Integrating Universities' Thesis and Research Deposit Mandates

Stevan Harnad
Canada Research Chair in Cognitive Sciences, Université du Quebec à Montréal
http://www.crsc.uqam.ca/
School of Electronics and Computer Science, University of Southampton
amsciforum@gmail.com

ABSTRACT: A growing number of universities are beginning to require the digital deposit of their thesis and dissertation output in their institutional repositories. At the same time, a growing number of universities as well as research funders are beginning to mandate that all refereed research must be deposited too. This makes for a timely synergy between the practices of the younger and older generation of researchers as the Open Access era unfolds. It also maximizes the uptake, usage and impact of university research input at all stages, as well as providing rich and powerful new metrics to monitor and reward research productivity and impact. It is important to integrate universities' ETD and research output repositories, mandates and metrics as well as to provide the mechanism for those deposits that may need to be made Closed Access rather than Open Access: Repositories need to implement the "email eprint request" Button for all Closed Access Deposits. Any would-be user webwide, having reached the metadata of a Closed Access Deposit can, with one click, request an eprint for research purposes; the author instantly receives an automatic email and can then, again with one click, authorize the automatic emailing of one copy to the user by the repository software. This feature is important for fulfilling immediate research usage needs during any journal-article embargo period, and it also gives the authors of dissertations they hope to publish as books a way to control who has access to the dissertation. Digital dissertations will also benefit from the reference-linking and book-citation metrics that will be provided by harvesters of the distributed institutional repository metadata (which will also include the metadata and reference lists of all university book output). Dissertation downloads as well as eprint-requests will also provide useful new research impact metrics.

What is Open Access?

Open Access (OA) means free online access to scholarly and scientific research (Harnad 1995). This does not necessarily mean all of scholarly and scientific research. The *sine qua non* for OA is that the item must be an *author give-away*. That means it is written primarily to be used, applied and built upon by other researchers rather than to be sold for royalty or fee income. Although there are many kinds of research output that would benefit from and may eventually seek to be OA, the only content that is an exception-free candidate for OA is the 2.5 million peer-reviewed journal articles that are published annually in the planet's 25,000 scholarly and scientific journals, across all disciplines and languages. Other potential candidates for OA are of course welcome, but they are riddled with exceptions: Not all scholarly/scientific book authors want to give away their texts for free online. Nor do all researchers want to make their research data accessible free for all, at least not immediately after they have collected it: they may want to have the exclusive opportunity to data-mine, analyze and report it first, sometimes for a number of years. And of course the creators of most non-scholarly content – trade books, music, video, software – may never want to make it OA at all.

What about theses and dissertations, then? This is the category of scholarly and scientific research output that is the closest to refereed research articles, OA's only exception-free target. The reason theses include exceptions is of course that some thesis authors hope to make their theses into books, perhaps even trade books. Hence they may fear that making them freely accessible online will either make it impossible for them to find a publisher, or it might reduce their prospects of royalty revenue, or both. These authors may or may not be right in their fears, but clearly they constitute exceptions to the rule that OA content must be author give-away content. Further complicating the special case of theses is the fact that they may contain portions that are the work of other authors, preventing the thesis author from making the full-text OA even when he wants to, because of copyright and intellectual property restrictions.

Fortunately, there are solutions for all of these potential problems and special cases, as I will describe below – solutions that sometimes also need to be applied to peer-reviewed articles as well (e.g., because of copyright issues). So it can be stated with assurance that theses too, without exception, can be treated the same way as other forms of university research output, insofar as OA is concerned.

Why Provide Open Access?

The first and foremost reason for providing OA is that it substantially enhances the uptake, usage, and impact of research findings, by maximizing access to them. In the case of journal articles, providing OA means that online access is no longer restricted to those users whose institutions can afford subscription access to the journal in which the article is published. In the case of theses, too, this means users not having to afford and to wait to order a hard copy, nor to pay for online access. As a consequence of removing these needless and dysfunctional paper-era access barriers to research findings – barriers that are already obsolescent in the online medium -- it has now been repeatedly demonstrated across all disciplines that OA increases citations, downloads, and other measures of research uptake, usage and impact (Gargouri & Harnad 2009; Hajjem et al. 2005; Hitchcock 2009), sometimes quite dramatically in the case of the most important and highest quality research (Seglen 1992).

Increasing the uptake and impact of research output is beneficial not only to the progress of research itself, but also to the career of the researcher (both in terms of employment, promotion and salary and in terms of future research funding). Maximizing research impact is also beneficial to the researcher's institution, and this is reflected in the fact that "publish or perish" is now based not merely on publication counts but also on citation counts, download counts and other metrics reflecting the degree to which the publications have proved useful, as reflected by their uptake by other scholars and scientists. Research funders, too, need to justify the tax-payer money that they have used to support research by showing that the research has had an impact.

For those authors of theses who are interested in making an academic or research career, the impact of their theses will be important in much the same way, and hence they too have an incentive to provide Open Access to their work (Brody et al 2007). Webmetrics are also becoming increasingly important measures of impact even beyond research and academics. Gratuitous access barriers to give-away work make as much sense as charging for access to advertisements.

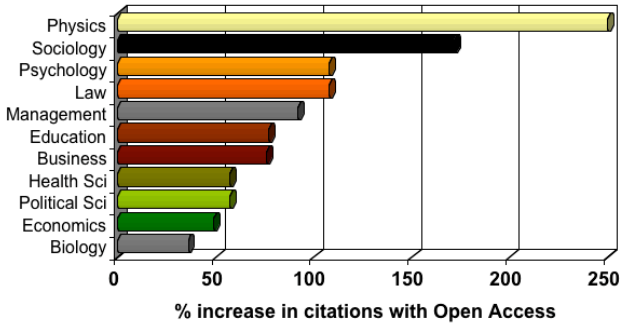

Figure 1: Comparison citation counts for articles published in the same journal and year that have or have not been made freely accessible online. Across all fields, free only access increases citations. (Harnad & Brody 2004; Hajjem et al. 2005)

How to Provide Open Access?

The way to provide OA is to create an OAI-compliant Institutional Repository (IR; ROAR 2009) and to deposit all peer-reviewed journal articles and all theses and dissertations therein. The problem is that only about 15% of peer-reviewed journal article authors deposit in their IRs spontaneously, and the figure is no doubt even lower for theses and dissertations – until and unless deposit is mandated by their institutions (ROARMAP 2009), in which case the deposit rate for articles approaches 100% within two years (Swan & Brown 2005; Sale 2006) and no doubt the compliance rate for compulsory deposit of theses is even higher.

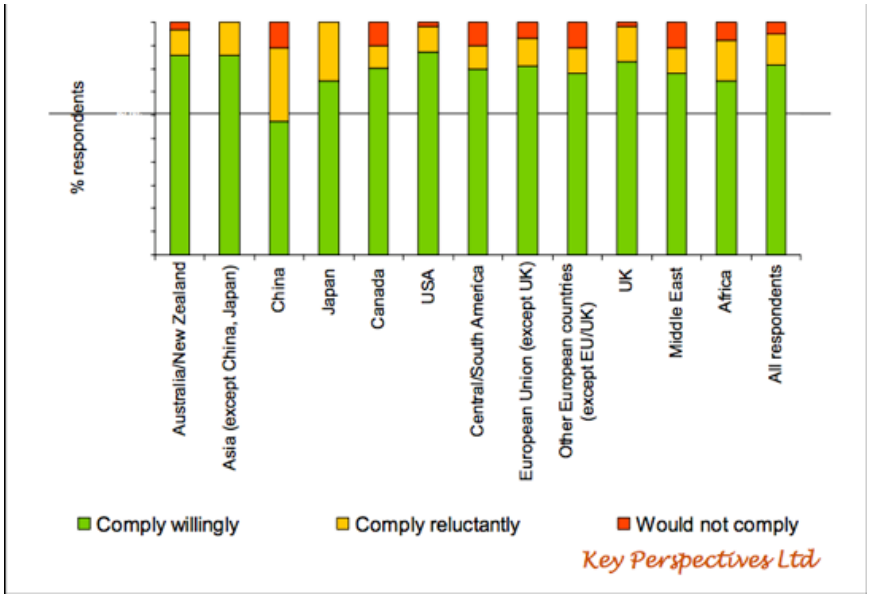

Figure X: Researchers responses to international survey on whether they would comply with deposit mandates (Swan & Brown 2005)

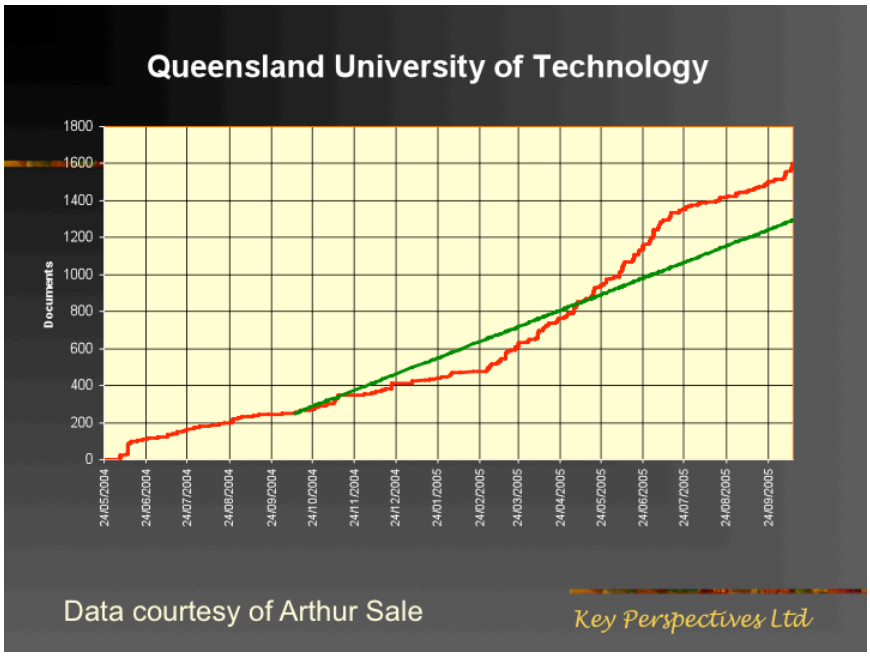

Figure X: Researchers' actual compliance rate when deposit mandate is adopted (Sale 2006): Deposit rate (red) is in line with research output rate (green).

Intergrating Repositories and Deposit Mandates for Articles and Theses

At the present time, however, these two "feeders" of OA content – articles and theses – although they are the research output of the same universities, not only proceed on different policy tracks (with some institutions mandating thesis deposit but not article deposit, and vice versa) but they often even have different repositories, the article repository usually being the

Institutional Repository (IR) and the thesis repository sometimes being the IR, sometimes a separate institutional thesis repository, and sometimes an institution-external central thesis repository.

The optimal solution, the most beneficial to all parties, and the fastest and simplest way to reach the universal OA for all articles and theses globally that all researchers and their institutions desire is to integrate both kinds of repositories and both kinds of deposit mandates: *Universities should have a unified deposit mandate for all research article and thesis output, and the locus of deposit should either be a single university IR or a portal that integrates all the university's repositories into one "virtual" IR.* Both kinds of mandates should also be integrated so as to pinpoint not only where the deposit should be made, but when: For articles, the deposit should be made immediately upon acceptance for publication, and for theses, immediately upon acceptance of the thesis. In both cases, the deposit should be the revised, corrected final draft. (Earlier drafts can be deposited too, but deposit should not be required.)

Integrating article and thesis deposit mandates in this way creates a synergy, encouraging the adoption of mandates covering both kinds of content. It also creates a culture with an established practice of providing OA at all stages of a research career, starting from the very beginning. I do not have data on the growth rate of thesis deposit mandates, but the growth rate of article deposit mandates has lately begun to accelerate. The very first deposit mandate was adopted in 2002 by the School of Electronics and Computer Science at the University of Southampton, who were also the designers of Eprints, the first free, Open Source software for creating OAI-compliant OA IRs (Tansley & Harnad 2000). But it was not until the adoption of the world's 44th deposit mandate, by Harvard's Faculty of Arts and Sciences in May 2008 that OA mandates began their growth spurt, nearly doubling their first five years growth to the current 83 in the year following the Harvard mandate (Swan 2009).

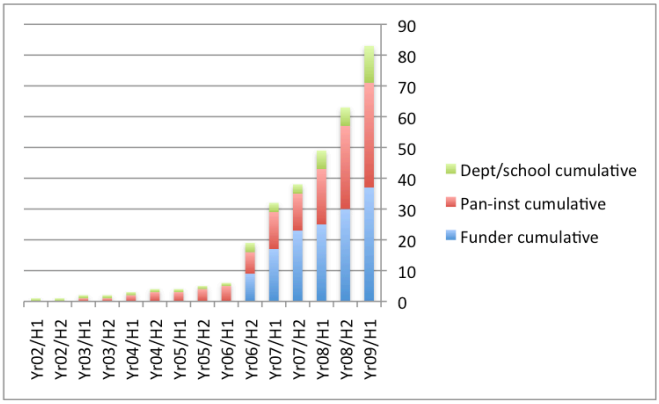

Figure 5. Accelerating rate of adoption of article deposit mandates by research funders, institutions, and departments (Swan 2009). (No data on adoption rate for thesis deposit mandates yet: please register in ROARMAP [2009].)

The Immediate-Deposit Mandate with the Closed-Access Option

None of the OA mandates today are ideal (ROARMAP 2009): The ideal mandate would be immediate-deposit (upon acceptance) and immediate-OA. Most journals (63%) endorse immediate-OA, but the rest request access-embargoes of various lengths. For article authors wishing to honor these access-embargoes, the deposit can be set as Closed Access, which means only the author can access it. The IRs, however, have a Button – the "email eprint request"

Button. In an extension of the longstanding research practise of mailing reprint requests to the author, the would-be user who reaches the metadata of a Closed Access deposit in an IR can cut/paste his own email address into a box and click. This sends the author an instant email requesting one copy of the eprint for research purposes. If the author wishes to honor the request, he need merely click on a URL within the email received, to trigger the immediate automatic emailing of the requested eprint by the IR.

The eprint-request Button does not provide OA, but it provides almost-OA. And not only can it tide over worldwide research needs during embargo periods almost as immediately and effectively as OA, but it provides the author with yet another impact metric, namely, the eprint request counts. It also makes it possible for all institutions and funders to adopt immediate-deposit mandates, without the need of exceptions, opt-outs or delays in the date of deposit; embargoes and other copyright restrictions apply only to the date the deposit is made OA, not to the date the deposit is made.

The Button hence also provides a solution for the minority of theses whose authors do not wish to make them OA for some reason, such as that they hope to publish them as a book: It is entirely up to the author whether to honor eprint requests. (For theses containing copyrighted material by other authors, another option is to deposit a Closed Access verson of the full text and an OA version of the text with the copyrighted portions omitted.)

And, needless to say, Closed Access deposit plus the discretionary Button is also an option for other research outputs, in addition to articles and theses, such as monographs, books, data, software, multimedia and of course unpublished manuscripts.

Thesis Impact Metrics and Book Metrics

In the online era there is no reason why thesis authors should not be credited and rewarded on the basis of the same kinds of usage and impact metrics used to assess and reward researchers and academics. Theses and books, along with articles, will be fully citation interlinked, generating metrics such as:

Citations (C)
CiteRank (recursively weighted, like Google)
Co-citations
Downloads (D)
C/D Correlations
Hub/Authority index
Chronometrics for C and D: Latency/Longevity
Interdisciplinarity: Endogamy/Exogamy
Book citation (and Amazon sales) index
Links
Tags
Commentaries
Journal Impact Factor
h-index (and variants)
Co-authorships
Publication counts
Number of publishing years
Semiometrics (latent semantic indexing, text overlap, etc.)

These metrics (and more) can in turn be combined and jointly cross-validated against peer rankings, as in the UK Research Assessment Exercise (Harnad 2009a), discipline by discipline. This will inititialize the weight on each metric (within each discipline). The weights can continue to be fine-tuned and calibrated on the basis of further testing and use as an aid in research assessment and comparative ranking of research, researchers, and institutions. Such batteries of multiple metrics are likely to be far more robust, predictive and equitable than the use of single, unvalidated metrics (such as citations alone) today.

References

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://eprints.ecs.soton.ac.uk/14418/

Gargouri, Y. and Harnad, S. (2009) Logistic regression of potential explanatory variables on citation counts. http://www.crsc.uqam.ca/yassine/SelfArchiving/LogisticRegression.htm

Hajjem, C., Harnad, S. and Gingras, Y. (2005) Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact. *IEEE Data Engineering Bulletin* 28(4) pp. 39-47 http://eprints.ecs.soton.ac.uk/11688/

Harnad, S. (1995) The Subversive Proposal. In: Ann Okerson & James O'Donnell (Eds.) *Scholarly Journals at the Crossroads: A Subversive Proposal for Electronic Publishing*. Washington, DC., Association of Research Libraries, June 1995. http://www.arl.org/sc/subversive/i-overture-the-subversive-proposal.shtml

Harnad, S. (2009a) Open Access Scientometrics and the UK Research Assessment Exercise. *Scientometrics* 79 (1) http://eprints.ecs.soton.ac.uk/17142/

Harnad, S. (2009b) Waking OA's "Slumbering Giant": The University's Mandate To Mandate Open Access. *New Review of Information Networking* (in press) http://eprints.ecs.soton.ac.uk/17298/

Harnad, S. & Brody, T. (2004) Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals, *D-Lib Magazine* 10 (6) June http://www.dlib.org/dlib/june04/harnad/06harnad.html

Hitchcock, S. (2009) The effect of open access and downloads ('hits') on citation impact: a bibliography of studies. http://opcit.eprints.org/oacitation-biblio.html

ROAR (2009) Registry of Open Access Repositories http://roar.eprints.org/

ROARMAP (2009) Registry of Open Access Repository Material Archiving Policies http://www.eprints.org/signup/fulllist.php

Sale, AHJ (2006) Comparison of IR content policies in Australia. *First Monday* 11 (4) http://eprints.utas.edu.au/264/

Seglen PO (1992) The skewness of science. Journal of the American Society for Information Science 43:628-38

Shadbolt, N., Brody, T., Carr, L. and Harnad, S. (2006) The Open Research Web: A Preview of the Optimal and the Inevitable, in Jacobs, N., Eds. *Open Access: Key Strategic, Technical and Economic Aspects*. Chandos. http://eprints.ecs.soton.ac.uk/12453/

Swan, A. (2009) Multiplying mandates. *Optimal Scholarship*, http://optimalscholarship.blogspot.com/2009/05/multiplying-mandates.html

Swan, A and Brown, S (2005) Open Access Self-Archiving: An Author Study. *JISC Technical Report*.

Tansley, R and Harnad, S (2000) Eprints.org Software for Creating Institutional and Individual Open Archives. *D-Lib Magazine* October 2000. http://www.dlib.org/dlib/october00/10inbrief.html#HARNAD