Collection of Paradata in CAPI System

Vesa Kuusela
Social Survey Unit
Statistics Finland
Background

- Paradata did not exist in paper-and-pencil era
- Computer Assisted Interviewing (CAI) systems brought along many possibilities to collect data about the process
  - CATI call schedule systems collect (automatically) data that can be used for quality control or for process development
  - In a CAPI system, paradata collection has to be arranged separately
    - Problem has been the limited telecommunications facilities
- Fast WWAN based telecommunications data transmission has opened new perspective for paradata collection in CAPI systems
Wireless Wide Area Network (WWAN)

- WWAN uses mobile telecommunications network for transferring data
- Based on third generation (3G) of telecommunication hardware standards
  - In Europe, Universal Mobile Telecommunications System (UMTS) is most common of the 3G technologies
    - In theory, UMTS supports up to 21 Mbit/s data transfer rates
  - General packet radio service (GPRS) is a mobile data service for the 2G mobile communication systems.
    - GPRS provides data transfer rates up to 114 kbit/s
- Pricing is based either on amount of transferred data or on fixed monthly fee
- For computer use, a slot for a SIM card is needed either on motherboard or in a connector to USB port.
Paradata from Blaise questionnaires

- All Statistics Finland questionnaires include a standard set of data fields for automatic recording of:
  - interviewer code
  - date and time of interview
  - end result (interview or non-response)
    - reason of non-response with standard coding
  - three start and end times to record the total processing time of each case to calculate the total treatment time
- At the office, paradata from interviews are processed and printed every day with a standard reporting scheme
- For specific needs it is possible to add automatic fields or questions for interviewers to collect paradata
Example of a standard report of a survey (LFS)

<table>
<thead>
<tr>
<th>Interviewer</th>
<th>All cases</th>
<th>Returned</th>
<th>Returned</th>
<th>Interviews</th>
<th>Non-resp.</th>
<th>Over coverage</th>
<th>Non-resp.</th>
<th>Mean time</th>
<th>Total time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0707, Mnnn</td>
<td>25</td>
<td>25</td>
<td>100</td>
<td>23</td>
<td>2</td>
<td>0</td>
<td>8,0</td>
<td>2,74</td>
<td>63</td>
</tr>
<tr>
<td>2147, Rrrrr</td>
<td>10</td>
<td>10</td>
<td>100</td>
<td>8</td>
<td>2</td>
<td>0</td>
<td>20,0</td>
<td>4,63</td>
<td>37</td>
</tr>
<tr>
<td>2668, Nnnnn</td>
<td>20</td>
<td>20</td>
<td>100</td>
<td>19</td>
<td>1</td>
<td>0</td>
<td>5,0</td>
<td>4,79</td>
<td>91</td>
</tr>
<tr>
<td>……</td>
<td>……</td>
<td>……</td>
<td>……</td>
<td>……</td>
<td>……</td>
<td>……</td>
<td>……</td>
<td>……</td>
<td>……</td>
</tr>
<tr>
<td>Total</td>
<td>2139</td>
<td>2139</td>
<td>100</td>
<td>1867</td>
<td>268</td>
<td>8</td>
<td>12,29</td>
<td>4,36</td>
<td>8136</td>
</tr>
</tbody>
</table>

This table is prepared at the end of field period
Same table can be produced also for all regions or based on respondents’ characteristics (e.g. gender or age)
Paradata from interviewer’s user interface

- In interviewers’ user interface is application for recording of all contact attempts with respondents
- Time stamps are automatically attached to each transaction
- Agreed appointments are added to the calendar application
Recorded contact history

- Each transaction is on one row showing:
  - Date and time
  - Contact media
  - Contacted person
  - Time of agreed appointment
  - Remarks

- Contact history is returned automatically with other case data
Possibilities provided by new technology

- Technically, the amount of collected paradata is no problem anymore
  - Very large paradata files can be sent, e.g. audit trail files, audio files and video files
- Many features of CATI case management can be installed in the new CAPI system
  - Some new possibilities for CAPI system design also emerge
What to do with the paradata?

- For what purpose is paradata collected?
  - Quality control and quality assurance?
  - To increase cost-efficiency of data collection?

- How is paradata collection arranged?
  - Automatically
    - Requires standardisation in data collection
  - Interviewers enter appropriate data
    - Causes extra cost if interviewers are paid by hour

- How is the reporting arranged?
  - Automatic reporting
  - Ad hoc –type reporting
Future scenarios

- Technically, CAPI case management with WWAN could be designed similarly as in a CATI system
  - Storing data real-time in a central data base is possible already (on-line interviewing)
- For example, it is possible to build a system where supervisors see in real time when an interviewer is conducting an interview and for what survey