The buzz around the word OPEN is getting louder

Open access …  Open Data…  Open Science…  Open Education…

I am very pleased to be here today to talk about the Canadian landscape and some of the ways libraries are getting involved in about open access and data sharing and what we are doing here at McGill Libraries.

Before I start my presentation, I want to ask you if the name Scott S. Reuben rings a bell?
Scott S Reuben is listed on the Retraction watch leaderboard list.

He was once held in high esteem for research he conducted while employed as a professor at Tufts University on anesthesiology and pain medicine.

In 2009 he was convicted for fabricating data in 21 major publications over a period of 15 years. (Shafer 2009)

Reuben’s work was and still is widely quoted. At the time of the scandal, he had been cited nearly 1200 times and his work was referred to in several guidelines and systematic reviews. (Rittner et al. 2009). With his now 25 retractions, he is not even close to being at the top of the Retraction watch leaderboard list.
The number one spot belongs to Yoshitaka Fujii with more than 180 retractions.

Like Reuben and Fijii, the others on the retraction watch list were widely respected researchers

- They publish prolifically
- Got hire, promoted and received awards; and
- They received more grant funding and speaking engagements.

That is .... until someone took a look at their data.
The number of retracted articles seem to be increasing. More than 2000 papers are currently indexed as retracted in PubMed.

According to a researcher named Fang, the potential reasons for retractions are errors, duplicate publication and plagiarism; more than 40% of retractions, however, are attributed to fraud (Fang et al. 2012).

In drug therapy studies this proportion is in excess of 70% (Samp et al. 2012).

New thinking is referring to the article as just the advertising of the research data. Basically, if we want to trust a discovery, we need:

- A summary of the findings (the peer reviewed article);
- The data; and
- The code
Arguments for sharing data

<table>
<thead>
<tr>
<th>Scientific argument</th>
<th>Ethical argument</th>
<th>Compliance</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Validation</td>
<td>• Greater research transparency</td>
<td>• Some scientific publications require that data be available for the scientific community before the article is published</td>
</tr>
<tr>
<td>• replication</td>
<td>• Results in better health care</td>
<td>• A growing number of funding agencies require that data be deposited in a public archive</td>
</tr>
<tr>
<td>• re-analysis</td>
<td>Public funding argument</td>
<td></td>
</tr>
<tr>
<td>• new analysis</td>
<td>• Taxpayer paid for the products of research.</td>
<td></td>
</tr>
<tr>
<td>• re-interpretation or inclusion into meta-analyses</td>
<td>• Efforts to ensure data are archived and preserved, increases the long term value publicly funded research</td>
<td></td>
</tr>
</tbody>
</table>

There are many arguments for why sharing is a good thing

For science
• Is means Validation
• replication
• re-analysis
• new analysis
• re-interpretation or
• inclusion into meta-analyses

Others argue that transparency benefits society
**And Taxpayers feel that they** paid for the products of research and as such ought to have the right to access and re-use that same data over time.

And in terms of compliance:
• Some scientific publications require that data be available at the same time as submission; and
• **More and more funding agencies** require that data be deposited in a public archive
In Canada, as elsewhere, data sharing practices are very discipline-specific. In some fields—such as genomics and proteomics—data archiving and sharing is the norm. In other fields, no such traditions exist. Regardless if the discipline there is, a growing awareness across the scholarly community that there are significant potential benefits in making research data available for re-use.

In 2004, Canada adopted the OECD Declaration on Access to Research Data From Public Funding. It declared publicly-funded research data as a public good that should be openly available.

- Canada did not take any real action at the time but other policies governing various aspects of the management of research data began to appear. (listed above)

Most recently, the Tri-Agency Statement of Principles on Digital Data Management

Around the same time as this Statement was probably being drafted. Libraries stated paying attention to the trend and in 2014, the Canadian Association of Research Libraries, launched Portage. Portage is a network that work within the library community to coordinate expertise, services, and technology in research data management, seeking to collaborate with
other research data management stakeholders.
The Portage website contains content and resources for researchers, librarians, and data stewards.

I will come back to the DMP assistant in another slide. (the orange bit)

so I want to focus quickly on the recent NEWS on the portage website.

On October 24th, Portage announced that they have joined forces with Compute Canada to build a national research data platform.
It will have three components: repository, preservation system, and discovery engine.

The Tri Agency documentation suggests that Policies that require data sharing will not be implemented without corresponding infrastructures and other support mechanisms being in place first.

There are some large scale international data repositories in certain fields, such as PubChem, GenBank, Protein Data Bank, BUT Repositories do not exist for all subject areas, and the vast majority of research data still rests on researchers’ hard drives or locked in cabinets. This is good news for Canadian researchers who do not have a easy solution for archiving their data.
On the McGill campus, we have a lot of work to do to get ready for any potential data sharing policies from funding agencies.

It is going to take a village. It is going to take the development best practices, workflows and the deepening of relationships.

The Library is engaged in making sure we do our part in getting ready

Our current strategy focuses on:

1. Upskilling Librarian/Staff
2. Relationship building (people in the circle above)
3. Providing Researcher support, advocacy, outreach, promotion and education about the RDM lifecycle and
4. connecting a core suite of research data services in one place...
Enter the new page on the library website for Library Research Data Services

The library has long history of offering services that provide access to data housed elsewhere (e.g., Statistics Canada data) through research data centres, *(the brown and green bits)*

For the most part though, data management and data sharing support through the library is in its infancy. *(the blue and red bits)*
We are putting action in to place so that librarians can assist researchers in

- **Planning for data management**
  data management plans & grant applications

- **Preparing data**
  coding, formatting, anonymizing, & more

- **Documenting data** - **METADATA!!**
  describing data so others can understand & use it

- **Archiving data** – **PRESERVATION!!**
  finding an appropriate repository or archive

- **Making data discoverable** – **DISCOVERY!!**
  promoting open access, indexing, helping assign a DOI

In the spring, the library built a customized version of the Portage Network DMP Assistant.

DMP Assistant is a tool that will guide a researcher or research team through a series of questions that need to be thoughtfully answered for a data management plan. McGill researchers can sign in with their McGill credentials and get started on creating a DMP.
If data sharing is something you wish to apply to your research, I encourage you to take the following steps:

1. incorporate data management best practices into your research
2. develop data management plans to guide you throughout the entire lifecycle of a research project and beyond
3. follow the requirements of your funding agency policies and disciplinary standards
4. acknowledge and cite datasets that contribute to your research
If you want to take a stance... consider creating a pledge like the one Erin McKiernan took when she was an early career research and a post doc.
If more people take pledge like hers, we may see fewer situations like Reuben.
Sources


Computer Canada https://www.computecanada.ca/renewing-canadas-advanced-research-computing-platform/rdm/


Portage Network. https://portagenetwork.ca


Sources and links above,.. My data ;-)