

# The use of institutional repositories for self-archiving in Canadian universities (Paper)

## Abstract or Résumé:

This study investigates the use of institutional repositories (IR) for self-archiving journal articles in the U15 universities as well as the presence of institutional policies and publisher embargos. While 45.1% to 56.6% of publications are available in open access (OA), only 0.5% to 10.7% are found in the IRs. We found only three university-wide OA policies, and embargo periods of 12 months or more for 25.6% of journal policies. This suggests that IR play a minor role in OA practices, and a need for more policies related to self-archiving and the use of IR specifically.

## Introduction

This study focuses on the Canadian universities' IR, which are digital storage facilities, locally or cloud-based, usually managed by the university library to preserve and make available research output, theses, teaching material, datasets, etc., produced within the organization. IR developed globally with usage differentiating across countries and regions, reflecting cultural responses to publication and dissemination needs. Growth rates are thought to be plateauing in North America, Germany, and the UK (Moskovkin et al. 2021) due to “saturation” (Pinfield et al. 2014). However, despite being an early signatory to the Budapest OA initiative and championed by Canadian researchers, Canada is not at the forefront of OA (Moskovkin et al. 2021; Simard et al. 2022): while Canada sits within the top 10 global repositories by count, its lack of OA participation from Canadian researchers has affected global OA ranking when compared with other high-income countries (Moskovkin et al. 2021). Contribution differences have also been highlighted across provinces (Larivière and Macaluso 2011; Paquet, van Bellen, and Larivière 2022). Canada's U15 IRs are predominately English, institutional, multidisciplinary, and use OAI-compliant metadata, factors which should contribute to their widespread usage (Pinfield et al. 2014). However, the U15 IRs experience low deposit levels (Paquet, van Bellen, and Larivière 2022) and more research is needed to analyze the effectiveness of initiatives and policies prescribing their use.

Despite the increased awareness of the importance and ethical push for OA (Creaser et al. 2010) and, more broadly, Open Science (Boulton et al. 2020; UNESCO, 2021), little attention has been given to the use of the IR by researchers within a Canadian university for the distribution of research results as Green OA, and how its use may be affected by university policies, journal self-archiving policies, or disciplinary cultures. While past research has looked at self-archiving rates in Canada (Paquet, van Bellen, and Larivière 2022) and globally (Simard et al. 2022), this work is, to our knowledge, the first to focus on the use of IR repositories specifically and to consider institutional and journal OA policies. Another unique feature of our work is its reliance solely on an open data source, OpenAlex, (Priem et al., 2022). We hope to understand the factors that may facilitate or inhibit OA practices by researchers and organizations, how disciplinary differences may be contributing, and how policies from the university and publishers may or may not be contributing to effective self-archiving practices.

The purpose of this study is to provide an overview of the uptake of OA in Canadian universities that are members of the U15 (an association of fifteen Canadian research-intensive universities) with a particular focus on the usage of their institutional repositories (IR) as a reflection on the effectiveness of

the university OA policy and the intersection of the publisher policy. More specifically, we aim to provide answers to the following research questions:

1. What OA policies are in place in the U15, and how are the policies acknowledging the IR?
2. What are the characteristics of the OA policies of journals in which the U15 publishes in terms of self-archiving in IR, the manuscript version allowed for deposit, and embargo periods?
3. What percentage of the research output from the U15 is available in OA, green OA, and in the IR?
4. Are there disciplinary differences in OA publishing and self-archiving?

## Methods

### *Data collection*

We used a relational database version of OpenAlex hosted by the Maritime Institute for Science, Technology, and Society (MISTS) to collect all journal articles published between 2016 and 2021 with at least one author affiliated with a Canadian U15 university. Since research articles and reviews are likely to contain more references than other document types and to have a Digital Object Identifier (DOI), we limited our dataset to works with a DOI citing at least ten other OpenAlex works<sup>i</sup> to mitigate this limitation. Our dataset thus constitutes a convenience sample for comparison purposes, and the numbers provided are not meant to be understood as absolute measures of the U15 universities' research output. Following Rivest et al. (2021), we used the Science-Metrix classification of scientific journals (Archambault, Beauchesne, and Caruso 2011) to assign a discipline to journals. The IR names and URLs were manually retrieved from each institution.

We used the Unpaywall<sup>ii</sup> API to obtain the OA status and locations of the works in our dataset. We then checked whether the IR of the institution was listed as one of the OA locations by searching for the repository URL in the *URL* field of the *OA Location* object of the Unpaywall data schema. We also use the OAI-PMH-enabled API to harvest all records from each repository. Titles of articles were matched using a fuzzy matching algorithm with low thresholds below 0.2 manually matched to the OpenAlex data. Any version of an article, (submitted, accepted, or published) found in the IR is counted as self-archived.

In July 2022, we manually searched the university websites to retrieve institutional policies on OA and open science and coded university OA policy content according to three criteria: 1) Policy type (whether deposit in the IR is mandated, suggested, or not mentioned), 2) manuscript version (the deposit of which version of the manuscript is mandated or suggested), and 3) timing (when does the policy require or suggest that the manuscript be deposited in the IR). To account for policies that may not be on university websites, missed in our search, or been published after Tummon & Desmeules's (2022) study, we contacted the scholarly communications librarians (or the role nearest to this) by email at each university in January 2023 for confirmation.

On July 20, 2022, we used the Sherpa Romeo API to collect data on the journal's self-archiving policies. We collected the following elements of information from Sherpa Romeo: 1) Whether self-archiving is allowed, 2) which manuscript version can be self-archived in an IR, and 3) the embargo period for archiving in an IR.

Table 1: Final dataset and variables

Category	Variable	Description	Source
Publication	paper id	Unique identifier of the publication in the OpenAlex.	OpenAlex
	DOI	Digital Object Identifier of the publication.	OpenAlex

	Publication year	Year of publication in the journal.	OpenAlex
	Discipline	Science Metrix Classification	OpenAlex + Science Metrix
	University	U15 institution of the author	OpenAlex
Institutional Repository	Repository name	Name of the institutional repository	University website
	Repository URL	URL of the institutional repository	University website
OA status and locations	OA status	Whether a publication is accessible in OA	Unpaywall
	Own institutional repository		Unpaywall + OAI-PMH API
	Version		Unpaywall
Journal policy	Green OA allowed	Key:value pair from JSON data	Sherpa Romeo
	IR deposit allowed	Key:value pair from JSON data	Sherpa Romeo
	IR version allowed	Key:value pair from JSON data	Sherpa Romeo
	IR version embargo	Key:value pair from JSON data	Sherpa Romeo
Institutional policy	Policy exists	Manually derived	University website
	Policy type	Manually derived	University website
	Version required	Manually derived	University website
	Timing	Manually derived	University website

## Results

Our findings show the percentage of OA items consistent with prior studies, with most of the U15 institutions around 50% overall and 40% Green OA. The proportions of Gold to Green OA are similar when compared to a recent study on Canadian OA (Paquet, van Bellen, and Larivière 2022), which also investigated OA by authors with Canadian university affiliations. In table 2, we find a low use of IR, with between 0.5% and 10.7% of peer-reviewed articles deposited in the IRs. In line with prior research (Larivière and Sugimoto 2018; Paquet, van Bellen, and Larivière 2022), we find differences in OA publishing and self-archiving across disciplines, with Health and Natural Sciences leading, followed by Applied Sciences, Natural Sciences, Economics & Social Sciences, and lastly Arts & Humanities, as shown in table 3. When considering the overall use of Green OA, the Applied Sciences and the Economics and Social Sciences seem to make slightly higher use of the IR for self-archiving. While our data covers 2016-2021 from a different data source and with limitations on scope, we confirm findings in the use of IR observed by Alexandre-Benavent et al. (2019), which they identify as a decrease over time.

Table 2. Number of publications and OA status for OpenAlex works by the Canadian U15 universities. The percentages shown are of the number of works attributed to each university.

Institution	Works	Open access (%)		Institutional repository (%)			Version of paper available in IR (Unpaywall only) (%)		
		All	Green	Unpaywall	OAI-PMH	Combined	Submitted	Accepted	Published
Dalhousie University	7,350	55.3	43.8	0.0	1.1	1.1	0.0	0.0	0.0
McGill University	18,483	58.1	48.3	0.0	4.0	4.0	0.0	0.0	0.0
McMaster University	13,115	55.3	44.5	0.4	1.0	1.1	0.3	0.0	0.1
Queens University	7,934	53.4	43.7	2.5	0.0	2.5	1.3	0.5	0.7
Université Laval	9,020	58.6	47.1	4.4	5.8	6.8	0.1	2.5	1.8
University of Alberta	19,990	47.1	36.3	0.0	1.4	1.4	0.0	0.0	0.0

University of British Columbia	25,787	55.8	45.6	0.0	5.1	5.1	0.0	0.0	0.0
University of Calgary	14,223	54.5	43.6	5.3	4.9	5.8	0.5	0.4	4.4
University of Manitoba	8,117	55.3	44.1	5.3	4.1	5.4	0.5	0.2	4.6
University of Montreal	12,286	59.5	49.4	2.3	2.2	2.9	1.4	0.2	0.7
University of Ottawa	12,142	53.4	42.1	0.0	5.2	5.2	0.0	0.0	0.0
University of Saskatchewan	6,812	48.6	37.4	0.1	0.5	0.5	0.1	0.0	0.1
University of Toronto	37,020	56.6	46.5	2.7	5.5	6.0	0.2	0.1	2.4
University of Waterloo	10,634	46.4	37.1	3.6	3.6	4.0	0.8	1.1	1.6
Western University	12,472	50.7	41.4	0.0	10.7	10.7	0.0	0.0	0.0

Table 3: Number of publications and OA status by domain, sorted alphabetically.

Institution	Works	Open access (%)		Institutional repository (%)			Version of paper available in IR (Unpaywall data)		
		All	Green	Unpaywall	OAI-PMH	Combined	Submitted	Accepted	Published
Applied Sciences	33,831	33.8	23.1	1.9	3.8	4.2	0.4	0.5	1.0
Arts & Humanities	2,448	37.5	21.0	1.3	2.2	2.5	0.3	0.3	0.7
Economic & Social Sciences	11,694	34.5	20.7	1.8	3.7	4.2	0.8	0.4	0.6
Health Sciences	98,209	61.5	51.6	4.0	5.5	6.0	0.3	0.2	3.5
Natural Sciences	35,260	53.8	44.8	1.7	3.3	3.8	0.4	0.6	0.7

Table 4 shows our findings on the publisher policies reveal that of the articles published by the U15, that of 15,400 journals with policy information, only 8,996 journals had complete information including self-archiving permitted, manuscript version allowed for deposit, and embargo periods. 24.6% of those journal policies had embargo periods over 12 months affecting accepted and published versions.

Table 4: Embargo period length for journal policies that specify IR as a location, sorted by embargo period and version permitted for deposition. Percentages are of the total N journals.

Embargo (months)	Submitted		Accepted		Published		Total	
	N	%	N	%	N	%	N	%
0	88	0.98	128	1.42	18	0.20	234	2.60
1	0	0.00	0	0.00	3	0.03	3	0.03
3	0	0.00	1	0.01	6	0.07	7	0.08
6	0	0.00	339	3.77	59, 2*	0.66	400	4.45
12	3	0.03	5912	65.74	160, 63*	1.78	6138	68.25
18	1	0.01	883	9.82	2	0.02	886	9.85
24	1	0.01	1278	14.21	0	0.00	1279	14.22

36	0	0.00	46	0.51	0	0.00	46	0.51
Total							8996	

\* Has OA fee requirements

## Conclusion & Discussion

Our investigation of institutional OA policies showed 3 with OA policies (one mandated, and two suggested), 2 with drafts in review but not published, and with the remainder without a policy. Those with policies emphasize compliance with funding organization policies. Despite the absence of university-wide OA policies, all institutional libraries offer extensive education and support for OA publishing, with many institutional departments adopting their own OA policy or commitment statement. More analysis is needed to understand the effectiveness and adoption of statements over policies and how it affects IR usage.

In summary, due to the lack of university-wide OA policy, there is insufficient evidence to suggest how IR usage by institutional members may be affected by qualities of an OA policy. We show that publisher policy information is incomplete on Sherpa Romeo and a quarter of policies (24.6%) with complete information exceed the 12-month limit set by funding organizations affecting accepted and published manuscripts versions. Disciplinary differences in IR usage reflect similar differences seen in prior studies investigating OA publishing and self-archiving. Our study contributes to knowledge of IR usage by Canadian institutions with 0.5-10.7% (mean 4.16%) of articles attributed to an institution found in their respective IR.

Unlike in the USA, where the lack of a centralized federal system can be seen as an obstacle to more efficient university policies (Mering 2020), Canada has the possibility to use Tri-Council and the FRQ as a lever to increase self-archiving in IRs. As the FRQ changed to the Plan S model in March 2023, this study may be useful as a prior data point in determining the future effectiveness of IR deposition of Green OA over longer time periods to address the limitations of a short time frame during recent policy implementations and change in Canada. Ultimately, the adoption of OA in general and particularly the use of IRs for self-archiving is a behavioural change problem, and future research will be able to shed light on the effectiveness of these policies.

## References

- Aleixandre-Benavent, Rafael, Antonia Ferrer Sapena, Silvia Coronado Ferrer, Fernanda Peset, and Alicia García García. 2019. "Policies Regarding Public Availability of Published Research Data in Pediatrics Journals." *Scientometrics* 118 (2): 439–51. <https://doi.org/10.1007/s11192-018-2978-1>.
- Archambault, Eric, Olivier H Beauchesne, and Julie Caruso. 2011. "Towards a Multilingual, Comprehensive and Open Scientific Journal Ontology." In *Proceedings of the 13th International Conference of the International Society for Scientometrics and Informetrics (ISSI)*, 66–77. Leiden, Netherlands: Noyons, B., Ngulube, P., & Leta, J. (Eds.).
- Boulton, Geoffrey, Heide Hackmann, David Kaplan, and Marcos Regis da Silva. 2020. "Open Science for the 21st Century." International Science Council. [https://en.unesco.org/sites/default/files/isc\\_paper\\_for\\_unesco\\_open\\_science\\_consultation\\_2020.pdf](https://en.unesco.org/sites/default/files/isc_paper_for_unesco_open_science_consultation_2020.pdf).
- Creaser, Claire, Jenny Fry, Helen Greenwood, Charles Oppenheim, Steve Proberts, Valérie Spezi, and Sonya White. 2010. "Authors' Awareness and Attitudes Toward Open Access Repositories." *New*

- Review of Academic Librarianship* 16 (sup1): 145–61.  
<https://doi.org/10.1080/13614533.2010.518851>.
- Larivière, Vincent, and Benoit Macaluso. 2011. “Improving the Coverage of Social Science and Humanities Researchers’ Output: The Case of the Érudit Journal Platform.” *Journal of the American Society for Information Science and Technology* 62 (12): 2437–42.  
<https://doi.org/10.1002/asi.21632>.
- Larivière, Vincent, and Cassidy R. Sugimoto. 2018. “Do Authors Comply When Funders Enforce Open Access to Research?” *Nature* 562 (7728): 483–86. <https://doi.org/10.1038/d41586-018-07101-w>.
- Mering, Margaret. 2020. “Open Access Mandates and Policies: The Basics.” *Serials Review* 46 (2): 157–59. <https://doi.org/10.1080/00987913.2020.1760707>.
- Moskovkin, Vladimir M., Tatyana V. Saprykina, Marina V. Sadovski, and Olesya V. Serkina. 2021. “International Movement of Open Access to Scientific Knowledge: A Quantitative Analysis of Country Involvement.” *The Journal of Academic Librarianship* 47 (1): 102296.  
<https://doi.org/10.1016/j.acalib.2020.102296>.
- Paquet, Virginie, Simon van Bellen, and Vincent Larivière. 2022. “Measuring the Prevalence of Open Access in Canada: A National Comparison.” *The Canadian Journal of Information and Library Science* 45 (1): 1–21. <https://doi.org/10.5206/cjilsresib.v45i1.14149>.
- Pinfield, Stephen, Jennifer Salter, Peter A. Bath, Bill Hubbard, Peter Millington, Jane H.S. Anders, and Azhar Hussain. 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): 2404–21. <https://doi.org/10.1002/asi.23131>.
- Priem, J., Piwowar, H., & Orr, R. (2022). *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts* (arXiv:2205.01833). arXiv.  
<https://doi.org/10.48550/arXiv.2205.01833>
- Rivest, Maxime, Etienne Vignola-Gagné, and Éric Archambault. 2021. “Article-Level Classification of Scientific Publications: A Comparison of Deep Learning, Direct Citation and Bibliographic Coupling.” *PLOS ONE* 16 (5): e0251493. <https://doi.org/10.1371/journal.pone.0251493>.
- Simard, Marc-André, Gita Ghiasi, Philippe Mongeon, and Vincent Larivière. 2022. “National Differences in Dissemination and Use of Open Access Literature.” Edited by Alberto Baccini. *PLOS ONE* 17 (8): e0272730. <https://doi.org/10.1371/journal.pone.0272730>.
- Tummon, N., & Desmeules, R. (2022). How Open Is the U15? A Preliminary Analysis of Open Access Publishing in Canadian Academic Libraries. *Journal of Librarianship and Scholarly Communication*, 10(1), Article 1. <https://doi.org/10.31274/jlsc.13831>
- “UNESCO Recommendation on Open Science.” 2021. 7, place de Fontenoy, 75352 Paris 07 SP, France: United Nations Educational, Scientific and Cultural Organization.  
<https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>.

---

<sup>i</sup> References to documents that are not indexed in OpenAlex are not included in the data.

<sup>ii</sup> [www.unpaywall.org](http://www.unpaywall.org)