

# Dataset Basics

ATIPP Office

Number 4 | Updated June 2016

## Background

Under the *Access to Information and Protection of Privacy Act, 2015*, public bodies are required to provide applicants with records in the format requested, where possible ([section 20](#)).

## What is a dataset?

In its simplest form, a dataset is a collection of data. When people refer to a dataset, they are most commonly referring to a single data table made up of columns (fields or variables) and rows (records).

Dataset can also describe a collection of related sets of data. They are composed of separate yet related elements that can be manipulated by a computer and distributed as a collection. An example would be a collection of tables (each individually a dataset) made available as a single collection.

Often users requesting datasets will look for lower level data (such as individual records or unrounded figures) that can be used with other datasets to complete research and analysis.

## What is the difference between a dataset and other documents?

A dataset is a collection of facts, figures, or numbers that can easily be manipulated by a user. Documents, such as PDFs can include data in them, but that data is not easily manipulated, and often is not in its raw form (e.g. it is a summary of the data vs. the actual data).

## What can you do with datasets?

There are many uses for datasets. Once they are made available, datasets can be used for research, statistical analysis and development, or for creating apps, among other things. For example, information from datasets can be used to find the most

popular birthday, average age of cars on the road, or which areas have less crime. Datasets often have policy and/or commercial applications.

## What do datasets look like?

Datasets can look different depending on the format in which they are made available. On the Government of Newfoundland and Labrador Open Data website (<http://opendata.gov.nl.ca/>), many datasets are available in a CSV format (similar to an Excel format), which is tabular, but does not require any specific software to use. Mapping data would look very different, and is often requested as a shapefile (a popular geospatial vector data format for geographic information system software).

## How do I get information in a dataset?

While there are datasets available on the Government of Newfoundland and Labrador Open Data website, you can also request information in a machine-readable format (i.e. dataset) when contacting a public body or submitting an access to information request. If it is available in the format requested, or it would be reasonable to provide in the requested format (and it would not unreasonably interfere with operations), the public body will provide the information in the requested format.

Datasets available under the Open Government License can be accessed through <http://opendata.gov.nl.ca/>.