

The Use of Alternative Data for Price Indexes: The Canadian Experience

Heidi Ertl, Statistics Canada
Heidi.Ertl@canada.ca

Overview

While the majority of Statistics Canada's published price indexes are still based on traditional survey data, massive efforts are being made to leverage administrative and alternative data sources to reduce response burden and program costs, while maintaining or improving data quality. However, this reliance on alternative data is not without challenges.

The Canadian experience

Of the 25 published price indexes, more than half use administrative/alternative data, either solely or in combination with survey data. This extensive use of alternative data has led to the development of innovative tools and new ways of managing, integrating, capturing, processing and estimating price data.

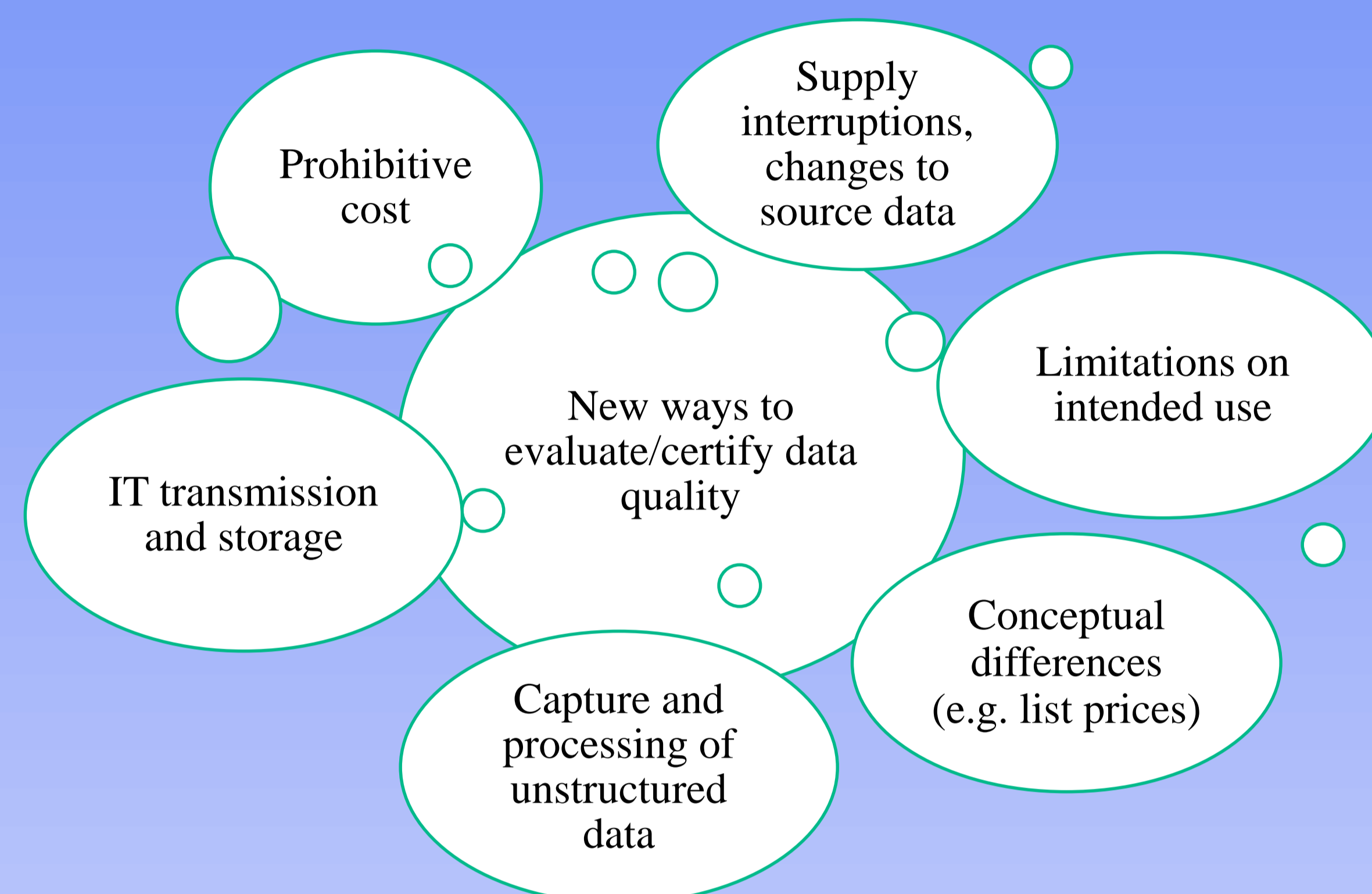
Why use alternative data?

Besides making use of existing data and saving on collection costs, using alternative data can:

- Reduce response burden
- Maximize number of price observations
- Minimize sample and survey bias

The challenges with alternative data

Alternative data as a replacement for survey data has many advantages, however there are also challenges and risks that may come into play:



Alternative data: types

Public data

Internet websites, subscriptions to catalogues or journals

Unstructured price data for specific products/services

Administrative data

Acquired under the *Statistics Act* or other formal agreements

Structured files containing financial data, corporate tax data, etc.

Data from other STC programs

Data from other survey programs used as inputs into another

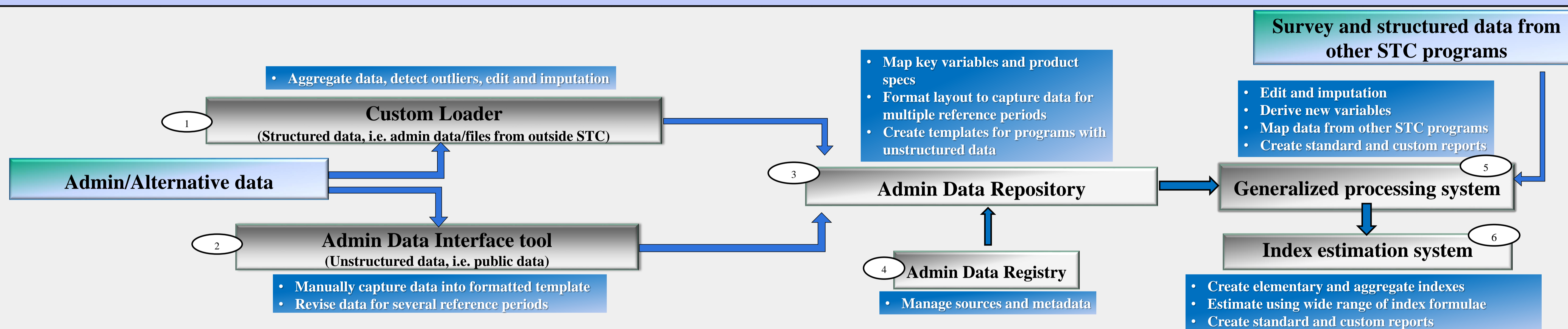
Structured data from other price indexes, and surveys

Data from third parties in the private sector

Data compiled by the private sector and acquired through contract or data sharing agreements

Commercially available structured data

Alternative data: tools and work flow



Price index program	Type of source data	Data capture	Processing	Estimation	Main challenges
Investment Banking Services	Data from private sector Other STC programs	Custom loader to Repository	Generalized processing system	Weighted average of annual price changes by product, chained together to form an index series	Product heterogeneity (within period and over time) Non-repeating transactions Shortage of transactions or missing data in some periods
Industrial Product and Raw Materials	Monthly electronic survey Other STC programs	Survey Admin Data Interface to Repository Custom loader to Repository	Generalized processing system	Price relatives weighted by sales of goods manufactured. Indexes are produced by commodity and industry	Multiple data sources, price concepts
Non-residential and Apartment Building Construction	Quarterly electronic survey Other STC programs Public data	Survey Custom loader to Repository	Generalized processing system Secondary processing system to handle multiple dimensions (product, building type, city)	Quarterly price movements for weighted representative products applied to building models to produce indexes by building, city and trade groups	Survey collects price movements, while alternative data collected is unit price. Additional processing in secondary system required to accommodate multiple dimensions
Commercial Software and Computers and Peripherals	Data from private sector	Admin Data Interface to Repository	Generalized processing system	Fixed sample for desktop, laptop, tablet, servers and smartphones. Matched sample for computer peripherals (printers, monitors) and commercial software. Also known as 'maximum overlap' method	Completely dependent on external data source Changes to data can impact weighting, number of price observations/products and index methodology Fixed sample does not fully maximize data from the supplier and requires hedonic quality adjustment to link discontinued and replacement models