

# JLSC

ISSN 2162-3309 | JLSC is published by the Pacific University Libraries | <http://jls-public.org>

**Volume 6, General Issue (2018)**

## **The Types, Frequencies, and Findability of Disciplinary Grey Literature within Prominent Subject Databases and Academic Institutional Repositories**

Wanda R. Marsolek, Kristen Cooper, Shannon L. Farrell, & Julia A. Kelly

Marsolek, W.R. (2018). The Types, Frequencies, and Findability of Disciplinary Grey Literature within Prominent Subject Databases and Academic Institutional Repositories. *Journal of Librarianship and Scholarly Communication*, General Issue(6), eP2200. <https://doi.org/10.7710/2162-3309.2200>



© 2018 Marsolek, et al. This open access article is distributed under a Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>)

# The Types, Frequencies, and Findability of Disciplinary Grey Literature within Prominent Subject Databases and Academic Institutional Repositories

Wanda R. Marsolek

*Virtual Reference and Digital Media Equipment Coordinator, University of Minnesota*

Kristen Cooper

*Plant Sciences Librarian, University of Minnesota*

Shannon L. Farrell

*Natural Resources Librarian, University of Minnesota*

Julia A. Kelly

*Science Librarian: Applied Economics, Ecology, and Horticulture, University of Minnesota*

**INTRODUCTION** In many disciplines grey literature, or works that are more ephemeral in nature and are not typically published through traditional scholarly channels, are heavily used alongside traditional materials and sources. We were interested in the type and frequency of grey literature in subject databases and in North American institutional repositories (IRs) as well as what disciplines use grey literature. **METHODS** Over 100 subject databases utilized by academic researchers and the IRs of over 100 academic institutions were studied. Document type, search capabilities, and level of curation were noted. **RESULTS** Grey literature was present in the majority (68%) of the literature databases and almost all IRs (95%) contained grey literature. **DISCUSSION** Grey literature was present in the subject databases across all broad disciplines including arts and humanities. In these resources the most common types of grey literature were conference papers, technical reports, and theses and dissertations. The findability of the grey literature in IRs varied widely as did evidence of active collection development. **CONCLUSION** Recommendations include the development of consistent metadata standards for grey literature to enhance searching within individual resources as well as supporting future interoperability. An increased level of collection development of grey literature in institutional repositories would facilitate preservation and increase the findability and reach of grey literature.

Received: 04/22/2017 Accepted: 02/04/2017

Correspondence: Wanda R. Marsolek, 60 Magrath Library, University of Minnesota, Minneapolis, MN 55455, mars0215@umn.edu



© 2018 Marsolek, et al. This open access article is distributed under a Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>)

## IMPLICATIONS FOR PRACTICE

1. In searching the commercial databases, grey literature was more prevalent and was found in more disciplines than expected. This speaks to the need for selectors in all areas to consider the role of grey literature as they build and maintain collections.
2. In contrast to commercial databases, the discoverability of grey literature in institutional repositories varied greatly. This study suggests that richer and more consistent metadata could address this issue and increase discoverability.
3. With the growing interest in systematic reviews beyond the health sciences and guidelines strongly suggesting the inclusion of grey literature, discoverability issues have implications for how complete a review can be.

## INTRODUCTION

Many disciplines depend on grey literature as a medium for both obtaining and disseminating information in an expedient way. Grey literature is more prevalent and commonly utilized in some subject areas, such as physics and economics, and disciplines also vary in their norms about how grey literature is integrated into the research process as well as how its creation is recognized and rewarded in academia (Creaser et al., 2010; Kling & McKim, 2000; Li, Thelwall, & Kousha, 2015; Pinfield, 2002; Velden & Lagoze, 2009).

The authors of this paper, librarians at the University of Minnesota – Twin Cities, have been actively collecting large amounts of forestry as well as agricultural and applied economics grey literature for a number of years in both paper and digital formats. Staff dedicate time to locating and cataloging grey literature because these materials are valued and utilized by researchers. Due to the ephemeral nature of grey literature, we worry that materials may disappear or that researchers will not be able to find them on their own. This concern prompted us to undertake a study exploring the findability and accessibility of grey literature across numerous subject databases and academic institutional repositories (IRs).

There have been various definitions of grey literature over time. In their “Grey Literature” chapter of *Encyclopedia of Library and Information Science*, Schöpfel and Farace (2010) look to a combination of definitions from sources such as the 1997 Third International Conference on Grey Literature, which describes grey literature as “that which is produced on all levels of government, academics, business, and industry in print and electronic formats, but which is not controlled by commercial publishers” (p. 2029). The definition was revised at the Sixth International Conference on Grey Literature in 2004, to add a

postscript that reads “. . . not controlled by commercial publishers, i.e., where publishing is not the primary activity of the producing body” (Schöpfel & Farce, 2010, p.2029). In 2010 Schöpfel proposed a new definition at the Twelfth International Conference on Grey Literature, stating that

grey literature stands for manifold document types produced on all levels of government, academics, business and industry in print and electronic formats that are protected by intellectual property rights, of sufficient quality to be collected and preserved by library holdings or institutional repositories, but not controlled by commercial publishers i.e., where publishing is not the primary activity of the producing body. (Schöpfel 2011, Concl. para. 5)

Among the most common types of grey literature used by academic researchers are conference proceedings, technical reports, theses and dissertations, working papers, and government documents. Many forms of grey literature, such as government documents and working papers, have a history of being openly available, even long before the current open access movement (Ćirković, 2017; Rizor & Holley, 2014).

The growth in the number of systematic reviews being conducted and librarian involvement in the process has brought increased attention to grey literature (Bonato, 2016; Di Cesare & Sala, 1996; Mahood, Van Eerd, & Irvin, 2014). Systematic reviews use a research method that involves formulating a research question, conducting reproducible searches, and selecting and appraising the resulting studies in order to make evidence-based decisions. The main guidelines used for conducting systematic reviews specifically mention grey literature and the need to add steps to the search process to help locate it (Higgins & Green, 2011; The Methods Group of the Campbell Collaboration, 2016).

The authors of this study are interested in the role of grey literature in academic research across disciplines. Assessing the findability and accessibility of grey literature across numerous academic subject databases and IRs is the first step in evaluating and understanding this landscape. We hope to gain an idea of the prevalence of grey literature in information sources that are commonly used by our researchers as well as those to which we might direct them when they are looking specifically for grey literature. We also hope to gain a better idea of how much grey literature there is in IRs and what types are the most common. Another area of interest is whether grey literature finds its way into IRs through active collection activities of the administrators of the repositories or via random uploading by contributors to the repository.

## LITERATURE REVIEW

### Importance of Grey Literature and Discoverability Challenges

Many scholars have noted the importance of grey literature. Seymour (2010) argues that grey literature's nontraditional publication process includes many benefits: "speedy distribution, presentation of abundant amounts of data, in-depth analyses, consideration of a range of methodological and theoretical issues using sizable datasets, and avoidance of many of the stifling political hurdles and time delays of traditional publishing" (p.228). In their paper on grey literature in institutional repositories, La Fleur and Rupp (2004) draw attention to the importance of grey literature in the scientific process when they say that the "quest for scientific knowledge is an evolutionary process in which every increment of new knowledge adds to, modifies, refines, or refutes earlier findings" (2004, para. 1).

Even as finding tools have moved from print to electronic, it is still more difficult to locate grey literature than more traditional publication forms such as books and journal articles (Gelfand, 2006; Okoroma, 2011). Grey literature may be published on obscure websites that are not well indexed, or it may be archived in repositories without appropriate metadata to facilitate discovery. Various authors have noted the challenges associated with discovering grey literature. Chowdappa, Devi, Ramasesh, and Shyamala (2011) found that researchers were interested in grey literature and requested orientations on how to find it. According to Schöpfel, Le Bescond, and Prost (2012), findability may hinge on good metadata. Both Lambert, Matthews, and Jones (2006) and LaFleur and Rupp (2004) argue that librarians ought to edit the metadata that authors have provided through self-deposit to IRs to make it more complete and findable.

### Grey Literature and Systematic Reviews

The importance of grey literature has received new attention recently with the increased creation and use of systematic reviews. In the past, systematic reviews were mostly conducted in the health sciences, but more recently disciplines in the humanities, social sciences, and agricultural sciences have also produced them. Many guides to systematic reviews, such as the *Cochrane Handbook for Systematic Reviews of Interventions*, emphasize the importance of including grey literature (Higgins & Green, 2011). Bonato (2016) notes that "information from unpublished studies, and the failure to identify trials noted in conference proceedings and other sources of grey literature might affect the results of a systematic review" (p. 252). Mahood et al. (2014) hint that there is no method for how to conduct grey literature searches, but that nonetheless, including grey literature in systematic reviews generates a stronger review.

## Grey Literature and the Shift to Electronic Publication

The shift from print to electronic access has the potential to have a large impact on grey literature in the areas of discoverability, access, and preservation. With the exception of theses and dissertations and possibly government documents, print versions of grey literature have not been collected by many libraries in any organized fashion (Gelfand & Lin, 2013). Following the U.S. Government's lead, many organizations have stopped producing print versions of their grey literature (Lyons, 2006). Organizations of all sizes have the ability to produce born-digital versions of technical reports, working papers, and many other forms of grey literature. Scanning is relatively inexpensive and widely available, so print versions of documents can be digitized easily. Sophisticated search engines like Google make locating individual grey literature items possible. Preservation of electronic copies of grey literature, however, remains a challenge (Lambert et al., 2006). While IRs and other platforms provide safe storage, persistent URLs, backup, and possibly migration if it is needed in the future, many groups simply keep their documents on a local server and link to them from a web page. Examples include the [Brookings Institute](#), the [Institute for Agricultural and Food Policy](#), and [Friends of the Earth](#).

## Grey Literature and Repositories

In seeking appropriate locations to make grey literature more discoverable and accessible, IRs have arisen as potential solutions, partly due to their stability (Gelfand, 2005). IRs commonly support open access archiving, which facilitates access by researchers without access to subscription resources. In a survey of French repositories, grey literature represented 18% of all documents (Schöpfel et al., 2012). Melero Abadal, Abad, and Rodríguez-Gairín (2009) found that 23% of documents in Spanish repositories were full-text grey literature. While IRs offer great potential for the storage of grey literature, they present certain challenges to users. Author self-deposit, where the author or author's staff upload documents and accompanying metadata, is a common strategy for populating IRs, which can lead to spotty inclusion of materials (Creaser et al., 2010). This can be a result of the varying deposit procedures and levels of mediation at different institutions. Individual authors may not be well versed in copyright issues concerning authorship of conference papers and government documents and may pass up opportunities to make deposits (Schöpfel et al., 2012). Conference organizers may not make arrangements to archive the materials themselves, and also may not be clear with authors about the copyright status of their papers, further inhibiting the use of IRs (Linde et al., 2011). Also, authors often add their own metadata, which can lead to inconsistency across deposits, and in turn, limited findability (Colati, Dean, & Maull, 2009; Costanza, Knight, & Lui-Spencer, 2009).

There are some subject or disciplinary repositories that specialize in indexing and archiving grey literature. Examples include [arXiv](#) (physics, math, and computer science), [PhilSci Archive](#) (philosophy of science), and [AgEcon Search](#) (agricultural and applied economics). Some but not all subject repositories include grey literature, and they may or may not enhance records with controlled subject terms.

### **Grey Literature and Commercial Indexes**

Unlike subject repositories, commercially produced indexes seldom focus on grey literature, although they may include various kinds of it. One factor that influences inclusion is the subject matter and whether a discipline produces, uses, and values grey literature. For example, Medline focuses almost exclusively on the journal literature, since that is what researchers in the health sciences mainly cite, while Compendex, which covers engineering topics, indexes over 20,000 conferences and technical reports series.

Commercial indexes often have robust search features that may ease the task of locating grey literature. Bonato (2016) studied Google Scholar and Scopus to assess their potential use in locating grey literature and found Scopus to have a superior search platform although neither was ideal for the task.

### **METHODS**

This study examines:

1. The types and frequency of grey literature within prominent subject databases and academic institutional repositories (IRs)
2. Which disciplines generate grey literature
3. How subject databases and IRs allow users to search for and find grey literature
4. Whether IRs actively collect grey literature

### **Database and institutional repository selection**

Databases were chosen from lists generated by University of Minnesota – Twin Cities liaison librarians in their associated database subject guides. Databases were restricted to those to which the University of Minnesota subscribes. Databases were excluded if they were known as predominantly grey literature databases (e.g., arXiv, Digital Dissertations), as the goal was to identify databases whose focus is not primarily grey literature. Compilations of books or journals from a single publisher (e.g., Springer eBooks) were also excluded, as the focus of this work was on indexes covering a broad range of material. This resulted in

a list of 173 databases. The authors made a decision not to include discovery tools (e.g., OneSearch, Primo) because we wanted information about individual databases. Discovery tools would not help us answer the question about whether particular databases included grey literature and if so, how much.

IRs chosen for inclusion are academic libraries (i.e., college, university, undergraduate, or professional libraries) that are also members of the Association of Research Libraries (ARL). We removed IRs from the list if they were affiliated with public libraries or organizations that are not colleges or universities (e.g., New York Public Library or National Agricultural Library), as we wanted to limit our study to peer institutions to provide a fair comparison. This resulted in a list of 115 IRs.

### **Types of grey literature**

We defined grey literature as works that are ephemeral in nature and published in nontraditional ways. We identified several types of grey literature by conducting a preliminary search and selecting those likely to be used by researchers, which included conference papers and posters, government documents, technical reports, theses and dissertations, and working papers (a complete list of types is provided in Table 1). We then specifically searched for these terms within the above databases and IRs.

### **Data collection and evaluation**

Evaluation of databases and IRs was split evenly among the authors. To maintain consistency in data gathering, we started with a small sample from the total to discuss differences, strengthen criteria, and establish guidelines for entering data; this was done to establish interrater reliability. Once agreement was reached, the remaining databases and IRs were divided among the authors to evaluate using agreed-upon ratings. Each author focused on a particular IR software (e.g., DSpace, Islandora) to highlight differences in implementation. We collected the following data about each database:

- Does it contain grey literature? (Y/N)
- Does it contain limiters to help the user find grey literature? (Y/N)
- What types of grey literature are included? (list them)
- Does the database publisher mention the inclusion of grey literature within the database? (Y/N)
- What is the primary discipline of focus of the database? Is it multidisciplinary?



To search for the presence of grey literature within a database, we used the advanced search feature and either a) looked for a limiter or filter to refine our search for the identified types of grey literature in the fields “document type” or “publication type”; or b) searched for representative types of grey literature using keywords such as “report,” “dissertation,” “conference,” or “working paper.” We actively searched for all types listed in Table 1. If we discovered other types of grey literature within the databases and IRs, that data was captured, but we did not specifically look for these types of grey literature going forward. After exhausting the list of grey literature types mentioned above and finding nothing, a database was counted as not containing grey literature. We also examined publishers’ websites to determine the disciplinary focus of the database’s content and to see if they mention grey literature as being included in the database. Working from [AAUW’s List of Academic Fields \(AAUW 2014\)](#), we created a list of eight broad disciplines (see Appendix). If databases appeared to cross more than one of these broad disciplines, they were classified as “multidisciplinary.”

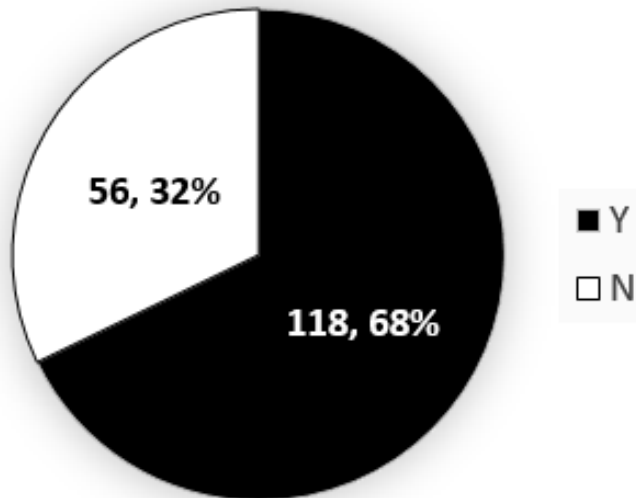
We collected the following data in each institutional repository:

- Does it contain grey literature? (Y/N)
- What types of grey literature does it contain? (list them)
- Can an author self-deposit documents? (Y/N)
- Does it include any grey literature series? (Y/N)

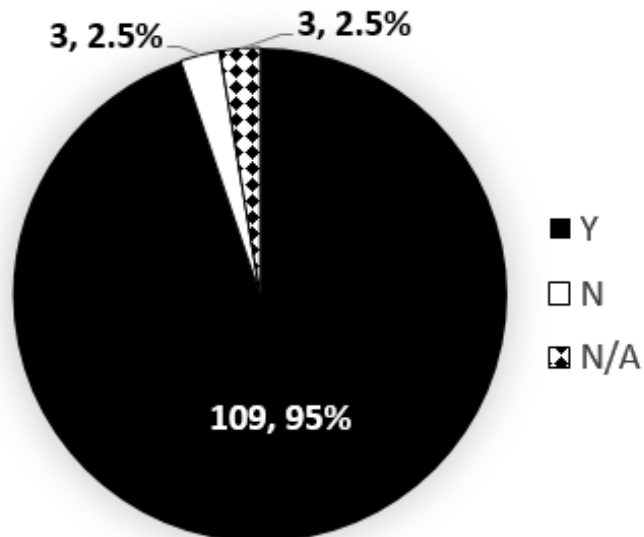
The search process to discover grey literature in IRs was similar to that of the databases. To determine if grey literature was present, we used the advanced search feature to search for representative types of grey literature in the “document type” or “publication type” field using keywords such as “report,” “dissertation,” “conference,” or “working paper.” Due to the variability in IRs, the search feature did not work to locate grey literature in some cases; when this occurred, we browsed collections to see if they included grey literature. As some disciplines are known to contain higher frequencies of grey literature (e.g., engineering), we started browsing there. Once any grey literature was found, we noted that the IR held grey literature and listed the types we had found. We also examined each repository’s policy about how documents are deposited: Can an author self-deposit? Must an author be invited to participate, or does a librarian have to deposit the work? To discover if IRs were actively developing collections, we examined whether an IR collected multiple volumes or years of grey literature such as “conference proceedings,” “newsletters,” or “technical reports.” If any of these criteria were met, an IR was recorded as actively developing collections. To be counted as actively developing collections, these volumes did not need to be complete as long as they showed evidence of attempted development.

## RESULTS

Of the 173 databases, 118 (68%) contained some form of grey literature (see Fig. 1). Of the 115 institutional repositories, 109 (95%) contained grey literature (see Fig. 2). Three IRs were listed as not applicable (N/A) because although the IR had an operating website, there was no content in the IR.



**Figure 1.** Number and percent of subject databases that included grey literature



**Figure 2.** Number and percent of institutional repositories that included grey literature

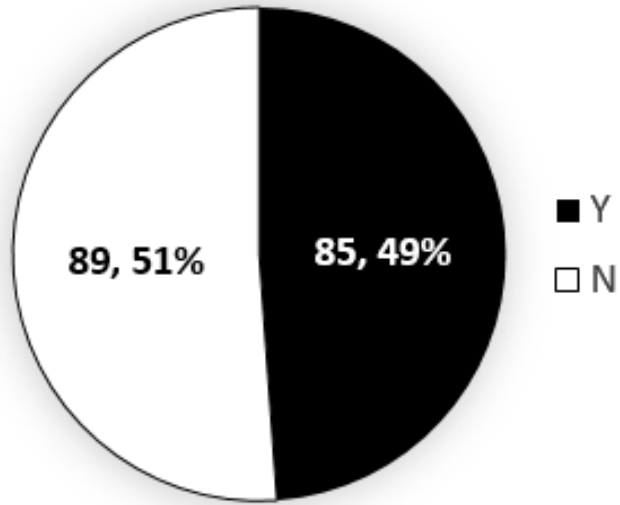
Subject databases and IRs varied in the types of grey literature that they contained. Table 1 illustrates the percentage of types of grey literature present in databases and IRs. The grey literature types most common in databases were as follows: conference proceedings, papers, reviews, and posters (41%), technical reports (33%), and theses and dissertations (28%). The most common types of grey literature in IRs were as follows: theses and dissertations (91%), conference proceedings, papers, reviews and posters (65%), technical reports (58%), and working papers (50%). Theses, dissertations, and working papers were far more common in IRs than in databases, a finding that reflects academic institutional policies to retain those items. We discovered more types of grey literature within the databases that were not present in IRs, including blogs, legal cases, and patents, as well as specifications, standards, and protocols.

Types	Subject Databases	Institutional Repositories
Bibliographies	16%	29%
Blogs and newsletters	2%	0%
Conference proceedings, papers, reviews and posters	41%	65%
Fact sheets, brochures and pamphlets	10%	32%
Government documents	10%	12%
Legal cases	1%	0%
Manuals and guides	1%	0%
Patents	1%	0%
Specifications, standards and protocols	6%	0%
Technical reports	33%	58%
Theses and dissertations	28%	91%
Trade literature	2%	0%
Working papers	9%	50%

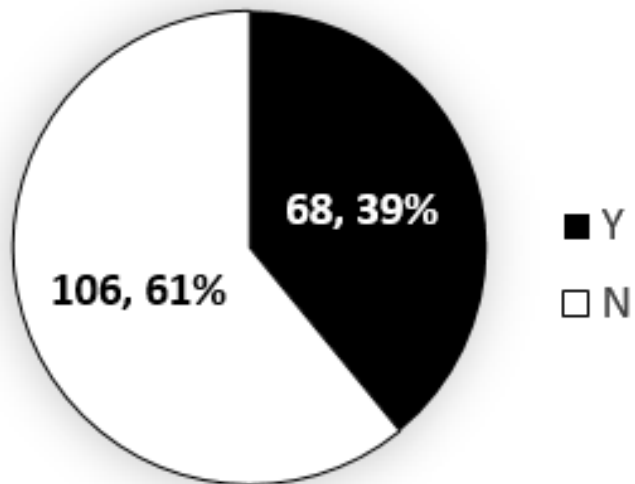
**Table 1.** Percentage of types of grey literature in subject databases and institutional repositories

Subject databases were fairly evenly split on whether or not they contained limiters for grey literature. Of the 173 databases, 85 (49%) had limiters while 89 (51%) did not (see Fig. 3).

Most publisher websites for the subject databases did not mention grey literature types in their descriptions of the content included in their databases. Even though 68% of the databases included some form of grey literature, only 39% of the associated websites referred to grey literature types of content (see Fig. 4).

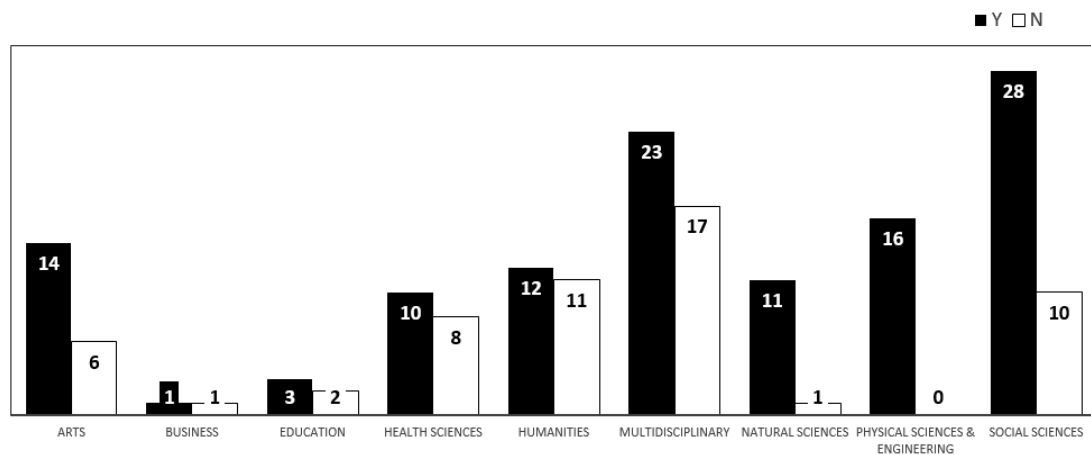


**Figure 3.** Limiters for grey literature in subject databases



**Figure 4.** Websites of database publishers that mention grey literature

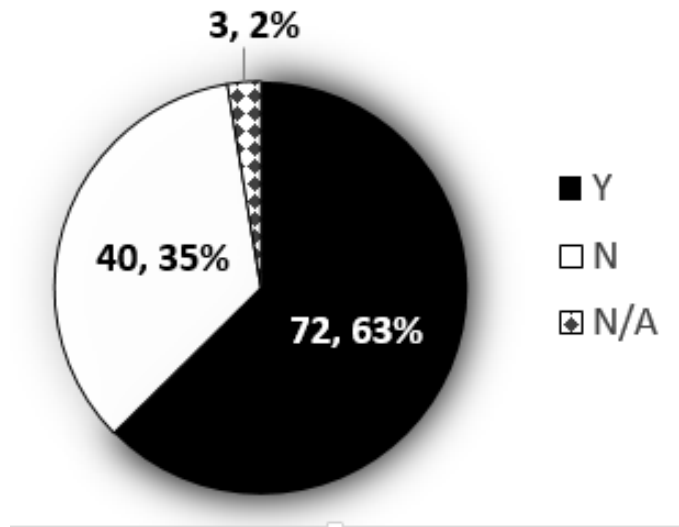
In all disciplines, databases with grey literature outnumbered those without (see Fig. 5). After classifying databases into subject categories, we found that the largest number of databases were multidisciplinary (40 total). Of these, 58% contained grey literature. When examining the databases by discipline, some notable patterns emerged. All of the physical sciences and engineering databases contained grey literature, as did the vast majority of the natural sciences databases (92%). Both social sciences databases and arts databases had high percentages of grey literature, 74% and 64% respectively. The health sciences and humanities databases were closer to 50%, at 56% and 52% respectively. Our sample of business (2) and education (5) databases was not large enough to illustrate a clear pattern.



**Figure 5.** Grey literature in subject databases by discipline

The vast majority of IRs (87%) allow users to self-deposit, although not all self-deposits are the same. Some IRs operate via a mediated self-deposit in which users deposit content, but the content is either reviewed by library staff before it appears in the IR or requires prior approval to deposit content. Of those that did not allow self-deposit, such as the University of Delaware’s UDSpace (<http://udspace.udel.edu>) and Duke’s Digital Repository (<https://repository.duke.edu>), users are required to fill out a form to request inclusion or must be invited by IR staff to deposit in a collection.

Most IRs are involved in actively collecting grey literature (see Fig. 6); 63% appear to collect series of conference proceedings, technical reports, or working papers. IRs may not have collected every paper, but appeared to have made a good-faith attempt to include the majority of papers in a series.



**Figure 6.** Institutional repositories that are actively collecting grey literature

## DISCUSSION

There were several limitations to this project. The first is that we looked at one institution’s list of databases, which makes generalizing our results to other institutions challenging. The second limitation is that when searching the IRs we did not contact or consult anyone involved in the repositories, but simply searched their public interfaces. Due to this, we have no detailed knowledge of the policies or personnel of the different repositories or the reasons behind those policies. The final limitation is the focus on only IRs found in North America. This leaves us with a knowledge gap on the state of grey literature in international repositories and prevents us from extrapolating our results beyond North America.

When we began the project we expected to find grey literature in both databases and IRs, with repositories having a higher concentration. We also expected to be able to search by publication type within both databases and repositories. While we did find grey literature in both locations, we were surprised by how much grey literature was contained in the databases. Additionally, while we expected some disciplines and their corresponding databases to have more grey literature than others, we found a large number of databases across disciplines that contained grey literature (see Figs. 1 and 5). We also noted that only half of the databases allowed users to limit or search by publication type, which has implications for the extent of search refinement that is possible. This comes into play for those conducting literature searches for systematic reviews. Guidelines note the importance of grey literature in a comprehensive search and suggest that authors seek out sources that cover the grey lit-

erature particular to their topic, or, if that is not possible, search the literature itself by hand.

The differences in the terminology used to describe grey literature publication types in commercial databases (e.g., “meeting” vs. “conference,” “pamphlet” vs. “brochure”) hindered our search process and impacted findability. The databases we searched utilized numerous metadata schemes, and end users may have to carefully evaluate how to search for grey literature as a result. Since commercial databases are proprietary platforms, librarians may offer suggestions for improvements in searching but have no control over contents or policies. As expected, we found grey literature in the vast majority (95%) of the IRs that we searched (see Fig. 2). However, the amount of grey literature in each repository was not as extensive as we thought it would be, nor did it appear to be as actively collected as we expected. As IRs evolve and their emphasis shifts away from open access journal articles and possibly toward other assets, they might consider intentionally focusing energy on the recruitment of grey literature (Lynch 2017).

There also appeared to be a greater variation in the level of organization and collection development than we expected. We found that many of the repositories did not allow searching or browsing by publication or grey literature type, and there was an array of terminology used to describe grey literature across institutions. Similar to the terminology differences present in the databases, this hindered the ability to search for grey literature and has implications for its findability.

In looking at IRs, we found great variation in how different institutions implemented the various features of the same repository software, such as limiting by publication type. In some cases, the contents of the IRs had to be accessed via the library catalog, making searching dependent on the level of cataloging (e.g., series rather than item level). Further, if a repository had conference proceedings, there was often no way of knowing what papers were included without actually looking at the items. We had hoped for a higher degree of uniformity of description and evidence of active collection of grey literature since many IRs have a high degree of involvement from library staff.

There are many known issues with the structure, implementation, and support of IRs that may impact inclusion and findability of grey literature. The first is that the time required for dedicated repository work can be substantial; librarians can invest a significant amount of time acquiring or soliciting grey literature from various content creators throughout their institution, whether it is eventually self-deposited or uploaded by IR staff (Childress, 2003). For some institutions these time requirements can be a significant challenge. In the Liberal Arts Scholarly Repository (LASR), a shared repository for several small liberal arts institutions, staff sizes are small, and repository duties are in addition to their current responsibili-

ties (Costanza et al., 2009). Examples like this could explain why some repositories used several broad categories instead of allowing users to search by document type. This could also explain the variation we saw in how the same repository software was utilized in different ways by different institutions. Another factor at many institutions is technology support. Server space or the support of a digital assets management system may not be among the top priorities for some institutions (Costanza et al., 2009).

Another potential issue is the application of terminology within an IR's metadata. Unqualified Dublin Core, with only 15 basic elements, is a metadata schema commonly used with IRs (Rychlik, 2016). While unqualified Dublin Core may seem simple and straightforward for repository use at first glance, Schopf et al. (2012) found that bibliographic control of grey literature remains poor due to flawed or incomplete metadata and variability in how the elements are applied. An example is how the same elements can be found referring to different concepts within a repository (Park & Richard, 2011). One study looking at the "dc.publisher" field found that it contained either the name of the creators' institution, the department to which they belonged, or no information at all. Some institutions use different fields all together, such as "dc.description.sponsorship" or "dc.contributor," to describe institutional affiliation (Costanza et al., 2009). This makes it difficult not only for users who are trying to locate the materials, but also for librarians who have to create or edit the metadata records. A clear, unified set of recommendations to which librarians consistently adhere for the application of terminology to grey literature using the Dublin Core metadata schema for IRs would allow librarians to develop full records for the grey literature in their repository in the shortest amount of time. This would also help institutions that have automated much of the process by allowing them to give individuals who provide content to the repository the clearest and most user-friendly description of the information to increase findability.

We believe that the lack of inclusion of grey literature in traditional collection development policies and IR scope statements may contribute to the varying degree of findability of grey literature. As Gelfand (2006) observed, "most collection development policies only address resources for which payment has been made, where formal acquisitions or licensing practices are observed" (para. 1) We discovered that 95% of repositories in our study contained grey literature, but only 63% appeared to be making an effort to actively collect that grey literature. Since we defined *active collection* as including multiple volumes of one or more document in the same title or series, our active collection percentage may actually reflect a very low level of collection. At our institution, there is no overarching policy about grey literature, print or electronic, leaving acquisition decisions in the hands of individual subject librarians. Situations such as these leave subject librarians with little guidance about grey literature. Along with this, self-archiving is most often the primary way of aggregating digi-



tal collections for an institution's repository (Xia, 2008). Having authors self-archive their own work can result in uneven amounts and coverage of grey literature across subject areas. Some faculty may practice self-archiving on a regular basis, while others may not practice it at all. Additionally, many faculty are unlikely to be metadata experts and may enter incorrect or incomplete information, resulting in poor metadata records that hinder findability. Finally, this issue can be exacerbated by potentially confusing ingestion guidelines.

The heightened interest in grey literature that is being fueled in part by the rise in systematic reviews makes addressing the issues of description and findability all the more important. In conducting systematic reviews, the ideal situation for researchers looking for grey literature on particular topics would be to have the ability to easily narrow by a particular publication type. They are often looking for particular conferences or items from a certain agency or association, or they may be anxious to be able to accurately state what grey literature was contained in the databases that they have searched.

Looking ahead to the future, the metadata standards and terminology used by IRs have ramifications for interoperability as well. As Riley (2017) points out, "everyone benefits when metadata can be transferred effectively; it reduces duplication of effort" (p. 39). The varying levels and types of description and use of terminology that we observed makes interoperability much more challenging. The level of organization needed for interoperability requires institutions to adhere to internal and extraorganizational standards and best practices such as recommended file types, metadata schemas, and controlled vocabularies (Moulaison & Dykas, 2016). Consistency in the application of metadata schema among IRs has many benefits to those searching for grey literature. It allows all searchers to find desired content more quickly and easily. For example, conference papers are a commonly searched-for type of grey literature, but metadata schemas do not make it obvious in which field to place important elements such as sponsoring organization or conference name. In some cases, this means this information may be placed in an inappropriate field or is left out completely. This inconsistency leads to issues with interoperability and makes it difficult for searchers to locate content. For those doing systematic reviews, this inconsistency results in an inability to identify particular content that they are seeking. The standard use of fields would allow better searching within the IR and potentially searching across numerous IRs. The ability to search across repositories instead of having to search each one individually will allow the resulting review to be as thorough as possible.

## CONCLUSION

Grey literature is a vital source of information for many disciplines. In searching subscription databases across disciplines, we found that they included more grey literature than we

expected. This means that researchers and students who are conducting literature searches on particular subjects will have an increased opportunity to discover grey literature while searching disciplinary or multidisciplinary databases.

The IRs that we investigated did not seem to systematically include most forms of grey literature, with the clear exception of theses and dissertations. There is an opportunity for repositories to increase the amount of materials they include and enhance findability and access to grey literature. The marriage between IRs and grey literature could elevate the value of IRs to the research community. IRs could make a substantial difference in ensuring grey literature's preservation, increasing its reach, and, in many cases, providing a form of legitimacy to these items published outside traditional realms. Therefore we make the following recommendations to increase grey literature's scope and findability across repositories. The first is to examine both current collection policies and IR scope statements to note where they might be expanded to include more grey literature. The next is to actively seek out grey literature that is produced on your campus or is important to the subject areas that are strengths at your institution, such as departmental newsletters, experiment station reports, policy briefs, or locally sponsored conferences. Another recommendation would be for IRs to adhere to existing metadata standards for use of fields and terminology important to the identification of grey literature, such as publication type. Including enhanced descriptions adequate to identify and locate grey literature would facilitate more successful searching. As interoperability between repositories grows, it will be beneficial for those specifically seeking grey literature to utilize common terminology and enhanced description.

## REFERENCES

- AAUW. (2014). List of Academic Fields. <http://aauw.scholarsapply.org/pdf/AAUW-AcademicFields.pdf>
- Bonato, S. (2016). Google Scholar and Scopus for finding grey literature publications. *Journal of the Medical Library Association*, 104(3), 252–254. <http://doi.org/10.3163/1536-5050.104.3.021>
- Childress, E. (2003). Going grey: Grey literature and metadata. *Journal of Internet Cataloging*, 6(3), 3–6. <http://doi.org/10.1300/J141v06n03>
- Chowdappa, N., Devi, L. U., Ramasesh, C. P., & Shyamala, V. R. (2011). Dependency on regional libraries for grey literature: Perceptions of researchers in engineering sciences and technology. In *Thirteenth International Conference on Grey Literature* (pp. 1–11). Retrieved from <http://eprints.uni-mysore.ac.in/14572/>

- Ćirković, S. (2017). Transition to open access and its implications on grey literature resources. *Grey Journal*, 13(2), 101–109. Retrieved from <http://web.ebscohost.com/ehost/detail/detail?vid=0&sid=357da624-84ff-4670-be31-92eff05ebae7%40sessionmgr120&cbdata=JnNpdGU9ZWlhvc3QtbGl2ZSZzY29wZT1zaXRl#AN=123453246&db=aph>
- Colati, J. B., Dean, R., & Maull, K. (2009). Describing digital objects: A tale of compromise. *Cataloging & Classification Quarterly*, 47(3–4), 326–369. <https://doi.org/10.1080/01639370902737331>
- Costanza, J., Knight, R. C., & Liu-Spencer, H. (2009). Metadata implementation for building cross-institutional repositories: Lessons learned from the Liberal Arts Scholarly Repository (LASR). *Journal of Library Metadata*, 9(1–2), 153–166. <https://doi.org/10.1080/19386380903176063>
- Creaser, C., Fry, J., Greenwood, H., Oppenheim, C., Proberts, S., Spezi, V., & White, S. (2010). Authors' awareness and attitudes toward open access repositories. *New Review of Academic Librarianship*, 16(S1), 145–161. <https://doi.org/10.1080/13614533.2010.518851>
- Di Cesare, R., & Sala, C. (1996). *The use of grey literature in the agricultural economics field: A quantitative analysis*. In Second International Conference on Grey Literature: Grey Exploitations in the 21st Century (pp. 157–168). Washington D.C.
- Elsevier. (2018). *Compendex Source List*. Retrieved from <https://www.elsevier.com/solutions/engineering-village/content/compendex>
- Gelfand, J. (2004). “Knock, Knock:” Are Institutional Repositories a Home for Grey Literature? *The Grey Journal*, 1(2), 61–66.
- Gelfand, J. (2006). Grey literature : Taxonomies and structures for collection development. In *8th International Conference on Grey Literature*. 11–19. Retrieved from [http://greyguide.isti.cnr.it/wp-content/uploads/2017/04/GL8\\_Conference\\_Proceedings.pdf#page=11](http://greyguide.isti.cnr.it/wp-content/uploads/2017/04/GL8_Conference_Proceedings.pdf#page=11)
- Gelfand, J., & Lin, A. (2013). Research life cycle: Exploring credibility of metrics and value in a new era of escholarship that supports grey literature. In *14th International Conference on Grey Literature* (pp. 123–134). Retrieved from <http://www.scopus.com/inward/record.url?eid=2-s2.0-84924135181&partnerID=tZOtx3y1>
- Higgins, J., & Green, S. (Eds.) (2011). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. The Cochrane Collaboration. Retrieved from <http://www.handbook.cochrane.org>
- Kling, R., & McKim, G. (2000). Not just a matter of time: Field differences and the shaping of electronic media in supporting scientific communication. *Journal of the American Society for Information Science*, 51(14), 1306–1320. [https://doi.org/10.1002/1097-4571\(2000\)9999:9999<::AID-ASI1047>3.0.CO;2-T](https://doi.org/10.1002/1097-4571(2000)9999:9999<::AID-ASI1047>3.0.CO;2-T)
- LaFleur, L., & Rupp, N. (2004). Making grey literature available through institutional repositories. In *Sixth International Conference on Grey Literature*. 30–34 Retrieved from [http://greyguide.isti.cnr.it/wp-content/uploads/2017/04/GL6\\_Conference\\_Proceedings.pdf#page=38](http://greyguide.isti.cnr.it/wp-content/uploads/2017/04/GL6_Conference_Proceedings.pdf#page=38)

- Lambert, S., Matthews, B. M., & Jones, C. (2006). Grey literature, institutional repositories, and the organisational context. *The Grey Journal*, 2(2), 85–90. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=21386583&site=ehost-live&scope=cite>
- Li, X., Thelwall, M., & Kousha, K. (2015). The role of arXiv, RePEc, SSRN and PMC in formal scholarly communication. *Aslib Journal of Information Management*, 67(6), 614–635. <https://doi.org/10.1108/AJIM-03-2015-0049>
- Linde, P., Eriksson, J., Kullman, L., Fathli, M., Karlsson, K., Sikström, M., . . . Tång, I. (2011). Accessibility and self-archiving of conference articles: A study on a selection of Swedish institutional repositories. *Information Services and Use*, 31(3–4), 259–269. <https://doi.org/10.3233/ISU-2012-0656>
- Lynch, C. (2017). Updating the agenda for academic libraries and scholarly communications. *College & Research Libraries*, 78(2), 126–30. <https://doi.org/10.5860/crl.78.2.126>
- Lyons, S. (2006). Preserving electronic government information: Looking back and looking forward. *The Reference Librarian*, 45(94), 207–223. [https://doi.org/10.1300/J120v45n94\\_12](https://doi.org/10.1300/J120v45n94_12)
- Mahood, Q., Van Eerd, D., & Irvin, E. (2014). Searching for grey literature for systematic reviews: Challenges and benefits. *Research Synthesis Methods*, 5(3), 221–234. <https://doi.org/10.1002/jrsm.1106>
- Melero, R., Abadal, E., Abad, F., & Rodríguez-Gairín, J. M. (2009). The situation of open access institutional repositories in Spain: 2009 report. *Information Research*, 14(4), 3. Retrieved from <http://www.informationr.net/ir/14-4/paper415.html>
- Moulaison, S. 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>
- National Library of Medicine. (2018). *List of serials indexed for online users*. Retrieved from <https://www.nlm.nih.gov/tsd/serials/lsiou.html>
- Okoroma, F. N. (2011). Towards effective management of grey literature for higher education, research and national development. *Library Review*, 60(9), 789–802. <https://doi.org/10.1108/00242531111176808>
- Park, E. G., & Richard, M. (2011). Metadata assessment in e-theses and dissertations of Canadian institutional repositories. *The Electronic Library*, 29(3), 394–407. <https://doi.org/10.1108/02640471111141124>
- Pinfield, S. (2002). Creating institutional e-print repositories. *Serials*, 15(3), 261–264. <https://doi.org/10.1629/15261>
- Riley, J. (2017). *Understanding Metadata: What Is Metadata, and What Is It For?: A Primer*. Baltimore: National Information Standards Organization. <http://www.niso.org/publications/understanding-metadata-2017>

- Rizor, S. L., & Holley, R. P. (2014). Open access goals revisited: How green and gold open access are meeting (or not) their original goals. *Journal of Scholarly Publishing*, 45(4), 321–335. <https://doi.org/10.3138/jsp.45.4.01>
- Rychlik, M. (2016). The role of the academic library in disseminating grey literature - Adam Mickiewicz University Repository as a case study. *The Grey Journal*, 12, 30–64.
- Schöpfel, J. (2011). Towards a Prague definition of grey literature. *The Grey Journal*, 7(1), 5–18.
- Schöpfel, J., & Farace, D. J. (2010). Grey literature. In M. J. Bates & M. N. Maack (Eds.), *Encyclopedia of library and information sciences* (3rd ed.) (pp. 2029–2039). Boca Raton, LA: CRC Press. <https://doi.org/10.1081/E-ELIS3-120043732>
- Schöpfel, J., Le Bescond, I., & Prost, H. (2012). Open is not enough: A case study on grey literature in an OAI environment. *The Grey Journal*, 8(2), 112–124.
- Seymour, D. (2010). In the trenches around the ivory tower: Introduction to black-and-white issues about the grey literature. *Archaeologies*, 6(2), 226–232. <https://doi.org/10.1007/s11759-010-9130-z>
- The Methods Group of the Campbell Collaboration. (2016). *Methodological expectations of Campbell Collaboration intervention reviews: Conduct standards*. Retrieved from <https://www.campbellcollaboration.org/mec2ir.html>
- Velden, T., & Lagoze, C. (2009). White paper: The value of new scientific communication models for chemistry, 1-70. Retrieved from [https://ecommons.cornell.edu/bitstream/handle/1813/14150/whitepaper\\_final.pdf?sequence=4](https://ecommons.cornell.edu/bitstream/handle/1813/14150/whitepaper_final.pdf?sequence=4)
- Xia, J. (2008). A comparison of subject and institutional repositories in self-archiving practices. *Journal of Academic Librarianship*, 34(6), 489–495. <https://doi.org/10.1016/j>

## APPENDIX

### Adapted from AAUW 2014-15 List of Academic Fields (AAUW 2014)

#### **Natural Sciences**

- Agricultural Sciences
- Animal Sciences
- Biochemistry
- Biology
- Cell and Molecular Biology
- Ecology
- Entomology
- Fisheries
- Food Sciences
- Forestry
- Genetics
- Horticulture
- Marine Biology
- Natural Resources Sciences
- Nutrition
- Parks and Recreation Management
- Plant Sciences
- Soil Sciences
- Wildlife Management
- Zoology

#### **Physical Sciences & Engineering**

- Astronomy
- Atmospheric Sciences
- Chemistry
- Civil Engineering
- Computer and Information Sciences
- Earth Sciences
- Electrical Engineering

- Industrial Engineering
- Materials Engineering
- Mathematics
- Mechanical Engineering
- Physics

**Health Sciences**

- Chiropractic Sciences
- Dental Sciences
- Environmental Health
- Epidemiology
- Health Science Administration
- Medical Sciences
- Nursing
- Occupational Therapy
- Optometry
- Pharmaceutical Sciences
- Physical Therapy
- Public Health
- Speech/Language Pathology
- Veterinary Medicine

**Social Sciences**

- Anthropology
- Archaeology
- Economics
- Home Economics
- Law
- Political Science
- Psychology
- Public Administration
- Social Work
- Sociology

## **Humanities**

- Communications
- English Language and Literature
- Foreign Language and Literature
- History (except Art History)
- Philosophy
- Religion and Theology

## **Arts**

- Architecture
- Art History
- Performance Studio Arts

## **Education**

- Curriculum and Instruction
- Early Childhood Education
- Education Administration
- Education Evaluation and Research
- Elementary Education
- Higher Education
- Secondary Education
- Special Education
- Student Counseling and Personnel Services

## **Business**

- Accounting
- Banking
- Business Administration
- Finance
- Management

## **Multidisciplinary\***

\*Includes more than one of the broad disciplines listed above



