ISSN 2162-3309 10.7710/2162-3309.2201

PRACTICE

Harvesting the Academic Landscape: Streamlining the Ingestion of Professional Scholarship Metadata into the Institutional Repository

Jonathan Bull
Scholarly Communication Services Librarian, Valparaiso University
Teresa Schultz
Scholarly Communications and Copyright Librarian, University of Nevada, Reno

INTRODUCTION Although librarians initially hoped institutional repositories (IRs) would grow through researcher self-archiving, practice shows that growth is much more likely through library-directed deposit. Libraries must then find efficient ways to ingest material into their IR to ensure growth and relevance. **DESCRIPTION OF PROGRAM** Valparaiso University developed and implemented a workflow that was semiautomated to help cut down on the time needed to ingest articles into its IR, ValpoScholar. The workflow, which continues to be refined, makes use of practices and ideas used by other repositories to more efficiently collect metadata for items and upload them to the repository. **NEXT STEPS** The article discusses the pros and cons of this workflow and areas of ingesting that still need to be addressed, including adding full-text items, checking copyright policies, managing student staffing, and dealing with hurdles created by the repository's software.

Received: 05/22/2017 Accepted: 12/07/2017

Correspondence: Jonathan Bull, Christopher Center, Room 265, Valparaiso University, Valparaiso, IN 46383-6493, Jon.Bull@valpo.edu

INTRODUCTION

In the past two decades, institutional repositories (IRs) have become a common part of the scholarly communication ecosystem, with 2,823 existing across the world (OpenDOAR, 2017). For IR administrators, though, the challenge of sustainability beyond the initial project launch is paramount. This can be an even greater challenge for smaller institutions with fewer resources to support an IR.

While many institutions have shown the impact of growing IR content, many IRs continue to struggle to show sustained growth (Cullen and Chawner, 2011; Kim, 2010; Marsh, 2015). Clifford Lynch recently addressed this stagnant growth by asking, "Is it important for institutions to maintain a local comprehensive record of their scholarly output through their IR?" (Lynch, 2017, p. 127). Because most institutions routinely create intellectual content for a world of information-seeking audiences, these questions of stagnating IRs and whether or not the institution needs a local collection of its scholarly record are interesting and complex, to say the least.

One possible reason for this stagnation and inability to capture the institution's scholarly record could be an institutional inability to move from the initial content of the pilot project phase to a sustained workflow for content recruitment. This paper presents an example of a workflow designed for library-supported IRs that seek to move from a pilot phase or a post-pilot stagnation phase to the use of a fully operational metadata archiving service. This service would be similar to Current Research Information Systems (CRIS), while still allowing for full-text availability when possible. The workflow was designed at Valparaiso University's Christopher Center Library Services (CCLS) to perform all tasks required to populate the IR, ValpoScholar, with limited staff involvement. Factors that contributed to the desire for a new workflow at CCLS included ineffective indexing, poor metadata, and lack of institutional buy-in as well as a lack of a CRIS. Although many kinds of works may be included in an IR, this workflow focuses specifically on scholarship metadata produced by faculty and professional staff—typically, these are journal articles, book chapters, and conference proceedings.

LITERATURE REVIEW

Several approaches to populating IRs with faculty scholarship may be found in scholarly literature on the topic. These include focusing on the implementation or pilot stage, self-archiving or direct deposit, retraining library staff/faculty to recruit and add content, and streamlining or automating workflows. In addition, several institutions have also experimented with interoperability and even merging the IR with CRIS.

Content Population via Initial Pilot Project Phase

For those who saw the need for and potential benefits of IRs, many struggled with how best to launch, integrate, and populate their institution's repository. Palmer, Teffeau, and Newton (2008) point out three different but common approaches to IR implementation, including a different content focus for each model (p. 149). In regard to IR content recruitment, many institutions created collection development policies as one of the first steps in adding content to the IR; these were similar to policies developed for traditional subject-specific library collections. These policies often determine the "purpose of the repository, scope of the collections, selection criteria and responsibilities, editorial rights, and how to handle challenges and withdrawals" (Wang, 2011, p. 83). The details of collection development policies vary from institution to institution. For example, some institutions' policies focused on which campus unit might be responsible for the IR or subsections of the IR (Baudoin & Branschofsky, 2003, pp. 36–37), what priority criteria of material should be added (Cohen & Schmidle, 2007, p. 289), or how to identify the faculty needs regarding and interest in repositories before recruiting content from the faculty itself (Makori, Njiraine, & Talam, 2015, p. 618).

Some institutions focused on electronic thesis and dissertation content first before moving to faculty and staff content (Wrenn, Mueller, & Shellhase, 2009), while others have identified initial content partners as well as a "collaborative IR model" across multiple institutions for content recruitment (Oguz & Davis, 2011, p. 14). These and other approaches have all shown varying levels of success at or near the implementation stage at a variety of institutions (Nykanen, 2011, pp. 17–18).

Content Population via Self-Archiving

Initially, many IR administrators approached faculty members to deposit their own work, or "self-archive," believing that faculty would want to make their work more visible. While the idea of the IR is usually well received, asking faculty to self-archive "did not translate into real content being deposited," resulting in stagnating growth (Mackie, 2004, n.p.). Several other studies found self-archiving rates to be low, with the majority of scholarship made accessible by someone other than the researchers themselves (Foster & Gibbons, 2005; Xia & Sun, 2007; Covey, 2009; Covey, 2011).

Further studies found similar inaction at other institutions. Davis and Connolly (2007) interviewed eleven faculty members in the sciences, social sciences, and humanities about their opinions on depositing their work in an IR. Some reasons for not using a IR include: learning curve, copyright concerns, publishing concerns ("Is a preprint in an IR considered

publishing?"), quality association, fear of plagiarism and being scooped, reputation and the importance of accuracy, and use of subject repositories ("Publishing Original Work"). If faculty members did deposit, they did so not "without a lot of coaxing." (Koopman & Kipnis, 2009, p. 115).

Another study examined if researchers in a discipline (physics) with a "familiarity with self-archiving" via a subject repository (Arxiv.org) would deposit into their respective institutional repositories at a higher rate than researchers without such familiarity and vice versa (Xia, 2008). The study suggested that "when an article has been presented in one repository, the author(s) will be hesitant to make it repeatedly available in a second repository" (p. 494).

Many institutions have addressed self-archiving through open access policies (or mandates) via associations such as the Coalition of Open Access Policy Institutions (COAPI, n.d.). However, initial buy-in for self-archiving practices resulting from an open access policy was limited, as many faculty members still lacked the technical skill or time to deposit, or were still concerned about copyright infringement (Xia et al., 2012), while others were worried about plagiarism or concerned about the newness or small scale of the repository (Singeh, Abrizah, and Karim, 2013). In addition to those concerns, faculty members also "do not see the benefit of open access reflected in the tenure process, so they fail to deposit items into the repository" (Xia et al., 2012, p. 98).

Content Population via Library Staff/Faculty Retraining and Outreach

While convincing faculty to self-archive has proven difficult, many proponents have advocated for library personnel to be retrained and handle marketing and outreach for faculty scholarship themselves. Bailey (2005) proposed reference librarians—considered by libraries to be the eyes and ears of their institutions—as possible servicers of content population (p. 266). Jenkins, Breakstone, and Hixson (2005) elaborate on the importance of librarians in getting the word out by "conveying [IR value] effectively to authors," which is only possible if those librarians "understand the culture of scholarly communication locally and beyond" (p. 315). Yet, few studies address the varying roles that reference librarians play from institution to institution, specifically small vs. larger institutions, and how those librarians would be able to take on these extra duties.

Reference is not the only branch of academic librarianship viewed as having an important part to play in IR content population. Depending on the repository platform, technical services personnel can play a direct role in content population, specifically collection development, cataloging/metadata, and preservation (Connell and Cetwinski, 2010; Aucock, 2012). In addition, the retraining of the library paraprofessional staff on repository-related tasks can

also have a positive impact on populating the repository, while giving these personnel additional professional development (Bull & Eden, 2014). Duranceau and Kriegsman (2013) reviewed a variety of ways libraries are responding to open access mandates and their effect on IRs. Overall, an OA mandate can help in a variety of ways, including motivating staff retraining and outreach, refueling faculty interest, and creating new partnerships with campus research offices and publishers. However, estimating and planning a workload for IR-related duties can still be difficult, due to the year-round, uneven publication cycle of academia.

Content Population via Workflows, Batches, and Automation

For achieving sustainable, systematic growth, adding batches of content through specified workflows shows strong potential. Mackie (2004) outlined different approaches for content recruitment in the short, intermediate, and long-term periods. For short and intermediate periods, the University of Glasgow (UG) identified prospective adopters via staff/faculty websites and previous publications and determined which journals might approve population and those materials already available via open access online. After finding that these "relatively small-scale approaches" for the short and intermediate term that were "not sustainable," UG investigated more "systematic" approaches to IR population, such as adding bibliographic records through their Reference Manager with the assistance of a PERL script or through "departmental/faculty publication databases" (Mackie 2004, "Long-Term Strategies"). These longer-term approaches to IR population workflow allow for a "significant percentage of the process to be automated," but with complete automation still out of reach ("Long-Term Strategies").

Madsen and Oleen (2013) outlined the importance of moving beyond a single-person-managed IR to a locally developed Workflow Management System (WMS). This system assigns the content population responsibilities to a variety of personnel, an approach that is similar to that of other institutions but also utilizes RefWorks and Local-Area Network (LAN) for greater IT-related efficiency. While this workflow is much more efficient and has added a great deal of value, it still is not automated or balanced.

Flynn, Oyler, and Miles (2013) added to this push to automate IR workflow by creating a Google Script that could check a publisher's copyright policy in SHERPA/RoMEO within a spreadsheet. Strauss and Miles (2014) built on this further by creating a full semiautomated workflow using Excel, scripts, faculty CVs, and Google Drive. Utah State University attempted a similar form of automation, trying to import publication documentation, including the scholarship itself, from Digital Measures into the IR (Wesolek, 2014). This batch import proved to be successful, but it still required some human interaction, despite limited staffing. However, using Digital Measures relies on faculty to input their works correctly.

Any mistakes they make would then be carried over into the IR, which might require additional human interaction. Kipphut-Smith (2014) also discussed Rice University's attempt to streamline repository population due to an open access policy, though much of the workflow is specific to the requirements of the repository platform (DSpace) and did not address other platform challenges.

Zhang, Boock, and Wirth (2015) also advocated for using existing citation data available in various databases, specifically Web of Science, to help populate an IR, as well as for using existing IR citation data to check how much of it is indexed in Web of Science. While this approach is useful, it relates only to DSpace repositories and is not without problematic citation data (pp. 5–7). Using XML and an XSL stylesheet, Li (2016) designed a process for ingesting citations from Web of Science into Digital Commons, increasing input from 50 to 1,000 records into the IR over the same period of time. However, this process did not check copyright or full-text acquisition.

Current Research Information Systems and Institutional Repositories

Parallel to the development of the IR, Current Research Information Systems (CRIS) were created to track faculty output and research documentation. After summarizing the historic need for CRIS, Joint (2008) first suggested that repository librarians "shadow these larger patterns of integration by examining the place of their own repositories within the local 'campus research information system,'" yet left the question of what the relationship of the IR and CRIS might look like unanswered, acknowledging some, but not complete, overlap (p. 571).

A few additional studies did consider this IR-CRIS question more fully. For example, Viner (2010) and Tate (2012) did develop respective possible workflows for migrating CRIS full-text submissions, mainly for dissertations/theses, into an institution's repository. However, both of these studies only discussed situations where an institution has both a CRIS and an IR. Most recently, Rybinski et al. (2017) suggested combining the functionality of a CRIS and an IR with a focus on the users' behavior and needs throughout the entire research cycle into a single platform. This idea is interesting but might be difficult to implement with limited staffing.

Case Study: A New Workflow for Harvesting Citation Data into the Repository

Background

Prior to the creation of the IR, Valparaiso University did attempt to implement a CRIS in

order to track faculty research, but that attempt failed to gain faculty buy-in and was phased out within a semester. When launching ValpoScholar, the institution's IR, this failed attempt to track faculty scholarship was cited as a possible obstacle to IR success. As a result, ValpoScholar was marketed as an IR, despite much of its initial workflow mimicking the function of CRIS. Since its creation in 2011, ValpoScholar's workflow has focused on creating metadata records prior to depositing articles. Records are created without concern for whether the accompanying item can be deposited. Instead of depositing the actual artifact, staff instead initially link to the article's web page with its journal or publisher. Only after a record for an item has been created in ValpoScholar do staff research the item's copyright and whether it is allowed to be deposited. If the item may be deposited, staff do so. However, if an item is under embargo, staff make note of when the embargo lifts. The process has been time-consuming, as standard procedure called for creating records on an individual basis, instead of in a batch. For the first four years, CCLS faculty and staff kept track of this entire process, including records created and copyright and deposit status for items, through a series of Excel worksheets.

As the repository grew, the worksheets became clunky, and in the fall of 2014, CCLS decided to investigate a new way to handle the workflow in the hopes of streamlining it. The project quickly grew, however, into a search for ways to help automate at least part of the process, especially as CCLS has few resources dedicated to the IR. Library staff perform the actual work of creating records and depositing items into the repository. No one person was dedicated to this task, however, and it was often up to staff to assist during their downtime. A new process that could make quicker work of record creation and depositing would mean that CCLS could increase deposit rates without requiring significantly more personnel hours.

Initially, staff hoped to base their workflow on Flynn et al.'s (2013) system, which helped automate part of the process at the College of Wooster. However, Wooster's repository is based on DSpace software, and Valparaiso decided that CCLS staff, who did not have any significant IT support for ValpoScholar other than through their vendor, bepress, did not have the ability to adapt the full process for their needs. The library did take advantage of one aspect of Wooster's workflow: the code that allows a user to search Google Sheets containing embargo policies from SHERPA/RoMEO.

A scholarly communication listserv also provided CCLS staff with other workflow ideas when a librarian for the Smithsonian noted that they had set up alerts for articles and then collected the metadata from these alerts into a citation manager. These alerts allowed the Smithsonian Libraries to passively collect articles by setting up email alerts with major journal publishers and databases (A. Hutchinson, personal communication, 2015). This

passive collection is well suited to a small library such as CCLS and can be easily set up. The library began collecting items from its email account and saving the citations to be pulled into the citation manager Zotero. Unfortunately, Smithsonian Libraries also use DSpace software for their repository, so CCLS determined that the rest of their workflow could not be directly adapted.

Both Wooster and the Smithsonian did provide ideas about how the process could be partially automated, however, including editing metadata for multiple records at once. Up until this point, CCLS staff created records individually for ValpoScholar, an easy but time-consuming task. They had not utilized Digital Commons' batch upload feature and thought that adapting the DSpace-specific process for editing metadata used by both Wooster and the Smithsonian Libraries could fit the batch upload process.

Creating the Process

Adapting the process was not easy, though, and CCLS staff ran into several starts and stops. For instance, an intern discovered that collecting citations for items through the reference manager EndNote allowed for easy editing of certain parts of the metadata. An initial workflow was then created that used a mix of EndNote and Excel to easily edit metadata in bulk. However, Valparaiso does not have full access to EndNote, and a concern was raised about the fact that EndNote is the product of a for-profit company; thus, using it goes against the open access ethos. CCLS then decided that it would be better to rely on open source programs and to try Zotero instead.

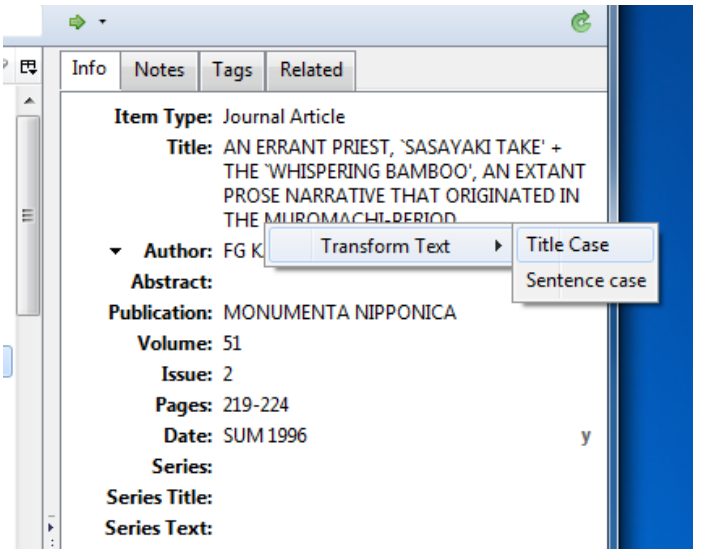


Figure 1. Using Zotero to change the title of an article to title case

Zotero has both the benefits and drawbacks of an open-source program, however, including offering fewer options than its for-profit rivals. Some of the options available to help edit metadata in EndNote do not exist in Zotero (or did not at the time of the project). This rendered the initial workflow useless, and CCLS had to study other options to replace some of the functionality that EndNote provided. CCLS had to rely on other, existing programs for all editing functions, as no staff had the coding knowledge needed to create their own functions and the library did not have the needed support from IT.

The workflow was slowly cobbled together over about a year, undergoing several overhauls as CCLS discovered better, more efficient ways to perform certain functions. It finally established a process that could be successfully replicated using a mix of editing functions in Zotero and Google Sheets. The overall process begins with the passive collection of items through email alerts for "Valparaiso University" that were set up with most of the major journal publishers, Web of Science, and Google Scholar. This process brings in about five to 10 citations a week. Care needs to be taken when gathering these items, however, as the items are not always recent research publications by VU faculty. Google Scholar, in particular, has shown to be hit-or-miss, as it often finds mentions of "Valparaiso University" in acknowledgments, captions, and other mentions within an item that do not refer to an author's institution.

Citations for these items are then collected once a week in Zotero and organized by department. This step continues until a decent number of articles—about 50 to 75—have been collected, and then the process proceeds. After formatting author names, journal titles, and article titles within Zotero to fit the repository's style (Fig. 1), the citations are exported as a CSV file and imported into Google Sheets (Figs. 2–3).

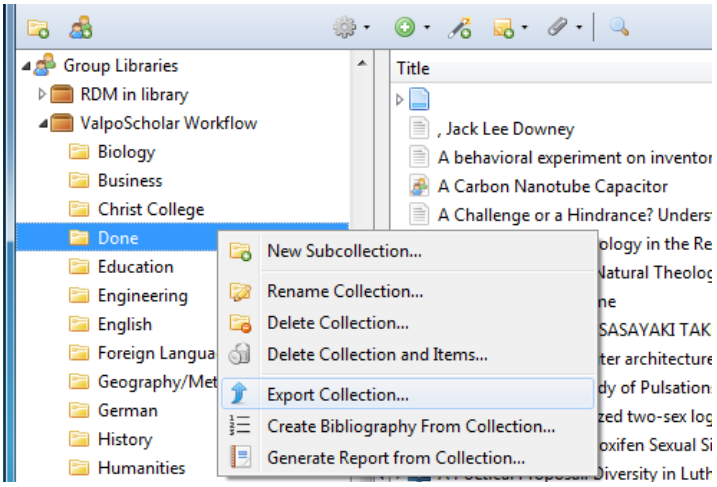


Figure 2. Step one of exporting a Zotero collection

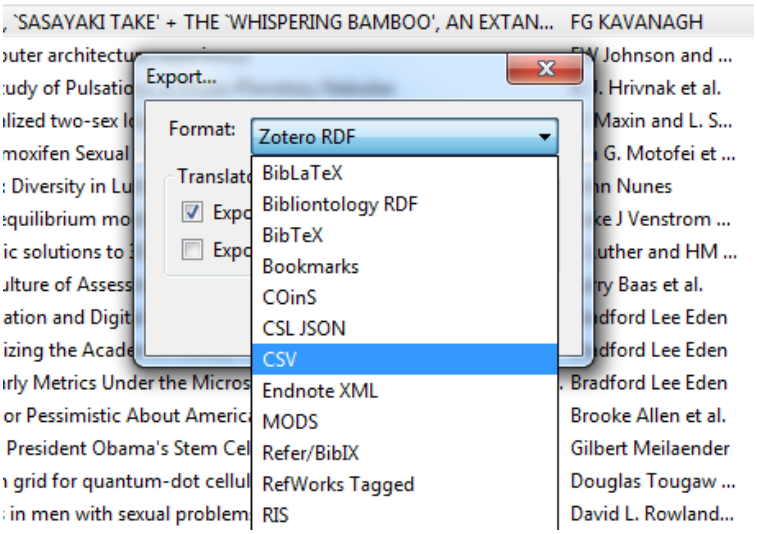


Figure 3. Step two of exporting a Zotero collection

These files come with a number of unnecessary columns, which are deleted, and more metadata editing is done in bulk using both the simple "find and replace all" function and the free add-on tool Power Tools (Figs. 4–5). Unfortunately, not all of the editing can be done in bulk; some items must be typed in manually. However, the process allows for much of the data to be added or edited at once instead of going by individual article.

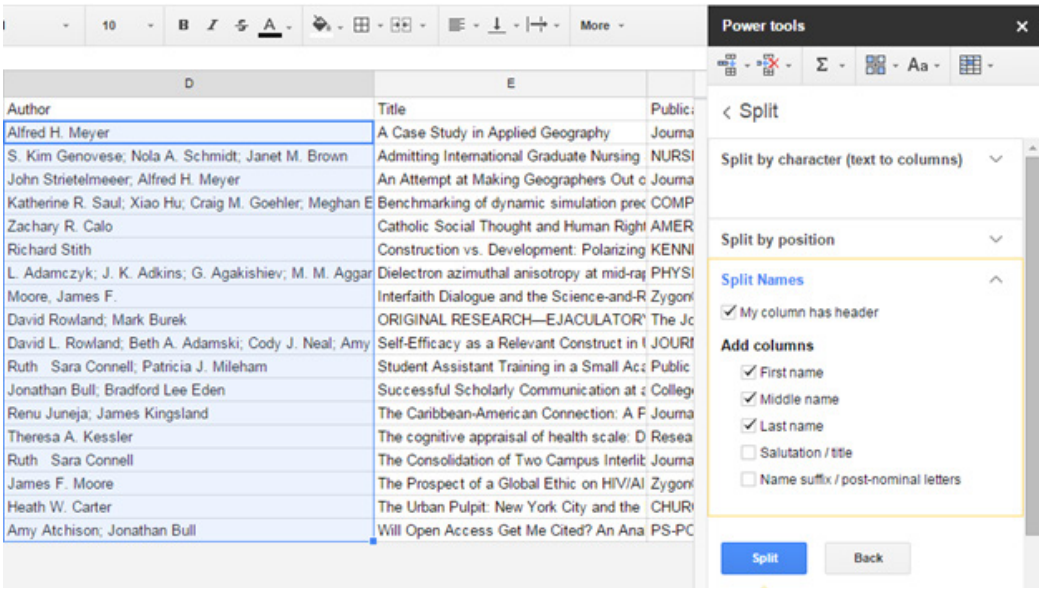
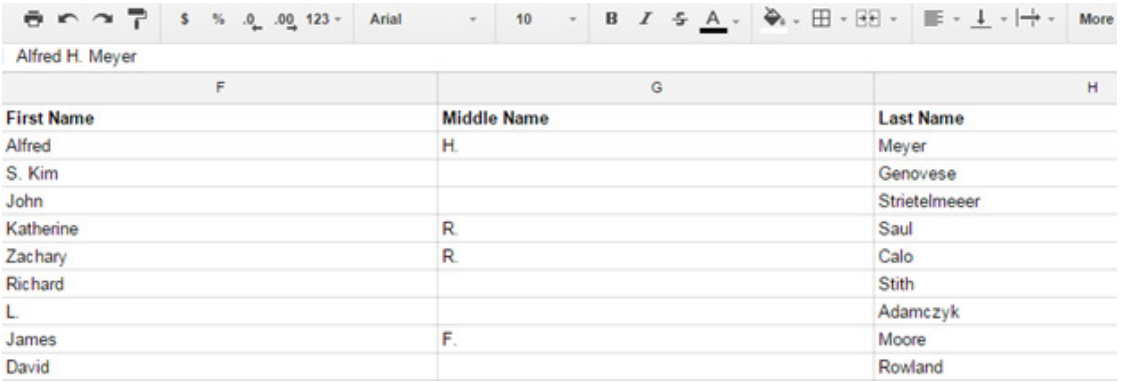


Figure 4. Using the Power Tool split function to separate names into separate cells



Alfred H. Meyer

F	G	H
First Name	**Middle Name**	**Last Name**
Alfred	H.	Meyer
S. Kim		Genovese
John		Strietelmeeer
Katherine	R.	Saul
Zachary	R.	Calo
Richard		Stith
L.		Adamczyk
James	F.	Moore
David		Rowland

Figure 5. Resulting columns after using the Power Tool split function

The metadata is then copied and pasted into the appropriate columns in bepress's Excel batch upload spreadsheet for uploading into ValpoScholar. From Google Sheets, Valparaiso runs Wooster's automated search of SHERPA/RoMEO for publishers' embargo policies. Staff then deposit articles where journal policies allow.

This process was created to collect new works of research produced by Valparaiso faculty, but it can also be used for older works. CCLS staff have performed historical searches of Web of Science and other large databases for older works by VU faculty and collect those articles into Zotero. The process follows the same path from there.

A training document was created to help teach new staff how to use this process. Initial testing on the process took place in summer and fall 2016.

Lessons Learned

Prior to the implementation of this workflow, ValpoScholar saw a decline in overall record creation after an initial period of significant growth (see Fig. 6). Since fully implementing this process in summer 2016, two batches (109 and 102, respectively) have been completed, contributing to an increase in the number of records created in 2016 compared to the two prior years. The workflow also helped maintain an increase in the number of records of faculty-related scholarship (see Fig. 7). Records created in SelectedWorks, our research profile service, were not included in the figures for this study, as many of those records were created outside of the traditional workflow and by the authors themselves.

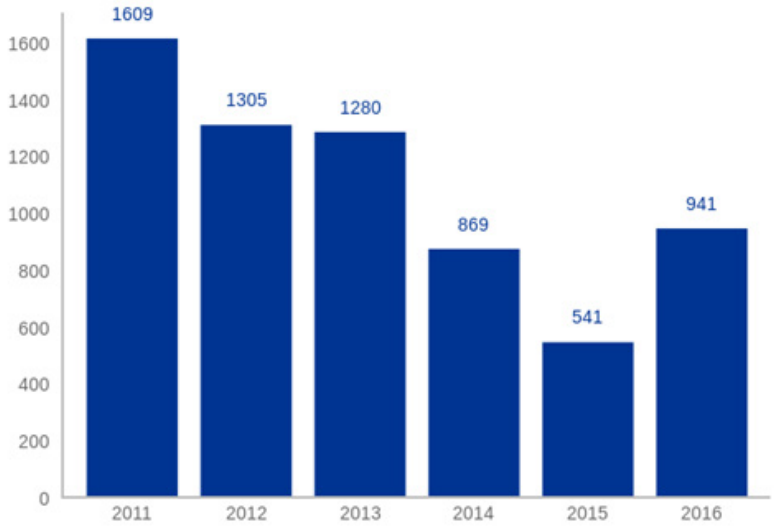


Figure 6. The number of all records created each year in Digital Commons, 2011–2016

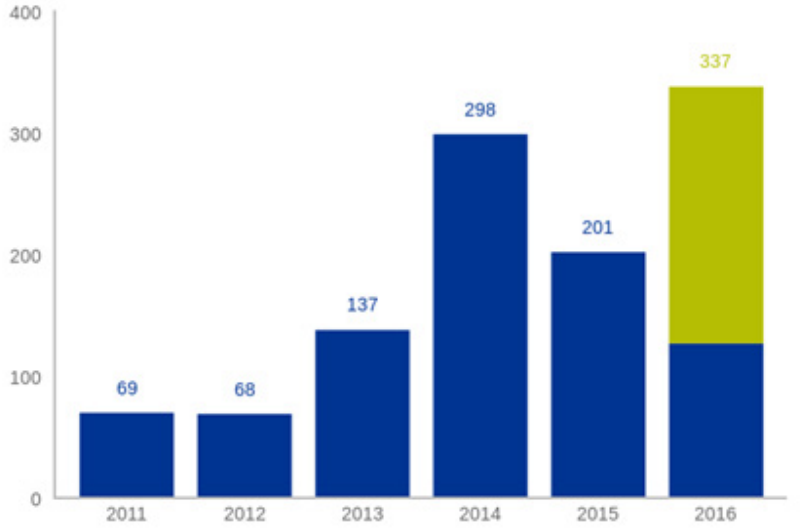


Figure 7. All records created by year for faculty works. The new workflow started in 2016 after 126 works (bottom of the bar) had already been uploaded. Works added using the new workflow are represented by the upper portion of the bar.

Although the new workflow did help increase the number of faculty records created compared to prior years, the process is by no means perfect. Adding 211 scholarship-related records as a result of the new workflow may be positive for a smaller institution, especially if it can be done in select batches. However, despite these numbers, the process proved to be more time-consuming and problem-laden than previously envisioned.

Bulk/Batch Editing vs. Individual Record Creation

Several obvious shortcomings still exist. While bulk/batch editing can be more efficient in many ways, it still takes several hours to go through the whole process. Because records must be uploaded one department at a time, the process is no more efficient than individual record creation for a department that has just one or two new items. However, for those departments that do see more activity, bulk editing of metadata reduces the time spent formatting records. Because the process also relies on programs created by third parties, especially for-profit Google, it is also at the mercy of any changes made to these programs. Google Sheets and Zotero handle most of the needed editing functions now, but that could change in the future. One add-on tool for Google Sheets that this process makes use of, Power Tools, was available entirely for free when this process started but has since begun to make some of its features available only under a paid version. Lack of coding knowledge also prevented CCLS staff from taking further advantage of Google's add-on programs. The company allows users to create their own codes to perform new functions and add them into Google Sheets, the same process used at College of Wooster. Coding knowledge would help buffer CCLS in the future should Google Sheets take away certain functionalities, and could help them create new functionalities not currently offered by Google Sheets.

CCLS also had to overcome certain issues because of the way the Digital Commons platform handles bulk uploads. As stated earlier, CCLS first creates a record for an item and links out to that item's webpage. This has not been a problem with individual uploads. However, the bulk upload Excel sheet does not allow URLs unless they link out to the full-text item on a publicly available server. When a link is included to an article's webpage where it lives behind a paywall, the upload is rejected. CCLS has thus had to eliminate including the URL when batch uploading records. The batch upload spreadsheet for Digital Commons also requires that each part of an author's name lives in its own cell. When Zotero exports metadata into a CSV file, it groups all authors' names into one cell, so a cumbersome step had to be added here to split these names up.

This new workflow process is also slowed down if there is any inconsistent or nondedicated staffing. While undergraduate or graduate student staffing can work, it can also become

problematic. Student workers will likely need to be retrained each time a new batch is prepared, unless they can make a multisemester commitment to the project and the position. The best solution for managing a semiregular project like this may be to dedicate a professional staff member who would have a long-term commitment to the project and be more comfortable with this project's level of human interaction. "Batch time" could be scheduled on the worker's yearly calendar. Depending on the size of the institution, determining a batch upload schedule could be tricky, but it is essential if the dedicated staff is to remain up-to-date on the batch upload procedure.

Unsolicited Submissions, Too Many Saved Searches, and De-Duplication

Another unresolved issue in this process is how to account for the independent submission of scholarship from faculty and staff members. These unsolicited submissions, while previously very welcomed, complicate this new workflow, as they can sometimes have been added in a previous batch, include erroneous or incomplete (author-supplied) information, or be delayed for inclusion in a future batch. For the workflow to succeed, the unsolicited submissions need to be added to the current batch-in-progress. In addition to the unsolicited submission possibly creating duplication, having too many saved searches could result in many duplicated notifications. For a small institution with limited experience in collecting faculty citation information or for the larger institution with tremendous scholarly output, trying to capture all of these notifications might be the only way to discover the institutional scholarship. However, it also creates a great deal of duplication, as many publishers list their publications in several aggregators.

Incomplete Metadata

Another problem that arises from semiautomated data collection is that the ISSN is not a standard piece of metadata included in citation information. This portion of metadata is essential to determining the embargo status of scholarship, yet it is rarely included in unsolicited submissions or in collected aggregator/publisher notifications. Even if the ISSN is easily discoverable or included, it does not mean that the journal's information will be in Sherpa/RoMEO or, if it is, that it will be accurate.

NEXT STEPS

Several challenges remain in order to make this process more efficient. These include, in particular, ensuring full-text availability, streamlining notifications, and providing more dedicated staffing.

Full-Text Availability Remains a Challenge

Even though adding full-text availability was outside the scope of this case study, it still needs to be addressed in the future. While much of this scholarship is available in full text through a library database, downloading these versions from a database might violate institutional licensing agreements with that vendor. When considering this, directly requesting full-text files from the authors themselves may be best, but even if they comply, many faculty members will still likely only send the publisher's final version, which may be most restricted in relation to IR use. A greater effort will need to be made to educate faculty members on why a preprint (manuscript) or postprint (post-refereeing) version of their work could be more easily added to the repository.

More Efficient Saved Searches and Notifications

Considering the amount of duplicated notifications CCLS received from multiple saved searches, the search and notification processes will need to change. In particular, an overlap analysis will need to be conducted for these saved searches and the duplicated notifications will need to be deactivated, which will result in a more efficient stream of publication and indexing notifications.

More Dedicated Staffing

While this workflow could work with student worker support, future student staffing will need to be more sustainable, in order to save staff the time of repeated student training. This may preference the hiring of early-career students (i.e. freshmen or sophomores) or a full-time staff member to administer the workflow. Either way, more long-term staffing will likely be needed, albeit very limited increased staffing.

CONCLUSION

Moving the IR from the pilot phase to a sustained library practice can be problematic, but it can also help the library meet the institutional need of capturing the scholarly record. Many institutions are currently not capable of or choose not to do this, especially if they do not have a CRIS. Many IRs start their collections with a pilot project model, randomly creating records and adding materials from whatever can be found and is free to post. However, this model of random addition may not collect much of the institution's scholarly output and may paint an incomplete picture of the institution's professional scholarship. This fractured collection development might stem from a lack of awareness or access or an institutional culture of not reporting one's scholarship. However, often this may be

due to a lack of staffing and/or a consistent, defined workflow.

As presented in this case study, one possible solution to capturing more of the institution's scholarship is to implement a workflow that would allow for more scholarly records to be created over time with less staffing. By using automated notifications, freely available tools such as Google Sheets, Power Tools, and Zotero, and an IR software batch upload option, more of an institution's academic landscape can be harvested without much additional staffing as compared to the pilot project phase or the cost of implementing a CRIS. While there is overlap with this new workflow between an IR and a CRIS, it is important to stress that they do not share all of the same functionality, including providing open access to scholarly research by including the full text with the metadata record when possible.

As also evidenced in this case study, the flow of IR submissions is likely to become more problematic than initially anticipated, due to staffing inconsistencies, unsolicited submissions, and duplicated uploads—to name just a few complications. As presented, this workflow will need to be amended in the future—despite an initial increase in metadata capture and record creation—to include more involvement from professional and graduate student staffing (and perhaps less undergraduate staffing), fewer saved searches, and a new procedure for accepting unsolicited items that will need to be implemented. If others wish to pursue a semiautomated method to populating one's IR, as presented in this case study, several factors need to be considered at length:

1. Technical Requirements of the IR: Do you have a vendor or locally hosted IR? What are the requirements for a batch upload? Does the vendor allow for changes to the batch upload requirements?
2. Addressing Quality Control for Record Creation: How does the IR handle large numbers of duplicate submissions and/or unsolicited submissions? Would a semiautomated method of submission complicate or improve this?
3. Staffing Concerns: Who works on the IR? Do you have dedicated professional staff or only student staff? Is there potential for high turnover (i.e. student workers)?
4. Stage of IR development: How old is your IR? How much financial, policy, and/or administrative buy-in does it have?

While more of the scholarly record is captured with this new workflow, problems remain, including a lack of full-text addition and copyright clearance, nondedicated or short-term staffing, incomplete metadata, and duplicated notifications and submissions. Limitations imposed by the Digital Commons batch-upload Excel file have also created difficulties,

both in adding steps to the metadata editing process and in the inability to include certain metadata.

However, even with these challenges, this workflow begins to address Clifford Lynch's recent challenge to "re-think about the real prospects and best approaches and roles for IRs" (Poynder, 2016, p. 13). It moves the IR from the pilot phase into a more standardized method of growth, further capturing the institution's scholarly record, while limiting reliance on the institutional patron (i.e. the faculty member) for submitting content and metadata.

REFERENCES

Aucock, J. (2012). Why not send a cataloguer? *Catalogue and Index*, *167*, 28–32.

Bailey, C. W. (2005). The role of reference librarians in institutional repositories. *Reference Services Review*, *33*(3), 259–267. https://doi.org/10.1108/00907320510611294

Baudoin, P., & Branschofsky, M. (2003). Implementing an institutional repository: The DSpace Experience at MIT. *Science & Technology Libraries*, *24*(1–2), 31–45. https://doi.org/10.1300/J122v24n01_04

Bull, J., & Eden, B. L. (2014). Successful scholarly communication at a small university: Integration of education, services, and an institutional repository at Valparaiso University. *College & Undergraduate Libraries*, *21*(3–4), 263–278. https://doi.org/10.1080/10691316.2014.932264

Coalition of Open Access Policy Institutions (COAPI). (n.d.). Retrieved November 8, 2016, from http://sparcopen.org/coapi/

Cohen, S., & Schmidle, D. (2007). Creating a multipurpose digital institutional repository. *OCLC Systems & Services: International Digital Library Perspectives*, *23*(3), 287–296. https://doi.org/10.1108/10650750710776422

Connell, T. H., & Cetwinski, T. (2010). The impact of institutional repositories on technical services. *Technical Services Quarterly*, *27*(4), 331–346. https://doi.org/10.1080/07317131003765993

Covey, D. T. (2009). Self-archiving journal articles: A case study of faculty practice and missed opportunity. *Portal: Libraries and the Academy*, *9*(2), 223–251.

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

Crow, R. (2002). The case for institutional repositories: A SPARC position paper. *Research on Institutional Repositories: Articles and Presentations*. Retrieved from http://digitalcommons.bepress.com/repository-research/27

Cullen, R., & Chawner, B. (2011). Institutional repositories, open access, and scholarly communication: A study of conflicting paradigms. *Journal of Academic Librarianship, 37*(6), 460–470. https://doi.org/10.1016/j.acalib.2011.07.002

Davis, P., & Connolly, M. (2007). Institutional repositories: Evaluating the reasons for non-use of Cornell University's installation of DSpace. *D-Lib Magazine*, *13*(3/4). https://doi.org/10.1045/march2007-davis

Duranceau, E. F., & Kriegsman, S. A. (2013). Implementing open access policies using institutional repositories. In P. Bluh and C. Hepfer (Eds.), *The institutional repository: Benefits and challenges* (pp. 75–97). Chicago: American Library Association. Retrieved from https://dash.harvard.edu/handle/1/10202474

Flynn, S. X., Oyler, C., & Miles, M. (2013). Using XSLT and Google Scripts to streamline populating an institutional repository. *Code4Lib Journal*, *19*. Retrieved from http://journal.code4lib.org/articles/7825?utm_source=tuicool

Foster, N. F., & Gibbons, S. (2005). Understanding faculty to improve content recruitment for institutional repositories. *D-Lib Magazine*, *11*(1). https://doi.org/10.1045/january2005-foster

Jenkins, B., Breakstone, E., & Hixson, C. (2005). Content in, content out: The dual roles of the reference librarian in institutional repositories. *Reference Services Review*, *33*(3), 312–324. https://doi.org/10.1108/00907320510611348

Joint, N. (2008). Current research information systems, open access repositories and libraries: ANTAEUS. *Library Review*, *57*(8), 570–575. https://doi.org/10.1108/00242530810899559

Kim, J. (2010). Faculty self-archiving: Motivations and barriers. *Journal of the American Society for Information Science and Technology*, *61*(9), 1909–1922. https://doi.org/10.1002/asi.21336

Kipphut-Smith, S. (2014). "Good Enough": Developing a simple workflow for open access policy implementation. *College & Undergraduate Libraries*, *21*(3–4), 279–294. https://doi.org/10.1080/10691316.2014.932263

Koopman, A., & Kipnis, D. (2009). Feeding the fledgling repository: Starting an institutional repository at an academic health sciences library. *Medical Reference Services Quarterly*, *28*(2), 111–122. https://doi.org/10.1080/02763860902816628

Li, Y. (2016). Harvesting and repurposing metadata from Web of Science to an institutional repository using web services. *D-Lib Magazine*, *22*(3/4). https://doi.org/10.1045/march2016-li

Lynch, C. (2017). Updating the agenda for academic libraries and scholarly communications. *College & Research Libraries*, *78*(2), 126–130. https://doi.org/10.5860/crl.78.2.126

Mackie, M. (2004, April 30). Filling institutional repositories: Practical strategies from the DAEDALUS Project. *Ariadne*, *39*. Retrieved from http://www.ariadne.ac.uk/issue39/mackie/

Madsen, D. L., & Oleen, J. K. (2013). Staffing and workflow of a maturing institutional repository. *Journal of Librarianship and Scholarly Communication*, *1*(3). https://doi.org/10.7710/2162-3309.1063

Makori, E. O., Njiraine, D., & Talam, P. (2015). Practical aspects of implementation of institutional repositories in Africa with reference to the University of Nairobi. *New Library World*, *116*(9/10), 610–640. https://doi.org/10.1108/NLW-10-2014-0125

Marsh, R. M. (2015). The role of institutional repositories in developing the communication of scholarly research. *OCLC Systems & Services: International digital library perspectives*, *31*(4), 163–195. https://doi.org/10.1108/OCLC-04-2014-0022

Nykanen, M. (2011). Institutional repositories at small institutions in America: Some current trends. *Journal of Electronic Resources Librarianship*, *23*(1), 1–19. https://doi.org/10.1080/1941126X.2011.551089

Oguz, F., & Davis, D. (2011). Developing an institutional repository at a medium-sized university: Getting started and going forward. *Georgia Library Quarterly, 48*(4). Retrieved from https://digitalcommons.kennesaw.edu/glq/vol48/iss4/5/

OpenDOAR. (2017). Growth of the OpenDOAR database – worldwide, institutional repositories. Retrieved from http://opendoar.org/onechart.php?cID=&ctID=&rtID=2&clID=&lID=&potID=&rSoftWareName=&search=&groupby=r.rDateAdded&orderby=&charttype=growth&width=600&height=350&caption=Growth%20of%20the%20OpenDOAR%20Database%20-%20Worldwide,%20Institutional%20Repositories

Palmer, C. L., Teffeau, L. C., & Newton, M. P. (2008). Strategies for institutional repository development: A case study of three evolving initiatives. *Library Trends*, *57*(2), 142–167. https://doi.org/10.1353/lib.0.0033

Poynder, R. (2016, September 22). Q&A with CNI's Clifford Lynch: Time to rethink the institutional repository? [Blog post.] Retrieved from https://www.richardpoynder.co.uk/Clifford_Lynch.pdf

Rybinski, H., Skonieczny, L., Koperwas, J., Struk, W., Stepniak, J., & Kubrak, W. (2017). Integrating IR with CRIS - A novel researcher-centric approach. *Program*, *51*(3), 298–321. https://doi.org/10.1108/PROG-04-2017-0026

Singeh, F. W., Abrizah, A., & Karim, N. H. A. (2013). What inhibits authors to self-archive in open access repositories? A Malaysian case. *Information Development*, *29*(1), 24–35. https://doi.org/10.1177/0266666912450450

Strauss, B., & Miles, M. (2014, July). *Automating workflow: From a trickle to a stream*. Paper presented at the Digital Commons Great Lakes User Group Meeting, Valparaiso University, Valparaiso, IN. Retrieved from http://engagedscholarship.csuohio.edu/msl_facpub/105

Tate, D. (2012). Implementing a Current Research Information System (CRIS) with an existing institutional repository (IR): A brief overview. *Catalogue & Index*, *167*, 24–27.

Viner, R. A. (2010). Pure silver: Reusing and repurposing bibliographic data in a current research information system and institutional repository. *Catalogue & Index*, *161*, 34–40.

Wang, F. (2011). Building an open source institutional repository at a small law school library: Is it realistic or unattainable? *Information Technology and Libraries*, *30*(2), 81–84. https://doi.org/10.6017/ital.v30i2.3008

Wesolek, A. (2014). Bridging the gap between Digital Measures and Digital Commons in support of open access: Or, how I learned to stop worrying and love human mediation. Retrieved from http://tigerprints.clemson.edu/cheer/scholcomm/all_scholcomm/4

Wrenn, G., Mueller, C. J., & Shellhase, J. (2009). Institutional repository on a shoestring. *D-Lib Magazine*, *15*(12). https://doi.org/10.1045/january2009-wrenn

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.acalib.2008.09.016

Xia, J., Gilchrist, S. B., Smith, N. X., Kingery, J. A., Radecki, J. R., Wilhelm, M. L., … Mahn, A. J. (2012). A review of open access self-archiving mandate policies. *Portal: Libraries and the Academy*, *12*(1), 85–102. https://doi.org/10.1353/pla.2012.0000

Xia, J., & Sun, L. (2007). Assessment of self-archiving in institutional repositories: Depositorship and full-text availability. *Serials Review*, *33*(1), 14–21. https://doi.org/10.1016/j.serrev.2006.12.003

Zhang, H., Boock, M., & Wirth, A. A. (2015). It takes more than a mandate: Factors that contribute to increased rates of article deposit to an institutional repository. *Journal of Librarianship and Scholarly Communication*, *3*(1), eP1208. https://doi.org/10.7710/2162-3309.1208