A Citizen’s Guide

to Big Data and Your Privacy Rights in Nova Scotia

Office of the Information and Privacy Commissioner for Nova Scotia

A Citizen's Guide to Big Data and Your Privacy Rights in Nova Scotia[1]

New tools for combining and analyzing information have made it possible for researchers to uncover hidden patterns and connections in large data sets that would have previously been unknown. Collectively, these large data sets and the analytical tools and practices used to identify trends are known as 'big data.' While private sector companies often use big data analyses to support marketing and product development, public organizations are attracted to it as a way to improve policy and program development and ensure it is supported by better evidence.

Big data has the potential to provide public bodies and health custodians with greater insights into the quality and effectiveness of services and programs such as healthcare, social services, public safety and transportation. However, it also raises concerns regarding privacy and the protection of individuals' personal information.

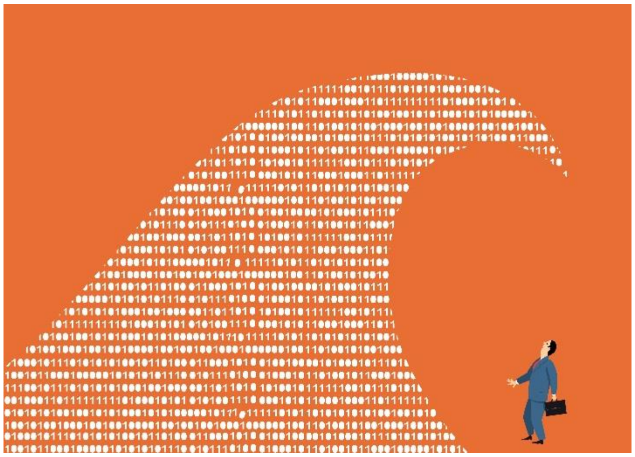
The Office of the Information and Privacy Commissioner for Nova Scotia (OIPC) is responsible for oversight of the ***Freedom of Information and Protection of Privacy Act***, the ***Municipal Government Act***, ***Part XX*** and the ***Personal Health Information Act***. Public bodies governed by these acts, such as government departments, municipalities, police services, health custodians, universities and school boards, must comply with these acts when collecting, using and disclosing personal information.

This fact sheet has been developed to help members of the public understand what big data is, and how it can have an impact on an individual's privacy.

[1] This Guide is based entirely on "Big Data and Your Privacy Rights" produced by the Office of the Information and Privacy Commissioner of Ontario available at: https://www.ipc.on.ca/wp-content/uploads/2017/01/fact-sheet-big-data-with-links.pdf. We are very grateful for their assistance and support in creating this Nova Scotian version of their fact sheet.

What is big data?

The term 'big data' is used for a lot of different things, and it can be difficult to pin down one single definition. In general, when the term 'big data' is used, it is referring to data collections that cannot be easily managed or understood using traditional means because of the size, irregularity or complexity of the data. It is often defined by three 'V's:



- **Volume** – there is so much data and so many data points that it can be difficult to sift through and comprehend
- **Velocity** – data is generated and added rapidly to the collection and can be updated in real time
- **Variety** – the data included in the collection can have many different forms and come from many different sources. For example, a big data collection might have structured data, like what you would find in a database, and unstructured data, such as collections of social media posts or blog entries

'Big data analytics' refers to the various methods and tools used to generate insights from big data. Algorithms[2] can be applied to large collections of data to identify patterns and connections, derive rules to automate decision making and even predict the results of a course of action, including a person's behaviour.

What impact does big data have on privacy?

Big data can, and often does, include personal information, which is any recorded information about an identifiable person, such as name, address, opinions or medical history. Just like any collection of data that includes personal information, Nova Scotia's privacy laws set out certain safeguards that public bodies and health custodians need to have in place to ensure that your personal information is appropriately collected, used, retained and disclosed.

[2] An algorithm is a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer (Concise Oxford English Dictionary). Algorithms specify a series of steps that perform a particular computation or task: http://fiftyexamples.readthedocs.io/en/latest/algorithms.html.

Big data, however, raises a number of unique privacy issues due to its size and complexity, and the types of analytical processes used by it.

Surveillance

Information collected by public bodies and health custodians can provide a great deal of detail about us, such as where we live, who we associate with and what our political or religious beliefs might be. When big data projects involve the collection and combination of large amounts of data (both publicly available and internal to organizations), it is possible to learn a lot about people and how they live their lives. When this happens, and when it is tracked over time, it becomes surveillance, which is any activity that monitors behaviours, actions or communications. There are many permissible uses for surveillance, such as for public safety and health care research. However, big data, when used irresponsibly, can lead to surveillance activities that are unwarranted. Whenever public bodies and health custodians decide to engage in big data activities that include personal information, the benefits of those activities must outweigh any potential privacy risk.

Accuracy of results

As a consequence of collecting and combining a lot of personal information that may not be related, it is possible for big data analytics to find patterns in the data that are merely coincidental. This may occur as a result of the sheer amount of information involved in the analysis. With so many variables at play, there are likely to be some data elements that appear to be related without actually being so.

Understanding the difference between correlation and causation is an essential part of evaluating the accuracy of big data analytics. Correlation means that the values of two variables in a data set are statistically related. For example, they tend to increase or decrease together. Causation is the stronger claim that the two variables relate by necessity and that a change in one always brings about a change in the other. A correlation, however, does not necessarily mean that the change in one variable was the cause of the change in the other variable. The two could simply relate by chance. For example, discovering a pattern in a data set that red haired people driving red cars have been in car accidents does not imply that being red haired and driving a red car caused those accidents. While there may be a correlation between these facts, that by itself is not sufficient to establish causation.

Bias in data sets

Big data is sometimes characterized as being more objective and unbiased than traditional data analyses because of its ability to analyze "all" the data. However, even if it were possible to collect all relevant data, biases can still be present in big data analytics. Data collection or generation practices can contain implicit biases that exclude or single out certain groups of people.

For example, if the hiring practices of an organization have resulted in people from similar backgrounds being hired more often, then the data on those hires will simply reflect those decisions. If that data is analyzed to find common attributes to be used to screen future applicants, the biases of the earlier hiring practices can be reinforced. Automatic decision making or application screening based on this data analysis can result in qualified people being automatically excluded from consideration due to an irrelevant attribute.

Unauthorized use

When we give our personal information to public bodies and health custodians subject to Nova Scotia's privacy laws, we should be told what that information will be used for and why. Big data projects, however, complicate things. Information that is collected for one reason may be combined with information collected for a different reason. The results of the combination and any analytics that are run on the combined data may be part of a study or program that the people who gave their information have never heard of. Generally speaking, under Nova Scotia's privacy laws, personal information should only be used in accordance with the purpose for which it was originally collected or for a purpose that has been authorized by law. If not properly managed, the collection, use and disclosure of big data sets may be contrary to Nova Scotia's privacy laws.

What is the OIPC doing to protect my privacy?

The OIPC is committed to ensuring that your privacy is protected. Big data creates new challenges and opportunities for Nova Scotia's public bodies and health custodians, but we believe that with diligence, proper planning and education on the unique issues raised by big data and analytics, it can be used in a privacy protective way. The OIPC will continue to work closely with public bodies and health custodians and will release guidance materials on privacy best practices and compliance in big data projects.

What rights do I have under Nova Scotia's privacy laws?

If you have any questions about what personal information a public body or health custodian is collecting or how it is being used, consult the public body's or health custodian's website or contact them for more information. If you would like a copy of the information that a public body or health custodian has about you, you can file a freedom of information request directly with that public body or health custodian.

To learn more about making a freedom of information request and what you can do to protect your privacy, visit the OIPC's website: www.foipop.ns.ca.

This document was produced by the Office of the Information and Privacy Commissioner for Nova Scotia. We can be reached at:

Office of the Information and Privacy Commissioner for Nova Scotia
509-5670 Spring Garden Road
P.O. Box 181
Halifax, NS B3J 2M4

Phone: 902-424-4684
Toll Free (NS): 1-866-243-1564
TDD/TTY: 1-800-855-0511
Fax: 902-424-8303

Website: www.foipop.ns.ca
Email: oipcns@novascotia.ca
Twitter: @NSInfoPrivacy