

## Re-Thinking Data Quality in a Modern Framework

Anil Arora, Chief Statistician of Canada, Statistics Canada Q2022 Conference, June 9, 2022



Delivering insight through data for a better Canada

established over 100 years ago

**high-quality** statistical information that matters world's leading statistical agencies

**Ensuring quality** management through our **Quality Guidelines and** Quality Assurance Framework

**Trusted Source of** information – accurate, objective, and unbiased

Remaining relevant by listening to our data users and adapting

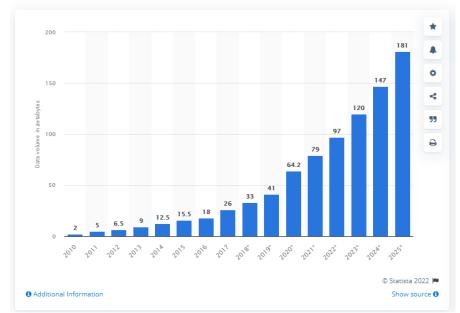
Statistics Canada

What is Statistics Canada?

## The New Data Ecosystem

Volume of data created, captured, copied and consumed worldwide from 2010 to 2025

(in zettabytes)









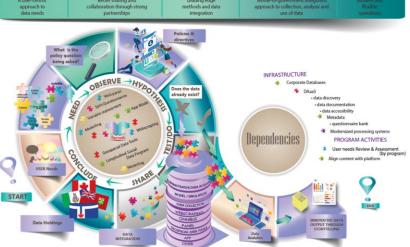


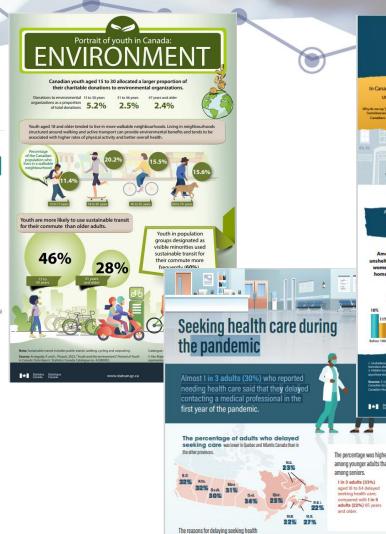






### Data Intake and Integration - How do we get there?









**BUSINESS CONDITIONS** 

IN CANADA, second quarter of 2022



Canada

care were often related to taking precautions

 Fear of possible exposure to COVID-19 outside health care settings (7%)

against COVID-19:

Butter Batter

# **Statistics Canada Modernization Journey**

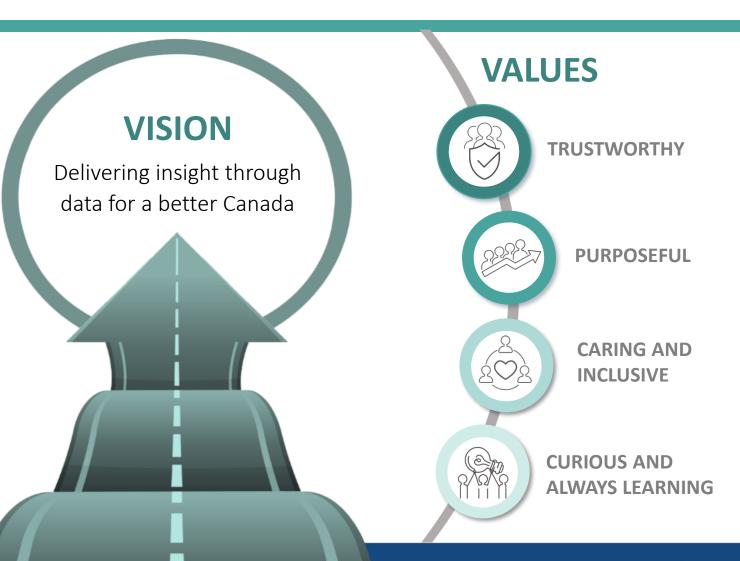
## **MISSION**

Serving Canada with high-quality statistical information that matters

## **MODERNIZATION PILLARS**

- 1. User-centric service delivery
- 2. Leading-edge methods and data integration
- 3. Statistical capacity building and leadership
- 4. Modern workforce and flexible workplace
- 5. Sharing and collaboration

Ensuring no one is left behind



# Frameworks

# Necessity and Proportionality Framework

#### MODERN STATISTICAL NEEDS



Collecting data for statistics is more than a century old and keeps evolving. In our Digital Age, your information needs are growing, and new data sources offer new possibilities for information.



#### COLLABORATION

Statistics Canada is working with experts from businesses and academia from around the world as well as the Office of the Privacy Commissioner to develop new statistical methods based on necessity and proportionality.



#### **NECESSITY**

Statistics Canada is mandated to produce data for Canada
—data that is essential for governments, municipalities,
businesses small and large, and individuals like you.
The agency measures our society, economy,
and environment.





When we plan surveys, our experts work to balance the volume and sources of data with the need to reduce the response burden on you, all while maintaining protection of your privacy.

#### A NEW FRAMEWORK



The new framework expands on these principles, which have always guided Statistics Canada. It assesses proportionality and data sensitivity, and ensures statistical values, such as the protection of privacy and confidentiality.

## Five Safes Framework (from UK)



## Foundations of Ethical Considerations: Guiding Principles

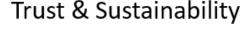


Benefits for Canadians
Privacy & Security





Transparency & Accountability





Data Quality

















## Approach to Data Stewardship



#### **Data Quality**

High quality, certified data maintains trust and transparency

Reduce time spent verifying and validating data



### **Data Standards**

Ensure interoperability, comparability and facilitate data linkage

Remove the burden of developing and adapting to standards yourself

Adhere to changing industry standards



## Data Management (

Ensure data is protected, used and shared in a fair. safe and equitable way

Ensure transparency of data use and centralized governance of assets based on rules and policies

**Subject Matter Expertise** 

Leverage our subject matter expertise to help measure social and economic issues and derive insights



### Data Sharing & Access

Alignment to FAIR principles: findable, accessible, interoperable, reusable

Reduce time spent finding data

Data stewarded by Statistics Canada under the Statistics Act



#### **Data Strategies**

Implement successful data strategies to solve problems and answer policy questions

Use data as a strategic asset

Ensure alignment to the Data Strategy for the Federal Public Service



#### **Data Science**

Meet evolving, data-driven needs

Implement a new scale of knowledge work automation

Review machine learning processes



### **Data Analytics**

Leverage world-leading expertise

Increase the value of insights/outputs leading to better outcomes for Canadians

Better understanding of how to use data



#### Data Governance

Clear roles and responsibilities around data

Ensure health of national statistical system

Access to national and international partnership networks



### **Data Protection**

Protection of privacy and confidentiality

Maintain trust and social license

Ensure data is protected and shared in a fair, safe and equitable way



### **Data Literacy**

Increase skills and competencies needed to understand and use data

Increase/optimize the use of data for decision and policy-making







# Data Quality Framework

Dimension	Definition
Access	The ease with which data can be discovered, processed, manipulated, and obtained by a user.
Accuracy	The degree to which data describes the real-world phenomena it is intended to represent.
Coherence	The degree to which data from one or more sources is comparable and linkable.
Completeness	The degree to which data values are sufficiently populated.
Consistency	The degree to which data is internally non-contradictory.
Interpretability	The degree to which data can be understood in its appropriate context.
Relevance	The degree to which data is deemed suitable to support an objective.
Reliability	The degree to which variability in data can be explained.
Timeliness	The amount of time between the end of the period to which data pertains, and the time at which that data is available to meet user needs.



## STATISTICS CANADA'S MISSION

Is to leverage valuable data assets by providing trusted data products, statistical services and insights to support the decision making of Canadian institutions, businesses and citizens - through robust quality, privacy protection, communications and measurement.





Sers

10

Φ

alu





## TRANSLATING TRENDS IN **CANADA**

#### INTO DATA POINTS AND KNOWLEDGE

COVID-19 Social and **Economic Impacts** 

Canadian Statistical Geospatial Explorer



Distribution of Household Accounts dashboard



Real-time local business conditions



...Many more on statcan.gc.ca

### PRIVACY AND CONFIDENTIALITY

AT EACH STEP OF THE DATA VALUE CHAIN

#### **LEGISLATION**



Statistics Act - mandate to collect, use and publish data, while protecting privacy and confidentiality (Privacy Act)

#### **DATA COLLECTION**



The Necessity and Proportionality framework ensures a balance between data needs and ethics.

#### **DATA PROCESSING**



The Quality Assurance Framework ensures we have rigor and standards for around the quality of data we produce

#### **ANALYSIS & MODELLING**



Framework for Responsible Machine Learning ensures no harm when using artificial intelligence to process data.

#### **DISSEMINATE & SHARE**



Five Safes framework ensures: Safe people. Safe projects. Safe data. Safe setting. Safe outputs

### Optimizing the value of Canada's Economic Data

- Point of Sale Data used to improve and automate CPI
- More Granular Household Accounts to inform pandemic income support measures and recovery
- Canadian Housing Statistics program to better understand Canada's evolving real estate market

### THE SCIENCE OF DATA **STEWARDSHIP Partnerships**

Telling the data story

- Disaggregated Data Action Plan
- Trust Centre
- · StatsCan Plus App
- · Eh Sayers Podcast
- · Data Literacy training
- Virtual Data Labs

## Analyze, Model

Continuously improving Canada's data story.

#### Explore, Clean, Describe

- Quality Assurance
- Metadata Standards
- Record Linkage
- Encryption,
- Deidentification Automation

- More timely releases: Nowcasting, Flash estimates Greater granularity

  - Cross-cutting analyses
  - Machine Learning and Al Microsimulation, modelling

### Define, Find, Gather

- Necessity and Proportionality
- Strategic Partnerships
- Data Stewardship
- Data Science, modern methods
- Modernizing sampling techniques to provide high quality disaggregated data
- Crowdsourcing understanding social trends sooner

data value chain



