**LEGAL AND ETHICAL ISSUES WHEN COLLECTING AND USING INFORMATION ON NON-RESPONDENTS**

Linn-Merethe Rød, NSD – Norwegian Centre for Research Data

**1 INTRODUCTION**

Over the last decade there has been a rapid growth and diversity of so called auxiliary data sources, ranging from commercial consumer data to official administrative records as well as information captured from the internet. The continuous development of new technologies, software and other resources makes linking of these new forms of external data to survey data both possible and more accurate. New opportunities to supplement survey data have naturally generated an increased demand for auxiliary data to be used in survey research. Judging from the literature, the most prominent advantages of auxiliary data can be made within nonresponse analyses, on the basis of the theoretical promise of auxiliary data to overcome the challenge of nonresponse (Krueger and West 2014). In fact, after decades of falling response rates, even in what can be considered as benchmark surveys, nonresponse is acknowledged as a considerable and increasing problem for most general population surveys (Brick and Williams 2013, Massey and Tourangeau 2013, Peytchev 2013). Additional information to adjust for nonresponse bias could for instance consist of neighbourhood contextual information such as population density, crime rates and poverty, geo location indicators with descriptions of geography including transport networks and local amenities such as the proximity to schools, libraries and hospitals, behavioural or attitudinal data captured from social media and administrative data on tax collection, pension or welfare benefit computations (Comber et.al 2011, Biemer and Peytchev 2013). However, the relentless expansion of new technologies into all aspects of our lives has not only increased the volume of personal data potentially available to researchers, but also introduced new types of personal information that must be protected (ESOMAR 2016).

The main issues requiring attention to ensure ethical research practice, center around the commitment to protect the rights of free will, privacy, confidentiality, and wellbeing of research participants. These ethical rights date back to the Helsinki Declaration (1948) and the Belmont Report (1979) (Singer 2008). Following the line of Helsinki and Belmont, the two key ethical principles that stand out as particularly relevant for survey researchers are to protect the confidentiality of the data collected and to ensure that the respondents are treated as autonomous individuals with the right to make informed voluntary decisions about participation (Couper and Singer 2013). Furthermore, researchers continuously need to protect their reputation through practices that ensure transparency for respondents, maintain confidence in provided information, and demonstrate consideration for research participants (ESOMAR 2016). Building of trust is essential for keeping individuals engaged in research. However, the principles of confidentiality and consent grounded on transparent, easily understandable information are getting more complex as the collection of data within survey research expands and the volume of what can be considered as personal data increases. In the following, legal and ethical issues relating to the processing of personal data on non-respondents will be explored in relation to ethical guidelines, current legislation and new framework conditions in Europe.

**2 HARMONISATION OF LEGAL DATA PROTECTION FRAMEWORKS**

Ethical obligations and guidelines are embodied in government laws and regulations that to some extent vary across national borders (Singer 2008). In Europe, the processing of personal data is currently regulated by national implementations of the Data Protection Directive 95/46/EC. Consequently, researchers working in a global context face a patchwork of national data protection laws (ESOMAR 2016). The need for a harmonised legal framework in Europe was thus one of the reasons why the European Commission in January 2012 proposed a comprehensive reform of the current data protection rules. The promise of the new General Data Protection Regulation (hereafter referred to as GDPR or Regulation) is to harmonise the current data protection rules, as well as to ensure a consistent and high level of personal data protection to provide legal certainty and trust (European Commission 2015).

Compared to the Directive, the GDPR imposes stricter obligations with regard to both consent and data security (Maldoff 2016). Unlike the Directive, the Regulation also promotes techniques such as anonymisation (removing personally identifiable information where it is not needed), pseudonymisation (replacing personally identifiable material with artificial identifiers), and encryption (encoding messages so only those authorized can read it) to protect personal data (Article 30). Furthermore, the GDPR introduces the concept of “data protection by design”. At the conceptual level, data protection by design means that privacy should be a feature of the development of a product, rather than something that is tacked on later (Maldoff 2016). Thus, Article 23 requires controllers to implement appropriate safeguards “both at the time of the determination of the means for processing and at the time of the processing itself.” The GDPR will be directly applicable to all EU and EEA countries, and shall apply from 25 May 2018.

**2.1 WHAT CONSTITUTES PERSONAL DATA?**

One of the essential elements of a harmonised legislation on privacy protection is a uniform understanding of the term personal data. The current Data Protection Directive defines personal data as “any information relating to an identified or identifiable natural person” meaning someone “who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to physical, physiological, mental, economic, cultural or social identity” (Article 2(a)). This definition has been implemented across Europe with various degrees of strictness. For instance the UK legislative has practiced a liberal interpretation, with personal data defined as personal only if the investigator can identify the data subject (Beyleveld et.al 2004:408). Likewise, the Dutch Data Protection Authority (Autoriteit Persoonsgegenes) states that data should not be considered as personal if actual identification is reasonably excluded because of effective measures such as encryption of the data and/or agreements about the access to the data. At the other end of the scale, in countries like Estonia and Norway, any data concerning an identified person is defined as personal data, regardless of the form or format in which such data exists. Hence, a lot more individual level data has been falling outside the scope of the privacy regulation in the UK and Netherlands than in Norway and Estonia (Bøe et.al 2014).

The definition of personal data in the new EU Regulation is similar to the existing one in the Directive, with the difference that location data, online identifiers and genetic factors have been included. Furthermore, the Regulation continues the account taken in the Directive of means reasonably likely to be used, when determining whether a person is identifiable. The Regulation states that account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments. Personal data which have been pseudonymised is explicitly stated to be considered as information on identifiable persons and therefore subject to the Regulation (cf. recital 26). Thus, even if a researcher or controller no longer have the ability to re-identify a data set, such data set may still be regulated under the GDPR if it could be re-identified with reasonable effort, by another person (Maldoff 2016). Furthermore, data is considered anonymous only when it does not relate to an identified or identifiable natural person (by any means likely to be used) (cf. GDPR, recital 26). To achieve a harmonised interpretation of the definition of personal data across Europe, the GDPR will probably lower the threshold for considering information as directly or indirectly identifiable, resulting in more research projects falling within the scope of the law, at least in countries which have a liberal definition today (Bøe et.al 2014, Kvamme et.al 2014).

**3 RESEARCH AS A LEGITIMATE BASIS FOR PROCESSING?**

In survey research, the respondent’s informed consent usually constitutes the ethical and legal basis for the processing of personal data. In the new EU Regulation, consent is similar to the current Directive required to be “freely given, specific, informed, and unambiguous” (GDPR, Article 4(8)). Furthermore, in line with the principle of transparency, the GDPR requires that any information and communication between researcher and the data subject should be easily accessible and easy to understand, by using a clear and plain language (cf. recital 58).

Consent to collect data on non-respondents is however typically not feasible to obtain. Thus, nonresponse analyses which involve processing of personal data at any stage depend on alternatives to consent-based processing for research purposes. Article 7 of the current EU Directive states that personal data may be processed only if the data subject has unambiguously given his consent or the processing is necessary for the public interest. Accordingly, the recitals (cf. 34) allow processing of sensitive categories of personal data without consent if important reasons of public interest justify it, as scientific research. The dimension of necessity and public interest as an alternative to consent implies a balancing of interests between the survey’s societal value and possible inconveniences for the data subjects. When it comes to non-respondents, probable disadvantages include the underlying premise that they are being studied despite their refusal to take part in the survey and presumably their expectance that the researcher does not include them at all. Furthermore, relevant factors of disadvantages are for instance whether data is considered to be sensitive, which data source is used (e.g., administrative data subject to professional secrecy or publicly available data), the degree of identifiability, implemented safeguards and who have access to identifiable data, for how long a period. When it comes to publicly available personal data, which are “manifestly made public by the data subject”, it is worth mentioning that the collection and use of such information is exempted from consent both in the old EU Directive and the new GDPR (cf. GDPR, Article 9(2)(e)). However, ethical considerations around the individuals’ autonomy and perception of privacy, still could request that information should be provided to survey participants on collection and use of e.g. their tweets, blogs or utterances on Facebook (The Norwegian National Committees for Research Ethics 2015, ESOMAR 2015).

Currently across Europe, the procedures regarding how, or whether, to permit non-consent based processing of (sensitive) personal data varies. Some countries require a special license for the processing, or do not allow the processing at all without consent, while others could consider the same collection of data to fall outside the scope of the law, dependent of the national definition of personal data (Bøe et.al 2014). In addition, the basic provisions of the current Directive concerning alternatives to consent are implemented in various ways, depending on whether consent is seen as the main rule for processing of personal data, or appears as one of a number of equal alternatives. For researchers conducting cross-national survey research, it is thus a challenging task to review and comply with the national and international data protection regulations of each country where they plan to collect or process data. Furthermore, they risk that in some countries, certain non-consent based projects will be impossible to carry out. These challenges should hopefully prove a lot easier in the future, as one of the promises of the new GDPR is to harmonise both legislation and practice, to improve the possibilities for data access, data sharing and in general cross-national research.

The Regulation has luckily turned out to be fairly research-friendly. The most important special provisions from the Directive are continued, clarified and strengthened. Amongst other this applies to the possibility to process personal data based on other grounds than consent. The Regulation states e.g. that special categories of personal data can be processed if necessary for research purposes (cf. Article 9(2)(j)), and that additional processing for research purposes generally will be considered compatible with the original purpose, and, therefore, exempt from specific consent (Article 5(1)(b)). However, these special exemptions for research purposes require appropriate safeguards to be implemented. In particular, technical and organisational measures must be in place in order to ensure that only the personal data necessary for the research purposes are being processed, in accordance with the principle of data minimisation. This involves that the personal data should be adequate, relevant and limited to what is necessary to the purposes for which they are processed (cf. Article 5(c)). Data minimisation measures may include pseudonymisation[[1]](#footnote-1), provided that the research purposes can be fulfilled in that manner (cf. Article 89(1)). The use of pseudonymisation techniques are thus not always required, but to a large extent encouraged, as a best practice when personal data are processed (Maldoff 2016). Hence, within survey research, one should make sure that a master file linking (non-)respondents’ names, addresses or phone numbers with their corresponding internally-generated ID numbers is kept secure with access limited to a small number, e.g. sampling or panel management staff (ESOMAR 2016). A short as possible retention period should apply to personal data about non-respondents. However, it is important not to delete the personal data before all quality checks regarding accuracy and data integrity privacy principle are performed (ibid).

**3.1 NOTICE REQUIREMENT**

Although controllers are not required to obtain the data subject’s consent for all processing for research purposes, they do however remain bound by the GDPR’s notice requirement. Article 12(1) states that “appropriate measures” must be taken to inform data subjects of the processing of personal data and the rights available to them. The information should be provided in writing, or by other means, where appropriately in electronic form. Controllers are required to provide this information in all circumstances, regardless of whether consent is the basis for processing, “in a concise, transparent, intelligible and easily accessible form, using clear and plain language”. Based on this provision, it would appear that the inclusion of information in the invitation letter on the use of various auxiliary data (on both non-respondents and respondents) in survey research to a large extent will become an explicit requirement.

The notice should include the controller’s identity and contact information, the intended purposes of the processing of personal data, and the data subject’s rights to access, rectification, erasures and object to processing, as well as the “period for which the personal data will be stored”, or “criteria used to determine that period” (cf. Article 13(2)). For researchers to fulfil this obligation, it will be decisive to find ways to inform the data subjects about the collection and use of auxiliary data without impeding the research. Ideally, the collection of auxiliary data to adjust for nonresponse bias should not result in a higher rate of non-respondents.

However, a researcher may claim exemption from the notice requirement if providing notice would be “likely to render impossible or seriously impair the achievement of the [research] objectives”, provided that appropriate safeguards are in place, “including making the information publicly available” (Article 14(5)(b)). On the same conditions, research could be exempted from the data subject’s rights to erasure of her/his personal data (Article 17(3)(d)), and a researcher could also override a data subject’s objection to processing if justified by public interest with basis in Union or Member State law (Article 21(6)). Otherwise, recital 33 states that controllers should act “in keeping with recognised ethical standards for scientific research”.

**4 DISCUSSION**

The new EU Regulation continues to give research a unique position in its general provisions. As long as appropriate safeguards are implemented, researchers may be exempted from quite a range of obligations and restrictions.

The Regulation however enhances the focus on completely transparency towards data subjects about intended collection and use of their personal data, regardless of whether the processing is based on consent. How will this potentially stricter notice requirement affect researchers’ collection of auxiliary data on identifiable non-respondents? How detailed information will be needed to ensure an adequately understanding of why these data are being collected? Will researchers most likely try to avoid giving information on this kind of data collection? Will providing information have negative effect on response rates? Furthermore, could researchers risk that non-respondents will object to such processing or demand erasure of their personal data? Should best practise, ethically and legally, be to exempt nonresponse analyses from the notice requirement provided that appropriate safeguards are implemented? Or, should best practice be to provide information, including or excluding, the rights for the data subject to objection and erasures?

**REFERENCES**

Autoriteit Persoonsgegenes (n.d). *Artikel 1 sub a Wbp*: *Niet identificeerbaar*. Retrieved from https://autoriteitpersoonsgegevens.nl/nl/over-privacy/wetten/wbp-naslag/hoofdstuk-1-algemene-bepalingen-art-1-tm-5/artikel-1-sub-wbp

Beyleveld, D., Grubb, A., Townend, D., Morgan, R. & Wright, J. (2004). The UK’s Implementation of Directive 95/46/EC. In D. Beyleveld, D. Townend, S. Rouillé-Mirza & J. Wright (Eds.), Implementation of the Data Protection Directive in Relation to Medical Research in Europe (pp. 403-428) Aldershot: Ashgate.

Biemer, P., & Peytchev, A. (2013). Using geocoded census data for nonresponse bias correction: An assessment. Journal of Survey Statistics and Methodology, 1, 24-44.

Brick, J. M., & Williams, D. (2013). Explaining Rising Nonresponse Rates in Cross-Sectional Surveys. ANNALS of the American Academy of Political and Social Science, 645, 36–59.

Bøe, M., Rød, L., Parra, C., De Smedt, K., Kvalheim, V., Segadal, K., … Samdal, G. (2014). Handbook on legal and ethical issues for SSH data in Europe, Part II. DASISH Work Package 6: Legal and Ethical Issues. Deliverable 6.5.

Comber, A., Brunsdon, C., & Radburn, R. (2011). A spatial analysis of variations in health access: linking geography, socio-economic status and access perceptions. International Journal of Health Geographics,10:44. Retrieved from http://www.ij-healthgeographics.com/content/10/1/44

Couper, M. P., & Singer, E. (2013). Informed Consent for Web paradata Use. Survey Research Methods 7 (1), 57-67.

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 (1995). Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML

European Commission (2015). Agreement on Commission's EU data protection reform will boost Digital Single Market. Retrieved from http://europa.eu/rapid/press-release\_IP-15-6321\_en.htm

ESOMAR World Research (2015). ESOMAR Online Research Guidelines. Retrieved from https://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ESOMAR-GRBN-Online-Research-Guideline-October-2015.pdf

ESOMAR World Research (2016). ESOMAR Data Protection Checklist. Retrieved from https://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ESOMAR-Data-Protection-Checklist\_update-April-2016.pdf

Krueger, B. S., & West, B. T. (2014). Assesing the Potential of Paradata and Other Auxiliary Data from Nonresponse Adjustments. Public Opinion Quarterly 78, 795–831.

Kvamme, T., Parra, C., De Smedt, K., Segadal, K., Kvalheim, V. (2014). Report about Preservation Policy-Rules (Preservation Challenges). DASISH Work Package 6: Legal and Ethical Issues. Deliverable 6.6.

Maldoff, G. (2016). Top 10 operational impacts of the GDPR: Part 3 – consent. Retrieved from https://iapp.org/news/a/top-10-operational-impacts-of-the-gdpr-part-3-consent/

Maldoff, G. (2016). Top 10 operational impacts of the GDPR: Part 8 – Pseudonymization. Retrieved from https://iapp.org/news/a/top-10-operational-impacts-of-the-gdpr-part-8-pseudonymization

Maldoff, G. (2016). How GDPR changes the rules for research. Retrieved from https://iapp.org/news/a/how-gdpr-changes-the-rules-for-research/

Massey, D.S., & Tourangeau, R. (2013) The nonresponse challenge to surveys and statistics. Annals of the American Academy of Political and Social Science, 645, 6-27.

Peytchev, A. (2013). Consequences of survey nonresponse. Annals of the American Academy of Political and Social Science, 645, 88-111.

Proposal for a regulation of the European parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (2012). Retrieved from http://ec.europa.eu/justice/data-protection/document/review2012/com\_2012\_11\_en.pdf

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (2016). Retrieved from http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L\_.2016.119.01.0001.01.ENG&toc=OJ:L:2016:119:TOC

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

The Norwegian National Committees for Research Ethics (2015). Ethical Guidelines for Internet Research: 1. Public or private? Retrieved from https://www.etikkom.no/en/ethical-guidelines-for-research/ethical-guidelines-for-internet-research/1.-public-or-private/

1. Pseudonymisation is defined as «the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution to an identified or identifiable individual” (Article 4(3)(b)). [↑](#footnote-ref-1)