

Open access and submission to the REF post-2014 – response from the Chartered Institute of Library and Information Professionals

The Chartered Institute of Library and Information Professionals (CILIP) welcomes the opportunity to comment and advise HEFCE on issues to inform the development of the consultation process on the role of open access publishing in the submission of outputs to the post-2014 Research Excellence Framework (REF).

CILIP is established by Royal Charter and is the professional body for library and information professionals in the UK. It has around 15,000 members working in all parts of the UK economy.

We support, as a matter of principle, measures to improve people's access to information. Many of our members have been leaders in advocating and implementing open access models, recognising the huge public benefits that come from the wider dissemination of research outputs. Equitable access to the latest research democratises knowledge, stimulates innovation and economic growth, increases research productivity and facilitates rapid knowledge transfer.

The standard model of scholarly communications whereby libraries subscribe to academic publications is unsustainable. The average cost of journal subscriptions has been rising at a rate far above inflation for decades, while budgets at academic libraries have remained fairly static. Figures from JISC estimate that UK university libraries spend around £110 million on their journal subscriptions, with subscriptions to the large journal collections accounting for up to 80% of a library's journals budget¹. Not even the

¹ JISC, 2010, The value of UK HEIs contribution to the publishing process <http://www.jisc-collections.ac.uk/News/Value-of-HEIs-to-publishing/>

most well-funded libraries can sustain this, which has led to a gap in the knowledge base. This has been a major driver in the open access movement.

Q.1: We welcome advice on our expectations for open access publications, as set out at paragraph 11.

CILIP welcomes both the intent of this consultation and the sensitivity it displays to the existing operating environment. As recognised in the Finch report, the research communications ecosystem is complex and for the foreseeable future no single mechanism will suffice to adequately and equitably expand access to knowledge. In our response to the BIS committee Inquiry into the implication of the government's preference for "gold" over "green", CILIP set out our concerns with the former model². We can see merit in hybrid solutions and support a mixed approach that enables repositories to continue to be an important part of the scholarly communications process. We therefore welcome HEFCE's proposal to accept material published via either gold or green routes, and the central role given to the repository in housing these outputs. Indeed, in addition to the valuable role repositories play in increasing sustainable and convenient public access to research, library and information professionals in several universities already recognize that the institutional repository is a good system for providing, electronically, a record of research outputs for the purposes of the REF and have been using it to meet this end for several years.

We welcome the proposal that publications should be presented in a form that allows for text-mining. Text and data mining (TDM) provides researchers with the opportunity to cover volumes of material, across sources and times, which could not be used or processed "manually". A JISC study on TDM found it could reduce "human reading time" by 80%³. The potential of TDM technology is therefore enormous, advancing research in a number of ways, including linking across data to develop new findings,

² <http://www.cilip.org.uk/get-involved/policy/responses/Documents/BIS%20Open%20Access%20Inquiry%20-%20response%20from%20CILIP.pdf>

³ Text Mining and Data Analytics in Call for Evidence Responses. UK Government <http://www.ipov.gov.uk/ipreview-doc-t.pdf>

suggesting new avenues for exploration to researchers, and sifting the very large quantities of research now being published. This will bring significant benefits for industry, citizens and governments as well as academia.

THE ROLE OF INSTITUTIONAL REPOSITORIES

Q. 2: We welcome further advice on repository use and on techniques for institutional repositories to cross refer to subject and other repositories.

The JISC funded Registry of Open Access Repositories (ROAR) site lists 250 UK university repositories⁴, many of which were created and managed by the library.

CILIP supports the proposition that, in order for an output to meet the REF open access requirement, it *has* to be accessible through a repository. At least 24 British universities have already adopted a “green mandate”. These include Durham, Birkbeck College, Bath, Nottingham, UCL and Edinburgh⁵. However, overall the spontaneous deposition of outputs by researchers is low. The newest data on this still points to a general base level of around 15%⁶. Making the deposit of publications in a repository a requirement of the REF would have a direct, immediate impact on repository use. International evidence shows that, where deposit mandates have been strengthened, compliance rates quickly exceed 70% of the total research output.⁷ This creates an invaluable resource.

It must be recognized, however, that researchers may deposit in either institutional or subject-based repositories and that in some disciplines the preference may be to deposit in subject-based repositories over the local institutional repository. This being

⁴ <http://roar.eprints.org/> [accessed 8 March 2013]

⁵ Cited in International Society for First World War Studies – *written evidence to the Science and Technology Select Committee Inquiry into the implementation of open access*:<http://www.parliament.uk/documents/lords-committees/science-technology/Openaccess/OpenAccessEvidence.pdf>

⁶ Gargouri et al, 2010 cited in Swan, A., 2011, *Institutional Repositories – now and next*

⁷ RLUK, 2013, *House of Commons Business, Innovation and Skills Committee: Inquiry into Open Access Policy. Submission by RLUK*

the case, it is important that research outputs deposited within a local institutional repository are still accessible via an appropriate subject-based repository, and vice versa, thus ensuring the widest possible impact for the research.

There is a cost to maintaining repositories. We welcome HEFCE's assertion that it intends to "support the development and use of repositories as far as possible" and look forward to more details about what this will entail as there is at present a lack of clarity over how these costs will be met.

With regard to cross-linking between both institutional and subject repositories, we wish to draw attention to the following issues and make several recommendations:

- There is considerable potential for duplication of effort, and of content. To mitigate this requires the structured implementation of persistent identifiers. Since the *Digital Object Identifier*⁸ standard has gained significant currency in recent years (albeit one with an associated cost), we would urge HEFCE to develop their plans with a strong commitment to agreeing a framework for persistent identification;
- The typical method for achieving structural interoperability is via OAI-PMH⁹, thus the use of this protocol should be considered mandatory (indeed, it is commonly used across repositories). As the updating and synchronisation of metadata content via OAI-PMH is not always consistent across platforms, we recommend that the standard application of these between repository platforms be encouraged.
- To address the risk of significant failings in semantic interoperability and discoverability we encourage HEFCE to investigate the development of a classification scheme or taxonomy, even if at a very high-level, to promote discoverability of the open access resources.

⁸ <http://www.doi.org/>

⁹ <http://www.openarchives.org/pmh/>

- The cross-referring of repositories ought to provide a mechanism for distributed digital preservation, which HEFCE ought to consider exploiting. An example of this is LOCKSS (Lots of Copies Keep Stuff Safe) methodology¹⁰. Sustaining the value of the investment for the long-term demands a clear view on digital preservation.

EMBARGOES AND LICENCES

Q.3: While we expect that sufficient clarity and reassurance on embargoes and licences will be achieved through the Research Council discussions, we welcome responses which address these issues

There is clear evidence that for publicly funded research the maximum economic and research impact is achieved when research outputs are readily available, accessible, and re-usable.

Embargoes

In the interest of making research outputs publically available, shorter and consistent, or no, embargo periods are the desired outcome. However, it is necessary to recognise that agreement has to be achieved between publishers, funding councils and libraries.

Because the longer “shelf-life” of social sciences, arts and humanities disciplines *might* mean that the business models of publishers that depend on generating revenue through journal subscriptions are adversely affected, CILIP welcomes RCUK’s flexible policy on embargo periods during the transition phase. We support the Finch Report’s recommendation that rules relating to embargoes on access to journals that are not funded in the main by APCs should be “kept under review in the light of the available evidence as to their likely impact on such journals” and RCUK’s commitment to

¹⁰ <http://www.lockss.org/>

undertake such a review. Researchers in all disciplines want access to the latest research, so policies that limit that access should be evidence based.

Licences

CILIP supports the RCUK position that, in order to maximize the opportunities for access to and reuse of content funded through APCs, research papers should to be made available using CC-BY. We also encourage liberal and enabling licences – ideally CC-BY – for material deposited under the “green” route. Crucially the CC-BY licence removes doubt about how research papers may be used, which encourages re-use and the sharing of scholarly outputs. A move to standardisation also cuts administrative costs, as libraries would have fewer variant terms and conditions to keep track of.

We acknowledge that publishers relying heavily on revenues from the sale of reprints have raised concerns about the impact of CC-BY, and we welcome RCUK’s commitment to actively monitor this impact. As with embargoes, policies on licences should be evidence based; the dissemination and re-use of publically funded material should not be restricted unless there is a solid reason for doing so.

We share the concerns raised by some stakeholders that the requirement for Gold open access to be accompanied by a CC-BY licence will lead some journal publishers to increase their Article Processing Charges (APCs). We hope that a reasonable balance in the business model is attained so that the level of APCs is controlled.

In some subject areas, especially in the humanities, research outputs include images and other elements that are reproduced under permission from a copyright holder. We welcome the commitment of RCUK and JISC to produce guidance on the identification and management of third party copyrighted material included within a paper licensed under CC-BY.

Research not funded by the research councils

It should be recognised that in some fields, notably the arts, humanities and social sciences, a large amount of research takes place that is included in the REF for which funding is outside of the funding councils. Research council guidance on embargoes and licences will therefore not address these outputs.

Where APCs are applied it may be difficult for researchers to access these funds within their own institutions as they are not in receipt of external funding. If these researchers are to be subject to an open access mandate for inclusion in the REF some sort of accommodation needs to be made that ensures that they are supported.

EXCEPTIONS

Q.4: We welcome advice on the best approach to exceptions and on an appropriate notice period. Any cases made for exceptions should be underpinned by clear evidence.

No comment.

MONOGRAPH PUBLICATIONS

Q. 5: We seek comment on when it may be thought inappropriate to expect repository deposit of monograph text. Alternatively, given the percentage of submitted material which is in monograph form, we ask for advice on whether an expectation of a given percentage of compliance would eliminate the need for a special-case exception for monographs.

CILIP shares concerns about the difficulties authors and publishers face in developing a secure future for monographs and we welcome the recognition given to this by HEFCE, together with the spirit of their proposal with regards to the repository deposit of

monograph text. However, we anticipate an increase in the number of open access monographs post 2014 REF as the publishing landscape evolves, so any special provision for monographs needs to be supported by evidence. We therefore call for HEFCE and others to do more to establish the evidence justifying any exemption.

OPEN DATA

Q.6: We invite comment on whether respondents feel this is the appropriate approach or whether they feel that sufficient progress has in fact been made to implement a requirement for open data as well. We will consider any representations that such a requirement may reasonably now be developed but would also need advice on how this might be achieved.

CILIP welcomes the consideration being given to the issues involved in increasing access to research data. There is no significant field of research that is not already rich in digital data. While the experimental sciences are particularly data intensive, social sciences and humanities enquiry can also generate large volumes of data¹¹. Many of the arguments for open access to publically funded research literature apply equally well to publically funded research data. The UK government is committed to open data.

Digital data can be mined and re-used to create new knowledge. McKinsey Global Institute reported in 2011 that effective use of “big data” in the US healthcare sector could be worth more than US\$300 billion a year, two-thirds of which would be in the form of a reduction in national health care expenditure of about 8%. In Europe, the same report estimated that government expenditure could be reduced by €100 billion a year¹². TDM has already enabled new medical discoveries through linking existing

¹¹ Swan, A., 2011. *Institutional Repositories – now and next*

¹² McKinsey Global Institute. 2011. *Big data: The next frontier for innovation, competition, and productivity*.

http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_f_or_innovation Cited in open letter to Commissioners Barnier, Geoghegan-Quinn, Kroes and Vassiliou, February 2013

drugs with new medical applications, and uncovering previously unsuspected linkages between proteins, genes, pathways and diseases¹³.

Publishers are increasingly requiring original datasets to accompany an article, especially in disciplines such as genomics, and some subjects, such as bioscience, have a long history of centralised, professionally-managed open data repositories (e.g. GenBank¹⁴)¹⁵. However, these facilities care for larger datasets – so called “big data” - “small” and “medium” data remains vulnerable and action is needed to ensure its archiving and preservation and to aid its discovery and re-use. Institutional repositories are a good candidate for housing these small and medium datasets.

However, CILIP welcomes the decision to delay making mandatory the provision of access to data pertinent to those publications submitted in a future REF. Datasets come in a wide variety of formats, sizes and complexities and the challenge of managing – and finding adequate space to store – these datasets has still to be addressed. In addition to the technical challenges, the legal, ethical, economic and competitor issues surrounding access to research data are complex. CILIP suggests that HEFCE take small incremental steps towards their broader ambition of increasing access to research data.

CILIP would like to see an increase in the number of data librarians employed to manage datasets. These professionals will have “an understanding of the provenance of research data, how they were created, how they can be manipulated and what must be done to store and preserve them properly”, and will be able to advise researchers on complying with data policies and requirements from research funders and from institutions.¹⁶

¹³Open letter to Commissioners Barnier, Geoghegan-Quinn, Kroes and Vassiliou, February 2013

¹⁴<http://www.ncbi.nlm.nih.gov/>

¹⁵Swan, A., 2011. *Institutional Repositories – now and next*

¹⁶Swan, A., 2011. *Institutional Repositories – now and next*