Open Access to Refereed Research Publications and Open Access to Research Data: A Crucial Strategic Distinction

Stevan Harnad
Canada Research Chair in Cognitive Sciences
Université du Québec à Montreal
&
School of Electronics and Computer Science
University of Southampton

(1) On the Open Access Impact Advantage for Refereed Research Reports

It has now been repeatedly demonstrated that refereed research articles that are made Open Access (OA) are used and cited significantly more in every scientific and scholarly field tested than those that are not made OA. It has now also been shown that this OA advantage is just as great for mandated OA as it is for self-selected OA. This means that the OA Advantage is not (as some have suggested) simply an artifact of selectively making higher-impact research open access: OA is the cause of the increased research impact. This finding greatly increases the importance and urgency of mandating OA for the sake of increasing and accelerating research uptake and progress.

Gargouri, Y., Hajjem, C., Lariviere, V., Gingras, Y., Brody, T., Carr, L. and Harnad, S. (2010) Self-Selected or Mandated, Open Access Increases Citation Impact for Higher Quality Research. *PLOS ONE* 5(10): e13636.
http://dx.plos.org/10.1371/journal.pone.0013636

(2) On the Importance and Potential of Open Access Data-Archiving

Although there is not yet enough OA data to be able to demonstrate that the same kind of impact benefits will be generated by OA to research data as those that have been demonstrated for OA to research articles, it is highly probable that that will prove to be the outcome. Moreover, the impact benefits of making research articles OA, and the rich new means of measuring research usage and impact that OA is generating will also serve as incentives to encourage researchers to provide OA to both their articles and their data.

Brody, T., Carr, L., Gingras, Y., Hajjem, C., Harnad, S. and Swan, A. (2007) Incentivizing the Open Access Research Web: Publication-Archiving, Data-Archiving and Scientometrics. CTWatch Quarterly, 3 (3).
http://www.ctwatch.org/quarterly/articles/2007/08/incentivizing-the-open-access-research-web/

(3) On the Crucial Differences Between Research Archiving and Data-Archiving -- And Why Immediate Data-Archiving Cannot be Mandated

There is, however, a crucial difference between providing OA to research articles and providing OA to data: Scientists and scholars are not primarily data-gatherers. They

gather data in order to data-mine, analyze, interpret and build further findings, theories and applications on it. Hence (except in the rare cases where the data speak for themselves), researchers cannot be expected (or mandated) to make their data OA immediately upon having collected or generated it, for all other researchers to data-mine and analyze. Researchers must be given sufficient time to data-mine their data, having invested the time and effort into collecting or generating it. And the length of the fair embargo interval on Open Access to data will vary depending on the nature of the data and the time, effort and ingenuity required to collect or generate it. This is fundamentally different from the case of refereed research reports, for which there is no justification whatsoever for embargoing Open Access once the paper has been peer-reviewed and accepted for publication.

Hence providing OA to refereed research reports can and should be mandated by researchers' institutions and funders, immediately upon acceptance for publication. Such immediate OA mandates cannot, however, be simplistically extended to research data (nor to unrefereed preprints of research reports) without generating the risk of needless and counterproductive conflicts of interest with the researchers that gathered the data. OA data-archiving, as soon as possible, should be strongly encouraged; in some cases embargo length limits can be set. But it cannot and should not be mandated (except in very special cases where the data-gathering itself is the research that is being funded.)

OA, OA self-archiving, OA publishing, and data archiving
http://openaccess.eprints.org/index.php?/archives/776-guid.html

Open Access and Open Data
http://openaccess.eprints.org/index.php?/archives/353-guid.html

On Not Conflating Open Data (OD) With Open Access (OA)
http://openaccess.eprints.org/index.php?/archives/733-guid.html

More on Potential Conflict of Interest with Open Data (OD) Mandates
http://openaccess.eprints.org/index.php?/archives/734-guid.html%A0

Don't Risk Getting Less By Needlessly Demanding More
http://openaccess.eprints.org/index.php?/archives/386-guid.html

How Green Open Access Supports Text- and Data-Mining
http://openaccess.eprints.org/index.php?/archives/310-guid.html

On Patience, and Letting (Human) Nature Take Its Course
http://openaccess.eprints.org/index.php?/archives/262-guid.html

Open Access: What Comes With the Territory
http://openaccess.eprints.org/index.php?/archives/259-guid.html

Text of presentation:

SS1 Hello from Montreal. I'm sorry I couldn't be here in person for the symposium on data-sharing in Washington today.

Greetings to Mike Lesk, Mike Carroll, and Heather Joseph, and greetings and thanks to Paul Uhlir for letting me participate virtually in this way.

I am sure that the importance and the many benefits of research data-sharing have already been described today. I apologize if some of the points I'll be making have already been made by the other speakers.

I'll be talking specifically about open access and open data, the theme of an international conference that will be taking place in Cologne, Germany on December 13th and 14th. **SS2 OA**

What made both open access to research articles and open access to research data possible was the advent of the online medium in which both articles and data could be shared in a way that was impossible in previous media. The open access movement came first (although researchers had already begun online sharing of both articles and data well before that). **SS3 OAtoWhat**

Open access or OA was focused on the 2.5 million articles a year that are published in the planet's 25,000 peer reviewed research journals, across all scientific and scholarly fields, in all languages and all nations. **SS4 cycle1** The idea of OA was that in the online era those articles no longer needed to be accessible only to users at institutions that could afford to subscribe to the journal in which they were published. **SS5 cycle2** They could now be made accessible to all potential users webwide by depositing them, free for all, on the web.

Institutional repositories were created with free OAI-compliant software that made them all interoperable, harvestable, navigable, searchable and useable as if they were all just one global repository, free and open to all. **SS6 OAadvantage**

The advantages to research and researchers of making all their articles OA in this way were reported in a growing number of studies showing that OA significantly increases research uptake, usage and impact as measured by downloads and citations. The message about the feasibility and benefits of OA spread to researchers worldwide. **SS7 repositories**

Institutional repositories were created, worldwide. **SS8 publishers**

Most publishers endorsed providing immediate open access.

And researchers began depositing their articles on the web. **SS9 bjork**

But it soon became apparent that the spontaneous deposit rate – currently about 20% -- was not high enough and not growing fast enough to make all, most, or even much of the planet's annual research output OA. **SS10 survey**

Researchers were surveyed, and their responses revealed that – because of worries about copyright and about the time and effort it might take to deposit – researchers would not provide OA until and unless their institutions and/or their funders mandated it. But if it was mandated, over 90% would comply, over 80% of them willingly. **SS11 outcome**

So the only thing needed was deposit mandates from institutions and funders, and these are now growing worldwide, including mandates from NIH, Harvard, MIT. **SS12 mandgrow**

But the reason it is possible for research institutions and funders to adopt OA mandates, and the reason they are being complied with willingly by researchers, is that, without exception, every one of the 2.5 million annual research journal articles that are the OA movement's primary target is written for the sake of research impact, which OA maximizes, and not for income or royalties from selling access to them to users. **SS13 OAadvantage** Access-tolls block research uptake and impact, and it is for uptake and impact that research is funded, and conducted; and it is on the uptake and impact of their research that researchers' careers depend. **SS14 OAtoWhat**

So much for research articles and why OA needs to be, can be, and is being mandated for them. But now what about research data? Surely uptake and impact need to be maximized there too, for much the same reasons. **SS15 NotDatagatherers**

The answer is, yes, but the case is somewhat more complicated with data, because, with the exception of some special cases, scientists and scholars are not data-gatherers. They are analyzers and interpreters of data. They do gather and generate data, to be sure, and often at the cost of much time and effort, but they do so in order to be able to exploit and data-mine the data they have gathered or generated. What they publish as articles is the result of their analyses; that is why they are researchers, and it is on that that their careers and rewards depend. **SS16 survey**

A mandate to make their refereed research articles freely accessible to all users immediately upon acceptance for publication is welcomed by most researchers, but a mandate to make their *data* freely accessible immediately upon being gathered, or even immediately upon publication of the results of the first analyses would not be welcomed by most researchers – perhaps not even by many. The reason is that researchers must be allowed first exploitation rights on the data they gather, if they wish it, otherwise there is the risk of destroying their motivation for gathering the data in the first place.

Some kinds of research and some researchers will be amenable to immediate data-sharing. They need merely be encouraged and given the technical means to do so.

SS17 1st exploitation But in general, researchers will need to be allowed exclusive first-exploitation rights on their data, even if the data-gathering was publicly funded.

The length of the allowable exclusive period will vary by the field and by the project and will need to be negotiated in advance, as part of the conditions of the funding. **SS18 metrics**

Mandating OA to the resulting research articles, together with new metrics that measure the uptake and usage of research data in particular will both increase researchers' motivation to make their research data OA as soon as possible. **SS19 motives** But this fundamental pragmatic and motivational difference between OA to research articles and OA to research data needs to be understood and taken into account in order to avoid needlessly generating a conflict of interest in researchers that could rebound not only against providing open data, but even against providing OA to the articles -- which although it is within reach, has not yet been grasped either. **SS20 refs**

Thank you for listening.