MONTANA STATE UNIVERSITY LIBRARY

Why So Many Repositories? Examining the Limitations and Possibilities of the Institutional Repositories Landscape

Authors: Kenning Arlitsch & Carl Grant

This is an Accepted Manuscript of an article published in Journal of Library Administration on March 27, 2018, available online: http://www.tandfonline.com/10.1080/01930826.2018.1436778.

Arlitsch, K., & Grant, C. (2018). Why So Many Repositories? Examining the Limitations and Possibilities of the Institutional Repositories Landscape. Journal of Library Administration, 58(3), 264–281. doi:10.1080/01930826.2018.1436778

Made available through Montana State University's ScholarWorks
scholarworks.montana.edu

Column Title: posIT

Column Editor: Kenning Arlitsch, Dean of the Library, Montana State University, Bozeman, MT
kenning.arlitsch@montana.edu

This JLA column posits that academic libraries and their services are dominated by information technologies, and that the success of librarians and professional staff is contingent on their ability to thrive in this technology-rich environment. The column will appear in odd-numbered issues of the journal, and will delve into all aspects of library-related information technologies and knowledge management used to connect users to information resources, including data preparation, discovery, delivery and preservation. Prospective authors are invited to submit articles for this column to the editor at kenning.arlitsch@montana.edu

Why So Many Repositories? Examining the limitations and possibilities of the IR landscape

KENNING ARLITSCH
Dean of the Library, Montana State University, Bozeman, MT, USA

CARL GRANT
Associate Dean, Knowledge Services & Chief Technology Officer, University of Oklahoma Libraries, Norman, OK, USA

Keywords

Institutional repositories; digital repositories; IT management; library administration; open access; collective action, network effect, enterprise institutional repositories

© Kenning Arlitsch and Carl Grant
Address correspondence to Kenning Arlitsch, Dean of the Library, Montana State University,
P.O. Box 173320, Bozeman, MT 59717-3320, USA. E-mail: kenning.arlitsch@montana.edu

Abstract

Academic libraries fail to take advantage of the network effect because they manage too many digital repositories locally. While this argument applies to all manner of digital repositories, this article examines the fragmented environment of institutional repositories, in which effort and costs are duplicated, numerous software platforms and versions are managed simultaneously, metadata are applied inconsistently, users are served poorly, and libraries are unable to take advantage of collective data about content and users. In the meantime, commercial IR vendors and academic social networks have shown much greater success with cloud-based models. Collectively, the library profession has enough funding to create a national-level IR, but it lacks the willingness to abandon local control.

Introduction

Most industries have learned to leverage the power of networks, but libraries still struggle to afford similar advantage to the realm of institutional repositories (IR). The theory of the “network effect” holds that a product gains value much faster as more people use it, and it is characterized by the use of cloud-based infrastructure and services. But by continuing to manage thousands of locally-installed repositories, libraries and their users have failed to benefit from this network effect. “Many institutions describe 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 IR.” (Coalition for Networked Information, 2017).

Academic libraries have helped to create a wealth of open access scholarly content, but it is a wealth that is difficult to count, analyze, and understand how it benefits users and the academy. The dispersed model facilitates local control, but it creates a variety of problems that include siloed content, duplication of effort, inconsistent application of metadata standards, discovery deficiencies, and increased strain on scarce resources. When characterizing the current condition of IR, some authors have also cited the lack of local grassroots support, poor usability, low usage, high cost, fragmented control, and distorted markets (Van de Velde, 2017).

The proliferation of individual repositories means that standards are often implemented a little differently from one repository to another, requiring crosswalks that lead to loss of granularity when attempts are made to aggregate content. Aggregators, such as Google Scholar, for IR, or the Digital Public Library of America (DPLA), for cultural heritage repositories, often struggle to harvest and normalize metadata from disparate repositories where “standards” were applied inconsistently. Meanwhile, commercial platforms like Digital Commons, from bepress, have used the power of the network to limit such inconsistencies, which advantages them and their customers. Academic social networks such as Academia.edu and ResearchGate, whose models of operation have leveraged the power of the network from their inception, have fared much better than library-managed IR, when measured by submission participation, use, and bulk of content. Library Systems Platforms, such as Alma from Ex Libris, or WorldShare, from

OCLC, also capitalize on this network effect and have shown their worth by rapidly acquiring share in that market.

The problem is not money, as David Lewis has recently pointed out in his proposal for a 2.5% commitment toward collective open access, open source, and open data efforts. Collectively, U.S. academic libraries control budgets of approximately $7 billion (D. W. Lewis, 2017), but we lack the will to pool our resources and act collectively, on this and other costly concerns, such as the continually escalating prices of scholarly journals (Wenzler, 2017).

Numbers and Growth

Tracking the number of IR worldwide can be difficult, as the few registries that exist tend to rely on self-registration. However, by examining those registries and the professional literature, a reasonable estimate of the number of IR can be calculated and their growth mapped.

Although generic digital repositories had existed for some time, repositories devoted to collecting the intellectual output of universities began to appear in the early years of this century, and by mid-2005 a survey found 305 IR in 13 countries, excluding the U.S. (Van Westrienen & Lynch, 2005). A similar survey taken in that same year, of the then-124 members of the Coalition for Networked Information (CNI), found that 40% had implemented an IR, while 88% of the rest had plans to do so (Lynch & Lippincott, 2005). A survey from the IMLS-funded MIRACLE Project cast a wider net to include directors and senior administrators from a variety of academic libraries in the spring of 2006, and found that only 11% of its 446 respondents reported implementing an IR, while 53% had done no planning that would take them in that direction (Markey & Council on Library and Information Resources, 2007).

Based on those studies, it is fair to say that the number of IR was low and growth was slow until around 2005. However, a boom in IR growth was about to occur, as a survey of repositories registered with the Directory of Open Access Repositories (OpenDOAR) showed that “the total number of repositories in OpenDOAR grew from 128 in December 2005 to 2,253 in December 2012 … [representing] a 1660% increase during the period” (Pinfield et al., 2014).

A recent report on next-generation repositories refers to “the globally-distributed network of more than 3000 repositories” (Confederation of Open Access Repositories, 2017), a number that seems conservative according to the current listings in OpenDOAR of 3,448 repositories (University of Nottingham, 2017). Numbers in the Registry of Open Access Repositories (ROAR) are even higher, listing 4,585 results on its website (University of Southampton, 2017). ROAR’s numbers, however, are potentially inflated as a result of its automated harvesting model, which “tends to pick up a significant number of invalid sites, including experimental implementations that have few records or those with metadata-only entries” (Pinfield et al., 2014).

Repository66, a “mashup of data from ROAR and OpenDOAR,” had the number at 3,045 repositories in April 2014, which was apparently the last time that data were last harvested (S. Lewis, 2014).

The DuraSpace Registry lists 2,288 public repositories, but that number includes only DSpace, Fedora, and VIVO repositories (over 2,000 of those are DSpace repositories), and there is little reason for non-DuraSpace members to list themselves in this registry (DuraSpace, 2017a). Numerous other repository platforms that appear in OpenDOAR and ROAR, including Digital Commons and ePrints, do not appear in the DuraSpace Registry.

Based on these surveys and registries, it seems fair to state a conservative estimate of the number of open access IR at 3,000-3,500 worldwide at the end of 2017. It is quite possible that the actual number is higher.

Management

There are positive aspects to distributed repositories. Local implementation may be desirable, for reasons of flexibility, control, customization, and “distributed networks are more sustainable and at less risk to buy-out or failure” (Confederation of Open Access Repositories, 2017). However, there is little direct discussion in the literature about the drawbacks of managing many instances of repositories; most of the literature focuses on content recruitment, staffing, or metadata issues (Moulaison Sandy & Dykas, 2016; Palavitsinis, Manouselis, & Sánchez-Alonso, 2017; Park, 2009; Wu, 2015; etc) rather than the challenges posed by managing multiple repositories.

A 2013 survey examines costs associated with IR and the additional services offered by some, but it is inconclusive due to a low survey response rate. However, the authors do raise concerns about the model of dispersed repositories in their conclusion, noting:

> *“…it remains to be seen if individual institutional management of a repository is the most efficient and effective means of operation. A question that should be asked of the users of repositories is whether their needs are met by the dispersed model of repositories that exists at the present time, or if some kind of unification (or at least unified search and retrieval capability) would be more useful” (Burns, Lana, & Budd, 2013).*

Seaman lists a variety of other challenges faced by IR: insufficient awareness and stakeholder engagement; content recruitment; perceptions of intellectual property loss; and submission difficulty. He also examines the a priori assumption in many libraries that IR were necessary and would succeed, which usually resulted in an absence of formal needs assessments prior to implementation (Seaman, 2017, p. 34).

Indirectly, Seaman raises the advantages of the network effect by observing the success of “academic social networks like ResearchGate and Academia.edu” (Seaman, 2017, p. 37), noting that they currently contain approximately 100 million (ResearchGate, 2017) and 18

million (Academia.edu, 2017) publications, respectively. It is important to acknowledge that the success of ResearchGate is tempered by a pending publishers lawsuit that seeks removal of up to 7 million articles that may be in breach of copyright (Van Noorden, 2017), and that Elsevier issued “2800 [takedown] notices to Academia.edu in 2013, but did not take the site to court” (Chawla, 2017). These actions reveal a strength of library-run IR, where the copyright status of publications is generally vetted before a paper is uploaded. Still, the papers subject to removal from ResearchGate and Academia.edu because of potential copyright breaches are small percentages of their respective totals. It is also useful to note that the decentralized nature of open access IR worldwide makes it virtually impossible for libraries to produce the total number of publications they contain, and which are so easily revealed in ResearchGate and Academia.edu.

Further evidence of the popularity and reach of academic social networks comes from a study that found ResearchGate contained more than half of the articles published by scholars at top Spanish universities, while only 11% of those same articles were available in the institutional repositories of the authors’ home institutions (Borrego, 2017). Borrego concludes that “IR may be fulfilling a role by helping universities to showcase their research activity and give an account of their research performance, but they are hardly playing a role in providing OA to content published in subscription journals.”

Of course, the task of managing repositories requires a resource precious to all libraries: staff labor. While the numbers can vary wildly, “many institutions spread the work over two main posts: a Repository Manager… <and..> a Repository Administrator...” (Wickham, 2011). The authors of this article note that the average FTE devoted to repository management between their two institutions alone is approximately 5 FTE. Multiplying this by the 3,500 repository count from above (which, again, could be low) provides a measure of the massive amount of resources our profession is allocating to repositories (8 hrs/day*22 days/month*12 months*5 FTE*3,500 institutions=36,960,000 person hrs/year). Of course, this number needs further study to identify what portion of labor is being devoted to support unique customizations compared to product maintenance, such as upgrading software versions, fixing bugs and security enhancements. If we could re-allocate these latter staffing resources to developing shared user-facing solutions, we could see some sizable boosts in functionality. In the current environment, the profession is wasting massive amounts of resources in duplicating shared IR infrastructure work.

Software Versions

The variance in software platform versions implemented in the community are dramatic. For example, the DuraSpace Registry listed 2,004 repositories that were running the DSpace platform as of December 2017. Of these, there is a wide distribution of versions, with almost 800 (nearly 40%) of the known versions showing 1.8 or earlier, representing a span of software releases dating back nearly 16 years (see Figure 1). The latest release of DSpace is version 6.x, and all DSpace versions prior to 5.x are unsupported as of January 2018 (DuraSpace, 2017b).

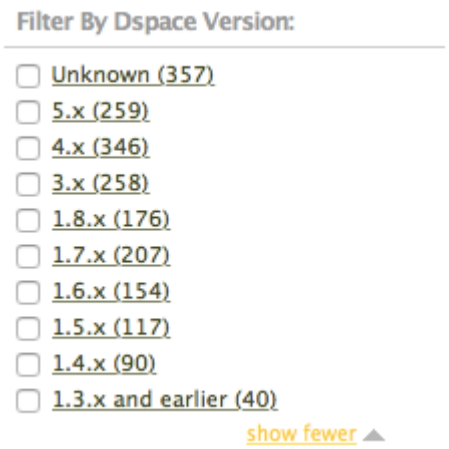

Figure 1: Screen capture from the DuraSpace Registry taken 2017-12-10, showing the number of registered repositories (in parentheses) running each version of DSpace.

The negative implications of this situation for the profession cannot be overstated. With each subsequent release of a new version of the software, these sites become more entrenched in the past as the difficulty of upgrading becomes more and more challenging. In addition, these sites increasingly put themselves at risk of security and ransomware attacks, both of which contain the potential to compromise the trust required by the community of users who invest their intellectual property in these very repositories. The community of developers is also spread more thinly as a result of trying to provide critical security upgrades for multiple versions of the software.

By any reasonable measure, this is a misuse of valuable resources that are largely focused on simply maintaining the status quo.

User Perspectives

Perhaps the biggest question to ask when considering the value in the current environment of multiple repositories is “how does this condition serve the user?” In answering this question, we can consider two types of users: those engaged in the discovery process; and those whom we hope will self-archive their intellectual output in a repository.

From the perspective of those engaged in discovery, there is little value in trying to search repositories individually. Far more effective is the use of aggregators like Google Scholar (GS), whose academically-inclined users are exactly the audience that are valued by IR managers, and which previous research has demonstrated to be the source of 48%-66% of referrals to IR that it has successfully harvested and indexed (OBrien et al., 2016). But previous research has also demonstrated that many IR struggle with the harvesting and indexing requirements of GS, resulting in IR content being poorly represented in its index (Arlitsch & O’Brien, 2012) and thus diminishing the value of the aggregation.

In the domain of self-archiving, Poynder notes that "...author self-archiving has remained a minority sport, with researchers reluctant to take on the task of depositing their papers in their institutional repository. Where deposit *does* take place, it is invariably hard-pressed intermediaries who do the work" (Poynder, 2016). It is interesting to note that MIT recently stated: "MIT remains a leader in open access, with 44 percent of faculty journal articles published since the adoption of the policy freely available to the world." (MIT Libraries, 2017). While we don't dispute the statement, we will note the that if capturing 44% - i.e. less than one half of faculty journal articles - makes one the leader (and it does, because most institutions see lower numbers), then, clearly we all have much work left to do. A recent CNI presentation outlined some of the challenges users face (including Researchers, VPRs, and Provosts) in dealing with self-archiving and other IR services (Grant, Soderdahl, & Beit-Arie, 2017). Included were:

1. Researchers
    a. Depositing in multiple repositories
    b. Entering metadata
    c. Maintaining multiple "profiles"
    d. Managing requests for copies of data associated with research
    e. Generating research data management plans
2. Vice Presidents of Research
    a. Minimizing workflows for researchers so they can focus on research.
    b. Helping research and researchers gain exposure and press coverage
    c. Needing to raise the profile of the institution
    d. Helping researchers comply with the data management plan milestones
    e. Generating reports on research data use and reuse
3. Provost
    a. Increasing faculty impact
    b. Helping with researcher retention
    c. Increasing community impact and understanding of the value of research
    d. Increasing the brand value of the institution (CNI Meeting, Fall 2017, Project Briefings)

These are sizeable challenges but not insurmountable, provided we come to the realization that allocating resources as we are won't take us there. IR are clearly not working very well for users, either on the discovery end or the submission end. It's time for a change.

Looking Forward

The rate of technological change makes it incredibly difficult to predict, with any accuracy, where we'll be 5-10 years from today. In his recent book, Thank You For Being Late (Friedman,

2017), Thomas Friedman noted that it was just over ten years ago that the following technologies were introduced:

- iPhone (2007)
- Hadoop (big data)
- GitHub
- Facebook (Sept 2006)
- Twitter
- Google bought YouTube (2006)
- Android
- Kindle
- IBM's Watson
- Intel's high-k-metal gate microchips

The rate of technological change really hasn't slowed since that time, nor is it likely to slow in the future. One thing that is clear is that this rate of change quickly obsoletes many existing technologies. The implications of the rate of change on scholarly communication will continue to be large and causes one to wonder if Van de Velde was correct when he surmised: "The Institutional Repository is obsolete. Its flawed foundation cannot be repaired. The IR must be phased out and replaced with viable alternatives." (Van de Velde, 2017). Of course, the definition of viable alternatives covers a great many possibilities and far fewer agreements on which one is correct. However, most discussions on this topic focus on newer, more comprehensive versions of the major IR software platforms, such as DSpace or Fedora Commons. Others focus on the needs for enterprise level, cloud-based solutions that offer massive scalability, multi-tenant software, network effects and more. Beyond this level, others are pushing into entirely different, innovative, but as yet untested ideas like those discussed by Herbert Van de Sompel in his Closing Plenary of the Fall 2017, CNI Meeting during the Paul Evan Peters Award & Lecture, titled: "Scholarly Communication: Deconstruct and Decentralize. A Bold Speculation" (Van de Sompel, 2017).

Changing the Mindset

Fifteen years ago, many academic libraries were still managing their own email servers. Today, email servers are rare in libraries; most of us use cloud-based email and calendaring services from Microsoft or Google. The idea of libraries running their own email servers now rightfully seems ludicrous, but at the time, abandoning them was often a cause of anger and argument, and some libraries gave them up only reluctantly.

Currently, the idea of a national-level institutional repository is ruled out almost reflexively. "No monolithic system could ever work" is a common response whenever someone is bold enough to suggest it. But why? Why has a monolithic system worked so well for ResearchGate and Academia.edu? Why did it work so well for Digital Commons, whose customers have long been

confident that their content was indexed in Google Scholar than most distributed DSpace repositories? Why has it worked for HathiTrust, whose collection of digitized books surpassed 16 million volumes in 2017 (Furlough, 2017), and whose sheer mass has created exciting text-mining possibilities in one of the largest literary datasets ever produced (HathiTrust Research Center, 2016)?

A new web service called the Repository Analytics & Metrics Portal (RAMP) was developed in 2017 by Montana State University and its partners at OCLC Research, the University of New Mexico, and the Association of Research Libraries (OBrien, Arlitsch, Mixter, Wheeler, & Sterman, 2017). The purpose of RAMP is to accurately count file downloads from IR, and it works with almost any platform. Aside from providing good reporting numbers, RAMP's greater value to the community is in the dataset it is accumulating from the nearly 30 repositories that are currently subscribed. The dataset provides non-personally identifiable information about search engine queries that surfaced repository content, and it includes the URL or handle of each file that was downloaded from a given repository, meaning that all the metadata for each publication could be subsequently mined from each repository.

The research value of this dataset is enormous and unique. It could, for the first time, provide a profile of the content available in IR, internationally; it could be used to conduct a gap analysis to determine what users are searching vs. what they're finding; an analysis could be conducted of metadata across repositories, and using machine learning techniques, could reconcile local metadata to national standards. The possibilities are endless.

In numerous presentations around the nation where the team has introduced RAMP, one of the most frequent questions we have fielded is whether the code will be made openly available. The technical answer is yes, it could be. There is nothing secret or commercially valuable about the code itself. But the question once again reveals the mindset of defaulting to local implementation and local control, while completely missing the value of the centrally-accumulated dataset. Again, ResearchGate, Academia.edu, and bepress figured out long ago the value of a dataset that could be generated from a centralized infrastructure.

We are approaching a crossroad where technological capability, increasing budgetary restraints, and sustainability concerns are converging to give critical mass to the idea of a collective infrastructure. David Lewis's 2.5% Commitment proposal provides potential funding for just such an idea.

At the moment, we see two simultaneous paths:

1. Existing or new organizations take on this challenge, which will take a great deal of time, resources and energy; and
2. Using technology adoption life cycle analysis to create a group of innovators to lead the way.

We believe these two things need to happen in parallel, as one is a longer term solution and the other is a shorter term solution and clearly we need both. We'll deal with the technology

adoption life cycle analysis (the shorter term solution) in a later section, and here we'll deal with the organizational topic.

Currently, we see two non-profit, membership organizations that seem appropriate for dealing with the challenges: OCLC and DuraSpace. Both already have repository products in this space and both are already running hosted (Software-as-a-Service) offerings of their repository software. Yet, both have challenges to overcome before they could take on what we're describing in this paper.

What's in the Way?

Collaborations that reflect our weaknesses rather than strengths

As a profession, librarians excel at the creation of collaborative organizations and/or efforts. We're a group that dreams big, but we don't seem to execute as well as we dream. How many of the press releases issued at the launch of collaborative efforts *actually achieve* the intended goals? Perhaps the marketing value of the press release is adequate return, but one wonders if some level of accountability for achieving the stated goals needs to be applied?

Perhaps we should more actively challenge our existing collaborative organizations to step up to these challenges. OCLC is a large, community-owned and governed organization. Yet, it isn't actively leading these conversations. Perhaps the leadership doesn't see the need? Or perhaps the governance structure doesn't allow for the need to be addressed? One has to marvel at the description of the governance structure[1] on the OCLC website:

OCLC Governance organization	Number of positions
Executive Management Team	11
Board of Trustees	15
Global Council of Delegates	48

A total of 74 positions governing an organization generating $208M in annual operating revenues. While this is certainly "community governance" one must observe that the 2017 Annual Report notes: "OCLC has operated at a loss due to restrained price increases combined with heavy strategic investment into new services, as well as technology upgrades, facility renovations, and a staff resource realignment."[2]

Operating at a loss is one possible explanation for OCLC's limited leadership in this area, but limited agility may be another reason. Perhaps, as "community members" we need to push this organization to address the critical need for IR reform.

[1] https://www.oclc.org/en/about/leadership.html?cmpid=md_ab_leadership
[2] https://www.oclc.org/en/annual-report/2017/home.html

A much younger organization than OCLC, DuraSpace represents a community of users and is focused on a relatively narrow suite of open access repository platforms and storage. That narrow focus is a strength compared to OCLC's array of products and services, but whereas OCLC has become a large, global organization, DuraSpace is small and its personnel are limited. Were an initiative like the 2.5% Commitment to become a reality, funds could be funneled to DuraSpace to manage a single platform for all its subscribers. Simultaneously, its members would need to support a move away from the current two versions (XMLUI/JSPUI) of its DSpace platform to a single, shared platform, ensuring that money is being applied for maximum benefit to all.

As a profession, we need to analyze what is working and what isn't, learn from it, and apply the lessons to an organization that addresses the need of a next generation service for open, scholarly communications.

Collaborative governance models/agreements that lack accountability

Many new collaborative or organizational models are, quite appropriately, based on signed agreements between the participants. These agreements can range from a Memorandum of Understanding (MOU) to a full contract, and templates for both can readily be found on the web. These agreements help to ensure the organization's intent carries across time and personnel changes, as well as to clarify what happens when something goes wrong. Generally, the parties will agree that if done properly, these agreements get dropped in a file and stay there. They're typically brought out only if something has gone awry. However, there are lessons we need to apply to these documents to ensure they deliver the intended results. As is well documented in the library literature, successful collaboration contains the following elements:

1. Clear and shared goals
2. Risks and rewards are real and shared
3. The time period for achievement is clearly defined
4. Everything written down

Accountability is a crucial part of that list. Those who work in academia are well aware that the best of plans are subject to buffeting by changes in university administration, local, state, or national politics and, of course, the ever present question of funding availability (Grant, 2010). For example, one of the scenarios that could arise is when a university administrator instructs the library dean to change direction in a way that requires the library to redirect resources previously committed to a collaborative effort. If the collaboration agreement is well written, the library should be obligated to cover the cost of another collaborative organization being found and/or hired to handle the work to which the library had previously committed. If this wording were in place, it would provide protection to: 1) position the library dean to negotiate with the administrator to cover the cost of the impact of their new idea and; 2) ensure that the work component gets done so the collaborative effort can still succeed as originally planned. In such a scenario, collaborations would be more likely to succeed since accountability would be maintained for all concerned.

What can we do we do now?

The scenarios envisioned above, if implemented properly, will take years of concentrated, collaborative effort. The needed governance models will require intricate and complicated discussions and a lot of compromise. The funding models such as envisioned by Lewis, if they can be agreed upon by librarians, will then be complicated by compliance needs at the local, national, and international legal levels. MOUs or contracts will need to be developed, vetted, and approved. People will need to be hired, trained, and put into place to manage the new collaborative organization.

Of course, the problem with most collaborative approaches, as alluded to above, is that the core idea can die the proverbial death-by-a-thousand-cuts, be they compromises to the idea or length of time passed. The authors wonder if in fact, the right question has even been asked at the starting line? Perhaps, the question that needs to be asked is: How does the profession of librarianship move from the identification of a clear, wide professional need (an idea) to the realization of a working solution to address that need? The problem is really not unique to this profession, and many publications have been written on this subject. One that has endured the test of time and seen multiple editions as a result, is Moore's *Crossing the Chasm*, which describes a "technology adoption life cycle that shows the stages of idea adoption, starting with those who create the idea (innovators), to those that become the early adopters and then on through the early majority, late majority and finally, the laggards." He developed a chart (Figure 2), that shows the well-known Bell Curve, but applied here with the Innovators at the far left and the Laggards at the far right, and the majority in the large hump in the middle. The 'chasm' comes into play between the early adopters and the early majority, in that many products don't successfully navigate that transition and fall into oblivion. The book's focus is analyzing what is required for a successful transition.

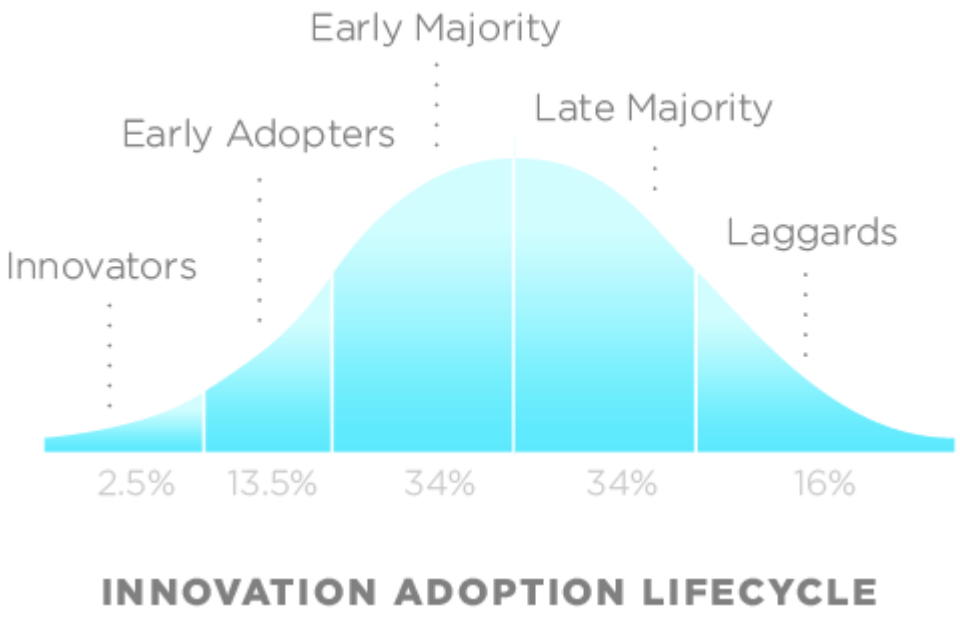

Figure 2: Technology Adoption Cycle taken CC BY 2.5, File:DiffusionOfInnovation.png, Wikipedia

There is a great deal of wisdom in this work for us to apply to the challenges being analyzed above and in trying to determine what we can do to meet them today and in the long term. One of the most important, for the purpose of this article, is to realize that major new ideas are created by a small group of “innovators” and “early adopters” and not by the early to late majority or laggards. As such, an approach we should consider is to identify who falls into the Innovators category, and then focus on enabling and empowering them to have access to the resources needed to bring their ideas at least to a working conceptual model. Examples of these types of people have been named throughout this work. Perhaps, as we look at creating a new organization or investing in an existing one to take on these challenges, this is a model we should utilize.

The technical infrastructure needed for any of the solutions discussed is available today. Lewis has shown that we have the money, collectively, so what remains is the will power. Multiple vendors now offer sophisticated, secure, agile, and scalable cloud-based hardware hosting platforms (IaaS - Infrastructure-as-a-Service) and at very reasonable costs. Many libraries run their repositories entirely on such cloud-based services and have found it superior in nearly every way to locally run hardware. Notre Dame University, for example, made this decision at the university level for the majority of their IT services (including the Libraries) noting: “It’s really one of our core IT strategies,” says Michael Chapple, senior director of IT Service Delivery at the University. “It’s the biggest technology change we’ve made in the last 10 years. We’re reinventing the way we provide IT service to the campus.” (Butler, 2015).

There are two important caveats that should be noted here: 1) The impact of the recent decision to end Net Neutrality could have severe consequences for cloud-based systems in higher

education, and thus the impact will have to be monitored closely. This decision was made within the last two weeks of the writing of this article, and thus, it's too early to predict; and 2) it bears repeating, we're only talking here about the IaaS (Infrastructure as a Service) layer, and not the SaaS (Software as a Service) total stack. Even then, it's not necessarily an enterprise solution, because often, all that is being done is moving the repository stacks into the cloud, but still with no network effect. For that to happen, the software must become true multi-tenant software, wherein the same software instance is supporting essentially all users. That is when the true economies of scale begin to apply; when a new version is released, the one instance is updated and therefore all users connected to that instance, instantly see the new functionality.

Summary

The proliferation of locally-run institutional repositories in academic libraries has created, in our estimation, problems that far outweigh the benefits. Of course, there are advantages to local control, but the thousands of extant repositories do not serve our users well, and they massively duplicate labor and other costs. The desire for local control has led to a fragmented landscape that is not sustainable, and in which we are unable to generate and analyze potentially powerful collective data about the content of our repositories and the interaction of users.

There is tremendous value to considering the possibility of consolidated data. All content, metadata, and use analytics are considered data in this paradigm, and the service and research possibilities are endless in such a paradigm. The artificial barriers of individual repositories must be eliminated to achieve this vision.

We must prepare for the rapidly evolving technological landscape needed to support scholarly communication. Agility is essential. Teams, organizations, infrastructure, and software tools must be prepared for this environment. Most importantly, we must determine a way forward that enables the innovators in our profession to rapidly build working models for deployment and refinement.

IR were born from a dream for open access to the intellectual output of the world's research. Libraries have made an extraordinary effort that has accumulated a massive amount of content, and that content must be protected and migrated to a more collective and useful platform. It is time to acknowledge that we have come up short, but we may yet achieve our dream if we can change course and re-architect to a large-scale, shared infrastructure that better serves our users.

Acknowledgement

Some of the thinking and research in this article was informed by previous funding from the Institute of Museum and Library Services. The authors would also like to thank Doralyn Rossmann for her review of the final draft.

References

Academia.edu. (2017). About Academia.edu. Retrieved December 10, 2017, from https://www.academia.edu/about

Arlitsch, K., & O'Brien, P. S. (2012). Invisible institutional repositories: Addressing the low indexing ratios of IRs in Google Scholar. *Library Hi Tech*, *30*(1), 60–81.

Borrego, Á. (2017). Institutional repositories versus ResearchGate: The depositing habits of Spanish researchers: Institutional repositories versus ResearchGate. *Learned Publishing*, *30*(3), 185–192. https://doi.org/10.1002/leap.1099

Burns, C. S., Lana, A., & Budd, J. M. (2013). Institutional Repositories: Exploration of Costs and Value. *D-Lib Magazine*, *19*(1/2). https://doi.org/10.1045/january2013-burns

Butler, B. (2015, December 14). How Notre Dame is going all in with Amazon's cloud. *Network World*. Retrieved from https://www.networkworld.com/article/3014599/cloud-computing/how-notre-dame-is-going-all-in-with-amazon-s-cloud.html

Chawla, D. (2017). Publishers take ResearchGate to court, alleging massive copyright infringement. *Science*. https://doi.org/10.1126/science.aaq1560

Coalition for Networked Information. (2017). *Rethinking institutional repositories: report of a CNI executive roundtable*. Washington, D.C. Retrieved from https://www.cni.org/wp-content/uploads/2017/05/CNI-rethinking-irs-exec-rndtbl.report.S17.v1.pdf

Confederation of Open Access Repositories. (2017). *Next generation repositories: behaviours and technical recommendations of the COAR Next Generation Repositories Working Group* (p. 32). Retrieved from https://www.coar-repositories.org/files/NGR-Final-Formatted-Report-cc.pdf

DuraSpace. (2017a). DuraSpace Registry [Non-profit]. Retrieved December 8, 2017, from http://www.duraspace.org/registry

DuraSpace. (2017b, September 20). All Documentation - DSpace Documentation [DuraSpace Wiki]. Retrieved January 14, 2018, from https://wiki.duraspace.org/display/DSDOC/All+Documentation

Friedman, T. L. (2017). *Thank You for Being Late: An Optimist's Guide to Thriving in the Age of Accelerations* (Reprint edition). New York: Picador.

Furlough, M. (2017, December 11). HathiTrust has reached 16 Million Volumes! Retrieved January 9, 2018, from https://www.hathitrust.org/16-million-volumes

Grant, C. (2010). A Partnership for Creating Successful Partnerships. *Information Technology and Libraries*, *29*(1), 5–7.

Grant, C., Soderdahl, P., & Beit-Arie, O. (2017, December). *Research Outputs: You Want Me to Do What?!? Finding a Way Forward for Librarians, Researchers and Other Stakeholders*. PowerPoint presented at the CNI Fall Member Meeting, Washington, D.C. Retrieved from https://www.cni.org/topics/user-services/research-outputs-you-want-me-to-do-what-finding-a-way-forward-for-librarians-researchers-and-other-stakeholders

HathiTrust Research Center. (2016, May 5). One of the world's largest digital libraries opens doors to text mining scholars. Retrieved January 9, 2018, from https://www.hathitrust.org/one-of-worlds-largest-digital-libraries-opens-doors-to-text-mining-scholars

Lewis, D. W. (2017, September 11). *The 2.5% commitment*. Indianapolis, Ind. Retrieved from http://hdl.handle.net/1805/14063

Lewis, S. (2014, April 20). Repository Maps. Retrieved November 29, 2017, from maps.repository66.org

Lynch, C. A., & Lippincott, J. K. (2005). Institutional Repository Deployment in the United States as of Early 2005. *D-Lib Magazine*, *11*(09). https://doi.org/10.1045/september2005-lynch

Markey, K., & Council on Library and Information Resources (Eds.). (2007). *Census of institutional repositories in the United States: MIRACLE Project research findings*. Washington, D.C: Council on Library and Information Resources.

MIT Libraries. (2017, December 30). MIT convenes ad hoc task force on open access to Institute's research. Retrieved December 30, 2017, from http://news.mit.edu/2017/mit-convenes-ad-hoc-task-force-open-access-research-0707

Moulaison Sandy, H., & Dykas, F. (2016). High-Quality Metadata and Repository Staffing: Perceptions of United States–Based OpenDOAR Participants. *Cataloging & Classification Quarterly*, *54*(2), 101–116. https://doi.org/10.1080/01639374.2015.1116480

OBrien, P., Arlitsch, K., Mixter, J., Wheeler, J., & Sterman, L. B. (2017). RAMP – the Repository Analytics and Metrics Portal: A prototype web service that accurately counts item downloads from institutional repositories. *Library Hi Tech*, *35*(1), 144–158. https://doi.org/10.1108/LHT-11-2016-0122

OBrien, P., Arlitsch, K., Sterman, L., Mixter, J., Wheeler, J., & Borda, S. (2016). Undercounting File Downloads from Institutional Repositories. *Journal of Library Administration*, *56*(7), 854–874. https://doi.org/10.1080/01930826.2016.1216224

Palavitsinis, N., Manouselis, N., & Sánchez-Alonso, S. (2017). Metadata and Quality in Digital Repositories and Libraries from 1995 to 2015: A Literature Analysis and Classification. *International Information & Library Review*, *49*(3), 176–186. https://doi.org/10.1080/10572317.2016.1278194

Park, J.-R. (2009). Metadata Quality in Digital Repositories: A Survey of the Current State of the Art. *Cataloging & Classification Quarterly*, *47*(3–4), 213–228. https://doi.org/10.1080/01639370902737240

Pinfield, S., Salter, J., Bath, P. A., Hubbard, B., Millington, P., Anders, J. H. S., & Hussain, A. (2014). Open-access repositories worldwide, 2005-2012: Past growth, current characteristics, and future possibilities: Open-Access Repositories Worldwide, 2005-2012: Past Growth, Current Characteristics, and Future Possibilities. *Journal of the Association for Information Science and Technology*, *65*(12), 2404–2421. https://doi.org/10.1002/asi.23131

Poynder, R. (2016, September 22). Q&A with CNI's Clifford Lynch: Time to re-think the institutional repository? Retrieved January 10, 2018, from https://poynder.blogspot.co.uk/2016/09/q-with-cnis-clifford-lynch-time-to-re_22.html

ResearchGate. (2017). About us fact sheet. ResearchGate. Retrieved from https://www.researchgate.net/press

Seaman, D. M. (2017). *Leading across boundaries: Collaborative leadership and the institutional repository in research universities and liberal arts colleges* (Dissertation). Simmons College School of Library and Information Science, Boston, MA. Retrieved from https://search.proquest.com/docview/1958944695?accountid=28148

University of Nottingham. (2017, November 23). Search or browse for repositories. Retrieved from http://www.opendoar.org/find.php?format=charts

University of Southampton. (2017, November). Welcome to the Registry of Open Access Repositories. Retrieved November 24, 2017, from http://roar.eprints.org

Van de Sompel, H. (2017, December). *Scholarly Communication: deconstruct and decentralize*. Internet presented at the Coalition for Networked Information fall member meeting, Washington, D.C. Retrieved from https://www.slideshare.net/hvdsomp/paul-evan-peters-lecture

Van de Velde, E. (2017, December 28). Let IR RIP [Non-profit]. Retrieved December 28, 2017, from http://scitechsociety.blogspot.com/2016/07/let-ir-rip.html

Van Noorden, R. (2017). Publishers threaten to remove millions of papers from ResearchGate. *Nature*. https://doi.org/10.1038/nature.2017.22793

Van Westrienen, G., & Lynch, C. A. (2005). Academic Institutional Repositories: Deployment Status in 13 Nations as of Mid 2005. *D-Lib Magazine*, *11*(09). https://doi.org/10.1045/september2005-westrienen

Wenzler, J. (2017). Scholarly Communication and the Dilemma of Collective Action: Why Academic Journals Cost Too Much. *College & Research Libraries*, *78*(2), 183–200. https://doi.org/10.5860/crl.78.2.16581

Wickham, J. (2011, October 5). Institutional Repositories: staff and skills set. University of Nottingham. Retrieved from http://www.rsp.ac.uk/documents/Repository_Staff_and_Skills_Set_2011.pdf

Wu, M. (2015). The Future of Institutional Repositories at Small Academic Institutions: Analysis and Insights. *D-Lib Magazine*, *21*(9/10). https://doi.org/10.1045/september2015-wu