# Achieving the Full Potential of Repository Deposit Policies

**Karla Hahn, Assistant Executive Director, Transforming Research Library Roles in Research, Teaching, and Learning, ARL**

Editor's note: A small group of individuals with expertise on author-rights policies, the campus policy environment, National Institutes of Health (NIH) deposit processes, and digital repository services met in Washington DC on January 9, 2009, under the auspices of ARL's Public Policy and Scholarly Communication programs. The group explored opportunities, desired outcomes, and policy issues involved in developing capabilities for institutionally mediated deposit processes and content transfer between institution-based and funder-based repositories, particularly PubMed Central. Based on that discussion, the group also identified potential strategies that would lead toward creating the needed rights-management environment and repository services. This essay reflects the January 9 discussions.

## Background

Digital repository services are a key component of research-focused cyberinfrastructure. Institutions are individually and collaboratively developing the capability to house, manage, and preserve a wide range of products of the research process. While institutions are acting to develop repository services for their scholars and research, other digital repositories are evolving to serve as national and international resources for particular research fields. ArXiv and PubMed Central are prominent examples, each providing services based on different types of content that are highly valued by their research communities (physics and biomedicine, respectively). The number of independent repositories will continue to grow, elevating the need for greater sharing and harvesting of materials among repositories as well as broader coordination and searching across them.

Congress recently strengthened PubMed Central's ability to aggregate peer-reviewed, published, medical literature by requiring that authors of articles

based on NIH funding deposit their works in the repository. The development of complementary repository services by research institutions and federal agencies supporting research should further act to enhance support for the research process. However, the creation of multiple repositories that serve the same researchers raises many issues regarding how best to coordinate the content and functions of those repositories. The NIH Public Access Policy creates an initial impetus for libraries, as the common mediators of institutional repository services, to concretely examine the need to develop close interactions between a large disciplinary repository, PubMed Central, and repositories housed at research institutions.

> …promising strategies include standard author-publisher contracts, a uniform rights-transfer addendum, direct institution-publisher negotiations on behalf of authors, or an institutional policy granting it limited rights to institutionally affiliated authors' works as they are created.

The NIH Public Access Policy requirement for funded authors to deposit their works into PubMed Central has been in place since April 2008. During this period many libraries have developed various approaches to supporting authors in completing their deposits. However, there is more that many libraries would like to do toward developing services that smooth PubMed Central deposit for authors, assist universities in monitoring compliance, and allow capture and innovative uses of the deposited content. Already it has become evident that the technology issues involved in developing these kinds of services are relatively mundane. Rather, the substantive issue is the creation of appropriate copyright licensing regimes to support authors and institutions.

## Key Questions

To identify the issues and concerns that must be addressed to coordinate author-centric repository services at the institutional and funder levels, it is helpful to pose some key questions drawing on what we have learned from the first implementation of funder-imposed deposit requirements:

- **What are early experiences with PubMed Central deposits, and where are there opportunities for libraries to help make the NIH Public Access Policy as successful as possible?**

  Currently three main input streams provide content covered by the NIH Public Access Policy: authors, publishers providing published versions of

articles, and author versions submitted by publishers. The ingest processes for author manuscripts differ from those for the published versions. When a publisher submits an author manuscript, authors must still provide grant information and review and approve their manuscript's accuracy following NIH's standardization of document formatting, to complete deposit and comply with the policy.[1]

Many authors are submitting their own manuscripts with little difficulty. Libraries are assisting some authors, and are finding that the process is simple enough that most authors can more easily and expeditiously deposit their works themselves.

Relatively few publishers are participating in NIH's Full Participation, Portfolio, or Selective Deposit programs,[2] whereby they deposit published versions of articles. This is a well-developed process and, for those publishers participating, it relieves their authors from needing to go into PubMed Central to complete deposit.

Some publishers are passing the author version to PubMed Central along with contact information for the authors. Although these publishers are ensuring that deposit begins, many authors are failing to review their articles to allow completion of deposit, possibly from an incorrect belief that the publisher's transfer of the manuscript to PubMed Central completes the deposit process rather than merely beginning it.

Libraries and institutions could better assist their authors with deposit if they could be notified concurrently with the author or could mediate notification of the author about the final deposit steps needed. General education of authors regarding their responsibilities for completing the deposit process when publishers submit manuscripts on their behalf is another service libraries could provide.

- **How could institutional support for NIH submission be broadened to include facilitating deposits into institutionally based repositories?**
  Institutions could potentially develop submission streams that are comparable to those that some publishers have created. While they would have the same issue regarding authors' need to complete the deposit process, institutions may be better positioned to work with authors to complete all steps of the deposit process.

- ## How do challenges sort out in terms of policy and technology?

There appear to be few or no technology barriers to developing mechanisms for institutions to harvest or submit content. With regard to submission, many publishers have successfully implemented workflows to transfer works to PubMed Central, and it should be straightforward for institutions to develop parallel services.

Similarly, harvesting works from PubMed Central should be technologically uncomplicated. Although only a small proportion of deposited works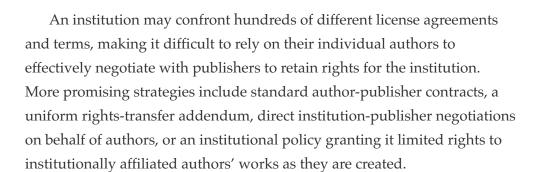 are open access, sufficient numbers are held in the archive to form a modest corpus of open access articles that repositories could begin harvesting and using for experiments with repository services based on harvested content.

Institutions (even though they are grantees) largely lack the limited rights they need to either submit or harvest works produced by their grant-funded authors. This concern is not limited to PubMed Central and the NIH policy, but would apply to any other funder's requirements. In fact, in many cases institutions do not necessarily have the limited copyright license they need to hold their authors' work in their own repositories.

> Repository services will achieve their full potential when they support the broadest possible dissemination of funded research and offer the fullest possible rights for reuse.

- ## What are the author rights required for repository deposit and how can institutions assist authors in conveying appropriate rights to institutions hosting repositories?

A wide range of rights-transfer agreements are used by publishers and these vary substantially regarding the extent to which authors retain the ability to grant their institutions limited licenses to store and disseminate their work through repositories. Some publishers grant authors the necessary rights automatically but many do not.

Institutions hosting repositories do not need the authors' full copyright or first-publication rights. They do need sufficient limited rights to hold, manage, use, and share works. Broad usage rights for the institution are important to support core activities around research and teaching. Reuse, text mining, and digital preservation are just a few examples of rights that are needed.

An institution may confront hundreds of different license agreements and terms, making it difficult to rely on their individual authors to effectively negotiate with publishers to retain rights for the institution. More promising strategies include standard author-publisher contracts, a uniform rights-transfer addendum, direct institution-publisher negotiations on behalf of authors, or an institutional policy granting it limited rights to institutionally affiliated authors' works as they are created.

- **Under what circumstances might institutional agents submit works for deposit on behalf of authors?**

Institutions (along with their authors) could benefit in various ways from mediating funder-mandated deposits. Ensuring compliance with funder requirements is a looming concern for grantee institutions. Demonstrating the institution's productivity and the value it is creating is another imperative for grantees.

> …actions leading to an environment with broad licensing of copyrights to various parties are an important early step toward the development of more powerful repository services.

Involvement with deposit and the ability to harvest authors' works allow institutions to capture important evidence of their success.

Institutions are rapidly acquiring diverse but often related content produced by their researchers and scholars. Research data, multimedia works, digital documents of all sorts, and new kinds of content emerging from the network context are increasingly going to be managed as institutional assets with a goal of encouraging their broadest possible use. The ability to hold, manage, and use manuscripts facilitates the development of an environment where the products of research process can be integrated, synthesized, and reused now and into the future.

## Strategies

In considering what best to do to strengthen authors' experiences with repositories and maximize the value and usefulness of articles in repositories, the overarching question is: What is the desirable future for repository services operating in an environment of funder expectations for public access to research results? Some answers that arose during the January discussion include:

- With a number of funders now requiring deposit of funded articles, an important measure of successful and effective repository services will be
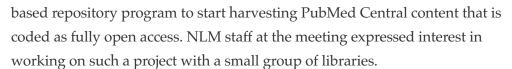
maximal compliance with requirements for repository deposit imposed by NIH, Canadian Institutes of Health Research, Wellcome Trust, Howard Hughes Medical Institute, and any future funder. The collective value of the deposited articles is greatest when there is full participation by funded authors and institutions.

- Repository services will achieve their full potential when they support the broadest possible dissemination of funded research and offer the fullest possible rights for reuse. Just holding and preserving articles in repositories, while valuable, is not enough.

- It is reasonable to expect authors to deposit articles resulting from research funding they have received, but their efforts will be most successful when they are able to rely on institutional capabilities to facilitate compliance with funder requirements and ensure ongoing dissemination of work through local repositories. Retaining copies of institutional-based works complements and feeds discipline-based collections and broadens experimentation, promotes service development, and enhances preservation of institutional assets.

- A culture of broad acceptance of granting limited license rights to institutions allowing them to obtain, retain, and disseminate copies of affiliated authors' works will be essential to enable institutions to support authors, act on their behalf, and work to ensure the broadest uses of funded research now and into the future.

## Actions to Pursue

Having looked at key questions and fostered agreement among the January meeting participants, several potential action arenas emerge:

1. Exchange of content between different repositories is a needed capability for research institutions, one combining both technology and rights issues. While rights issues limit much of what can be done to develop the technology infrastructure, there are some technology issues that could be addressed now. For example, one step toward expanding locally based repository services is to harvest content as it is deposited in a disciplinary repository. A small pilot project could be developed to explore the capabilities of existing repositories to capture PubMed Central content and identify high-value uses that can be made once the content is ingested into an institutionally based repository. It is already feasible for an institutionally

based repository program to start harvesting PubMed Central content that is coded as fully open access. NLM staff at the meeting expressed interest in working on such a project with a small group of libraries.

2. With copyright concerns limiting so much that can be done, actions leading to an environment with broad licensing of copyrights to various parties are an important early step toward the development of more powerful repository services.

One approach is to develop a framework of elements that describe the limited license rights that institutions need to support exchange of content between repositories, ongoing dissemination from repositories, and reuse of content to support further research and scholarly exchange.

Similarly, and possibly simultaneously, there is a need for a "universal addendum" for author-publisher agreements that facilitates the grant of a limited license to an author's funding organization and affiliated institution. This would be a valuable tool for creating the rights environment needed to move content between repositories and allow institutions to provide deposit services.

Another way to advance toward the desired copyright-sharing environment would be for libraries to engage in conversations with publishers about appropriate rights-management practices on behalf of the authors at their institution. One avenue where this could occur is through negotiations libraries engage in with publishers to license journal products.[3] Particularly with large publishers, including discussion of rights assignments for works authored by affiliates of the licensing institution could be an efficient approach.

In addition, as many journals are published by scholarly societies, this opens an opportunity for librarians to pursue conversations with campus faculty who are members of those societies to ensure that the societies understand the importance of granting limited licenses of author copyrights to academic and other institutions that support researchers and scholars.

There is a continuing role for advocacy, both nationally and locally, to preserve existing deposit requirements and expand opportunities for funded research to be placed in disciplinary and institutionally based repositories.

3. As researchers and institutions are beginning to adapt to an environment where article deposit is routine, it is also important to pursue steps that

advance compliance with funder requirements. Repositories cannot function effectively nor interact successfully where initial content submissions are not advanced to completion of the deposit process. A variety of strategies could promote maximum completed deposit rates. There is still a need for institutions to educate researchers regarding their responsibilities for participating in the deposit process, especially where a publisher is providing an author manuscript to PubMed Central—a step that starts but does not complete the deposit process. For instance, libraries can educate authors regarding the different mechanisms publishers use to contribute works on their behalf and the additional steps they need to take when their publisher is depositing author manuscripts rather than publisher versions. Institutions will also be able to better educate and support their researchers when they receive data on their compliance rates and the status of individual deposit processes. When institutions obtain such information, they can notify researchers regarding problems and work with researchers on completing deposits, or even complete deposits on researchers' behalf.

Another idea is that a library could submit manuscripts to PMC on behalf of its authors, similar to what some publishers are doing. However, in this case the submitting library would receive the notices from NIH's submission system about necessary reviews and approvals, in addition to (or instead of) their going to the respective authors. This would allow the library to follow up with the authors to ensure that they complete the process. NLM has indicated that it is willing to try this process with one institution and to make it more widely available if the experiment is a success.

One of the challenges hindering effective communication with authors and a more efficient deposit process is the lack of some basic tools for identifying authors and institutions. Promoting efforts to develop institutional and author identifiers would facilitate deposit, compliance tracking, rights management, and content exchange between repositories.

## Conclusion

Effectively balancing policy and technological developments is required to achieve the full potential of repositories to collect and disseminate new knowledge. If an integrated and interworking multi-repository environment can be created and operate within a copyright policy environment that allows

ongoing dissemination and reuse of content, there are opportunities to advance the research enterprise and share its fruits far more widely than has ever been possible.

The NIH Public Access Policy is advancing PubMed Central's repository services and together the policy and the repository infrastructure are beginning to demonstrate the value of new approaches to managing and sharing research results. However, a further suite of policy developments—at institutions and other organizations—will be required to allow the emergence of a next generation of linked repositories and services. The technology elements are largely already in place or relatively easy to develop. Creating the necessary environment for assigning limited copyrights to institutions will require a more complex and multi-faceted series of investments by a range of stakeholders, especially research institutions. Many of the most important next steps have now been clarified thanks to the ongoing implementation process for the NIH policy. What remains is to act on the ideas presented here. ARL will be working with member libraries on how best to move closer to the ideal repository environment, one that effectively incorporates the requirements of research funders as well as the interests of research institutions.

[1] For more details on the PubMed Central ingest process, see "Submission Methods," National Institutes of Health Public Access, http://publicaccess.nih.gov/submit_process.htm.

[2] For descriptions of these programs, see "What are a journal's options for depositing articles in PubMed Central?" in "PMC Frequently Asked Questions (FAQs)," PubMed Central, http://www.pubmedcentral.nih.gov/about/faq.html#q15. For a list of participating journals, see "Journals That Submit *All* NIH-Funded *Final Published Articles* to PubMed Central," National Institutes of Health Public Access, http://publicaccess.nih.gov/submit_process_journals.htm.

[3] See Ellen Duranceau and Ivy Anderson, "Author-Rights Language in Library Content Licenses" in this issue of *RLI*.

**To cite this article:** Karla Hahn. "Achieving the Full Potential of Repository Deposit Policies." *Research Library Issues: A Bimonthly Report from ARL, CNI, and SPARC*, no. 263 (April 2009): 24–32. http://www.arl.org/resources/pubs/rli/archive/rli263.shtml.