Office of the
Information & Privacy
Commissioner
for British Columbia

Protecting privacy. Promoting transparency.

CHECK AGAINST DELIVERY

KEYNOTE ADDRESS TO

SELKIRK COLLEGE

NOVEMBER 17, 2016

DREW MCARTHUR
ACTING INFORMATION AND PRIVACY COMMISSIONER FOR B.C.

Thank you to Selkirk College and Ian Parfitt for the invitation to speak with you today. I am pleased to be here to help mark Global Information Systems Day.

I also understand that you recently celebrated your 50th year – congratulations for all you've done and will continue to do to provide educational opportunities to your students.

I know we have a variety of people in the room: students, faculty, government employees, and GIS practitioners.

First, so we are all on the same page, I want to talk briefly about open government.

You probably have heard that term quite a lot lately.

The purpose and values of open government relate directly to access to information legislation. FIPPA, as I mentioned, gives the public a right to access government records. This usually occurs through a freedom of information – or "FOI" – request.

However, FIPPA does not limit the discretion of public bodies to **proactively** disclose records related to their programs, services and operations.

Open government has many benefits. People are better informed about their government and communities and can identify priorities, help solve problems, and participate in decision making.

And governments benefit too. Allowing access to data with few restrictions promotes entrepreneurship and innovation, which can help the economy.

To demonstrate a commitment to open government, my office has recommended that records be made public by default. The data provided should be structured, machine readable, and freely shared.

Any exception to releasing records should be limited and specific. One of these exceptions is privacy, which I will talk about shortly.

Open Government in B.C.

I'm told that your project is about open government in rural B.C.

Local governments have actually led the way in endorsing open government and in developing and publishing open data catalogues.

Municipalities are closest to the citizens, so they are a natural place for the open data movement to thrive. By proactively publishing information, like property assessment data, local governments may reduce the need for people in rural areas to travel to city halls or municipal centres. It's a way for government to come to them.

Some of our municipalities are doing quite well:

- This year, four local governments in BC were included in the Public Sector Digest's list of the top 20 open cities in Canada: New Westminster, Surrey, Vancouver, and Victoria.

In 2011, the Province joined local governments in the open government movement when the Premier announced a **policy on open information and open data.**

The policy provides guidelines for the public release of government information and data. Briefly, these are:

Proactive disclosure of responses to general FOI requests: This means that the response letter and responsive records are posted publicly at least 72 hours after they are released to the applicant. This is, of course, subject to any applicable FIPPA exceptions.

The policy also provides for what it calls "routine release": Ministries are encouraged to routinely make information available that may be useful or interesting to the public unless limited by law, contract or policy.

The data tracked by the B.C. government is fascinating and diverse. While I was writing this speech, I did a quick check: 3,678 datasets were available on the DataBC website on a range of topics, including public sector compensation, land and resource use, and government's budget.

From commissioning aerial photography of debris along the B.C. coastline after the 2011 tsunami in Japan…

… to tracking forest harvesting throughout the province to analyze the impact on wildlife and forest structures…

… to mapping the province's cellular and high-speed internet coverage.

You can even look up all physical locations and web addresses of government-owned highway webcams.

And, interestingly, B.C. was the first jurisdiction in Canada to make its budget available in a machine-readable form.

For data-savvy users, government provides tools to create projects using the published data. DataBC offers a number of tutorials and a forum for users to request more information.

Almost half of these datasets (about 1,600) are published under the province's Open Government Licence, which is perpetual, worldwide, royalty free, and non-exclusive.

Users can copy, adapt, publish or otherwise distribute the data, even for a commercial purpose, and they are only subject to limited terms, such as acknowledging the source of the data.

The remaining datasets have greater use restrictions because they contain personal information or information that must be withheld under FIPPA **or** they are subject to third party rights such as intellectual property rights.

Users can still download these datasets but they are subject to some restrictions on **using** the data.

For our part, at the OIPC, we have also adopted an open data license for datasets published on our website.

Privacy implications

So, I've been talking a lot about access to information, but what about privacy rights? FIPPA, as I mentioned, has **two parts** – access to information and privacy protection.

Public bodies are prohibited from disclosing personal information, except as authorized by FIPPA.

When a public body considers publishing data or information, they need to determine if it includes personal information.

FIPPA defines personal information as information about an identifiable individual. This includes where an individual is directly identifiable **and** where the information could reasonably be used, either alone or with other information, to identify someone.

It's this "combination" that is often the issue. As more data enters the public domain and more powerful tools are developed to analyze and connect datasets, it becomes easier to combine information that can result in the ability to identify individuals.

Take for example the case of an IP address. Normally, we would associate an IP address with a particular computer at a point in time; however if we know who owns the account associated with the internet service, we may be able to determine the individual who was likely using the computer.

This is often referred to as the mosaic effect, where non-identifying data can become "personal information" when combined with other data or when placed in a particular context.

We've seen this happen time and again, sometimes with significant consequences.

You might remember, in 2006, AOL released the search history of over half a million users. Though the information was theoretically de-identified, each search was associated with a numerically identified account. The New York Times successfully identified several individuals, which led to people at AOL losing their jobs, including the CTO.

In 2014, New York City officials released information about 173 million unique taxi trips. The poorly de-identified data revealed detailed information about individual drivers' locations and work performance. In just a few hours, you could work out a driver's home address, income, and movements throughout the city. Even the private information about **their passengers'** journeys!

And one of the most notable cases is when the Massachusetts Group Health Insurance Commission released "anonymized" data on state employees, showing all of their hospital visits. William Weld, governor at the time, promised that people's privacy was completely protected because names, addresses, and Social Security numbers had been removed.

Well, not long after, a graduate student at MIT, Dr Latanya Sweeney, obtained the hospital discharge data, compared it with publicly available voter registration information, and quickly identified the health records of Governor Weld himself. Her work led to a hasty change in state policy.

All that to say: the risks of re-identification are more than theoretical.

I will be honest and say it is an issue we have struggled with. As my colleague Timothy Pilgrim, Australia's Access and Privacy Commissioner, likes to say: **this isn't rocket science… But it kind of is.**

So let's take a look at the B.C. government and see how they are managing this difficult challenge of promoting open government while protecting privacy.

OIPC Evaluation of Open Government

Two years after the open government initiative was announced, my office took a close look at government's performance.

In our report, my predecessor acknowledged that the province's open information website was a **big step forward**.

Creating a single online space for information that government **already** publishes on 600 unique websites makes a lot of sense.

And in May this year, the B.C. government announced **additional** measures to enhance openness and transparency. Government started proactively releasing information on directly awarded contracts, calendars of senior officials, regular summaries of gaming grants, and the status of all active FOI requests.

I think these are important and positive steps. But when it comes to public bodies generally, there is still progress to be made.

For example, under section 71 of FIPPA, public bodies **must** establish categories of information for proactive disclosure **without** requiring an access request. Many public bodies are failing to meet this requirement.

And on the privacy side, the B.C government is not quite measuring up there either.

Many datasets proactively disclosed through open data programs have nothing to do with individuals… data related to land use, finances or natural resources, for instance.

But of course, many datasets **do** include data about people, which can be very useful for social policy development, health research and the like.

Ministries are responsible for de-identifying datasets before they become open data. And ministries must also conduct a privacy assessment, according to the government's own policy.

In our 2012 report, we advised the government to create standardized guidance for all ministries to use to de-identify datasets.

We also recommended government re-assess the DataBC site on a routine basis to determine whether the risk of re-identification has changed as a result of more information entering the public domain.

But these important measures have not been implemented yet. Although I was told last week that the government is working toward it.

You'll recall that we made these recommendations to government over three years ago now, so I think it's fair to say that they have some work to do in this area too.

So how do we move forward?

As you know, once you start removing variables, you risk losing the utility of datasets. But these steps may be necessary to create and preserve anonymity.

De-identification is challenging: it always comes with a risk that personal information can be re-identified.

Public bodies need to assess whether a dataset that they plan to publish includes personal information.

If it does, then they should first check FIPPA and see if there's a provision for the data to be disclosed. In most cases, this is pretty unlikely.

But certain personal information **can** be disclosed, such as the remuneration of an officer, employee or member of a public body. There's a healthy appetite for this information – the salary information dataset is within the top 30 downloaded out of over 3,600 available.

FIPPA also allows personal information to be disclosed, if that information has been designated as a category of records under section 71, which I commented on earlier. But only if that information would not be an unreasonable invasion of the personal privacy of the individual. There is a test in the Act for determining whether or not this is the case.

But in most situations where personal information is in a dataset, it will need to be de-identified. This can be especially challenging for local governments, because when we are dealing with smaller populations, there is a greater likelihood that someone can be identified based on indirect or quasi identifiers.

Doing this properly means regular and robust risk assessments, **before and after** the data is released, because technological developments and new public data sets can increase the possibility of re-identification.

My colleagues in Ontario, Australia and the UK have created guidelines about some of the ways to assess the risk of re-identification and some common de-identification techniques.

Australia is going even further: they are planning to change their *Privacy Ac*t to make it a criminal offense to re-identify anonymized government data. They would also make it an offence to publish or communicate any re-identified dataset.

My office will be keeping an eye on the impact of this unique approach.

I mentioned earlier that municipalities have actually led the way for open government at the provincial level here in B.C. With your project, you are in a great position to offer guidance and best practices to any local government looking to implement an open data program.

It’s not an easy endeavour, to be sure. But I think there are some useful lessons to learn – of what to do and what not to do – from cities like Vancouver as well as from the provincial government.

And of course, my office is here to answer any questions you might have as you proceed with your three-year project.

Thank you for inviting me to speak with you.