Summary Report:

Where Next for Repositories? An open national forum sponsored by CARL in association with COAR

November 10, 2016 – Ottawa, Ontario

In conceiving of this open national forum, the Canadian Association of Research Libraries (CARL) and the Confederation of Open Access Repositories (COAR), in collaboration with the institutional repository (IR) community of Canada, aimed to strengthen the role and momentum for a Canadian network of repositories, to inform the community about the current state of repositories internationally, to discuss new and evolving roles for repositories and how they can support library operations, and to develop an agenda for the future of repositories in Canada.

In addition to being the first pan-Canadian meeting entirely devoted to discussing institutional repositories, a deliberate decision was made to bring together library directors as well as those working as repository managers and scholarly communications librarians. This allowed the discussion to span beyond the administration of the repository to reimagining of what such a space could become. The enthusiasm generated by this unprecedented event was extraordinary, with participants recognizing that repository managers often feel isolated in their work, and expressing a strong desire to develop a community of practice, and to build strategic collaborations to move Canadian repositories forward.

The event was organized by Kathleen Shearer (COAR), Jeanette Hatherill (University of Ottawa), Leah Vanderjagt (University of Alberta), Susan Haigh and Lise Brin (CARL), under the guidance of CARL's Advancing Research Committee. Hatherill and Vanderjagt planned the preceding Repository Managers' Meeting on November 9, which served as an opportunity for more than forty repository managers from across the country to share successes, strategies, concerns and questions.

Leveraging the Network of Repositories to Create Change in Research, Teaching, and Community Engagement – Leslie Chan

In launching into his keynote, Chan suggested an alternate title for his talk: "Leveraging the Network of Repositories to Create Change in Scholarly Communication." He went on to discuss how this could be accomplished, advocating for a "repository by the people for the people," one that is built to serve what universities generally claim as their largest mission – embracing public enquiry for the public good. He advocated for "changing the narrative from

the IR as storefront to a university's research prowess to

the IR as a showcase of a university's commitment to its public mission."

Chan went on to suggest that reimagining the institutional repository requires that we simultaneously reimagine scholarship itself: that the notion of scholarly outputs needs to be expanded and updated. He took up the arguments put forward by Sir Peter Medawar, who suggested that the scientific paper is an easily commodified entity, and urged us to broaden the scope of materials welcomed into institutional repositories to include all forms of teaching and scholarly output.

International Trends in Open Access and Repositories – Kathleen Shearer

Much like Leslie Chan advocated in his talk, COAR's vision is to create global knowledge commons through repositories. Shearer spoke of recent efforts by COAR's Next Generation Repositories Working Group to identify future directions and functionality for repositories and figure out how we can collaborate internationally with a common vision. Technical interoperability (standard vocabularies and metadata) and the creation of communities of practice are essential to such collaborations. Data exchange across regional repository networks (e.g. OpenAire, SHARE and LA Referencia) is also an important strategy for building in sustainability, encouraging common standards, and creating a brand within a national or regional context.

Shearer suggested that smaller, local journals have more in common with repositories than with commercial international journals, and that given this common mission we should endeavour to work together. With international journals influencing what subjects are seen as worthy of publication, the global south is feeling acute pressure to publish research that is not relevant to their region. Many COAR members are not in favour of article processing charges (APCs) as the main model for open access, which is why COAR advocates for strengthening and adding value to repositories. Value-added services could include: enabling data synching, building peer review into repositories, ensuring that repositories are better integrated into workflows, and standardized usage measures that iteratively replace what exists in big publishers.

Canadian Repositories Now: A National Snapshot – Jeanette Hatherill

Hatherill reported the results of a survey sent out in advance of the forum to those who were participating in the repository managers' meeting, with some comparisons to similar data gathered by CARL in 2004:

2004

- 7 CARL institutions had IRs

- 5 out of 7 of these had fewer than 200 items

- TSpace was the third largest DSpace repository in the world in terms of content, with 2,400 records

2016

- 33 institutions (not all CARL members) have repositories

- These 33 IRs collectively contain approximately 939,165 items

- Two of the repositories (UBC and Canadiana.org) have over 200,000 entries each

These figures clearly demonstrate how much the Canadian repository community has expanded and evolved over the past 15 years. When asked to list their biggest concerns, repository managers reported that content recruitment remains a huge barrier. IT infrastructure (especially in terms of hosting new types of materials) was also mentioned. Other concerns raised include: university community and administration buy-in; workflow and resourcing; metadata and interoperability.

Concordia University Open Access Senate Resolution – Guylaine Beaudry

Beaudry described the origins of Concordia's open access mandate, stemming from its hosting of the Federation of Social Sciences and the Humanities' Congress in 2010, and the desire to retain materials from Congress for posterity. She described how Gerald Beasley, Concordia's University Librarian at that time, presented at numerous committees, and guided the discussions and concessions that resulted in the final text, which requires deposit for peer-reviewed journal articles only. Open access at Concordia now exists in three distinct layers: the repository, an authors' fund, and now a fully open access university press. Beaudry asserted that the Concordia repository, Spectrum, is doing very well (with 14,000 documents and over 1 million downloads per year), and that open access is very much part of the discourse at Concordia – in fact, the President recently stated that "OA is part of Concordia's DNA."

Approaches to Tri-Agency Mandate on Open Access – Pascal Calarco

In May 2015, Windsor passed a resolution at Senate that requires deposit of tri-agency funded articles and encourages the deposit of everything else into Windsor's institutional repository, Scholarship at UWindsor. The policy allows for simultaneous deposit in both institutional and subject repositories, and recognizes the difficulties posed by embargoes.

Drawing on Windsor's experience and considering what needs to be done to apply the resolution at scale, Calarco suggested that libraries have a responsibility to lessen the burden of compliance for their institutions' authors. One of the biggest challenges currently is that the post-print is effectively an ephemeral document, "so the clock is ticking to try to get that version." His advice for winning over faculty members is to advertise to them the benefits of publishing open access, the importance of depositing for knowledge mobilization and towards the common good.

Against Integers or, What We Are not Doing – Dale Askey

Askey's talk focused on the download statistics that are generated from within repositories, and cautioned us to question their accuracy. When he started looking closely at those generated by his institution's IR, MacSphere, he became convinced that the bot detection and filtering was far from adequate. Askey recommended that instead of contenting ourselves with distributing "feel good integers," we should develop standard filtering rules for repository data (as is being done within IRUS-UK) to ensure that it be reliable and comparable across institutions.

Institutional Repositories & the Current Research Information System (CRIS) – Rosarie Coughlan

Coughlan described efforts at Queen's University to integrate the institutional repository, QSpace, with the Research Office's newly adopted Current Research Information System (CRIS). This project, which is still at the development phase, will allow researchers to deposit into the IR via a "somewhat mediated service" (library staff will take care of related tasks such as embargo checks with publishers). Similarly, some integration across the CRIS, IR and Canadian Common CV (CCV) is planned, allowing a hyperlink to QSpace to be entered for publications in CCV. Coughlan hopes this interoperability will address some barriers to deposit (reduce time and admin burden) while also adding incentives for researchers.

UBC's Open Collections: Improving Discoverability and Delivery of Repository Contents – Bronwen Sprout

Sprout gave an overview of the four repositories that are jointly searchable and discoverable via Open Collections, UBC's locally implemented platform: ContentDM, DSpace, AtoM (for finding aids and some archival collections), and

Dataverse. The platform uses Elasticsearch to search metadata across the four databases, and provides a consistent view of objects based on file format rather than original source, leading to a better user experience. Years of metadata cleaning along with applied user testing were required to make this clean user experience a reality.

National IR Managers Meeting – Leah Vanderjagt

Vanderjagt provided an overview of the Repositories Managers' Meeting held November 9, 2016 at the University of Ottawa. This event, organized by Vanderjagt and Jeanette Hatherill, brought together 43 repository managers from institutions across Canada. Several would describe themselves as "lone rangers" while others work as part of large repository teams. The repository managers' meeting on November 9th covered a very large range of topics. A detailed summary from the IR management community in Canada is forthcoming. Sample observations from the day include:

- More face-to-face national meetings are needed to reduce work isolation and promote collaborative problem-solving;

- It is important to have a collaborative working space for sharing challenges and successes;

- There is some concern re: the potential existence of "OA fatigue" among librarians who work outside of scholarly communication given other new emerging library services;

- There is a need for reliable statistics that can be used as a tool for buy-in, with caveat that repository assessment is much more than download counts;

- Mediated deposit work shouldn't have to live solely within the repository unit – there is a need to operationalize new services across library organizations, and break down work 'silos' around IRs;

- IR managers would like to align repository services with university Research portfolio goals and services at their universities;

- Integrating IR services with campus CRIS is of tremendous interest;

- Interoperability discussions at a national level need to address the need for local development time to help implement needed specifications and standards;

- Repository managers work very closely with CVs/citation list processing workflows and should be included in discussions around ORCID;

- Everyone is looking for automated ingest workflows;

- SHERPA/RoMEO categories aren't always that useful or sufficiently nuanced, and can create difficulties around automatic ingest;

- Acquiring post-prints is a difficult content recruitment problem – and some researchers don't want these shared publicly;

- There seems to be a lot of success in getting unpublished material into repositories, and these collections are valuable. In thinking about future roles for repositories, grey literature should be part of the discussion.

Group Discussions

Each table was asked to reflect on two of the following four questions and to report their findings back to the full group. The attendees' comments are summarized below each question.

1. What would a Canadian network of repositories look like? What are the mechanisms that would connect the network? And, how would we go about connecting internationally?

    - It would be important to rely on standards (for metadata, interoperability);

    - A Canadian network should use international standards so that linking internationally is a possibility from the outset, regardless of whether we end up doing so (national infrastructure based on international standards);

    - A network should use linked open data, ORCID IDs, good authority files;

    - A network could be a "nodal library," a "network of networks" like ARCA;

    - Creating redundancy for Google Scholar would be wise;

    - A question was raised about whether non-academic repositories (e.g. government) should be part of this network.

2. How do we develop a strategic vision for repositories in Canada? Building on ideas arising today or emerging in other settings, what do we need to do collectively to develop our vision and who will lead?

    - This effort should be Canada-wide and bilingual;

    - Similar international efforts should be studied;

    - The project should start with a small demonstration project (Ontario has some consolidation so could be a good place to start);

    - UBC's Open Collections could be a good model for inspiration;

    - It is important not to undermine, or to neglect to consider, subject repositories;

- Leadership is an important concern and a possible role for CARL, but the project needs to include non-CARL institutions, and should have non-library partners such as [CARA](#) (Canadian Association of Research Administrators);

- The Slack channel and community of practice are important elements to keep people connected;

- This initiative should be framed and informed by user needs; academics and repository managers should be at the table.

3. Should we expand the scope of repository operations to include a broader range of content? If so, how can we do this so that we develop collections of real value to research community? What should be our priorities, and how do we tackle these?

- The answer to the first part of this question was a resounding yes – make the IR an "intellectual archive" with all sorts of materials, including grey and white literature, published or unpublished, successful or not, open textbooks, learning objects, oral histories, multimedia, maps, technical reports, digital humanities projects;

- We should develop means for capturing versioning within repositories;

- Perhaps there could be a unified collection strategy across the country;

- We need to work to respond to researcher needs; whenever possible we should try to not turn people away;

- Ingest first, without necessarily exposing the objects if there are ethical or copyright issues;

- There was an acknowledgement that even though DSpace does not accept many types of content, we shouldn't let technology drive content.

4. How do we ensure that our repositories are not isolated within our institutions, and that they are integrated with other systems such as CRIS, researcher profiling systems, research data repositories, etc.?

- It's important to be at the table, to position the library for its strengths;

- We need to know what's going on our campuses; and we need to have conversations with IT, the Research Office, and Graduate Studies, and understand the scope and expertise of those departments.

Conclusion

The day's discussions were stimulating, hopeful, and purposeful. A renewed energy and enthusiasm within our community came to light, and the vision put forward early in the day by Leslie Chan – "repositories for the people, by the people" – provided an underlying principle that guided the day's explorations.

Participants came away from the Forum seeing their repositories in a new light, as nodes within a national and potentially global scholarly communication network, where the contents could encompass much more than articles.

There were two major areas identified for repository development:

1. **Improve the functionality of repositories** with a focus on interoperability and interconnectedness

2. **Support and encourage the development of value added services**, such as commentaries and peer-review, usage measures and tools.

Specific ideas identified for follow-up action

- Develop standards for IR usage data (block same bots, count usage consistently in order to have comparable data across institutions);

- Coordinate uptake of interoperability standards across Canadian repositories;

- Identify best practices for collection development, possibly leading to a unified national collection strategy for Canadian institutional repositories;

- Look to develop a national aggregator

- Develop and strengthen a community of practice around repositories and scholarly communication;

- Develop strategies to demonstrate the value of repositories (e.g. track how the IRs are being used, and for what purposes).

In the coming months, the CARL Office will be working with the leads from the managers' meeting as well as with COAR to determine how best to advance these suggestions.