Methodological, legal and technical perspectives on the feasibility of web survey paradata in German official statistics

5th ESRA Conference in Ljubljana, 16th July 2013
Session: Use of Paradata for Production Survey Management
Chair: Kyle Fennell

Sabine Sattelberger
Questionnaire Testing, Survey Methodology
sabine.sattelberger@destatis.de
Web data collection at Destatis

- Until now: ~ 200 of 380 statistics p.a. also covered online (mostly in business statistics)
- Future self-obligation to provide secure online reporting systems as first mode for mandatory surveys

- Implementation by in-house maintained IT-application (IDEV 4)

- Quality assurance in statistical production starts with instrument:
  - Standardized guidelines for appropriate web survey design
  - Iterative testing and further development
Multi-stage approach of qualitative testing

1. Expert review (functionality & technical performance)

2. Usability testing under laboratory conditions:
   - Observation
   - Cognitive interviewing
   - Eye-tracking
     - What probands do,
     - what they think,
     - what they do (not) perceive

3. Questionnaire evaluation (data quality & response burden)
   (based on Tries 2011)
Potential use of web survey paradata (I)

**Definition:**
- Very detailed quantitative process / observational data
- Generated in computer-assisted data collection, here: CAWI
  - Server-side: log files documenting every server access
  - Client-side: JavaScript recording all events within page
  (based on Callegaro 2013 and Heerwegh 2011)

**Status quo at Destatis:**
- IT uses log files to monitor run of application and server security
- Methodologists have not yet used any paradata for pretesting
Potential use of web survey paradata (II)

Main benefits as a supplement to qualitative testing:

- Application for business statistics
- In pilot studies or statistical production
- Self-completion in natural environment
- Low reactivity
- High sample sizes: scope and severity of detected problems (based on Dumas/Redish 1999)
Why have web survey paradata not yet made the final leap to a well-established testing method?
Resisting the “Babylonian confusion”

Different in-house perspectives need to be harmonized before starting to implement web survey paradata.

- Methodologists: What is useful?
- IT: What is feasible?
- Legal experts: What is permitted?

In the end:
We do talk the same language (strategic goals of NSI).
Some methodological challenges

- Quality standards for recording (e.g. timestamps)
- Analysis techniques to distill informative value and recognize patterns:
  - Theory-driven indicators (e.g. for data quality or response burden)
  - Creative visualization (e.g. of navigation profiles)
  - Multiple levels of analysis (e.g. questions – respondents – devices; chronological – anomalous events)
  - Flexible benchmarks (e.g. to identify problematic questions or outliers)
- Evidence of lasting added value for new and established surveys
- Combination of methods to overcome limits in gain of knowledge (e.g. no reasons for detected problems, local-temporal distance to respondents)
Some technical challenges

- Overview of available web survey paradata

- New functionality to be embedded into given IT-environment:
  - Constant URL and other NSI specialties
  - Performance of complex web questionnaires
  - Server stability

- Automated output adapted to methodological needs:
  - Manageable raw datasets
  - Ready-to-use key indicators
  - Quality assurance

- Solutions for omitting paradata of respondents who opt out
Some legal challenges (I)

- Web survey paradata – personal data?

- Right to “informational self-determination” strongly forms German public opinion about data protection (Tschwerwinka 2012)
  - e.g. Census, Google Street View, Facebook’s privacy policy, Microsoft Xbox One, Prism, EU General Regulation…

- Following national law and “professional ethics” (e.g. International Statistical Institute 2010, UN Statistical Commission 1994)
  - Need to adapt rules of etiquette to non-reactive methods

- Informed consent prior to each survey participation
Some legal challenges (II)

- “Exit strategies” to disconnect survey participation and paradata use, especially in mandatory surveys:
  - Paper questionnaires in mixed-mode data collection
  - Explicit opt-out option integrated in informed consent

- “No” to informed consent refers to methodological use only (probably neither possible to stop recording nor technical use)?

- Analysis for submitted web questionnaires only (probably neither for lurkers, drop-offs, time-outs nor intermediate storages)?
Conclusions: Preliminary assessment of web survey paradata

**Efficient testing method?** Input ⇔ output compared to qualitative testing:
- Similarly high implementation costs that pay off over time
- Complementary, not replacing insights into self-completion
- Potential use for additional purposes (e.g. adaptive design, mixed-mode research)

**Feasible testing method?**
NSIs should consider technical possibilities and legal regulations as soon as possible to solve practical challenges early in process

In times of limited resources, much depends on long-term, in-house support (e.g. by management and subject-matter statisticians)
➡ Pilot studies to test new method and specify benefit-cost-estimates
Thank you for your attention.

Any questions or comments?

Sabine Sattelberger
Questionnaire Testing, Survey Methodology
sabine.sattelberger@destatis.de
Acknowledgments

My sincere thanks for a gainful exchange of experiences to

Corina Teichmann, Anja Tilp and Kay Sommer (Destatis)
Lars Kaczmarek (DE)
Mattijn Morren and Ger Snijkers (NED)
Gustav Haraldsen and Øyvin Kleven (NO)
Mick Couper and Rachel Horwitz (USA)
References


Dumas, Joseph S. / Redish, Janice C., 1999: A practical guide to usability testing. Intellect Ltd.


