Putting the IR in RIMS: Towards an Automated Integration Between Institutional Repositories and
Research Intelligence Systems.

Jamie Wittenberg

Indiana University, Bloomington

Introduction

Institutional repositories are at a turning point. There have been several public and contentious assertions
that the institutional repository (IR) is dead[1], but it is more accurate to say that the IR may not continue to
exist in the way that we currently conceive of it. In 2017, Ellen Catz Ramsey, Director of Scholarly
Services, wrote a blog post addressing why UVA launched a new repository at a time when the value of
institutional repositories was being questioned, even by those who initially supported them. She wrote
"As an option for authors whose disciplines are not congregating around an international discipline-based
archive, or whose work doesn't (yet) fit existing scholarly archives, every good research institution will
always need the safe haven of a local repository...Put it in the IR, poof, it's in the library's catalog,
Google Scholar, and has a persistent link you can cite." However, Ramsey also writes that IRs have not
served their function as clearinghouses for research at an institution. In contrast, Novak and Day at the
University of Nevada, Las Vegas assert that "After reading the literature and a self-examination of our
repository situation, we believe a new role exists for the IR, a research administrative one" (2018). This
contentious, contested new role is inexorably tied to the rise of research information management
products in the higher education sector. Libraries have historically collected and analyzed publication data
in order to improve services and collections. This data has taken on new significance in the age of data-
driven university administration. Publication quantity, venue, and citation counts are often used as a
proxy for measuring the impact of research. Thus, publication data enables universities to assess research
impact, productivity, co-authorship with other institutions, etc. Furthermore, it allows systems that
integrate publication data with funding data to mine publications for keywords that can be matched to
grants and, ideally, assist research administration offices in suggesting appropriate funding opportunities
to faculty authors. Commercial entities have developed sophisticated software that links faculty
biographical data with data on past grant and award activity, publications, co-authorship, and more. The

---

[1] Eric Van de Velde wrote in a 2016 blog post "With the IR at a dead end, Green OA must pivot towards alternatives that have viable paths
forward: personal repositories, disciplinary repositories, social networks, and innovative combinations of all three"
http://scitechsociety.blogspot.com/2016/07/let-ir-rip.html. See also: Poynder, R. (2016, September 22). "Open and Shut?: Q&A with CNI's
Clifford Lynch: Time to re-think the institutional repository?" https://poynder.blogspot.com/2016/09/q-with-cnis-clifford-lynch-time-to-
re_22.html

collection and monetization of this data on research activity makes up a lucrative research intelligence market.

It is no secret that academic publishers are making headway into the research intelligence market, and it is a logical progression to then develop or acquire faculty activity reporting systems. These systems, often referred to as research profiling systems, research information management systems (RIMS), or current research information systems (CRIS), are systems that collect and manage data about research activity.[2] Elsevier, for example, announced in 2016 that they were rolling out Faculty and Academic Activity Reporting functionality in Pure, their RIMS which enables administrators to track faculty research activity by integrating faculty profile, funding, and publication data ("Pure Faculty Activity Reporting: Making data-based strategic decisions," 2016). Shortly following Elsevier's reporting tool, Bryant et. al. noted that "RIM adoption [is] growing in countries without strong national reporting mandates, driven by reasons other than compliance, such as improved decision support and improved researcher services" (2017). Commercial ownership of preprint servers and institutional repositories (SSRN, Bepress, Esploro) coupled with the rise of RIMS and their consolidation with faculty profile, reporting, and funding operations systems strategically targets research administration and compliance offices as new customers for commercial publishers. As much as these developments serve commercial publishing interests (and provide potentially chilling data to school administrators), this adoption and consolidation also presents new opportunities for libraries in automating the repository population process and positioning library repositories as central services in the university's research ecosystem. Libraries are uniquely poised to lead development of community-owned infrastructure that challenges commercial publishers' strategic consolidation of the apparatus of research and scholarly communication (Schonfeld, 2017).

Indiana University Case Study

At Indiana University, the campus' response to the implementation of a new open access policy in 2017 created an opportunity for the University Libraries to develop an in-house RIMS built to facilitate the automated deposit of green open access content into an institutional repository. In 2016, roughly one year before the policy was passed, a new annual reporting system was implemented for the university. This reporting system, Digital Measures Activity Insights (DMAI), is used across the university to standardize faculty annual reports. Each year on January 15th, all university faculty report on their teaching, research, and service activity. As part of the 2017 implementation of a campus Open Access Policy, the University

---

[2] There are numerous research profile systems that have been adopted by universities. One resource for comparison of these systems is the "Comparison of research networking tools and research profiling systems" Wikipedia article, available here: https://en.wikipedia.org/wiki/Comparison_of_research_networking_tools_and_research_profiling_systems

Libraries established a workflow with the DMAI team, administered by the Office of the Vice Provost for Faculty and Academic Affairs. As part of this workflow, the Libraries run a report each year after the annual reporting deadline that results in a CSV file containing citations for every article reported by a faculty member as accepted for publication or published during the last calendar year. In 2018, the IU Libraries Scholarly Communication Department processed 1,591 unique citations with IU Bloomington authors, and in 2019, they processed 2,193. To manage these records, the IU Libraries developed a tool: Bloomington Research Information Tracking Engine (BRITE). The BRITE tool is designed to support librarians in processing faculty annual report citations for articles subject to the Open Access Policy in order to ingest metadata and assets for those citations into the institutional repository IUScholarWorks Open. This repository is one of three managed by the Scholarly Communication department and uses Tind, a cloud-hosted spin-off of open-source CERN software. In order to ingest records, the Scholarly Communication Department must first determine which version of the article is acceptable for deposit, procure the article, then ingest the article and metadata into the repository. This workflow leverages several APIs to automate as much of the process as possible.

Metadata Enhancement

The Crossref API is used to add crucial attributes to incomplete records. Because citations are often user-entered, information may be incomplete, incorrect, or out of order. The Crossref integration enables the Scholarly Communication department to copy accurate metadata to an incomplete record. This is particularly important in ensuring correct author order, ISSN, and DOI for each record.

Open Access Check

When an author has already published an open access version of their article, IUScholarWorks Open points to the open access version of record, rather than archiving the asset. In order to determine whether an open access version is (or will be) available, BRITE integrates with three APIs: Unpaywall, Directory of Open Access Journals, and PubMed. The Unpaywall API harvests metadata from over 50,000 publishers and returns information to BRITE based on whether the article is open access. The Directory of Open Access Journals, a registry of open access journals, returns information to BRITE based on whether the journal is open access. The PubMed API is used to determine whether an article is in PubMedCentral, indicating that it was funded by NIH and subject to their public access requirements. It returns information on PubMedCentral ID to BRITE.

Rights Check and Content Recruitment

In 2018 and 2019, about one-third of IU-authored articles have had an open access version available on the web. For the articles that do not have an open access version available, BRITE leverages the SHERPA/RoMEO API to determine the most permissive version of an article that can be shared in an institutional repository. The record is then tagged with the most permissive version (unknown, preprint, postprint, or offprint) that can be deposited. Depending upon which version can be shared, an email template is sent to the IU author requesting that version for deposit. The email address, author name, journal title, and article titles are pre-populated by BRITE and emails are sent through the BRITE tool.

Deposit

Once authors respond to requests for their final versions, assets are stored in a Box directory using a naming convention that relates each asset to its associated record by way of a unique ID. Metadata-only records, which point to articles previously published openly, and records with these assets are then batch deposited into IUScholarWorks Open, mapping the fields used in DMAI and BRITE to the MARCXML used in the repository. Once the repository is populated, each record is issued another unique ID within the repository. These IDs are then harvested and appended to each record in BRITE. This ensures that every record in BRITE links to a record in the institutional repository. It also enables the Scholarly Communication department to send a final confirmation email to each author, providing them with a link to their repository record.

Conclusion

The modern institutional repository is a local effort but also a collaborative one, integrating with data sources stewarded by institutional and national stakeholders. It serves as a clearinghouse for published and unpublished research output - a site of preservation and of records management, a local solution for persistent access to institutionally-authored work and a research information management system. It offers a locally-provided solution, often designed to support specific institutional needs, without giving away institutionally-authored work to commercial publishers. By leveraging faculty annual report citation data to populate an institutional repository, the workflow used by Indiana University models new possibilities for institutional repositories as clearinghouses of research output. While this model is not fully automated - it does still require quality control, intervention, escalation of complex cases, and record management - it provides a blueprint for the development of a repository that is part digital asset management system, part records management system, part research information management system. This workflow was designed to address a specific institutional repository use case, but resulted in the development of a RIMS that stores a complete record of every reported publication on the Bloomington

campus. Because this workflow is not dependent upon commercial software, it offers several advantages to institutions:

1.   It gives campuses local control over quality control when disseminating metadata for faculty-authored articles.

2.   It enables institutions to be flexible in selecting or de-selecting research administration workflow components without disrupting the entire administration and compliance ecosystem.

3.   It lessens the likelihood that faculty-entered data in RIMS that is ingested into public-facing profile systems will be monetized by commercial entities.

4.   It populates the institutional repository and provides a single record for research output on campus while broadening access to faculty-authored research.

Bryant et. al. wrote in a comprehensive 2017 report on RIMS that "While RIM systems and IRs overlap in functionality, there are characteristic differences. The main purpose of collecting publications as part of RIM is to collect and validate institutional research outputs…the main purpose of an IR is to facilitate open access and reuse of publications..." (p. 52). Although these systems have traditionally run on separate platforms and served different institutional priorities, a future-facing institutional repository is likely to be at least minimally integrated with a RIMS and it is conceivable that going forward there will be wider adoption of systems that offer fully integrated RIMS/repository platforms. There are new technologies and opportunities for a modern IR that revitalize its role and broaden its value at an institution. These products are already emerging in the commercial sector – it remains to be seen whether locally-owned, open alternatives that prioritize the critical scholarly communication role of libraries will surface.

References

Bryant, B., Clements, A., Feltes, C., Groenewegen, D., Huggard, S., Mercer, H., Wright, J. (2017). Research Information Management: Defining RIM and the Library's Role. Retrieved June 21, 2018, from https://www.oclc.org/research/publications/2017/oclcresearch-defining-rim.html

Novak, J., & Day, A. (2018). The IR Has Two Faces: Positioning Institutional Repositories for Success. *New Review of Academic Librarianship*, *24*(2), 157–174. https://doi.org/10.1080/13614533.2018.1425887

Pure Faculty Activity Reporting: Making data-based strategic decisions. (2016). Retrieved August 19, 2019, from https://www.elsevier.com/research-intelligence/campaigns/AIR-registration-june-2016

Ramsey, Ellen. (2017, 11). We Launched a New Institutional Repository. Here's why. Retrieved from The Taper: Copyright and Information Policy at the UVA Library: http://thetaper.library.virginia.edu/2017/08/11/we-launched-a-new-institutional-repository-here-s-why.html

Schonfeld, Roger. (2017). Strategy & Integration Among Workflow Providers. Retrieved August 19, 2019, from The Scholarly Kitchen: https://scholarlykitchen.sspnet.org/2017/11/07/strategy-integration-workflow-providers/