



Repository Quick Submit and CV Scraping

Citation

Revin, Deborah, and Colin B. Lukens. 2020. Repository Quick Submit and CV Scraping. *Against the Grain* 32, no. 5: 66-68.

Published Version

<https://www.charleston-hub.com/2020/12/v325-biz-of-digital-repository-quick-submit-and-cv-scraping/>

Permanent link

<https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37366564>

Terms of Use

This article was downloaded from Harvard University's DASH repository, WARNING: No applicable access license found.

Share Your Story

The Harvard community has made this article openly available. Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

***Against the Grain* - November 2020**
Biz of Digital Column
By Deborah Revzin and Colin B. Lukens

A significant challenge in administering an institutional open-access repository is acquiring local scholarly content to distribute and build the repository. Complicated licensing and author re-use rights can sometimes be viewed as a barrier by authors who are looking to deposit their work. Paired with the challenges of communicating the benefits of repository deposit and the rights afforded by institutional open-access policies, limited resources, or lack of administrative support, repository managers often struggle to build a broader culture around deposits outside of open-access advocates. A proactive, mediated, and collaborative publication review program can mitigate or solve some of these issues. By reviewing an author's publication list or CV with an eye towards repository deposit, repository managers and scholarly communication librarians can demystify the process and educate depositors on licensing and open-access policies. Here, we outline such an effort at Harvard University.

The Harvard Library Office for Scholarly Communication (OSC) was founded in 2009 to support the first of many open-access policies adopted by the University.¹ The Digital Access to Scholarship at Harvard (DASH) repository was inaugurated shortly thereafter to support these policies.² Throughout the OSC's history, Open Access Fellows - graduate students from Harvard University and Simmons University's School of Library and Information Science program - have and continue to identify scholarship to deposit into DASH, determine licensing for works in the repository, and communicate with faculty about DASH deposits. These Fellows are invaluable resources in the OSC's mission to build scholarly content in DASH and to promote open access at Harvard.

Since its inception, the OSC has provided DASH depositing assistance to all members of the Harvard community. In the early years, Fellows contacted faculty to inform them of open-access policies and guide them through the depositing process. To minimize the number of steps in the depositing process and to ensure greater licensing compliance, a Quick Submit form for depositing works into DASH was created. This form has become a much faster way for authors to deposit a work, shifting metadata reconciliation and licensing determination responsibilities to Fellows and the OSC staff. The Quick Submit option on DASH's user-facing interface asks depositing authors to sign an Assistance Authorization (AA), which provides the range of licenses used for works in DASH and gives approved proxies permission to make DASH-related license

¹ <https://osc.hul.harvard.edu/policies/>

² <https://osc-harvard.pubpub.org/pub/2m1q3hm6/release/2>

choices on the author's behalf.³ In the case of a Harvard-affiliated faculty member, AAs re-affirm the University's open-access policy and serve as the opt-in for the new Harvard Individual Open-Access License.⁴

In addition to the self-deposit Quick Submit option, the OSC offers affiliates of the University a mediated deposit service known as a CV Scrape. This service increases deposits into DASH and has become a vital part of making more scholarship at Harvard open-access. The CV Scrape program also helps the OSC foster a culture and understanding around repository deposits at the University. Knowledge of and interest in this service is driven by presentations given to faculty and faculty assistants by OSC staff, library peers advocating this service to authors, word of mouth, and the OSC's webpage outlining the CV Scrape service.⁵

The first step in the process is to acquire an author's CV and a signed AA. From there, Fellows begin working on the scrape using a custom spreadsheet template to capture data from the author's CV. The columns of the spreadsheet act as a workflow to guide Fellows through a series of decision points. First, citations are collected and works are crosschecked with those already in DASH, to prevent duplicate entries. Next, Fellows search for publisher DOIs or manuscript URLs and lastly, SHERPA/RoMEO is used to determine the publishers' copyright and re-use policies. When all the necessary data has been recorded on the spreadsheet, Fellows make deposits into DASH those works for which a publisher or University open-access policy allows for its distribution. Finally, the Senior Repository Manager (SRM) is informed of the completion of the CV scrape. From this point, the SRM sends a distillation of the spreadsheet to the author, which acts as an Outcomes Report. This report is color-coded to indicate what the Fellow was able to make available in DASH, what already exists in the repository, and if certain versions of work are needed from the author to be deposited into DASH. As well as offering next steps to authors, the Outcomes Report empowers them to make more of their works available in the repository while also creating an opportunity for the OSC to educate authors on open access, licensing, and copyright and to engage in fruitful discussions on scholarly communication. These discussions position the library as the

³ Recently, the OSC has piloted and adopted a collaborative distributed workflow program that divides the work of repository deposits, metadata review, and licensing amongst groups of helpers located within the Harvard Library, the University's school library units, and academic administrative staff. This program is called D3, or Distributed DASH Deposit. A 2018 DLF Forum presentation outlined this program: bit.ly/D3_DLF

⁴ In 2018, Harvard adopted an opt-in open-access policy for all non-faculty Harvard-affiliated authors, thereby giving all affiliates the same rights as those afforded under faculty-approved policies. Details of this new policy are outlined here: <https://osc.hul.harvard.edu/policies/ioal/>

⁵ <https://osc.hul.harvard.edu/authors/cv-faq/>

knowledge center for these issues, helping to create and cultivate a culture of open access and continued depositing into DASH.

The CV Scrape service has a number of proven benefits. The primary benefit is the increase in the amount of Harvard scholarship that is open access and available in DASH. The more personal patron service model provides flexibility for authors with varying levels of familiarity with open access. Authors who are new to depositing scholarship into a repository and the concept of open access often feel unsure of where to begin. The CV Scrape process gives authors a starting point by initiating the creation of their collection in the repository. It is hoped that this beginning produces favorable statistics that can act as a motivator in turn, encouraging more regular deposits into DASH. For authors who are familiar with the repository and open access, the CV Scrape process mitigates complications and streamlines depositing.

Even though this process has many benefits, it also poses a few challenges. Some of the most common obstacles Fellows face when working through a CV Scrape include a lack of communication from authors, an author's assumption that all scholarship on a CV can be made available in DASH, licensing that can often be difficult to determine (especially for older journals), and the labor intensive nature of the process. Some scrapes demand a lengthy amount of time, which diverts attention away from other projects and pulls from Fellows' limited availability.

The OSC continues to monitor ways to improve its mediated depositing service, in part by testing new applications and processes. Last year, Fellows tested the Open Access Permission Checker,⁶ then in its beta-form, and provided feedback and commentary on using the tool in the CV Scrape process. The Open Access Permission Checker was created by Our Research, a joint venture nonprofit co-founded by Heather Piwowar and Jason Priem. The Checker was one aspect in the team's larger project, Unpaywall, created to "help make scholarly research more open, accessible, and reusable."⁷ The OSC also stays up to date on what processes are used by other colleges and universities to build and enhance their repositories. Sharing feedback on ways to improve repository collections serves all institutions. The OSC would love to hear from you on how your scholarly communication office or repository managers are using CV Scrapes or other processes to populate collections in your institutional repository. Share your projects with us at OSC@harvard.edu.

⁶ <https://shareyourpaper.org>

⁷ <https://unpaywall.org/team>