The Ottawa Hospital
Centre for Journalology
Affiliated with uOttawa

uOttawa

CREATING AN AUDIT AND FEEDBACK LOOP TO MONITOR AND ENHANCE DATA SHARING IN CANADA
A WHITE PAPER

Prepared by:

Dr. Felicity Tayler (University of Ottawa Library)

Dr. David Moher (Ottawa Hospital Research Institute & University of Ottawa)

Dr. Kelly Cobey (Ottawa Hospital Research Institute & University of Ottawa)

Submitted to:
NDRIO: Call for white papers on Canada's future DRI ecosystem
14 December 2020

DOI: http://dx.doi.org/10.20381/s9px-zq30

Executive summary

Data sharing is an integral part of open scholarship and research integrity (i.e., trustworthiness of research). Canada is embracing data sharing practice and will roll out data management mandates in the near future. The success of these policies can be assured if discipline specific training in data sharing is provided, local infrastructure is created, and a data sharing dashboard is created (and assessed) to monitor the implementation of data sharing practices. Such actions will place Canada at the top of the innovation and global leadership list. NDRIO is perfectly positioned to play a central role to enable these activities.

This white paper makes the following recommendations:

- That NDRIO initiate a funding competition open to the researcher community to create a dashboard to monitor data sharing and other open scholarship practices.
- That a parallel to funding competition be created to for researchers to initiate discipline specific national training resources.

Identifying an auditing gap between DRI tools, policy, and data sharing behaviour

A foundational aspect of research integrity is trustworthiness. It is only possible to examine trustworthiness if research outputs can be seen (open access) and reproduced (data sharing). The Core Trust Seal for data repositories is an example of a certification process that ensures trustworthiness in the international DRI ecosystem. **This white paper will argue that the research community in Canada needs an automated digital dashboard to audit open scholarship practices, including open access publishing and data sharing, thereby promoting trustworthiness not only in repository infrastructure, but in research(er) practice.**

The Canadian Tri-Agencies have an open access mandate for their grantees. This ensures that research is made freely available for others to read and build upon. They also have a draft Research Data Management Policy which will be rolled out shortly. This policy will require researchers to generate data management plans (DMPs) for their projects, with a strong preference for open data sharing as part of the lifecycle of publicly funded research. The value of these policies is clear: they help ensure transparency and foster innovation through reducing barriers to access information and data. However, there is limited knowledge on the degree to which Canadian grantees and the larger research ecosystem adhere to these and other open science practices. In the absence of audit and feedback of these policies, we don't know if we are doing a good job, we can't compare ourselves internationally, and we can't set benchmarks or implement interventions to drive improvements in our practices.

We maintain that the research community needs an automated digital dashboard to audit open scholarship practices, including open access and data sharing. Drawing on existing work in Canada (e.g., CARL; Portage Network) and Europe (e.g., the CARL-OpenAIRE collaboration) the dashboard could provide a real time audit to funders as to what fraction of their grantees are meeting open science mandates and whether education and training interventions are needed. The tool would also benefit other stakeholders

in the research ecosystem. Academic institutes could track their researchers open scholarship practices. The proposed dashboard would enable funders and institutions to monitor the degree to which its members employ OS practices (e.g., data sharing; see **Figure 1,** top panel). Customizable to a university or funders unique priorities and needs, the dashboard could enable better tracking of progress toward achieving realistic transparency targets among its members (e.g., achieving 50% open data by 2025) and to identify department(s)/division(s)/researcher members that are (not) meeting these transparency standards (see **Figure 1,** middle panel). The dashboard would allow individual researchers to gauge the success of their own open scholarship practices. Once an automated digital dashboard to track open scholarship practices was developed and implemented, it would be imperative that it be evaluated. There is substantial economic and intellectual waste if an element of key digital research infrastructure is implemented in the absence of evaluation.

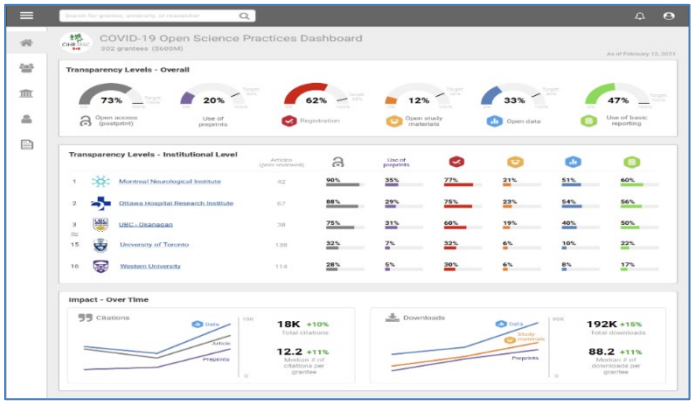


***Figure 1**. Visualization of the types of information the proposed dashboard could track from CurateScience.org and LeBel et al. (2018). Here, we represent five institutions hypothetical performance on a range of open scholarship practices, including open access and data sharing. The dashboard represents an audit of these institutions open scholarship practices for CIHR-funded projects addressing COVID-19. The left column highlights options to view more detailed data at the institutional and individual researcher level.*

Barriers to data sharing that would be addressed if this DRI gap was filled

Tenopir and colleagues (2011) have shown that researchers are not reliably sharing their data and associated information (e.g., analytical code) nor are they consistently sharing research outputs in an open access format. The COVID-19 pandemic has highlighted this problem; at a time when open data sharing could greatly accelerate discovery to reduce global suffering, data sharing is not happening sufficiently. Torres-Salinas and colleagues (2020) have shown that since the pandemic began, rates of open sharing have gradually reduced with paywalls restricting access to information becoming more common. If study findings and data are shared behind a paywall, this limits its potential usefulness and drives inequity. Other concerns include inappropriate data sharing leading to privacy breaches or liability risks: examples of biomedical researchers inadvertently sharing identified patient data have been discussed, see the *Sharing, Discovering and Citing COVID-19 Data and Code* (2020) NIH webinar for some alarming examples. No researcher wants this to happen, and this is a clear signal that there is a need for discipline specific training for researchers. Among studies which do share research data, data is often shared in a format that is not FAIR (Findable, Accessible, Interoperable, and Reusable). The *RDA Covid-19 Guidelines and Recommendations* (28 May 2020) and the Portage Network *Guide to COVID-19 Rapid Response Data*

Sharing and Deposit for Canadian Researchers (21 September 2020) are two examples of initiatives from RDM service providers that hastily mobilized in a time of global crisis to address these significant researcher training gaps.

A fundamental principle of implementation science is that there needs to be an audit and feedback loop to change behavior. The development and implementation of an automated open scholarship digital dashboard to track behaviours can provide this missing link. Let's imagine university "A" wants 80% of their researchers to share their data but the digital dashboard indicates that the university is only performing in the 50% range. In essence, the proposed tool could provide an audit and feedback loop regarding a gap in behavior. As part of the development and implementation of the dashboard we think it would be essential to provide access to open educational resources to improve open scholarship skills including data sharing. Educational interventions (and other social psychological ones) can be implemented and continually monitored. There are existing educational resources on data sharing within the Canadian ecosystem: Research Data Canada and the Portage Network have diverse educational resources available to help researchers adopt data sharing practices, but they lack a coordinated curriculum, and these available training materials were not created with a researcher-centred discipline-specific approach. These existing educational resources could be leveraged to train researchers as part of the roll-out and monitoring of the proposed open scholarship dashboard. However, a recent literature review of best practices for RDM train-the-trainer models by Tayler and Jafary (2020) indicates that future iterations of national training entities within the context of the NDRIO must meaningfully engage with researchers to identify barriers and facilitators, unique to distinct disciplines, with regards to implementing data sharing. In tandem with a tool such as the proposed dashboard, this information can be used to evaluate gaps in the existing educational resources and to create new discipline specific content, when relevant.

The literature shows that researchers must have clear incentives to undertake necessary data management training. The current system of academic incentives places emphasis on the quantity or research outputs, rather than their quality. This system is in direct conflict with behaviors that support transparency and maximizing the societal benefits of research. The dashboard would also provide the infrastructure institutions need to incorporate open scholarship quality practices into researcher assessment (e.g., considered in promotion and tenure file) and establish incentives and rewards. For example, the Technical University of Delft (The Netherlands) allows 'data champions' to include data sharing mentoring and applications as part of their promotion and tenure portfolio. The Hong Kong Principles developed at the 6th (2019) World Congress of Research Integrity advocates for research organizations incentivizing and rewarding open scholarship practices as a central part of researcher assessments for promotion and tenure. Even though the principles were only published during the summer (2020), individual researchers and research organizations, including the Ottawa Hospital Research Institute, have endorsed them. Incentive and reward structure are crucial to embed data sharing values.

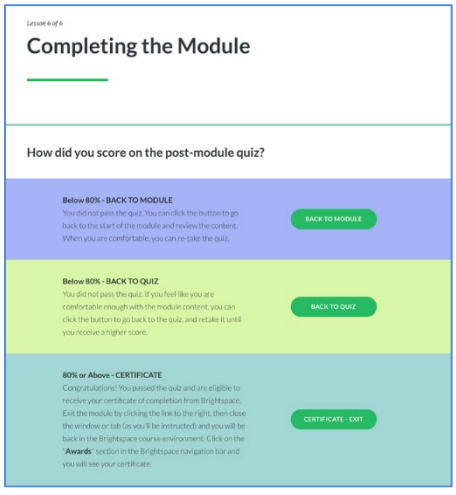
Imagining a future for trustworthy DRI ecosystem that includes auditing

NDRIO and other leaders in the Canadian data sharing ecosystem need to prioritize audit and feedback of their policies, tools, and training to ensure they are moving toward achieving their strategic goals. Similarly, any policies, tools and training implemented requires evaluation. There is little point in introducing 'interventions' that do not have their intended effect. Larivière and Sugimoto (2018) indicated that CHIR's open access policy decreased 20% between 2014 and 2017 due to less enforcement and infrastructure.

We recommend that NDRIO initiate a funding competition open to the researcher community to create a dashboard to monitor data sharing and other open scholarship practices. The successful applicants must be able to show that the tool they have produced accurately captures the metrics they are measuring and reporting upon. The successful applicants would require disciplinary knowledge of bibliometrics and be able to produce a prototype working dashboard, implement it across stakeholders (e.g., funders and research organizations), and evaluate its utility.

While the proposed dashboard is an essential tool to provide evidence-based audit and feedback regarding data sharing, the implementation of data sharing practices will be more assured when the Canadian research ecosystem has access to comprehensive (bilingual) open online discipline specific educational resources. For example, study registration is a core open scholarship practice in biomedicine but likely of less relevance in planetary science.

Data sharing educational modules might expand on current Portage Network and Research Data Canada training resources to include a range of researcher-driven disciplinary-specific topics. A disciplinary lense can be applied to training on topics such as the current status of data sharing in Canada, new RDM initiatives and recommendations, what are the FAIR principles, developing a data management plan, data sharing ethics and law, and funder guidance on data sharing. Each module should include a pre-test and post-test of data sharing knowledge. Learners achieving an 80% knowledge threshold would be offered a data sharing certificate. This certification could be included as part of the researcher's promotion and tenure portfolio. The tests provide a baseline of knowledge (pre-test) and knowledge acquisition (post-test) about data sharing and ensures a minimum standard of knowledge among the learners. The training modules could look similar to the existing OHRI Data Sharing Course (screen shot to left in **Figure 2**).



We further recommend that in parallel to the funding competition for the development of an auditing dashboard, the best way to engage the expertise of researchers in the creation of training resources would be for NDRIO to create a funding competition for national training. This incentivizes researchers through a model of award and career advancement that is recognized by their disciplinary peers. The successful applicants would be adept at articulating the training gaps of their disciplinary community, completing a

scoping review to ensure existing data sharing resources are included in the training platform, include training modules on Indigenous data sovereignty, equity, diversity and inclusiveness, and incorporate an evidence-based assessment of the proposed training.

Funders might promote a data sharing budget line item for grant applicants in much the same way as Author Processing Charges are now included. However, data sharing requires more local infrastructure to help enable this practice. Even if researchers are keen to share their data, the particulars of how to do this are not currently available at most Canadian research organizations. For example, if the data sharing budget line item was $1500 dollars per application, an equal dollar amount would be matched by the applicant's institution. These monies would be used to help build a comprehensive data sharing infrastructure, such as data champions. The matched funding period to be limited, such as for three years by which time research organizations would have built the necessary infrastructure.

DRI audit and feed-back loop call to action

NDRIO has the ability to play a key role in driving the development of the proposed dashboard and in coordinating Canadian-wide discipline specific training services related to data sharing and other related open scholarship practices. We have identified few public funding opportunities to fund this type of core infrastructure. CANARIE and CFI funding streams are narrowly focused on technical infrastructure development, without provision for the necessary researcher training and incentivization to encourage user engagement with data sharing infrastructure; disciplinary-specific funding streams like CIHR or the COVID-19 Rapid Response Funding are limited because the disciplinary peer-community is focused on research, with less emphasis on pedagogy and infrastructure concerns; the SSHRC RDM Capacity Building Initiative is a temporary funding stream, and while the focus on knowledge mobilization through an event-based model lends itself to training initiatives, the envelope is not a large enough. This is despite the clear benefit an initiative such as the proposed dashboard would provide to several stakeholders (e.g., funders, research organizations and researchers) in the Canadian research ecosystem.

A researcher-centred grass-roots approach is essential to drive the culture change needed for the broad adoption of data sharing. We view the automated open scholarship dashboard as a catalyst to creating a network of national training entities that are responsive to discipline specific data communities, as defined by Springer and Cooper (2020). These communities could represent a network of researchers within a specific discipline dealing with the same type of data or using the same methodological approaches. With baseline information provided by the dashboard about how various disciplines are doing in regard to data sharing, including cross-disciplinary and international discipline comparison data, we can start a discussion about how we drive improvements to meet the diverse needs of distinct disciplines or to address nuances in barriers to data sharing. If researchers are empowered to lead the discussions around barriers/facilitators to data sharing within their discipline, and to highlight, modify, or create training for their peers, it is more likely to resonate within their community. At present many researchers are skeptical about data sharing or don't see it as relevant to them. Competitive funding needs to be made available to support teams of researchers to establish discipline specific national training entities. An important requirement of funding for discipline specific national training entities on data sharing would be to embed assessment in their offerings to measure the quality of the training.