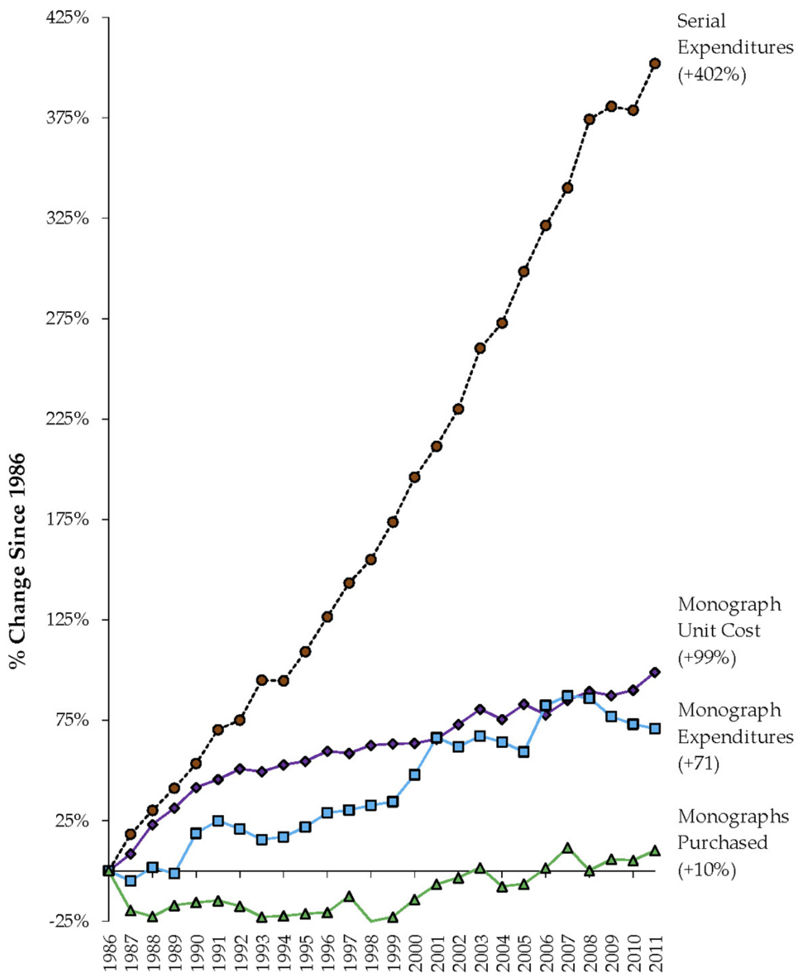
Key Issue
Out of the shadows: the case for a national repository of open content

Looking back on the history of the open access (OA) movement in the UK from a library perspective, we can see that it has been a journey towards a not fully clear destination, motivated differently at different times, and yet nonetheless a journey undertaken with seriousness and a determination to see change. Working in the scholarly communications field back at the time of the Budapest Initiative in 2002, it seemed to me and to colleagues in other research university libraries throughout the UK that we had found the solution to the besetting problem of our new digital age, the accelerating rise in the cost of journals, or the 'serials pricing crisis'. Librarians on both sides of the Atlantic soon became very familiar with early versions of this graph (Figure 1), produced by the Association of Research Libraries. It clearly shows the price of serials rising frighteningly and unsustainably, illustrating both apparent profiteering by the publishers concerned, and the prospect for libraries of having to make cancellations on a large scale.

JOHN MACCOLL

University Librarian and
Director of Library Services
University of St Andrews



NOTE: Data for monograph and serials expenditures was not collected in 2011-12.

Source: *ARL Statistics 2010-11* Association of Research Libraries, Washington, D.C.

Figure 1. Monograph and serial costs in ARL Libraries, 1986-2011 (including e-resources from 1999–2011)

The year prior to the Budapest Initiative had seen the publication of Jean-Claude Guédon's masterly essay, In Oldenburg's Long Shadow[1], which had uncovered the mechanism by which smart publishing companies had begun to dominate the marketplace and drive up the prices to the point of crisis. Guédon talked of a 'quality pump' which had been 'primed' to ensure that particular titles were emerging as the highest-impact journals in their field, thereby attracting more citations to their papers, thus retaining their top spots and so turning into cash cows for their publishers. At around the same time we had witnessed the arrival of the 'big deal', whereby publishers made title-by-title sales significantly more expensive than offerings of journal bundles, and with one (master) stroke thereby reduced their own costs and increased their income. Any publisher that could command a series of high-impact journals across even a few disciplines could ensure that no self-respecting academic institution, via its library, could avoid a subscription to the bundle which contained them. If that publisher was a commercial enterprise, responsible for maximizing profit to its shareholders, then – until something radical happened to the market – libraries would be condemned to paying hefty above-inflation prices year by year for the journals concerned. And this has sadly been true, and remains the case more than a decade later.

"… open access has regained the spotlight as the virtuous and ethical way to publish scholarly outputs."

But hope was on the horizon, in the shape of work that had been done initially within the high-energy physics community in the US, at the Los Alamos National Laboratory, the original home of ArXiv, a 'preprints server' for the scholarly publication needs of high-energy physicists and an increasing range of researchers in cognate fields. Here was a new model of scholarly publication, which prized the velocity of publication above the 'authorized version' that would eventually be recorded in the journal literature. It gave rise to the view that an alternative scholarly publishing model could be created, putting together the ownership of the scholarly publishing process by academic authors (they write the papers, referee the papers, and edit the journals) with the instantaneous access afforded by the internet. What was needed was a means of allowing interoperability, so that any number of preprint servers of the ArXiv type could interwork in a way that would eventually make the publisher-controlled system of very expensive journal publication redundant. Thus we saw the Open Archives Initiative, and the arrival of the Dublin Core metadata standard – the means by which the academy, via its libraries, would seize the means of control from the publishers. Of course there would still be some cost, but it would be much lower, and justifiable. We even speculated on what we would do with the massive 'windfall savings' we would make in our serials purchasing budgets.

Fast-forward to the last two or three years, and we see the case for open access, which had lost some of its collective dynamism by the end of the decade, return in the service of a new focus upon the costs of research, not this time driven by librarians with their serials pricing crisis graphs, but by research funders – supported by the UK government – seeking to make publicly accessible the outputs paid for by taxpayers. Led initially by Research Councils UK (RCUK), in the report of a working party chaired by Dame Janet Finch[2], and now more recently by the UK's Higher Education Funding Councils as a new policy within their system for allocating research funds to universities on the basis of their research strength profiles (the Research Excellence Framework, or REF)[3], open access has regained the spotlight as the virtuous and ethical way to publish scholarly outputs. Whereas back at the time of the Budapest Initiative what was envisaged was a revolutionary and subversive reform of scholarly publishing, this time around an attempt is being made to develop a system which works in a complementary way to the publisher-controlled journal publication system.

"Publishers hold the trump card of peer review …"

Publishers hold the trump card of peer review, and the value of this card has strengthened over the past ten years. We did think at one time that we might take peer review into the academy by translating whole editorial boards away from unacceptably priced journals, and setting up rivals. The Scholarly Publishing and Access Resources Coalition (SPARC) initiative was born from a desire to do that. But the failure of the SPARC model to work in anything more than a very limited way illustrated the real trump card which publishers hold, and that is the intellectual properly vested in the journal brands. We now understand that academia is a deeply conservative world. We see more clearly that academics are time-poor and need to spend all of their available time producing papers for publication in high-impact journals, which have become the prized brands owned – in many cases – by commercial publishers, and that very few of them have sufficient incentive to pause or slow down and assist libraries with their ongoing problems of serials pricing.

The Finch and the recent HEFCE report posit models of open access which encourage publishers to retain their journal brands and their publishing practices, but promote OA by virtue of additional payments for the release of publisher-branded material into the OA world (this is 'gold' open access, in which publishers are paid to provide it). The idea behind this approach is to create a market mechanism by means of these additional payments, known as article processing charges (APCs). The alternative approach of 'green' open access is managed by the universities themselves, involving publishers only by having them agree to delayed publication in repositories, following embargo periods, and often with 'final author' versions of papers that are not quite the finished, published article. Appearing mainly in institutional repositories managed by libraries, these repositories have largely been seen as unthreatening by publishers, many of whom have continued to play the game of exploiting the academic marketplace by claiming fear of wide subscription cancellation and the collapse of their income models if anything other than these two models – gold open access to publisher content at a price, or green open access to an unthreatening shadow corpus – should be countenanced.

The cost to libraries for journal content has remained high, and the arrival of gold OA has pushed it even higher, since the gold route involves new APCs on top of continued above-inflation subscriptions. For the research library community, the hope has been that this increase should be a temporary aberration which will disappear once the business model changes, and the new costs will be offset by falling subscription prices. But already there is evidence that the publishers of the most expensive journal bundles are not acting reasonably in this new world of gold and green, and as a result we are returning to the original goals of open access, with a renewed determination.

The fact is that 14 years of campaigning for and developing infrastructure for open access have had little effect on journal pricing. A recent British Academy report puts this very well: 'If journal prices are unsustainable in library budgets, then there either have to be fewer journals or else journals have to be cheaper. These are pressures which are independent of any open-access policy.'[4] Most librarians realized long ago that the idea of an open access replica corpus of the world's journal literature was not going to happen. Libraries will continue to buy journals, and since the OA movement has not succeeded in bringing down their prices, we will have to buy fewer of them. Publishers might then ask why we are still beating the drum for open access. The answer is because we want to continue to exercise some rights over what should be our own property. Publishers provide a very necessary service to academia – seen in its exemplary form in the case of learned society journals. They are entirely justified in selling the content they publish, and making a fair surplus or profit from it. But what they are selling, in the quality product that a journal represents – branded, clearly produced and easily citable – is also the product of the academy; it is our stuff. And we want to retain the rights – in addition to the publishers' intellectual property rights – to give it away for free to those who are interested but can't afford university-level subscriptions, to preserve it as a safeguard against digital archive loss in private hands, and to text mine it, data mine it, and use it as a research corpus.

We also want to reassert the scholarly publishing order, which has been turned on its head by the factors identified by Guédon. What publishers sell to their subscribers – including largely the libraries of the same institutions that provide their raw material – is what the institutional homes of the authors allow them to sell. But most institutions don't realize this. Ideally, publishers would be licensed to publish by the institutions of their authors. How could this be achieved? The answer could be an aggregation of repository content, what we in RLUK have called an 'Open Mirror'. In our existing system of institutional repositories, the open access material is very much under the radar. A proliferation of over 100 different systems, using variable metadata standards, and with varying degrees of compliance and resource dedicated to filling them, splits attention, diffuses gravitational pull, and serves to keep OA largely hidden from view. What we should now seek to do is to become much more open about our objectives as a library community in relation to this material. This could be done by aggregating it into a single, managed repository that has visibility across the UK.

It is time to make a bolder statement, and to link our repository holdings together into an aggregation with clout, gravitational pull and presence, a repository with strong production values and its own attractive branding. This is the essence of recent feasibility work undertaken by Jisc and RLUK into an Open Mirror. Based upon the holdings from the disparate configuration of institutional repositories, an Open Mirror would do what libraries instinctively do with collections and the metadata that describes them – it would aggregate them into a whole greater than the sum of its parts. In this way, the academy could take control of its own output, building an aggregation which functions as a preservation layer and a research corpus, and which includes material additional to the published version: findings of negative results; underlying data; refutations and other satellite outputs from research practice. It would not replicate the publisher product, but it would offer free access to a version of it, where the tested and peer-reviewed ideas could be found. Where the publisher product services the need for academic reputation and citability, the Open Mirror would be the corpus that truly represented the academy at work. Its establishment would need academic authors to require of publishers as a condition of rights transfer that it was a destination for their outputs at the expiry of embargoes. Learned society publishers in particular should see the mutual advantage.

> "It is time to make a bolder statement, and to link our repository holdings together into an aggregation with clout …"

The idea is not original. In Australia, the National Library has been harvesting the content of university institutional repositories into what is now called its Trove service for the past nine years. And in the US in the past few months, in response to forthcoming government requirements for open access to federally funded research issued by the Office of Science & Technology Policy (OSTP) in 2013, two aggregation initiatives have emerged. One, from the Association of American Publishers, is for a new portal service (CHORUS – Clearing House for the Open Research of the United States) for all federally funded research outputs, which links them through to the papers on the publishers' websites. The other initiative has come from the universities themselves. SHARE (Shared Access Research Ecosystem) proposes a federation of existing institutional repositories, with records upgraded with the metadata necessary to identify the research funder and programme. As with the proposed UK Open Mirror, SHARE will require the retention of intellectual property rights by universities. This is a challenge, but is in many ways the key to success for an academy seeking to regain control of scholarly publishing, and to correct the distortions that we first observed as a community back in the late 1990s. The HEFCE policy on open access in the REF makes reference to this: 'We further recommend that institutions fully consider the extent to which they currently retain or transfer the copyright of works published by their researchers, as part of creating a healthy research environment.' The problem can only be addressed if universities work collectively, and on behalf of their libraries, in this endeavour. A common approach across both sides of the Atlantic would undoubtedly be even stronger.

> "… the key to success for an academy seeking to regain control of scholarly publishing …"

The open access journey has not yet ended, but a historical view can be helpful along the route. Thirteen years ago, Jean-Claude Guédon advised us that it would take some time for the way to become clear: 'Presently, many of the legal and political skirmishes that we observe are probably little more than stages toward the slow, painful, meandering, and sleepwalking invention of new legal categories and institutional settings related to digital publishing'.[5] But if we walk in the shadow of the secretary of the Royal Society and founder of the *Philosophical Transactions*, whose concerns were for the integrity of science and its record, we can be confident that true values will prevail in the end.

References

1. Guédon, J-C, 'In Oldenburg's long shadow: librarians, research scientists, publishers, and the control of scientific publishing', Association of Research Libraries, 2001.
2. Finch, J, Working Group on Expanding Access to Published Research Findings, Accessibility, Sustainability, Excellence: How to Expand Access to Research Publications, 2012, London, UK Government, http://www.researchinfonet.org/wp-content/uploads/2012/06/Finch-Group-report-FINAL-VERSION.pdf (accessed 21 May 2014).
3. Higher Education Funding Council for England 'Open access in the post-2014 Research Excellence Framework', 2014.
4. Darley, R, Reynolds, D and Wickham, C 'Open access journals in Humanities and Social Science: A British Academy Research Project', 2014.
5. Guédon, J-C, ref. 1, p.7.

To cite this Key Issue:

MacColl, J, Key Issue: Out of the shadows: the case for a national repository of open content, *Insights*, 2014, 27(2), 205–209; DOI: http://dx.doi.org/10.1629/2048-7754.167