



Statistics Canada
www.statcan.gc.ca

FedCASIC 2012 Plenary Panel

The Use of Paradata to Improve Survey Quality: Organizational Approaches and Challenges

François Laflamme

Introduction

- Statistics Canada has been collecting, recording, processing, storing and analyzing a wide range of paradata for all CATI and CAPI surveys for several years
 - ◆ Many paradata sources are automatically recorded
 - ◆ Standardized by type of surveys (social, agricultural and business)
 - ◆ Single entry point for paradata storage, processing and access
 - ◆ Starting to work on paradata for Web surveys
- Paradata are used before, during and after data collection
 - ◆ Before: e.g. survey costing, survey planning (based on historical info.)
 - ◆ During: e.g. reporting, survey & interviewer monitoring, responsive design
 - ◆ After: e.g. post mortem analysis , non-response adjustments
- Paradata are also used to meet both operational and methodological research objectives

Paradata research objectives



- ❑ Better understand data collection process and practices
- ❑ Identify potential operational efficiencies
- ❑ Evaluate new data collection initiatives
- ❑ Provide timely and relevant feedback and information
 - ◆ Data collection survey management (Active Management)
 - ◆ Responsive Collection Design
- ❑ Maintain and improve data quality
- ❑ Improve the way surveys are conducted and managed
 - ◆ New challenge multi-mode surveys

What type of paradata are recorded / collected?

- Transaction files - Call and contact information
 - ◆ CATI - Blaise Transaction History (BTH) file & CAPI - Case event file
 - Audit trail (interviewer key strokes)
 - Administrative & Interviewer time reporting information (payroll)
 - Interviewer monitoring (e.g. listening (CATI) & audio recording (CAPI))
 - Interviewer notes (no interviewer observations)
 - Web/EQ paradata (in development)
-
- *Transaction and audit trail - automatically recorded*
 - *Collection division is responsible for paradata (centralized)*
 - *Historical information since 2003*
 - *Generally updated on daily basis*

Standardized paradata and key indicators

- Are standardized paradata being collected/recorded across modes and/or surveys?
 - ◆ Yes for the vast majority of paradata
 - e.g. for BTH: common set of variables for social surveys and few surveys specific variables - always at the end of file
 - ◆ The main sources are integrated into a single large database
 - ◆ Transaction information for EQ paradata are more likely to be different than other modes

- Are there key indicators used to assess survey process quality?
 - ◆ Indicators to monitor survey progress and effort
 - ◆ In addition to response rate, representativity, cost, productivity, responding potential of in-progress cases are used during collection
 - Develop in a research perspective
 - Essentially used for Responsive Collection Design (RCD) surveys

What are the challenges to:



- ❑ Recording / collecting and processing paradata?
 - ◆ Paradata are automatically recorded - no cost or interviewer burden
 - ◆ Data storage (very large files) and analysis are the main challenges

- ❑ Providing tailored indicators ?
 - ◆ Customized product (often case-by case) - needs to be merged with sample design information and other sources (e.g. RCD)
 - ◆ Pressure for more standardization - conflicting objectives

- ❑ Implementing responsive design ?
 - ◆ Main challenges are to get access to timely (24-hours delay) paradata; to produce relevant active management reports and to have the opportunity to react/intervene during collection
 - ◆ Successfully implemented for some CATI surveys since 2009

Other issues

□ How are these challenges being met?

- ◆ Single entry point for paradata storage, processing and access
- ◆ Collaboration between data collection, methodology and subject matter
- ◆ Sharing experiences and developing analytical skills - paradata course
- ◆ Senior management support

□ What challenges remain?

- ◆ EQ/Web paradata (design, analysis and integration with other modes)
- ◆ Continue to convince management to invest in paradata research

□ What are the overhead costs associated paradata and dashboards, and how are they justified?

- ◆ No real collecting cost; storage, development (database and system) and analysis are the main components of the cost
- ◆ Justifications:
 - Should not be “nice to know” information
 - Need to demonstrate benefits to improve cost-efficiency or quality

Summary

- Paradata has been the cornerstone of data collection research at StatCan and continues to be extensively used
 - ◆ To meet operational and methodological research objectives
 - ◆ To operationalize, monitor and assess new data collection initiatives and survey management strategies
 - ◆ To improve the link between data collection and post data collection activities

- Future research and implementation will focus on “Strategies to improve the way data collection is conducted and managed”. Hence the research will need to:
 - ◆ Be sound and operationally viable
 - ◆ Lead to more cost-effective collection and sample design strategies
 - ◆ Lead to data quality improvements