



Statistics Canada  
www.statcan.gc.ca

# Using Paradata to Monitor Survey Quality in Regional Data Collection

Mike Maydan  
Statistics Canada  
FEDCASIC March 2009

# Introduction

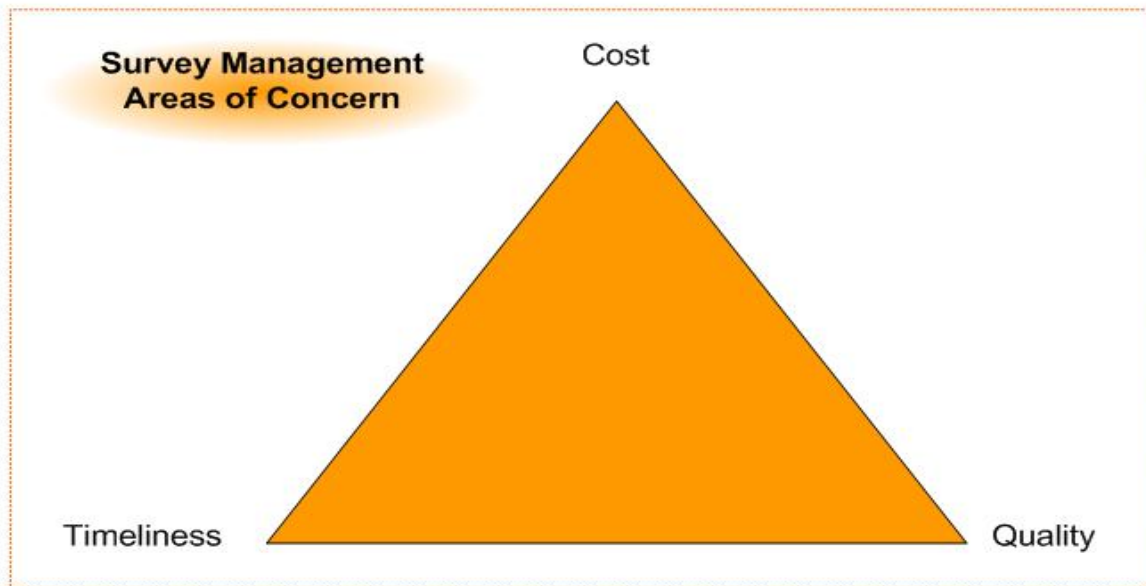


- ❑ The increasing complexity of survey design demands better tools to assess the data collection process
- ❑ CAI has opened the door to a wide range of process data (“paradata”)

# Data Collection Management

- ❑ Data Collection is a key element of the survey process
- ❑ Data Collection management is a trade-off between: Timeliness, Quality (data accuracy) and Cost
- ❑ Management information requirements depend on:
  - ◆ Managements levels
  - ◆ Point in time during data collection
  - ◆ Focus on
    - Production units and/or
    - Interviewer performance

## Regional Office Data Collection Management Dimensions



### Levels of Management Information Needs

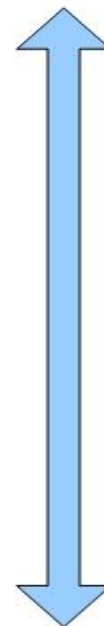
Regional Director

Assistant Director Operations RO

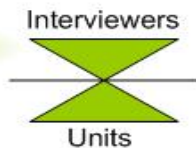
SI-05 Programme Manager

AS-03 Collection Manager

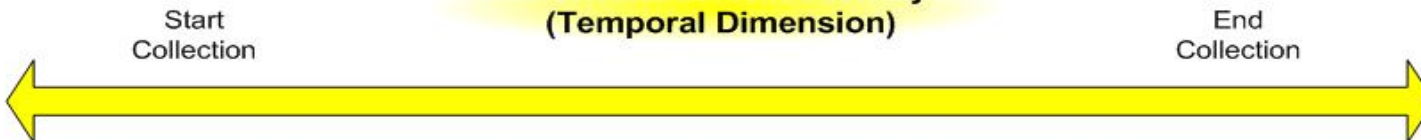
Senior Interviewer



### Survey Production Management Focus



### Point of Time in Collection Cycle (Temporal Dimension)



# What is MIS?



Definition: Management Information System

- An automated system designed to provide progress and status information to management as an aide to decision making
- “data and records ...which are designed as a common pool of information to assist managers in carrying out their responsibilities”

source: OECD / USBC

# MIS Issues for STC Regional Collection



There is “no single system”

- ❑ Interconnected and disconnected data collection and archive systems
- ❑ Automated/predefined and ad hoc reports
- ❑ National and regional aspects

Management challenge

- ◆ competing priorities
- ◆ Varying needs - dimensions

# Survey Reporting



Monitoring of survey progress is typically done through reports  
Survey/cycle level is the starting point for collection MIS

- ❑ Management by mode (CAPI or CATI)

Two way street

- ❑ Break down by span of control: region, CATI site, field interviewer team, individual interviewer, case etc.

- ❑ Roll up reporting for higher level management: groups of multiple ongoing surveys

- ❑ All based on case level paradata

# Case Level Drivers



Two main paradata sources

- ❑ BTH files from CATI
- ❑ CAPI case management event details  
(case event file)

Source of timing, routing and outcome details

Combined with pay system and financial survey metadata to provide measurement tools



# Data Integration Tool



- **Data Integration and Production Planning (DIPP):**  
Integrate and consolidate administrative and operational data

Report interface

3 main reports

- [National Progress Report](#)
- [Survey Summary Report](#)
- [Survey Performance Report](#)

# Data Quality Reporting



Paradata provides collection managers with reports to measure data accuracy e.g.

- Response & non response rates
  - ◆ overall & by component (strata)
- Non response follow up
- Refusal conversion
- Tracing & research

# Audit Trail – Intra Case Level Data



- ❑ Over time, the emphasis for audit trail analysis has changed from evaluating the application to evaluating the act of collection.
- ❑ Used alone or in conjunction with BTH it enhances the measurement of an interviewer's performance

# POINT System



- ❑ POINT – Pace of Interview information
- ❑ Uses survey audit trail data to identify irregular calls during production
  - ◆ Pace – field changes per minute
  - ◆ Item non-response (“don’t knows” and “refusals”)
- ❑ Quality assurance - objective measure of interviewing practices
- ❑ Help identify need for additional interviewer training or corrective action

## Survey Summary by Site

<b>Survey Name</b>	Surv 1 CP 1	<b>Survey Pace</b>	6.7 Field Changes / Minute
<b>Date Range</b>	All collection dates	<b>Irregular Calls Boundaries</b>	Amended Pace > 11.7 FCPM
<b>Site(s)</b>	All sites		Item Non-Response > 25.0%
		<b>Other</b>	Long-Visit Field > 3.0 minutes

Site	# of Ints	Evaluated Calls			Field Changes	Amended Interviewing Pace		Item Non-Response		
		Type	#	%		Time (min)	FldChg /Min	DK	RF	%
<b>TOTAL</b>	<b>32</b>	<b>All Calls</b>	<b>1,672</b>	<b>100.0</b>	<b>60,361</b>	<b>8,855.2</b>	<b>6.8</b>	<b>1,360</b>	<b>91</b>	<b>2.4</b>
TOTAL		Regular	1,639	98.0	59,397	8,774.3	6.8	1,260	66	2.2
TOTAL		Irregular	33	2.0	964	80.9	11.9	100	25	13.0
<b>Site 1</b>	<b>10</b>	<b>All Calls</b>	<b>653</b>	<b>100.0</b>	<b>25,246</b>	<b>3,705.1</b>	<b>6.8</b>	<b>355</b>	<b>55</b>	<b>1.6</b>
Site 1		Regular	643	98.5	24,950	3,681.7	6.8	347	46	1.6
Site 1		Irregular	<u>10</u>	1.5	296	23.4	12.6	8	9	5.7
<b>Site 2</b>	<b>22</b>	<b>All Calls</b>	<b>1,019</b>	<b>100.0</b>	<b>35,115</b>	<b>5,150.1</b>	<b>6.8</b>	<b>1,005</b>	<b>36</b>	<b>3.0</b>
Site 2		Regular	996	97.7	34,447	5,092.6	6.8	913	20	2.7
Site 2		Irregular	<u>23</u>	2.3	668	57.5	11.6	92	16	16.2

# Other Uses of Audit Trail Data



Files are processed to provide case level call details to RO managers for:

- Repairing technical data losses
- Post collection validation and verification of collection
- Augmenting CAPI monitoring to include interaction of interviewer and survey instrument
- Time per section report based on ADT data – post survey analysis & budgeting [TSR example](#)

# Tracing Respondents



- ❑ “Black hole “for survey managers
- ❑ Combination of online & offline interviewer activity
- ❑ Challenges
  - ◆ Measure & manage individual interviewer effectiveness
  - ◆ Identify wasted effort “chasing dogs”  
[Tracing effort](#)

# Refusals



- ❑ Declining response rates in CATI
- ❑ Societal vs training issues
- ❑ Refusal avoidance
- ❑ Management challenge is to identify problem interviewers

[Summary refusals](#)

[Detail Report](#)



# Summary/Issues



- ❑ Increasing demand for data and reports
- ❑ Challenge for operational areas to keep pace with client and methodology drivers
- ❑ Integrating additional metadata
- ❑ Integrating audit trail & BTH data in a standardized fashion to provide relevant and useable quality reports

# Summary/Issues



## Human Issues

- Developing user documentation on understanding Audit Trail files
- Need for user training on MIS reports, teaching managers how to analyse outputs

# Work in Progress/Next Steps



## Technical Issues

- ❑ Data warehouse to replace DIPP – proof of concept for early spring 2009
- ❑ Consolidating paradata holdings to rationalize storage and manage data access
- ❑ Next generation production reports from Survey Master Control System – dashboard approach – prototype late 2009

# Questions



For more information contact:

Mike Maydan

[mike.maydan@statcan.gc.ca](mailto:mike.maydan@statcan.gc.ca)

Thank you