Institutional repositories have work to do if they're going to solve the access problem

Blog Admin

The Green Open Access route, which encourages the use of institutional repositories for depositing manuscripts as drafts or after a publisher's embargo period, seems to many in the Humanities and Social Sciences like a more agreeable way to widen access to scholarly work. ***Mike Taylor*** *is not convinced that the Green route is able to adequately address the key concerns of the access problem. He outlines five limitations that Green access will have to address if institutional repositories are to become truly useful.*

The progressive RCUK policy on open access has recently come under fire, particularly from humanities scholars, for favouring Gold Open Access (OA) over Green. For various reasons — and I won't, for now, go into the question of which of these reasons are and aren't sound — they favour an approach to open access where publishers keep final versions of their papers behind paywalls, but drafts are deposited in institutional repositories (IRs) and people who want to read the paper can have access to the drafts.

It's appealing to think that this relatively lightweight way of solving the access problem can work. Unfortunately, I'm not convinced it can, for several reasons. I'll discuss these below, not so much with the intention of persuading people that Gold is a better approach, but with the hope that those of you who are Green advocates have seen things that I've missed and you'll be able to explain why it can work after all.

1. Two-class system

Most fundamentally, I worry that Green OA creates a two-class system which I can't approve of. It does this in two ways:

First, it necessarily creates two classes of papers: author's draft and publishers' final versions. These will differ in some respects, and it's hard in general to know what those respects are. Of course pagination will differ — which means you can't cite page-numbers reliably. But other changes are possible as well. For example, Matt and I have a paper in press now for which a whole additional figure — an important one — was added at the proofing stage. In our case, the paper will be OA anyway, but if it were not then the authors' manuscript would be a poor substitute.

And by implication, Green OA creates two classes of researchers — a potentially harmful division between those privileged few who have the "proper" papers and an underclass who have only manuscripts. (It doesn't help that for stupid historical reasons, our manuscripts are often butt-ugly: double-spaced, line-numbered, all the figures at the end instead of where they're needed, etc.) Admittedly, the two-classes-of-researcher problem is not created by Green OA: it already exists, in a worse form where the underclass doesn't have access to *any* version of the paper. But whereas Gold OA solves this problem (everyone has exactly the same access to PLOS and PeerJ papers), Green doesn't.

(To me, it's obvious that democratising access is a good thing. But now that I've made the notion explicit, I can't help the uncharitable thought that there may be those out there who want to *maintain* a two-class system — to retain a notion that they are "in" while others are "out". I hope I'm wrong, and I'm certainly not accusing Green OA advocates of having this motivation. It just seems like it might be implicit in some of the broader struggles over access. Anyway, let's not confuse this separate *potential* problem with access with the *actual* problems with Green OA I'm addressing in this post.)

2. Expense of continuing subscriptions

I find it baffling that people keep talking as though Green OA is cheaper than Gold. It isn't, at all. As I've shown previously, the cost to the world of a paywalled paper (aggregated across all subscriptions) is about $5333. There is no reason to think that will change under the Green model, in which we continue to give the final and best version of our work to publishers. By contrast, even the publisher-influenced Finch estimates typical Gold APCs as £1500-£2000 (about $2270-$3030), which amounts to 43%-57% as much. (Conveniently, the midpoint of the Finch range, £1750, is about $2650, which is almost exactly half of what we pay by the subscription model.

But the true cost of Gold OA is much, much less. Follow the link for the detail, but one credible banner figure starts with the observation that half of all Gold OA articles are published at no cost to the author and that the average APC of the other half is $906, to arrive at a true average APC of $453 — about **one twelfth** of the cost for a paywalled article. So for purely pragmatic financial reasons, Green seems like a silly path. There's a very short-term saving, sure, as we avoid paying APCs. But we have to look further ahead than the next five years.

3. Embargoes

Now there is nothing intrinsic to Green OA that means embargoes must be in place. It's perfectly possible, and manifestly desirable, that no-embargo Green-OA mandate should be enacted, requiring that authors' final manuscripts become available immediately on publication. But for whatever historical reasons (and I admit I find this baffling) there are few or no Green-OA mandates that do this. Even the best of them seem to allow a six-month delay; twelve months is not uncommon (and Michael Eisen worries that the new White House policy with further establish twelve months as the norm.

I will have more to say about embargoes in a subsequent post. (SPOILER: it's not going to be pretty.) But for now it suffices to say that any system that makes research freely available only a year after it's published is wholly inadequate. Not to mention stupid. Stupid and inadequate. So if Green OA is going to be the solution we need, it has to break free from embargoes.

4. Non-open licences

Similarly, there is no intrinsic reason why Green OA should mean non-open licences and Gold OA should mean truly open (BOAI-compliant) open access. And yet history has brought us to a point where is often how things are. For example, the RCUK policy (even before its progressive erosion got properly under way) says of its Gold arm that "The CC-BY license should be used in this case", but contains weasel words in its Green arm:

> *"the journal must allow deposit of Accepted Manuscripts that include all changes resulting from peer review (but not necessarily incorporating the publisher's formatting) in other repositories, without restrictions on* ***non-commercial*** *re-use."*

This just won't do. It's not open access. To quote Heather Piwowar's pithy statement once more, "We do basic research not only to know more, but to *do* more". Non-commercial licences impede the *use* of research, and that's not to the benefit of wider society. (I won't labour this point now, because I'll have more to say on non-commercial clauses in a subsequent post.) So as with embargoes, if Green OA is going to be the solution we need, it has to break free to its habitual acceptance of non-commercial clauses.

5. Practical failings

On top of the fundamental problems already discussed (two-class system, expense of continuing subscriptions, embargoes and non-open licences), the repository system *as it exists today* suffers from a suite of practical problems that render it pretty inadequate.

- Many institutions don't even *have* an IR; or if they do, it doesn't work.
- Many scholars aren't associated with an institution and so don't know where they should reposit their manuscripts. (That this is overlooked is a symptom of an unfortunate elitist tendency among academics.) [**UPDATE 4th March**: thanks to Neil Stewart, whose comment below points out Open Depot as a solution to this.]
- The use of IRs involves an institution-by-institution fragmentation, with different user interfaces, policies, etc.
- For whatever reasons, many scholars do not bother to reposit their manuscripts in institution repositories.
- Even when mandates are in place, compliance is often miserable, to the point where Peter Suber considers the 80% NIH compliance rate as "respectable". It really isn't. 100% is acceptable; 99% is respectable.
- Many IRs have abject search facilities, often for example lacking the ability to restrict searches to papers that are actually available.
- Many IRs impose unnecessary restrictions on the use of the materials they contain: for example, Bath's repo prohibits further redistribution.
- There is no central point for searching all IRs (at least not one that is half-decent; I know about OAIster).
- The quality of metadata within most IRs variable at best
- Use of metadata across IRs is inconsistent — hence many of the problems that render OAIster near-useless.

… and, I am sure, many more that I've not thought of right now.

Could these issues be addressed? Yes, probably; but ten years have unfortunately not done much to resolve them, so I don't feel all that confident that the next ten will. Do the IR advocates have a plan for solving these problems? Because they are much more political/sociological than technical, and those always seem to be the hardest ones to solve.

This article was originally published on Mike Taylor's blog, and is published here with permission.

About the author:
Mike Taylor *is a dinosaur palaeontologist, computer programmer and open access advocate, affiliated with the University of Bristol. He has named two new dinosaurs, Xenoposeidon ("alien earthquake god") and Brontomerus ("thunder thighs") and written other papers so boring that his wife fell asleep while he was explaining one of them to her.*

Related posts:

1. Can LinkedIn and Academia.edu enhance access to open repositories?
2. Open access repositories are beginning to push academic publishers off their previously unreachable perch.
3. University libraries, repositories and Open Access should be seen as crucial tools in improving the impact of academic research
4. Digital visibility is king but what colour is our Open Access future?
5. Hybrid gold open access and the Chesire cat's grin: How to repair the new open access policy of RCUK