Free and Open-Source Automated Open Access Preprint Harvesting

Jack E. Peplinski, Joanne Paterson, Courtney Waugh & Joshua M. Pearce*

This article underwent semi-anonymous peer review in accordance with JLSC's peer review policy.

Free and Open-Source Automated Open Access Preprint Harvesting

Jack E. Peplinski
Department of Electrical & Computer Engineering, Western University

Joanne Paterson
Courtney Waugh
Western Libraries, Western University

Joshua M. Pearce[*]
Department of Electrical & Computer Engineering, Western University; Ivey Business School, Western University

ABSTRACT

Universities are attempting to ensure that all of their research is publicly accessible because of funding mandates. Many universities have established campus open access (OA) repositories but are struggling with how to upload millions of manuscripts under numerous license agreements while also linking metadata to make them discoverable. To do this manually requires around 15 minutes per manuscript from an experienced librarian. The time and cost to do this campus-wide is prohibitive. To radically reduce the time and costs of this process and to harvest all past work, this article reports on the development and testing of a free and open source (FOSS) JavaScript-based application, *aperta-accessum*, which does the following: 1) harvests names and emails from a department's faculty webpage; 2) identifies scholars' Open Researcher and Contributor IDentifiers (ORCID iDs); 3) obtains digital object identifiers (DOIs) of publications for each scholar; 4) checks for existing copies in an institution's OA repository; 5) identifies the legal opportunities to provide OA versions of all of the articles not already in the OA repository; 6) sends authors emails requesting a simple upload of author manuscripts; and 7) adds link-harvested metadata from DOIs with uploaded preprints into a bepress repository; the code can be modified for additional repositories. The results of this study show that, in the administrative time needed to make a single document OA manually, *aperta-accessum* can process approximately five entire departments worth of peer-reviewed articles. Following best practices discussed, it is clear that this open-source OA harvester enables institutional library's stewardship of OA knowledge on a mass scale for radically reduced costs.

INTRODUCTION

In order for science to progress optimally, it is understood that scientific knowledge must be commonly owned (Sismondo, 2010, p. 24); however, scientific progress is impeded by restricting access to copyrighted scientific literature without payment (Gibbons, 1994; Heise & Pearce, 2020). This collateral damage of intellectual property concept (Boldrin & Levine, 2008) has the unintended consequences of copyright laws and paywalls restricting access to the scientific peer-reviewed literature (Lewis, 2012) to the point that even wealthy Harvard University is challenged to pay for it (Sample, 2012). This has divided scientists into the "haves and have nots" (Chagas, 2018). Fortunately, the open access (OA) movement is flourishing, making up at least 28% of the literature (Piwowar et al., 2017), and is moving toward the point that the peer-reviewed literature could be universally accessible (Johnston, 2008; Joseph, 2013; Liesegang, 2013). Lewis even argues that OA is inevitable based on the empirical growth rates (2012). The benefits of OA are well established in the literature and include the following: 1) the pragmatic advantage of, by being freely and easily on the Internet, an author's work is available to the widest possible audience (Lewis, 2012); 2) increased citation rates for OA publishing (Antelman, 2004; Harnad & Brody, 2004; Hajjem et al., 2005; Eysenbach, 2006; Swan, 2010; Niyazov et al., 2016); 3) OA provides a means of access to relevant literature for making significant advancements in knowledge (Boote & Beile, 2005; Webster & Watson, 2002); and 4) both increased efficiency and effectiveness of science (Partha & David, 1994). Poynder summarizes "…it is no longer rational, or even necessary, for subscription paywalls to be built between researchers and research" (2011).

Not surprisingly, a growing list of funders demand OA for studies that they finance. There is a particularly strong case for OA for public funding of science (Suber, 2003; Suber, 2012; Heise & Pearce, 2020) based on the simple idea that, if the public funds research, the public should at the very least be able to read it. Globally, 87 major funders and 57 funder-research organizations already demand OA of work that they fund (ROARM, 2021). Finally, over 850 universities and research organizations have also mandated that researchers share their work OA (ROARM, 2021). As universities navigate how to transition to OA, they are developing OA policies and then moving to support their researchers in making OA possible. A popular method in some areas has been to finance gold OA fees for a single publisher (e.g., Sweden's deal with Springer/Nature publishing group that allows all Swedish academics to publish their work OA for free in the entire publishing line, which increases OA and simplifies workflows for researchers but risks conserving the high costs associated with the status quo) (Olsson et al., 2020). For universities in nations that do not offer such a program, individual universities have also experimented with funding article processing charges (APCs) for OA for their faculty members (Gyore et al., 2015). In contrast, the most-common method to provide OA for universities has been to develop university-specific institutional OA

repositories (Pinfield et al., 2014; Liauw & Genoni, 2017) and encourage (or require) that the faculty members deposit their work there. This method supports green OA, which refers to the growing self-archiving (Swan & Brown, 2005) of a version of the article (normally not the final published version) on an institutional repository (IR). For example, Western University is considering a strong OA policy (Western Libraries, 2021) that would grant the university non-exclusive permission to archive and disseminate articles via Western's IR. Under the policy, university community members agree to publish in OA publications and/or deposit scholarly work in Western's OA IR, Scholarship@Western, or in a disciplinary repository such as arXiv as early as possible, ideally sometime between the date of acceptance and the date of publication. This will obviously be a boon for making Western University's scholarship more accessible and help accelerate science worldwide, as those without library subscriptions must currently pay about $35 per article for those behind paywalls. Providing free access to all of a university's scholarship this way (although no additional funds go to the publishers) does, however, involve an enormous amount of work with associated internal costs for the millions of articles. To radically reduce the time and costs of this process and harvest all past work, this article reports on the development and testing of a free and open-source (FOSS) JavaScript-based application *aperta-accessum*, which does the following: 1) harvests names and emails from an academic department's faculty webpage; 2) identifies the scholars' Open Researcher and Contributor IDentifiers (ORCID iDs); 3) obtains the full list of the digital object identifiers (DOIs) of scholarly publications for each scholar; 4) determines whether the articles already exist in an institution's OA database; 5) identifies the legal opportunities to provide OA versions of all of the articles not already in the OA database; 6) sends authors an email requesting a simple upload of the author's manuscripts; and 7) adds link-harvested metadata from the individual DOI and an uploaded preprint into a bepress repository; the code can be modified for additional repositories. The time savings that *aperta-accessum* provides librarians in facilitating making articles OA in bulk are quantified, best practices are reviewed, and the ability of this OA harvester is discussed in the context of the future of institutional library's stewardship of OA knowledge.

There is a widespread concern among professors (Beaubien & Eckard, 2014) that requiring OA would be onerous because publication in high-impact journals is an important component of demonstrating expertise for grants, tenure, and promotion, and many OA journals, because they are, in general, newer, carry lower impact factor scores (Pearce, 2022). This challenge can be overcome in two ways. First, nearly all publishers and journals allow preprint/accepted manuscript posting in traditional subscription journals. Second, authors can pay APCs to OA journals or OA fees to hybrid journals (e.g., subscription journals that allow authors to pay for OA for a specific article). Over 17,000 journals offer a means of OA, and over 12,000 have no APCs (Directory of Open Access Journals, 2021). To recruit faculty participation, which is critical for the

success of even this automated process, care must be taken (Otto, 2016) with the necessary marketing needed to make faculty aware of OA in general (Colla, 2020; Kakai, 2021); this initiative in particular is important because of the risks of filling faculty inboxes with preprint requests.

METHODS

Open-source software design and operation

JavaScript was used to write the scripts for *aperta-accessum*, which is released under the GNU General Public License (GPL) version 3 (GNU, 2007) and is available on GitHub (https://github.com/jackpeplinski/aperta-accessum). A registry of this article's version is available on the Open Science Framework (https://doi.org/10.17605/OSF.IO/7MECZ). These scripts were written to be run using Node.js. The upload site for *aperta*-accessum was created primarily using a variety of JavaScript libraries and frameworks, including React, Material-UI, Emotion, and React Dropzone. A potential workflow using *aperta-accessum* is shown in Figure 1.

The administrator is the person responsible for the maintenance of the IR. In Figure 1, a bepress repository was the IR used.

Potential workflow description

Stage 1 of *aperta-accessum*: administrator article identification. This stage provides the administrator with a comma-separated values (CSV) file of article titles; DOIs; and authors' first names, last names, and email addresses. This CSV can be used to send emails to prompt researchers to upload their articles to the IR. The administrator can begin the workflow by executing the command "node sendEmail.js" from the command line. This command will run the sendEmail.js script.

The "getPeople(scrapeURL)" function is executed first. This function scrapes emails and first and last names from a specified URL (e.g., a faculty directory page).

The "getORCIDID(fName, lName, institution)" function uses these scraped first and last names to get the ORCID iDs of all authors with the same first and last name at the specified institution from the ORCID database. These parameters were recommended by Western librarians as sufficient to determine whether the works of the ORCID iD should be included in the university's IR.
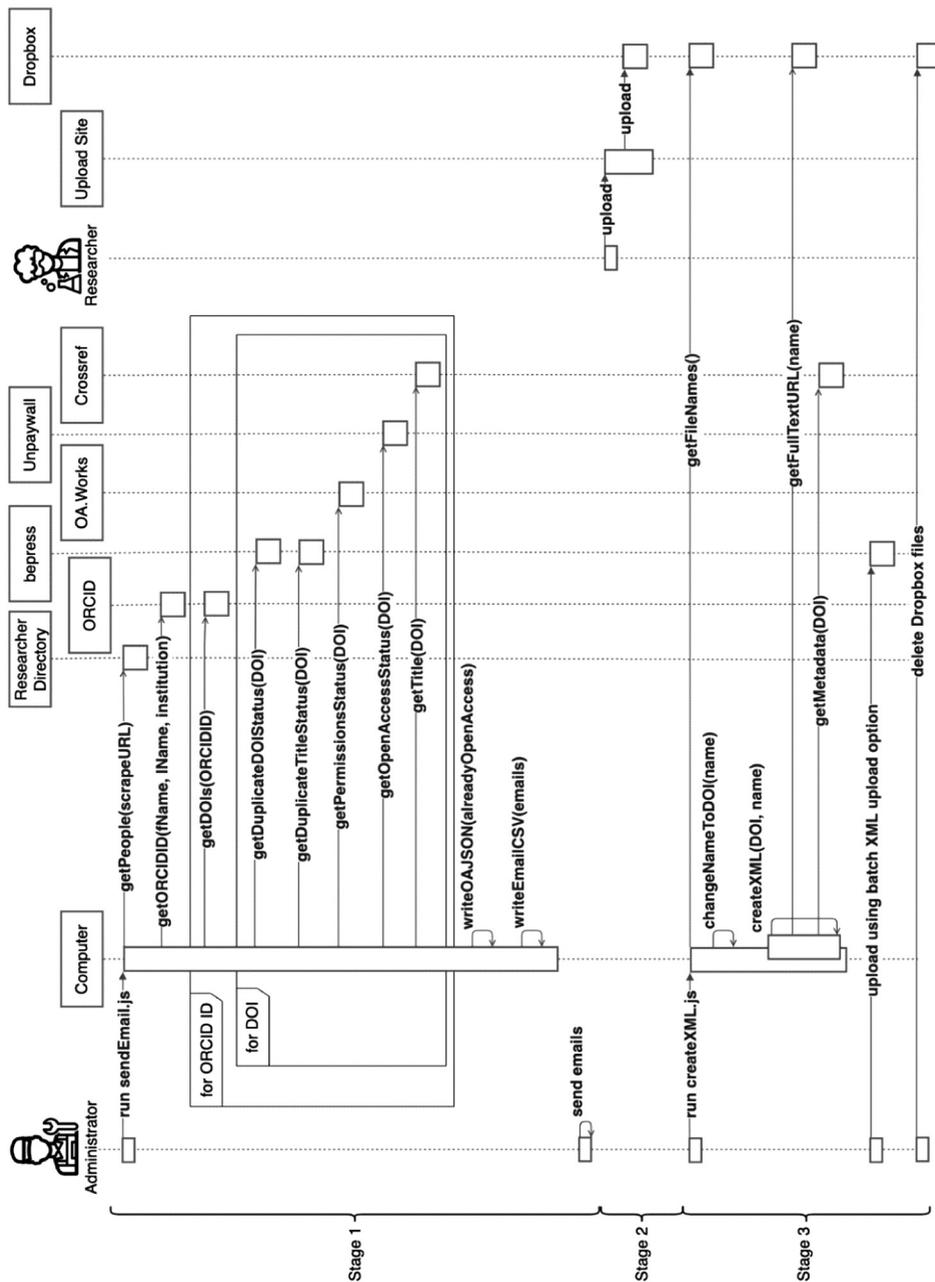
Figure 1. A potential workflow using *aperta-accessum*.

For each ORCID iD, the "getDOIs(ORCIDID)" function is executed. This function gets the DOIs for works listed for the ORCID iD from the ORCID database.

For each DOI, the "getDuplicateDOIStatus(DOI)," "getDuplicateTitleStatus(DOI)," "getPermissionsStatus(DOI)," "getOpenAccessStatus(DOI)," and "getTitle(DOI)" functions are executed. These functions use application programming interfaces (APIs); the bepress API is used for the first two function calls and then OA. Works, Unpaywall, and Crossref, respectively, as shown in Figure 1, were used to determine whether the article is already present in the IR, whether the article is able to be OA, and whether the article is already OA, as well as to get the title of the article.

If the article is not already present in the IR but is eligible to be OA, and if the title is available, an entry in the CSV is created. If these previous conditions are met and the article is already OA, the DOI and URL where the file is available are added to a JSON (JavaScript Object Notation) file. JSON is a lightweight data-interchange format, which is both easy for humans to read and write as well as easy for machines to parse and generate.

The administrator can then use the CSV to email professors to prompt them to upload their articles to the IR. This can be done using Outlook's mail merge feature, for example.

Stage 2 of *aperta-accessum*: professor submission of articles. Once the professor has received an email, they will click the custom URL and be directed to a webpage. The webpage is included in the *aperta-accessum* repository and needs to be deployed only once; further instructions are included in the Github Readme. The webpage will display the DOI and title of the article; the professor will be prompted to upload the article and click a "submit" button.

When the submit button is clicked, the file is automatically uploaded to a specified Dropbox folder. The Dropbox folder needs to be configured only once; further instructions are included in the Github Readme. The name of the file is the article's DOI, modified to comply with Dropbox's file naming rules. For bepress, uploading the file to a cloud service such as Dropbox is required because bepress requires a public URL ending in ".pdf."

Stage 3 of *aperta-accessum*: administrator database inclusion. The administrator should execute the command "node createXML.js" using the command line when either their Dropbox storage is reaching capacity or few new uploads are being made. This command will run the createXML.js script and create an XML file, upload.xml,

that can be uploaded to bepress, which will enter the articles in Dropbox into the bepress database.

The "getFileNames()" function is executed first. This function gets all of the names of files in the specified Dropbox folder. The names of these files are the article's DOIs, which are modified to comply with Dropbox's file naming rules.

The "changeNameToDOI(name)" function changes the file's name from the modified Dropbox form back to the proper DOI.

The "createXML(DOI, name)" function uses this proper DOI and name to create an XML file. This function includes two other function calls, "getFullTextURL(name)" and "getMetadata(DOI)." The information from both of these functions is used to create the XML.

The "getFullTextURL(name)" function gets a URL from Dropbox as the content of the "full-text-url" XML tag.

The "getMetadata(DOI)" function gets metadata from Crossref for the DOI. This metadata is used as the content for the "title," "publication-date," and "author" XML tags.

When the createXML.js file execution has been completed, the administrator can upload the upload.xml file to bepress, which concludes the workflow.

Experimental trials

To compare the time savings of *aperta-accessum* with processing publications manually, it was run for a department, the Western University's Electrical and Computer Engineering Department (https://www.eng.uwo.ca/electrical/people/faculty/index.html), and a high-productivity professor. The times for the administrator to complete the stages they were responsible for, Stage 1 and 3, were logged and repeated three times by one librarian at Western.

Institutional context

Western University is a large research-intensive university located in London, Ontario, Canada that is part of the U15 Group of Canadian Research Universities (U15, 2022). Western prides itself on research excellence and the success of its 40,000-member student body. Western identifies "Greater Impact" as one of the key pillars on which it anchors its latest strategic plan. One important way of achieving this goal is to

accelerate research, scholarship, and creativity "to serve not only individual disciplines but also the public good – by advancing knowledge and sharing it…" (Western University, 2021). As previously noted, OA is one way of serving the public good, as it facilitates equitable access to academic scholarship by eliminating access and financial barriers.

In 2019, Western established the Provost's Task Force on Open Access, which undertook a campus-wide consultation process toward an institutional policy on OA. As of this writing, the draft policy is still under review, and workflows are being established to support faculty in depositing their work under the new policy.

Limitations and future work

aperta-accessum is only as accurate as the APIs that it relies on. If there is incomplete or incorrect data in the databases that *aperta-accessum* uses, it will not catch these issues. For example, *aperta-accessum* searches the ORCID database using an author's first and last name and institution to get a list of publications; if authors do not have their publications linked to their ORCID profile or if these publications do not have DOIs, *aperta-accessum* cannot send emails for these missing articles to be uploaded.

Additionally, bepress does not support an API to upload or revise articles. If bepress adds this functionality in the future, it would streamline the *aperta-accessum* process significantly by removing the need for Dropbox, the creation of an XML file, and the editing and upload of a revised CSV file.

The primary current limitation of *aperta-accessum* is that it only works for bepress, which has been criticized as it is now owned by a for-profit academic publisher (McKenzie, 2017). The code of *aperta-accessum* could be expanded to enable its use in other commonly used open-source OA stacks such as D-Space, Islandora, and Samvera. Customizing *aperta-accessum* for these other applications would require modifying the code for Stage 2 and Stage 3. The difficulty of this modifications depends on the specific application, but, in general, open-source applications have more API functionality, which could remove the need to upload the XML, and delete the Dropbox files, which would make modifying the code easier.

Numerous other improvements could be made to *aperta-accessum*, including the following:

1. Creating a user interface to receive API tokens, any other required parameters, and buttons to run the scripts on click. A user interface would make *aperta-accessum*

more accessible to administrators with no or minimal command line experience. If a user interface was built, it would likely need to be either a desktop or cloud application because bepress does not allow cross-origin resource sharing (i.e., API calls cannot be made to bepress from within the browser of a non-bepress URL).

2. Automating the upload of the XML file containing new articles to bepress.

3. Automating the upload and creations of a CSV file for revision of articles already in bepress.

4. Automating the scripts to run at set intervals (e.g., run "sendEmail.js" bimonthly).

5. Providing better terminal output (e.g., color code key words, add loading animations, etc).

6. Testing the tool with a larger department or faculty.

Finally, Willinsky points out that there is a convergence between open source, OA, and open science (2005). There are well-known benefits not only for science but also for researchers for using open research practices: increases in citations, media attention, potential collaborators, job opportunities, and funding opportunities (McKiernan et al., 2016). To optimize this convergence, *aperta-accessum* could be expanded to other areas of open science (Spellman et al., 2017), including the following: data sets (Chen et al., 2018; Kazmi et al., 2021); FOSS (Von Krogh & Von Hippel, 2006; Von Krogh & Spaeth, 2007); free and open source hardware (FOSH)(Pearce, 2013; Pearce, 2015; Pearce, 2016; Maia Chagas, 2018); and new Elsevier companion journals such as *Data in Brief, MethodsX, SoftwareX,* and *HardwareX.*
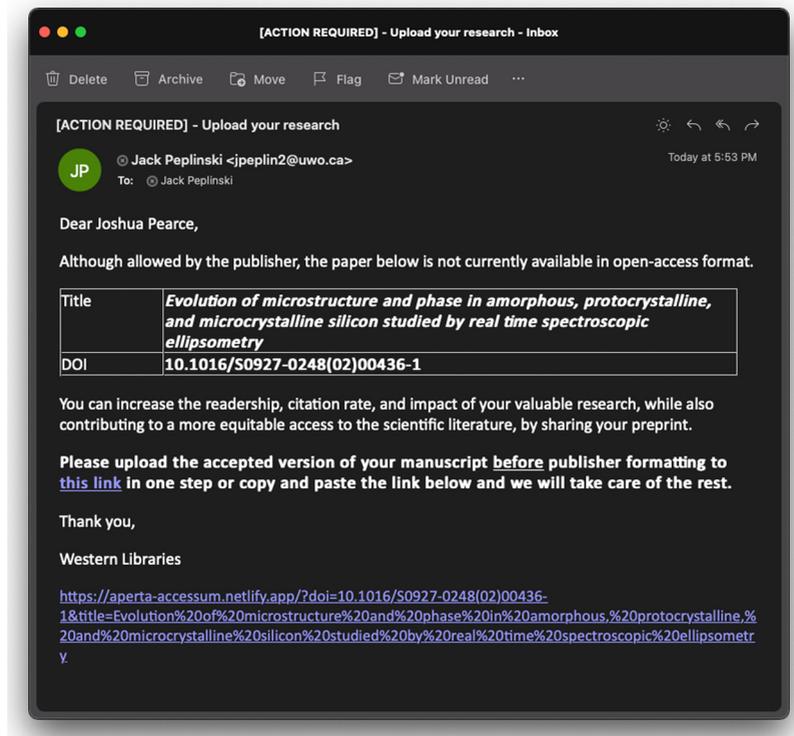
RESULTS

The open-source code for *aperta-accessum* was successfully developed for the workflow described in the Methods section. Figure 2 shows the results of the first script of Stage 1 when used on Western University's Electrical and Computer Engineering Department directory page. As Figure 2 shows, 37 people were found on the directory page, eight ORCID iDs were found for these people, 905 identifiers (e.g., DOIs) were found for these ORCID iDs, 280 DOIs were already OA, and 12 DOIs were found and were ready to email. Figure 3 shows an email generated to prompt researchers to upload their articles. Figure 4 shows the custom website that a professor is sent to upload the preprint or accepted manuscript of a specific article, and Figure 5 shows the successfully uploaded page. Of the 12 emails sent, 10 responses were received.
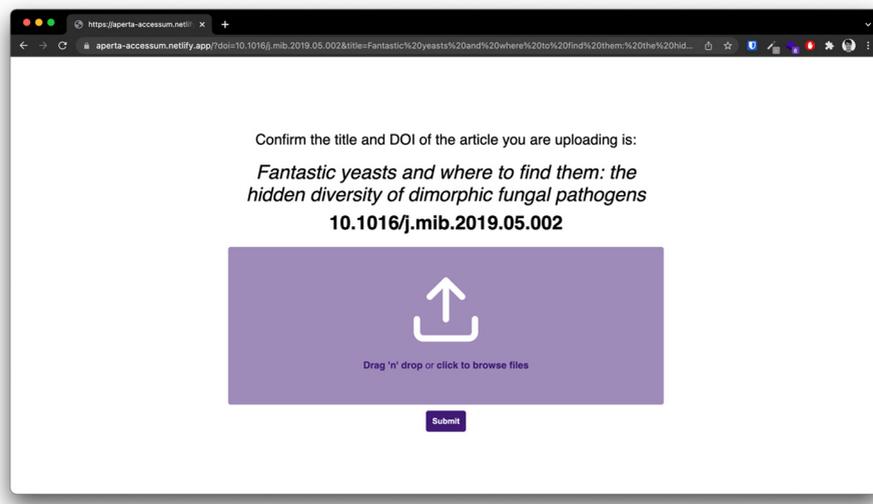
Figure 2. Results of Stage 1 when used on Western University's Electrical and Computer Engineering Department directory.

Figure 3. A sample email generated to prompt researchers to upload their articles.

Figure 4. A website page to upload articles.
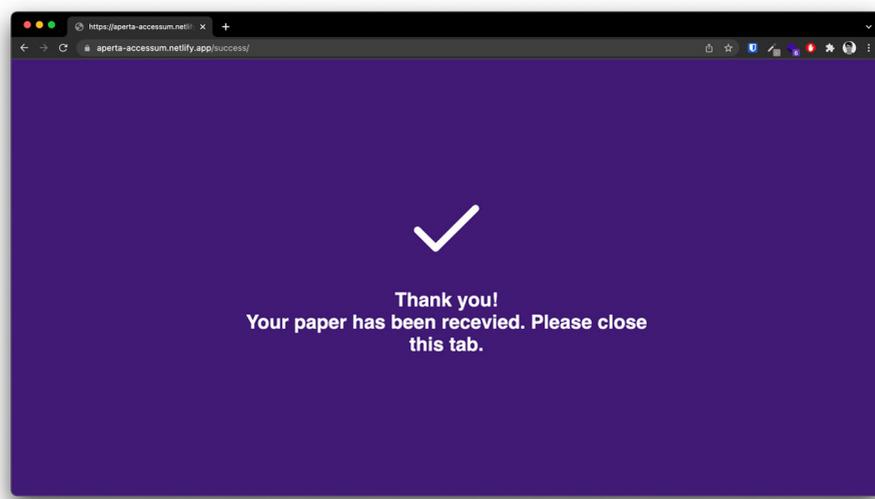
Figure 5. A success page after upload.

The results of the time trials are shown in Tables 1 and 2 for a department and individual researcher, respectively. The administrator can start running a script and let it complete in the background. The time the administrator spent directly working with *aperta-accessum* (e.g., starting a script, uploading a file, etc.) was classified as administrator time. The total time to complete the stage (i.e., administrator time plus the time to complete any background tasks) was classified as total time. As can be seen, the administrative time invested in running a department is under 3 minutes, and the total time is under 20 minutes.

Trial	Stage	Administrator time (mm:ss)	Total time (mm:ss)
1	1	01:38	16:30
	3	01:04	01:09
	1 and 3	02:42	17:39
2	1	01:36	17:02
	3	01:07	01:09
	1 and 3	02:43	18:11
3	1	01:36	16:24
	3	01:05	01:10
	1 and 3	02:41	17:34

Table 1. Timed-trial results of stages when using *aperta-accessum* on Western University's Electrical and Computer Engineering Department directory page

Trial	Stage	Administrator time (mm:ss)	Total time (mm:ss)
1	1	01:40	04:46
	3	01:02	01:06
	1 and 3	02:42	05:52
2	1	01:35	04:43
	3	01:04	01:07
	1 and 3	02:39	05:50
3	1	01:36	04:44
	3	01:05	01:09
	1 and 3	02:41	05:53

Table 2. Timed-trial results of stages when using *aperta-accessum* on a high-productivity researcher

As can be seen in Table 2, the administrative time to run an individual researcher is approximately the same as to run a department; however, because a department has many more papers to process, the total time is less for a researcher (e.g., < 5 min even for high-productivity individuals).

DISCUSSION

The results of the development and testing of *aperta-accessum* will first be discussed in terms of its performance and ability to make an entire institutions' scholarship OA. The best practices to roll out *aperta-accessum* are outlined. Next, the limitations are detailed, and future work is presented. Finally, the potential long-term impact of widespread adoption of this FOSS tool is discussed.

Performance

Tables 1 and 2 show that total administration time (i.e., the time a person would have to spend using *aperta-accessum*) is approximately 3 minutes, regardless of the number of papers. Therefore, this is a substantial improvement, as 15 minutes per manuscript is needed to accomplish the same process manually.

The librarian using the tool received an hour-long training session from the student who developed the software. After this hour-long training session, the librarian was able to complete all stages of the workflow. The Electrical and Computer Engineering Department was told verbally, by another professor, that they may receive emails asking them to upload research. Of the 10 responses, 2 responses attached documents instead of using the upload link, but these responses were from the same professor. All other respondents uploaded their documents correctly, indicating that they understood the process. There were no false-positives or issues with the uploads, but rolling the software out to a larger group would require some adjustments, which are detailed in the following section.

Rollout best practices

Capitalizing on the launch of Western's OA policy will be key in the rollout and success of *aperta-accessum*. Another important consideration will be to target departments and faculties where OA publishing is the default while simultaneously working with the broader campus community to raise awareness of the benefits of OA and Western's IR, Scholarship@Western. Acknowledging that uptake of OA is often dependent on unique disciplinary publishing cultures (Severin et al., 2018), it will be important to tailor messaging that meets the diverse needs of scholars. As *aperta-accessum* relies on data from ORCID, parallel efforts to support faculty in populating their ORCID profiles will be critical for maximizing the utility of *aperta-accessum*. This can be seen in the results from Figure 2, as most faculty currently do not have ORCID profiles. This process will take time and require coordinated communication between campus stakeholders, but the potential payoffs in terms of demonstrating research impact and public accountability are great. Currently, only 39% of Western University's scholarship is OA (COKI, 2022), which puts Western just under the average of the U15, as shown in Table 3. The effectiveness of the rollout can be determined by monitoring this statistic.

The *aperta-accessum* software is an exciting tool that will complement existing workflows as it is intended to capture previously published research articles, thus making it easier for interested faculty to upload all of their scholarship to the repository.

Institution	OA %	Total publications	Total open	Total citations
University of Alberta	37	116k	44k	3.4m
University of British Columbia	44	153k	67k	5.8m
University of Calgary	40	82k	33k	2.4m
Dalhousie University	40	44k	18k	1.3m
Université Laval	46	55k	26k	1.5m
University of Manitoba	40	47k	19k	1.3m
McGill University	45	116k	53k	4.0m
McMaster University	41	77k	32k	2.6m
University of Ottawa	41	70k	29k	1.9m
University of Saskatchewan	35	38k	13k	873k
University of Toronto	44	220k	96k	8.1m
University of Waterloo	33	65k	21k	1.7m
Western University	39	73k	29k	2.1m

Table 3. Percentage of OA publications for Canada's top 15 research universities.

Potential long-term impact

As *aperta-accessum* is a completely FOSS tool, it demands that any person or institution that adapts it to their own repository must reshare the code to benefit the overall global community following the tenants of the GNU GPL. There is a need for this to occur because this article only reports on the ability of *aperta-accessum* to function for the proprietary bepress-based IRs. As outlined in the previous section, there are many types of IRs commonly used. If only a handful of institutions make the relatively minor investment in adapting *aperta-accessum* to meet their own repository's requirements, it is theoretically possible that the entirety of academic output would at least provide relatively easy metadata-tagged OA to the literature. The political feasibility of this seems reasonably possible, as many academics are calling for unlimited access to the entire peer-reviewed literature (Budapest Open Access Initiative, 2002). There is already considerable evidence that both the number of platinum OA journals (free to read, with no APC for authors) as well as platinum OA with impact factors is growing rapidly (Pearce, 2022). There is also widespread interest in vastly expanding transparency in all aspects of the scientific knowledge-generating process (European Commission, 2015). Furthermore, recent research has shown a clear willingness of academics to expand OA, which would hasten scientific progress while also making science more just and inclusive (Pearce et al., 2022a, 2022b). The results of this article indicate that functionally doing this would not be prohibitively expensive or time consuming for the past literature and would provide a legal means to provide the same level

of access that Sci-Hub provides illegally (Bohannon, 2016). This will likely put economic pressure on the current business models of scientific publishers, which have been heavily criticized for profiteering from predominantly publicly funded research (Eisen, 2003; Monbiot, 2011; Buranyi, 2017). The conflict between green OA (self-archiving) and gold OA (APCs) is not yet resolved (Albert, 2006). A small percentage of all academic articles have been self-archived, but universal online access may be more readily available because of the use of *aperta-accessum* and the self-interest of scholars to have an easy way to have their work read and cited more often, thereby increasing their prestige. In addition to the rise of platinum OA journals with impact factors that enable academics to gain academic status and fully share OA without charges (Pearce, 2022), *aperta-accessum* also offers another path to relatively easy OA in journals with impact factors. This will add price pressure on journals as well as the need to find new business models in academic publishing.

CONCLUSION

This study has provided an economic method for universities that have established campus OA repositories to upload millions of manuscripts under numerous license agreements while also linking metadata to make them discoverable. The development and testing of a FOSS JavaScript-based application, *aperta-accessum*, was described in detail. The results show that *aperta-accessum* is capable of radical time savings for harvesting OA articles. In the administrative time that it takes to manually make a single document OA, an administrator using *aperta-accessum* can now process approximately five entire departments worth of OA articles. This study demonstrated *aperta-accessum* for a single type of repository. There is future work needed to adapt it for all OA repositories to enable universal OA to the peer-reviewed literature.

ACKNOWLEDGMENTS

DISCLOSURES

The authors declare that they do not have any conflict of interest, financial or otherwise.

REFERENCES

Albert, K. M. (2006). Open Access: Implications for scholarly publishing and medical libraries. *Journal of the Medical Library Association*, *94*(3), 253–262.

Antelman, K. (2004). Do open-access articles have a greater research impact? *College & Research Libraries*, *65*(5), 372–382. https://doi.org/10.5860/crl.65.5.372

Beaubien, S., & Eckard, M. (2014). Addressing faculty publishing concerns with open access journal quality indicators. *Journal of Librarianship and Scholarly Communication*, *2*(2), eP1133. https://doi.org/10.7710/2162-3309.1133

bepress (2021). Digital Commons Customers. bepress. https://bepress.com/products/digital-commons/why-digital-commons/customers/

Bernius, S., Hanauske, M., König, W., & Dugall, B. (2009). Open access models and their implications for the players on the scientific publishing market. *Economic Analysis and Policy*, *39*(1), 103–115. https://doi.org/10.1016/S0313-5926(09)50046-X

Bohannon, J. (2016). Who's downloading pirated papers? Everyone. *Science*, *352*(6285), 508–512. https://doi:10.1126/science.352.6285.50

Boldrin, M., & Levine, D. K. (2008). *Against intellectual monopoly* (Vol. 78). Cambridge University Press.

Boote, D. N., & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. *Educational Researcher*, *34*(6), 3–15. https://doi.org/10.3102/0013189X034006003

Budapest Open Access Initiative. (2002). *Budapest Open Access Initiative—Erklärung*. https://www.budapestopenaccessinitiative.org/boai-10-recommendations

Buranyi, S. (2017, June 27). *Is the staggeringly profitable business of scientific publishing bad for science?* The Guardian. https://www.theguardian.com/science/2017/jun/27/profitable-business-scientific-publishing-bad-for-science

Chagas, A. M. (2018). Haves and have nots must find a better way: The case for open scientific hardware. *PLOS Biology*, *16*, e3000014. https://doi.org/10.1371/journal.pbio.3000014

Chen, X., Gururaj, A. E., Ozyurt, B., Liu, R., Soysal, E., Cohen, T., Tiryaki, F., Li, Y., Zong, N., Jiang, M., & Rogith, D. (2018). DataMed–an open source discovery index for finding biomedical datasets. *Journal of the American Medical Informatics Association*, *25*(3), 300–308. https://doi.org/10.1093/jamia/ocx121

COKI. (2022). *Curtin Open Knowledge Initiative*. https://open.coki.ac/

Colla, E. (2020). Creating institution-wide awareness of, and engagement with, open scholarship. *Journal of Librarianship and Scholarly Communication*, *8*(1), eP2387. https://doi.org/10.7710/2162-3309.2387

DOAJ. (2021). Directory of Open Access Journals. https://doaj.org/

Eisen, M. (2003). The open access movement in scholarly communication. *Emerging Visions for Access in the Twenty-first Century Library*, 56.

Eysenbach, G. (2006). Citation advantage of open access articles. *PLoS Biology*, *4*(5), e157. https://doi.org/10.1371/journal.pbio.0040157

European Commission. (2015). *Digital agenda for Europe: New trends in open science*. http://ec.europa.eu/digital-agenda/en/open-science

Gibbons, M. (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. SAGE.

GNU (2007, June 29). GNU General Public License version 3.0 - GNU Project - Free Software Foundation. GNU. https://www.gnu.org/licenses/gpl-3.0.en.html

Gyore, R., Reeve, A. C., Cameron-Vedros, C., Ludwig, D., Emmett, A. (2015). Campus open access funds: experiences of the KU "One University" open access author fund. *Journal of Librarianship and Scholarly Communication*, *3*(1), eP1252. https://doi.org/10.7710/2162-3309.1252

Hajjem, C., Harnad, S., & Gingras, Y. (2005). Ten-year cross-disciplinary comparison of the growth of open access and how it increases research citation impact. *IEEE Data Engineering Bulletin*, *28*(4), 39–47. https://doi.org/10.48550/arXiv.cs/0606079

Harnad, S., & Brody, T. (2004). Comparing the impact of open access (OA) vs. non-OA articles in the same journals. *D-lib Magazine*, *10*(6). https://doi.org/10.1045/june2004-harnad

Heise, C., & Pearce, J. M. (2020). From open access to open science: The path from scientific reality to open scientific communication. *SAGE Open*, *10*(2), 2158244020915900. https://doi.org/10.1177/2158244020915900

Johnston, W. (2008). *Open access journals: the global movement and local publishing*. Routledge. https://atrium.lib.uoguelph.ca/xmlui/bitstream/handle/10214/1770/openAccessEjournals.pdf

Joseph, H. (2013). The open access movement grows up: Taking stock of a revolution. *Plos Biology*, *11*(10), e1001686. https://doi.org/10.1371/journal.pbio.1001686

Kakai, M. (2021). An analysis of the factors affecting open access to research output in institutional repositories in selected universities in East Africa. *Journal of Librarianship and Scholarly Communication*, *9*. https://doi.org/10.7710/2162-3309.2276

Kazmi, H., Munné-Collado, Í., Mehmood, F., Syed, T. A., & Driesen, J. (2021). Towards data-driven energy communities: A review of open-source datasets, models and tools. *Renewable and Sustainable Energy Reviews*, *148*, 111290. https://doi.org/10.1016/j.rser.2021.111290

Liauw, T., & Genoni, P. (2017). A different shade of green: A survey of Indonesian higher education institutional repositories. *Journal of Librarianship and Scholarly Communication*, *4*, eP2136. https://doi.org/10.7710/2162-3309.2136

Lewis, D. W. (2012). The inevitability of open access. *College & Research Libraries*, *73*(5), 493–506. https://doi.org/10.5860/crl-299

Liesegang, T. J. (2013). The continued movement for open access to peer-reviewed literature. *American Journal of Ophthalmology*, *156*(3), 423–432. https://doi.org/10.1016/j.ajo.2013.04.033

McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., McDougall, D., Nosek, B. A., Ram, K., Soderberg, C. K., Spies, J. R., Thaney, K., Updegrove, A., Woo, K. H., & Yarkoni, T. (2016). How open science helps researchers succeed. *eLife*, *5*, e16800. https://doi.org/10.7554/eLife.16800

McKenzie, L. (2017, August 3). *Elsevier expands footprint in scholarly workflow*. Inside Higher Ed. https://www.insidehighered.com/news/2017/08/03/elsevier-makes-move-institutional-repositories-acquisition-bepress

Maia Chagas, A. (2018). Haves and have nots must find a better way: The case for open scientific hardware. *PLoS Biology*, *16*(9), e3000014. https://doi.org/10.1371/journal.pbio.3000014

Monbiot, G. (2011). Academic publishers make Murdoch look like a socialist. *The Guardian*, *29*, 8.

Niyazov, Y., Vogel, C., Price, R., Lund, B., Judd, D., Akil, A., Mortonson, M., Schwartzman, J., & Shron, M. (2016). Open access meets discoverability: Citations to articles posted to Academia. *PloS One*, *11*(2), e0148257. https://doi:10.1371/journal.pone.0148257

Olsson, L., Francke, H., Lindelöw, C., & Willén, N. (2020). The first Swedish Read & Publish agreement: An evaluation of the Springer Compact pilot. *LIBER Quarterly*, *30*(1). https://doi.org/10.18352/lq.10309

Otto, J. J. (2016). A resonant message: aligning scholar values and open access objectives in OA policy outreach to faculty and graduate students. *Journal of Librarianship and Scholarly Communication*, *4*. https://doi.org/10.7710/2162-3309.2152

Partha, D., David, P. A. (1994). Toward a new economics of science. *Research Policy*, *23*, 487–521. https://doi.org/10.1016/0048-7333(94)01002-1

Pearce, J. M. (2013). *Open-source lab: how to build your own hardware and reduce research costs*. Elsevier.

Pearce, J. M. (2015). Quantifying the value of open source hardware development. *Modern Economy*, *6*, 1–11. http://dx.doi.org/10.4236/me.2015.61001

Pearce, J. M. (2016). Return on investment for open source scientific hardware development. *Science and Public Policy*, *43*(2), 192–195. https://doi.org/10.1093/scipol/scv034

Pearce, J. M. (2018). How to perform a literature review with free and open source software. *Practical Assessment, Research & Evaluation*, *23*(8). https://doi.org/10.7275/jjhz-sz75

Pearce, J. M. (2022). The rise of platinum open access journals with both impact factors and zero article processing charges. *Knowledge*, *2*(2), 209–224. https://doi.org/10.3390/knowledge2020013

Pearce, J. M., Pascaris, A. S., & Schelly, C. (2022a). Professors want to share: Preliminary survey results on establishing open-source-endowed professorships. *SN Social Sciences 2*, 203 (2022). https://doi.org/10.1007/s43545-022-00524-3

Pearce, J. M., Tiwari, S., Pascaris, A. S., & Schelly, C. (2022b). Canadian professors' views on establishing open source endowed professorships. *Cogent Education 9*:1, https://doi.org/10.1080/2331186X.2022.2122255

Pinfield, S., Salter, J., Bath, P. A., Hubbard, B., Millington, P., Anders, J. H., & Hussain, A. (2014). Open-access repositories worldwide, 2005–2012: Past growth, current characteristics, and future possibilities. *Journal of the Association for Information Science and Technology, 65*(12), 2404–2421. https://doi.org/10.1002/asi.23131

Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., Farley, A., West, J., & Haustein, S. (2017). The State of OA: A large-scale analysis of the prevalence and impact of open access articles. *PeerJ Preprints, 5*, e3119v1. https://doi.org/10.7287/peerj.preprints.3119v1

Poynder, R. (2011). *Suber: Leader of a leaderless revolution*. Information Today, Inc. https://infotoday.com/it/jul11/Suber-Leader-of-a-Leaderless-Revolution.shtml

ROARMAP. (2021). The Registry of Open Access Repository Mandates and Policies. http://roarmap.eprints.org/

Sample, I. (2012). Harvard University says it can't afford journal publishers' prices. *The Guardian, 24*, 2012.

Schweik, C. M. (2007). Chapter 10: Free/open-source software as a framework for establishing commons in science. In: Hess, C. & Ostrom, E. (Eds.), *Understanding knowledge as a commons: From theory to practice* (pp. 277–309). The MIT Press.

Severin, A., Egger, M., Eve, M. P., & Hurlimann, D. (2018). Discipline-specific open access publishing practices and barriers to change: An evidence-based review. *F1000 Research, 7*, 1925. https://doi.org/10.12688/f1000research.17328.2

Sismondo, S. (2010). *An Introduction to Science and Technology Studies*. (2nd ed.). Wiley-Blackwell.

Spellman, B., Gilbert, E., & Corker, K. S. (2017). Open science: What, why, and how. *PsyArXiv* https://psyarxiv.com/ak6jr/

Suber, P. (2003, September 4). *The Taxpayer Argument for Open Access*. SPARC Open Access Newsletter. http://dash.harvard.edu/bitstream/handle/1/4725013/suber_taxpayer.htm

Suber, P. (2012). Ensuring open access for publicly funded research. *BMJ, 345*, e5184. https://doi.org/10.1136/bmj.e5184

Swan A, Brown S. (2005). Open access self-archiving: an author study. JISC Committee for the Information Environment; Open Access Institute. *Key Perspectives* 4385 http://cogprints.org/4385/

Swan, A. (2010). The Open Access citation advantage: Studies and results to date. Technical Report, School of Electronics & Computer Science, University of Southampton, 2010. http://eprints.ecs.soton.ac.uk/18516/

U15. (2022). U15 Group of Canadian Research Universities. *About Us.* https://u15.ca/about-us

Von Krogh, G., & Von Hippel, E. (2006). The promise of research on open source software. *Management Science, 52*(7), 975–983. https://doi.org/10.1287/mnsc.1060.0560

Von Krogh, G., & Spaeth, S. (2007). The open source software phenomenon: Characteristics that promote research. *The Journal of Strategic Information Systems, 16*(3), 236–253. https://doi.org/10.1016/j.jsis.2007.06.001

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly, 26*(2), xiii-xxiii.

Western Libraries. (2021a). Scholarship@Western. https://ir.lib.uwo.ca/

Western Libraries. (2021b). Task Force on Open Access and Scholarly Communication. https://www.lib.uwo.ca/scholarship/task_force.html

Western University. (2021). Towards Western at 150: Western University Strategic Plan. https://strategicplan.uwo.ca/pdf/Western-Strategic-Report.pdf

Willinsky, J. (2005). The unacknowledged convergence of open source, open access, and open science. *First Monday, 10*(8). https://firstmonday.org/ojs/index.php/fm/article/download/1265/1185