



# Paradata - from by-product to standard documentation



12th Annual European DDI User Conference - 01-12-2020

Ingo Barkow (FHGR), Geneviève Michaud (Sciences Po), Anja Perry (GESIS), David Schiller (FHGR), Wendy Thomas (IPUMS)

DOI: [10.5281/zenodo.4299076](https://doi.org/10.5281/zenodo.4299076)

# Paradata - from by-product to standard documentation

- first definition
  - modes of collection
  - what paradata are or are not
  - examples
  - potential usage
  - barriers to publication
  - use cases
  - primary use
  - potential of reuse
  - the call for a model
  - from by-product to standard documentation
  - DDI framework
  - next steps for this group
-

# first definition

There is no clear definition of paradata, but it is mainly viewed as **auxiliary data collected in a survey as a by-product that describe the collection process**. Although such a concept is not limited to a specific discipline, it is often discussed **linked to survey research**.

# how paradata are collected

Depending on the mode of data collection:

- automatically, manually
- at different levels:
  - interviewer or survey agency
  - respondent level (self-administered surveys)
  - item level (passive data collection)

# what paradata are or are not

- paradata are not auxiliary data
  - stratum identifiers, demographic features of Census tracts, data from commercial sources
  - data from different sources
- some metadata are in fact paradata
  - response rates for example

# paradata examples

- call records, contact history and disposition codes
  - successful interview, hard refusal, non-contact
- audio-recordings, verbal paradata
  - pauses, voice pitch
- response behavior in online survey
  - keystrokes, navigation within the survey, correction of answers
- sensor data
  - such as GPS data

# potential paradata usage

- improve survey quality
- field monitoring, and costs control
- included in total survey error measures
- post-survey assessment or corrections of errors
- paradata-driven basis to improve survey quality in a longitudinal study

# barriers to paradata publication

- post-processing is resource-consuming,
- no standard format,
- ethics and legal concerns:
  - in European studies, GDPR requires respondent's "specific and informed" consent prior to data collection,
- dramatic increase in **re-identification risk**.



# paradata stories and use cases

- Multiple modes of paradata generation
  - PIAAC
- Log files from wearables
  - SensoMot
- Survey methodology
  - CRONOS
  - ELIPSS

# paradata stories and use cases

- Paradata on clinical trials
- Capture protocols
  - clinical trials template
- Integration/Harmonization
  - IPUMS experience
- Privacy Preservation Environment - Statistics Canada

# paradata primary use

- verification,
- quality control content management,
- data processing,
- international normalization,
- data analysis,

Overall, **paradata support the evaluation of the primary data collection** and the processes used to capture it.

## potential of reuse

What our uses cases show is a huge potential for reuse and analysis.

In fact, paradata is often not released.

To fully realize this potential, we need a model framework, i.e basic requirements.

This model would be used as a guideline for researchers during the first steps of study design.

# the call for a model

This would help to move paradata from a totally internal information resource to a research object that can be published and reused.

## from by-product to standard documentation

This would help to move paradata from a totally internal information resource, an under-(re)used, internal by-product, to a research object that can be published and reused.

# DDI framework

Given the role of DDI as a common form for the "primary" data, **a DDI expression of a paradata model** is both appropriate and very desirable for structuring, accessing and using paradata.

## next steps for this group

We are currently considering three papers to start the discussion (all working titles):

- Leveraging paradata from capture to analysis (in preparation)
- Formalizing paradata as a standard (only title, no content)
- Paradata, the DDI view (only title, no content)



## next steps for this group

Furthermore, we are waiting for the new scientific board to start operating. We are considering to apply to become a new official working group about paradata.

If you are interested in joining, please contact David Schiller

[david.schiller@fhgr.ch](mailto:david.schiller@fhgr.ch)

**Thank you!**