Guest Editorial

Updating the Agenda for Academic Libraries and Scholarly Communications

Clifford Lynch

January 19, 2017

This issue of *C&RL* is focused on scholarly communication, and it seems appropriate, in this invited guest editorial, to step back and examine the broader agenda that academic and research libraries need to consider today in engaging with scholarly communications as a way of framing the issue. My view is that this agenda is ripe for re-thinking. The overall environment has changed significantly in the last few years, underscoring the growing irrelevance of some long-held ideas, and at the same time, clearly identifying new and urgent priorities. What I hope to do here is to summarize very succinctly my thoughts on the most pressing issues and the areas most needing reconsideration. Articles in this issue touch upon aspects of many of these topics; I hope that future authors may also find topical inspirations here. You'll note that many of these are issues that have been important to the CNI agenda in recent years, and I've included a few references to some of these materials.

In the United States and elsewhere both public and private research funders have fundamentally altered long-standing discussions about open access (OA) with their various public access or OA mandates. In the US the idea of a relatively *comprehensive* "green" (local institutional repository [IR] based) OA to the scholarly journal literature through author self-archiving is now pretty much unrealistic at most institutions;[1] funder (and funder-blessed disciplinary) journal article repositories will play this role, with an additional role for publisher-based initiatives like Chorus which capitalize on the fact that the most natural and useful place to provide public access for articles is going to be in the context of the venue where they were published. Funder mandates are not going to force faculty or publishers to populate institutional repositories, and in most cases institutional policies alone have so far seen very limited success; note also that these institutional policies are the only way to ensure deposit of most unfunded research (extensive in the humanities, some social sciences and other areas). Funders to date have shown little interest in enabling or mandating large-scale automatic replication from funder or funder-blessed repositories to institutional ones. It is at best unclear if institutionally-driven initiatives to automatically replicate from mandated repositories to institutional ones is likely to be either legally or technically feasible at scale, even after embargo periods have expired and articles are open for public access. Developments like SHARE and institutional research information systems (CRIS) will certainly allow automated IR population for *metadata* covering a growing proportion of faculty publications.

At least in the near term, much of this public access may be delayed by up to a year; though it is clear that top quality OA journals, typically funded though author fees, are now well established and flourishing. While it would be wonderful to have all faculty research published in immediate OA journals, it's very unclear how to fund this shift,[2]

doi:10.5860/crl.78.2.126

or how to motivate faculty to do so (though certainly education and advocacy has had some genuine and steady success), and I am skeptical about our ability to "flip" commercial journals at any scale, at least in the near term. There have certainly been many successes with scholarly societies that have decided to move to open access for some or all of their journals, which are often among the most respected in their disciplines. I suspect that we'll see slow, gradual progress towards even more widespread journal article OA in the coming years.

I want to be very clear here. I believe articles absolutely have a place in the IR for faculty who want to place them there; but existing US funder mandates are not going to get them there en masse, and, even in the absence of strong institutional policies and high compliance with the policies, even libraries that are willing to do the work can't populate the local IRs due to rights and permissions barriers. Some faculty will doubtless want to work with the library to build and curate as comprehensive a collection of their own works as possible in the IR, or of selected works on a particular topic, and they will be mindful of these barriers when selecting and dealing with publishers; the library may also be able to help negotiate permissions for older, already published works. We must honor, welcome and support this activity. Such faculty collaboration is important and very desirable, and it even makes sense to reach out to faculty and encourage them to do this if they are willing; it is also a great way to bring in, and to enhance the value of both the journal articles and other complementary material that at present is going unstewarded and that I think ought to be strategic priorities for inclusion in the IR.

While on the subject of opening up the journal article, I note that whatever policies are put in place going forward, there's a huge mass of historic material locked up under copyright and where funders, authors, and academic institutions have very limited leverage other than paying to license the materials for their local communities. We need to start talking much more seriously about how to make this older material publicly accessible, and indeed to better understand the current state of such public access (which has been opened up by some scholarly societies, for example, and by some publishers).[3]

Is it important for institutions to maintain a local comprehensive record of their scholarly output through their IR? This is likely to be very expensive and the coverage is likely to be spotty, and I am skeptical in most cases of its value; it also, I think, absolutely requires very strong institutional faculty OA mandates and a very strong culture of compliance with and support for these mandates. Perhaps a few institutions can succeed with this. It is interesting to note that the recent report of the task force on the future of the MIT libraries[4] strongly reaffirms the vision of the institution maintaining such a comprehensive record, and perhaps MIT (and a handful of other institutions) can successfully construct this and effectively exploit it with the future systems that they envision. I really hope they succeed and once more "invent the future at MIT." But how common and typical will such an undertaking be, at least in the near future?

For most institutions, I think this is not a strategic near-term priority and isn't going to be a good investment. The linkage between journal article open access and institutional repository agendas has been a mistake, and one that has resounded to the detriment of both agendas.

And this takes us to a re-consideration of the purposes of IRs. I believe these must be disconnected from the OA agenda for journal articles, and re-positioned in the broader context of managing and preserving institutional community assets. There is so much that needs to be stewarded; a large-scale random sampling of local faculty journal publications is not, in my view, the strategic priority for the institution. IRs can play a huge role in facilitating and expediting the transition to scholarly communication that

is genuinely designed for the digital environment; many of these works are not well accommodated or cared for by established journals today. They can be instrumental in advancing the electronic theses and dissertations movement, or the growth in open educational resources.

Preprint servers (such as arXiv) have played a transformative role over the last 20 years in accelerating the creation and dissemination of knowledge, and in democratizing these processes. Recently the idea of preprint servers seems to be gaining new traction in scholarly disciplines that have previously largely rejected them. Anything that academic and research libraries can do to accelerate and support this trend will have very high payoffs in terms of better, more effective scholarship.

Before I leave scholarly journals, let me also note that licensing these materials is now not only a very complex economic negotiation with publishers, but must also incorporate many other important non-economic contractual terms dealing with topics ranging from reader privacy to usage reporting[5] and statistics, archiving policies, accessibility, and even the growing dangers of malware being vectored through third party advertising that may be hosted on publisher sites. We could fruitfully give more attention to sharing issues and best practices in these areas.

There are many other areas related to open access that urgently need attention and advocacy. One obvious area involves access and re-use policies for institutional special collections, museums, and related materials. A few institutions have taken really important early leadership positions in this area, such as Cornell. But this is an institutional policy choice (as opposed to an effort to gather faculty support) that should be relatively straightforward, should be commonplace, and it's a choice that can have a huge impact. This should include revisiting terms and conditions for accepting new special collections that facilitate digitization, access and reuse. University press policies in areas like quotations permissions and fair use are also fair game here.

A second area is developing a collective vision (driven by scholars, but with the active engagement of libraries and publishers) about what we expect or want in terms of OA to scholarly monographs, both prospectively and retrospectively. Is immediate or delayed OA to newly published monographs even a desirable goal? If so, how can we get there in terms of business models? I do not believe there is anything close to a rough consensus on the first question, and certainly there are a huge number of largely untested proposals for new and mostly untested business models.

Related to this is the movement to shift university press management to the library, and the overall growth of library publishing (or dissemination) programs for scholarly works.[6] There is a great deal of work to be done in figuring out how university presses should fit into broader university strategies for the global dissemination of scholarly work, and in fully understanding how to think about university press budgets in this new environment.

Let me move on to what I think are the really serious crises. I think that academic and research libraries need to be spending a lot more time considering the changing nature of the scholarly record, the broader cultural record that underlies it and that enables future scholarship, and how we can collectively exercise effective long-term stewardship over this. There's a huge problem with public or OA materials on the web: everybody relies on them, but nobody wants to take responsibility for curating and preserving them. The essential work of EDINA's Keepers Registry and the broader Keepers Community here is enormously critical in tracking and quantifying this situation: I cannot urge our community strongly enough to find ways to track, support and advance this program.[7] Here I'd also note that the evidence suggests that the traditional commercial scholarly journals are pretty safe from a stewardship perspective, but the new components of scholarly communication–video, data, software,

OA journals, websites, blogs, perhaps preprint servers–are typically very much at risk. These belong in institutional, disciplinary, and stewardship community repositories, and they are going to be increasingly important parts of the scholarly record going forward. It's urgent to develop better strategies for these materials. There are very interesting experiments such as library-led institutional engagements with the Open Science Framework from the Center for Open Science,[8] or connections linking digital preservation and digital humanities or broader digital scholarship centers with the library to not only disseminate but also to preserve faculty work.

Research Data Management is clearly an enormous ongoing problem, and not just because of funding challenges. We need much more serious work on discovery, and description (metadata) that advances discoverability. We need to get realistic about the need to re-appraise datasets periodically in deciding what to keep, and who should keep it. And we must move conclusively beyond the current focus on open or publically accessible data; the reality is that a very large proportion of data in the biomedical and social sciences, and a great deal of other data, cannot be simply made public. The challenges are to share it as effectively and frictionlessly as possible, to ensure that it is preserved, and that it is not ultimately orphaned and rendered useless by privacy constraints. Some progress has been made in biomedicine and genetics, but there is tremendously more work to do, and this must begin by abandoning the notion that open data is a panacea. Open data is great when it works, but often it can't.

Moving beyond the scholarly record, there's an enormous crisis in preserving and curating the broader cultural record that will be needed to support future scholarship: social media, news in the digital age, and so many more things: popular e-books for example, as well as the new digital "gray literature." We don't even have any good ways to measure how badly we are failing at this. The current copyright laws are going to be part of the problem here because they totally fail to recognize the broad societal need to preserve the digital cultural record as an essential and critical priority; dealing with this is going to be a very broad public policy discussion and advocacy challenge, not a narrow and technical series of legal arguments, and we need to be prepared to advance this discussion.

I worry a great deal about quantitative metrics (both well-established and so called "alt-metrics"). Some nations have become intoxicated with these (the UK and Australia, for example), and they are gaining increased attention by administrators and funders in the US. While bibliometrics (webmetrics to use the newly-fashionable name) are of genuine ongoing interest as fascinating research tools for understanding information flows within scholarly communities (among other phenomena), I am deeply concerned about the potential quantification of scholarly impact.[9] This is a terrible, and morally bankrupt, idea. There are also more technical issues; for example, I think it's essential that any metrics that receive broad adoption be fully transparent and reproducible.[10]

Finally, in setting an agenda for scholarly communications, we need to be profoundly mindful that for virtually all faculty and graduate students, the dissemination of their scholarly work has become a complex, confusing, time-consuming morass of funder mandates, institutional policies, choices about publishing venues, article processing charges, and questions such as whether or not to release preprints at various stages of the development of their work. We have got to find ways to simplify and streamline this mess and to honor the increasingly scarce time of researchers, hopefully in ways that lead to very desirable outcomes for the broad scholarly community and for stewardship institutions. For example, could we work with publishers to badge journals as meeting various funder and institutional compliance policies automatically on behalf of authors publishing in these venues (assuming that authors identify their institution and funders, which the usually do already)? A number of publishers already do this

for authors with regard to funder mandates, but it isn't always clear to authors. Rather than just telling faculty authors about complex funder mandates that they must respond to in interminable detail, could we just tell them that if they publish in some particular set of journals everything is taken care of? If it's too difficult for publisher badging to include institutional policies as well as funder mandates, could the local institution develop subsets that meet these requirements and list them as recommended, no-worries venues? I believe that we need a lot more creative thinking along these lines.

In closing, I want to remind readers of the truly central challenge and opportunity for our era: to develop appropriate new genres of scholarly communication for the digital environment. Scholars own this challenge, though librarians (and many others) can help. But libraries must recognize the emergence of these genres in a timely manner, and respond by collecting, organizing, curating, and preserving them. This must be a core part of the scholarly communications agenda, and will certainly be one of the most difficult, starting with the problem of recognizing emergence. But it must not be overlooked.

My thanks to Joan Lippincott and Diane Goldenberg-Hart for their help with this.

Notes

1. See Richard Poynder, "Q&A with CNI's Clifford Lynch: Time to Rethink the Institutional Repository?" *Open and Shut?* Sept. 22, 2016, http://poynder.blogspot.co.uk/2016/09/q-with-cnis-clifford-lynch-time-to-re_22.html.
2. MacKenzie Smith, "The Cost of Open Access to Journals: Pay It Forward Project Findings," Presentation at the Coalition for Networked Information Fall 2016 Membership Meeting, https://www.cni.org/topics/assessment/the-cost-of-open-access-to-journals-pay-it-forward-project-findings.
3. See the 2004 *Washington DC Principles for Free Access to Science*, www.dcprinciples.org.
4. *Institute-wide Task Force on the Future of Libraries: Preliminary Report*, October 24, 2016, https://future-of-libraries.mit.edu/sites/default/files/FutureLibraries-PrelimReport-Final.pdf; see also *The Future of Libraries* website, https://future-of-libraries.mit.edu/.
5. See Clifford Lynch, "The Rise of Reading Analytics and the Emerging Calculus of Reader Privacy in the Digital World," forthcoming.
6. See *Institutional Strategies and Platforms for Scholarly Publishing*, Report of a CNI Executive Roundtable Held Dec. 10, 2012 (Feb. 2013), https://www.cni.org/wp-content/uploads/2012/12/CNI-Executive-Roundtable-Report-December-2012.pdf; Mary Rose Muccie, et al, "Across the Great Divide: Findings and Possibilities for Action from the 2016 Summit Meeting of Academic Libraries and University Presses with Administrative Relationships (P2L)," http://www.arl.org/storage/documents/across-the-great-divide-2016-p2l-summit.pdf.
7. See https://thekeepers.org/about; Peter Burnhill, et all, "Transnational Strategies for Stewardship of Our Shared Scholarly Record (and of Each Nation's Published Heritage): Both Open and Subscribed Content," Presentation at the Coalition for Networked Information Fall 2016 Membership Meeting, https://www.cni.org/topics/digital-preservation/transnational-strategies-for-stewardship-of-our-shared-scholarly-record-and-of-each-nations-published-heritage-both-open-and-subscribed-content; Peter Burnhill, "Taking the Long View: International Update from the Keepers Registry," Presentation at the Coalition for Networked Information Fall 2015 Membership Meeting, https://www.cni.org/topics/digital-preservation/taking-the-long-view-international-update-from-the-keepers-registry.
8. Matthew Spitzer, "An Open Science Framework for Solving Institutional Challenges: Supporting the Institutional Research Mission Across Departments and the Full Project Lifecycle," Presentation at the Coalition for Networked Information Spring 2016 Membership Meeting, https://www.cni.org/topics/ci/an-open-science-framework-for-solving-institutional-challenges-supporting-the-institutional-research-mission-across-departments-and-the-full-project-lifecycle; Andrew Sallans and Natalie Meyers, "The Open Science Framework (OSF) at Notre Dame: Connecting the Workflow and Supporting the Research Mission," Presentation at the Coalition for Networked Information Fall 2015 Membership Meeting, https://www.cni.org/topics/scholarly-communication/the-open-science-framework-osf-at-notre-dame-connecting-the-workflow-and-supporting-the-research-mission.
9. Yves Gingras, *Bibliometrics and Research Evaluation: Uses and Abuses,* MIT Press, 2016.
10. See the work related to, and enabled by, the Jisc Open Citation Project, opencitations.net.