Point & Counterpoint: The Purpose of Institutional Repositories: Green OA or Beyond?

Rebecca Kennison, Sarah L. Shreeves, Stevan Harnad

Point & Counterpoint

The Purpose of Institutional Repositories: Green OA or Beyond?

Institutional Repositories—So Much More than Green OA

Rebecca Kennison *Director, Center for Digital Research and Scholarship, Columbia University*
Sarah L. Shreeves *IDEALS Coordinator and Scholarly Commons Co-Coordinator, University of Illinois at Urbana-Champaign*

Institutional repositories (IRs) have a conflicted history in terms of purpose. Although always closely associated with the open access movement, in particular open access to the published research through self-archiving ("Green" OA), an approach long championed by Stevan Harnad (e.g., Harnad, 1999) and others, some of the most influential and visionary early essays on IRs speak of them as providing infrastructure for the stewardship of a wide range of institutional output (Lynch, 2003) and as a new way for libraries to support publishing functions (Crow, 2002). And while many libraries have concentrated on green OA to fill their IRs—with or without mandates, always with mixed success— many more have slowly but surely built successful, thriving IRs by providing stewardship of and access to the grey literature, the theses and dissertations, the undergraduate research, and the research data produced on their campuses. In fact, we would argue that libraries are better placed to implement green OA resolutions and mandates when their IR is already well populated and well used with other critical institutional content. An IR should focus on the "I"—on the output of the institution, created by individual researchers producing much more than published peer-reviewed articles.

We, of course, believe that provision of open access to the published literature is an important and critical role of the IR. The argument before us is whether it should be the *primary* role, and we believe that even in the case of an institutional mandate to archive some version of the published literature, doing so should only be one priority of the repository, not its sole purpose. In addition to providing free access to (some) published articles, under the terms allowed by publisher licenses, IRs can fill a critical need for preservation of and access to research output other than published journal articles, provide useful support for pedagogical initiatives such as formal undergraduate research programs, and provide infrastructure for publishing initiatives.

With the rise of the digital, scholarship and research have been shifting in fundamental ways. Scholarly output has always gone well beyond traditional articles in traditional journals, and this is even more the case in a rapidly changing digitally driven—dare we say "DIY"?—academic environment. For some disciplines, such as economics, computer science, and business, technical reports and working papers are a critical component of the scholarly communications landscape. These pieces of so-called "grey literature," which libraries used to receive from departments and research centers in paper, now often exist only on the

web; the risk of loss is great if there is not an archival system like an IR in place. Theses and dissertations now often exist only electronically (as ETDs); it is often the IR that preserves and provides access to these valuable scholarly and institutional records. In addition, the IR can provide support for a wide variety of "supplementary" files for ETDs, crucial information that because of its format cannot be included in the PDF submitted to ProQuest; these files often include datasets, software, and multimedia content, as well as research protocols. In both these cases (technical reports and ETDs), libraries traditionally held responsibility for preservation and access to these materials; we cannot and should not abandon that role in the digital environment. Many institutions have focused on digitization of older grey literature as well as theses and dissertations and have placed these within their IRs; this activity often gives new life to older, but still important and applicable, research. Whether recent or historical, it is this "original" content that is most valued by users, who cannot find this material anywhere else. For example, the top downloads for IDEALS, the IR at the University of Illinois at Urbana-Champaign, are regularly techncial reports (such as http://hdl.handle.net/2142/9750) and masters theses (such as http://hdl.handle.net/2142/16211).

Beyond the standard grey literature such as technical reports, working papers, theses, and dissertations, one finds blogs, video and audio and other multi-modal work, research instruments and protocols, and datasets cited and used within the literature. How can institutions ensure long-term persistent access to such material? Institutional repositories can in many cases provide the infrastructure to support long-term preservation and access. Both of the authors have been approached several times in the past year by authors who need to provide open, citable access to datasets and supporting software that are to be published alongside a journal article; the publishers asked them to approach their institutional repository to provide this service. Columbia has taken a proactive role whenever journals (e.g., *Journal of Neuroscience* or *Journal of Experimental Medicine*) have announced they will no longer take "supplementary material" to reach out to authors in those journals to offer the IR as a place to place the data, providing in return a permanent URL.

Beyond the IR's service to researchers, for many public universities in the United States—particularly land grant institutions—there is an explicit expectation that much of what we do within the university, including research, is for the public good of the citizens of the state and beyond. IRs are often employed within this mission to make research output, whether packaged for the general public or presented in the form of technical reports, openly available. At the University of Illinois at Urbana-Champaign, for example, this can be seen in the reports, maps, and other material produced by the Prairie Research Institute (https://ideals.illinois.edu/handle/2142/10683). These studies of land use, wildlife populations, water quality, et cetera are available for all citizens of Illinois and beyond, and are extraordinarily valuable for those making decisions about stewardship of land, water, and wildlife. While a private institution, Columbia likewise "expects all areas of the university to advance knowledge and learning at the highest level and to convey the products of its efforts to the world," and the university administration looks to its IR to help in that effort.

The place of IRs in the teaching mission of the university or college is often underexplored or utilized. We believe that IRs have a role to play here specifically in supporting undergraduate programs. The Council on Undergraduate Research (CUR) defines undergraduate research as "an inquiry or investigation conducted by an undergraduate student that makes an original intellectual or creative contribution to the discipline" (http://www.cur.org/about.html). The use of IRs to disseminate and publish original undergraduate research is becoming a critical component of these formal undergraduate research programs. For example, IDEALS provides access to original student ethnographic or archival research produced through the Ethnography of the University Initiative (http://www.eui.illinois.edu/). Within this program, each student uses the archived work to ask their own research questions and design methodologies to answer those questions; they then can decide to publish this work in IDEALS. Similarly, at Columbia, Academic Commons, Columbia's IR, provides the home for the issue briefs created by all the students year after year in an ongoing political science research seminar called "Majority Rule and Minority Rights" (http://bit.ly/117q35v). The students build on the work from previous semesters while at the same time they add their own work to this growing and well-cited collection, work that would otherwise have

ended up in the professor's files somewhere, especially precarious in this case, as the professor in question is an adjunct. Academic Commons likewise provides a platform for publishing senior theses and other student work (http://bit.ly/13r0Vty), the most-used content within the collection in terms of number of views per item (Figure 1). Thus the IR becomes a critical part of building the student's understanding of the scholarly research publishing cycle and allowing them a place in it.
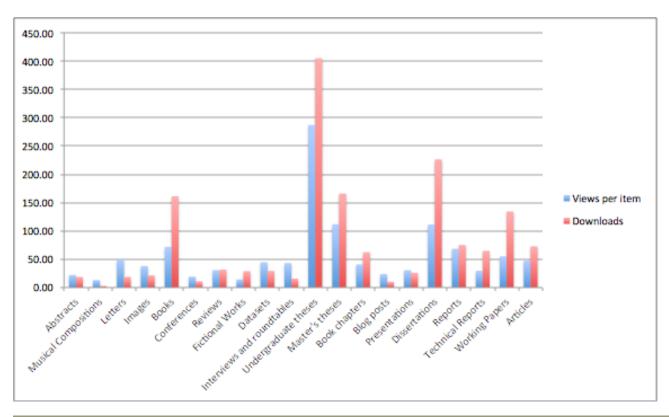
This is not to say that we believe that IRs should become places where any and all institutional documents should go. At both Illinois and Columbia, IDEALS and Academic Commons, respectively, explicitly do not accept administrative documents or institutional records that might more properly be maintained by the university archive, maintaining instead a focus on material that reflects the intellectual life of the university. The descriptions of both IRs make this distinction, that the content within the IR a visitor can expect to find is to faculty-, student-, and staff-produced "research and scholarship" (IDEALS) or "scholarly work and research" (Academic Commons). Our IRs might therefore be more properly designated

as "research" rather than "institutional" repositories (see Kennison, 2008), although we claim the broader term in this piece. We would venture to say most IRs are likewise research repositories, and, like ours, go well beyond solely including "green OA" materials, although only rarely are they what might be called fully *institutional* repositories.

All of these examples are to illustrate that shifting the primary purpose of the IR from green open access to providing persistent and reliable access to the full range and diversity of scholarly output of a research institution can mean a vibrant, well used, and well understood system. As studies have shown (e.g., Charbonneau & McGlone, 2013; Gargouri et al., 2010; Poynder, 2012; Xia et al. 2012), without teeth (e.g., funder monies withheld), a mandate does not result in more content being deposited than does unmandated deposit. Positioning the IR in the way we have described, by including in it all content that any researcher may wish to preserve, connects it much more closely to the needs and workflows of faculty, departments, research centers, and students; it becomes part of the fabric of the university. Then, if and when a mandate comes into play —whether institutional or funder—the IR, already

Figure 1. Columbia University Academic Commons Content Use
Average Per Item Views and Downloads by Content Type in Columbia University's Academic Commons, April 1, 2012–March 31, 2013

a known and trusted entity, is in the best possible place to aid in the fulfillment of that mandate.

Rebecca Kennison and Sarah Shreeves serve on the JLSC editorial board. The opinions expressed here are those of the authors, and do not necessarily reflect JLSC's position.

REFERENCES

Charbonneau, D. H., & McGlone, J. (2013). Faculty experiences with the National Institutes of Health (NIH) public access policy, compliance issues, and copyright practices. *Journal of the Medical Library Association*, *101*(1), 21–25. http://dx.doi.org/10.3163/1536-5050.101.1.004

Crow, R. (2002). *The case for institutional repositories: A SPARC position paper*. Washington, D.C.: Scholarly Publishing & Academic Resources Coalition. Retrieved from http://sparc.arl.org/sites/default/files/ir_final_release_102.pdf

Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., Brody, T., & Harnad, S. (2010). Self-selected or mandated, open access increases citation impact for higher quality research. *PLoS ONE*, *5*(10): e13636. http://dx.doi.org/10.1371/journal.pone.0013636

Harnad, S. (1999). Free at last: The future of peer-reviewed journals. *D-Lib*, *5*(12). Retrieved from http://www.dlib.org/dlib/december99/12harnad.html

Kennison, R. (2008). Much ado about everything: data, publications, and the role of repositories. Presentation at *American Psychological Association's Fourth Annual Science Leadership Conference*, 4 October. Retrieved from http://hdl.handle.net/10022/AC:P:29740

Lynch, C. A. (2003). Institutional repositories: essential infrastructure for scholarship in the digital age. *ARL: A Bimonthly Report*, 226 (February 2003), 1-7. Retrieved from http://www.cni.org/wp-content/uploads/2003/02/arl-br-226-Lynch-IRs-2003.pdf

Poynder, R. (2012). Open access mandates: ensuring compliance', Open and Shut? 18 May. Retrieved from http://poynder.blogspot.com/2012/05/open-access-mandates-ensuring.html

Xia, J., Gilchrist, S. B., Smith, N. X. P., Kingery, J. A., Radecki, J. R., Wilhelm, M. L., . . . Mahn, A.J. (2012). A review of open access self-archiving mandate policies. *portal*: *Libraries and the Academy*, *12*(1),85-102. http://dx.doi.org/10.1353/pla.2012.0000

Ordering Institutional Repository Priorities vs. Breaking Through Open Doors

Stevan Harnad *Canada Research Chair in Cognitive Sciences, Université du Québec à Montréal*

Let me start by pointing out the part of Kennison & Shreeve's (K&S's) argument with which one can easily agree: Yes, Institutional Repositories (IRs) can and should be used for other kinds of content besides an institution's own peer-reviewed journal article output—including its grey literature, theses, undergraduate research, and research data, along with their digital preservation. Why not? There's plenty of empty space in IRs, and these are all valid and worthwhile contents.

But over a decade has gone by since those "most influential and visionary early essays" recommending IRs "for the stewardship of a wide range of institutional output (Lynch 2003; [cf. Harnad 2003])" and as "ways for libraries to support publishing functions (Crow 2002; [cf. Harnad 2002])." If the presence of those other kinds of content did make "libraries...better placed to implement green OA...mandates," then that

effect has not been very big, because most IRs still have a lot of empty space where institutional peer-reviewed journal article output ought to be. (Library publishing function support remains close to nil—but that continues to be irrelevant to the problem of making institutional refereed research output OA.)

There is also nothing to disagree with in K&S's belief "that provision of open access to [an institution's own peer-reviewed journal article output] is an important and critical role of the IR." But we may have more than a semantic disagreement about what K&S mean by "priorities," and we certainly have a substantive disagreement about the order of those priorities, as well as a profound strategic disagreement about how to achieve them.

K&S correctly state that "[t]he argument...is [about]

whether it should be the *primary* role of the IR" to provide OA to the institution's own peer-reviewed journal article output. K & S believe that "even in the case of an institutional mandate… [OA] should only be one priority of the repository, not its sole purpose." And there's the semantics: Arguing that OA should be an IR's *first* priority is not the same as arguing that it's an IR's *sole* priority.

So let's get to the point: the reason OA to refereed articles should be the first priority for an IR is that OA is (1) urgent and (2) faces obstacles from publishers, whereas depositing the grey literature, undergraduate research, and research data faces no opposition from publishers: the doors are wide open, and authors need merely walk in, if they wish. With articles, they feel they can't. (Theses are a special case, because making them OA may reduce their chances of being accepted for publication as a book, but there the solution is simple and obvious: deposit them as Restricted Access (author only) instead of OA if you plan to try to publish them as a book).

The reason OA is urgent is that potential research uptake, usage, and impact[1]—hence applications, progress and productivity—are being lost, daily, cumulatively, some of it probably irretrievably, because the only users with access to journal articles are those whose institutions can afford subscription access to the journals in which the articles are published (and most institutions cannot afford access to most journals[2]). This needless access-denial and research impact loss has been going on ever since the online medium has made it possible to put an end to it. Among the losers are not just research, researchers and their institutions, but the tax-payers who fund the research.

The obstacles that OA faces are journal OA embargoes[3] and authors' fears that providing OA violates copyright. There are no such obstacles for grey literature, undergraduate research, and research data (nor for theses not aspiring to be published as books). Yet there are ways to overcome both obstacles—if institutions adopt the right OA mandate.

K&S write that "many libraries have concentrated on green OA to fill their IRs—with or without mandates, always with mixed success." (K&S don't explain why it's "green" OA, and it's not even clear why they use the

color-term, originally coined to distinguish "green" OA self-archiving from "gold" OA publishing, which K&S do not go on to discuss at all.) It may be true that many libraries without mandates have concentrated on filling their IRs with OA articles without success. But it can't be true that many have concentrated on filling them with mandates, because out of the more than 2500 universities and colleges in the US, only 25 of them[4] (1%) as yet have any OA mandate at all, let alone the right OA mandate (Gargouri et al 2012).

K&S rightly note that "as studies have shown,…a mandate…without teeth (e.g., funder monies withheld)…does not result in more content being deposited than does unmandated deposit." But instead of going on to propose concentrating on the right solution—promoting the adoption of mandates with "teeth"—K&S recommend concentrating on filling IRs with the other kinds of contents mentioned until it "becomes part of the fabric of the university…Then, *if and when a mandate comes into play* [emphasis added]— whether institutional or funder—the IR, already a known and trusted entity, is in the best possible place to aid in the fulfillment of that mandate."

The reasoning here is not apparent to me (though it does call to mind the proverbial persistent search for the key by the well-lit lamp-post). Yet there's no reason to be in the dark about the right strategy; it's known, it's been tested, and it works:

As the sole means of submission for institutional performance assessment, and as a condition of research agency funding, it needs to be mandated that the final refereed draft of all articles must be deposited in the author's institutional repository *immediately upon acceptance for publication* (Harnad 2011; 2013a,b). Setting access to the deposit immediately as OA should be strongly recommended, but if the author wishes to comply with a publisher OA embargo, access to the immediate-deposit can instead be set as Restricted Access (author only). During the embargo, the repositories have a facilitated eprint-request Button with which users can request and authors can provide an individual eprint for research purposes with one click each (Sale et al 2012).
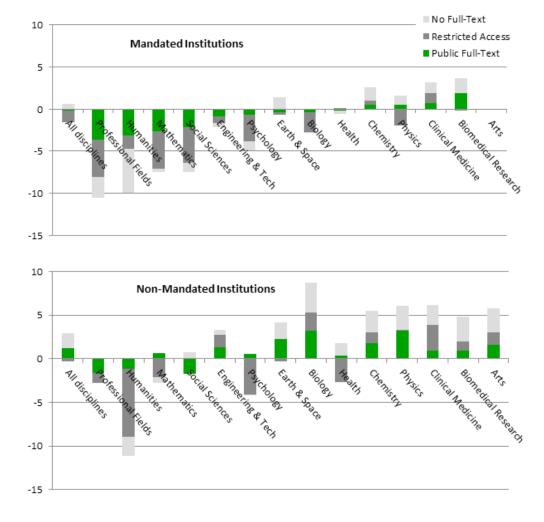
The immediate-deposit mandate (Rentier & Thirion 2011)—called the "Liège model," after the first

university to adopt it—works. It raises the deposit rate from the baseline for unmandated deposit (about 20%) to 60% within a year or two, and then it continues to climb toward 100%. The University of Liège[5] immediate-deposit mandate is complemented by an immediate-deposit mandate by the Belgian funding council, FNRS[6] (Fonds de la Recherche Scientifique). This is the mandate model recommended by BOAI-10[7] as well as HEFCE[8] (Higher Education Funding Council for England) and BIS[9] (Department for Business Innovation & Skills) in the UK. Immediate-deposit is also a clause in the Harvard[10]/MIT[11] (copyright retention) mandate.

All the evidence suggests that there is no point in just continuing to collect other kinds of contents in the hope that they will somehow lead to an OA mandate and compliance. The first and foremost priority of those who seek to fill IRs with their primary intended content should be to work toward the adoption of the Liège model mandate by their institutions as well as their funders and then to implement an effective monitoring system to ensure compliance: Alongside their refereed final drafts, authors should be asked to deposit the dated acceptance letter so as to verify immediate deposit (within, say, six weeks of acceptance). Evidence already suggests that compliance will be timely (Figure 1).

Figure 1. Average Deposit Latency for Unmandated and Mandated UK Repositories for Publication Year 2012, By Discipline

The average deposit date minus publication date for Public FullText (OA), Restricted Access, and metadata only (No FullText) deposits in UK repositories was +3 months for unmandated deposits and -1.8 months for mandated deposits. Titles for all articles published in 2012 by UK first authors indexed by Thomson-Reuters Web of Science (WoS) were searched for in all UK Institutional Repositories indexed by ROAR.[12] Results are presented by discipline. Note that publication dates are later than acceptance dates and actual date of appearance may be even later. Note also that this figure only shows size of delay, not the number of deposits; however, for Arts there were no Mandated deposits at all in this sample. (Data from Gargouri, Y., Lariviere, V., Gingras, Y., Brody, T., Carr, L., & Harnad, S., in preparation)

IN-TEXT LINKS

[1] http://opcit.eprints.org/oacitation-biblio.html

[2] http://fisher.lib.virginia.edu/cgi-local/arlbin/arl.cgi?task=setupstats

[3] http://www.sherpa.ac.uk/romeo/

[4] http://roarmap.eprints.org/cgi/search/archive/advanced?screen=Search&dataset=archive&eprintid=&organisation_merge=ALL&organisation=&keywords_merge=ALL&keywords=&datestamp=&geoname=geoname_2_US&type=institutional_mandate&satisfyall=ALL&order=-datestamp&_action_search=Search

[5] http://roarmap.eprints.org/56/

[6] http://roarmap.eprints.org/850/

[7] http://www.budapestopenaccessinitiative.org/boai-10-recommendations

[8] http://www.hefce.ac.uk/media/hefce/content/news/news/2013/open_access_letter.pdf

[9] http://www.publications.parliament.uk/pa/cm201314/cmselect/cmbis/99/9902.htm

[10] http://roarmap.eprints.org/75/

[11] http://roarmap.eprints.org/122/

[12] http://roar.eprints.org/

REFERENCES

Crow, R. (2002). *The case for institutional repositories: A SPARC position paper*. Washington, D.C.: Scholarly Publishing & Academic Resources Coalition. Retrieved from http://sparc.arl.org/sites/default/files/ir_final_release_102.pdf

Gargouri, Y., Lariviere, V., Gingras, Y., Brody, T., Carr, L., & Harnad, S. (2012). Testing the Finch Hypothesis on green OA mandate ineffectiveness. *arXiv* [preprint] arXiv:1210.8174. Retrieved from http://arxiv.org/pdf/1210.8174

Harnad, S. (2013a). Comments on HEFCE/REF Open Access Mandate Proposal. Open access and submissions to the REF post-2014. Retrieved from http://eprints.soton.ac.uk/349893/

Harnad, S. (2013b). Evidence to BIS Select Committee Inquiry on Open Access. *Written Evidence to BIS Select Committee Inquiry on Open Access*, Winter Issue. Retrieved from http://eprints.soton.ac.uk/348483/

Harnad, S. (2011). What is to be done about public access to peer-reviewed scholarly publications resulting from federally funded research? (Response to US OSTP RFI). Retrieved from http://eprints.soton.ac.uk/273080/

Harnad, S. (2003). Cliff Lynch on institutional repositories. *American Scientist Open Access Forum* 2744. Retrieved from http://users.ecs.soton.ac.uk/harnad/Hypermail/Amsci/2744.html

Harnad, S. (2002). Comments on Raym Crow's (2002) SPARC position paper on institutional repositories. *Open Access Archivangelism* 671. Retrieved from http://openaccess.eprints.org/index.php?/archives/671-guid.html

Lynch, C. A. (2003). Institutional repositories: essential infrastructure for scholarship in the digital age. *ARL: A Bimonthly Report*, 226 (February 2003), 1-7. Retrieved from http://www.cni.org/wp-content/uploads/2003/02/arl-br-226-Lynch-IRs-2003.pdf

Rentier, B., & Thirion, P. (2011). The Liège ORBi model: Mandatory policy without rights retention but linked to assessment processes. Retrieved from http://orbi.ulg.ac.be/handle/2268/102031

Sale, A., Couture, M., Rodrigues, E., Carr, L., & Harnad, S. (2012). Open access mandates and the "fair dealing" button. In: Rosemary J. Coombe & Darren Wershler (Eds.), *Dynamic fair dealing: Creating Canadian culture online*. Retrieved from http://eprints.ecs.soton.ac.uk/18511/