Using Audit Trail Data to move from a Black Box Process To a Transparent Data Collection Process
Black Box Process

*process or system whose inputs and outputs are known*

*But*

*where the internal structure is not well known or understood*
Sample Households

Interviewers

Questionnaire

Household Survey
Data Collection
No Visibility of Data Collection Process

Sample Households

Interviewers

Questionnaire

Household Survey Data Collection

Survey Data

“What gets measured gets managed.”
“If you can't measure it, you can't improve it.”
- Peter Drucker
Transparency of Data Collection Process
- Call History Data

Sample Households

Interviewers

Questionnaire

Contact day/time

Duration

Appointment

Survey Data
Transparency of Data Collection Process
- Audit Trail Data

Sample Households

Interviewers

Questionnaire

Survey Data

Contact day/time
Appointment
Durations
Response burden
Route
Interviewer performance
No. sessions
Contact Outcome
Questionnaire performance
Date/time
Transparency of Data Collection Process

Sample Households

Interviewers

New Insights

Process Improvements

Survey Data

Performance Indicators

- Improve operational performance
- Focus on continuous and sustainable improvements

- Contact day/time
- Appointment
- Durations
- Response burden
- Route
- Interviewer performance
- No. sessions
- Contact Outcome
- Questionnaire performance
- Date/time
HOUSEHOLD Surveys Development Project

Traditional face to face interviewing to recruit Households and for more complex surveys

Introducing telephone interviews for LFS repeat waves

Focus on efficiency and innovation through automation and consolidation of processes

Call centre contracts genesis of business problem to be solved
Problem Evolution

**Initial Business Problem**

- Payment for telephone interviews based on interview duration
  - Pricing models vary: costs per call, per agent hour, inclusive per campaign
  - Trade off between quantity and quality
- Initial test phases uncovered substantial problem calculating accurate interview durations
- Conducted quality measure on data
  - Could not calculate duration for 30% of cases using the time stamps
  - Approx. overpayment

**Opportunity**

Alternative to time stamps, use of Audit trail data
- accurate interview durations BUT in a difficult format
  (very large, messy, complex, semi-structured text files)
- previous use limited to recreating corrupt interviews & investigating suspected inappropriate behaviour
Consensus of potential benefits of data but considered too difficult to harvest as data is “large, messy & semi-structured”

– Records vary in length and composition
– Analysis is time consuming and messy
Documented problems converting and processing semi-structured data

- Processing described as difficult, ad-hoc and time consuming
- No optimal solutions
  - data silos
  - extracting durations only
  - pre defined queries
  - sub-setting the data by interviewer or case
Optimum Solution

• Convert audit trail data into a searchable format that can be integrated with other survey admin & paradata for richer information & reports

• Redesign data returns processes to manage audit trail data

• Combination of generic reports and self-service functionality for expert users to explore the data
Planned Use of Audit Trail Data

• Duration as a quality indicator
  – By interview
  – By sessions
  – By question

• Activity analysis
  – movement through the survey instrument, revisiting Qs, correcting answers etc.

• Quality/performance indicators
  – Survey instrument
  – Interviewer performance
  – Data quality
## Development Evaluation

Based on key performance drivers for good data management

- Easy user access & availability of data
- Timely updates
- Integration with other data
- Efficient parsing process

<table>
<thead>
<tr>
<th>Goal</th>
<th>Criteria</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality Attribute</td>
<td>Data Accuracy</td>
<td>Accurate interview durations can be calculated</td>
</tr>
<tr>
<td></td>
<td>Data Timeliness</td>
<td>Data can be updated as part of the nightly processes</td>
</tr>
<tr>
<td></td>
<td>Data Range</td>
<td>Complete set of audit trail data is available to query and explore e.g. interviews by day, time, session, module, actions taken etc.</td>
</tr>
<tr>
<td>Efficacy – capable of</td>
<td>Interview duration</td>
<td>An interview duration report can be created for the call centre manager</td>
</tr>
<tr>
<td>producing desired effect</td>
<td>report</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Performance Measures</td>
<td>Data can be used to manage and measure key performance measures – queries can be written to extract key performance variables</td>
</tr>
<tr>
<td>Utility – usefulness of</td>
<td>Data access</td>
<td>Data is stored in a central repository that can be used in a self-service manner by a variety of users</td>
</tr>
<tr>
<td>solution</td>
<td>Integration with case</td>
<td>Data import process can be integrated with the overnight Case Returns process</td>
</tr>
<tr>
<td></td>
<td>processing</td>
<td></td>
</tr>
<tr>
<td>Validity – execution</td>
<td>Data integration</td>
<td>Data is integrated with other survey admin and para data that can be used to create useful information</td>
</tr>
<tr>
<td>soundness</td>
<td>Efficient process</td>
<td>Integrating the import process with the overnight case returns process does not impact the process performance</td>
</tr>
<tr>
<td></td>
<td>Optimal data searches</td>
<td>Data is in a format that allows optimised searches and queries; with large volumes of data, response times can be slow</td>
</tr>
<tr>
<td></td>
<td>and queries</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rating</th>
<th>Performance Descriptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Meets evaluation expectation</td>
</tr>
<tr>
<td>Medium</td>
<td>Falls short of the evaluation expectation is some degree</td>
</tr>
<tr>
<td>Low</td>
<td>Below evaluation expectation</td>
</tr>
</tbody>
</table>
**Solution Implementation**

**Design, Build, Evaluate**

---

Image of a diagram titled "Solution Implementation" with sections labeled "Design, Build, Evaluate" and "Audit trail file".

---

**Audit trail file**

---

**Data Graph**

---

**Datamodel Iterations**

---

**v0.1**

---

**v0.2**

---

**v0.3**

---

**v0.4**

---
Benefits associated with good quality paradata

- Better, more efficient management of data collection operations

- Measurements for key process variables
  - Instrument performance
  - Interviewer performance
  - Interviewing Activity

- Monitor & Min. known survey errors
  - Interviewer error
  - Non-response error
  - Measurement error

- Improved quality of processes and statistical data
Conclusion

- Improved information management
  - Raised business awareness of value of paradata and provided easy access to the data
- Value of solution extended from calculation of interview duration to transparency around the interviewing process
- Data can be used to improve the
  - design & testing of survey instruments
  - operational performance of data collection processes

Mindset change for business users from ‘data creators’ to ‘data users’