

RegData Canada Technical Documentation

Purpose:

RegData Canada is part of the RegData project that looks to quantify the regulation in a given jurisdiction's regulatory code and other regulatory documents. RegData Canada focuses on the regulatory codes of the federal Canadian government and the provinces and territories within the country. The quantified regulatory codes will allow researchers to directly test the effect of regulation on economic outcomes in Canada.

Content:

RegData Canada quantifies both regulations and statutes/acts for federal Canada and the eleven provinces and territories within Canada. The main metadata file for each jurisdiction includes restriction counts and word counts for each regulatory document. These datasets can be combined with the industry datasets (labeled as {jurisdictions}__naics07_030.csv) to produce "industry-relevant restrictions," or the amount of regulation that regulates a certain industry. The complexity datasets include information on the complexity of each regulatory document. Each of the complexity metrics have different strengths and weaknesses and are most informative when used together.

RegData Canada 2.0 is an upgrade over the 1.0 release because it adds complexity metrics, quantifies statutes/acts, and adds a second time series data point to each jurisdiction. The federal data now spans from 2006-2019 and the province/territory data on regulations spans from 2018-2019. Provincial data on statutes only includes 2019.

Technical Notes:

- RegData Canada uses the newest 3.0 version of the industry classification algorithm. Therefore, these datasets should not be directly compared to any jurisdiction datasets that use previous estimator versions.
- Users can safely aggregate document restriction counts, but mathematically should not aggregate "industry relevant restrictions" across multiple industries as this will invoke either some form of double-counting or under-counting.
- The regulatory data for federal Canada includes Statutory Instruments (SI's), Statutory Orders and Regulations (SOR's), and older versions of regulations labeled as C.R.C.'s. The statute data at the federal level looks at the Consolidated Acts posted on the [Justice Laws Website](#).
- A small portion Quebec's data is missing due to the government only publishing some regulations in French.
- Reference the below website links and the metadata files to verify that the data aligns with the files in interest. This project did its best to collect regulatory text for each province/territory, but every province/territory is different and has unique methods for publishing regulations and related materials. The data for this project is informative, but not necessarily all-encompassing or a complete picture of a jurisdiction's legally binding documents. Users should also keep in mind that all provinces/territories may not define "Acts", "Statutes", and "Regulations" in the same manner.
- The following websites were used to obtain the text that was quantified for RegData Canada:
 - [Canada Federal](#), [Canada Federal statutes](#), [Alberta, Alberta statutes](#), [British Columbia, British Columbia statutes](#), [Manitoba, Manitoba statutes](#), [New Brunswick, New Brunswick statutes](#), [Newfoundland & Labrador, Newfoundland & Labrador statutes](#), [Northwest Territories, Northwest Territories statutes](#), [Nova Scotia, Nova Scotia statutes](#), [Nunavut, Nunavut statutes](#), [Ontario, Ontario statutes](#), [Prince Edward Island, Prince Edward Island statutes](#), [Quebec, Quebec statutes](#), [Saskatchewan, Saskatchewan statutes](#), [Yukon, Yukon statutes](#).

Variable Descriptions:

Dataset	Variable	Description
Metadata	document_id	The unique identifier for a document. This variable appears in each dataset and should be used to merge data.
	date_collected	The date the document was collected from its source.
	year, date_consolidated	The date in which the data should be aggregated to allow for annual analysis. Aggregating on this variable will provide a picture of the jurisdiction's regulatory code for that date.
	published_date	Date the document was published.
	reg_id, act_id, title	The title or jurisdiction-given identifier of the regulation or the act.
	words	Total number of words in the document.
	shall, must, may not, required, prohibited	Legally binding terms that signal a regulation. These columns show the total occurrences of each of these terms in the given document.
restrictions	Total aggregate amount of the above five restrictive terms.	
Industry	document_id	The unique identifier for a document. This variable appears in each dataset and should be used to merge data.
	label	NAICS industry code number.
	probability	Probability that the document is relevant to the given NAICS industry.
Complexity	document_id	The unique identifier for a document. This variable appears in each dataset and should be used to merge data.
	conditionals	Total count of the number of "branching words" such as "if", "but", and "provided" that identify logical branches in a document.
	flesch_reading_ease	Readability of a document with a higher score meaning the document is more readable. Metric information found here .
	sentence_length	Measurement of the average length of sentences in a document.
	shannon_entropy	Measurement of the likelihood of encountering new words and concepts in a given document.