

## **Collecting and publishing paradata on the German Internet Panel**

*Annelies Blom*

University of Mannheim

The increase in the collection of paradata in recent years is based on the assumption that they have predictive power to explain survey errors in coverage, nonresponse, measurement or coding. In addition, their use is frequently justified by the ease with which they can be collected during the survey process.

In probability-based online panels, the potential for collecting paradata is large – spanning paradata on the recruitment process in general and, at each wave, on the respondents' computer settings, nonresponse patterns, question time stamps and key strokes. While the collection of paradata is typically connected with low additional costs, the publication of some types of paradata requires detailed data preparation. Furthermore, warranting respondents' anonymity and informed consent to data collection demands from researchers that they deal carefully with these data; this might inhibit the publication of some potentially available paradata.

This presentation showcases how the German Internet Panel (GIP) faces the challenge of paradata in terms of their collection and publication. We demonstrate how some types of paradata are easily made available to the research community, while others are associated with high additional costs and ethical concerns.

The GIP ([http://reforms.uni-mannheim.de/internet\\_panel/home/](http://reforms.uni-mannheim.de/internet_panel/home/)) is based on a random probability sample that is recruited offline and represents both the online and offline population aged 16 to 75 in Germany. It was first set up in 2012 with a refresher sample in 2014. Once online, panel members are interviewed every two months via web-questionnaires on social, political and economic topics. The GIP data are published via the GESIS Data Archive for the Social Sciences within 6 months of their collection.