Partnership
The Canadian Journal of Library and Information Practice and Research

vol. 10, no. 1 (2015)

# The State of Canadian Library Data

Brian Jackson
Assistant Professor
Mount Royal University Library
bjackson@mtroyal.ca

## *Abstract*

At a time when the Canadian library community advocates for, educates about, and develops infrastructure in support of open data, we have been less than successful in establishing principles and practices that ensure our own data is open and useful. Data about libraries in Canada are limited in scope and often restricted in distribution. If we expect government agencies and academic researchers to devote the resources required to make their data open, we must be willing to do the same. Fortunately, the library community is well equipped to implement a program of open data that would benefit both libraries and the broader community.

## *Keywords*

open data; academic libraries; public libraries; library data

★ ★ ★

There is no shortage of bodies collecting information about libraries in Canada. Academic library data are collected by the regional university library consortia, the Canadian Association of Research Libraries (CARL), the Canadian Association of University Teachers (CAUT), Universities Canada (formerly the Association of Universities and Colleges in Canada), the Canadian Association of University Business Officers, and the Association of Community Colleges of Canada. Individual municipalities, and all of the provincial and territorial departments responsible for library services, collect data about public libraries. Public library data are also gathered by the Canadian Urban Libraries Council (CULC) and the Provincial/Territorial Public Library Council. Data about school and special libraries are gathered by an array of bodies, often in an inconsistent and ad hoc manner. And Statistics Canada publishes a smattering of data about all types of libraries.

Yet there is no authoritative aggregate service providing comprehensive and meaningful data about libraries in Canada. Decentralized data gathering and the resultant fragmentation in distribution, as well as protectionism, privacy concerns, and an absence of data collection standards have all contributed to a situation whereby Canadian library data cannot be easily employed to shed light national library contexts, if the data can be obtained at all.

Both libraries and the Canadian public are currently balancing two frames of discourse about data. In one frame, our online activities form part of the big data gathered by the Communications Security Establishment Canada (CSEC), the Canadian Security Intelligence Service (CSIS), the National Security Agency (NSA), Google, Facebook, and Amazon, and countless other shadowy governmental and private organizations. In this frame, the data are collected for purposes ranging from simple business intelligence, to the development of targeted marketing, to the monitoring and prosecution of potential or perceived criminal activity. With less transparency in the collection and dissemination of these data comes greater suspicion around the intentions of the players involved.

The other frame sees a cultural shift towards large scale data sharing as a boon to intellectual discovery. Data have found their way into mainstream rhetoric about open information; as a result, research data repositories are popping up at higher education institutions, funding agencies encourage grant recipients to release their data, and governments at all levels have complied to some extent with requests that their data be made freely available.

The first frame positions data as a threat; the second, as a public good.

Where do libraries fit when it comes to our own data? The library community has long been mindful of both the imperative of individual privacy and the benefits of open information. In the absence of threats to privacy, though, the community has generally held that the availability of large volumes of free information makes us all better off. The Canadian Library Association's (CLA) Position Statement on Intellectual Freedom states that "it is the responsibility of libraries to guarantee and facilitate access to all expressions of knowledge and intellectual activity" (CLA). Librarians have been fervent in their mission to fulfill this responsibility by fighting censorship, championing open access to scholarly publications, and holding accountable those public agencies that restrict access to information. In a similar vein, librarians now rightly argue that open data can only contribute to the economic and social welfare of Canadians. This discourse would seem to position libraries firmly in the second frame, but in some ways we bear a closer resemblance to the shadowy organizations of the first frame.

## *Barriers to Use*

With such a wide array of organizations collecting information about libraries and an immense volume of resources dedicated to harvesting and organizing data, one might wonder why obtaining national statistics about libraries is such a difficult endeavour. But the current model of data gathering is one of the reasons for the paucity of national data on libraries. Libraries regularly report to the governing or consortial entities to which they belong, but there is inconsistency across reporting channels and no national voice currently taking responsibility for statistics collection and dissemination. So while the data provided by, for example, provincial governments, CULC, or CAUT might be very good, it is always limited in scope, prohibiting comparison across provinces, rural and urban contexts, or university size.

In the United States, while there is a similar diversity of reporting channels, there are organizations that collect data on a national scale. These include the American Library Association and its Association of College and Research Libraries (ACRL) division, the Institute for Museum and Library Services, and the National Center for Education Statistics, among others. All told, it is relatively easy to gather American national information, not only on library characteristics, but on users and use as well.

While the United States does not suffer from the same limitation of decentralized data collection that Canada does, both countries experience varying degrees of another, more perplexing barrier, particularly with respect to academic library data. This barrier stems from a reluctance or inability of bodies that hold data to distribute them openly. As much as libraries have encouraged governments and researchers to make their data available, citing the benefits of informed citizenry, accountability, and reproducibility of research results, we have not put forth the effort ourselves to ensure that interested parties can access our data. In some cases where the barrier of decentralization is not present, access is limited by paywalls or member-only restrictions, or summarized in a way that restricts deep investigation. A number of organizations, including CARL, the Council of Prairie and Pacific University Libraries, the Council of Atlantic University Libraries, ACRL, and Association of Research Libraries protect data in these ways.

In some cases, access restrictions appear as if they are due simply to protectionism and secrecy, although it is more likely that resource and privacy concerns are the primary culprits—organizations that charge for library data are trying to recover the costs associated with data gathering and analysis, while salaries and some other operational data might come with assumptions of privacy from libraries supplying the data. There is no doubt that collecting and organizing data while ensuring anonymity is a resource-intensive process, but we need to ask ourselves if the time of researchers and government departments is any less valuable than ours. If the answer is no, then the least we can do is commit the resources to making our data open as we have asked others to do.

## *Why Open Library Data?*

Apart from the persuasive case for open data made by demonstrating open practices, making library data more widely available would confer a number of benefits on library, academic, and civic communities. There is a vast body of literature the shows how data can be used to demonstrate the value of libraries. In the academic context, Megan Oakleaf's *The Value of Academic Libraries* makes a comprehensive case for tying together library and institutional data to demonstrate the library's impact on students and faculty. Literature on public libraries, as well, is replete with data-driven research on the impact and sound operation of libraries. The Martin Prosperity Institute, for example, was able to show that the Toronto Public Library created more than one billion dollars in economic impact in one year. A national statistical program would open the doors to this type of research on a broader scale.

This research is important because libraries do not exist in isolation; they are inseparable from the communities to which they belong. Opening library data would

greatly expand the breadth of inquiry into library value and operations to include community partners and disciplines that could bring perspectives currently underemployed in the use of library data. We should not underestimate the value of library data to the wider community, nor should we claim a monopoly on the interpretation of those data.

The library community has a long history of grappling with questions around library data. Which data should we gather? How do we define data? How do they define us? How do they limit us? We often fear that our data do not do justice to the context in which we work, that they do not truly define us, and that others will not understand us if they see us only through our data. In a 2000 paper in *IFLA Journal*, G.E. Gorman called for:

> a greater awareness among library professionals that meaningful data are contextual; and that meaning depends on interpretation; that they are derived from variables that are complex and difficult to measure; and that understanding is an inductive process (118).

Our context is our communities, though, and opening the data to our communities is an opportunity to add context that might not exist when we are continually gazing at our navels. We are not the sum of our data, and we have no reason to believe that others will see us only in that way if we grant them access to our data.

In the same article, Gorman also noted:

> Every library or information service is accountable to someone else in that they depend on that someone for funding, for their very raison d'être...If external stakeholders are allowed to dictate data collection needs and presentation standards, then it is totally realistic to expect them to structure these for their own interests rather than for those of the library - and why shouldn't they? (118).

A national program of open data, planned and managed by the library community, would help free us from external definitions while simultaneously permitting deeper and broader reflection about ourselves.

## *Opening and Standardizing the Data*

There is no simple solution to the problem of library data access, but there are three clear steps to ensure that our data are truly working for us. If we want library data to be open and useful, we need to: (1) identify organizations with the resources and scope to undertake a statistical program on libraries; (2) establish national protocols for data reporting among both public and academic libraries (some of which already exist informally); and (3) establish workflows that ensure data are complete, free of information that might compromise individual privacy, and readily available to the public. Fortunately, the library community is well positioned to pursue these goals:

- We already commit a massive amount of resources to the collection and analysis of data. Comprehensive data collection is a requirement of all libraries, the

process of which a standard would be unlikely to substantially change. Any protocol with a chance of widespread adoption would consider and incorporate the work that we are already doing.

- We have expertise in library assessment, data analysis, preservation, information dissemination, and other areas required to create useful, open data. While it is not a small request to ask the community to devote time and expense to an undertaking such as this, there are bodies like CARL and the CLA that are national in scope with the reach to bring the library community together toward a shared understanding of what we should do with our data, and to which members already give of their expertise.

- The infrastructure needed to store, archive, and distribute the data is already available. Storage is cheaper than ever, data repositories have been established at a growing number of universities, and the online spaces of most large library organizations incorporate content management systems that make open distribution simple.

We have much to gain from making library data national in scope and openly available, and little to lose, if done with consideration of what we value and how we are valuable. We cannot continue to cite the benefits of open data and cry foul when public bodies are anything less than fully transparent with their data if we refuse to be transparent with ours. If we do, we convey the message that we have something to hide, and run the risk of being forgotten in conversations about what is most important to the greater community.

## *Works Cited*

Canadian Library Association. “Position Statement on Intellectual Freedom.” *Canadian Library Association*. 18 Nov. 1985. Web. 8 Jun. 2015.

Gorman, G.E. “Collecting Data Sensibly in Information Settings.” *IFLA Journal* 26.2 (2000): 115-119. Web.

Martin Prosperity Institute. *So Much More: The Economic Impact of the Toronto Public Library on the City of Toronto*. Martin Prosperity Institute, 2013. Web. 12 Jun. 2015.

Oakleaf, Megan. *The Value of Academic Libraries: A Comprehensive Research Review and Report*. Chicago: American Library Association, 2010. Print.