Methods for Populating Scholarly Profiles with Repository Data

Law Repositories Caucus Sandbox Series 2021 – Aug. 2, 2021

Pam Brannon

Background

For over a decade, the Georgia State University College of Law Library has maintained a database of faculty publications. The goal is to include metadata for every publication authored by a College of Law faculty member. This database was originally housed in a homegrown system, but is now a part of the Law Library's Digital Commons repository. The faculty publications database contains a mixture of full-text and metadata-only records. These records are the primary source used to populate faculty members' SelectedWorks profiles.

Serious discussion of ORCID adoption began at Georgia State University in 2016, and the goal from the beginning was to use the faculty publication information maintained by the Law Library to populate faculty ORCID profiles. At the time, however, there was no ORCID integration with Digital Commons; although there is an integration currently in development, that integration is designed to pull information from ORCID into Digital Commons, rather than push information from Digital Commons to ORCID. So we had to figure out how to move that information ourselves.

The focus of our efforts has been on populating ORCID profiles with publication information. However, Academia.edu also accepts batch upload via BibTeX, and I have used the file to add publication information to Academia.edu for faculty.

Potential Methods for Populating Scholarly Profiles

ORCID API

ORCID offers an API that can be used to populate profiles. In this method, the faculty member authorizes the institution to change/update their profile. However, there are several barriers to effectively using the ORCID API. First, the API to change/update profiles is only available to ORCID member institutions. Second, developing a system using the ORCID member API often requires more IT development resources than are available, particularly when the repository system housing the information does not offer a programmatic method for retrieving information.

BibTeX Import

Alternatively, ORCID allows for batch import of bibliographic information formatted in BibTeX. The BibTeX file for each faculty member can then either be sent to the faculty member to import themselves, or can be imported by a librarian who has been designated a "trusted individual" who can add/update profile information in ORCID.

Formatting Repository Metadata in BibTeX

Zotero/Juris-M/EndNote

One relatively low-tech way to gather bibliographic information is to use a reference manager such as EndNote, Zotero, or Zotero's law-focused offshoot, Juris-M. Once bibliographic information is in Zotero, it can be exported in BibTeX format and imported into ORCID. This method has the advantage of using

software that is relatively easy to install and use, so it doesn't have a steep learning curve. However, there are a few problems with this method. First, Zotero/Juris-M is inconsistent. Zotero does not retrieve multiple records at once from our Digital Commons repository, so publication metadata has to be retrieved item by item. Second, Zotero may not retrieve all metadata for a publication. This makes using Zotero a slow and somewhat error-prone way of creating a BibTeX bibliography.

Coding in Python, R, etc.

Another way to create a BibTeX file that can be used with ORCID is by scripting in a programming language such as Python or R. Our Digital Services Librarian at the time, Jonathan Germann, is a Python programmer, so he fairly quickly developed a Python script that creates a BibTeX file for a faculty member using the information contained in the Digital Commons batch revise Excel file. This method has a number of advantages, primarily because the script is adaptable and customizable. If we did not want to include a field or if we want to include an additional field represented in our batch revise file, we can modify the script to do that. I am currently adapting the script to work with information retrieved using the bepress Digital Commons API. Another advantage is the speed with which a file can be created. With a working script, a BibTeX file containing metadata for dozens of publications can be created and imported into ORCID within minutes.

Perhaps the most significant barrier to this method is technological expertise. At the time the script was developed, we had a librarian who was comfortable using Python. However, that librarian left GSU, and so I had to learn how to use and modify the script. Since he left, I and several of my research assistants have made several modifications to the original script, and, as mentioned above, I am currently working to adapt the script for use with the Digital Commons API rather than the batch revise file.

Current Workflow

Below is my current workflow for populating ORCID profiles:

1. Prompt the faculty member to create an ORCID profile. In the prompt, ask the faculty member to add me as a "trusted individual."
2. Download a current version of the Digital Commons batch revise file.
3. Save the file as facpub.csv in CSV UTF-8 format.
4. Open and run the script in Python. The script prompts for the faculty member's email address and last name, and creates a new file named [lastname]_bibtex.bib.
5. Go to my ORCID profile and switch to the faculty member's profile.
6. In the Works section of the faculty member's profile, select "Add Works" and then "Import BibTeX"
7. Load the file created by the Python script and save the records imported.

Next Steps

As mentioned above, I am currently rewriting the script for use with the Digital Commons API. As the Digital Commons API does not currently make available first and last page information, we will not move to use the API until this is fixed.

I am also working to modify the script so that it only retrieves publication metadata entered after a certain time. The script currently retrieves information for *all* publications by a specific faculty member.

Retrieving only publications entered after a specified date will allow me to update the ORCID profile for the faculty member without having to first delete earlier publication information.