Breaking It Down: A Brief Exploration of Institutional Repository Submission Agreements

Breaking It Down: A Brief Exploration of Institutional Repository Submission Agreements

Amanda Rinehart [a,*], Jim Cunningham [b]

[a] Ohio State University, 175 W. 18th Avenue, Columbus, OH 43210, USA
[b] Illinois State University, 201 North School St., Normal, IL 61790-8900, USA

ABSTRACT

Institutional repositories typically have a submission agreement that is meant to protect the institution hosting the repository and inform submitters of their rights and responsibilities. This article examines how various libraries have created submission agreements, enquires as to issues surrounding them, and identifies commonalities and unique statements. The authors deployed a survey to institutional repository administrators listed in OpenDOAR in the United States. Approximately 7% of the 304 potential institutional repository managers responded. Library administrators, institutional repositories managers/architects, and legal counsel were the most likely to have input into the creation of the submission agreement; scholarly communications librarians were involved only 20% of the time. Although submission agreements averaged 282 words arranged in 9 sentences, their reading complexity requires a university degree. Commonalities include characterizing the agreement as a non-exclusive license, indicating the submitter's responsibility for obtaining permissions for any content that they did not produce, and confirming the right of the submitter to enter into the agreement. Submission agreements are generally complex and do not accommodate the common practice of mediated submission. Sharing submission agreements publicly may lead to simplified and standardized language and reduce barriers to submitters.

© 2016 Elsevier Inc. All rights reserved.

1. INTRODUCTION

At the turn of the century, the initial experiments to implement institutional repositories (IRs) were actively ongoing. Crow (2002) differentiated IRs from other repository types as having four primary characteristics: institutionally defined, scholarly in nature, cumulative and perpetual, and open and interoperable. It was predicted that the institutional and open components would result in increased prestige, as well as capture new and emerging forms of digital scholarship (Crow, 2002; Lynch, 2003). The cumulative and perpetual component addressed concerns regarding the ephemeral nature of the digital world and the need to preserve it, while the open and interoperable component represented a welcome alternative to traditional, monopolistic and rigidly controlled commercial journals (Duranti, 2010; Robertson and Borchert, 2014; Crow, 2002; Bastos, Vidotti, and Oddone, 2011; Bergstrom, Courant, McAfee, and Williams, 2014). The advent of disruptive technologies allow the collection and distribution of material at a lower cost (Blythe and Chachra, 2005; Heath, 2009; Odlyzko, 1997) and the sheer increase in the overall volume of scholarly material

(Budd, 2006; Rawls, 2015) made IRs an attractive addition to the academic library's suite of services.

In 2005 about 40% of research institutions in the US had IRs, a rate that rose to 62% in 2007 (Lynch and Lippincott, 2005; Markey and Council on Library and Information Resources, 2007). As of 2015, OpenDOAR listed 455 IRs in the US, which is an estimated 10% of degree-granting post-secondary Title IV institutions having an IR (University of Nottingham, 2015; U.S. Dept. of Education, 2013). This would indicate that the vast majority of higher educational institutions have either chosen not to implement an IR or may still be faced with that decision.

1.1. SUBMISSION AGREEMENTS

A key part of launching an IR is creating the submission agreement document. This is a formal legal agreement that "defines the relationship between the individual submitting the content and the institution that is operating the repository" and grants "the repository the necessary rights to disseminate an author's work while affording the institution a measure of protection against submitted content that may violate legal or ethical boundaries" (Gilman, 2013, section Repository Submission Agreements and Contracts and Licenses). It may be a paper copy, a check box on a 'click-through' agreement, or other digital signature mechanism (Jones, 2007). These submission agreements may also be referred to as a license or a deposit agreement. For simplification of

---

* Corresponding author at: Ohio State University, 175 W. 18th Avenue, Columbus, OH 43210, USA.
E-mail addresses: Rinehart.64@osu.edu (A. Rinehart), jlcunni@ilstu.edu (J. Cunningham).

language, the term 'submission agreement' or SA will be used for the remainder of this publication. As well, the term 'author' will refer solely to the generator of the material, while the term 'submitter' refers to the person that signs or approves the SA, regardless if they are the actual author.

The primary motivation for an IR SA is to minimize the legal risk to the host while maximizing the ability to re-use the material. The SA should "clearly indicate that the repository is not responsible for any mistakes, omissions or infringements in the deposited work" (Jones, Andrew, and MacColl, 2006, p. 150). In particular, it should state that:

> "In the event of a breach of intellectual property rights, or other laws … the repository … is not under any obligation to take legal action on behalf of the original author, or other rights holders, or to accept liability for any legal action arising from any such breaches" (British Library, n.d., section liability).

While protecting the institution, SAs may also address a number of other components, such as: submitters' rights and responsibilities, end-user permissions for both full-text and the metadata created during submission, and various policies and laws. Gilman (2013) recommends that the SA cover, at a minimum, the submitter's rights to enter into the agreement, granting of a license to the institution and assurances from the author of the legality of the content. However, many SAs include other submitter's rights and responsibilities, as well as end-user permissions. Although SAs are commonplace, the importance of them should not be underestimated. As early as 2002, a coalition of library organizations found that more "Work is needed on models for obtaining copyright clearance and models for contracts or agreements between rights owners/producers and archives/libraries" (RLG/OCLC Working Group on Digital Archive Attributes, Research Libraries Group, and OCLC, 2002). As well, the SA is mentioned six times in the Trustworthy Repositories Audit and Certification: Criteria and Checklist and is also a key part of the administrative function of the Open Archival Information System Reference Model (Center for Research Libraries and OCLC, 2007; Consultative Committee for Space Data Systems, 2002).

1.2. SUBMITTER RIGHTS

Submitters may have the right to determine when end-users may have access to the material (embargo period), the circumstances for removal, and the terms of any re-use (Jones et al., 2006). Since one hallmark of the IR is openness, a re-use license is necessary to delineate how end-users may access, re-use and distribute the material (Jones et al., 2006). This license is typically separate from the SA, and may range from the traditional 'All rights reserved' statement, which indicates that end-users cannot use the material for any purpose without permission from the owner, to more liberal Creative Commons licenses (Creative Commons, n.d.). While creative commons licenses have become more popular, there is some controversy over which one constitutes true Open Access (Andersen, 2015). Although the different re-use licenses allow for flexibility, they add to the number of new decisions that a submitter may be confronted with when agreeing to a SA.

1.3. SUBMITTER RESPONSIBILITIES

It is the intention that the SA be signed or approved by someone who holds the copyright to the work or has permission from the copyright holder. While authors initially hold the copyright to their works, they often transfer the copyright to a commercial company during the publication process. IR submitters "are expected to read the licence carefully and ensure that they have the right, as confirmed by the publishers of a paper that might have appeared in a journal, to deposit the item in the IR" (Tedd, 2006, p. 252). Most SAs then ask the submitter "to grant the institution a license to use the work in question…[and]…in all cases, the grant of rights should be *nonexclusive*" (Gilman, 2013, section Grant of License to the Institution; author's emphasis). By placing

copyright clearance responsibilities on the submitter, the host of the IR is legally protected.

However, the already-complicated and confusing world of copyright gets even further so when dealing with various types of items: articles, books, raw research data, unpublished reports, works for hire, etc. Different types of material may be subject to other laws in place of, or in addition to, copyright law. A few examples include export control laws, the Federal Educational Rights and Privacy Act, Americans with Disabilities Act, and HIPPA (U.S. Department of State, n.d.; U.S. Department of Education, n.d.; U.S. Department of Health and Human Services, n.d.).

In addition, there may also be local policies that appear to be, or are, conflicting with the SA. It is common for higher education institutions to have an existing intellectual property policy that states what types of works can belong to an individual and which are the property of the institution. Often these intellectual property policies were created prior to the advent of technology that allows for easy distribution of digital material and without consideration to the diverse types of scholarship that now exist. Therefore, the intellectual property policy may need additional interpretation for researchers to understand when they must seek permissions from their employer prior to submission. Some SAs seek to remind the submitter of these additional obligations.

1.4. END-USER PERMISSIONS - METADATA

Metadata that makes the content discoverable is created during the submission process, by the submitter or the IR staff. There is growing interest in mining this metadata with automated computer scripts (Swanson and Rinehart, 2016). As such, "The host repository may, or may not, wish to claim copyright in any additional data created during the submission and subsequent archiving of the work" (Jones et al., 2006, p. 150).

> "It is advisable to state an explicit re-use policy for metadata, otherwise people will have to make assumptions – the failsafe being that they do not have permission … OpenDOAR goes further in recommending that you even allow your metadata to be reused commercially. This is because any loss of potential revenue is far outweighed by the benefits accrued from the additional exposure of your material." (JISC, n.d.a, section re-use of metadata).

As interest grows in bibliometrics and meta-analysis, metadata ownership and automated access is becoming more important (Dollar, King, Knight, and Leonard, 2014). However, specifying who may access and re-use metadata, particularly when it is created by multiple people, adds to the complexity to the SA.

1.5. END-USER PERMISSIONS - FULL-TEXT

In addition to metadata mining, web robots are increasingly used to harvest full texts for various purposes. Some of these purposes are beneficial to the IR - the obvious case being indexing by search services such as Google. However, the IR may or may not want a third party to make cache copies of complete works, particularly as a collection of works often has greater value than the works individually. As well, IRs may decide to harvest works from each other, triggering a number of unexplored questions. OpenDOAR recommends that repositories allow transient harvesting of full items by robots for benign or beneficial purposes (JISC, n.d.a). Since re-use statements almost always assume that end-users are individuals, this practice may not be considered in most SA documentation.

1.6. COMPLEXITY OF THE SA

Because of these complex issues – institutional protections, submitter rights and responsibilities, end-user permissions, and the

plethora of laws and policies that may apply – it is easy for an SA to become a lengthy morass of legal language. The necessary level of detail and complexity must be balanced with the desire to keep it understandable. As Jones et al. (2006) state, "these notices should be of a simple design, which can be easily understood" (p. 152) and Gilman recommends that libraries "construct clear and effective submission agreements that will grant the repository the necessary rights to disseminate an author's work while affording the institution a measure of protection " (2013, Chapter Contracts and Licenses). If users do not understand the SA, then they may be deterred from submitting their work, or submit in violation of law or policy.

Due to the complexity of the issues outlined above, it is not uncommon for legal counsel to be consulted during the creation of the SA (Barton and Waters, 2004). As well, it is understandable that "some people have felt unable to sign the licence as we ask them to agree that they are the copyright holder and/or have the right to grant the licence – this has been overcome by offering advice and support on copyright policies" (Barwick, 2007). It is now common practice in the US for IRs or their institutions to offer some level of support for copyright clearance (Hanlon and Ramirez, 2011).

1.7. MEDIATED SUBMISSION; THE END RUN AROUND THE SA

In addition to copyright clearance support, some authors rely on other people to submit their material. Submitting works on behalf of an author is termed a 'mediated submission' or 'mediated deposit', and prevents the author from agreeing to the SA (JISC, n.d.b, section Mediated Deposit). Hanlon and Ramirez (2011) found that about 36% repositories provided mediated submissions and another 54% provided a blend of self-submission and mediated submissions. Mediated submissions place the legality of the SA into question, and therefore, "repository administrators may wish to consider collecting additional proof of the permissions being granted when deposit is taking place via a third party… While [these do not] constitute a complete legal defence, they are likely to demonstrate a responsible approach" (JISC, n.d.b, section Mediated Deposit). As Markland and Brophy (2005) stated "the legality of 'click through' licenses [is not] always clear, particularly when it is the repository manager, not the academic author who is depositing" (p. 35). In the case of mediated submission, Gilman recommends that "adequate information [be] gathered by the repository platform to allow reasonable certainty as to which individual completed the agreement …[and]… If work is submitted on behalf of a another person (e.g. an assistant), … the language in the submission agreement [allows] that proxy to agree on behalf of the author"(2013, section Repository Submission Agreements). If this language is not present in the SA, it is possible that this practice reduces, or even eliminates, its meaningfulness.

1.8. MAKING SAS EASY

In 2003 Lynch speculated that IRs would "offer the opportunity for bottom-up, community-driven, consensus development about rights and permissions" and would result in a "relatively small number of sets of terms and conditions that can cover the majority of the materials" (p. 7). However, a few years later, Markland and Brophy (2005) noted that "Repository managers have differed in their attitudes towards what needs to be done by whom to keep within strict compliance with the law, a reflection perhaps on the vagueness of available information" (p. 35). Indeed, Gilman's excellent work on the topic calls for librarians "to construct clear and effective submission agreements", while simultaneously noting that "it is necessary to be aware not only of international differences, but of differences between states' laws and in the varying applicability of laws between private and public educational institutions", "defamation, obscenity, privacy and accessibility laws also vary between countries", and that SAs are "shaped also by the local institutional setting – for example, issues such as liability risk tolerance and the institutional insurance coverage" (2013, sections

Contracts and Licenses, Conclusion: Context changes, but ethics remain). All these considerations make the SA a difficult document to construct. Despite the many examples of IR SAs, independent evaluation of SAs are difficult to find and peer-reviewed literature on the topic is scarce. This knowledge gap is problematic for US institutions that wish to craft or change a SA. This brief exploration of SAs and the practices surrounding their creation is a proof-of-concept effort in identifying issues surrounding SA creation and specific IR SA language. Ultimately, these methods may be used to 1) aid those in the position of crafting an IR SA, and 2) provide empirical data to inform larger discussions regarding the legal partitioning of rights within the scholarly communications landscape.

2. MATERIALS AND METHODS

Using the e-mail distribution service from OpenDoar (University of Nottingham, 2014), the authors deployed a survey (Appendix A) to approximately 304 IR administrators in the United States. The survey was deployed via Survey Select software, and due to logistics, IR administrators were only contacted once. The survey was deployed Wednesday, October 9th, 2013 and responses were accepted for four weeks. The survey solicited information about the creation and deployment of the SA and general information about the IR and host institution. The results were summarized and analyzed using descriptive statistics and concept mapping.

For initial text analysis, all SAs were loaded into Voyant Tools (https://voyant-tools.org/) and frequencies of common words (such as 'non-exclusive', 'rights', and 'intellectual property') were noted. After initial exploration, each SA was read and entire phrases were classified into common concepts, such as 'copyright permissions', 'infringement of rights – general' and 'infringement of rights – privacy'. Therefore, some concepts bear great similarity to each other, but are listed separately, as they represent different SAs. A spreadsheet was used to track frequencies of each concept. Due to complex wording, some phrases included multiple concepts. In these instances, the concept category was either widened to be more inclusive, or the phrase broken down and counted in more than one category. The final concept categories and frequencies were then confirmed by the second author and any discrepancies resolved through discussion. These concepts were then further sorted into four categories: protecting the host institution from unlawful activities, delineating the rights of the submitter, making specific promises to the submitter, and reminders of specialized circumstances or options. For SA language analysis, 22 complete SAs were submitted. However, two institutions submitted SAs that are specific to student work and these SAs are considered separately.

To aid in visualization, the shortest and longest SA are compared to all of the SAs by removing any proprietary names and creating collocate graphs in Voyant Tools. A collocate graph is a network graph that represents higher frequency terms (in blue) and terms that appear in proximity to them (in orange).

3. LIMITATIONS

The term 'submission agreement' was not defined in the survey, although the SAs that were collected all fit the basic definition given in the introduction of this work. Due to the modest response rate, the results presented here are exploratory in nature. As well, the majority of responses are from doctorate institutions, which skews the observations towards that community. However, considering that all institutions are under similar legal burdens, these results may well apply to smaller, and/or non-research institutions. The survey was limited to the United States, as SAs are heavily influenced by copyright laws, and these laws differ by country. Therefore, the results of this survey are only generalizable within the United States. As well, since only IRs are addressed in this study, other types of repositories may require different SAs, which are not within the scope of this article.

4. RESULTS

4.1. DEMOGRAPHICS

Twenty-two of the potential 304 IRs responded completely, and another 13 responded partially. This yielded a 7% complete response rate. However, all responses were included when possible, resulting in an 11.5% response rate for some questions. Responses were relatively even across all four regions of the country, with 18% originating from the Midwest, 21% in the West, 29% in the South and 32% in the East. About two-thirds, or 68%, of the responses were from doctorate granting universities, 19% from Baccalaureate colleges, 6% from master's colleges and universities, and 6% from medical or special schools. Dspace is the most commonly employed IR software (42%), with Digital Commons coming second (35%), and the remaining 22% switching software, using a homegrown platform, IR + or Fedora/Islandora. On average, the IRs have been operational for about six years, although the entire range spans from two years to ten (Fig. 1).

4.2. WHO HAD INPUT?

As would be expected, library administrators had the most input into the creation of the SA (21 or 70%, Fig. 2). This was closely matched by IR managers/architects (19 or 63%), followed by legal counsel (15 or 50%), and other library personnel (11 or 37%) (Fig. 2). Interestingly, only 6 scholarly communications librarians had input (20%, Fig. 2). In 10 of the SAs, the legal department had no input into the creation of the SAs. In another eight, the legal department only had a little input. The remaining SAs are split between moderate legal department input (5), no legal department (2), and significant legal or entirely legal department input (4).

4.3. SELF-SUBMISSION AND FORMAT

With regards to the SA format, 18 are click-through (56%), with another 7 being both click-through and paper (22%). Four have no SA and three have a paper agreement. Nearly two thirds, or 22 of the responses indicated that less than 25% of the material in the IR was self-submitted (Fig. 3). Another 6 reported no material that was self-submitted (17%, Fig. 3). Only four have self-submissions as more than 50% of their

Fig. 2. What entities had input into the submission agreement(s)? (n = 30).

material (11%, Fig. 3). Therefore, most IRs use mediated submission processes for much of their material.

Of the six repositories that did not report self-submission, two still had a SA (Fig. 3). In contrast, there were two repositories that reported self-submission, but no SA (Fig. 3). All of these repositories contained both journal articles and book chapters, indicating probable commercially published content, and thus a probable need for the legal protection of a SA (data not published). However, for the two repositories without self-submission, but with submission agreements, it is unclear if the SA was created and then not used, or if the SA operates in a mediated submission process.

4.4. IR CONTENT

Most of the IRs (over 80%) contain journal articles, theses/dissertations, unpublished reports/working papers, multimedia/audio-visual materials, conference/working papers, and books, chapters and sections (Fig. 4). About two-thirds collect undergraduate work and other special item types (Fig. 4). It was less common to collect teaching materials (34%), datasets (29%) and software/computer code (11%) (Fig. 4). Only a few IRs were dominated by one type of material: namely, four IRs consist primarily of theses/dissertations, three consist primarily of journal articles, two primarily consist of books, chapters and sections and one is more than 76% unpublished reports/working papers
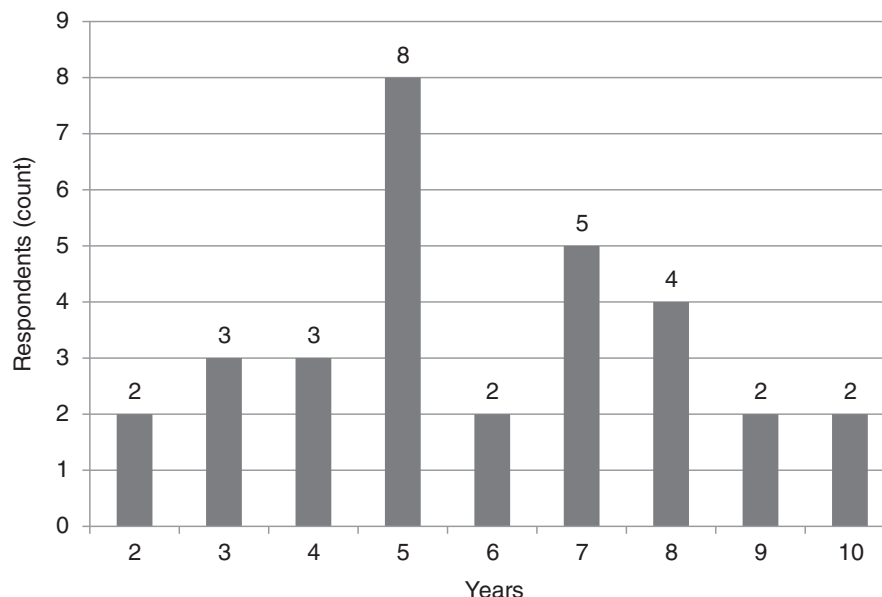
Fig. 1. Approximately how many years has your institutional repository been operational? (n = 31).
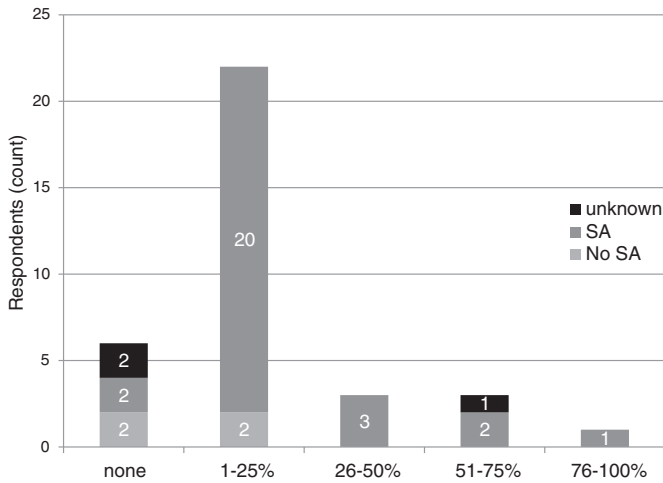
Fig. 3. The approximate percentage of material in the IR that was self-submitted by presence or absence of a SA (n = 35).

(Fig. 4). Overall, about 90% of IRs had less than 25% of any one type of materials, indicating that most IRs have diverse collections (Fig. 4).

4.5. SA TEXT

In general, the 20 SAs included in text analysis ranged from 101 to 596 in word counts, with one lengthy outlier at 1190 words. The average was 282 words arranged in 9 sentences (without the outlier). However, common terms and their collocates varied widely, as illustrated by collocate graphs of the shortest, longest, and total SAs (Fig. 5). Additionally, the Flesch Reading Ease averaged 27 and the Flesch-Kincaid Grade Level averaged 17, regardless of length of the SA. Therefore, SAs generally require a university degree to be understood (a score of 1–30) or approximately 17 years of education (Kincaid, Fishburne, Rogers, and Chissom, 1975).

Common concepts are in bold, followed by the frequency in which the concept appears in the 20 SAs and representative phrases (Tables 1–4). It should be noted that none of these concepts, or the phrases that represent them, are intended to be used in isolation.

4.6. POTENTIAL CHANGES

Only five respondents clearly indicated that they would change anything about their SA. Two clearly indicated a desire to simplify the agreement:

– "I wish it sounded less alarmist about copyright. We have had faculty refuse to submit because of issues with the language. But the language was drafted by Univ. legal counsel."
– "Consolidate it"

Another three indicated that their agreements were lacking:

– "Will include firmer embargo sign-off by ETD author and advisor."
– "I'd like to have an easy way to capture author(s) permission when library staff deposit on behalf of author(s)"
– "Would like option to separate it out from individual submissions so that agreement can still be electronic but so we don't have to require it for each submission."

5. DISCUSSION

5.1. SA CONTRIBUTORS

Surprisingly, scholarly communications librarians appear to rarely be involved in the creation of the IR SA. Considering that they are often tasked with interpreting and explaining the document, this seems a strange disconnect. However, it is possible that the SA was created prior to the hiring of a Scholarly Communications librarian. Other explanations may be the lack of a scholarly communications librarian entirely or the use of another title by the person that fills this role. In these cases, it is quite possible that the person responsible for the interpreting the SA for end-users was involved in its creation, but that this was not measurable by this survey. Only more detailed data collection from this community will confirm any disconnects between the crafting of the SA and the responsibility of interpretation and explanation.

Considering that a primary motivation behind the SA is legal protection, the lack of legal counsel's involvement in the creation of the majority of the agreements is also surprising. When these results are combined with the total lack of responses to the question regarding any changes to the SA, the length of time that some of the IRs have
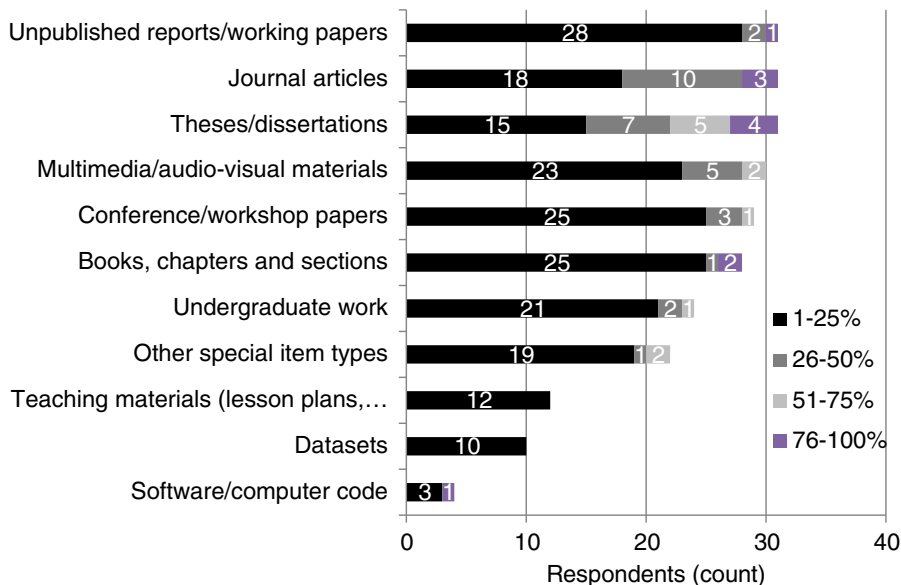

Fig. 4. Approximately how much and what type of material is currently in your institutional repository? (n = 35).
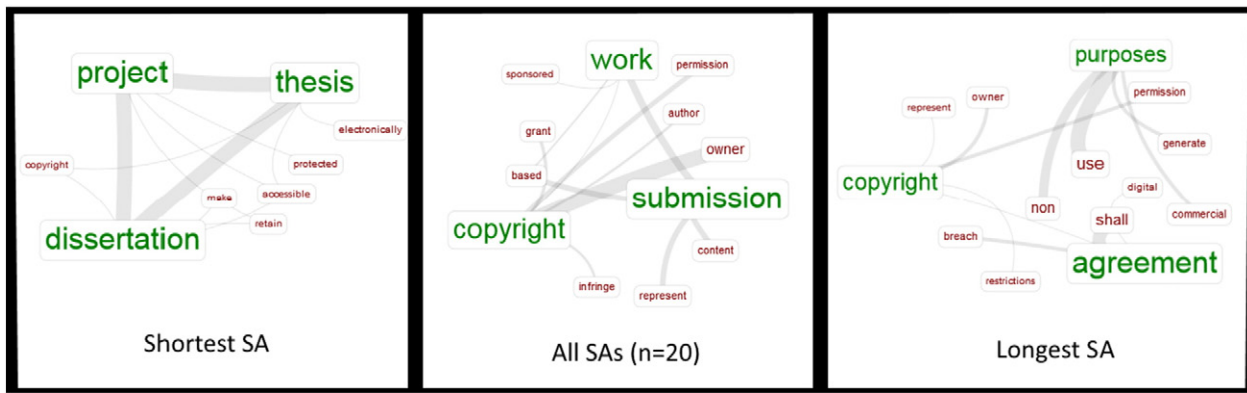
Fig. 5. Collocate graphs of the shortest SA, all SAs collected by this study (n = 20), and the longest SA.

been operational, and the lack of any known legal issues in the literature, this seems to indicate that the SAs are quite durable despite lack of legal counsel input. Anecdotally, many of the SAs appear to be patterned off of robust templates from either other institutions or IR software providers and these may have been initially crafted with legal counsel input. Therefore, it is possible that additional legal input is not needed. Indeed, if an SA has been created largely from a vetted, robust example, then a variety of personnel may not need to be involved in SA creation. Contrarily, any flaws in SA language, including those based on initial assumptions about the use of the IR, would be perpetuated. Therefore, caution should be used in re-using SA language. Unfortunately, it is not possible to determine from this brief survey if an individual SA has been copied or modified from another institution's SA. However, legal issues are not necessarily readily volunteered and we must not forget that while we have an absence of evidence, this cannot be construed as evidence of absence. It is possible that there have been significant legal issues surrounding SA language, but they have been handled in such as manner as to not have made it into the literature or be common knowledge. Only future research on the topic can provide conclusive evidence.

5.2. LACK OF SELF-SUBMISSION

Originally, IRs were intended to accept content from the original authors, requiring them to both know, and understand, the copyright implications of posting their work freely on-line. However, 28 of the respondents, or 79%, indicated that less than 25%, or none, of their material was self-submitted. As one respondent commented:

"At present, we search major 3rd party databases to identify articles written by authors. We contact authors for permission to deposit, but the actual submission (and clicking through the license agreement) is done by library staff. We retain the emails from faculty authors giving us permission to deposit."These results are consistent with other findings of low self-submission (Royster, 2007; Watson, 2007; Salo, 2008). Indeed, it's been suggested that self-submission may be discouraged due to administrator reluctance to relinquish control (Dubinsky, 2014). Considering our results, and the recent finding that self-submission may increase IR costs (Burns, Lana, and Budd, 2013), we anticipate that mediated submission will continue to outpace self-submission.

Table 1
Phrases to protect the institution from unlawful activities.

Concept Phrase	Usage rate (n = 20)
Copyright or third party permission, if no copyright to the material	90%
"I warrant that I have the copyright or other intellectual property right or permission to grant these rights" or "I have obtained and uploaded… written permissions statement(s) from the owners of each third party copyrighted matter"	
The submitter has rights to enter into the agreement	75%
"You represent and warrant that (1) you have all of the rights necessary to grant the license" or "You represent that … you have the right to grant the rights contained in this license."	
The work is original	55%
"You represent that the submission is your original work"	
No infringement on anyone else's copy or intellectual property rights	55%
"You also represent that your submission does not, to the best of your knowledge, infringe upon anyone's copyright." or "your submission does not, to the best of your knowledge, infringe upon any third party's copyright or other intellectual property rights;"	
All third-party work is identified and acknowledged	55%
"If the submission contains material for which you do not hold copyright, or for which the intended use is not permitted, or which does not reasonably fall under the guidelines of fair use, you represent that you have obtained the unrestricted permission of the copyright owner to grant [*the IR*] the rights required by this license, and that such third-party owned material is clearly identified and acknowledged within the text and content of the submission." or "If the submission contains material for which you do not hold copyright and that exceeds fair use, you represent that you have obtained the unrestricted permission of the copyright owner to grant [*the university*] the rights required by this license, and that you have identified and acknowledged such third-party owned material clearly within the content of your submission."	
No infringement on others rights, in general	25%
"I agree to hold the Institution, Department, [*repository name*] and their agents harmless for any liability arising from any breach of the above warranties or any claim of intellectual property infringement arising from the exercise of these non-exclusive granted rights." or "The work is original and does not infringe upon the rights of others,"	
No infringement on others rights, or contain any unlawful material, specifically named as libelous, privacy, and/or confidentiality	20%
"The work is original and does not infringe upon the rights of others, does not contain any libelous content, and does not invade any privacy or confidentiality of third parties." or "You represent and warrant that …the submission contains no libelous or other unlawful matter and will not violate the right of privacy or constitute an invasion of any other rights of any person or third party."	
No infringement on others rights, specifically named as patent or trade secret	10%
"The Work does not infringe any copyright, patent, or trade secrets of any third party."	

Table 2
Phrases that delineate the rights of the submitter.

Concept Phrase	Usage rate (n = 20)
Non-exclusivity, making no exclusive claim on the use of the material	90%
"I hereby grant to [*repository name here*] and its agents the non-exclusive license to…"	
The submitter or author retains all current ownership rights, including copyright	35%
"You will retain your existing rights to your work, and may submit the work to publishers or other repositories without permission from [*the institution*]." or "You retain all of your ownership rights in all materials submitted to [*the IR*]."	
Leave the copyright with the submitter or author, specific to publication elsewhere and/or copyright transfer	20%
"I retain all ownership rights to copyright of the thesis, dissertation or project. I also retain the right to use in future works (such as articles or books) all or part of this thesis, dissertation, or project." or "I understand that the Work's copyright owner(s) will continue to own copyright outside these non-exclusive granted rights."	
Option to choose an embargo and/or expiration date(s)	10%
"I understand that accepted papers may be posted immediately as submitted, unless the submitting author requests otherwise or submits a revision." or "I do not want my project(s) available to the larger scholarly community upon deposit in [*repository name here*]. I want my project(s) available only to the [*repository name here*] community, permanently or until (specify date) when it may be made available to the larger scholarly community."	

As noted elsewhere in the literature, this lack of self-submission may result in very few authors actually seeing the SA (Hanlon and Ramirez, 2011; Markland and Brophy, 2005). Indeed, one respondent noted that "So far, only some students and one faculty member are doing their own submissions. We do everything else". Therefore, it would appear that the legality of the SA may be undermined by the practice of mediated submissions. As no SA analyzed in this study mentioned mediated submission, it is unclear how this practice is being accommodated. Due to this problematic issue, the authors did not attempt to craft a simplified SA from existing phrases, as this may inaccurately represent an ideal SA.

5.3. UNIQUE WORDING

Emerging trends include characterizing the SA as a non-exclusive license, specifying the submitter's responsibility for obtaining rights or permissions for any content for which they did not produce themselves or rights that were previously assigned to another entity, and confirming the right of the submitter to enter into the agreement. However, some SAs accommodate a few unique circumstances. For instance, two repositories submitted SAs that are specific to student work and mentioned FERPA.

As well, one respondent indicated that they "would like [the] option to separate [the SA] out from individual submissions so that [an] agreement can still be electronic but so we don't have to require it for each submission". Another SA addressed this very issue:

"In the event that You may, from time to time and subject to acceptance by [*the university*], deposit additional papers and other materials, this Agreement shall be applicable to the additional deposit. A description of the additional papers and materials shall be provided by You and attached as an exhibit to this Agreement."

This is the only instance we received of one SA being used to cover multiple submissions. Interestingly, this SA also addressed the assignment of responsibility after death. "[*The university and IR*] may assume custodial responsibility for previously accepted [*works*] orphaned by the death or dissolution of the Submitter and not formally assigned to the custody of another agency." These clauses address often overlooked issues that may merit future exploration.

6. CONCLUSION

Crafting a SA is often a balance between educating the submitter and protecting the institution. While it may be tempting to attempt comprehensiveness, it may not be advisable, or possible, to craft an all-inclusive SA. Even a lengthy and complex SA may not accommodate the changing scholarly communications landscape and unnecessary complexity, as noted by one respondent, deters submissions.

In recent years, IRs have been under increasing pressure to manage a diverse array of material in repositories, including datasets and student work (Newton, Miller, and Bracke, 2011). Dataset sharing has been spurred by the OSTP Public Access Memo of 2013, which called for "each Federal agency with over $100 million in annual conduct of research and development expenditures to develop a plan to support increased public access to the results of research funded by the Federal Government" (Holden, 2013). These results are specified as both publications and datasets (Rinehart, 2015). Datasets are particularly confusing as copyright may, or may not, apply (Enimil and Rinehart, 2014). Additionally, other legal responsibilities, such as export control laws, human subjects' privacy, and patent claims may apply. Outside of federal pressure, students have readily adopted IRs (Passehl-Stoddart and Monge, 2014; Rozum, Thoms, Bates, and Barandiaran, 2015). However, the standard SA does not accommodate FERPA, which is necessary to properly ingest student work. When considering these new demands, it readily becomes apparent that attempting to encompass all potential circumstances in a single SA is improbable.

All of the SAs were dense documents. This is appropriate if the reading audience is graduate students or faculty, but undergraduates may find the documents too difficult to understand. For instance, many SAs had language that accounted for unlawful content; libelous, invading privacy, violating confidentiality, etc. Some specified particular circumstances such as human subjects or interviews. Various user groups,

Table 3
Phrases that imply specific promises to the submitter.

Concept Phrase	Usage rate (n = 20)
Digital preservation of the material	90%
"to migrate the item (i.e., copy it) for preservation purposes."	
Digital preservation of the material, detailing back-up and security	50%
"You agree that [*the IR*] may, without changing the content, translate the submission to any medium or format, as well as keep more than one copy, for the purposes of security, backup and preservation"	
The IR will not alter or change the work outside of the agreement	55%
"[*The institution*] may not alter the content of the work." or "[*The institution*] will clearly identify your name(s) as the author(s) or owner(s) of the submission, and will not make any alteration, other than as allowed by this license, to your submission."	

Table 4
Phrases that remind the submitter of specialized circumstances or other options.

Concept Phrase	Usage rate (n = 20)
Reference to sponsored research (or other contractual) obligations	45%
"IF THE SUBMISSION IS BASED UPON WORK THAT HAS BEEN SPONSORED OR SUPPORTED BY AN AGENCY OR ORGANIZATION OTHER THAN [*the institution*], YOU REPRESENT THAT YOU HAVE FULFILLED ANY RIGHT OF REVIEW OR OTHER OBLIGATIONS REQUIRED BY SUCH CONTRACT OR AGREEMENT"	
State law obligations	20%
"This Agreement shall be governed by and interpreted in accordance with the laws of the State of California." or "This agreement will be interpreted and governed in accordance with applicable Federal law and the laws of the State of Maryland without reference to its conflict of laws rules."	
Additional requirements for IRB or human subjects research	10%
"If your work includes interviews, you must include a statement that you have permission from the interviewees to make their interviews public." or "No Social Security numbers are included in the work and all use of human subjects in support of the work has been cleared by the …University Institutional Review Board."	
Printing of the SA	10%
THE ABOVE CONSTITUTES A LEGAL AGREEMENT BETWEEN YOU AND [*the university*], SO PLEASE PRINT OUT A COPY OF THIS AGREEMENT FOR YOUR RECORDS.	
Creative Commons licenses	5%
"If you wish to attach a Creative Commons license to your work, please check which one"	
Student rights	N/A
"NOTICE OF SUBMITTED WORK AS POTENTIALLY CONSTITUTING AN EDUCATIONAL RECORD UNDER FERPA: Under FERPA (20 U.S.C. § 1232 g), this work may constitute an educational record. By signing below, you acknowledge this fact and expressly consent to the use of this work according to the terms of this agreement. I further understand that I am waiving some of my rights to anonymity under the Family Education Rights and Privacy Act of 1974 (FERPA) and am granting [*the university*] the right to display my name and year of birth in connection with any contribution of which I am an author or co-author."	

including younger students and international researchers, may not be overly familiar with these concepts. Unfortunately, when submitters are faced with a complex and confusing SA, they may ignore it, decline to participate, or look to their peers for possibly misleading interpretations.

Complexity of the SA can easily result in confusion. Consider the basic purpose of SAs: copyright. SAs generally require that the submitter understand and know whether they have assigned their copyright to a publisher. This is typically stated by two different methods: the statement that the work is original and that any third-party works are accounted for properly. The distinction between original work and third-party work could be confusing should a submitter interpret the statements together: how can an (entirely) original work contain third-party work? Similarly, is a co-authors work included in the original or treated as third-party work? Co-authors may assume that as a shared copyright holder, every author must give permission in order to submit the work (personal experience, redacted for review). Indeed, some legal counsel representatives have recommended this practice, possibly increasing this confusion (personal experience, redacted for review). Similarly, many submitters may be confused about when they must seek permission from a third party and when they can simply provide a citation. While these issues may be readily transparent to those who commonly discuss copyright implications, they may not be so to the average submitter. Similarly, reminding submitters of their other obligations in all capital letters can deter them from participation. While it may be seen as an added emphasis, it can be taken as yelling or shouting.

The ideal SA is a simple, readily understandable document that requires no added emphasis. Practices for simplifying the SA could include requiring the submitter to properly attribute or cite any work that is incorporated into the whole work, and if the copyright to the work has been assigned to another party, to request permission from that party. Additionally, instead of listing all possible violations, it may be more beneficial to use the blanket ban of anything 'unlawful'. This would protect the university without attempting to predict every future possibility. Standardization of language between institutions would also make SAs a great deal easier for mobile academics to understand.

However, a clear and simple SA may not cover the nuances that may be required for submitters to understand their rights and obligations. While this survey did not explore current educational resources that are associated with SAs, it is common for librarians to provide additional education on these topics. These activities may benefit from being embedded in ongoing education practices. For instance, particular 'unlawful' circumstances may be better explained in annual Responsible Conduct of Research training or in an FAQ. Ongoing education, along with public availability, may increase submitter understanding. As one

respondent noted: "We make the submission agreement available for people to see before they get into the submission process and we explain what it means". Because of this educational role, it is beneficial if the person who will be interpreting the SA for end-users, and providing accompanying education, be involved in its creation. More research within the communities of practice will reveal the most effective SA statements and accompanying education practices.

Another consideration for the SA document is the practice of mediated submission. Although the initial development of IRs anticipated self-submission of material, these results bolster the observation that self-submission is relatively uncommon. The practice of mediated submission means that the SA may not be serving the purpose to which it was intended. Perhaps a two-step workflow, with an initial agreement that the author signs to delegate the actual submission to another person, would resolve this dilemma.

From this brief exploration, we conclude that the while majority of SAs cover the same common basic concepts, efforts to be comprehensive result in complex, dense documents, with variability in language and detail. If IR managers, library administrators, and library legal departments exchanged SA language and experiences, a crowd-sourced effort may rapidly optimize SA language. This would particularly benefit those that are striving to launch a new IR, as it would actively reveal any re-purposing of SA language. Re-using language that has already been vetted may reduce or eliminate the need for a plethora of personnel, such as legal counsel, to contribute. Alternatively, re-purposing SA language may not consider the changing scholarly communications landscape or unique local needs. Regardless, sharing SA language would reveal the possibilities and pitfalls of current SA language.

ACKNOWLEDGMENTS

APPENDIX A

E-MAIL MESSAGE

"We'd really appreciate it if you could participate in a survey to document current practices in institutional repository submission agreements. If there are any questions, please contact Rinehart.64@osu.edu". The survey link is below:

URL: Redacted
Contact: Rinehart.64@osu.edu

INSTITUTIONAL REPOSITORY SUBMISSION AGREEMENTS SURVEY

The intent of this survey is to document current practices in institutional repository submission agreements. Participation in this survey is entirely voluntary. There are only thirteen questions and it should take less than twenty minutes. If there are any concerns regarding this survey please contact Rinehart.64@osu.edu.

E–mail message.

"We'd really appreciate it if you could participate in a survey to document current practices in institutional repository submission agreements. If there are any questions, please contact [Redacted for review]. The survey link is below:

URL: Redacted for review

Contact: Redacted for review

Institutional repository submission agreements survey.

The intent of this survey is to document current practices in institutional repository submission agreements. Participation in this survey is entirely voluntary. There are only thirteen questions and it should take less than twenty minutes. If there are any concerns regarding this survey please contact [redacted for review].

1. What approximate percentage of material in your institutional repository was self–submitted by depositors?

  ◉ None

  ○ 1–25%

  ○ 26–50%

  ○ 51–75%

  ○ 76–100%

2. If your depositors agree to a submission statement, please copy and paste it below:

3. Approximately how much and what type of material is currently in your institutional repository?

	0%	1–25%	26–50%	51–75%	76–100%
Journal articles	◉	○	○	○	○
Theses/dissertations	◉	○	○	○	○
Undergraduate work	◉	○	○	○	○
Unpublished reports/working papers	◉	○	○	○	○
Books, chapters and sections	◉	○	○	○	○
Conference/workshop papers	◉	○	○	○	○
Multimedia/audio–visual materials	◉	○	○	○	○
Other special item types	◉	○	○	○	○
Teaching materials (lesson plans, syllabi, etc.)	◉	○	○	○	○
Datasets	◉	○	○	○	○
Software/computer code	◉	○	○	○	○

Page 2. IR submission agreement creation and changes

4. Please select all that apply to your Institutional Repository's submission agreement(s):
  ☐ It is a click–through agreement
  ☐ It is a paper agreement
  ☐ There is no agreement required for submission
  ☑ Other, please specify

5. What entities had input into the submission agreement(s)?
  ☐ Library administrators
  ☐ University or college administrators
  ☐ Scholarly communications librarian
  ☐ Institutional repository manager/architect
  ☐ Legal council
  ☐ Other library personnel, not individually listed here
  ☐ Other campus–level entities, not individually listed here
  ☐ Other, please specify

REFERENCES

Andersen, R. (2015). CC BY and its discontents: An OA challenge. *Library Journal, 140*(5), 16.

Barton, M. R., & Waters, M. M. (2004). *Creating an institutional repository: LEADIRS workbook.* Massachusetts: MIT Libraries Retrieved from: http://dspace.mit.edu/bitstream/handle/1721.1/26698/Barton_2004_Creating.pdf?sequence=1

Barwick, J. (May 01, 2007). Building an institutional repository at Loughborough University: Some experiences. *Program: Electronic Library and Information Systems, 41*(2), 113–123.

Bastos, F., Vidotti, S., & Oddone, N. (2011). The university and its libraries: Reactions and resistance to scientific publishers. *Information Services & Use*, *31*(3/4), 121–129.

Bergstrom, T. C., Courant, P. N., McAfee, R. P., & Williams, M. A. (January 01, 2014). Evaluating big deal journal bundles. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(26), 9425–9430.

Blythe, E., & Chachra, V. (September 01, 2005). The value proposition in institutional repositories. *Educause Review*, *40*(5), 76–77.

British Library. (n.d.). EThOS toolkit: A guide to using and participating in EThOS. Deposit agreements. Retrieved from: http://ethostoolkit.cranfield.ac.uk/tiki-index.php?page=Deposit+Agreements

Budd, J. M. (2006). Faculty publishing productivity: Comparisons over time. *College & Research Libraries*, *67*(3), 230–239.

Burns, S. C., Lana, A., & Budd, J. M. (2013). Institutional repositories: Exploration of costs and value. *D-Lib Magazine*, *19*(1/2). http://dx.doi.org/10.1045/january2013-burns.

Center for Research Libraries (U.S.), & OCLC (2007). *Trustworthy repositories audit & certification (TRAC) criteria and checklist.* Chicago: Center for Research Libraries.

Consultative Committee for Space Data Systems (2002). *Reference model for an Open Archival Information System (OAIS).* Washington, D.C: CCSDS Secretariat.

Creative Commons. (n.d.). About the licenses. Retrieved from: http://creativecommons.org/licenses

Crow, R. (2002). The case for institutional repositories: A SPARC position paper. *ARL Bimonthly Report. 223.*

Dollar, D., King, L., Knight, P., & Leonard, P. (November 7, 2014). Data mining on vendor-digitized collections [presentation]. *2014 Charleston conference: Issues in book and serial acquisitions.* NC: Charleston.

Dubinsky, E. (2014). A current snapshot of institutional repositories: Growth rate, disciplinary content and faculty contributions. *Journal of Librarianship and Scholarly Communication*, *2*(3), eP1167. Retrieved from: http://dx.doi.org/10.7710/2162-3309.1167.

Duranti, L. (2010). The long-term preservation of the digital heritage: A case study of universities institutional repositories. *Italian Journal Of Library & Information Science*, *1*(1), 157–168.

Enimil, S., & Rinehart, A. (December 3, 2014). Can I copyright my data? [webinar]. *Presented for the Association for Library Collections and Technical Services, a division of the American Library Association.*

Gilman, I. (2013). *Library scholarly communication programs: Legal and ethical considerations.* Oxford: Chandos Pub.

Hanlon, A., & Ramirez, M. (2011). Asking for permission: A survey of copyright workflows for institutional repositories. *Libraries and the Academy*, *11*(2), 683–702.

Heath, F. (2009). Documenting the global conversation: Relevancy of libraries in a digital world. *Journal of Library Administration*, *49*(5), 519–532. http://dx.doi.org/10.1080/01930820903090896.

Holden, J. P. (February 22, 2013). Increasing access to the results of federally funded scientific research. Retrieved from: http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

JISC. (n.d.a) Repositories support project: Data re-use policies. Retrieved from: http://www.rsp.ac.uk/start/policies-and-legal-issues/re-use-policies/.

JISC. (n.d.b) Repositories support project: Submission policies. Retrieved from: http://www.rsp.ac.uk/start/policies-and-legal-issues/submission-policies/.

Jones, C. (2007). *Institutional repositories: Content and culture in an open access environment.* Oxford: Chandos.

Jones, R., Andrew, T., & MacColl, J. (2006). *The institutional repository.* Oxford: Chandos Publishing.

Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease formula) for navy enlisted personnel. *Research Branch Report 8–75.* Naval Air Station Memphis: Chief of Naval Technical Training.

Lynch, C. A. (2003). Institutional repositories: Essential infrastructure for scholarship in the digital age. *Portal: Libraries and the Academy*, *3*(2), 327–336.

Lynch, C. A., & Lippincott, J. K. (September 01, 2005). Institutional repository deployment in the United States as of early 2005. *D-lib Magazine*, *11*, 9.

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

Markland, M., & Brophy, P. (2005). *SHERPA project evaluation, final report.* Manchester: CERLIM (Centre for Research in Library & Information Management) Retrieved from: http://www.sherpa.ac.uk/documents/SHERPA_Evaluation.pdf

Newton, M. P., Miller, C. C., & Bracke, M. S. (January 01, 2011). Librarian roles in institutional repository data set collecting: Outcomes of a research library task force. *Collection Management*, *36*(1), 53–67. http://dx.doi.org/10.1080/01462679.2011.530546.

Odlyzko, A. (1997). The economics of electronic journals. *First Monday*, *2*(8). http://dx.doi.org/10.5210/fm.v2i8.542.

Passehl-Stoddart, E., & Monge, R. (2014). From freshman to graduate: Making the case for student-centric institutional repositories. *Journal of Librarianship and Scholarly Communication*, *2*(3), eP1130. Retrieved from: http://dx.doi.org/10.7710/2162-3309.1130.

Rawls, M. (2015). Looking for links: How faculty research productivity correlates with library investment and why electronic library materials matter most. *Evidence Based Library and Information Practice*, *10*(2), 34–44 Retrieved from: http://ejournals.library.ualberta.ca/index.php/EBLIP/article/view/24163

Rinehart, A. (April 27, 2015). Federal public access initiatives update [blog]. Retrieved from: https://library.osu.edu/researchcommons/2015/09/03/funding-for-data-science-research/

RLG/OCLC Working Group on Digital Archive Attributes, Research Libraries Group, & OCLC (2002.). *Trusted digital repositories: Attributes and responsibilities: An RLG-OCLC report.* Mountain View, CA: RLG.

Robertson, W. C., & Borchert, C. A. (2014). Preserving content from your institutional repository. *Serials Librarian*, *66*(1–4), 278–288.

Royster, P. (2007). The institutional repository at the University of Nebraska-Lincoln: Its first year of operations. *OCLC Systems and Services*, *23*(2), 183–189. http://dx.doi.org/10.1108/10650750710748487.

Rozum, B., Thoms, S., Bates, S., & Barandiaran, D. (2015). *We have only scratched the surface: The role of student research in institutional repositories in in creating sustainable community: The proceedings of the ACRL 2015 conference.* Portland, OR: Association of College and Research Libraries, 804–812 Retrieved from: http://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/2015/Rozum_Thoms_Bates_Barandiaran.pdf

Salo, D. (2008). Innkeeper at the roach motel. *Library Trends*, *57*(2), 98–123. http://dx.doi.org/10.1353/lib.0.0031.

Swanson, J., & Rinehart, A. (2016). Data in Context: Using Case Studies to Generate a Common Understanding of Data in Academic Libraries. *Journal of Academic Librarianship.*, *42*(1).

Tedd, L. (Ed.). (2006). *Institutional repositories.* Bradford, England: Emerald Group Publishing.

U.S. Department of Education, National Center for Education Statistics (2013). Chapter 2. *Digest of education statistics, 2012 (NCES 2014-015).*

U.S. Department of Education. (n.d.). Family Educational Rights and Privacy Act (FERPA). Retrieved from: http://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html

U.S. Department of Health and Human Services. (n.d.). Understanding health information privacy. Retrieved from: http://www.hhs.gov/ocr/privacy/hipaa/understanding/

U.S. Department of State. (n.d.). Overview of U.S. export control system. Retrieved from: http://www.state.gov/strategictrade/overview/

University of Nottingham (2014, April 24). OpenDOAR Email Distribution Service. Retrieved from: http://www.opendoar.org/tools/emailservice.html

University of Nottingham (January 8, 2015). OpenDOAR; The Directory of Open Access Repositories. Retrieved from: http://www.opendoar.org/index.html

Watson, S. (2007). Authors and university IRs. *Serials*, *30*(3), 225–230.