University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Copyright, Fair Use, Scholarly Communication, etc.

Libraries at University of Nebraska-Lincoln

5-2017

Rethinking Institutional Repository Strategies

CNI Executive Roundtable

Clifford Lynch

Follow this and additional works at: https://digitalcommons.unl.edu/scholcom

Part of the Intellectual Property Law Commons, Scholarly Communication Commons, and the Scholarly Publishing Commons

Rethinking Institutional Repository Strategies

**Report of a CNI Executive Roundtable
Held April 2 & 3, 2017**
Published May 2017

Background and Synthesis

At the spring 2017 CNI meeting in Albuquerque, New Mexico, we held an Executive Roundtable on the topic of *Rethinking Institutional Repository Strategies*. We knew this was a timely topic for many of our member institutions, and we anticipated that many CNI member institutions would seek to attend this roundtable. To meet the demand, we offered two sessions of the roundtable (with different institutions participating in each) on sequential days, and even then had to turn away many who expressed interest. At the main CNI meeting, CNI's executive director Clifford Lynch offered a summary of the roundtables during one of the breakout sessions.[1] This report expands upon the themes Lynch described and adds more details about participants' observations and current activities.

CNI and its member institutions have played important roles in the development, shaping, and growth of institutional repositories (IRs) since the early 2000's. In fact, one of the earliest CNI Executive Roundtables was on this topic.[2] While some, especially in various European countries, have viewed IRs primarily or even solely as a means of curating bibliographies, preprints or open versions of institutional faculty articles, CNI has promoted the view that IRs are a "set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members."[3] IRs may include such content as electronic theses and dissertations (ETDs), research data, videos of campus performances and events, open educational resources (OERs) created at the institution, and other materials. They provide for systematic curation of a wide variety of content and a mechanism to showcase the institution's contributions, creativity, and research output. The content in most IRs is open access (OA),

---

[1] A recording of Lynch's talk is available at https://www.cni.org/topics/repositories/institutional-repository-strategies-what-we-learned-at-the-executive-roundtables.

[2] *Institutional Repositories*, report of a CNI Executive Roundtable held Dec. 8, 2003 (Jan. 2004), https://www.cni.org/go/institutional-repositories-fall-2003-executive-roundtable.

[3] Clifford A. Lynch, "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age," ARL Bimonthly Report 226 (February 2003), 1-7, https://www.cni.org/wp-content/uploads/2003/02/arl-br-226-Lynch-IRs-2003.pdf. Reprinted in Portal: Libraries and the Academy 3:2 (2003), pp. 327-336. For more on this ongoing discussion see my interview with Richard Ponyder, "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, and Clifford Lynch, "Updating the Agenda for Academic Libraries and Scholarly Communications – Guest Editorial," *College & Research Libraries*, 78:2 (February 2017), p. 126-130, doi:10.5860/crl.78.2.126.

but in some cases restrictions (embargoes) are in place for some materials due to publisher policies, pending patents, or for security or privacy reasons.

A great deal has changed since the first early-adopter institutions set out to develop their institutional repositories.

Disciplinary repositories have been well established since the early 2000s. They are often focused on preprints and rapid dissemination of research results, and were formed in several different sectors, generally the sciences, and sometimes funded by substantial grants. These disciplinary repositories often contain content from researchers around the globe. Two of the oldest and best established are the arXiv at Cornell University (originally at Los Alamos National Lab) and SSRN (recently acquired by Elsevier). But recently there has been renewed interest and activity around disciplinary preprints from a number of new disciplines, notably biology, and sophisticated platforms have become openly available through platforms like the Open Science Framework, greatly reducing barriers to entry by other new disciplines.

At times it is difficult to explain when faculty should place materials in an IR – or indeed, where, if anywhere, they should deposit them, and who (if anyone) will take care of the process. The rise of "open access" and (in the United States) "public access" funder mandates and institutional policies, and the recent emergence of what are specific requirements to deposit into funder repositories in the US have further complicated matters. Increasingly, faculty are also depositing materials in commercially run systems that offer social networking features, informative "walled garden" analytics on the use of the materials (citations and other measures), and other benefits; the misleadingly-named academia.edu, or ResearchGate, are examples of these. Are these in competition with IRs or complementary? What deposits are required, and which ones are optional? There's a great deal of confusion.

Finally, repository software and service platforms, as well as institutional views of functional requirements for these systems, have evolved substantially, as has the broader systems landscape.

Participants in these roundtables included librarians in administrative, scholarly communications, and technical roles as well as individuals from companies who develop institutional repository products, research labs, non-profit organizations, and others. Representatives from the US and the UK participated.

Making Distinctions between Repositories and Other Digital Collections

Institutions are struggling with the scope of the IR, particularly trying to disentangle a set of demands for various types of digital collection management platforms from things we traditionally think of as repositories. Many of the institutions have items coming out of digitization programs, such as newspapers, image collections, and special collections materials, where there is a consistent need for platforms where one can store the materials, provide metadata, annotate, curate, and offer (sometimes quite specialized) discovery and possibly analysis services. Libraries have invested significantly in the materials they have digitized, including adding item-level metadata.

This is rather different from an environment where one takes in contributions from the campus community and represents the items in much the same way that they are deposited. The archetypal examples here are document-like materials in various formats. Cases where a community member wants to deposit a few videos or sound recordings, computer programs,

or modest data files are very similar. There are other cases that are more ambiguous and nuanced: when a researcher approaches the library with a collection of photographs that she wishes to add to the IR, it would be atypical for that collection to arrive complete with metadata for each item; further, if the collection really is fully documented with detailed and perhaps specialized metadata the researcher might want retrieval and manipulation tools that go far beyond the standard IR offerings. Some other platform may be more appropriate (and there are also questions about the ongoing costs to the library of committing to long-term support of the material at this level of functionality). We will return to this very important issue below.

A key question that came up repeatedly at the roundtables was whether an institution should try to have one platform that they use for all digital collections or whether they use separate ones for different types of content.

Addressing how to develop repositories for a variety of disparate digital collections is a genuinely difficult problem made more challenging because of constraints and architectural choices implicit in the platform solutions that are now embedded in many institutions, especially if they began their digital collections/repository in the early 2000s when there were a very limited number of platforms to deal with digital collections or institutional repository collections. Now there are many alternatives for curatorial work around collections such as ArchivesSpace and Omeka, and even Artstor Shared Shelf. Many of the institutions at the roundtable were currently grappling with how to make distinctions among digital collections and whether those distinctions should be addressed in terms of finding solutions for them as separate or common needs. Many institutions described a situation where they have as many as five different platforms (and perhaps as many as 20 or more actual independent instances of one of these multiple platforms) that have characteristics of IRs, and it is increasingly impractical to operate all these separate platforms. In addition, and importantly, it is also difficult to explain to the user communities what is in each of these siloes, and to provide discovery tools that can span diverse systems and collection characteristics. Some institutions are now shifting collections from one platform to another in order to consolidate, rationalize, reduce or eliminate redundant systems. This is a slow, expensive process, which also exacts opportunity costs.

Some participants noted that it's not just the nature of the collection and the way it is presented, but also the mix of media types that may drive platform strategies. Some institutions have invested heavily in video or recorded audio and if they are building those digital collections at scale, existing popular general-purpose repository platforms are often not a good fit. Therefore, the institutions have employed media-type-specific platforms for those collections, and it is unlikely that that will change soon, although there is a question of volume; if the collection comprises only a modest number of audio files or videos, and there are few or no requirements for transcoding or streaming delivery, it is not much of an issue and those materials can be accommodated in most repository platforms.

Difficulties in clearly delineating the boundaries and definition of a repository surfaced in relation to other cases in addition to those centered on media. For example, most of the institutions represented at the roundtables were doing some kind of publishing activity, such as supporting journals (using the Open Journal System [OJS] platform or others) or collaborating with university presses on projects to digitize out of print or backlist monographs and make them available, most commonly publicly accessible. In both cases, there's a clear need for a place to manage and preserve these materials (though the precise level of preservation commitment on behalf of the institution may vary, or be somewhat

unclear). Similarly, a wealth of digital and digitized archives from various sources (internal and external, institutional and personal), which may or may not be part of the library, may also have a need for preservation, and where this fits into many institutions is not resolved, nor is the role of the library in this strategy. These issues are also relevant to records management; it may be to the advantage of those programs to use the expertise the library has in long-term preservation. However, it is unclear whether institutional archives and records should go into the institution's IR or into some other platform. The platforms specifically developed for archives and records management may come with different access, security and discovery apparatuses than one would typically find with IR platforms.

Many of the institutions present are also grappling with another complicated scope and relationship question: how research data management relates to an IR, in particular different models about whether data sets being kept for the long term go into an IR or into some data-specific repository and data preservation environment, perhaps run by the campus or perhaps externally operated. In many instances, disciplinary data repositories will play an important role, but there is a long tail of data sets that will need institutional curation. Are traditional IR platforms well equipped to handle them? It varies based on the institutional needs and ambitions. If the institution handles fairly small, straightforward datasets such as Excel spreadsheets, they could probably put them in any platform. If the institution needs to manage and curate much larger datasets, they probably need something more specialized.

Some organizational issues must also be considered; most research institutions are building teams of data managers who work with faculty. How do these teams relate to teams building repositories or to scholarly communications librarians? Since repositories are often part of the scholarly communications unit of the library, this can give the IR some biases about what its priorities are (a strong focus on publication-like things, OA journals, or shifting journals to OA models), which may be very different from a focus on research data, the evolving scholarly record, or archives broadly. IRs cast strongly in the tradition of supporting a very traditional view of scholarly communications with an OA overlay can function as a highly conservative force. It is important for institutions to carefully consider where the repository fits in the spectrum of services that are emerging as significant shapers of the stewardship strategy.[4]

It is interesting to note that in one of the two roundtables the needs of research data management were seen as a much greater driver for future repository strategies than in the other.

Relationships between Research Information Management (RIM) Systems and Repositories

One could think of RIM (sometimes also called a current research information system [CRIS]) as a comprehensive faculty bibliography. An RIM may be separate from a faculty tracking system (annual activity reporting in support of tenure and promotion) though there is growing interest in how to integrate the two, in part to reduce the amount of redundant faculty time and effort, though doing so involves transcending some very long-standing organizational silos. There is an interesting school of thought surfacing that says that a CRIS might serve much of the function of an IR, and could be used to span faculty work in the

---

[4] Deanna Marcum, Roger C. Schonfeld, and Sarah Thomas, *Office of Scholarly Communication: Scope, Organizational Placement, and Planning in Ten Research Libraries* (New York: Ithaka S+R, 2015), https://doi.org/10.18665/sr.275206.

published literature as well as what is in the local repository and could be a way to provide access to work that is open; Elsevier and CHORUS are introducing APIs that allow IRs to operate in this fashion. Systems like Symplectic Elements, and even feeds from SHARE can serve as sources for automatically populating the bibliography. Big publishers will sometimes push bibliographic information back to the institution as well as pointers to the articles themselves; Pure from Elsevier is a very good example of this approach. The institution doesn't typically get the full text back locally and the publisher controls access. That is an interesting model and one that people are thinking about. For institutions that are determined for one reason or another to internally build up a collection of faculty publications as part of their repository activities, such a bibliography also serves as a guide for the kind of material that the institution may seek to accession. These discussions further frame questions about what problems we're trying to address with the IR. Note that a relatively traditional faculty bibliography does *not* provide a full inventory of what ought to be in an institutional repository under the broader views of their scope, or even a good guide to priority accession targets.

What Problem Are We Trying to Solve? IRs and Open Access Strategies

CNI has advocated the position that the priority for IRs is to capture all types of content produced by the institutional community, particularly material at greatest risk of being lost. Others say that the primary purpose is to capture a record of the institution's scholarly output and place an intense emphasis on the journal literature; for some scholarly communications-driven repositories, that is the goal, and it is not being met very well for US institutions, in part due to the way that public access funder mandates have worked out in the US. The situation in the US is developing very differently when compared to some other countries such as the UK, where IR deposit compliance rules are strong, vigorously enforced, and tightly coupled to both funder regulations and to institutional (departmental and above) as well as individual faculty evaluation.

In the US, 60% compliance rates are considered high; even these levels of compliance are typically obtained only by negotiating arrangements where the publishers will help populate the institutional repository directly, usually as part of "big deal" negotiations and for a fee. This favors the concentration of scholarly publishers and publishing platforms (e.g. HighWire, Atyphon) because it requires a lot of institutional overhead for such activities as setting up and managing publisher feeds and negotiating custom arrangements; therefore, institutions can usually only afford to set up and pay for arrangements with the publisher(s) who gives their institution the biggest return on investment.

A very interesting conversation on the vexed issue of the OA-IR strategy relationship threaded through both roundtables. Participants were very mixed about whether they had an OA institutional policy, and if so, whether it had any teeth. Some institutions that have an OA policy are grappling with the question of to what extent that implies an obligation for the library (or indeed the institution more broadly) to invest to make that policy work. There is a big difference between having a policy on the books and making a genuine institutional investment to seek to fulfill it. It is meaningful when the institution puts resources behind implementation of the policy. The investment to do this credibly is significant, far in excess of simply establishing and running an IR. There is a considerable disconnect in this area, and it needs to be considered by provosts and faculty senates, not just libraries. At present, funds supporting an IR are usually reprogrammed or shifted from some other part of the library budget, though many libraries did receive some limited-time startup new funding. Funding responsibilities and sources for OA mandates and meeting funder mandates is not well

settled. Each institution had a different answer, but if there was an emerging trend among institutions participating in the roundtables, it was that they are more willing now than they were five years ago to actively put processes in place to add materials to the IR, or, perhaps more accurately and more broadly, to help faculty to meet new funder mandates related to public access; this may result, in part, from the complex provisions of new funder mandates, and also from developing clearer processes for ingesting the permitted version of author contributions.

Repositories in the UK

In one of the roundtables, a representative from the Jisc, which has a broad view of developments in higher education in the UK, provided a summary of how repositories had developed in UK universities, what challenges they are facing, and how he views the way forward; we present this summary of his comments with attribution, specifically with their permission.

Initially, repositories in UK universities focused on metadata due to specifications of government funding and reporting requirements. In 2013, the Research Councils, which provide much of the research funding to UK universities, began requiring research outputs to be made open access, with deposit into local institutional repositories being a common method of compliance. More recently, the Research Councils have required access to research data as a condition of funding, with the Engineering and Physical Sciences Research Council (EPSRC) requiring deposit into an institutional repository.

Since libraries in the smaller institutions don't have the capacity to provision the services in-house, many are using cloud services, including using remote instances of EPrints and DSpace. In addition, some are using figshare or Hydra/Fedora solutions, hoping these will act as off-the-shelf solutions that are customizable. The largest and most research-intensive universities, such as Oxford and Cambridge, may have different strategies. In general, there are two strands of IR development: one that is technical, seeing the content as a group of files on the web with an efficient, easy workflow for ingest, and the second that views IRs as an essential piece of the scholarly communication infrastructure. In the future, some of the likely focus for Jisc and UK universities will be towards moving open science into IRs, capturing born-digital content, and moving beyond PDFs into new forms of scholarly outputs. In addition, Jisc is actively working on a research data shared service that will incorporate discoverability and data curation.

Next Stage Strategies

At the roundtable, we sought to understand how institutions were revising their repository platform strategies, especially in cases where they faced the legacy of choices made a decade or more earlier. In general, there were three (not mutually exclusive) strategies that various institutions were taking for moving their repository(ies) to the next stage. One was to reduce the number of platforms and collections and consolidate them into one or a fewer number of collections/platforms. A second strategy was to migrate a number of collections on various platforms to a platform new to the institution; often Hydra/Fedora was mentioned in this case. A third strategy being considered and implemented by some when their institution has a multiplicity of platforms was to implement a cross-platform discovery tool. Note that this third strategy often also seeks to provide discovery that subsumes other collections maintained by the institution beyond the library, such as museums and archives, as well as

faculty-created resources. All of these strategies are resource-intensive, especially in terms of staff time and expertise.

An IR migration is a variable process depending on what the institution is trying to achieve. A main focus, for example, may be to either disentangle or consolidate collections. There are specific questions such as what metadata can be preserved in a migration; this is particularly serious for systems housing digital collections, where variable and specialized metadata is commonplace, and there are often special search, presentation and use requirements. A significant number of institutions are staking a lot of hope on migration to Hydra/Fedora implementations but are concerned by the amount of technical staff it takes to support that in both migration and ongoing maintenance. It is an environment that is gaining mind share. It is also clear that institutions using early versions of platforms such as DSpace, which have been extensively locally customized, are facing nasty migration problems, even to migrate to a current version of the same platform; this is not a new problem or exclusive to DSpace, of course, but is perhaps most visible there because DSpace was such a major platform for early adopters (there were very few alternatives at the time).

There are subsidiary problems (although not much addressed in the roundtable) including the need for local authority control when running a quality repository system, and the reality that the institution's own faculty may not be in the library's existing authority control mechanisms. There are some developments underway at a national level to begin to address these challenges, but there is still going to be a substantial local level system (and data) integration challenge here.

Institutional Perspectives

- Libraries are still debating whether a repository should be focused on discovery, access, and/or preservation. As the outputs of the scholarly record have evolved in the time since the advent of repositories, there are many additional components of the scholarly record such as webpages and new media, and there is no cohesive discussion of the intersection of repositories and these materials, much less social media and emerging genres like annotation.

- It's clear that special collections/IR boundaries and relationships need more attention and analysis, especially in light of the growing emphasis on both accession/acquisition of digital materials and digitization programs for existing materials in the special collections world.

- The Digital Humanities/IR relationship is very important. For faculty aggressively embracing DH work, and producing new genres of scholarly output that don't fit within the existing publishing (and hence stewardship and preservation) systems, the ability to shift access and preservation of their work into an institutional setting is essential to legitimizing this work.

- It appears that IR systems can serve as an early warning system for identifying new classes and genres of digital scholarly outputs that will need stewardship. Notable examples that are emerging through this are augmented reality, virtual reality, and 3-D object imaging and architectural models.

- A number of participants noted that as they strategized about the directions their repository should take, they had institutional goals in mind. For example, in an institution that is aiming to expand its research profile, the library wants to understand how it can assist with helping the university be accountable for research, bring together outputs of research, and showcase those results.

- Not all participants noted how long their institution had had a repository, but for those that did, the range was from 4 to 15 years; there were very significant differences between early and late adopters in terms of strategy issues and development.

- Some institutions stated that they have tens of thousands of items in their repository; in some cases, those were consortial or state-wide repositories. One state-wide repository has over 10 million items; this is also the digital library program for the state. There is still little agreement about how to count, measure and compare or benchmark the contents, use, and roles of repositories and digital collections.

- Many institutions commented on their efforts in outreach and education to encourage submission of materials to their repository. One participant lamented that without continual marketing, it is hard to sustain faculty interest in depositing their materials in the repository.

- Another participant shared that when a faculty member wants to deposit complex material in the IR (such as a database), they do a mock-up to show how the content will actually look in the IR platform, and it is often not satisfactory for the faculty member's purposes. This very important observation helped to underscore that there is a broad set of faculty needs: often they want a platform to mount a collection of materials mainly for access and near-term use, and are much better served by something like Omeka, Artstor Shared Shelf, or other tools. Little work (other than a few research papers and studies by long-standing digital humanities centers about what to do with old DH projects[5] has been done on developing strategies to deal with the rapidly developing long-term stewardship challenge for various kinds of faculty collections and similar projects. Closely related here is the role of IRs in accommodating the websites and other digital assets of de-commissioned institutional units. Also relevant here but not discussed in these roundtables is longer-term stewardship of faculty materials currently being made available on external public platforms: Instagram, Flickr, YouTube, SlideShare, etc.

- In institutions that have an open access mandate passed by their faculty senate, librarians are interested in assisting the faculty make good on their policy, in some cases by providing "white glove service," selecting and vetting items for the repository. However, the librarians see this as an unfunded mandate since no funding supports the kinds of services they provide or could develop.

- One institution reported that the provost's office did fund new positions for the repository but they were targeted for data management, not specifically for open access.

---

[5] See Digital Preservation for Social Sciences and Humanities (DPASSH) 2017, "Preserving Abundance: The Challenge of Saving Everything," University of Sussex, Brighton, June 14-15, 2017, http://dpassh.org/.

- In the US, a small number of libraries, primarily at large, research universities are working with publishers or trying to work with them in an automated way for papers of their faculty to be ingested into the institutional repository. One university reported this type of negotiation has helped them achieve a two-thirds rate of faculty publications represented in their repository, the highest percent reported by anyone at the two roundtables. The wider adoption of the use of ORCIDs will facilitate this process; the prospect of institutional identifiers is also promising here.

- Librarians worked with a wide variety of institutional partners in developing their repository strategy. These partners included the information technology unit, the Vice President (VP) for Research office, the university public relations office, and the university press. In one university, the library worked with the office of research to jointly offer orientation sessions for faculty on a variety of issues, including applying for grants, compliance, and deposit of research outputs. In another institution, a new policy requiring data deposit with dissertations (ETDs) is propelling a closer relationship between the library and the VP for Research office.

- There were many varieties of governance structures for repositories as well as institutions with no formal governance structure. Some IRs are administered solely with library oversight. In consortial repositories, generally each library has a representative on a policy committee. In some universities, other units are involved in governance; those mentioned were the information technology unit and the research office.

- Only a few institutions discussed assessment of repository programs and no institution present had stated goals or metrics for what constituted success; this omission is obviously closely related to the lack of clarity about the objectives and goals of repository efforts. One participant stated that his institution was not seeing a high return for their investment in repositories and that the statistics generated by the systems are poor and do not demonstrate impact. Another individual asked whether we are trying to assess value for the researchers themselves, the institution, global scholarship, the general public, or a combination of those groups. Researchers and librarians at several universities are working to make analytics on use of items in IRs more reliable.

- A number of participants mentioned that Open Educational Resources (OERs) are gaining a lot more interest and importance at their institution, especially since they are able to document substantial, clear, immediate savings for students. Including OERs in repositories can give them good visibility, which is a service to the institution and the students. OERs are also an important use case to help broaden thinking about the scope and objectives of an IR.

- Changes in strategies about undergraduate teaching and engagement with primary research materials are giving rise to new discussions about the extent to which the institution should take responsibility for stewarding, preserving and providing access to materials being created by the student community at both the graduate and undergraduate levels, and the role that the IR should play in providing these services.

- In a liberal arts college that has a program of curricular grants for incorporating digital humanities methods into student assignments, the library has curated many of the

student materials created in those courses as well as syllabi in their repository. They would like to promote this type of student collaboration with faculty to newly accepted students in order to demonstrate the kinds of opportunities students have at their institution.

- Although we are aware that at least one CNI member institution has begun using figshare as its repository platform, none of the participants in the roundtables reported that development at their institution. However, we know that many individual researchers in other institutions are choosing to deposit research outputs in figshare or other platforms rather that locally; one participant argued that libraries need to think of repositories as active hubs where social interaction for researchers could be built in. It is interesting to consider work that is going on in surrounding environments like the Open Science Framework and their connections to institutional library-based systems in this regard.[6]

- University PR/marketing/recruitment/government relations organizations are becoming important partners and stakeholders in IR programs. Showcasing faculty expertise to the external world can be a very important function. bepress has recently advanced some very interesting offerings here.

- Faculty conceptions about the relationships and connections between IRs and OA programs can actually be detrimental, as they can discourage faculty deposit of a great deal of very desirable material (conference papers, grey literature, data, etc.) that is not part of the published journal literature.

- If faculty submit publisher versions of articles to the IR, perhaps in an attempt to satisfy an institutional OA mandate, this presents a serious dilemma for the library and a potential liability to the institution as a whole: to what extent should IR managers attempt to police and vet author-publisher agreements as part of managing the IR deposit stream?

- Materials that need to remain "dark" or inaccessible in the long term are a real issue, and institutions need to think about their commitments to preserving these materials; long-term darkness is a virtual guarantee that the materials won't have much impact.

- It is clearly possible to reflect a nearly comprehensive and complete record of the theses and dissertations produced by a university (ETDs) in an IR. This is now clearly routine and non-controversial. But, by itself, it doesn't seem to be an adequate justification for establishing an IR service.

- For institutions that are doing systematic web archiving (either for institutional reasons, such as archiving images of the institutional web space as part of the local archive, or in support of faculty research initiatives) this work is typically siloed away from the IR strategy.

---

[6] See "The Open Science Framework (OSF) at Notre Dame: Connecting the Workflow and Supporting the Research Mission" by Andrew Sallans (Center for Open Science) and Natalie Meyers (University of Notre Dame, presentation (video) from the Fall 2015 Membership Meeting of the Coalition for Networked Information (CNI), Dec. 2015, https://www.cni.org/topics/scholarly-communication/the-open-science-framework-osf-at-notre-dame-connecting-the-workflow-and-supporting-the-research-mission.

- A massive and largely un-discussed issue is what to do with athletics-related materials (which may include extensive, massive video files), and how this relates to institutional asset management and stewardship strategy and governance.

- There's a clear path of evolution in architectural thinking, moving from the IR as platform to IR as a service to IR as a bundle of related services.

Concluding Thoughts

Some interesting observations emerge when reflecting on things that we expected to hear more about but did not. Most notable here were the very limited discussion of cloud based multi-tenant or even large-scale consortial repositories (though there are efforts such as the Digital Preservation Network that are considering some related efforts) as platform solutions, even though hosted repository instances are readily available through Duraspace and bepress is certainly growing its customer base. While these were little discussed, it was clear that organizations such as flagship state universities were also often functioning as hubs or hosts for many smaller, ad-hoc "consortia" involving local historical societies, branch campuses and other small organizations. Questions about consolidation, market development, and support strategies for smaller institutions is going to become increasingly critical as research data management grows in importance.

We have often framed the repository as a library/archives program, and particularly as we have started to talk about research data management, or indeed even serious support for an institutional OA mandate, it is clear that most libraries lack funding to support those kinds of resource-intensive services. Ongoing conversations about how to fund research data management at some institutions are now involving the IT unit, the VP for Research's office, the Provost's office, and academic departments. The point is that stewardship and the support of faculty are now being positioned not as library problems but as institutional problems arising from the changing nature of scholarly practice and the institution needs to step up to it. The library may be involved in stepping up to address the problem but fundamentally it is an institutional problem.

This theme was emphasized by several representatives in one of the roundtables, and they used the term "enterprise repository" to represent this concept. A case may be made for trying to move towards an enterprise repository that encompasses information researchers are actively developing in the process of their work but that that libraries would not usually consider in scope of an institutional repository because it is still too unstable and requires very different management strategies. This is a part of a broader strategy by which libraries might engage faculty early in, and throughout, the research process, rather than just receiving "outputs" from this process. Another approach would be to broaden the scope and content of general-purpose (usually commercial) digital asset management systems, which many institutions have invested in, but usually have not included objects curated by the library. It may be time for libraries to take a look at these systems and figure out how they fit into the landscape and how to move towards enterprise repositories that accommodate a much more diverse and inclusive set of institutional content assets. As is the case with IRs, at this point, generally there is no institution-wide funding for such a system, and for progress to be made, it is likely that the provost's office would need to make some serious commitment to the process, which would involve information stewards across the campus as well as faculty. We should seek to develop and use systems that allow us to leverage

workflow at an enterprise level and to have systems that will work and interoperate at a consortial level. A very modest beginning to one aspect of this is going on in the context of the SHARE initiative.

Going forward, library administrators need to focus more on how development is being handled on some open source repository platforms and whether development agendas are genuinely being driven by the strategic needs of the institutions or have become dominated by the technical agenda of developers themselves. We should be asking ourselves this question. It is easy to get captured by the agenda of frontline developers, which can diverge from strategic priorities and goals. In addition, there can be disconnects between the developers of the repository infrastructure and the librarians who provide the outreach to the university community.

Roundtable participants encouraged CNI leadership to continue to identify and analyze the main issues that should be addressed in the repository discussion and to assist institutions with moving towards better solutions. They asked whether we could come close to a global vision for repositories and then develop the steps needed to achieve that goal. One particular issue that was noted but not explored deeply in the discussions reported here, but that is receiving attention in other venues (for example, COAR in its recent discussion of next-generation repositories[7]), is the relative lack of success in developing a genuine global network of linked repositories through federated search, metadata exchange, and selective automated replication of data among repositories. Standards and best practices, as well as software platform support for these developments as they become established and obtain consensus, has been very slow in developing to enable this type of inter-repository linkage.

CNI Executive Roundtables, held at CNI's semi-annual membership meetings, bring together a group of campus partners, usually senior library and information technology leaders, to discuss a key digital information topic and its strategic implications. The roundtables build on the theme of collaboration that is at the foundation of the Coalition; they serve as a forum for frank, unattributed intra and inter-institutional dialogue on digital information issues and their organizational and strategic implications. In addition, CNI uses roundtable discussions to inform our ongoing program planning process.

The Coalition for Networked Information (CNI) is a joint program of the Association of Research Libraries (ARL) and EDUCAUSE that promotes the use of information technology to advance scholarship and education. Some 230 institutions representing higher education, publishing, information technology, scholarly and professional organizations, foundations, and libraries and library organizations, make up CNI's members. Learn more at cni.org.

---

[7] Confederation of Open Access Repositories (COAR), *Next Generation Repositories* (February 7, 2017 – draft for public comment), https://www.coar-repositories.org/files/COAR-Next-Generation-Repositories-February-7-2017.pdf. See also Next Generation Repositories COAR program website at https://www.coar-repositories.org/activities/advocacy-leadership/working-group-next-generation-repositories/.