Staffing and Workflow for Institutional Repositories

Stephen Craig Finlay

Follow this and additional works at: https://digitalcommons.unomaha.edu/crisslibfacpub
Please take our feedback survey at: https://unomaha.az1.qualtrics.com/jfe/form/SV_8cchtFmpDyGfBLE
Staffing and Workflow for Institutional Repositories

Craig Finlay, Murray State College


As we enter the 2020s Open Access Institutional Repositories have graduated from being “well established components of many academic libraries,” as Madsen and Oleen wrote in 2013, to being standard among universities. The Directory of Open Access Repositories lists over 5,200 IRs, all types, as of December 2019, while the Registry of Open Access Repositories lists over 4,100. The growth in repositories is sufficient that a conversation has started about whether the preponderance of IRs has the potential for confusion and diminishing returns (Arlitsch & Grant, 2018). For those universities and colleges who have not yet developed an IR, or are running a small IR with only part-time staff, this means that the problem of developing workflows and staffing policies for a has largely been solved, standardized and documented in the literature by the many, many librarians who have come before. This is good news those librarians and staff at smaller universities and colleges either in the process of starting an IR or who have been slowly populating their dSPACE collection with a few items here and there as time permits. There is no shortage of literature on which to rely for wisdom and advice.

Most of the librarians interested in this chapter will be those at these smaller or mid-sized institutions. Among larger universities, institutional repositories are commonplace and workflows have already been developed and staffing solutions are in place. For a broader discussion discussing IRs at smaller academic institutions, see Wu’s article, “The future of institutional repositories at small academic institutions,” (2015).
This chapter also assumes developing workflows in an in-person campus environment. For a description of IR workflows in an online/remote environment, see Dickman’s presentation, “Institutional Repository Workflow, Project Management, and Outreach in an Remote/Online Library Environment” (2019). Development of staffing and workflow of a new or young institutional repository will likely follow a predictable path. A single librarian, whether full or part time, will be tasked with the initial development. Software will be considered and chosen and implemented. An environmental scan will be conducted to determine low-hanging fruit for easy population of the IR. Once these projects are completed and a product is in place to showcase to faculty, work will begin on larger, more ambitious projects and the development of either an Open Access policy or at least a coherent plan to capture as much scholarship as possible. Once the work for these later-phase-goals is underway a repository can be considered to be firmly in the maturing phase, and it will become necessary to redefine job responsibilities and workflows to ensure continued growth (Madsen & Oleen, 2013).

The IR at IU South Bend was started in July 2013 and marked by the hiring of the campus’ first scholarly communication librarian. At the time, IU South Bend Libraries was already using Archives of Institutional Memory and thus some of the staff were not unfamiliar with using software built on dSPACE. For the purposes of the IR, the decision was made to use the existing infrastructure of IU ScholarWorks at Indiana University Bloomington. This decision was made simple by the fact that no budget existed for subscription to a service such as Bepress. IU Bloomington has a well-established dSpace instance and was happy create a community for IU South Bend to use as its institutional repository. Thus, IU South Bend was able to piggyback off of IU Bloomington, easing our foray into the world of IRs.
For many universities, neither of these (proprietary or piggybacking) will be an option. There may not be room in the budget and at the same time no larger parent institution willing to share resources. If this the case, and the only option is to go with a free IR platform, then the development of workflows should wait until software is implemented and the staff have become familiar with it. Any system will have its quirks that reflect in workflows. Similarly, if a subscription-based service is on the table, with the accompanying tech support and added features one might expect, that will reflect in workflows as well. “Who does one go to for tech support”, for example, is a workflow question with very different answers depending on the type of IR.

**Establishing roles**

Most university institutional repositories will probably start with a single individual responsible for everything. Large, mature repositories will likely have multiple full time staff with specialized responsibilities. For the purposes of getting something off the ground, a single individual is usually where it begins. As such, the establishment of clear workflows is important, especially if that individual takes another job. A mature institutional repository will have documentation for training, policies and procedures whereas a new IR may exist for a time in a vulnerable state, many of the workflows internalized in the mind of a single individual for whom the work becomes intuitive. If that individual fails to document these workflows and then leaves, then the institution may find itself in a bind. Creating well-documented workflows and role definitions will help ensure the continued growth of the institutional repository through multiple successions.

If a single individual is tasked with starting a new or overseeing a very young IR, roles and workflows will overlap. For the repository at IU, a single scholarly
communication librarian was brought in to start a full scholarly communication program. In such a case, the role with change as the workflow progresses. The roles and workflow of a new or young IR will closely resemble the following:

- Identification of items for deposit, including copyright checking
- Gathering of items
- Deposit of items
- Metadata application
- Removal of items
- Answering complaints
- Removal and deletion of items
- Tracking usage
- Communicating value

The major challenge facing any new IR, aside from which platform to use, is that of populating the IR with items. It has been noted that the hopeful idea of “if you build it, they will come,” quoting Field of Dreams, or the Tom Sawyer-and-the-fence, “make it seem fun,” are not terribly effective strategies for large-scale utilization by faculty (Giesecke 2011, p. 531). Similarly, appeals to altruism are often met with polite agreement but no real action (Odell, 2013). Ultimately, repository staff will likely need to retrieve the bulk of deposable items themselves. The task of identification of items will initially benefit from an environmental scan to identify low-hanging fruit for deposit. At this point the IR functions more as a silo for storage than anything else and needs a population of items to begin counting usage and demonstrating impact. At IU South Bend, one of the first items identified by the environmental scan was the collection of print master’s theses in the University Archives. This decision was influenced by the work of Bruns et. al (2014) in building an institutional repository at Eastern Illinois University. Additionally, the annual IUSB Undergraduate Research Conference was identified as a ready source of items. For the theses, a student employee was tasked with
scanning the theses and for the conference posters, PDFs were obtained directly from the on-campus wide-format printing center. The initial workflows then, were very simple. The student employee would scan a thesis, the scholarly communication librarian would check the student’s work, and the scholarly communication librarian would submit the item to IU ScholarWorks.

The third step of the workflow, metadata, was assigned to the library cataloging department, a partnership with technical services. This was a natural fit and an easy sell as the number items deposited at a small campus (the average number of annual faculty publications is around 75) would not be so much as result in a burdensome amount of extra work. Within IU ScholarWorks, items are deposited at the collection level, and collections live within sub-communities or communities. The IU South Bend repository is itself a community within the IU Bloomington Community. Then, for example, student publications are a sub-community within that, then there are further sub-communities for master’s theses, the Undergraduate Research Conference, a library research award, etc. Drilling down further from theses, there are collections for each master’s program.

Within both communities and collections, specific roles must be assigned. The scholarly communication librarian was initially assigned the role of submitter and approver, but this proved redundant as it simply meant he was approving his own submissions. The roles were thus changed so that the cataloging librarians approved the submissions and applied metadata, before adding an entry to IUCAT, the Indiana University-wide online catalog, pointing to the item.

This kind of low-hanging fruit is a ready way to populate the institutional repository with few staff. The single scholarly communication librarian was only able to devote a portion of his time to the repository, given other job responsibilities. The next
population of items examined were faculty publications, for which a more detailed workflow is necessary due to additional copyright concerns. Copyright is a central consideration for depositing faculty publications into an institutional repository. The workflow will initially look similar to this:

Identify faculty publications > determine copyright permissibility > gain author permission > gather items > submit items > apply metadata > approve item > create catalog entry

Copyright permissibility is usually determined via services such as SHERPA/RoMEO (Hanlon & Ramirez, 2011). A search of a journal title in this database will bring up a list of the publisher’s policies regarding self-deposit into an institutional repository. SHERPA/RoMEO applies color grades based on the level of permissibility. Green denotes a journal that allows deposit of both pre-print and post print or the publisher’s version/PDF. Blue denotes a journal that allows for deposit of the post-print or the publisher’s version/PDF. Yellow is assigned to journals that allow for pre-print only and white is for journals that do not formally support archiving. For example, Figure 1 displays the SHERPA/RoMEO entry for College and Research Libraries:

---

1 It should be noted that many journals do not have entries on SHERPA/RoMEO. In this case an added workflow task should be to contact the journal editor directly inquiring about the journal’s policy. In many cases this has resulted in the journal giving written permission to deposit an article.
C&RL is graded a blue journal due to a prohibition on depositing the pre-print. Both post-prints and publisher’s PDF versions are allowed. The different types of copyright permissions create further workflows within the larger workflow. Take, for example, Figure 2, *Library and Information Science Research*:
Figure 2: The SHERPA/RoMEO report for *Library and Information Science Research*

In this case, the journal allows for deposit of the post-print but not the publisher’s PDF. From a workflow perspective, post-print-only permissible deposits require extra steps and more work. A publisher’s PDF is a comparatively simple matter. For the post-print, a copy must be obtained from the author. This adds steps to the workflow that are not present for PDF permissible deposits. As such, it is recommended to keep track of the SHERPA rating for each publication in an excel sheet and divide them into groups. This enables one to attack each population of documents one at a time, moving from publisher PDFs to post prints and then to preprints if so desired. At IU South Bend we benefitted greatly from the annual publication by the Department of Academic Affairs of the faculty publication list, gleaned from the faculty and instructor annual reports. This saved a great amount of time. Identifying faculty publications for deposit simply meant cross referencing the listed faculty publications list with SHERPA. However, what if no such curated list existed?

<table>
<thead>
<tr>
<th>Journal: <strong>Library and Information Science Research</strong> (ISSN: 0740-8188)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RoMEO:</strong> This is a RoMEO green journal</td>
</tr>
<tr>
<td><strong>Paid OA:</strong> A paid open access option is available for this journal.</td>
</tr>
</tbody>
</table>

| Author’s Pre-print: **✓** author can archive pre-print (ie pre-refereeing) |
| Author’s Post-print: **✓** author can archive post-print (ie final draft post-refereeing) |
| Publisher’s Version/PDF: **✗** author cannot archive publisher’s version/PDF |

**General Conditions:**
- Authors pre-print on any website, including arXiv and RePEC
- Author’s post-print on author’s personal website immediately
- Author’s post-print on open access repository after an embargo period of between 12 months and 48 months
- Permitted deposit due to Funding Body, Institutional and Governmental policy or mandate, may be required to comply with embargo periods of 12 months to 48 months
- Author’s post-print may be used to update arXiv and RePEC
- Publisher’s version/PDF cannot be used
- Must link to publisher version with DOI
- Author’s post-print must be released with a Creative Commons Attribution Non-Commercial No Derivatives License
Depending on the size of the university, contacting individual faculty members may not be an option. For those at institutions without the luxury of a curated annual faculty publication list, the workflow step of “identify faculty publications” becomes more labor intensive. Options may include Scopus affiliation searches or scanning faculty annual review reports and assembling a list from that. Another option is to contact schools directly and inquire as to internal lists. It would be advisable to start at the highest levels, contacting deans, then drill down to faculty chairs and then finally to individual faculty if need be. This could be a good way to bring subject librarians into the project by tasking them with broaching the subject of during department visits. Bringing in subject librarians to gather publication lists or cataloging librarians to aid in metadata are both good examples of moving forward the mission of the IR with existing areas of expertise. Myers et. al (2014), recommends taking this a step further and finding existing initiatives in the library with which the IR can partner. These would ideally one-and-done projects that introduce staff and faculty to the repository and not open-ended projects that individuals might be less likely to join due to workload concerns. Another option is automation: see Heller (2016) for a discussion of using automation to increase productivity in IR workflows.

Returning to the added work required of post-prints, this is generally approached by emailing the publishing faculty directly and asking for a copy. This may be met with only middling success. Faculty may have misplaced their post print or otherwise fail to provide a copy. At this juncture, some institutions are employing a strategy of de-formatting the publisher’s PDF in order to create their own post-prints. Remember that post-prints refer to post-peer-review. The article has gone through peer-review, been accepted by the reviewers, but has not yet been passed on to final editing and layout.
This strategy is problematic and I will discuss the merits and risks.

If an article has only undergone formatting changes post-peer review then it can be safely de-formatted and turned into a post-print. Even the charts and graphs can arguably be used as there is a principle in copyright that simple expression of facts are not copyrightable. For this reason, a simple chart or graph that only displays information is not protected under copyright. The problem is that it is difficult to know how much further editing went into the article post peer-review. Sometimes very little editing occurs post-peer review, sometimes a great deal of editing occurs post-peer review. The degree to which it is problematic is a function of how extensive those editorial changes were during layout. The only way to be sure is to get a copy of the post-print and compare the two, in which case one would have a copy of the post-print anyway.

Once the identification and copyright steps have been cleared, the next step is to gain author permission. At this step in the workflow the scholarly communication librarian at IU South Bend emailed each author of an identified permissible deposit article to obtain consent for upload. How many times to do so is a matter of individual preference. It is probably a good rule that after three unanswered emails the IR should move on from that individual. The labor-intensive nature of obtaining faculty consent for upload into the IR is the reason why Open Access policies or resolutions are highly recommended. There are several OA policies online that universities and colleges can adapt for their own use. For a good list visit: https://osc.hul.harvard.edu/policies/

2 However, visual displays of information that can be considered highly creative can be protected under copyright. Recall Charles Joseph Minard’s famous visualization of the strength of Napoleon’s armies during the disastrous campaign in Russia. While that piece is not under copyright by virtue the date of creation, it is sufficiently creative (even artistic) that if it had been made today it would likely be under copyright.
The OA policy adopted at IU South Bend is a “lighter” than average policy because it does not call on the author to provide a copy of their own publication to the IR staff. The publication output here is sufficiently small that we can manually retrieve articles along the way while checking the annual publication list for copyright permissibility. For a larger university such activity may be more difficult and labor intensive, particularly for R1 or R2 schools. Whatever the size of the school, though, the presence of an OA policy simplifies the workflow by removing the requirement of faculty consent for deposit. The population of guaranteed deposits into the IR expands to include, as a given, every article for which publisher PDF deposit is permissible.

Which brings us to deposit. At IU South Bend all deposits were initially done by the scholarly communication librarian. After several years a student was trained to deposit and given privileges in IU ScholarWorks. These sort of repetitive tasks are not a good use of administrator time. When considering such “grunt work,” consider training hourly employees to do so and then checking over the quality of the work later. Application of metadata, the next step, is not simple grunt work and best left to professional librarians and staff. The head of cataloging at IU South Bend, Scott Opasik, authored the following chapter: “Metadata for Institutional repositories.” Please see that chapter for a detailed discussion of this step in the workflow. Following metadata the only tasks are checking over the final item one last time and creating a catalog entry.

There are two roles described above that are not present in the deposit workflow: complaints and removal of items. The OA policy we adopted at IU South Bend is an opt-out policy. A faculty member or student can request their article or thesis not be included.

---
3 The IU South Bend OA policy may be found at: https://library.iusb.edu/research-support/schpub/OAPolicy.pdf
in the IR and we will remove it. This role, once defined along with the circumstances under which an item may be removed, is straightforward. The final two roles mentioned above, documenting use and communicating value, now come into play. At this stage, once workflows and roles have been established for ready population of faculty and publications, the IR has moved into adolescence. The IR still serves as a silo designed to aggregate and provide access to finished scholarship or as-close-to-finished scholarship as the IR staff can find. A product has been created that can be showcased to faculty and university administration to justify further support (think, making the case for an OA policy).

The assessment role is, depending on the IR software, a cinch or a labor-intensive bore. Bepress features automatic download reports and users can see a world map showing downloads per country. IU ScholarWorks has no such automatic reporting feature. We had the option of collecting data manually by clicking through each item in the IR and viewing individual usage statistics or sending requests or bothering the administrators down at IU Bloomington. Either way, documented downloads were an important component in communicating value of the IR to faculty in the years leading up to bringing an OA policy before the faculty senate. This is another way to bring in the subject librarians – once a workflow has been developed for gathering usage data on a regular basis, the subject librarians can take that message to their respective departments as they visit. For a detailed take on marketing and outreach, please refer back to Chapter two, Communicating Value and Building Relationships: Marketing and Outreach for Institutional Repositories, by Harrison Inefuku, Iowa State.

In summation, then, the initial workflow for a new or recently-established Institutional Repository will likely contain the following steps:
Step 1: Identify faculty publications that can be deposited. In this workflow step, faculty publication lists are generated and relevant student scholarship such as ETDs are identified as targets for population.

Step 2: Determine copyright permissibility. In this step, publisher deposit polices are researched and the publications coded for deposit (publisher’s PDF, post-print, pre-print)

Step 3: Gain author permission: In this step the authors are either directly contacted to seek permission else an Open Access policy or resolution is in place granting automatic permission unless opted out. For a full discussion of copyright issues here, see Chapter 7, “Copyright and Institutional Repositories,” by Benjamin Keele.

Step 4: Gather items for submission. Items are either produced via digitization, gathered manually by staff or (much less frequently) supplied directly by the publishing faculty themselves. Items are separated into different folders according to type permissible for deposit.

  4A. Publisher PDF-permissible can be immediately uploaded
  4B. Post-prints should be solicited from faculty, though some institutions are undertaking deformatting (see above).
  4C. Pre-prints, also solicited from faculty.

Step 5: Apply metadata: Either in partnership with cataloging or undertaken by a designated individual in the department, appropriate metadata is applied to help ensure better discoverability via the web.

Step 6: Approve item: Check over the item for any mistakes.

Step 7: Create a catalog entry pointing to the new item.
Step 8: Develop workflow for continual assessment of item usage. Initially this will take the form of hits and downloads. It is also advisable to select a population of articles to monitor long-term for citations to document impact.

Step 9: Outreach and promotion of product. Using assessment data, undertake regular outreach to promote the IR and promote future use of the IR by faculty. This leads back into Step 1.

The first part of this chapter described the development of staffing and workflow during the founding and initial growth of and institutional repository and into the adolescent phase as a functioning silo of scholarship. As an IR moves out of adolescence, staffing and workflow will need to evolve to allow for continued growth. The silo model is a flat model. Once the workflows are in place to capture end-product faculty and student publications, running the IR becomes a job of maintenance. Moving into data curation takes the library closer to facilitation of scholarship and a step nearer what Clifford Lynch describes as the nature of the mature repository:

I believe that a mature and fully realized institutional repository will contain the intellectual works of faculty and students – both research and teaching materials – and also documentation of the activities of the institution itself in the form of records and events and performance and of the ongoing intellectual life of the institution. It will also house experimental and observational data captured by members of the institution that support their scholarly activities.4

Looking at this definition, Madsen & Oleen (2013) set about approaching workflow and staffing for a maturing repository. This process included moving from a single-individual overseeing a light stream of deposits to defining job responsibilities of a

---

well-staffed IR. They conceived for four distinct categories of job responsibilities including collection development (determining appropriate content), promotion, pre-processing (publisher policies, obtaining items, handing off to metadata staff) and metadata creation (Masden & Olsen, 2013). To accommodate growth for a larger staff, the pair envisioned three primary areas of development:

“Creating folders and subfolders … to store manuscripts, procedures, sample letters, and work product related to metadata creation.

Refining a local wiki for publishers’ policies so that information was clearly formatted for easy interpretation.

Creating a workflow management system (WMS to allow easy sharing of responsibilities as an article moved through the processes from identifying an item to…final deposit” (Madsen & Oleen, 2013, p 5).

The last point has been of particular relevance to the work at IU South Bend as the scholarly communication librarian accepted a job offer elsewhere. Having started and managed the repository and other scholarly communication initiatives since July 2013, a number of workflows had been developed with which the rest of the staff was unfamiliar. As such, the scholarly communication librarian needed to create a number of detailed workflows documents from scratch to ensure that work could continue through the search for a replacement and into the beginning of that individual’s tenure. This work was coupled with workshops for various faculty and staff who agreed to take on some of the departing librarian’s responsibility. It is recommended that creation and upkeep of workflow and staffing documents become a priority for any IR.

It is also important to remember that as your repository grows and matures, different workflows will need to be created for different collections and communities of users. Depending on user needs, copyright restrictions or collection aims, workflows
may have fewer, more, or simply different steps (SPARC, 2002, p. 31). For example, if a university has an article processing charge fund, extra steps may be involved in terms of deposit permissibility (OAWAL, 2020). Likewise, funder mandates should be considered, both in terms of outreach and in the actual workflow of deposit (ibid.). The danger here is that workflows may become compartmentalized. If workflows begin to differ greatly and different systems or software are employed then one group of staff may not have the knowledge to help another group of staff on certain projects. To address this, Matthies (2011) recommends distributed workflows in which staff and student employees are all trained on common tasks, software and interfaces, enabling the project manager to assign editors as needed.


Bruns, Todd A.; Knight-Davis, Stacey; Corrigan, Ellen K.; and Brantley, J. Steve, "It Takes a Library: Growing a Robust Institutional Repository in Two Years" (2014).Faculty Research & Creative Activity. 98. http://thekeep.eiu.edu/lib_fa


OAWAL, The Effect of Gold on Workflows and Staffing. (2020) Retrieved from:


Odell, Jere. “Building, growing and maintaining repositories” (2014). [Conference presentation delivered at Michiana Scholarly Communication Librarianship Conference]