Introductory Remarks on Metadata – Paradata – Data and Quality in Comparative Survey Research

Lars Lyberg
Stockholm University
and
Peter Ph. Mohler
Mannheim University
CSDI 2011 Workshop
London March 2011
Part 1
Meta-Data

Peter Ph. Mohler
Overview

- Definitions
  - Data, meta-, para-
  - Meta-Data
- DDI Alliance - a mega story
- DDI and Comparative Survey Methodology
  - Towards a Meta-Data UFO
  - Towards quality indicator driven comparative survey documentation
facts and **statistics** collected together for reference or analysis: *there is very little data available*

- the **quantities**, characters, or symbols on which operations are performed by a computer...

- philosophy: **things** known or assumed as **facts**, making the basis of reasoning or calculation.

Oxford Online Dictionary 2011
Two Types of Data in Survey Research – A First Approach

- Data
  - Extensive Data
    - Indicate the amount or extension of a fact
    - Numerals
      - Example: frequency of a response in a population
  - Qualitative Data
    - Indicate qualitative facts
    - Non-numerals
      - Example: Job descriptions (ISCO items)

*Note: questions often ask for a qualitative fact (education) that is then coded on the spot into a numeral*
Meta- & Para- Dictionary Entries

- "Meta"
  - "denoting something of higher or second-order kind"

- "Para"
  - "besides, adjacent to"
  - "something that protects or wards of"
Metadata

“a set of data that describes or gives information about other data”

Extensive & Qualitative Metadata

- Example extensive: response rate
- Example qualitative: name of PI, description of quality procedures
The Mega-Story of DDI

- 1999 Miller, Kenneth, and Pasqualino Titto Assini "The XML Files: Developing Data UFO"
- 2001 Ryssevik, Jostein, and Simon Musgrave "From Data Graveyards to Knowledge Greenhouses"
- 2006 Thomas, Wendy "Locating the Geographic Center of DDI 3.0 - Part of DDI for the Next Decade: Toward Version 3.0, to Presentations from the 2nd Annual European DDI Users Group Meeting, Utrecht, Netherlands, December 2010".
New in Comparative Survey Methodology

- Integration of production process quality into the survey life cycle paradigm
- Paradigm change from life cycle to complex production process systems
- Extensive guidelines since 2010
  - [http://ccsg.isr.umich.edu/](http://ccsg.isr.umich.edu/)
Turning the Life Cycle into a Survey Production Process System

Data Dissemination

Data Processing and Statistical Adjustment

Data Harmonization

Data Collection

Pretesting

Survey Quality

Survey, Organizational, and Operational Structure

Tenders, Bids, and Contracts

Sample Design

Questionnaire Design

Adaptation of Survey Instruments

Translation

Ethical Considerations in Surveys

Instrument Technical Design

Interviewer Recruitment, Selection, and Training

Publishing, Archiving

Design of Analyses

Instrument & Sample

Research Question

Adaptation of Sample Design

Copyright Mohler & CSDI
Simplistic View on Survey Production Process

Iverson, Jeremy "Metadata Driven Survey Design"
Presentations from the 2nd Annual European DDI Users Group Meeting, Utrecht, Netherlands, December 2010. (Collectia Sales Talk, PM)
Example Monocultural Integrated Process Quality Monitoring and Control
Meta-Data Comparative UFO

- Acknowledgment of survey process complexity
- Concentration on process stages incl. their quality
- Simultaneous 50+ surveys documentation (and control)
Quality Indicator Documentation

- Quality – see Lars Lybergs presentation
- Just one example of information reduction
- As always is the ESS plagued by “the enemy of the good is the better”...
E23 (ACQ30BS): Category 55 ("It depends") was not fielded in Croatia.

| COUNTRY      | FIELDWORK PERIOD | N  | RESP. RATE* | DEVIATIONS IN DATA
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>13.11.08-20.03.09</td>
<td>1 760</td>
<td>58.9</td>
<td>F6 (EDULVLA), F6 (EISCED), F36 (EDULVLPa), F49 (EDULVLFA), F55 (EDULVLMA)</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>06.03.09-31.05.09</td>
<td>2 230</td>
<td>75.0</td>
<td>F32 (HINCTNTA), F6 (EDULVLA), F36 (EDULVLPa), F49 (EDULVLFA), F55 (EDULVLMA), Gq (IPRSPOT)</td>
</tr>
<tr>
<td>Croatia**</td>
<td>22.12.08–31.03.09</td>
<td>1 484</td>
<td>45.7</td>
<td>E23 (ACQ30BS, E24 (ACQ70BS), F23 (TPORGWK)</td>
</tr>
<tr>
<td>Cyprus</td>
<td>29.09.08-21.12.08</td>
<td>1 215</td>
<td>78.7</td>
<td>F6 (EDULVLA), F32 (HINCTNTA), F36 (EDULVLPa), F49 (EDULVLFA), F55 (EDULVLMA)</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>08.06.09-08.07.09</td>
<td>2 018</td>
<td>69.5</td>
<td>C18 (RLGDNM), C20 (RLGDNME), F6 (EDULVLA) F36 (EDULVLPa), F49 (EDULVLFA), F55 (EDULVLMA)</td>
</tr>
<tr>
<td>Denmark</td>
<td>01.09.08-11.01.09</td>
<td>1 610</td>
<td>53.9</td>
<td>E23 (ACQ30BS), E24 (ACQ70BS), F36 (EDULVLPa), F49 (EDULVLFA), F55 (EDULVLMA), F70 (CHLDHHE)</td>
</tr>
<tr>
<td>Estonia</td>
<td>05.11.08-11.03.09</td>
<td>1 661</td>
<td>57.4</td>
<td>F6 (EDULVLA), F6 (EISCED), F62 (MARITALA), F63 (LVGHWA), F64 (LVGPTNA), F66 (LVGPTNE), F68 (DVRCDDEV)</td>
</tr>
<tr>
<td>Finland</td>
<td>19.09.08-05.02.09</td>
<td>2 195</td>
<td>68.4</td>
<td>B 11 (VOTE), C18 (RLGDNM), C20 (RLGDNME), E11 (P20CNEC), E13 (P70CUST)</td>
</tr>
</tbody>
</table>

The underlying national questionnaire items have changed since ESS3-2008.
Control Chart (example from Lars Lyberg)

- **Natural Variation**
- **Non Natural Variation**
- **Special Cause**
- **Common Cause**
## Special Cause (Non-Natural) Variation

### Comparative Meta-Data Documentation

<table>
<thead>
<tr>
<th>Country</th>
<th>Question</th>
<th>Non-Natural Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>My Country</td>
<td>Q5</td>
<td>Response Category &quot;it depends&quot; missing</td>
</tr>
<tr>
<td></td>
<td>Q200</td>
<td>Question not asked</td>
</tr>
<tr>
<td></td>
<td>Q313</td>
<td>Question did not pass translation quality assessment</td>
</tr>
<tr>
<td>Your Country</td>
<td>Q1</td>
<td>Question not asked</td>
</tr>
<tr>
<td></td>
<td>Q25</td>
<td>Wrong filtering resulting in inadequate target population</td>
</tr>
<tr>
<td></td>
<td>Q32</td>
<td>Response categories reversed</td>
</tr>
<tr>
<td></td>
<td>Q212</td>
<td>Translation comment part of questionnaire</td>
</tr>
</tbody>
</table>
Conclusions

- A Meta-Data UFO (or better Hubble) will be a quantum leap in transparency, measurement and survey quality.
- It requires accepting survey production complexity.
- Complexity of comparative surveys is a power function of number of populations or languages under observation.