Data Service Infrastructure for the Social Sciences and Humanities

EC FP7

Grant Agreement Number: 283646

DASISH Work Package 6: "Legal and Ethical Issues"

Deliverable: D6.3
Deliverable Name: "Exemplary analyses of confidential paradata"
Deadline: 31st December 2014
Nature: Demonstrator

Responsible: MEA (MPG)
Work Package Leader: MEA (MPG)

Contributing Partners and Editors: Daniel Schmidutz, MEA (MPG)
Johanna Bristle, MEA (MPG)

Exemplary Analyses of Confidential Paradata:

Ethical and Legal Considerations

CONTENT

1 Introduction

The deliverable "Exemplary analyses of confidential paradata" of work package 6 "Legal and Ethical Issues" (WP6) of the "Data Service Infrastructure for the Social Sciences and Humanities" (DASISH) project deals with ethical and legal aspects that have to be considered when analysing confidential paradata that is generated in the process of survey production. It builds on and continues the work of deliverable D6.2 ("Sample merged paradata sets") of the DASISH project.

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. In particular, in connection with the use of computer-assisted interviewing (CAPI) techniques and the implementation of web surveys much paradata, i.e. micro-level data about the process of survey production[1], 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 and are used for different purposes, ranging from the evaluation and improvement of survey instruments to a better understanding of respondents and their answers in surveys (cf. Couper and Singer, 2013: 57). Consequently, a strong demand from researchers and, in particular, the survey methodology community to make paradata of surveys available can be observed recently.

In deliverable D6.2 the extent to which this increasing and more structured collection, processing and use of different types of paradata impose legal and ethical challenges to survey researchers has been explored. A central finding of deliverable D6.2 is that legal and ethical issues that are connected to the collection, processing, use and re-use of paradata require a nuanced approach. Since there are different kinds of paradata that can be collected (depending on the survey mode and the technical system in place) and specific kinds of paradata are/can be used for certain analyses only, legal and ethical questions need to be answered on a case-by-case basis (cf. Schmidutz and Bristle, 2013: 16-19).

The current demonstrator takes account of this finding and discusses ethical and legal issues in relation to a few concrete practical examples of paradata usage. In doing so, a special focus of deliverable D6.3 lies on paradata which can be classified as 'confidential data'[2] and on the questions of how these data may be used and made accessible for re-use to researchers of the scientific community. In order to demonstrate how certain ethical and

[1] In this report we refer to a broad concept of *'paradata'*, which includes *'process paradata'* as well as *'auxiliary paradata'*. For a detailed definition and differentiation of paradata please see chapter 3 of this report. For a detailed explanatory description please see deliverable D6.2 of the DASISH project (Schmidutz and Bristle, 2013; available at: http://dasish.eu/deliverables/).

[2] *'Confidential data'* can be understood as information, which is protected against unwarranted disclosure for issues pertaining to personal privacy or for proprietary considerations.

legal aspects in relation to confidential paradata may be considered on a case-by-case basis, three practical examples from the Survey of Health, Ageing and Retirement in Europe (SHARE)[3] are introduced and discussed.

Example 1 focuses on the use of paradata for fieldwork monitoring purposes. It addresses the questions of how contact related paradata about cases in which either no contact to a target person could be established or in which the target person refused to participate in a survey may be used for analyses, although no explicit consent has been obtained from the target persons. Furthermore, the use of item-level time stamp data as an indicator for standardised data collection is addressed. Example 2 concerns the framework conditions under which paradata documenting the outcomes of contacts with target respondents (i.e. information on the cooperation process) may be used as information about interviewers' work performance as part of research of survey methodological interests. Finally, example 3 discusses the ethical and legal aspects connected to the analysis of keystroke paradata as information about respondents in the context of substantial scientific research, i.e. when being used in order to enhance the data provided by respondents in the course of a survey.

2 Scope and Objectives

2.1 Overall Objectives of Work Package 6

WP6 addresses various legal and ethical issues that modern research in the social sciences and humanities (SSH) is confronted with. Following the "Description of Work" (DoW), Annex 1 to the Grant Agreement of the DASISH project, WP6 has the following main objectives:

- To identify the legal and ethical issues, constraints and requirements for all data types occurring in the SSH domain as result of data integration and linking,
- To cope with legal and ethical challenges imposed by the new data types emerging in the social sciences and humanities,
- To look for professional long-run preservation strategies and policy-rules that can be applied to data collections in the social sciences and humanities.

2.2 Scope of Deliverable D6.3 in the Context of Task 6.1

Task 6.1 of WP6 of the DASISH project particularly is concerned with data types imposing "New Ethical and Legal Challenges" to the social sciences and humanities. It concentrates on the identification of new ethical challenges and legal requirements related to the various data types being recorded in modern research in the SSH domain.

[3] For further information see the website of the SHARE project: http://www.share-project.org/.

This demonstrator, as part of Task 6.1, builds on and continues the work of deliverable D6.2. ("Sample merged paradata sets"). While deliverable D6.2 focussed on the compilation and linkage of paradata and aimed to identify legal and ethical issues connected to the collection and use of paradata in general, deliverable D6.3 focuses on a few concrete practical examples of paradata usage based on SHARE data. It aims to demonstrate in an exemplary manner ethical and legal considerations in relation to the use and re-use of confidential paradata on a case-by-case basis.

As in deliverable D6.2, paradata are broadly defined as ***micro-level data about the process of survey production***, including [a] data about the process of survey production recorded as a by-product in the course of conducting a survey (*'process paradata'*), such as listing information, keystrokes, contact data and gross sample data, as well as [b] 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 (*'auxiliary paradata'*), such as interviewer observations, information on the interviewers, external supplementary data about the sample cases, etc. (cf. Schmidutz and Bristle, 2013).

While paradata themselves cannot be considered as a new type of data in the field of population-based survey research, obviously "a more structured approach in choosing, measuring, and analyzing key process variables is indeed a recent development" (Kreuter, 2013: 2; cf. Couper and Lyberg, 2005). It can be said that paradata only recently attracted the full attention of researchers conducting field surveys when realising the methodological and scientific value of this data. However, since 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), while on the other hand legal and ethical issues remain unclear, it is of increasing importance to systematically investigate the ethical and legal aspects related to different types of paradata.[4]

In WP6 special attention is given to this topic in the context of this deliverable and the previous deliverable D6.2.[5] With regard to these two deliverables WP6 closely cooperates with WP3 ("Data Quality"). As in the case of the compilation of a merged paradata set (D6.2), for which existing data sources from SHARE were used, practical examples from SHARE are also used in order to illustrate the analyses of confidential paradata that require legal and ethical considerations (D6.3).

[4] It is noted that this currently is a contested area – while some authors claim that the collection and use of paradata is an issue of ethical concern, others argue that the collection and use of paradata does not entail ethical issues at all. In this regard, it is assumed that, if there are claims that the collection and use of paradata is an issue of ethical concern, this subject at least needs ethical consideration.

[5] Besides, the "Report about new IPR Challenges" also addresses legal and ethical issues related to the collection and the use of paradata in the context of transnational survey research. Cf. chapter 7.3.2 "Using and Releasing Paradata (SSc)" of DASISH deliverable D6.1 (Schmidutz et al., 2013; available at: http://dasish.eu/deliverables/).

The outcome of deliverable D6.3 will add practical examples to the discussion of ethical and legal issues related to paradata of deliverable D6.2 by illustrating how confidential paradata may be used and re-used in different settings taking into account relevant ethical and legal aspects on a case-by-case basis.

3 Ethical and Legal Issues related to Paradata[6]

While not only the amount of paradata collected in surveys has increased but also an increasing use of paradata by survey researchers can be observed over the recent past, many legal and ethical issues related to paradata still remain unclear. Especially with regard to the release of paradata "unclear legal and ethical considerations" (Kreuter, 2013: 8) remain. According to Kreuter, up to now, only a few researchers have started to address this issue and even these authors state that "[e]xisting ethical codes are not very clear on the issue of paradata" (Couper and Singer, 2013: 58). Moreover, from a legal perspective, it is in many cases not clear under which conditions paradata should be collected and how they may be used and released.

In general, when collecting, using and releasing paradata, two key ethical principles of survey research should be taken into account[7]: first, to assure the autonomy of the respondents, which means obtaining informed consent of respondents prior to data collection, and second, to protect respondents from harm, which in survey research typically means ensuring the confidentiality of the participants' data (cf. Singer, 2008: 85).

Since existing types of paradata as well as the ways of collecting them differ substantially from each other, ethical and legal consideration requires a nuanced approach. Furthermore, depending on the specific type of paradata concerned, the measures to be taken in order to ensure appropriate acknowledgement of the key ethics principles and legal requirements may differ from case to case. For example, with regard to paradata that are unavoidably collected in the process of survey production[8] usually the only relevant question is whether respondents would consent to their 'use' (cf. Couper and Singer, 2013: 65), while with regard to paradata that are obtained separately from external sources or with a specifically targeted effort[9] the question whether additional[10] consent of the respondents to their collection has to be obtained is of relevance as well.

---

[6] This chapter summarises the findings of deliverable D6.2 of the DASISH project (Schmidutz and Bristle, 2013: 11-19; available at: http://dasish.eu/deliverables/), which provide the theoretical background for the following exemplary discussion of ethical and legal considerations in relation to paradata usage.

[7] It is noted that ethics principles lied down in 'Codes of Ethics' for survey researchers do not constitute rules. According to Denscombe,"[t]he point is not that each principle should be followed, but that it should be taken into account and considered. [...] The principle should be acknowledged." (Denscombe, 2002: 176)

[8] I.e. *'process paradata'*.

[9] I.e. *'auxiliary paradata'*.

In most cases of paradata collection and use, however, survey researchers face two ethical and legal questions: Firstly, the questions of whether, how and to what extent participants should be informed about the capture and the use of paradata and of how much detail should be provided to them. And secondly, the question of how and under which conditions different types of paradata may be released for scientific re-use. In relation to both issues, particularly the 'intended use' of the paradata in question appears to be crucial.

TABLE 1 summarises important ethical and legal aspects regarding the general issues of obtaining informed consent and ensuring confidentiality as well as specific questions that should be considered when collecting, using, processing and releasing certain types of paradata.

TABLE 1: ETHICAL AND LEGAL ISSUES RELATED TO PARADATA COLLECTION, PROCESSING, USE AND RE-USE.

Stages of the Research Process	Data Collection and Usage of Paradata	Data Processing and Release of Paradata
General Issues	***Obtaining Informed Consent***	***Ensuring Confidentiality***
Important aspects with regard to general ethical and legal issues	‣ Process paradata: unavoidably collected as a by-product of survey production; may be implicitly covered by consent to participate in a survey; in certain cases, however, they may be used as information about the respondents ‣ Auxiliary paradata: additionally collected; may or may not constitute information relating to the respondents	‣ Paradata sets may include both, direct identifiers and indirect identifiers ‣ If paradata sets are linked (to survey data, e.g.) relational data might lead to a disclosure of respondents' identities ‣ Certain types of paradata may be classified as sensitive or confidential; interviewers' rights might be concerned as well
Specific ethical and legal questions to be considered on a case-by-case basis	‣ Process paradata: Would respondents consent to the (intended/anticipated) use of the paradata? ‣ Whether, how and to what extent should participants be informed about the capture and use of paradata? ‣ Are respondents aware of the capture and use of paradata and how will such awareness impact on their behaviour? ‣ Auxiliary paradata (furthermore): Should additional consent be obtained?	‣ 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/nation-al/regional legal regulations been taken into account in relation to all data subjects?

[10] It is assumed that respondents have consented to participate in the survey and therefore to the collection of the information provided by them in the course of the interview.

Considering that different kinds of paradata exist, which can be used for certain kinds of analyses only, these questions need to be answered on a case-by-case basis, taking into account the specific kind of paradata, the concrete context in which these data are collected and the way in which they are or are intended to be used and released.

4 Analyses of Confidential Paradata

Chapter 4 describes three concrete examples of different kinds of analyses of confidential paradata from SHARE. For each example it describes the content and purpose of the analysis as well as the specific type of paradata that is used for this analysis. Furthermore, it shows why confidential paradata is needed or of added value for the specific research purpose. The examples are taken up again in chapter 5 as part of the exemplary discussion of ethical and legal considerations in relation to paradata usage.

In SHARE four different types of paradata are collected and processed:

- Item-level time stamp data (which are based on keystroke data),
- Contact information (day, time, outcome),
- Interviewer observations (including information on building type, accessibility of the building and some additional neighbourhood characteristics such as vandalism and public transportation) and
- Interviewer characteristics (i.e. additional information on the interviewers).

These data are used for fieldwork monitoring purposes and in order to evaluate and improve the survey instruments. Up to the present day, only interviewer observations and interviewer demographics from the first wave of SHARE are released. They were released together with the survey data since these data also have been collected by asking questions in the course of the survey and therefore were part of the SHARE questionnaire. Subsequently, most of these data were not collected as part of the survey anymore and no paradata has been released, apart from interviewer observations on the interviewer-respondent relation and characteristics of the building[11].

Chapter 4.1 provides an example on how paradata can be used during fieldwork for monitoring purposes. Besides monitoring current fieldwork, paradata are also frequently used to analyse survey participation and survey quality in the retrospective. Data from previous surveys and survey waves are used to inform survey practice of future surveys and waves. In chapter 4.2 an example of such research of survey methodological interest is shown. Finally, paradata can be used to enhance survey data for substantial research purposes. Chapter 4.3 provides an example of such research, which investigates cognitive

[11] These characteristics include information on the building type, on the area where the building is located, as well as information on the number of floors of the building and steps to the entrance.

decline in older age and – in order to do this – uses paradata to enhance the information provided by respondents in the course of the SHARE survey.

4.1 Example 1: Fieldwork Monitoring

Most commonly paradata are used during survey production for monitoring the fieldwork, including the data production progress and the evaluation of interviewer performance on a regular basis. The main purpose of monitoring fieldwork is to learn more about what happens during fieldwork and to enable survey managers to intervene if problems occur or room for improvement is spotted. If up-to-date paradata is available it can, e.g., be used for implementing responsive designs to guide data production efficiently and improve data quality (Groves and Heeringa, 2006). At this, paradata are essential in order to develop successful strategies for improvement of the fieldwork during the fieldwork phase since the information they contain about the process of data collection provides indicators for the assessment of quality of the ongoing fieldwork and therefore the collected survey data.

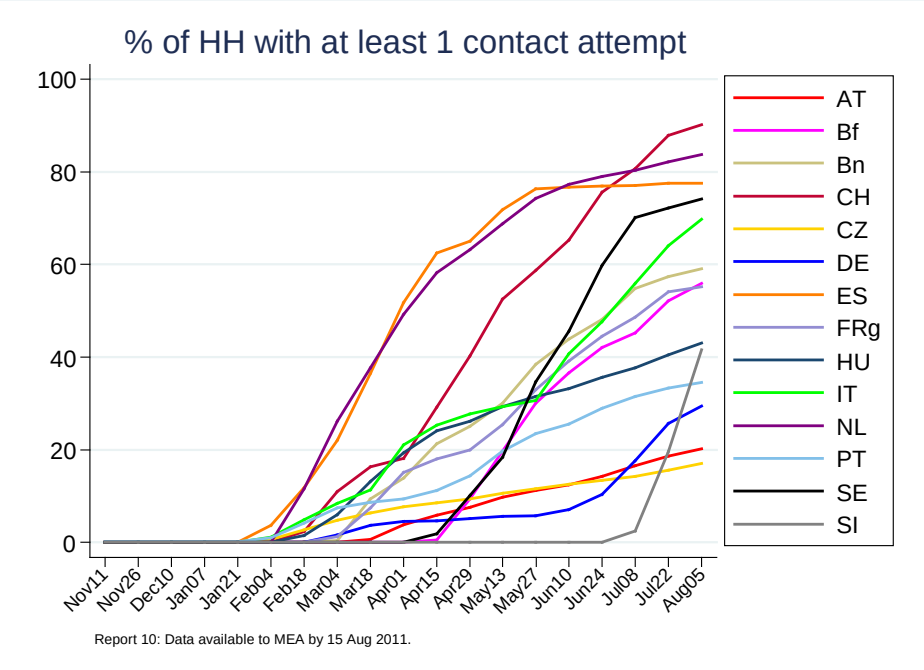
The main data source for fieldwork monitoring across countries in SHARE[12] is contact information from the sample management system (SMS). The SMS tool is used by the interviewers to document every contact with a household or individual respondent or enter result codes for every contact attempt that was not successful (e.g. "no contact", "contact-try again", or "refusal"). Most analyses for fieldwork monitoring purposes in SHARE are based on this *contact information*. Exemplarily, two indicators from this data source which are used to assess fieldwork progress are presented below: [a] the development of contact attempts over time as an indicator for countries' strategies of contacting households and [b] the outcomes of the contact attempts as an indicator of willingness to participate in the survey in the different countries. In addition, [c] *item-level time stamp data* is used to capture interview length or item length of introduction texts. The latter is an indicator for standardised data collection and may be used to assess if interviewers read out introductions properly.[13]

[12] The Survey of Health, Ageing and Retirement in Europe (SHARE) is a cross-national panel study. In the 4th wave of data collection the study was conducted in 19 countries. It is noted that this example only discusses the fieldwork monitoring that is carried out by the central coordination team of SHARE. Further monitoring based on country-specific paradata is conducted by the SHARE country teams (responsible for the implementation of the survey in their own country). Every country team is able to access the files of their own country only. The results are distributed in form of a fortnightly report among the members of the survey infrastructure, i.e. the participating country teams and their survey agencies, and some of them are presented at internal SHARE meetings. In Addition selected results are published as part of the first results books on the methodology of conducting SHARE (cf. Börsch-Supan and Jürges, 2005; Schröder, 2011; Malter and Börsch-Supan, 2013b).

[13] Further details on fieldwork monitoring and the use of keystroke data for this purpose can be found in chapter 4 of the DASISH deliverable D3.7 "Keystroke Analysis and Implications for Field Work" (Bristle and Halbherr, 2014; available at: http://dasish.eu/deliverables/).

[a] Information on contact attempts is analysed descriptively over time. As part of fieldwork monitoring the aggregated results are provided to the SHARE country teams, which are responsible for the implementation of the survey in the single countries. They are displayed graphically and show progress over time and across countries. **FIGURE 1** shows the percentage of households in the SHARE gross sample with at least one contact attempt. This does not imply an interaction between the interviewer and the potential respondent yet, but merely describes if an interviewer started working on a case. The time range reported on the x scale covers roughly two third of the fieldwork period of SHARE wave 4.[14] Countries "differed in their strategies of contacting households. Some countries had very steep increases from the get-go, whereas others only very gradually increased their contact attempts." (Malter, 2013: 130). Ideally, at the end of fieldwork all countries should at least have attempted to contact each household once (i.e. achieve 100 % in **FIGURE 1** below); especially with regard to the longitudinal sample. Since attempting to contact is the first step in the process of survey participation (which is followed by establishing contact, and finally cooperation), only if this step is taken there is a chance to obtain respondent's cooperation in the survey. This first step is exclusively in the sphere of influence of the interviewers and the survey agency. Monitoring this process enables survey managers, who aim to minimize non-participation, to intervene if necessary and directly influence participation at this first step in a favourable manner.

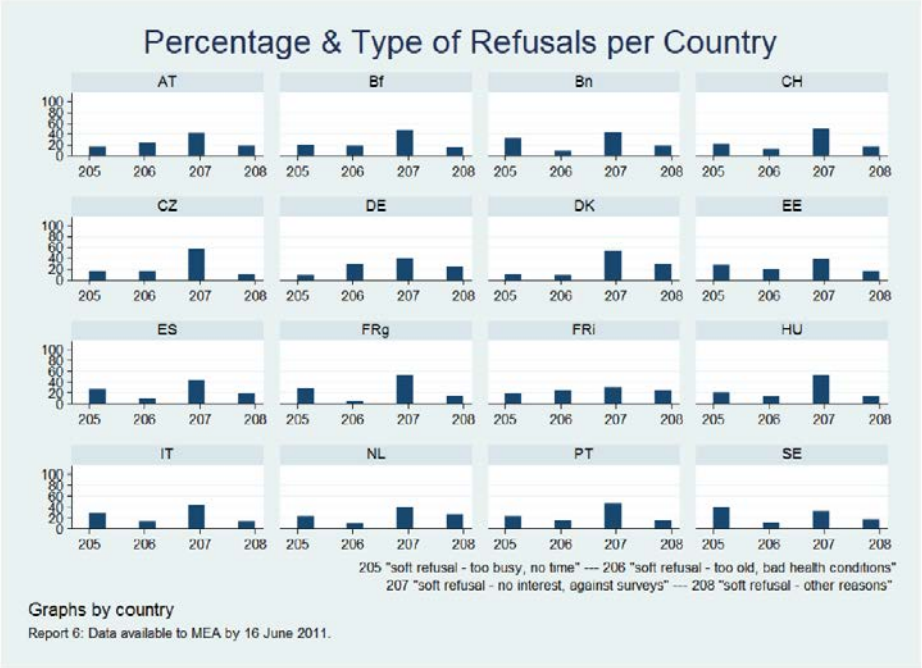
FIGURE 1: COUNTRIES' STRATEGIES OF CONTACTING HOUSEHOLDS. (SOURCE: MALTER, 2013: 130.)



[14] The countries with low percentages of contact attempts in August were not finished with the fieldwork at this point in time. E.g. Austria continued fieldwork until December 2011 and Germany until February 2012. In the Czech Republic the rates shown are low because an extremely large gross sample was drawn and the number of contacts attempts was reported in relation to the size of the originally drawn gross sample.

[b] Fieldwork monitoring further looks at the main reasons why respondents refuse to participate in the survey. Response rates and cooperation rates are often used as key numbers for data quality.[15] Collecting and analysing data about the reasons for non-cooperation can help to tailor and improve strategies on survey participation. In SHARE, interviewers code the outcome of every contact attempt according to a pre-defined list of contact codes in the SMS. Several of them refer to different reasons for refusals, which is valuable information for further contact attempts. The history of refusal outcomes helps the interviewer to tailor the next attempt to obtain respondent's cooperation in accordance with the code that was set previously. In **FIGURE 2**, the percentage and type of refusals is documented per country. The main reasons given were related to the categories "too busy, no time", "too old, bad health conditions", "no interest, against surveys" or "other reasons". In almost all participating countries, most persons who refused stated that they are not interested in the survey and therefore did not participate in an interview. Analysing the refusal codes during fieldwork can help to understand the reasons for reluctance of target persons and to take adequate measures, such as retraining of interviewers, if necessary.

FIGURE 2: PERCENTAGE AND TYPE OF REFUSALS PER COUNTRY. (SOURCE: MALTER, 2013: 132.)

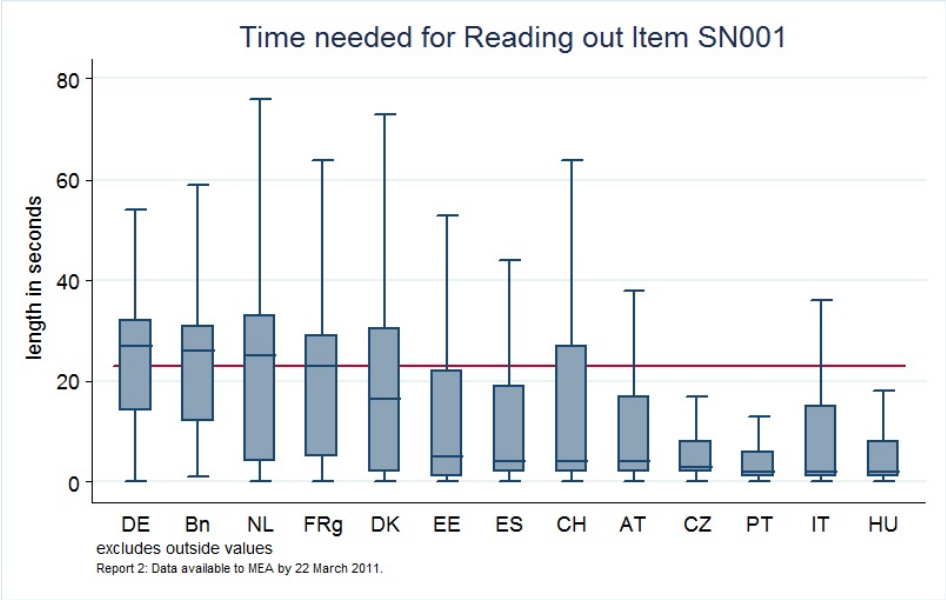


[c] Proper reading of entire introduction texts in interviews is commonly considered as a characteristic of good interviewer behaviour, and is an indicator for compliance with standardised data collection. For fieldwork monitoring purposes, times spent on reading out introduction texts were compared with normative standards (cf. red line in **FIGURE 3**). Ideally

[15] According to Heerwegh paradata are used "to describe and classify response behavior […] or to relate response behavior to data quality" (2002: 2).

there would be little variability between interviewers within a country[16] as well as between countries. FIGURE 3 below shows that the median of four countries are higher than the normative reading time of 23 seconds (which is the time it takes to properly read out the English generic text), while most of the countries show a very low median and a highly right-skewed distribution. This clearly shows that interviewers either cut the introduction text or skipped it completely since "[l]anguage differences alone cannot explain these stark differences and right-skewed distributions" (Malter, 2013: 137). Such findings suggest that there is room for improvement. If available in a timely manner, survey managers are able to adjust their strategies or take necessary actions (e.g. retraining of interviewers) within or across countries. Being able to monitor fieldwork through the use of real-time paradata clearly provides the basis for improvement with regard to standardised interviewing.

FIGURE 3: TIME NEEDED TO READ THE INTRODUCTION TO THE SOCIAL NETWORKS MODULE (VARIABLE SN001). (SOURCE: MALTER, 2013: 137.)



4.2 Example 2: Research of Survey Methodological Interest

In general, paradata is used for understanding and improving survey management. One key indicator often used for determining the quality of survey data is response rates, which has been illustrated in example 1 to some extent already. Since there is a trend that response rates are decreasing worldwide, and especially in Europe, it is important to put more effort into understanding nonresponse and response patterns. For such analyses that are needed beyond fieldwork monitoring in order to improve survey management, survey methodologists also mainly rely on paradata.

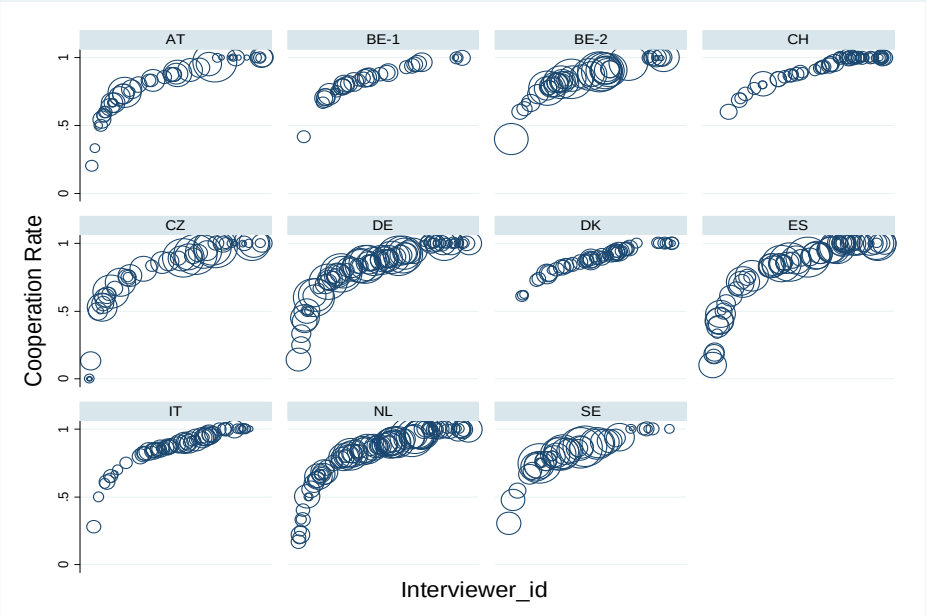
[16] Proper reading of the entire text would result in a boxplot which is rather high and short. This means it would be centred on a rather high median (close to the red line; cf. FIGURE 3) and show a short interquartile range and rather short whiskers.

An example of how paradata is being used in survey methodological research is the paper by Bristle et al. (2014) on "The Contribution of Paradata to Panel Cooperation in SHARE". The paper aims to understand respondents' and interviewers' behaviour in the process of survey participation in analysing cooperation in the fourth wave conditional on participation in the previous wave of SHARE. In order to understand (1) how previous interview experience shape current response behaviour and (2) how interviewers and country-specific fieldwork strategies influence respondents' decisions, paradata is needed.

Using multilevel models, the authors find that factors at different levels (survey agency coordinator, interviewer and respondent) influence cooperation. At the highest level, they highlight the importance of everyday communication between survey agency coordinators and interviewers to gain cooperation. At the interviewer level, even if a sizable part of the variance remains unexplained, interviewers' quality of work and experience significantly affect cooperation propensity. Respondents' prior interview experience and the interviewer-respondent interaction therein has a large influence on the re-cooperation decision overall. Such insights about nonresponse processes are of a great value for researchers in survey methodology and survey practitioners.

As a concrete example for paradata usage in the paper, the descriptive analysis of interviewer effects based on *contact information* can be highlighted. **FIGURE 4** shows cooperation rates of interviewers. Here, each subgraph represents one survey agency (conducting the fieldwork in the different SHARE countries) and each circle represents one interviewer. Interviewers are ordered according to their cooperation rate (0 = 0% cooperation rate; 1 = 100% cooperation rate). The size of each circle represents the interviewer's workload in terms of number of cases contacted.

FIGURE 4: INTERVIEWER-SPECIFIC COOPERATION RATE BY SURVEY AGENCY. (SOURCE: BRISTLE ET AL., 2014: 13).



From these data three conclusions can be drawn: (1) Interviewers differ substantially in their cooperation rates. (2) The differences between interviewers vary across countries.[17] (3) The fieldwork strategy in terms of workload assigned per interviewer varies across countries.[18] Based on the first conclusion, the authors conducted multivariate analyses, which support the descriptive finding of interviewer effects and highlight driving factors of the cooperation processes on the interviewer-level (cf. Bristle et al., 2014).

Interviewer information[19] (i.e. additional paradata) has been used here to investigate which characteristics of interviewers are related to success in gaining cooperation. To understand better the role of the interviewer with regard to the cooperation process is important for all survey researchers who employ interviewers (whether in face-to-face or in telephone interviews). New insights gained in this field through the use of paradata can help survey managers to make investments into training and to make selections based on empirical evidence.

4.3 Example 3: Research of Substantial Interest

Besides using paradata as pure information about the survey process as this is the case in the examples described in chapters 4.1 and 4.2, paradata may be of a great value for substantial research as well. An example in which paradata from SHARE is used to enhance survey data is a paper by Mazzonna and Peracchi (2012), in which the authors investigate the relationship between "Ageing, Cognitive Abilities and Retirement".

Following the human capital theory, the authors expect cognitive decline to increase after retirement: "The fact that retired individuals lose the market incentive to invest in repair activities may cause an increase in the rate of cognitive decline after retirement" (cf. ibid., 2012: 692). The SHARE questionnaire measures several dimensions of cognitive abilities and includes tests on orientation, immediate and delayed recall, fluency and numeracy. In addition, the authors use paradata, here *keystroke data*, to enhance the scores of the cognitive tests with information about the time the respondent needed to perform the test. Based on theories related to cognitive decline, Mazzonna and Peracchi (2012) argue that the concept of the cognitive abilities can be measured more accurately when taking a time measure into account. They state:

[17] There are cases in which cooperation rates range from 0.5 to 1 and only very few interviewers show poor performance. For other survey agencies, interviewers differ more in their cooperation rates.

[18] Looking at the size of the circles, it can be noticed that in some countries the workload is equally distributed among interviewers (e.g. BE-1, CH, DK or IT) while in other countries there are survey agencies where the workload assigned per interviewer varies (e.g. AT, DE, ES or SE).

[19] The interviewer information that has been obtained from the survey agencies "includes demographics (year of birth, education, gender) and [interviewers'] previous experiences in conducting SHARE interviews" (Bristle et al., 2014: 7). Interviewers' education level was provided by some survey agencies only, which have been ISCED-97-coded and exploited to run robustness analysis with this subsample of agencies.

> "We use the time spent on cognitive questions in a novel way, namely as a measure of a respondent's processing speed, a second dimension of cognitive abilities evaluation. As argued by Salthouse (1985), ageing is associated with a decrease in the speed at which many cognitive operations can be executed. The keystroke files allow us to capture this characteristic of cognitive deterioration." (Mazzonna and Peracchi, 2012: 693)

During the interview every entry via the keyboard is captured in keystroke files. Every time a key is pressed on the keyboard of the laptop, this action is registered and stored by the software in a text file. From these text files, time stamps on item-level can be computed. Mazzonna and Peracchi (2012) use this information to adjust the cognitive ability score. In **TABLE 2** the raw scores of the cognitive tests are displayed next to the adjusted scores, which take keystroke information into account. The added value of using keystroke data is obvious: The adjusted scores provide a more precise measure. The variance of the measure is increased and subtle distinctions can be revealed.

TABLE 2: RAW AND ADJUSTED COGNITIVE SCORES. (SOURCE: MAZZONNA AND PERACCHI, 2012: 695.)

Mean and standard deviation (S.D.) of raw and adjusted cognitive scores.

	Mean	S.D.	Correlations				
Raw scores							
Orientation	3.88	.37	1.000				
Recall imm.	5.21	1.67	.106	1.000			
Recall del.	3.80	1.90	.118	.663	1.000		
Fluency	20.56	7.29	.081	.374	.345	1.000	
Numeracy	2.61	1.02	.125	.326	.297	.315	1.000
Adjusted scores							
Orientation	3.50	.45	1.000				
Recall imm.	4.89	1.69	.112	1.000			
Recall del.	3.43	1.87	.129	.635	1.000		
Fluency	20.56	7.29	.110	.386	.350	1.000	
Numeracy	2.25	1.04	.152	.331	.302	.327	1.000

Mazzonna and Peracchi (2012) use the adjusted scores in their further analyses. They begin with describing the deterioration of cognitive abilities and then test their hypotheses with multivariate models. As a demonstration, the descriptive results of educational differences in cognitive decline are plotted in **FIGURE 5**.

FIGURE 5: AGE PROFILES OF TEST SCORES. (SOURCE: MAZZONNA AND PERACCHI, 2012: 696.)

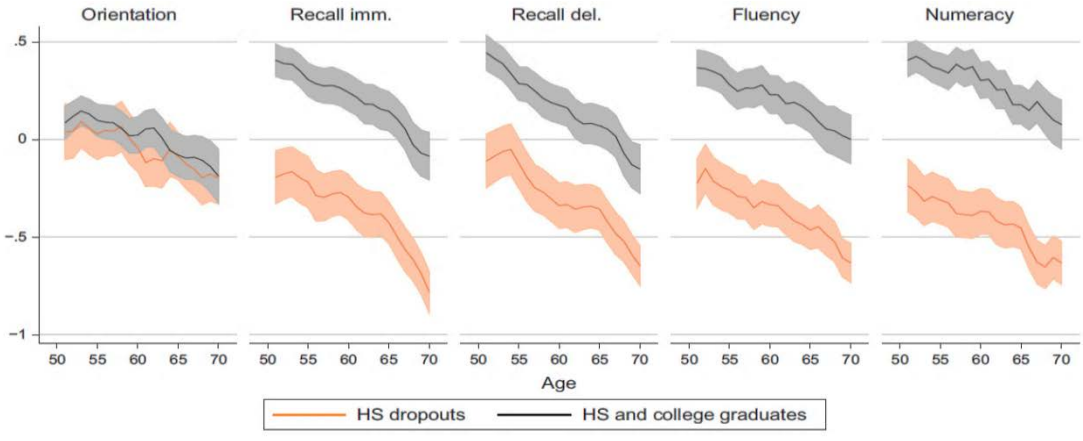


Fig. 2. Age-profiles of average test scores by education level.

Over all five measures of cognitive abilities a clear decline over the age span is apparent. Furthermore, the figure shows a level difference between respondents with high school or college degree (HS and college graduates) and those without a high school degree (HS dropouts) for all measures except for orientation. The same age-specific trend can be predicted using multivariate analyses and is robust towards several robustness specifications. In conclusion, their "findings show an increase in the rate of decline of cognitive abilities after retirement. In the light of [their] theoretical framework, this reflects the reduced incentives to invest in cognitive repair activities after retirement" (Mazzonna and Peracchi, 2012: 709).

5 Ethical and Legal Considerations

In connection with the usage and analyses of SHARE paradata as described in the previous chapter, the ethical and legal aspects related to the specific kinds of paradata and the concrete cases of paradata usage have been explored.

In accordance with the main finding of Schmidutz and Bristle (2013) legal and ethical questions have been considered on a case-by-case basis taking into account the specific paradata concerned, including the way of paradata collection as well as actual and potential use cases, as well as the data environment in which collection, processing, usage and release of the paradata are taking place.

At the beginning of these considerations regarding the concrete practical examples the paradata used have been classified in accordance with the differentiation of Schmidutz and Bristle (2013: 6-7), bearing in mind that the measures to be taken in order to ensure appropriate acknowledgement of the key ethics principles and legal requirements may differ from case to case. In all examples *'process paradata'* – i.e. data that are unavoidably collected as a by-product of survey production – have been used for analysis. Only in one of the examples *'auxiliary paradata'* are used (in addition to process paradata). Since process paradata in general do not capture respondents' behaviour outside the survey[20], and the auxiliary paradata used in example 2 only consist of interviewer characteristics[21], i.e. of information on interviewers and not about respondents, no additional consent (besides consent to participate in the survey) of the respondents to their collection has to be obtained. Therefore, with regard to the key ethics principle of assuring respondents' autonomy, the main question is whether the respondents concerned would consent to their use (cf. Couper and Singer, 2013: 65). In this connection, the questions of whether, how and

[20] According to Couper and Singer, the capturing of process paradata can be understood as "nothing more than collecting information about the process of completing a survey that is already covered by the informed consent statement for the survey itself" (2013: 59).

[21] Year of birth, gender, (ISCED-97-coded) education and previous experiences in conducting SHARE interviews.

to what extent participants should be informed about the capture and the use of paradata have been considered.

With regard to the second key ethical principle of protecting respondents from harm the question of how and under which conditions the paradata used in the different examples can be released for scientific re-use has been carefully examined. In relation to this question not only the 'intended use' of the paradata in question (such as fieldwork monitoring) but also all potential uses that might be made of such data appear to be crucial.

5.1 Paradata as 'Information regarding the Survey Process'

Both example 1 and example 2 (chapters 4.1 and 4.2) are concerned with methodological aspects of survey research: In example 1 different kinds of paradata (namely: contact information and item-level time stamp data) are used for fieldwork monitoring purposes; in example 2 contact information is used in connection with survey methodological research. They are similar in that they take paradata for what they are by definition: (micro-level) data about the process of survey production. In both cases paradata are considered and used in their capacity as information regarding the survey process. In none of the illustrated cases paradata are turned into data, i.e. information about respondents (cf. Couper and Singer, 2013: 57). Therefore both cases are rather unproblematic regarding the question of whether respondents would consent to the use for the illustrated purposes. It can be argued that this kind of paradata usage is covered by respondents' consent when participating in the survey.

However, as far as persons are concerned who did not consent to participate in the survey, which is the case in example 1, when contact related paradata about cases in which either [a] no contact to a target person could be established or [b] in which the target person refused to participate in the survey are being used, this questions might need further consideration. In the first case, which concerns contact attempt data only, it can be argued that this information (if processed in an anonymous form, i.e. without any details of the target persons[22]), does not constitute personal data[23] of the target persons as defined in the European "Data Protection Directive" (95/46/EC)[24], but rather information about the person/institution who made the contact attempt. Thus, data on contact attempts can be

[22] It is noted that in SHARE, all direct identifiers such as names, addresses, postcode information, telephone numbers are removed from the datasets before these are made available to researchers (even before being processed to the SHARE team). Paradata are processed in the same way and are only available in pseudonymised form.

[23] In the Directive, personal data is broadly defined and refers to "any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity" (95/46/EC, Article 2a).

[24] "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data".

used and made available for re-use as far as it concerns respondents' rights and the ethical responsibilities of survey researchers towards them.[25]

In the second case, however, even though, the paradata concerned (i.e. reasons of refusals as outcomes of contacts) does not constitute information that poses a risk of inflicting harm on the target persons, it could be argued that the ethical principle of assuring the autonomy of human subjects is touched when using and, especially, releasing such data on a micro-level. Regarding the use for pure fieldwork monitoring purposes this argument, however, cannot be made if the target persons voluntarily provide information about their refusal (especially if being asked to specify the reason for their refusal); if this is the case, it can be assumed that they consent to the further processing of this information in the process of survey production. The argument however may still apply as far as it concerns the public release of this data, since it cannot be necessarily assumed that the provision of this information includes consent to its further dissemination. Therefore, up to the present date, this information has been classified as confidential[26] in SHARE and consequently no micro-level data on refusals has been released. Only a selection of aggregated results has been published as part of the first results books on the methodology of conducting SHARE (cf. Malter, 2013: 132) and the SHARE Compliance Profiles (cf. Malter and Börsch-Supan, 2013a) in order to make the process of survey production more transparent to data users and the survey data community.

5.2 Paradata as 'Information on Interviewers'

In the third case that has been illustrated as part of example 1 [c], item-level time stamp data is used as an indicator for standardised data collection as part of the fieldwork monitoring. This case as well as example 2, in which contact information is used for an analysis of cooperation rates, constitutes a special case of paradata usage: In both cases, paradata are not only used as 'information about the survey process' but also as 'information on the interviewers'. Here as well as with regard to the use of interviewer characteristics, which are collected in order to enhance the information on the survey production process, very specific ethical and legal issues have to be considered (cf. Schmidutz and Bristle, 2013: 17-18). It is obvious that the interviewers themselves are data subjects and have to be considered in this role as well. According to Schmidutz and Bristle,

> "researchers do not only have to ensure the confidentiality of the data collected in the survey and obtain informed consent of their respondents, but also have to

[25] Since contact information also can be used in order to assess interviewers' work performance (cf. example 2), however, interviewers' rights and needs have to be considered as well (see below).

[26] The classification of data as *'confidential'* means, that these data are considered as information, which is protected against unwarranted disclosure. The data may be regarded as confidential for reasons pertaining to personal privacy or for proprietary considerations.

> consider these issues with regard to the interviewer. Furthermore, besides ethical issues and data protection requirements, in some cases (depending on the information included in the interviewer profiles and on the way in which these are obtained) national employment legislation has to be considered." (Schmidutz and Bristle, 2013: 18)

In SHARE, these issues are mainly the responsibility of the contracted survey agencies, which employ the interviewers. Firstly, the survey agencies provide the central coordination team of SHARE as well as the respective national SHARE country team with this information and are responsible to ensure that this data processing is performed in compliance with all relevant European and national legal provisions, including national employment legislation (which usually includes certain provisions regarding the protection of employee data).[27] Secondly, the national contracts that are concluded with the survey agencies with regard to the SHARE data collection previous to each survey wave explicitly include fieldwork monitoring. In accordance with the SHARE model contract, the survey agencies have to ensure that

> "[i]nterviewers are closely monitored for timeliness, data quality and falsifications. Each interviewer is assigned a unique identification number, which she or he will use when working with the SMS or the CAPI. This will be assessed by SHARE using SMS data at the end of fieldwork and published in the SHARE Compliance Profiles." (Source: Specifications and Deliverables 2014, Annex 1 to the Main Data Collection Contract for SHARE Wave 6: 8)

This provision, amongst others[28], ensures that the SHARE fieldwork management team is provided with all necessary data in order to be able to perform the fieldwork monitoring in an appropriate manner. It includes the use of item-level time stamp data (from the CAPI) for the assessment of the performance of interviewers as well as the use of information on the cooperation process (from the SMS) on an interviewer-level. Besides, the monitoring procedure is made transparent from the very beginning:

---

[27] Regarding this issue, the Main Data Collection Contract for SHARE Wave 6 (2014: 5) specifies: "[...] Each of the parties shall comply with the provisions of EC Council Directive 95/46/EC and any recent associated national laws (as amended from time to time) relating to the protection of personal data. [… A]ll data delivery must be in accordance with the national and European Data Protection Laws of the country specified in section 1 of this contract. […] The SHARE country team leader has the overall responsibility that national legal requirements of data confidentiality laws are fulfilled. In addition [the survey agency] is responsible that the legal requirements of data confidentiality laws are fulfilled as long as the data is collected, processed or used by [the survey agency]. If implementation of the European regulations concerning data confidentiality has not (yet) taken place in the respective country, the European regulations shall be applied directly."

[28] With regard to the delivery of interviewer information, e.g., the following provision is included in the SHARE data collection contracts with the survey agencies: "Interviewer CAPI experience, training attendance and further details on interviewers will be demonstrated to SHARE Coordination by SURVEY AGENCY submitting the interviewer roster (Deliverable SA11, based on Deliverable SHARE12) which must contain data on all trained interviewers independent of their actual activity for SHARE." (Source: Specifications and Deliverables 2014, Annex 1 to the Main Data Collection Contract for SHARE Wave 6: 8)

> "During field work of the main test SHARE will send out Fieldwork Monitoring Reports every other week to inform all survey agencies and country teams about the current state of fieldwork and suggest solutions to problems with fieldwork progress or data quality." (Source: Specifications and Deliverables 2014, Annex 1 to the Main Data Collection Contract for SHARE Wave 6: 9)

Furthermore, regarding the collection and use of contact information, e.g., the following provisions are included:

> "All specifications on contacting households will be assessed with SHARE SMS data after end of fieldwork and published through the SHARE Compliance Profiles." (Source: Specifications and Deliverables 2014, Annex 1 to the Main Data Collection Contract for SHARE Wave 6: 12)

> "For each telephone or in-person contact or contact attempt with the sample member or members of their household, or other informants such as neighbours, interviewers shall record the date of the call or visit, time of the call or visit, result code, which describes the call or visit outcome, contact type (telephone versus in-person), and interviewer comments about the call or visit if necessary. This information shall be entered into the electronic SMS. Interviewer comments should be sufficiently detailed so that someone other than the interviewer can understand the sequence and nature of calls and visits to a sample respondent. A set of standard result codes will be provided for classifying the outcome of each case. All quality control will be based on data provided to SHARE through the SMS." (Source: Specifications and Deliverables 2014, Annex 1 to the Main Data Collection Contract for SHARE Wave 6: 12)

These provisions also include that cooperation rates may be made publicly available, whether this is done in order to make the process of survey production more transparent to data users and the survey data community or as part of research of survey methodological interest, as illustrated in chapter 4.2.[29] It should be noted, however, that if such contact information is released on an interviewer-level (even if this is done in an aggregated form as regards the respondents) all measures necessary to ensure data privacy have to be taken. In this connection, anonymisation and pseudonymisation are central security measures to ensure confidentiality of the data in their capacity as 'information about the interviewer'.

[29] As far as item-level time stamp data are concerned, the question of how and under which conditions these paradata can be released is addressed, when considering if and how keystroke data may be made accessible for re-use in the following chapter (5.3). It has turned out to be of crucial importance to consider different and even potential cases of paradata usage in relation to this kind of paradata, before a final conclusion can be drawn.

5.3 Paradata as 'Information about Respondents'

In example 3 (chapter 4.3) paradata, namely item-level time stamp data (computed from keystrokes), are used to enhance survey data in the context of substantial scientific research. When considering ethical and legal issues connected to this example one aspect of this use of paradata is of crucial importance: the fact that item-level time stamp data are being used as 'information about respondents'. This means that the keystroke data are being used to enhance other information provided by respondents in the course of SHARE.

While, according to Couper and Singer, in general there is "no consensus on whether, or under what conditions, respondents should be informed that paradata are being collected and may be used[; a]rguably, they ought to be informed if researchers plan to use such data in conjunction with other information provided by respondents in order to make inferences about individuals. In other words, as the paradata (information about the process) are turned into data (information about respondents), informed consent issues may arise" (Couper and Singer, 2013: 57).

Against the background of these findings, prior to the realisation of this research, it has been examined if the use of paradata in case of example 3 "rises to a level needing explicit mention to respondents" (ibid., 2013: 66). At the time when the process paradata was collected during fieldwork the use of the keystroke data in this form has not been intended yet. Therefore, no specific consent for the analyses of keystroke paradata as information about respondents in the context of substantial scientific research on cognitive decline has been obtained and no specific information about the capturing and usage of paradata has been given to the respondents back then. In the light of this situation, the relevant question that needed to be answered (before carrying out this substantial research) is whether respondents *would have* consented to the use for the purposes of the research illustrated in example 3. In order to answer this question in an appropriate manner, several aspects have been considered: First, the content of the concrete research, including the different kinds of data used for analyses. Second, the data provided by respondents in the course of SHARE in combination with the consent they give with regard to a later use of this data. And third, the ratio between the two.

As mentioned in the description of example 3 in chapter 4.3, in SHARE several tests measuring the cognitive abilities of participants are included, namely tests on orientation in time, memory (immediate and delayed recall), fluency and numeracy. All respondents in SHARE are volunteers and the entire data collection is based on informed consent. Consent to participate explicitly includes the use of the data provided by respondents for scientific research purposes. This holds on two levels. Before starting the interview of each wave, each respondent's consent is obtained with regard to his/her participation. Additionally, during the interview answers to all questions are voluntary; each single question or test can be skipped if an individual does not want to answer a specific question or participate in a

specific test. Thus, respondents who participated in the various tests have consented to the use of the data collected in the course of these tests for scientific research related to cognitive abilities. These data provide the basis of Mazzonna's and Peracchi's research on the relationship between "Ageing, Cognitive Abilities and Retirement". When using "the time spent on cognitive questions […] as a measure of a respondent's processing speed [they use it to measure] a second dimension of cognitive abilities evaluation" (Mazzonna and Peracchi, 2013: 693); i.e. they are not using the paradata to measure and analyse something other than cognitive abilities – but only enhance their analyses, which is covered by respondents' consent, by adding a supplementary dimension in order to measure cognitive abilities more accurately.[30] In this regard, it is most unlikely that the additional piece of information, that not only the answers provided by respondents but also the speed of answering are used for analyses of cognitive abilities, would have resulted in a decision not to participate in the tests of those respondents who participated in them.

Since there are no plausible or reasonable assumptions why respondents who consented to participate in the survey in general and in the tests in particular – and thus decided to provide researchers with a lot of information about themselves and their cognitive abilities – would object to the use of the collected keystroke paradata for the concrete purpose of the illustrated substantive research, it can be concluded that paradata may be used in this concrete case.

With regard to the question of how and under which conditions item-level time stamp data can be released for scientific re-use, however, it is crucial not only to consider the actual or intended use of the paradata in question but also all potential uses that might be made of the paradata. This, however, is difficult to assess:

> "Since making paradata available to the public or the entire scientific community, would indeed not only make it necessary to consider the 'intended use' but also to consider all ways in which the released paradata possibly could be used […] it appears to be difficult for survey researchers to assess the most appropriate way of releasing certain paradata." (Schmidutz et al., 2013: 52)

Furthermore, the nature of the data has to be considered. Concerning item-level time stamp data, the example of Mazzonna's and Peracchi's use[31] shows that 'sensitive information'[32] can be concluded from this kind of paradata, when being used as information about respondents. It may be interpreted as health-related information with regard to an individual, which according to European data protection laws is regarded as sensitive. For

[30] It is noted that respondents who refused to participate in these tests as a matter of course have been excluded from the analyses of Mazzonna and Peracchi. Neither data nor paradata has been collected in these cases.

[31] In contrast to the use of item-level time stamp data that is illustrated in example 1 [c].

[32] 'Sensitive data', can be understood as being of a particularly risky nature with regard to possible negative outcomes when being revealed to unauthorised others.

these reasons, this information has been classified as confidential in SHARE so far. Accordingly, no keystroke data is publicly released on a micro-level.

However, as illustrated in relation to example 3, keystroke paradata may be used for certain kinds of substantial research. In order to enable such research, other levels of access providing for special access restrictions, such as on-site use[33], remote data access (RDA)[34], special usage restrictions, such as 'end user licences', or a combination of both, may provide an option. To facilitate research that includes the use of confidential paradata, SHARE currently offers the possibility to conduct certain paradata analyses during a visit as a guest researcher, dependent on a prior evaluation of the concrete research project and subject to special conditions of use[35], which are tailored to the intended use of paradata in the context of the respective research project. This concerns research of methodological interest as well as research of substantial interest.

6 Summary and Concluding Remarks

Only few researchers have started to address ethics and legal aspects in relation to the collection, use and release of paradata, even though these are still unclear in many cases. Chapter 5 of this demonstrator builds on the theoretical work of deliverable D6.2 (Schmidutz and Bristle, 2013) and tackles ethical and legal questions on a case-by-case basis with respect to specific examples of paradata usage from SHARE (as illustrated in chapter 4).

The outcomes of the ethical and legal considerations with regard to the exemplary analyses of confidential paradata support the finding of Schmidutz and Bristle (2013) that ethical and legal questions cannot be answered in relation to paradata in general but need to be explored on a case-by-case basis. Different ethical and legal aspects have to be considered depending on [1] the specific kind of paradata, [2] the way of their collection, [3] the group of human subjects about which they provide information and [4] the purpose, for which they may be used.

With regard to the examples that have been discussed, the following differences within these 4 dimensions could be identified:

[1] Three different types of paradata have been used in the examples: Contact data (contact attempts, outcomes of contacts), item-level time stamp data

[33] I.e. analyses of data in separate secure workplaces for guest researchers.

[34] RDA allows researchers to submit their own computer programs to research data centres (RDCs). At the RDCs, these will be run on the confidential micro-data sets. Subsequently, after having been scrutinized for confidentiality, the results are returned to the researchers.

[35] Guest researchers are required to fill out and sign a "Statement concerning the use of internal SHARE data including paradata" and an "Obligation of confidentiality" in accordance with national data protection law.

(based on keystroke data) and interviewer characteristics (interviewer demographics and interviewing experience).

[2] In principle two different ways of paradata collection have been identified: Paradata recorded as a by-product in the course of conducting a survey (i.e. process paradata), and additional paradata obtained separately from external sources or with a specifically targeted effort (i.e. auxiliary paradata). At this, only one external source has been considered; namely the survey agencies, which provided the information on their interviewers.

[3] The different types of paradata have the potential to provide (resp. reveal) information about different groups of human subjects: Contact information may provide information about sampled target persons, respondents and interviewers. Keystroke data may constitute information about respondents or interviewers, and interviewer characteristics obviously only constitute additional information on interviewers.

[4] Three general fields of paradata usage have been considered: Fieldwork monitoring (which tries to understand and improve strategies of contacting households, response rates and cooperation rates, standardised data collection, e.g.), research of survey methodological interests (interested in the cooperation process and interviewers' work performance, e.g.) and substantial scientific research (with the purpose of enhancing the data provided by respondents, e.g. on cognitive abilities).

With regard to both use and release of paradata the two key ethics principles of assuring data subjects' autonomy and of protecting them from harm have been considered. Besides, some specific legal aspects concerning the examples have been touched upon. E.g., with regard to paradata that constitute information on interviewers, it has been highlighted that permissible use and release of certain paradata depends on national data protection and employment legislation as well as on contractual agreements between survey managers and survey agencies or interviewers (cf. chapter 5.2). While making use of paradata is possible in all of the discussed examples, the outcomes with regard to the question of how and under which conditions the different kinds of paradata used in the examples can be released for scientific re-use vary from case to case.

As far as paradata also constitute personal data (besides being data about the process of survey production), when processing these data "particular importance has to be placed on the compliance with European and national/regional data protection law as well as on the safeguarding of sensitive data and confidential information" (Schmidutz and Bristle, 2013: 14-15). While in this connection, anonymisation and pseudonymisation are central

measures[36] that can be taken by researchers in order to ensure data confidentiality, with regard to the re-use of sensitive or confidential information data access and usage restrictions may be considered as additional safeguard measures.

In general, finding out how and under which conditions paradata can/may be released for scientific re-use, appears to be the most challenging task in relation to different kinds of paradata; in particular, since this issue is closely related to the question of whether paradata may be used in substantive research. Example 3, in which item-level time stamp data are used as information about respondents, clearly shows that all ways in which released paradata possibly could be used should be considered before releasing any paradata. If with regard to this kind of paradata only the use as an indicator for standardised data collection (cf. example 1) would be considered prior to releasing keystrokes on a micro-level, the fact that sensitive information can be concluded from this kind of paradata may be overlooked.

Especially, when paradata include sensitive information about respondents, making them available for re-use subject to certain special data access and usage restrictions only, may provide a solution, which enables scientific research and at the same time ensures an appropriate level of data protection.[37] This also holds for other kinds of paradata that are classified as confidential for other reasons (such as proprietary considerations, etc.). In general, making confidential paradata available for re-use subject to special data access and usage restrictions appears to be possible in most of those cases in which releasing micro-level paradata to the scientific community or the entire public seems to be problematic. In all cases that have been explored so far, however, aggregated anonymised research results (e.g. from fieldwork monitoring) can be made available publicly.

7 References

Börsch-Supan, A. and Jürges, H. (Eds.) (2005): "The Survey of Health, Ageing and Retirement in Europe – Methodology." Mannheim Research Institute for the Economics of Aging (MEA), Mannheim.

Bristle, J., Celidoni, M., Dal Bianco, C. and Weber, G. (2014): "The Contribution of Paradata to Panel Cooperation in SHARE." SHARE Working Paper (19-2014). Munich Center for the Economics of Aging (MEA), Munich.

[36] A more detailed discussion of data protection measures is provided in chapter 5.2 of deliverable D6.2 of the DASISH project (Schmidutz and Bristle, 2013; available at: http://dasish.eu/deliverables/).

[37] It is noted that with respect to the release of sensitive information in a cross-country scenario, differences in the level of data protection between different EU member states have to be taken into account.

Bristle, J. and Halbherr, V. (2014): "Keystroke Analysis and Implications for Field Work". DASISH, Work Package 3, Deliverable D3.7. Retrieved December 15, 2014, from http://dasish.eu/deliverables/.

Couper, M.P. and Lyberg, L. (2005). "The Use of Paradata in Survey Research." Proceedings of the 55th Session of the International Statistical Institute.

Couper, M.P. and Singer, E. (2013). "Informed Consent for Web Paradata Use." Survey Research Methods 7(1), pp. 57-67.

Denscombe, M. (2002). "Ground Rules for Good Research: A 10 point guide for social researchers." Open University Press, Buckingham.

Groves, R.M. and Heeringa, S.G. (2006). "Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs." Journal of the Royal Statistical Society 169(3), pp. 439-457.

Heerwegh, D. (2002). "Describing response behavior in websurveys using client side paradata." Paper presented at the International Workshop on Websurveys, pp. 17-19 October 2002, Mannheim, Germany.

Kreuter, F. (2013). "Improving Surveys with Paradata: Analytic Uses of Process Information." Wiley, Hoboken, NJ.

Malter, F. and Börsch-Supan, A. (2013a): "SHARE Compliance Profiles – Wave 4." MEA, Max Planck Institute for Social Law and Social Policy, Munich. Retrieved December 15, 2014, from http://www.share-project.org/fileadmin/pdf_documentation/SHARE_Wave4_ComplianceProfiles_v15.pdf.

Malter, F. and Börsch-Supan, A. (Eds.) (2013b): "SHARE Wave 4: Innovations & Methodology." MEA, Max Planck Institute for Social Law and Social Policy, Munich.

Malter, F. (2013): "Fieldwork Management and Monitoring in SHARE Wave Four." In Malter, F. and Börsch-Supan, A. (Eds.): "SHARE Wave 4: Innovations & Methodology." MEA, Max Planck Institute for Social Law and Social Policy, Munich, pp. 124-139.

Mazzonna, F. and Peracchi, F. (2012): "Ageing, Cognitive Abilities and Retirement." European Economic Review, Elsevier, vol. 56(4), pp. 691-710.

Salthouse, T.A. (1985): "A Theory of Cognitive Ageing." North-Holland, Amsterdam.

Schmidutz, D. and Bristle, J. (2013): "Sample Merged Paradata Sets: Ethical and Legal Issues." DASISH, Work Package 6, Deliverable 6.2. Retrieved December 15, 2014, from http://dasish.eu/publications/projectreports/D6.2_Paradata.pdf.

Schmidutz, D., Ryan, L., Gjesdal, A. and De Smedt, K. (2013): "Report about New IPR Challenges: Identifying Ethics and Legal Challenges of SSH Research." DASISH, Work

Package 6, Deliverable 6.1. Retrieved December 15, 2014, from http://dasish.eu/publications/projectreports/D6.1_final.pdf.

Schröder, M. (Ed.) (2011): "Retrospective Data Collection in the Survey of Health, Ageing and Retirement in Europe. SHARELIFE methodology." Mannheim Research Institute for the Economics of Aging (MEA), Mannheim.

Singer, E. (2008). "Ethical Issues in Surveys." In: De Leeuw, Edith D. / Hox, Joop J. / Dillman, Don A. (Eds.). International Handbook of Survey Methodology. Psychology Press, Taylor & Francis, New York, pp. 78-96.

8 Annex: Acronyms and Abbreviations

CAPI – Computer-assisted personal interviewing

CATI – Computer-assisted telephone interviewing

DASISH – Data Service Infrastructure for the Social Sciences and Humanities

DoW – Description of Work, Annex 1 to the Grant Agreement of the DASISH project

EC – European Commission

ESS – European Social Survey

EU – European Union

ISCED – International Standard Classification of Education

RDC – Research Data Centre

SHARE – The Survey of Health, Ageing and Retirement in Europe

SMS – Sample management system

RDA – Remote Data Access

SSH – Social sciences and humanities

WP(#) – Work Package(Number)