Blaise 5 Paradata requirements

Rebecca Gatward

Lisa Wood

Patty Maher

Gina-Qian Cheung

International Blaise Users Conference

September 2013

Survey Research Operations - Survey Research Center - Institute for Social Research

Focus of this presentation

- In 2011 a working group was formed with the purpose of specifying the Blaise community's paradata requirements for Blaise 5 .

- The group comprised representatives from Blaise User Corporate license holders and the paper was presented and discussed at the Blaise Corporate License User Board meeting in January 2013.

- This presentation will provide -
  - a summary of the requirements (more detail in the paper),
  - some practical examples to illustrate how paradata can be used to inform decision making throughout the survey process and assist in fieldwork monitoring.
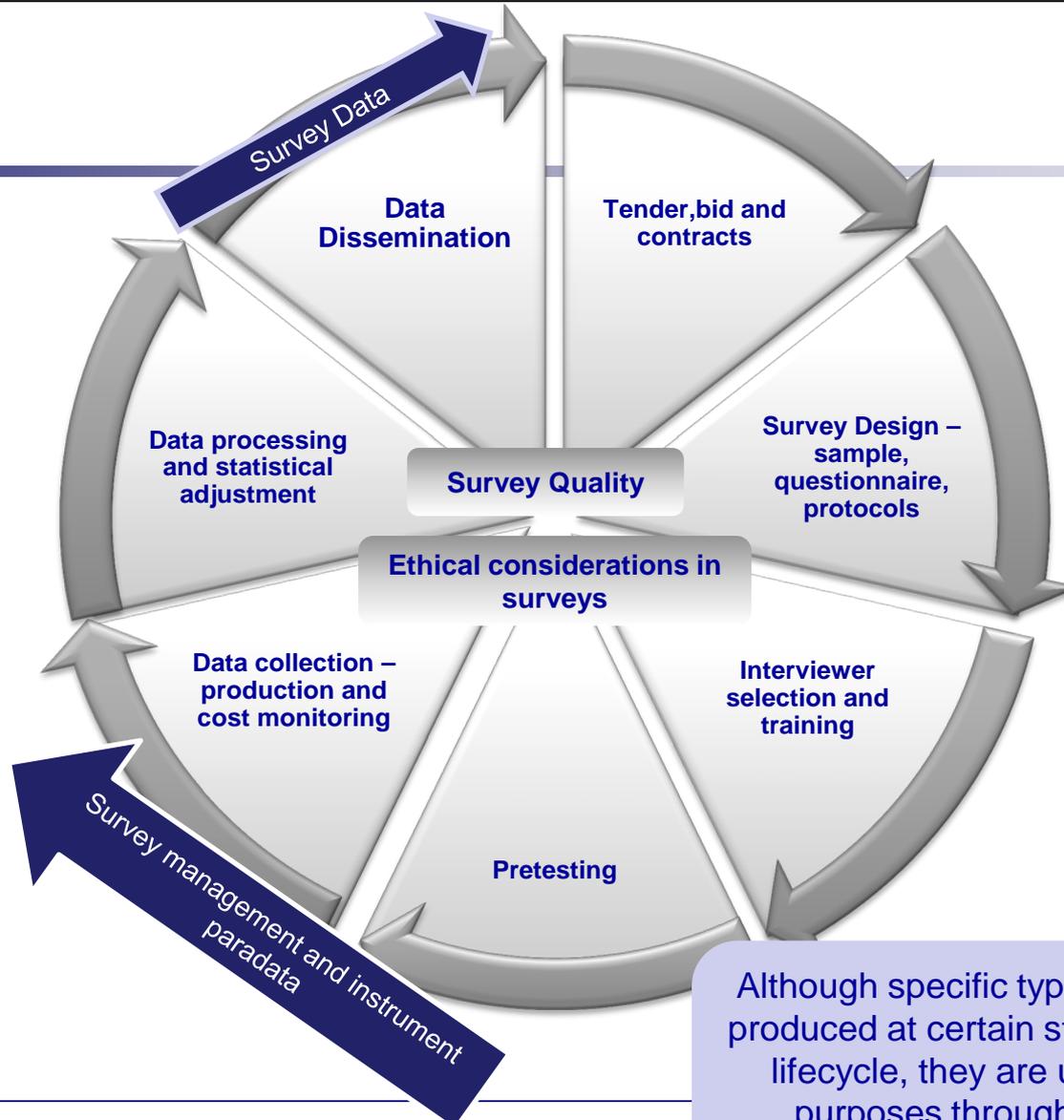
Also to note…

- This paper concentrates on paradata which are linked directly to the administration of a survey instrument.

- The source of this type of paradata are primarily:
    - audit trails
    - recordings of interviews using Computer Assisted Recorded Interviews (CARI),
    - server and client side data for web collection.

- Another type of paradata are collected or stored in survey and sample management systems.This type of paradata describe the actions and processes involved in data collection,
    - For e.g. data about contact attempts, the interview, observations from the interviewer, the sampling frame and respondent related administrative procedures, such as compliance to complete consent forms and payment of incentives.
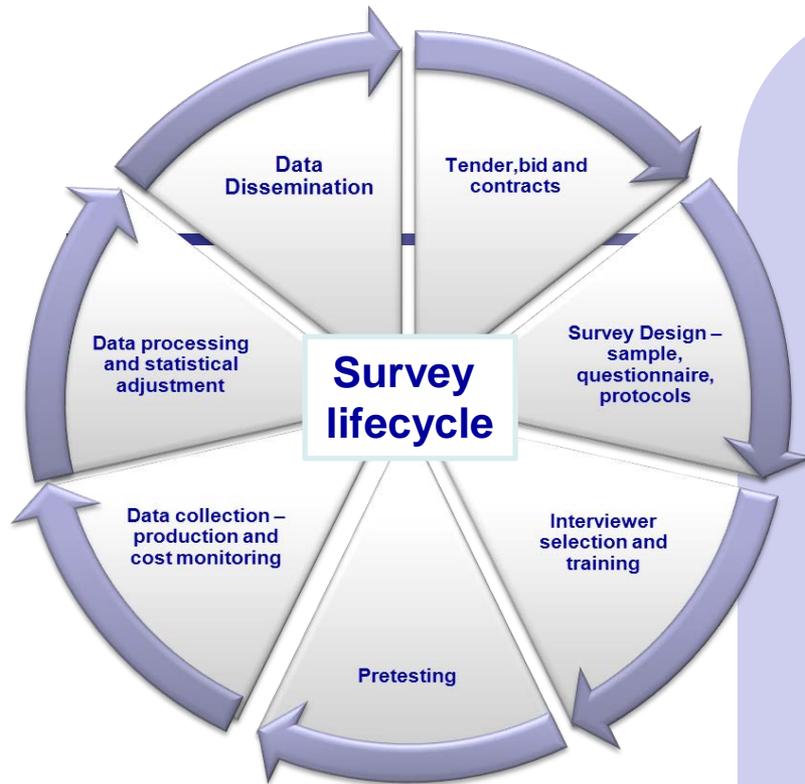
Paradata are used throughout the survey lifecycle –
from tender or budgeting , through the design and piloting phase, during data collection and when writing final documentation.

Survey Data

Data Dissemination

Tender, bid and contracts

Data processing and statistical adjustment

Survey Quality

Survey Design – sample, questionnaire, protocols

Ethical considerations in surveys

Data collection – production and cost monitoring

Interviewer selection and training

Pretesting

Survey management and instrument paradata

Although specific types of paradata are produced at certain stages of the survey lifecycle, they are used for multiple purposes throughout the survey process.
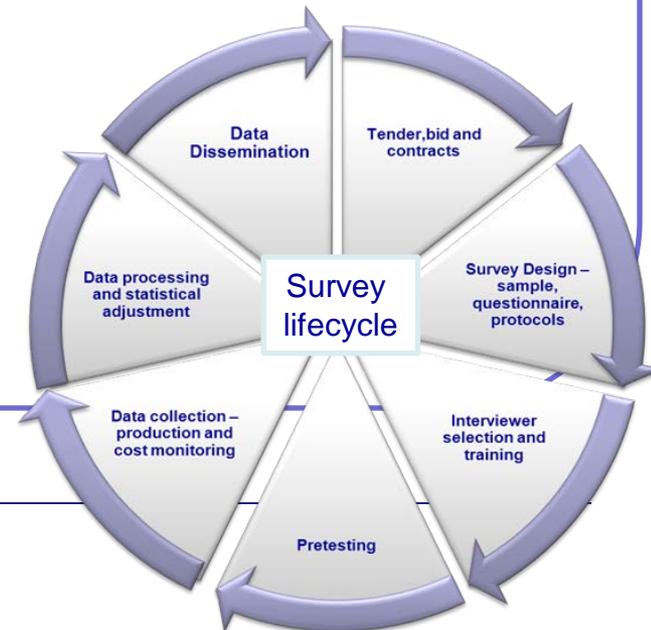
SRO - SRC - ISR

Survey lifecycle

- Tender, bid and contracts
- Survey Design – sample, questionnaire, protocols
- Interviewer selection and training
- Pretesting
- Data collection – production and cost monitoring
- Data processing and statistical adjustment
- Data Dissemination

The primary uses of paradata are to:

- monitor data collection and other study processes,
- identify methodological error in design,
- reduce survey error,
- improve the quality of survey processes and products,
- aid analysis when reviewing survey measures,
- and improve organizational quality.

Survey management tool
Some examples…

- Timings data – interview, session, block or question level and possible comparisons by mode, interviewer or respondent characteristics.

  (e.g. Case study in paper by Joel Devonshire).

- Identify areas in the questionnaire that are "trouble-spots" through the analysis of key stroke data, such as the frequency of the use of remarks by question or the use of short cut keys assigned to access any question specific help screens.

- Observations collected by interviewers (in Blaise) about the environment in which the survey was completed - add context, for e.g. –
  - whether the interview completed in privacy,
  - safety concerns about the location,
  - any impediments to entry to the address,
  - any concerns raise by respondent, or
  - Interviewer rating of the quality of responses.

Data Dissemination

Tender,bid and contracts

Data processing and statistical adjustment

Survey lifecycle

Survey Design – sample, questionnaire, protocols

Data collection – production and cost monitoring

Interviewer selection and training

Pretesting

An example – using Blaise paradata to monitor data quality at an Interviewer level

- Audit trail data.
- The factors were created three "factors" based on an analysis of data from a previous round of data collection.
  - Factor 1: Too Fast
  - Factor 2: Many Error Checks
  - Factor 3: Many 'Don't Know' and 'Refused' responses
- Nine individual performance indicators.

Paradata elements used…
- Field time
- Error escapes, suppressions, jumps
- Backups
- Don't know and refused responses
- Help key use
- Remarks used

Review and action…
- ✓ Small production group meets every two weeks to review interviewer performance.
- ✓ Identify interviewers with indicators that look troublesome.
- ✓ Decide on interviewer-level intervention.
- ✓ Monitor outcomes—look for improvement.

Sample management tool

Some examples of how paradata can be used to inform key decisions …

- Resource planning – how many interviewers are required (Interview length)
- Determining the cost of a survey

Quality Assurance tool

Some examples…

- Evaluate interviewer performance - for example:
    - An analysis of ADTs may detect suspicions of interview falsification.
    - Recordings (CARI) used to undertake detailed assessment of interviewer interviewing skills and adherence to protocol - for e.g. consent question.
    - Interview length – by interviewer, mode, language etc.
- Recover data using ADTs.
- Can facilitate bug identification during testing – for e.g. to replicate a problem, source of the issue – specific system, survey instrument or the user.
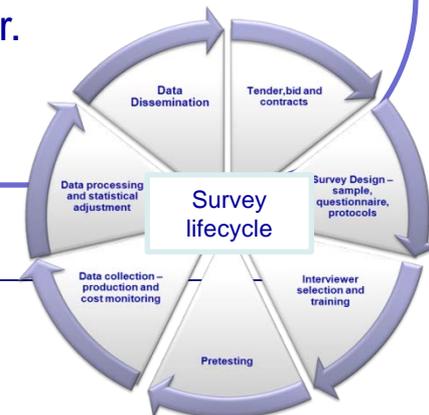
Adhoc analysis

To investigate specific research questions or
to explore an issue in greater depth.
<u>Some</u> examples…

- CAI design improvements – particularly useful at the pilot stage or for question testing.
- To examine the actions on each page of the instrument and the impact these actions have on survey outcomes.
    - For example, on a web self-administered questionnaire;
        - Did the respondent have to scroll?
        - Were questions answered in the order presented (if a grid is used)?
        - Did the respondent use a mouse or keyboard?
        - What was the total number of keystrokes?
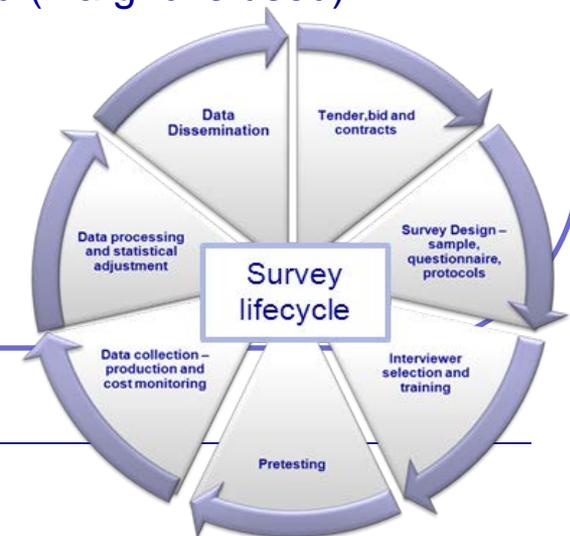
General requirements

- relate to the collection and storage of the paradata rather than specific elements of paradata.

- Retain current level of detail in ADT
- Security -
    - stored securely,
    - user can define which paradata elements are collected,
    - store the paradata separately but retain ability to merge with survey data for analysis.
- Data structure must be consistent across all modes.
- Define key project-specific variables to include in the paradata – for example, stratum, project name or outcome variables

CATI sample management system are not included in the requirements paper – only CAPI and Blaise IS

Specific requirements for CAPI

- Instrument paradata is required for each respondent at the variable, page, block, session and interview level.
- The following are paradata which are required about an interview or specific part of the interview.
  - Mode of data collection at a page or item level as well as at a session or form level.
  - Geographic information about location of the respondent (e.g., GPS coordinates).
  - CARI specific information (e.g., log of questions recorded, name of sound file(s) and/or screen shot(s)).
  - Information about the CARI set up or recordings (e.g., log of questions recorded, name of sound file(s) and/or screen shot(s)).

Detail of specific paradata elements are in an appendix to the paper.

Specific requirements for web data collection

- Information about the environment in which the survey was administered - for example, the type of device, operating system, browser, connection speed and screen or browser size.

- Paradata for web surveys should capture data from both the server and client side:
  - Server side includes submissions to server, page-level times and break-offs.
  - Client side for example, capturing user actions like changed responses, response latencies, keystrokes, mouse movement - for example, mouse coordinates every 'nth millisecond, how the keyboard and scrolling keys were used - such as direction of scrolling.

Detail of specific paradata elements are in an appendix to the paper.

In summary…

- User requirements and extensive use throughout the survey lifecycle emphasise the importance of paradata.

- Consistency in format of paradata and ability to turn and off elements will help tailor the analysis and usefulness of paradata.

- Paradata will continue to be a central part of survey collection operations, the further development of its use and the developments or efficiencies gained through the analysis of paradata are something we can share amongst the Blaise community.

UNIVERSITY OF MICHIGAN

Thank You