions and repositories.
2. Improve the tools. The current state-of-the-art tools for depositing articles, searching and indexing, and linking inputs (grants, funding sources) to outputs (papers, proceedings, etc.) is not yet good enough. Rather than place additional burdens on the users, we ought to be able to generate data standards and automated text-mining tools that take the pain out of this process. This is starting to happen and I have seen useful examples on various sites, but unfortunately less so for the mandated repositories that academics are being compelled to use by funders. Scraping funding codes, author names and IDs, article metadata and DOIs from publications and the like ought to be a solvable problem. Community-driven data standards for reporting and dissemination are another way to improve the situation, as has been shown in many areas of my own field, bioinformatics. Ideally, all we would need to do is sanity-check the auto-generated records, possibly add a few items to fix any omissions, and then press ‘go’.

3. Ask for things once. A recurrent problem for the academic at present, at least in perception, is being asked repeatedly for the same information relating to our funding, outputs and impact from multiple stakeholders, probably all in slightly different but subtly incompatible formats. As you can imagine, this is not appreciated and patience is wearing thin. Data standards, as noted above, will clearly help improve things, which will lead to better data exchange models between funders, libraries, journals, IRs, etc. Couple this to simple, occasional auto-reminders (e-mail is fine by me) and unify and simplify the message.

4. Go for gold. Aim for near 100 per cent compliance on gold OA across disciplines and journals. Although I suspect the opinion that gold is good is not one shared throughout all of academia, we could and should be aiming for it. HEFCE’s estimate of a 96 per cent theoretical compliance rate appears to be an entirely reasonable target to aim for. It will simplify the process, remove ambiguity from deposition and reporting and, if properly communicated, I do not think you will find too much resistance in most of the academic community.

Conclusion

More needs to be done to facilitate OA deposition and compliance, by all parties involved, but asking the academics to do much more is unlikely to be very productive. However, since this seems to be a technologically solvable problem, the future is bright, the future is … golden.

Competing interests
The author has declared no competing interests.

References
6. The University of Manchester Library: http://www.library.manchester.ac.uk/ (accessed 19 August 2015).


15. ORCID: http://orcid.org/


17. Creative Commons, About The Licenses: http://creativecommons.org/licenses/ (accessed 2 September 2015).


20. See the interesting reaction on the blogosphere to HEFCE’s review of this area by Wilsdon: http://www.hefce.ac.uk/rsrch/metrics/ (accessed 19 August 2015).