Better together:
Tracking research outputs & assessing accessibility

Dipa Barua, Research Assistant; **Sue Bengtson**, Collection & Systems Librarian; **Cassandra Larose**, Librarian, Scholarly Communication; **Gwen Thompson Hill**, Manager of EDI in Research; **Melanie Wrobel**, Research Development Coordinator

Find us at CCRA 2026 in Richmond, BC! We will have handouts to share on **replicating this process** and the **early results of Phase 3**.

The mission

To find out the extent to which research outputs from Royal Roads University are accessible

The scope

- **Who:** Core faculty and librarians only
- **What:** Publications with RRU affiliation in OpenAlex & Scopus
- **When:** Between January 2020 and July 2025

The people

Two staff from Research & Innovation, two from the Library, and one spectacular Research Assistant.

The limitations

- Doesn't capture everything (e.g., blog posts, interviews, op-eds)
- Doesn't capture publications from non-core faculty or staff, even if they list their affiliation as RRU

The dream

A way to **track, collate, and view** our faculty's research outputs, and to **tie those outputs to publisher accessibility standards**.

The unexpected

Working together across departments increased our understanding of each other's roles and challenges, leading to new collaborations between our departments.

Overview

This project began with a 'simple' goal: To assess the accessibility of our research outputs. When our researchers publish, how accessible are those publications? What standards do the publishers require, and what support do they offer for people with access needs? The Research office and Library staff decided early on to collaborate on this project, combining their skills in research, bibliometrics, and carving out time for important projects in busy schedules. Together, the two departments ran a competition for, and hired, a research assistant: Dipa Barua, a Master of Library and Information Studies student at the University of Alberta.

The team immediately identified a complication to the goal of assessing accessibility: RRU did not have an easily searchable database of faculty publications. Library staff confirmed that this was a project that would need to be done at some point for a variety of aspects of their work; the database thus became a new intended output of our project and will be the focus of this poster presentation. Using Scopus and OpenAlex, Barua and the team built a database of the research outputs of RRU's "core" (tenured or tenure-track) faculty and librarians over a period of five years. Along the way, Barua's documentation of the process ensured that we would have a replicable process for future use. The database was then used to identify the most frequently appearing publishers, which we then assessed to identify and compare accessibility standards.

Methodology

Phase 1: Preparation

Faculty & Librarian list. Human Resources staff were able to provide us with a confidential list of faculty and librarians who had been at RRU over the past 5 years. We included faculty and librarians who left the university during that period.

Access to systems. Barua required access to RRU systems that are not automatically made available to Research Assistants.

System decisions. The team discussed which systems to use for the database, and elected to use Excel, with which all team members are familiar. The low learning curve for busy team members was a key factor in this decision.

Scopus and OpenAlex. Larose and Bengtson worked with Barua on navigating Scopus and OpenAlex. Barua independently also took online training for both systems, as well as some specific Excel training.

Phase 2: Data collection

Data extract test 1: OpenAlex, Scopus, and Excel Power Query

Barua's initial test used two faculty members (referred to as Fa & Fb). The initial filters were:

1. Author Name
2. Year Range (2020-2025)
3. Institution/Affiliation - Royal Roads University

Results included 39 rows of results for Fa on OpenAlex and 10 rows for Fb on Scopus. Using Excel's Power Query tool, these files were merged; the column names and orders were adjusted to align, showing a result of 22 unique entries. The following column headings were chosen:

ID, DOI, Title, Publication Date, Type, Indexed In, Author IDs, Journal Name, ISSN, Indexed in Scopus, Publisher, Open Access, OA Status, APC Value, APC Currency, APC Paid, APC Paid Currency, Volume, Issue, Authors, Author Affiliations, and Grant Funder.

The DOI column was used as a unique identifier to remove duplicates from both databases. Example of initial sorting results:

id	DOI	Title
2-s2.0-85211586224	10.3389/frai.2024.1466321	Predicting financial distress in TSX-listed firms using machine learning algorithms
2-s2.0-105009347236	10.1108/JAL-11-2022-0112	The application of machine learning to study fraud in the accounting literature
2-s2.0-85195506474	10.1080/00207543.2024.2361434	Supply chain fraud prediction with machine learning and artificial intelligence
2-s2.0-85136846898	10.1080/19361610.2022.2114744	Predicting Money Laundering Using Machine Learning and Artificial Neural Networks Algorithms in Banks
2-s2.0-85217419646	10.1016/j.bar.2025.101560	Reprint of: The use of machine learning algorithms to predict financial statement fraud
2-s2.0-85199055824	10.1016/j.bar.2024.101441	The use of machine learning algorithms to predict financial statement fraud
2-s2.0-85148518373	10.1080/13504851.2023.2176435	Predicting money laundering sanctions using machine learning algorithms and artificial neural networks
2-s2.0-85171161584	10.1111/capa.12541	From IDA to IIROC: Has self-regulation in the Canadian investment industry evolved?
2-s2.0-85172675879	10.1177/21582440231197305	Two Decades of Accounting Fraud Research: The Missing Meso-Level Analysis
2-s2.0-85150506185	10.1007/s42001-023-00202-1	Incorporating machine learning in dispute resolution and settlement process for financial fraud
In Original Export Files		
	Means it's in both OpenAlex & Scopus exports	
	Means it's only on Scopus export	
In Appended File		
	Shows up in Appended file - but twice (one under the Scopus ID and the other under OpenAlex ID - FAIL)	
	Does not show up in Appended file - FAIL	
	Shows up in Appended file 3 times in different publications, under different DOIs (successful)	
	Shows up in Appended file 2 times in different publications, under different DOIs (successful)	
	Shows up in Appended file ONCE - this one was successful	

Refining this data included appending files again, removing duplicates, and cross checking against raw data files. Excel Power Query seemed to give too many errors and required a lot of manual cross checking; which led to:

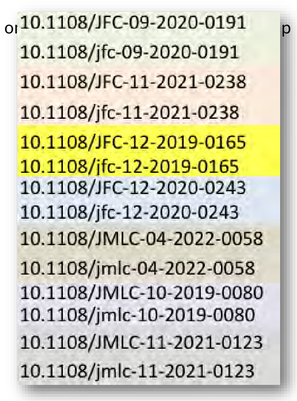
Data extract test 2: OpenAlex, Scopus, and Bibliometrix

"Bibliometrix is an open-source tool for quantitative research in scientometrics and bibliometrics that includes all the main bibliometric methods of analysis[1]". Switching to Biblioshiny, the user-friendly web app of Bibliometrics, reduced the manual labour required and increased our accuracy. Using the same examples (Fa & Fb), and the same filters from Data extract test 1, Scopus and OpenAlex extracts were combined. All Scopus results were in the final document.

Not all OpenAlex results were in the final document. Results with the same title but different DOIs (e.g., in one the Final Published version, and in another chose the Repository version) had to be manually removed.

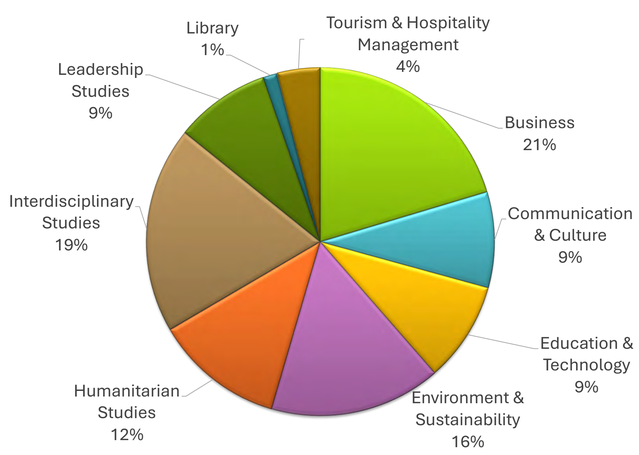
It was decided to keep both the Final Published and Repository versions. The same test was run again on Bibliometrix to merge and remove duplicates for Fa and Fb results from OpenAlex and Scopus. After the merge and removal of duplicates, a manual check of the exported results on Excel showed that 7 DOIs remained in the results that were repeats due to inconsistent formatting of the DOIs. Bibliometrix is case-sensitive and is therefore unable to differentiate between DOIs with uppercase and lowercase, thus, it kept both versions, as shown in the image on the right of this text.

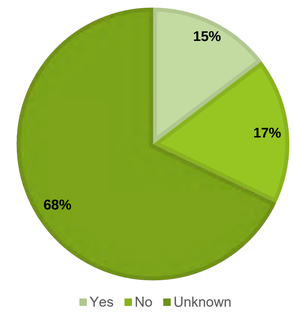
The Bibliometrix results on Excel were sorted alphabetically which made it easy to remove the duplicates manually from the file. An Excel book was created which included a tab for each faculty member. Each tab contained the research output results exported from OpenAlex and Scopus and merged using Bibliometrix with duplicate DOIs removed. Finally, the tabs were merged to create one spreadsheet for all the results.



[1] https://www.bibliometrix.org/home/index.php/layout/bibliometrix

What can our new database tell us?





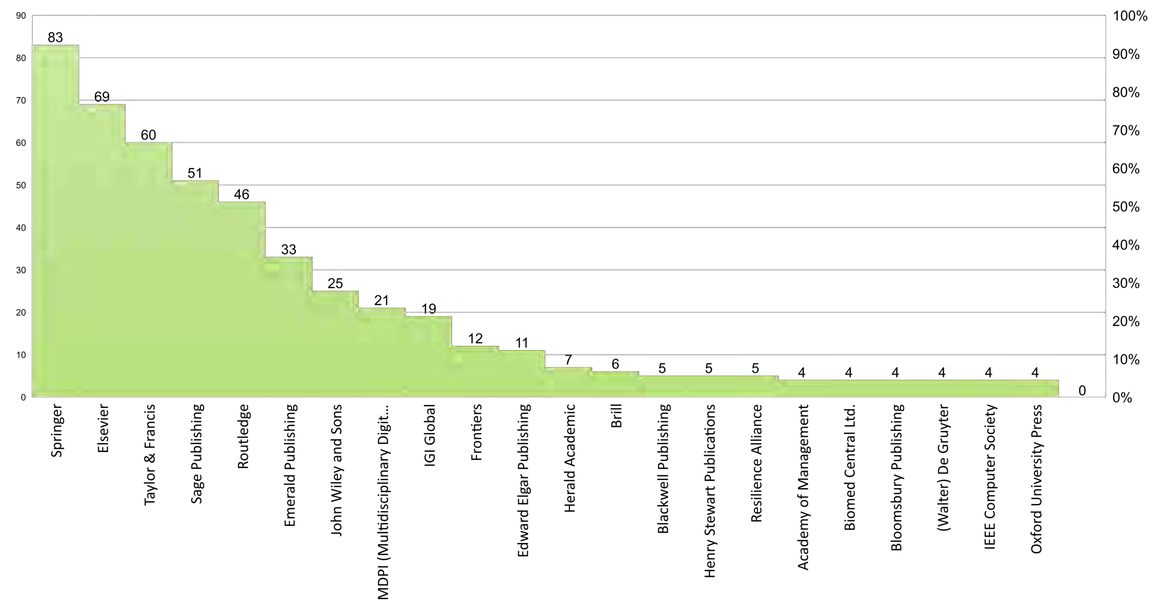
Open access status

21% of publications included non-RRU affiliations

96% of publications had unique DOIs

89% of publications had abstracts available

Publication Distribution Across Schools



Where do we publish the most? (n <4)

Note: Research Square, a pre-print platform in the USA, was removed from this data.

Phase 3: Assessing publisher accessibility

We began this third phase in early 2026 with a complete, sortable, searchable database. The database has allowed us to determine where RRU Faculty and Librarians are publishing the most often, and we have used that data (see bar graph, above) to begin assessing publisher accessibility standards, moving from most-used publishers to least.

3.1 - Initial data gathering via a review of publicly available information
Initial scans of publisher websites will provide us with information such as explicit accessibility commitment statements, accessibility team information and contact, accessibility guides for authors and editors, inclusive language guides, and more. These will be compiled with hyperlinks and connected to the publisher identity.

3.2 - Secondary data gathering via direct contact with publishers
We are aware that our initial review may miss information – publisher websites are complex and often contain a great deal of information, and we do not want to miss the inclusion of an accessibility consideration. Once the initial scan is complete, we will share each publisher's information with them and ask that they provide any information we have missed, or that is not publicly available (e.g., guides that are provided to authors only once an article has been accepted, etc.).

3.3 - Development of matrix for assessment
Once we have gathered data per 3.1 and 3.2 above, we will need to assess the accessibility factors of publishers. Our intent is to do this via the development of a matrix in partnership with experts in the field, but **we are open to suggestions and feedback from readers of this poster.**

Please contact us via Gwen at gwen.3hill@royalroads.ca .

This research was undertaken thanks to funding from the Canada Research Chairs Program.

Canada Research Chairs / Chaires de recherche du Canada

Canada

Royal Roads University