

Introduction

The Royal Society of Edinburgh Young Academy of Scotland working group on Open Data is pleased to respond to the open letter of 25/02/2013 from David Sweeney, Higher Education Funding Council for England (HEFCE), requesting advice to help shape the consultation on Open Access and submissions to the Research Excellence Framework (REF) post-2014.

Open Access to Research

On 16th July 2012, the UK Government announced that it had accepted the recommendations of the Working Group on Expanding Access to Published Research Findings chaired by Dame Janet Finch (herein referred to as the Finch report). HEFCE indicates preference for 'gold' open access publishing at least in the long term, where publication costs are borne by researchers (ultimately financed by research funders) rather than being paid by the readership/libraries (HEFCE letter, paragraph 9). This approach aims to ensure that publications are accessible to all and as soon as the work is published. While the Finch report supports the gold model, it also notes that "over the next few years, no single channel can on its own maximise access to research publications for the greatest number of people" and also supports "hybrid journals" (p7) - thus including the 'green' model where publication costs are borne by the readership and articles made available from a public repository, typically after an embargo period. For example, some researchers might not easily obtain open access fees (e.g. within companies, emeritus researchers).

It is difficult, though not impossible, to argue that publicly funded research should not be made available to all. Caveats include areas where security or economic interests conflict with openness; for example, knowledge relevant to making nuclear warheads or bioterrorism raises serious concerns [e.g. Science 2009;324:1499, Nature 2012;485:5]. There is also a public interest in protection of intellectual property (IP) generated from taxpayer funded research, to support economic growth and because a significant proportion of revenue typically benefits public institutions. Nevertheless, open access publishing energises research (both public and private) [Col. Res. Libr. 2004;65:372], may help to increase public engagement with research, and ensures that information critical to public policy decisions is available to all. Furthermore, there are economic, scientific and ethical drawbacks to the current IP system [Prometheus 2011;29:325].

Expectation for open access publications in post-2014 REF

We believe that the expectations set out in paragraph 11 of the HEFCE letter are broadly reasonable, with the caveat that making research available retrospectively may need to be admissible if the notice period referred to in paragraph 21 is not long enough to encompass the long lead-times for Humanities and Social Sciences publications. Additionally, we are concerned that the cost of open access publishing will impact on funding for the research work itself - unless additional resources are made available for this. Indeed, "In the period of transition there are bound to be additional costs as both systems exist side by side" (Finch report). We note that the high-cost transitional period is likely to be extended given that USA are putting gold access 'on hold' [<http://www.nature.com/news/gold-on-hold-1.12490>].

Economics of openness and research quality

The economic implications of open access warrant careful consideration; publication charges offer a new route for immediate economic gain by publishers. The RSE Young Academy of Scotland includes members from a wide variety of disciplines. Across disciplines, we have experienced increased correspondence over the last few years from open access publishers inviting us to publish in their journals, for a fee, which has also been noted by others [N Engl J Med 2013; 368:791-793]. Gold open access provides economic incentive for publishers to increase the number of articles published, which may impact on quality. A related concern with the gold model is that journals may drive down publishing costs at the expense of scrutiny on published research, potentially leading to less stringent editorial control and/or peer review [N Engl J Med 2013; 368:791-793]. On the other hand, journals that publish lower quality research may fizzle out in the longer term. These considerations lend weight to the 'green' model, which does not suffer from the above issues; indeed repositories appear to provide an attractive, stable point of access. However, the green model could lead to loss of revenue and eventual closure of many journals if the embargo period were not long enough. Indeed, the Finch report states "Where an appropriate level of dedicated funding is not provided to meet the costs of open access publishing, we believe that it would be unreasonable to require embargo periods of less than twelve months." We note that some journals currently operate embargo period as little as 6 months and even provide a repository (e.g. http://www.nature.com/authors/author_resources/deposition.html), energised by policies of research funders in the UK (e.g. BBSRC, Wellcome Trust, MRC) and abroad (e.g. NIH, CIH). Thus, while the gold model has the attraction of immediate access to all, we believe that it should be adopted alongside the green model - which is already common practise in some disciplines [<http://uk.arxiv.org> , <http://www.ncbi.nlm.nih.gov/pmc>]. Indeed, we urge caution to ensure research quality does not suffer, in line with the Finch report's guiding principles "Neither the quality and standing of the UK research community nor the underpinning of high-quality publishing channels must be put at risk" (p17).

Furthermore, while welcoming HEFCE's assurance that 'research in all subjects has equal importance', we also anticipate that a 'one-size-fits-all' approach may impact on research quality due to fundamental differences between current practice across disciplines that necessitate different models. For example we welcome the recognition that the research communication ecosystem around monographs differs to that of journal articles - this is further explored in the sections 'Exceptions and appropriate notice period' and 'Open access for non-journal outputs' below.

Research Councils UK announced on 16th July 2012 that they would provide support for Article Processing Costs (APCs) associated with gold access publishing . However, roughly 40% of researchers are not directly supported by grants [<http://tinyurl.com/40pcOA>], and in certain disciplines, the percentage is far higher. In one respect research funded by grants will have undergone considerable peer review, which may include modifications to methodology. On the other hand research funding is extremely competitive with many unsuccessful submissions rated as fundable by reviewers. If the researchers go on to conduct this research, or part of, through other means e.g. through a charity grant or the involvement of a PhD student, they might not have sufficient funding to publish. These issues are particularly relevant during a transitional phase where institutions would need to cover charges for both publisher subscriptions as well as open access. There are also additional inherent costs in any model where material is archived in both journal and institutional repositories.

Further, in the Humanities and Social Sciences, it is often possible to undertake substantive, high-quality research at low cost. In some of these disciplines, those undertaking grant-funded research are a minority; the majority are financed via their salaries, and whatever support (whether library or IT infrastructure, or research expenses grants) their institutions can provide. How will the cost of the gold model be financed for these researchers? More profoundly, publishing in the gold model might lead to a shift in the nature of the research being undertaken, and the kinds of outlets it is disseminated through. Planning for costs to the universities (ie APCs) to be ultimately balanced by lower journal subscription fees may flounder where journals adopt a 'hybrid' mixture of green and gold articles; for example in international journals where UK authors are in a minority. It seems unlikely that gold open access will be cheaper for universities in the near future.

Repository use and techniques for institutional repositories to cross-refer to subject and other repositories

We believe that a requirement for all HEIs to establish and maintain an institutional repository for research by their staff is broadly reasonable. However, in some areas, repositories are already widely used (e.g. arXiv [<http://uk.arxiv.org>], PubMed central [<http://www.ncbi.nlm.nih.gov/pmc>], <http://europepmc.org/>) and therefore institutional repositories would be secondary. We note that COPAC (copac.ac.uk) has brought together the library records of 70 UK HEIs into a single searchable site, and has overcome the different structures and formats of the individual institutional records. Indeed, if HEIs were required to ensure that the contents of their electronic repositories were fully and completely included in their library catalogues, and if COPAC were extended to all UK HEIs, then COPAC itself could become a UK research portal. However, for those who are searching more broadly on a general topic, seeking papers of whose existence they do not yet know, it will be essential that all UK research should be fully indexed in all the main search engines, academic and otherwise.

Of course, open access to research implies that material may be found by researchers all over the world. An alternative, possibly more economic, solution would be to negotiate international support for repositories along subject lines - such as those already in operation, noted above.

Exceptions and appropriate notice period

In addition to the problem of reproducing works of art (HEFCE letter, paragraph 20), there may be similar legal issues with the reproduction (either as an image or in transcription) of hitherto unpublished archival materials (e.g. historic correspondence, ledgers and diaries). Rare books libraries and archives may be willing to change the standard terms of their copyright permission forms to allow use in open access (OA) publications. However, in both the above cases existence of copyright protection over small sections of the work need not prevent that the bulk of the research from being OA. Publishers might provide a version of the published paper with the images or transcriptions blanked out (like censored mail), if the rights holders did not agree to their inclusion in an OA format - although this does not seem ideal and could be a useful question to examine in the consultation planned for later this year.

Regarding reasonable notice (HEFCE letter, paragraph 21), lead times in the arts and humanities are typically longer than in STEM subjects, even for journal articles. It can take over a year, from the point of acceptance, for a humanities journal article to appear; and two years for a monograph to appear. The proposed notice period should be at least two and a half years.

Open access for non-journal outputs (e.g. monographs)

The current debate about open access (OA) has largely been expressed in terms of journal articles, which are typically available electronically already (though not necessarily freely). It is clear that moving towards a full OA policy for journal articles is possible in the near future. However, it is well-recognised that different disciplines have different publishing patterns, and not all use journal articles as extensively. HEFCE's statement acknowledges the issue of monograph publications; monographs are not the only non-journal format in which written research outputs appear. In History, English, and other humanities disciplines, researchers may plausibly have more articles in 'edited volumes' (books containing essays by a variety of authors) than in journals. And even 'monographs' come in a range of forms, from academic to textbook to trade. Confining the discussion to 'conventional monograph publication' risks overlooking the full complexity of non-journal publishing.

Academic book publishers are already struggling with the relationship between ebooks and printed books. Unlike journal articles, academic monographs are not necessarily electronic; to make them so, let alone freely, will take a significant culture shift among publishers. It will also mean shifts in how many scholars read and work, which may not necessarily be advantageous. The equivalent of an article processing charge (APC) for monographs would compensate the publisher for the lack of revenue from the ebook. Yet, with so much humanities and Social Sciences research funded by HEIs not funding agencies, who is going to pay? Green OA, with an author's final version in a repository, appears to be a more achievable aim.

Once we move beyond academic monographs to textbooks and trade books, further issues arise. It is

difficult to imagine publishers willingly giving up income from electronic editions of either of these categories. Textbooks may be of less relevance to future REF cycles; but trade books will be. In English, History and similar disciplines, academically-respectable research can be published as a trade book, and there is congratulation if a researcher's book is issued as trade - for it will thus be able to fulfil the desire to communicate research to wider audiences. An academically-respectable trade book should be eligible for future REF exercises; but it seems highly unlikely that it would be OA. Insisting on OA for all book publications could therefore work against HEFCE's aim to promote the public understanding of research, at least in Humanities and Social Sciences (HSS).

The suggestion (HEFCE letter, paragraph 18c) of allowing institutions to submit up to 20% of non-OA outputs does not seem appropriate to deal with the challenges posed by non-journal publications. We are concerned that there may be issues of rationing and gatekeeping within institutions as is already raised by the limited funds available within institutions for paying APCs. In HEIs with strong HSS faculties, 20% may not be enough; and it is not difficult to imagine that the HSS faculties could 'swallow' the entire 20% of an institution's allowance, leaving no possibility for non-standard outputs from other units in the institution.

It appears that there are significant challenges to overcome with implementing OA for non-journal research outputs such that we are concerned that this may be difficult to include in the next REF cycle.

Requirement for open data

We are pleased that HEFCE support the principle of open sharing of raw data and note paragraph 24 of the letter states that they "...do not consider it feasible at present to make this a formal requirement". We would welcome as minimum a formal requirement for data sharing where an appropriate, stable public repository (or better, multiple repositories) can be identified such as ArrayExpress and GEO [www.ebi.ac.uk/arrayexpress, www.ncbi.nlm.nih.gov/geo/]. A model where research funders both require and are involved in policing of data sharing seems useful - subject to appropriate exemption guidelines (e.g. on ethical grounds). For example, UK-DA archives data from approximately 70% of ESRC funded projects [ESRC *pers. communication*]. We also recognise that in certain cases there would be important restrictions on the nature and extent of data-sharing. Additionally, open data sharing would be energised by further development of automated systems to facilitate assessment of compliance with data sharing requirements (e.g. for peer reviewers), including appraisal of provenance information, which is fundamental for effective reuse of data.

Models where individuals grant access to data prior to publication can raise concerns due to the current bibliometrics-driven assessment of research output. Making data directly citable could help this to some extent; for example companies such as Reuters could track the use of data (e.g. accession numbers, images) in research publications. Data transfer agreements may also be appropriate - especially where there are ethical concerns about how data is used. However, in some areas data are made available as soon as practicable, with a time-limited public disclosure embargo on results based on the data. For example, the Encyclopedia of DNA Elements consortium (ENCODE, [<http://www.nature.com/encode>]) operated an embargo for 9 months from the date of data release, which precludes any disclosure of results at seminars or to electronic servers such as journal submission systems. We would also welcome more consistency in data sharing policy amongst funding bodies, for example, researchers funded by multiple sources might be conflicted by differing data sharing policies. Nevertheless, we recognise that implementation of data sharing would likely require carefully developed guidelines tailored to specific subject areas. Broadly, 'large-scale' projects involving expensive equipment and/or many collaborators seem to inherently promote a culture of data sharing; communities of this type include Astrophysics, Particle Physics, and Functional Genomics. Funding bodies are evaluated at least in part through publication of research that they have funded, as well as citation of those publications. Therefore, a consideration with policing of data sharing by funders is potential for conflict between the funding bodies' own reporting objectives and any policing framework [<http://tinyurl.com/openDataYAS>]. For example where citations to raw data replace citations to journal articles. If source data are not retained, then the value of the associated research activity is greatly diminished. Moreover, availability of raw data is a critical component in independent evaluation of research findings. Therefore we believe that archiving of data is fundamental and that archiving costs should not be levied as an argument against open data sharing. Where costs of data archiving are limiting, alternative strategies (e.g. off-line access, minimising

redundancy, access-on-demand) should be considered. We would welcome scope for further examination of issues around open data sharing in the anticipated consultation referred to in the HEFCE letter.

Public engagement and open access

Sharing information with non-specialists remains a significant challenge, largely due to the implicit knowledge gap, leading to possible misinterpretation of results - and sometimes misrepresentation of academic endeavour [e.g. tinyurl.com/3sf495t]. However, mistranslations and misunderstandings are also an issue where investigators from different disciplinary fields come together to discuss their work. This issue is further complicated by mainstream media (e.g. newspapers) seeking to present research findings in ways that are attractive and compelling to readers; of course, university staff and press offices have key roles to play in enhancing the appropriate translation of their research into the media.

The Internet means that communication is almost instant, with wide national and international coverage. For example, the British Psychological Society Press Office 'tweet' about conference presentations during or shortly after presentations, giving details of their content. Conference presentations often include research that is unpublished, unpatented, and minimally peer-reviewed, if at all. The increase in the use of blogs among academics is another area where the lack of peer review can become an issue. One prominent example is Satoshi Kanazawa's controversial Psychology Today post on "Why are Black women physically less attractive than other women?", which presented unreviewed analyses that were later shown to be highly flawed [tinyurl.com/cejz6ml]. Although subsequent work can resolve incorrect findings, the initial reporting can often cause serious damage in the meantime, as exemplified by the MMR controversy. Indeed, cases of measles and mumps have risen significantly in recent years compared with 1998 levels [tinyurl.com/5uylxdc], potentially as a direct result of media-induced fears about a link between MMR vaccines and autism. The above issues highlight emerging pitfalls, and the importance of ensuring that the status of findings and analyses are made clearly apparent to the public. Furthermore, the wide availability of studies and commentaries on research adds impetus to arguments for open access to original research.

Closing remarks

Research funders, both public and private, play key roles in setting the norms in academic culture, including attitudes and behaviours around open access publishing. Of course, funders are in a unique position to enforce and reward open access publishing. Key points to consider:

- A hybrid gold and green model has advantages and may be most appropriate even in the long term.
- The significant challenges with open access for non-journal outputs, especially in humanities, suggest that requirement for open access in this area may be difficult in the next REF cycle.
- Different disciplines require fundamentally different approaches in the transition towards full open access.
- The cost of 'gold' open access must be weighed against the impact on available research funding.
- Developing a culture of open access publishing and data sharing may require established methods of recognition to be combined with tangible rewards as well as possible penalties.
- Ethical concerns must be addressed robustly, and this is a particular concern with sharing raw data. Principles of confidentiality and anonymity sit alongside the issue of ownership and the mechanisms by which benefits are realised.
- Open data sharing should be required in the next REF cycle where an appropriate repository is already in place, with option to apply for exemption (e.g. on ethical grounds).
- Security issues should be carefully considered to minimise misuse of research due to a lack of understanding or malicious intent.
- There are concerns about greater availability of information to the public via the media, including the internet. Serious efforts will need to be made to avoid potentially harmful consequences; for example, the premature release of information relating to healthcare decisions or global climate.
- Sharing research in the widest sense remains a significant challenge, partly due to the implicit knowledge gap between specialists and non-specialists.

- Widespread and systematic training of researchers in communication and engagement seems important as a foundation to encourage greater appreciation of research studies, as well as promoting dialogue between academics and wider publics.

The RSE YAS working group members were: Ian Overton (chair), Sinead Rhodes, Aileen Fyfe, Job Thijssen, Martyn Pickersgill, Lisa DeBruine, Alan Gow and Des Balmforth.
<http://www.youngacademyofscotland.org.uk/our-work/open-data.html>

This paper does not necessarily reflect the views of all RSE Young Academy of Scotland members or the RSE.