

Ethical and Legal Issues related to Paradata*

Background and Description

With the increasing use and further development of technological means in the context of survey-based data collection, the amount of information collected about the process of survey production has increased. Especially, in connection with the use of CAPI techniques and the implementation of web surveys much *paradata* are generated. Furthermore, survey researchers are also increasingly making use of paradata, such as keystroke data or contact protocols.

Paradata are key data for analysing data quality in survey production; they are used "to evaluate and improve survey instruments but also to understand respondents and how they answer surveys" (Couper and Singer, 2013: 57).

So far, however, paradata are rarely made available to the scientific community (an exception is the ESS¹). "[One] obstacle to releasing paradata are unclear legal and ethical considerations." (Kreuter 2013: 8). Existing codes of ethics are not very clear on the issue of paradata (Couper and Singer 2013: 58), and also from a legal perspective in many cases it is not clear under which conditions paradata should be collected and how they may be used and released.

Deliverable 6.2 explores the extent to which the increasing and more structured collection, processing and use (including linkage and release) of different types of paradata impose *legal and ethical challenges* to survey researchers. The collection and processing of paradata in SHARE² serves as a case study for this analysis.

Definition and Differentiation

In order to cover the entire range of legal and ethical issues that might be connected to the collection, processing and use of paradata, the deliverable relies on a broad working definition of paradata, which includes all *micro-level data about the process of survey production*.

This definition covers two types of paradata:

- a) Process paradata: data about the process of survey production recorded as a by-product in the course of conducting a survey, such as listing information, keystrokes, contact data and gross sample data.
- b) Auxiliary paradata: additional data about the process of survey production obtained separately from external sources or with a specifically targeted effort to enhance the information on the survey production process, such as interviewer observations, information on the interviewers and external supplementary data about the sample cases.

Paradata in SHARE

In SHARE, for example, four different types of paradata are collected, processed and linked currently (item-level time stamp data, contact information, interviewer observations and interviewer demographics). They are used for fieldwork monitoring purposes and in order to evaluate and improve the survey instruments. The practical example from SHARE illustrates the increasing importance to systematically investigate the ethical aspects and legal requirements related to different types of paradata.

* DASISH WP6 Deliverable 6.2: Schmidutz, D. and Bristle, J. (2013). Sample Merged Paradata Sets: Ethical and Legal Issues. Available at: <http://dasish.eu/deliverables/>.

¹ European Social Survey. For further information see: <http://www.europeansocialsurvey.org/>

² The Survey of Health, Ageing and Retirement in Europe. For further information see: <http://www.share-project.org/>

Findings and Conclusions

While on the one hand "[t]he number of surveys that collect and provide paradata is growing quickly, and [...] new applications and monitoring systems are develop[ed currently]" (Kreuter, 2013: 8), on the other hand legal and ethical issues remain unclear. In this regard, it is assumed that, if there are claims that the collection and use of paradata is an issue of ethical concern – even if this currently is a contested area – the subject at least needs ethical consideration.

In general, since paradata by definition are always closely related to the process of survey production, two key ethical principles of survey research have to be considered when collecting, using and releasing paradata: *to assure the autonomy of the respondents* and *to prevent respondents from harm*. In practice this means that researchers have to obtain informed consent of their respondents prior to data collection and that they have to ensure the confidentiality of the participants' data.

As there are several types of paradata that can be collected as well as different ways of recording and linking these data, however, the legal and ethical issues that are connected to the collection, processing, use and re-use of paradata *require a nuanced approach*.

Stages of the Research Process	Data Collection and Usage of Paradata	Data Processing and Release of Paradata
<i>General Issues</i>	<i>Obtaining Informed Consent</i>	<i>Ensuring Confidentiality</i>
Important aspects regarding the context of data collection and use and the types of paradata involved	<ul style="list-style-type: none"> ▸ <u>Process paradata</u>: unavoidably collected as a by-product of survey production; may be implicitly covered by consent to participate in survey; in certain cases, however, may be used as information about the respondents ▸ <u>Auxiliary paradata</u>: additionally collected; may or may not constitute information relating to the respondents 	<ul style="list-style-type: none"> ▸ <u>Paradata sets</u> may include both, direct identifiers and indirect identifiers ▸ If <u>paradata sets are linked</u> (to survey data, e.g.) relational data might lead to a disclosure of respondents' identities ▸ <u>Certain types of paradata</u> may be classified as sensitive or confidential; interviewers' rights might be concerned as well
Specific legal and ethical questions to be considered on a case-by-case basis	<ul style="list-style-type: none"> ▸ <u>Process paradata</u>: Would respondents consent to the (intended/anticipated) use of the paradata? ▸ <i>Whether, how and to what extent should participants be informed about the capture and use of paradata?</i> ▸ Are respondents aware of the capture and use of paradata and how will such awareness impact on their behaviour? ▸ <u>Auxiliary paradata</u> (furthermore): <i>Should additional consent be obtained?</i> 	<ul style="list-style-type: none"> ▸ Have all direct identifiers been removed from the paradata sets as early as possible? ▸ Has the entire data environment been considered carefully prior to the linking and release of paradata sets? ▸ What is the appropriate level of anonymisation/access in relation to the type/s of paradata concerned? ▸ Have all relevant European/national/regional legal regulations been taken into account in relation to all data subjects?

Further Reading

- Couper, M.P. and Singer, E. (2013). "Informed Consent for Web Paradata Use." *Survey Research Methods* 7(1), pp. 57-67.
 Kreuter, F. (2013). "Improving Surveys with Paradata: Analytic Uses of Process Information." Wiley.