Addressing Institutional Repository (IR) Underutilization:

Ingredients for Countering Valid and Perceived Copyright,

Plagiarism, and Low Value-to-Effort Ratios that Inhibit Engagement

Ryan Spellman

School of Information, Kent State University

LIS60633 Digital Curation

Dr. Karen F. Gracy

05/01/2022

Introduction

Institutional repositories (IRs) can trace their roots all the way back to the 1990s (Bicknese, 2003; Sabharwal, 2021). A more formal, wide-spread adoption of IRs began to sprout up across research-focused institutions in the United States and United Kingdom following a Scholarly Publishing and Academic Resources Coalition position paper by Ryam Crow that defined the IR as a means to ingest and preserve the scholarly work of university communities (Abrizah et al., 2014; Shukla & Ahmad, 2018). At the time of this writing, this point of more mainstream emergence of the IR concept was some 20 years ago, yet IRs are still concerning themselves with a lack of deposits from scholars (Shukla & Ahmad, 2018). In fact, in most situations IR content seems to primarily come from harvesting and mediation efforts by IR staff (Covey, 2011).

But what is the problem with IR staff taking responsibility for collecting the content for the IR? Part of the issue lies in the dwindling staff resources that research libraries, archives, and museums (LAMs) are facing. The Association of Research Libraries' (ARL) (n.d.) statistics on the ratio of FTE (Full Time Equivalency) library staff per student has steadily dropped to a -44% difference between 1998 and 2018. These numbers include professional staff, support staff, and student assistants, so this represents a significant shift. Relying on staff to harvest and mediate content for IRs is not sustainable. Additionally, as we will discuss later, merely amassing content is not sufficient for a successful IR. Through a review of IR literature about direct experiences and user studies, the following paper examines three major inhibitors to designated community engagement with IRs, approaches that positively impact rates of self-archiving and utilization of IRs, and how these insights can contribute to a recipe for a healthy IR.

Three Reasons Users Do Not Engage IRs

Stakeholders contributing content to IRs at academic research institutions affirm the need for the IR (Abrizah et al., 2014). Sustaining contributions is paramount to an IR's success, and engagement with its designated community is the impetus (Carr & Brody, 2007). First, we will review some reasons scholars shy away from engaging with IRs. Through this lens, we will explore three inhibitors to IR engagement that also lead to reasons scholars are using their IRs, including plagiarism concerns, fears of plagiarism, and a perceived low value-to-effort ratio. As we delve into these concerns, some benefits float to the surface that risk going unrealized by a IR's designated user community, feeding into our recipe for a healthy IR.

Plagiarism Concerns

According to OpenDOAR (n.d.), the United States alone contains 920 open access (OA) repositories. Once a fringe concept, the OA movement is now firmly in the mainstream (Banks, 2015). Upcoming IRs are likely to take on OA policies, and many currently established IRs are already there. OA is a key motivation for digital curation, and helps contributors maximize the impact of their research (Oliver & Harvey, 2016). However, OA also brings concerns over plagiarism into the mind of would-be IR contributors (Abrizah et al., 2014; Davis & Connolly, 2007; Covey, 2011; Shukla & Ahmad, 2018). Arduous work goes into scholarly research, so concern over data and findings being plagiarized is understandable.

Copyright Concerns

In addition to plagiarism concerns, scholars commonly purport that contributing to an IR may risk infringing copyright policies (Abrizah et al., 2014; Covey, 2011; Davis & Connolly, 2007; Shukla & Ahmad, 2018). This concern is rooted in confusion surrounding what they are permitted to do with their work already published in other journals (Davis, 2007). These are valid concerns that can make the decision appear as an either-or type of situation, unless users put the time into sorting out what is permissible under their contracts. The problem with this is that time and effort itself can be an inhibitor to IR engagement, especially for busy scholars and researchers.

To counter such confusion over copyright issues, one could suggest skipping the journal publication process altogether. However, scholars and researchers see publishing in journals as a way to spread their work widely and benefit their reputation (Shukla & Ahmad, 2018). One study featuring 500 questionnaires sent to 4 different research laboratories in India suggests that 58.86% of scientists and 49.78% of scholars prefer to publish their research in scholarly publications rather than submit to IRs, Open Access Journals, Conferences, or in-house publications combined (Shukla & Ahmad, 2018). In a University of California, Berkeley study, scholars were in support of publicly available research, but concern over advancing their stature in their discipline through peer-reviewed publications superseded any altruistic support for open access (Davis & Connolly, 2007). As Shukla and Ahmad (2018) expertly state, peer-reviewed prestige creates “inertia” against IR participation. It is no wonder that IRs are often second in line if this inertia is amplified by confusion over copyright.

Even if the intention and willingness is there, and the research is not yet published, the logistics of being dually published in peer-reviewed journals and IRs can be murky. Some journals will not accept previously published work, so does this mean that submitting to an IR will thwart any future peer-reviewed journal prospects? Some faculty have this concern, exemplifying how tricky these aspects of disseminating scholarly works can be (Covey, 2011; Davis & Connolly, 2007). What we see here is a snowball effect surfacing out of just the singular topic of copyright issues that can quickly make contributing to an IR less and less appealing.

Onerous Value-to-Effort Ratios

Additionally, faculty sometimes see self-archiving in an IR as time consuming and too much work (Covey, 2011; Abrizah et al., 2014). Especially beyond publishing solely in the IR, contribution can feel like double the work for little return if they are already publishing their work elsewhere or self-archiving on a personal website. Teperek et al. (2016) conducted information sessions in which prior to learning more about IRs, faculty clearly felt participation was just another thing to do on top of everything else. Scholars and researchers are busy, and IR staff should be able to help ease this burden. However, as touched on earlier, relying solely on staff to build and maintain an IR's content is not sustainable and not true to the spirit of the IR.

It is important to note that, as covered in more detail in the recipe for a healthy IR, there is indeed a benefit to harvesting and mediating content to build up an IR's foundation. However, relying exclusively on these measures is not sustainable for institutions with shrinking numbers of library staff. Additionally, exclusively aggregating content through efforts of library staff rather than encouraging self-archiving by researchers hurts the level of engagement and suggests that the value of the IR is either genuinely lacking or perceived to be in some regard.

The Importance of Addressing These Inhibitors of IR Engagement

Successful IRs should be utilized rather than serve as dark storage for the sake of storage alone. If this were the goal, the OA movement would not be so firmly in the mainstream. Granted, it is common to gauge the merit of an IR by comparing how many information objects they hold (Carr & Brody, 2007). But this is only part of the story and may not accurately represent the IR's actual utilization by a designated community. For instance, content that is batch uploaded can skew such content level and growth statistics (Davis & Connolly, 2007). An IR's success should be determined by both its use *and* usefulness to scholars rather than purely the amount of content they have amassed (Shearer, 2003, as cited in Abrizah et al., 20014).

Copyright concerns, plagiarism issues, and negative value-to-effort ratios are easy to empathize with if we put ourselves into the designated community's point of view. However, through working to address these issues, IRs are finding ways to spur engagement in ways that bring massive improvements to their services and resources. To help draft our recipe for a healthy IR, we will investigate some of these approaches and reasons researchers and scholars are engaging their IRs.

Recipe for a Healthy IR

Solutions to these inhibitors of IR engagement are sprinkled throughout the following recipe. That said, it is important to note that this recipe takes a special focus on aspects that will *help* cook up an IR that is appetizing to the user; far from the full and only recipe needed. This is to be viewed as ingredients to a small component of a healthy IR, this component in turn being just one ingredient among many in the larger recipe. For instance, recipes for policies, sustainable funding and access, audit practices, and much more must come into play for a healthy IR. Even so, this mini recipe for user engagement can serve as a starting point for deeper discussions around how an IR can better meet the needs of its designated community.

Build a Good Base: Taste Testing

There is no doubt that an IR must reach “a critical mass [...] to attract users and additional content” (Covey, 2011, p. 2). There are many ways to build up to this critical mass. For instance, Carnegie Mellon University took it upon themselves to harvest work from their faculty that was self-archived on their websites as long as the publisher was allowing OA and included a SHERPA / RoMEO database URL of their OA policies (Covey, 2011). Even with our discussion of dwindling staff resources, an initial investment of time to amass content can make an IR more appealing than asking researchers and scholars to participate with something that looks desolate.

Starting with a similar approach, Oregon State University Libraries and Press (OSULP) staff were seeking out faculty articles to deposit into the IR on their behalf (Zhang et al., 2015). Later, OSULP innovatively began using Web of Science, a citation database, to filter publications down to faculty at their institutions. This metadata was then put into a log that indicated what was OA, what was currently under an embargo and for how long, with staff subsequently sending email requests as articles became available for deposit (Zhang et al., 2015). What is so innovative about this approach is that it automates the work of seeking content while eliminating concerns about copyright. In OSULP’s example, they had a deposit rate increase of 45% after running this Web of Science project for just one year (Zhang et al., 2015).

We must note once again, though, that the level of content is only part of the picture. But how can an IR get users to self-archive into their collection? As it turns out, our recipe is not quite sequential in the way you might follow a recipe to your grandma's famous chocolate chip cookies. You need this baseline mass of content, but the IR itself needs to be built first. This is where we start inviting more cooks into the kitchen. For example, Teperek et al. (2016) examined results of a mandate to deposit to an IR at the University of Cambridge. Next, they began pulling in users to help shape RDM services while providing education opportunities about IRs, and investigated the impact this user involvement had on these engagement rates. Their conclusion was that mandates are helpful, but it is the mixture of mandates combined with involving the user in IR decisions that brings the most benefit to the IR (Teperek et al., 2016).

User-Informed Design: Let the User Help Write and Revise the Recipe

User involvement is a key component of the recipe. Not just in the sense of using the IR, but actually *building* it. A critical mass of users is necessary for an IR to be successful (Abrizah et al., 2014). During a series of interviews with users of the Emblematica Online digital library, one scholar makes the point that IRs are designed by computer experts who have the technical expertise but have a lack of subject knowledge that can cause accessibility issues (Green & Lampron, 2017). If looking to have this critical mass of actively engaged users, we need to keep such insights close to heart. Users can help guide the construction or revision of an IR in ways that will tip the scales on the value-to-effort ratio in favorable ways, returning greater value than previously expected.

It is also important that IRs do not get hung up on a stereotypical designated user base. When thinking about IRs, thoughts of scientific data and disciplines can dominate one's mind. Seaman (2011) points out that if we fail to incorporate information needs of non-science disciplines such as the humanities, the IR will suffer low content levels. This further speaks to how integral bottom-up approaches that embrace community feedback are to this recipe. IRs cannot just expect those in the humanities to jump right on board, though. IRs need to learn what they need and make it available to them. Therefore, decisions need to be made with interdisciplinarity in mind so that the conceptual designated user base is broadened.

Seaman (2011) expands this point by sharing how including digitization services, mass data storage, tagging features, and user commentary can make IR participation more attractive to humanities scholars – all of which are revelations that came from engaging with a previously disengaged humanities community. The recipe for a healthy IR is going to require services and resources that users want; features that will encourage users to both contribute and consume. No one can better inform such features than the IR's designated community, and this community must have an interdisciplinary scope to maximize engagement.

In another example, Sweeney et al. (2017) identified a need for satisfying both presentation and preservation of their community's research and met with digital scholars to pin down exactly what the best solution would be. The project ended in a custom API that connected repository data to a highly customizable presentation layer via a custom WordPress plugin. In this way, they are providing stable long-term storage, and providing a desired user service that will promote IR contributions (Sweeney at al., 2017).

Interviews, focus groups, surveys, and other user-centered design approaches will not only reveal what users want, but why they might shy away from an IR. These reasons can also help shape an IR's design and feed into education and outreach features. If IR staff can address reasons users might resist the IR, or pinpoint misconceptions surrounding things like copyright and plagiarism, outreach and education efforts can go a long way toward improving user engagement.

Education and Outreach: Put it on the Menu and Promote the New Savory Dish

The last component of this little recipe is incorporating education and outreach services for researchers and scholars. Outreach efforts are the stage from which LAM staff can advocate for OA and IR participation. One way to begin such conversations is to discuss why IRs are being used. For instance, as part of their study on the underutilization of Cornell University's DSpace installation, Davis & Connolly (2007) unearthed some insightful reasons that faculty were using the resource. These included: the permanence afforded by the IR's stewardship of keeping the data accessible (i.e., format conversions), meeting requirements by grants that manuscripts and supplemental data be publicly accessible, the fast turnaround on disseminating their work, and using the IR as a place for immediately registering a new idea to "stake a claim" on it rather than waiting for a conference publication to be completed (Davis & Connolly). Spreading awareness about some of these examples can help shift the line of thought that reluctant researchers and scholars are set upon.

A brilliant example of education and outreach benefits shines in the work of Teperek et al. (2016). Through direct discussions with users, structured interviews, surveys, and open meetings with funders, they were able to pull together a remarkably successful plan for research data management (RDM) services that improved and sustained user engagement. In short, this resulted in a centralized web resource for information about RDM services, a simple form to deposit content, and an RDM training and support program that powerfully aligns with this topic of education and outreach.

Teperek et al.’s (2016) training and support program went so well that it led to developing a need to recognize researchers that excelled with RDM services they were learning about. These “Data Champions” as they were called became involved in the RDM training efforts, which in turn helped with discipline-specific training and sustaining IR engagement levels among their peers (Teperek et al., 2016). Although not included in Teperek et al.’s report, one can imagine how such opportunities would be an excellent avenue to help dispel and alleviate copyright and plagiarism concerns that inhibit participation with an IR.

Conclusion

Library staff might help build a good baseline of content for an IR, but it is not sustainable and not how a healthy IR operates. There are many concerns that inhibit users engaging with an IR, such as copyright, risks of others plagiarizing their work, and a perception of low returns on effort. However, these can be addressed with the proper recipe. Building a good baseline of content to encourage participation, making the design and revision of IR resources and services user-driven, and taking time to conduct worthwhile education and outreach services to the designated community can cover a lot of ground toward lowering these barriers to IR participation. Although not a complete recipe, it is a recipe to an important ingredient of the final course.

References

Abrizah, A., Hilmi, M., & Kassim, N. A. (2015). Resource-sharing through an inter-institutional repository. *Electronic Library*, *33*(4), 730–748. https://doi.org/10.1108/EL-02-2014-0040

ARL. (n.d.). *Graph 5: Resources per student in ARL university libraries, 1998-2018.* https://www.arl.org/wp-content/uploads/2019/09/resources-per-student.pdf

Banks, M. (2015, September 8). Open access publishing: What it is and how to sustain it*. American Libraries.* https://americanlibrariesmagazine.org/2015/09/08/open-access-publishing/

Bicknese, D. (2003). Institutional repositories and the institution's repository: What is the role of university archives with an institution's on-line digital repository? *Archival Issues, 28*(2), 81–93. http://www.jstor.org/stable/41102081

Carr, L., & Brody, T. (2007). Size Isn't Everything: Sustainable Repositories as Evidenced by Sustainable Deposit Profiles. *D-Lib Magazine*, *13*(7/8), 1–1. https://doi.org/10.1045/july2007-carr

Covey, D. T. (2011, April 18). Recruiting content for the institutional repository: The barriers exceed the benefits. *Journal of Digital information, 12*(3). https://journals.tdl.org/jodi/index.php/jodi/article/view/2068

Davis, P. M., & Connolly, M. J. L. (2007). Institutional Repositories: Evaluating the Reasons for Non-use of Cornell University's Installation of DSpace. *D-Lib Magazine*, *13*(3/4), 3–3. https://doi.org/10.1045/march2007-davis

Green, H. E., & Lampron, P. (2017). User Engagement with Digital Archives for Research and Teaching: A Case Study of Emblematica Online. *Portal: Libraries & the Academy*, *17*(4), 759–775. https://doi.org/10.1353/pla.2017.0045

OpenDOAR. (n.d.). *OpenDOAR statistics: An overview of the data held in OpenDOAR.* https://v2.sherpa.ac.uk/view/repository_visualisations/1.html

Sabharwal, A. (2021) Institutional repository engagement framework: Harnessing resources, structure, and process for strategic plan support in higher education. *Journal of Electronic Resources Librarianship, 33*(3), 137-155, https://doi.org/10.1080/1941126X.2021.1949150

Seaman, D. (2011, March/April). Discovering the information needs of humanists when planning an institutional repository. *D-Lib Magazine, 17*(3/4). https://doi.org/10.1045/march2011-seaman

Shukla, P., & Ahmad, N. (2018, February). Impact of Institutional Repositories on Scholarly Practices of Scientists. *Library Philosophy & Practice*, 1-18. https://digitalcommons.unl.edu/libphilprac/1631

Sweeney, S. J., Flanders, J., & Levesque, A. (2017). Community-Enhanced Repository for Engaged Scholarship: A case study on supporting digital humanities research. *College & Undergraduate Libraries*, *24*(2-4), 322–336. https://doi.org/10.1080/10691316.2017.1336144

Zhang, H., Boock, M., & Wirth, A. (2015). It Takes More than a Mandate: Factors that Contribute to Increased Rates of Article Deposit to an Institutional Repository.

Journal of Librarianship & Scholarly Communication, *3*(1). https://doi.org/10.7710/2162-3309.1208

Zheng Ye (Lan) Yang & Yu Li. (2015). University Faculty Awareness and Attitudes towards Open Access Publishing and the Institutional Repository: A Case Study. *Journal of Librarianship & Scholarly Communication*, *3*(1). https://doi.org/10.7710/2162-3309.1210