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

March 27, 2012

Ananth Koppikar

MATHEMATICA
Policy Research

Overview

I. What paradata are collected?

II. Are standardized paradata being collected across modes and/or studies?

III. Is there a set of key indicators used to assess survey process quality?

IV. Conceptual paradata system architecture at Mathematica

V. Operational challenges

VI. Costs and justification

MATHEMATICA
Policy Research

I. What paradata are collected?

- Paradata indicators currently collected and reported on include: cost per complete, outbound calls per hour, clerical support, survey operations support, supervision hours per complete, and monitoring and mileage per complete (as appropriate by mode).

- Paradata indicators have been classified in three broad functional categories of interest: (1) production and cost, (2) data quality, and (3) business development.

- All indicators are associated with specific operational questions, such as: Are Survey Operations Center (SOC) and supervision hours per complete in line with budget expectations?

MATHEMATICA
Policy Research

I. What paradata are collected? (continued)

- Paradata targeted for the future phases include:
  - Date and time an interview was started (each CAI mode)
  - Date and time an interview was completed (each CAI mode)
  - Questionnaire responses
  - Average time per complete or average number of attempts
  - Item-level Don't Knows/Refused/Blanks
  - Key strokes – changing answers or backing up
  - Contact data
    - Date/time
    - Time slice
    - Contact result codes/response rates
  - Number of interviewers and number of hours worked

- In addition, we may capture and study the correlation of some paradata indicators to certain project metadata such as characteristics of interviewer and respondent, incentives, calling rules, etc.

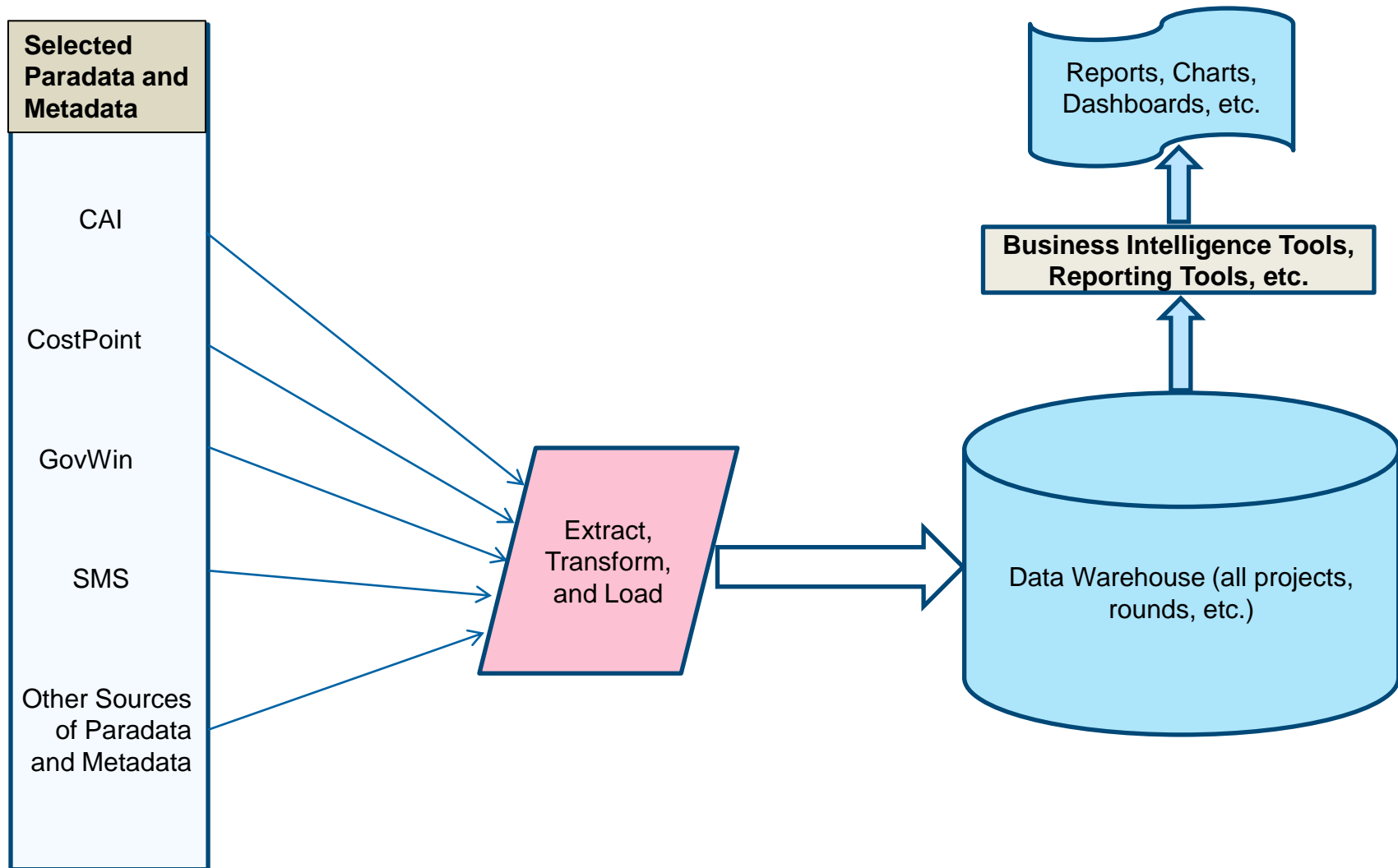## II. Are standardized paradata being collected across modes and/or studies?

- The paradata that we have identified for inclusion in 2012 are captured in standardized formats across projects.

- Standardized call details and history are available on our Blaise CATI system.

- For other modes, most projects use our Survey Management System (SMS), and its underlying database structure permits standardization of sample and operational data across projects.

- Other standard data sources include Deltek's Time & Expense system and GovWin.

- While standard elements are being captured in key systems, important challenges remain in developing standardized usage/indicator values.

MATHEMATICA
Policy Research

III. Is there a set of key indicators used to assess survey process quality?

- Key indicators by mode:
    - Cost per complete (TINT, Field, Web)
    - Clerical hours per complete (TINT, Field)
    - Hours per complete (TINT, Web)
    - Mileage per complete (Field)
    - Monitoring percentages (completes, cases per interviewer, TINT hours)
    - Outbound calls per hour (TINT)
    - SOC support hours (TINT, Field)
    - Supervision percentage (TINT)

- These paradata shed light on the following operational questions:
    - Are costs per complete higher or lower than expected?
    - Are support, on-call, clerical, monitoring, and supervision hours in line with the project budget?
    - Are hours per complete higher or lower than expected?
    - Is call volume higher or lower than expected?
    - Is mileage per complete higher or lower than expected?
    - Is outbound call volume higher or lower than expected?

MATHEMATICA
Policy Research

IV. Conceptual Paradata System Architecture at Mathematica

Selected Paradata and Metadata
- CAI
- CostPoint
- GovWin
- SMS
- Other Sources of Paradata and Metadata

Extract, Transform, and Load

Data Warehouse (all projects, rounds, etc.)

Business Intelligence Tools, Reporting Tools, etc.

Reports, Charts, Dashboards, etc.

MATHEMATICA
Policy Research

V. Operational Challenges

- To collecting and processing paradata:
  - Designing a flexible, scalable data warehouse that can store selected paradata
  - Standardizing values across projects
  - Missing data

- To collecting standardized paradata and producing standard indicators across modes and/or studies:
  - There is a cultural and educational challenge to the adoption of standards
    - Overcoming the fear that standardization leads to loss of important detail and that project-specific information needs will go unmet
    - Overcoming the fear that standardization will be too costly and will not provide a good return on investment
    - Once the case is made for standards, the challenge remains of which standards to adopt
      - We are investigating adoption of emerging paradata standards

MATHEMATICA
Policy Research

V. Operational Challenges (continued)

- To providing tailored indicators to address study-specific design issues:
  - Existing paradata reports are not widely or consistently used
  - Development of new report templates that are readily interpretable and contain actionable data has proven difficult

- To Implementing responsive designs based on process quality indicators
  - A model for implementing responsive survey design is being developed as part of the paradata initiative.
  - This model is based on the <u>Total Survey Error (TSE)</u> approach, which assumes a tradeoff between survey error and survey cost and provides a framework for reducing respondent selection error, response accuracy issues, and survey administration errors, as well as managing cost.
  - Achieving consensus on an organization-wide model is a time-consuming process
  - Training staff to understand paradata and use it appropriately

MATHEMATICA
Policy Research

VI. Costs and Justification

- Costs associated with developing a paradata system will vary based on the complexity of the system.

- Costs include labor associated with researcher and programmer time, hardware (e.g., servers), and software (e.g., BI tools).

- Initial phase at Mathematica (developing overall framework; formulating research questions; drilling down to identify indicators; specifying requirements for indicators and reports; designing and building the data warehouse, ETL programs, and reports; piloting and then deploying to production) is expected to cost over $200K. For this initial phase, we will be using existing hardware and not incorporating any BI tools.

MATHEMATICA
Policy Research

VI. Costs and Justification

- Regular, readily interpretable reporting on paradata provides the foundation for the implementation of responsive survey design

- Anecdotal evidence exists at Mathematica that paradata have been used on some projects to reduce data collection costs and improve response rates.

- Senior management recognizes that paradata would be an invaluable tool to improving data quality and optimizing costs, among other things.

- Costs are viewed by senior management as an investment and not an expense.

- Approaching the paradata initiative in phases – helps manage costs and expectations, and provides senior management with a greater degree of control on the investment.

MATHEMATICA
Policy Research