Introduction

This white paper summarizes and builds on a discussion of several key questions on roles and strategies for institutional repositories (IRs) at Cornell University. The idea for the forum originated with the recognition that we manage an increasingly distributed and complex suite of services and platforms, and that some key decisions will need to be made in the not-too-distant future. The HLM (Hospitality, Labor, and Management) Library is expanding its use of BePress’ Digital Commons to Hospitality. At the same time, the information technology group at CUL (CUL-IT) is examining Hydra¹ as a possible solution for managing multiple repositories and collections that are currently managed in separate systems, including legacy collections on the DLXS platform, eCommons, CULAR, and other collections. These decisions raise issues beyond technological ones, including matters of policy, service provision, scholarly communication, and even the general scope and purpose of digital repositories. Beyond Cornell, evolving mandates such as those related to the February 22, 2013 announcement that major US research funding agencies must develop policies to ensure free and open access to the outputs of the research they fund², may have significant implications for eCommons and other repositories, and suggest the time is also right to consider where IRs fit into open access strategies.

The purpose of the forum was to bring together a wide range of stakeholders to engage with these issues, identify themes, and raise new questions. The following questions were used as the basis for discussion:

1. What are the advantages and disadvantages of having a range of services (current situation) versus creating common services (such as a campus-wide DigitalCommons license)?
2. What are the current trends in IR from usability, functionality, and technology perspectives?
3. What are the motivating factors for CU faculty to use IRs? What are the discouraging factors?
4. What are the trends in creating and managing academic profiles of faculty (faculty pages vs. common IR pages)?

¹ https://confluence.cornell.edu/display/hydra/
² http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research
What are CUL’s goals and priorities in regard to capturing, disseminating, and archiving content created by Cornell’s scholars?

Participants formed small groups with each group addressing one question. Groups reported back and the discussion continued with the entire group. This paper attempts to capture and build on the key issues that emerged from that discussion.

March 4, 2013 Meeting Participants: Oya Rieger and Curtis Lyons (co-conveners), Mira Basara, Dan Blackaby, Femi Cadmus, Kathy Chiang, Jon Corson-Rikert, Jim DelRosso, Eileen Keating, Jason Kovari, George Kozak, Wendy Kozlowski, Dean Krafft, Holly Mistlebauer, Liz Muller, Jean Pajarek, Michelle Paolillo, Jaron Porciello, Steve Rockey, David Ruddy, John Saylor, Marty Schlabach, Gail Steinhart, Carissa Jean Vogel, Simeon Warner

Emergent Themes

We group the themes and ideas that emerged during the forum into three main areas: scope and purpose of institutional repositories, trends and issues in IR use and functionality, and a discussion of desired characteristics for IR infrastructure at CUL.

Scope and Purpose of Institutional Repositories

This issue was addressed specifically in the fifth discussion question, but it also permeated the discussions of virtually all the other questions. Even if we agree (and we may not) that it is a fundamental responsibility of CUL to ensure the continued availability of Cornell’s scholarly output, this basic statement raises multiple questions that touch on the library’s scope of collecting and choice of preservation strategies, as well as issues related to discovery and access:

- How do we define scholarly output and, based on that definition, what is in scope?
- Does ensuring availability of Cornell’s scholarly output mean that CUL must preserve a copy of everything, or simply know about copies held by other (presumably trustworthy) entities?
- What are CUL’s responsibilities and strategies with respect to ensuring discovery and access for scholarly output, particularly for content managed outside the library, or content with access restrictions of any kind?
- Given interdisciplinary and inter-institutional patterns in scholarship, how do we define “institutional”? What about unique content held in our special collections or other materials that are not limited to Cornell scholarly outputs?
- How will emerging open access mandates shape CUL’s strategic goals for IRs, especially within the context of subject repositories?
Defining scholarly output, defining what it is we need to preserve, and ensuring we’ve adequately captured it, raise still more questions. Policies and institutional obligations such as retention of university records that are increasingly born-digital (such as the course catalog) impose some specific collection requirements, and in some cases, security and access considerations that will be new for us. Even the question of what content we need to know about sometimes leads to the seemingly natural assertion that it is the library’s responsibility to maintain a record of the scholarly output of an institution, yet historically, we’ve maintained records only of what we hold in our collections. Regardless of how broad a responsibility we assume with respect to “knowing about” Cornell’s scholarly output, choosing to work with other entities to ensure preservation and access likely presents important economies of scale and relieves us of at least some of the responsibility for developing and supporting end-to-end preservation solutions. However, it also forces us to think carefully about how to support discovery and access, and challenges us when we wish to present coherent views of widely distributed content.

*Trends and Issues in Institutional Repository Use and Functionality*

The scholarly communication preferences and needs of faculty, and the factors that encourage or discourage them from using IRs, also present some important considerations. When library staff are available to do the work of depositing content in an IR, that may make that an attractive option for faculty. Copyright services, assistance with metadata, and the possibility of using IRs for gray literature and other content with no discipline-specific home can also make IRs attractive to faculty. Global, open access to IR content appeals to some faculty, while others wish to control access in some fashion, challenging the library’s mission to provide access as widely as possible. The possible relationship between IRs and reporting is another factor that can work for or against faculty interest in using them; connecting IRs and reporting systems can make the use of both more efficient, but faculty may also resist the notion of linking IRs and reporting systems.

Working against IRs are the facts that for some faculty they are relatively unknown, deposit may require some additional work on the part of faculty, and IRs may compete with preferred disciplinary alternatives. The Social Science Research Network (SSRN\(^3\)) is one such example. SSRN downloads are an important measure of impact in some disciplines, and faculty would be understandably reluctant to do anything that might decrease their SSRN download count. Finally, archiving their work may simply not be a priority for faculty who are satisfied with using personal websites to distribute content, and some administrators also do not consider it a priority. It’s also worth noting that while for the most part, we have worked to promote the use

of IRs, as use increases, the need to exercise the option of curatorial discretion may also increase, which will change how we market IRs to faculty.

We also have an opportunity to take a broader view of IRs and their utility. For example, some repository platforms have integrated publication functions (Digital Commons, for example, has a journal publication module). Even without that specific function, it’s possible to layer on top of and exploit IR infrastructure to support library publication activities. IRs could also be used to support the development of digital special collections, much as ILR makes use of Digital Commons to collect and make available material of interest to Cornell scholars, but created elsewhere.

The public presentation of faculty profiles presents another set of considerations. While this may seem unrelated to IRs, in fact, such profiles are an integral component of systems such as Digital Commons, and an optional feature in others such as Google Scholar, Mendeley, and Thomson-Reuters’ ResearcherID. Bibliographic information on researchers' publications form the heart of these systems. Cornell’s own VIVO also exposes information about publications, as well as institutional affiliations, grant information, and more. These systems play a role in establishing the public identity of researchers, reveal their publication and collaboration networks, can be useful for outreach, administrative, and reporting purposes, and can promote discovery of people, research activities, and their outputs.

Desired Characteristics of IR Infrastructure at CUL

We spent considerable time discussing the advantages and disadvantages of common architectures, and single and multiple repository instances. Related issues and choices include whether to build and manage systems locally or to contract for IR services, the advantages and disadvantages of specific platforms, and general approaches to IR architectures. Characteristics of IRs and their relationship to scholarly communication practices that motivate or discourage faculty to use them, as well as user and usability considerations, also informed our discussion, which can be distilled into the following priorities and issues for IR infrastructure and services:

- User needs must have a very high priority as we make these decisions. A user-driven and user-oriented approach to repository services and infrastructure is essential, if our repository strategies are to be successful. The term “users” encompasses both content creators and consumers of IR content.
- There is a general trend toward three-layer architecture for IR systems (a user interface layer, a middleware layer, and a repository layer) that can be developed and managed relatively independently. For locally maintained systems, this is a desirable direction in
which to move because platforms with this architecture are more sustainable and practical to develop and maintain.

- We need cost-effective solutions, which may argue in favor of a single platform (or at least a reduced number of platforms). A single platform also offers the possibility of pooling development efforts, providing uniform discovery and delivery, and better supporting discovery across multiple instances. Such a strategy may involve real trade-offs so we need to be able to prioritize our purposes and be ready to acknowledge compromises to scope, if needed.

- The ability to customize systems and interfaces to support distinct collections, branding, repository policies and diverse user needs is very important.

- Should we choose to do significant development in-house, we need to develop a realistic view of what we can support and sustain.

- That said, content users “don’t care where stuff is,” so IR infrastructure must support other means of discovery. There was some discussion of supporting search across multiple repositories, but it would probably be productive to consider whether we optimize discovery for users by focusing on cross-repository search, or by exposing repository metadata for discovery by more generic means.

- It is essential that repository design and function do not undermine faculty’s preferred methods of scholarly communication. This may mean performing collection and deposit functions for faculty rather than expecting faculty to do the work themselves, supporting a mix of local content and records of content stored elsewhere (such as SSRN), and offering repository data for reuse by other applications such as departmental or personal faculty web pages.

**Continuing the Conversation**

The forum conversations raised several important questions that would be helpful to consider as we continue to develop repositories and supporting services. Some questions and key pieces of information that could help advance our thinking and planning include:

- We found a fairly wide range of perspective on the scope and purpose of IRs at Cornell, their role vis-à-vis subject repositories (see earlier comments about SSRN on pages 3 and 5 of this document), as well as scope of content for IRs. We would benefit from a thorough and thoughtful consideration of our scope and purpose (multiple scopes, more likely) in establishing and maintaining digital repositories. Specifically, we need to decide both what we will accept (passive recruitment of content) and what we will actively collect (active recruitment of content), and allocate resources in such a way that those strategies are fully supported. Such an approach will require that we
move away from a narrow IR approach and assume a broader and more inclusive ‘digital repositories’ framework.

- CUL’s plan needs to be informed by actual and emerging needs of users as well as the Library’s vision and goals in this program area. Also critical will be to consider the long-term maintenance and development issues upstream before we develop a new system. Now that the initial CUL-IT Hydra assessment process is completed, the next step will be forming a group with representation from both IT and services providers to discuss an implementation and maintenance plan.

- DigitalCommons has proven itself to be an effective system and emerged as an important library service area at ILR and the Law School. Due to the economies of scale, CUL should consider the virtues and impediments of acquiring a campus-wide license so that other units can utilize the system for their needs. This decision will also be based on the results of a requirements analysis that will identify how Hydra can expand (or fulfill) the services currently provided by eCommons and DigitalCommons.

- The interplay between repositories and other systems such as researcher profiling (and discovery and access) systems is an interesting area that deserves some deliberate investigation and attention.

- A repository is only as good as its content. We must continue discussing and experimenting with service models and workflows that will allow us to maximize our content and acceptance to faculty, while keeping in mind cost limitations.

- It is not just content that makes an IR useful. Metadata can reveal and connect resources in important ways, and time and energy will need to be put towards the creation and management of metadata.

- We will continue to be faced with the need to make decisions about repository projects, in advance of the possibility of a more unified approach. Ideally, such decisions take future (prospective) developments into account, and allow for maximum flexibility.

Forum participants were enthusiastically engaged in the conversation. Several participants also noted the importance of engaging users directly. We clearly have a community of interested stakeholders, even just within the library. Possible approaches to continuing the conversation include holding another forum in the future, or possibly establishing a CUL repository advisory group or task force in order to encourage an ongoing strategic discussion and planning process. The newly established Scholarly Communication Working Group and the existing CUL groups such as the Data Executive Working Group will be instrumental in encouraging an ongoing strategic discussion and planning process to continue the conversation.4

---

4 The Scholarly Communications Working Group will provide leadership in developing, maintaining, and assessing a library-wide program in scholarly communications, including program areas such as publishing, intellectual property, digital repositories, open access, and digital media creation and preservation. In addition, this group will develop and support professional development opportunities and assess the scholarly communications needs of the Cornell community.