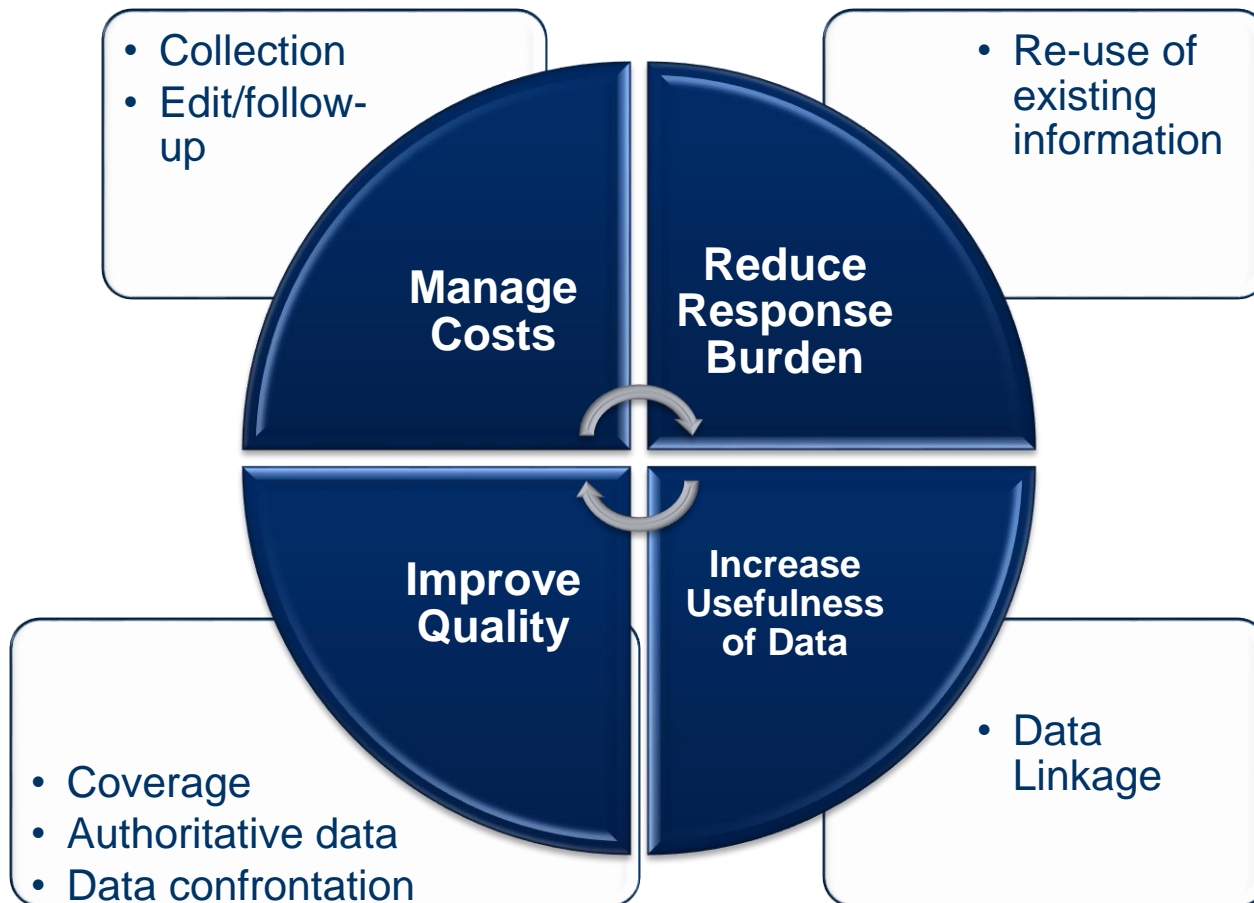# Administrative Data
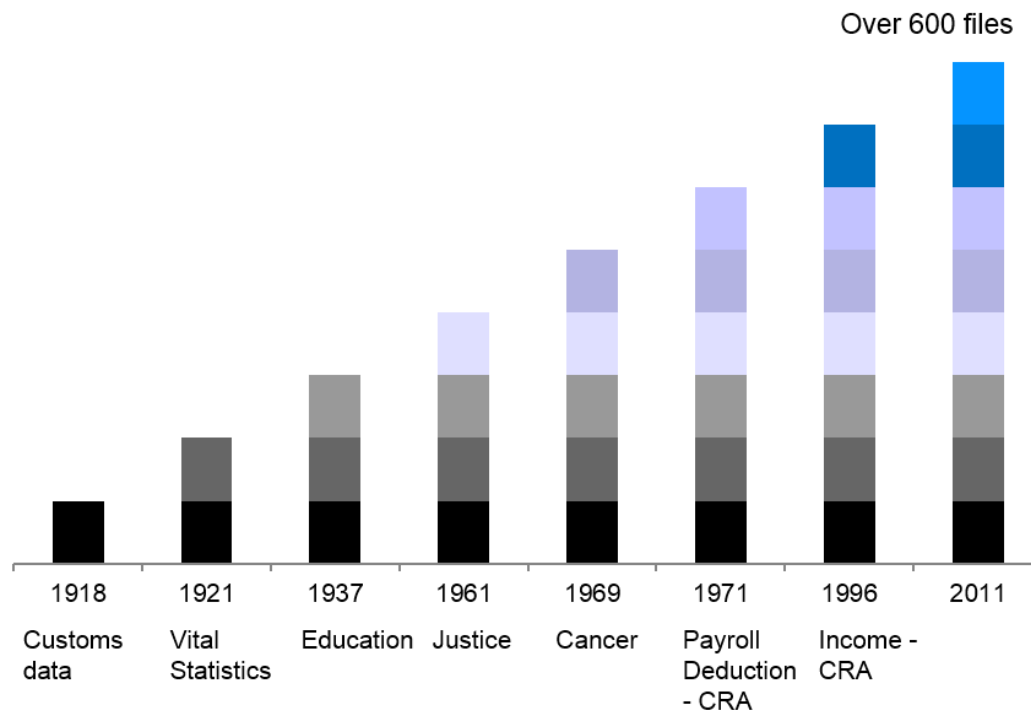
# A long journey
# A new beginning

## André Loranger

ISFP forum in Ecuador – Strengthening Statistical Management Practices
November 16, 2015

# Why use administrative data?



- Collection
- Edit/follow-up

**Manage Costs**

**Reduce Response Burden**

- Re-use of existing information

**Improve Quality**

**Increase Usefulness of Data**

- Coverage
- Authoritative data
- Data confrontation

- Data Linkage

Statistics Canada • Statistique Canada

# A tradition of using administrative data

- *As a stand-alone source to produce national statistics*
- *To support surveys*
  - *Creation of the statistical registers*
  - *Survey collection*
- *To partially replace survey data*
- *To provide new analytical opportunities*
- *Edit and imputation*
  - *Benchmarking, calibration, calendarization*
- *Data quality evaluation*
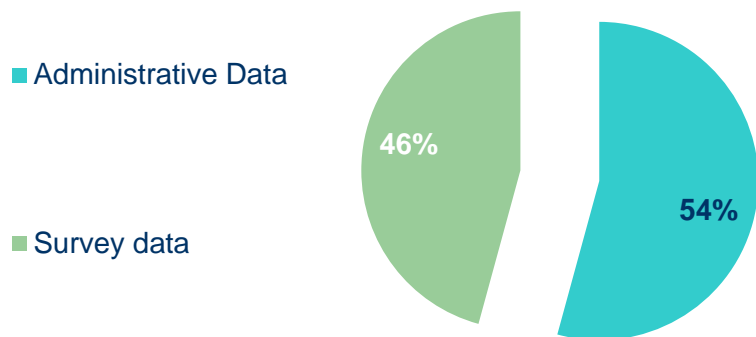- *Enhancing survey data – data integration*

Over 600 files

| 1918 | 1921 | 1937 | 1961 | 1969 | 1971 | 1996 | 2011 |

Customs data · Vital Statistics · Education · Justice · Cancer · Payroll Deduction - CRA · Income - CRA

**Principle: administrative data to be used when it leads to a better balance between relevance, quality, cost and respondent burden**
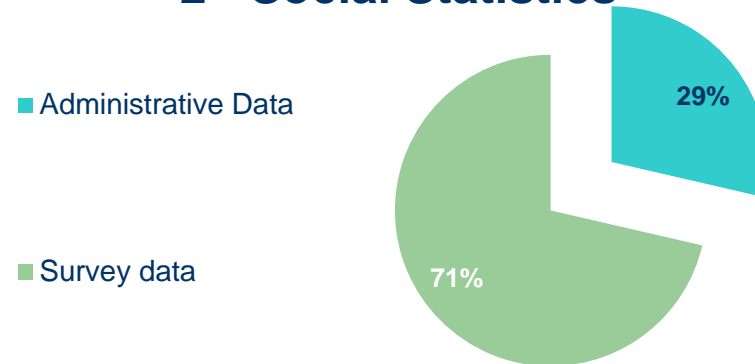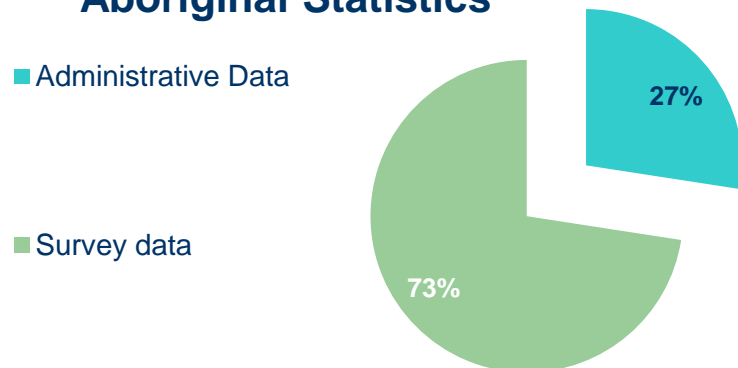
# Current use of administrative data

## 1 - Economic Statistics

- Administrative Data
- Survey data

46%
54%

## 2 - Social Statistics

- Administrative Data
- Survey data

29%
71%

## 3 - Census, Demography and Aboriginal Statistics

- Administrative Data
- Survey data

27%
73%

## Total

- Administrative Data
- Survey data

42%
58%

4

09/11/2015

# Need for a change

Surveys

Admin Data

Response burden

Collection Costs

Declining response rates

Increasing demand

Growth
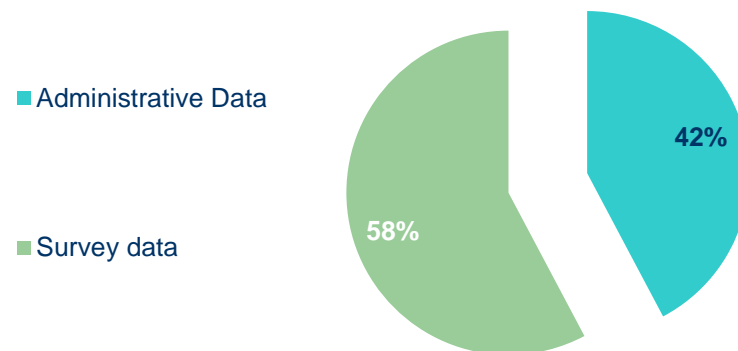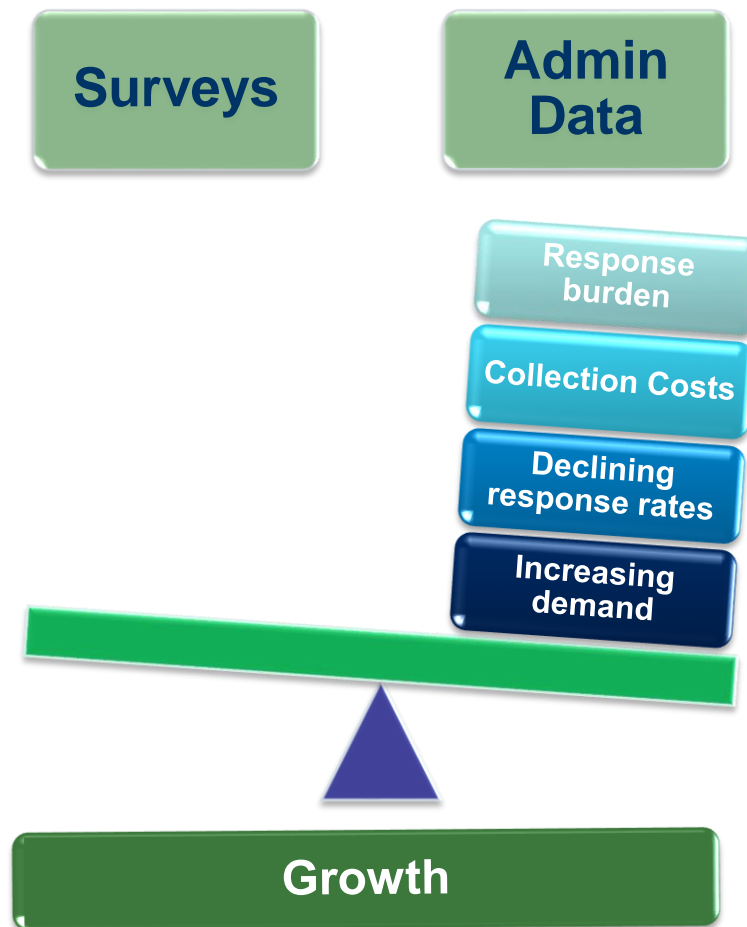
# Five pillars to a long term strategy

*Statistics Canada influences, has access to and makes maximum possible use of administrative and alternative data in its statistical programs, resulting in reduced response burden, lower costs, improved quality and new data series for users, while maintaining public trust.*

## Governance

- Policies, directives and business rules ensure a coherent approach

- Organizational changes ensure focused responsibility and accountability for corporate approaches

## Partnerships

- With Source organizations - governance facilitates influence

- Providing skills and expertise in exchange for data

- With collaborators – shared risk in research

- Best practices

## Communications

- With citizens promotes transparency to address privacy concerns

- Within NSO promotes successful strategies and best practices

## Knowledge Skills, Tools

- Understanding the potential and limitations of new data

- Recruiting and developing "data scientists"

- Tools and processes to accelerate use of new data e.g. quality evaluation framework

## Research & Development

- Using existing data in new ways

- Using new data sources to assess feasibility and quality

- Avoid the "all or nothing" approach

# Quality Evaluation Framework

Time is the most significant challenge in adopting new sources

- A quality evaluation framework assists in making a business case for investing time, capacity and resorces in acquiring a new data source

- Increases probability of usefulness of potential sources

- Establishes consistent practices in evaluating potential source data
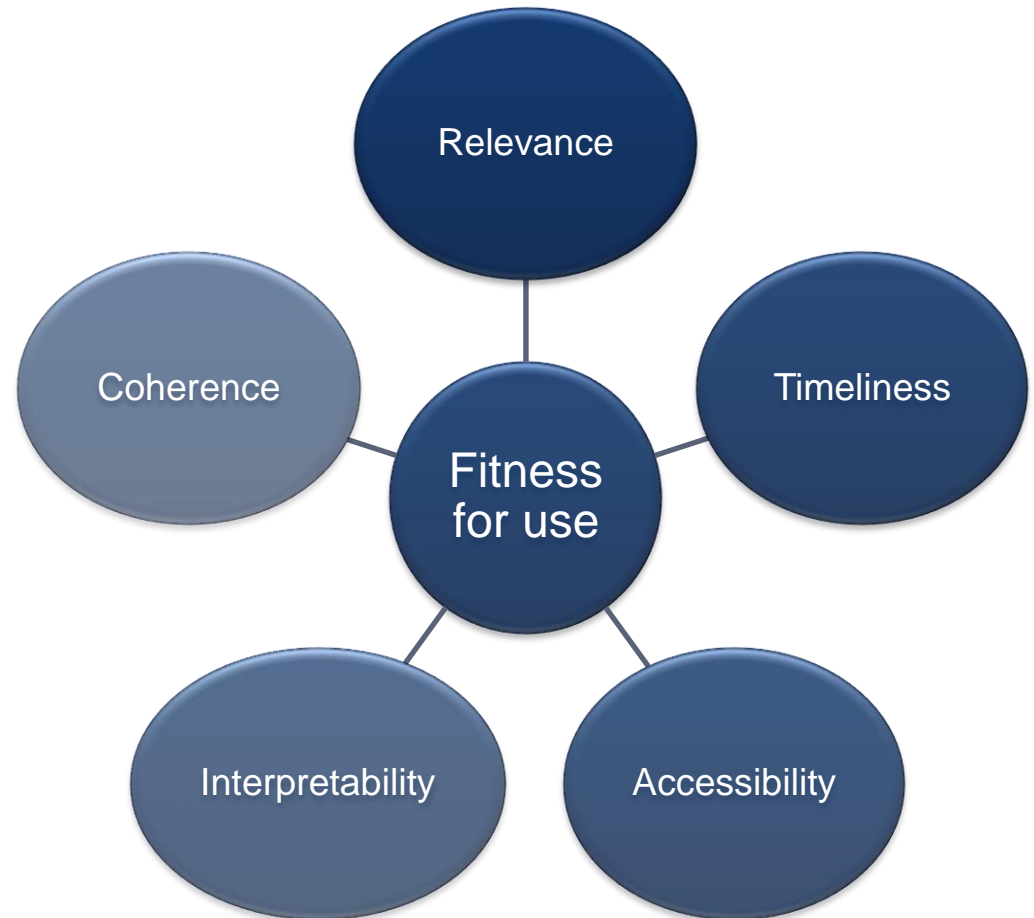
# Quality Evaluation Framework

**Exploration phase – no data**

- Assess dimensions of quality
  - Relevance
  - Timeliness
  - Accessibility
  - Interpretability
  - Coherence
- 3 components:
  - Institutional environment
  - Quality of file for use
  - Impact on statistical process

**Initial acquisition phase – some data**

- Evaluation of accuracy
- Validate information from Phase I

Phase I
Exploration

Fit for use?

Yes

No

Phase II
Initial
Acquisition

Assess accuracy

Document Stop

Accuracy Ok
Develop business case

Accuracy not sufficient
Document and stop
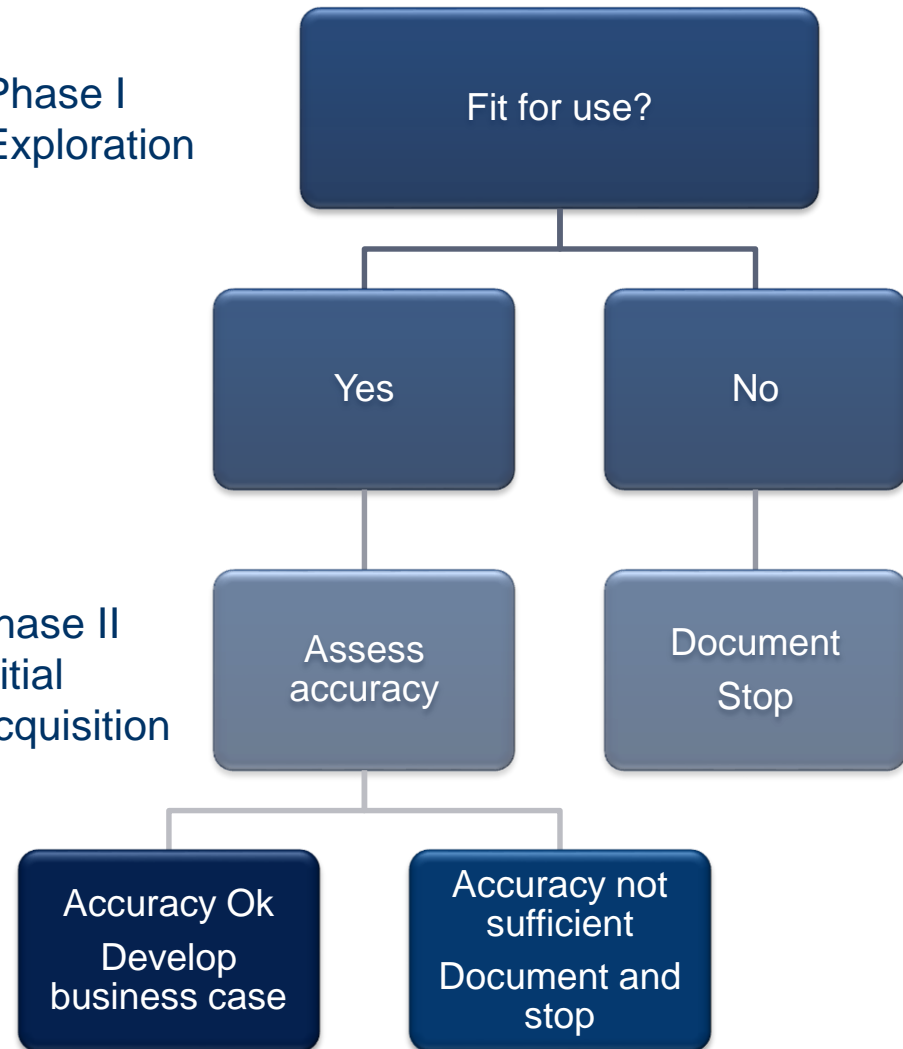
# A sample of current and future work

| **Governance** | **Partnerships** | **Communications** | **Knowledge Skills and Tools** | **Research and Development** |
|---|---|---|---|---|
| Directive on administrative data | Scanner data | Communications strategy | Community of practice – Big Data | Virtual population register |
| Administrative Data Division | Cell phone client files | Broadcast process integrating corporate needs | Administrative data inventory | Smart meters |
| Senior management committee | Credit card data | Community of practice on Big Data | Pre-processing of admin data | Webscraping |
| Federal code of practice | International – Big Data sandbox | Guides and templates on acquiring data | International - Big Data testing | Unique statistical identifier |

Statistics Canada • Statistique Canada

# Risks and Challenges

Citizens are the **Privacy and Public Trust** most important stakeholders

Are they compatible? **NSO and Partner Legislation** **Lead Time** Negotiations are rarely simple or quick

For free or fee? **Acquisition Costs** **Fitness for use** How good is good enough?

Will programs be at risk in the long term? **Reliability of Source** **Technical Capacity** Big Data requires special tools and IT infrastructure

# Additional contacts

| Claude Graziadei<br>Director General<br>Statistical Infrastructure | Julie Trépanier<br>Director<br>Administrative Data |
|---|---|
| 613-951-6128 | 613-951-9437 |
| Claude.Graziadei@canada.ca | Julie.Trepanier@canada.ca |