

Published: March 2015

Editorial: Will new data sharing policies feed the rich at the expense of the poor?

John Hargrove - Senior Research Consultant at SACEMA.

Scientists offering papers for publication will be becoming increasingly aware of a significant change in the attitude of journals to the publication of the data used to reach conclusions drawn in their manuscript. Whereas, in the past, there has been little or no requirement to make data available, new regulations are moving rapidly and uncompromisingly towards an opposite extreme policy where all of the data, and related metadata, required to replicate the reported findings must be made freely available to the world at large.

The attitude of PLoS in their new Data Policy to become effective on 1 March 2015, is perhaps typical of the new attitude: "Data availability allows replication, reanalysis, new analysis, interpretation, or inclusion into meta-analyses, and facilitates reproducibility of research, all providing a better 'bang for the buck' out of scientific research, much of which is funded from public or non-profit sources."

There is much to be said in favour of this argument, but one wonders whether journals, and funders behind the scenes who are apparently pulling their puppets' strings, have thought through some of the ramifications of the new policy.

Thus, whereas PLoS is forthrightly blunt about the required behaviour on the part of the person(s) who are publishing papers in their journals and who will be forced to make their data public, there is nothing in their document suggesting that there will be any restriction of any kind on any third parties who use those data. In particular, there is no mention in their policy statement of any requirement on the part of any person who utilises deposited data to do any of the following:

- Ask the opinion of the person or persons who generated the data as to the validity of the analyses to be published from their data.
- Allow, or offer, those persons sight of draft manuscripts before they appear in the press.
- Offer a co-authorship, regardless of circumstances, to any of those who collected the data.

It is clear that this policy aims, in part, to allow the general scientific community to check that

published work is indeed supported by the data quoted and used by the author(s) of a given paper. This is well and good and as it should be. But, as I read their policy, third party users of the data are not required even to check with the original authors that they have correctly understood the situation *before* they publish what could be a completely inaccurate analysis based solely on a misunderstanding of what the original authors had done. This policy could lead, at best, to a huge waste of time and effort and at worst to incorrect and, depending on the circumstances, even dangerous misrepresentations.

It is not unreasonable to think that this could lead to the publication of material damaging to the reputation of the original authors who need not, under the new policy, even be informed of the calumny before it is published. Said original authors would then, of course, have to publish a rebuttal. On the one hand this is a serious waste of time, as the rebuttal will hardly be read, and on the other the damage to their reputation has already been done. Mud sticks: and there will be subsets of the scientific community and the public at large who will choose to accept, and even to promulgate, the faulty interpretation – particularly if it suits their individual and/or political ends.

Moreover, the new data policy goes well beyond the simple process of checking the validity of the published work. Once the data are in the public domain they may be used for the publication of papers that have nothing to do with the work already published. The policy can thus be hugely unfair to the people who have generated the data, and it will be particularly unfair to those data generators who are from small, poorly funded institutions, and those from the Third World in general.

Suppose, for example, a person from the Third World, or small institute anywhere, has carried out a very large experiment over many years, and that the data can realistically be expected to give rise to numerous papers. Under the new policy that institute will be required – *at the time that the first paper is published utilising those data* – to make freely available, to the world at large, data, and meta-data, that could be used to publish further papers.

At that point there would be nothing at all to prevent a professor from a well-funded institution of putting half a dozen PhD students onto cannibalising the data and publishing papers which the person who, often at great personal cost, collected the data had in mind to publish him/herself.

Of course when the boot is on the other foot and original data are produced by the professor from the wealthy institute, things are rather different. Said professor will, or at least should, also be forced to make public with half a dozen PhDs in tow, is in a very much stronger position to make further use of the data to produce additional papers before the data scavengers descend.

The reader will follow that there is an asymmetry here. What the new policy is effectively doing is stacking the odds of publishing scientific work ever more strongly in favour of the rich and powerful and, yet further, marginalising the poor and weak – particularly those in the Third World.

What the policy will also be doing is to discourage everybody, but particularly those from disadvantaged institutions, from carrying out large studies. Who wants to spend/waste years of work knowing full well that they are essentially working for somebody else – and that the “somebody else” is not required in any way to share the fruits of their labour, save perhaps an acknowledgement that can be hidden away in the Supplementary Information?

What recourse have we in the face of these developments? One way is simply to move away from peer-reviewed journals and to publish in spaces such as *arXiv*. This has the great advantage that one is not faced with the increasingly crippling costs, particularly again for the Third World authors, of publication in the “Open Access” family. It has the further advantage that, whereas there is no peer review, one can simply forward the papers to *all* of one’s peers – and thereby open a lively and open discussion, which ought to achieve the same end as the system currently in use by open access journals.

Alternatively, and this is a very real danger, authors will turn to fly-by-night journals, from whom one gets numerous invitations to offer papers, whose sole intention is to make money, and who will not have the slightest qualm about publishing papers without a requirement for the publication of data.

These are not the ways that I – and I assume my colleagues – would like to go, but I believe that scientists as a whole will in future move away from journals with poorly or incompletely thought-out policies on data sharing.

John Hargrove - Senior Research Consultant at SACEMA. Areas of interest: *tsetse* and *trypanosomiasis* biology and control.
jhargrove@sun.ac.za