

Leveraging Open Services to Enhance Institutional Research Tracking Workflows

Yasmeen Alsaedi, King Abdullah University of Science and Technology (KAUST), yasmeen.alsaedi@kaust.edu.sa; Daryl Grenz, King Abdullah University of Science and Technology (KAUST), daryl.grenz@kaust.edu.sa; Mohamed Baessa, King Abdullah University of Science and Technology (KAUST), mohamed.baessa@kaust.edu.sa

Abstract

In order to support the implementation of an institutional open access policy, adopted in 2014, the KAUST University Library developed publications tracking processes with the additional aim of making the institutional repository a reliable source for bibliographic information about the university's research outputs.² The code underlying these processes has now been updated and combined into an open source software application, the Institutional Research Tracking Service (IRTS), with a public version planned for release at <https://github.com/kaust-library/irts> prior to the conference. This presentation will briefly introduce the structure and functionality of the IRTS application and walk through the different workflows it supports, before focusing on recent enhancements to the service, especially our use of the beta version of the new Sherpa Romeo API⁶ and of the Unpaywall API¹¹. We will also present an assessment of the initial effect of these improvements on the workflow, and the impact they have had on deposit rates of full text materials to the repository. Finally, we will give examples of how this service reinforces the interest of other university stakeholders in reusing the research information from the repository for purposes such as annual reporting, research evaluation, and maintenance of up-to-date publication lists on websites.

Keywords

Institutional research tracking, open access deposit rates, Unpaywall, Sherpa Romeo

Audience

Repository managers, developers and librarians.

Proposal

In national or regional contexts where maintenance of a university bibliography is necessary for assessment and funding exercises, many institutions have established systems to track the university's research outputs.⁴ However, where such external motivations are weaker and institutional tracking is instead implemented largely with the goal of improving implementation of an institutional open access policy, libraries have developed an array of approaches^{3,7,8,9,10}, that often prove difficult to sustain.

We are releasing the Institutional Research Tracking Service (IRTS) code as open source software in hopes that it will be useful to other libraries seeking to make their institutional repository a more comprehensive and up-to-date resource about research outputs affiliated to or funded by their institution, while improving open access deposit rates. The IRTS combines information harvested from multiple sources and feeds it into a library-mediated metadata review and file deposit process that minimizes the burden on researchers. It currently supports integration with repositories using DSpace with the REST API enabled. Patching the DSpace REST API¹ is required if the functionality supporting transfer of ORCID iDs in the item metadata will be used.

We have integrated the Unpaywall service into both the daily review process for ingesting new publications and into a periodic review process that checks whether an item for which we have a metadata only record has any new open access versions since we last checked. One unexpected benefit of using Unpaywall has been better identification of existing arXiv preprint records, especially when there is some variation in title between the two versions, helping us avoid issues with duplicate record creation. The presentation will

report on the results of using Unpaywall, especially the results of a full repository review for all metadata only items in our repository, including discussion of the remaining weaknesses we have found in the Unpaywall data in terms of file and version identification.

The decision to move forward with using the beta version of the new Sherpa Romeo API was driven by the need for machine-readable metadata that would simplify the permissions step of the deposit workflow, especially the assignment of embargos. While we also use the ScienceDirect Article Hosting Permission API⁵ to assign embargos for Elsevier publications, it does not always immediately have embargo information for a publication that has just been accepted, which is when we are processing the deposit and need the information, and in any case is limited to publications from only one publisher. Upon learning about the Orpheus tool developed by Cambridge⁷, we also considered integrating it into our service, but chose to focus on switching to the updated Sherpa Romeo service instead. The presentation will demonstrate how Sherpa Romeo's new data format has improved our embargo handling process, as well as what we see as remaining areas for improvement in the Sherpa Romeo data quality and coverage.

Time allowing, we will also walk briefly through how we use IRTS at KAUST to collect information about research output types, such as patents, datasets and software, that may not typically be included in research tracking workflows developed around open access policy implementation. In conclusion, we will then discuss examples of how research information in the repository is being reused within our university, and how feedback from other stakeholders has been taken into account in the development of both IRTS and our repository service as a whole.

References

1. Atmire. (2019). Expanded ORCID support for DSpace (Patch). *GitHub* [Software]. Available at: <https://atmire.github.io/expanded-ORCID-support>
2. Baessa, M. A., Grenz, D. M., & Wang, H. (2016). Towards a Comprehensive and Up-To-Date Institutional Repository: Development of a Publications Tracking Process. *Open Repositories 2016* [Presentation]. Available at: <http://hdl.handle.net/10754/615855>
3. Bjork, K., Cummings-Sauls, R. & Otto, R. (2019). Opening Up Open Access Institutional Repositories to Demonstrate Value: Two Universities' Pilots on Including Metadata-Only Records. *Journal of Librarianship and Scholarly Communication*, 7(1). Available at: <http://doi.org/10.7710/2162-3309.2220>
4. Bryant, Rebecca, Anna Clements, Pablo de Castro, Joanne Cantrell, Annette Dortmund, Jan Fransen, Peggy Gallagher, & Michele Mennielli. (2018). Practices and Patterns in Research Information Management: Findings from a Global Survey. Dublin, OH: OCLC Research. Available at: <https://doi.org/10.25333/BGFG-D241>
5. Elsevier. (no date). Article Hosting Permission API. Available at: <https://dev.elsevier.com/documentation/ArticlePermissionsAPI.wadl>
6. Field, Adam. (2019). Show and Tell: SHERPA Romeo's New API. *Open Repositories 2019* [Presentation]. Available at: <http://doi.org/10.5281/zenodo.3553819>
7. Sartori, André Fernando. (2019). Automating repository workflows with Orpheus, an Open Source database of journals and publishers. *Open Repositories 2019* [Presentation]. Available at: <http://doi.org/10.5281/zenodo.3553971>
8. Sergiadis, A. D. (2019). Evaluating Zotero, SHERPA/RoMEO, and Unpaywall in an Institutional Repository Workflow. *Journal of Electronic Resources Librarianship*. Available at: <https://doi.org/10.1080/1941126X.2019.1635396>
9. Smart, R. (2019) What Is an Institutional Repository to Do? Implementing Open Access Harvesting Workflows. *Publications*, 7(37). Available at: <https://doi.org/10.3390/publications7020037>
10. VanTuyl, Steve, Zhang, Hui, Gum, Josh, & Imholt, Sarah. (2019). Towards Easy Deposit: Lowering the Barriers of Green Open Access with Data Integration and Workflow Automation for an Institutional Repository. *Open Repositories 2019* [Presentation]. Available at: <http://doi.org/10.5281/zenodo.3553829>
11. Unpaywall. (2019). REST API. Available at: <https://unpaywall.org/products/api>