(Un)informed Consent: To What Degree are Research Participants 'Informed' by Common Consent Procedures in Psychology under EU Data Protection Law?

Malte Elson[ab], Dara Hallinan[c], Annika Külpmann[a], Franziska Boehm[c]

[a] Faculty of Psychology, Ruhr University Bochum
[b] Horst Görtz Institute for IT Security, Ruhr University Bochum
[c] FIZ Karlsruhe — Leibniz Institute for Information Infrastructure

ABSTRACT

There is reason to believe that consent forms may routinely do not fulfill the requirements for consent outlined in EU data protection law. Where this is the case, the legitimacy of the conduct of research may be undermined and could result in restrictions on the subsequent conduct of research, obligations to delete data, or obligations to limit the sharing of psychological research data. However, so far, there are no empirical data to support the proposition that compliance may not be the norm. We propose a study design in which we draw a random sample of psychological research reports and systematically compare the research practices (i.e., reported data collection procedures, sharing practices) with the details provided in the respective participant information and consent form and compare each of these with the legal requirements outlined in EU data protection law.

INTRODUCTION

Informational self-determination – the authority of individuals to decide when, and to which degree, data about themselves can and should be communicated with others – has become an increasingly important issue in the public sphere. In the information age, with rapid advances in computer technology, questions such as to who collects data and why, and how effective tools for individuals to (re)gain control over the processing of their personal data can be designed, have received a lot of attention.

Issues of information self-determination are salient for data processing by governments and large corporations, but naturally they also extend to academic research in which data is collected from human subjects. Scientists are held in high regard by the general public (Krause et al., 2019), and this trust naturally extends to the handling of personal data. If research participants are asked for consent to legitimate the processing of their personal data in research, they can, and should, expect that they are fully informed by scientists as to what personal data are collected, for which purposes, and who is given access to the data.

This is becoming more and more relevant as, for example, there has been a growing demand in the social and behavioral sciences to store data sets in public repositories (such as the Open

Science Framework[1], Zenodo[2], or PsychArchives[3]). Journals increasingly adopt data sharing policies that encourage researchers to share their data, and even make sharing their default, requiring authors to upload datasets when manuscripts are published or submitted, or explain in a written statement why it is not possible to do so (see, for example, the TOP Guidelines[4]). Recent developments, such as the partnership between the American Psychological Association (APA) and the Center for Open Science[5], push transparency beyond previous recommendations to share data upon request by making data sharing part of the research workflow, and more broadly, of *normal* science. The rationale and the benefits of increased sharing and transparency of data – including of open data – has been discussed at length (Wilkinson et al., 2016), and we, too, support these developments.

Research participants are often granted the right to decide whether their personal data are used in research, and, in this regard, generally have the right to be fully informed about which of their personal data are being collected and why. Study participants must be able to trust psychologists that they have been adequately informed about the implications of personal data collection and use in a consent transaction – including, for example, as to the potential for broad sharing of data in public repositories, as they may be used for a wide range of purposes that likely are beyond the original intent of the researcher who collects the. In essence, participants' rights to informational self-determination must be respected.

EU Data Protection Law and the General Data Protection Regulation (GDPR)

EU Data Protection law, organized under the General Data Protection Regulation (GDPR) elaborates the principles of data protection applicable to the processing (i.e., collecting, storing, analyzing, or transferring) of personal data. Personal data, as defined in GDPR Art. 4(1), are any information relating to an identifiable person – including physiological, cultural, or social data – all of which are routinely collected in psychological research. EU Data Protection law is applicable in the European Union (EU) and the European Economic Area (EEA), as well as to the transfer of personal data outside the EU and EEA.

In principle, the GDPR supersedes national laws that regulate the processing of personal data. In a number of instances, however, the GDPR still permits specifications and derogations to its provisions to be made by Member States in their national laws. The goal of EU data protection law is to establish principles concerning who will have control over whether the processing of personal data can happen, and as to how processing should subsequently happen – including the elaboration of principles giving individuals specific control rights over the processing of personal data, and principles elaborating a range of data subject rights and data controller obligations applicable regardless of who has control over processing.

One of the principles of EU data protection law is that personal data may only be collected and processed if there is a lawful basis. In many instances of social science research, the consent of the research subject will itself provide the legal basis. Even when it does not, and

[1] https://www.osf.io
[2] https://www.zenodo.org
[3] https://www.psycharchives.org
[4] https://www.cos.io/initiatives/top-guidelines
[5] https://www.apa.org/news/press/releases/2017/08/open-science

consent is nevertheless obtained as some form of supplemental ethical requirement, EU data protection law will likely still elaborate obligations relevant to the content of consent – for example, the range of types of information to be provided to the research subject.

Informed Consent under EU Data Protection Law

For consent to provide a valid, lawful basis under EU data protection law, it must be given in accordance with certain conditions outlined in the law – e.g., consent must be freely given, specific, informed, and unambiguous. In relation to the 'specific' criterion, for example, consent is, in principle, required to be given only in relation to limited, specifically defined and explicated purposes of research that justifies the data processing (although broader forms of consent are possible in some circumstances).

As in scientific research the exact purpose is not always clear prior to data collection, and other purposes may later arise, there are some exemptions for research and science regarding the reuse of data. For a further discussion of the GDPR in a research context, see e.g. (Hallinan, 2020, 2021; Mondschein & Monda, 2019).

The fact that the conditions of consent under EU data protection law will usually be – in some way – applicable to processing in psychological research done on the basis of consent, does not necessarily mean that all consent collected in the psychological research context necessarily complies with these conditions. Indeed, we see several reasons to doubt compliance is the norm.

EU Data Protection Law, Consent and Processing for Psychological Research

The person responsible for the research will usually also be responsible for making sure the research is compliant with the relevant conditions of EU data protection law. Research psychologists may not have the legal training required to formulate a comprehensive consent form which fulfills all relevant criteria.

In turn, the context in which researchers must look for answers is complex and confusing. Whilst the EU law (Regulation No 536/2014. Clinical trials on medicinal products for human use, European Parliament, Council of the European Union) and most EU Member State national laws mandate an ethics vote just for medical research, this is not the case for social science. Consequently, a patchwork of practices with regards to consent forms and participant debriefing has emerged – as there are no overarching legal authorities governing or advising researchers as to how to formulate a consent which fulfills all relevant requirements. Although there are attempts to offer guidelines and templates (e.g., by the German Psychological Society, 2021), these are not used throughout the entire research community, leading to substantial national, regional, or even local differences. Even such diligently drafted templates can merely offer a general structure, likely leaving a lot of specific information to be filled in for each new study.

Unfortunately, these circumstances have also, arguably, led to research practices that may neither be in the best interests of study participants nor of science (Meyer, 2018). For example, consent forms are often unnecessarily complex and not written in plain language,

which may overwhelm research participants and may even been argued to render their consent essentially uninformed (Foe & Larson, 2016). At the same time, descriptions of measures relating to the preservation of anonymity need to be chosen carefully as they may inadvertently and negatively affect participants' agreement to data sharing protocols (McGuire et al., 2011). Similarly, opportunities to make data publicly available are also often limited by overly and unnecessarily restrictive anonymity promises (van den Eynden, 2008).

Taken together, the above is reason to believe that consent forms may routinely fail to fulfill the requirements for consent outlined in EU data protection law. Where this is the case, the legitimacy of the conduct of research may be undermined and could result in restrictions on the subsequent conduct of research, obligations to delete data, or obligations to limit the sharing of psychological research data – a practice, as discussed above, that has seen growing support over recent years.

However, so far, there are no empirical data to support the proposition that compliance may not be the norm. We propose a study design in which we draw a random sample of psychological research reports and systematically compare the research practices (i.e., reported data collection procedures, sharing practices) with the details provided in the respective participant information and consent form and compare each of these with the legal requirements outlined in EU data protection law.

METHOD

Sample

Our sampling procedure follows three steps. First, we use a broad database search based on a few limiters to define a *population* of papers to sample from. Second, we draw random *screening samples* from this population, further excluding papers not matching our research question. Third, we contact the authors of those candidate papers and ask for participant information sheets and consent forms. Only those papers of which we end up with a full set (i.e. published paper, consent form, and any further participant information material, where applicable) enter our *final sample* that we examine in detail. The codebooks for the screening samples, the final sample, and the consent forms can be found in the PsychArchives project folder.

Population

To define a population of research papers, we retrieve a list of entries in *PsycInfo*[6], a comprehensive database of the psychological literature maintained by the APA, via *EbscoHost*[7], a database provider. As consent-based data processing should be part of every corner of psychology, there are only few restrictions with regards to research topics or methodology for practical and logistical reasons.

[6] https://www.apa.org/pubs/databases/psycinfo
[7] http://search.ebscohost.com/

SEARCH LIMITER	VALUE
Publication type	Peer-Reviewed Journal
Language (text)	English
Published date	06/2019 -
Methodology	Empirical study
Population Group	Human
Age	Adulthood (18 yrs & older)

Note that the earliest publication date of 06/2019 was chosen to allow a publication lag of about 1 year since GDPR has been implemented (May 25, 2018). While it is still likely that some data published after 06/2019 were collected even before 06/2018, the GDPR has constituted a point of reference for data protection in Europe since May 24, 2016. The publication lag should remove any ambiguity with regards to delays in the publication of manuscripts due to administrative reasons on the publisher's side. We also restrict the population to adult samples as studies with minors are subject to different regulations and usually require consent by multiple parties.

As the GDPR only applies to data processing in the EU and EEA, we exclude any study conducted by a data controller (likely the researcher) located and doing the research outside the EEA. This area is, regrettably, further restricted to those 9 EEA countries whose official language(s) any one of us is familiar with, as we need to be able to understand the consent forms and study materials. This includes Austria, Belgium, France, Ireland, Liechtenstein, Luxembourg, the Netherlands, the United Kingdom[8], and Sweden, which make up approximately 52% of the EEA population and approximately 65% of the EEA's total publishing activity in psychology (Scimago Lab, 2021). To this end, we restrict the population to research whose location (information provided by PsycInfo) or author affiliation matches any of those countries by using the following search term:

```
((PL (Austria OR Belgium OR Germany OR France OR Ireland OR
Liechtenstein OR Luxembourg OR Netherlands OR Sweden OR United Kingdom
OR UK) OR (AF (Austria OR Belgium OR Germany OR France OR Ireland OR
Liechtenstein OR Luxembourg OR Netherlands OR Sweden OR United Kingdom
OR UK OR Wales OR England OR Scotland))
```

Screening Sample

From this population, we randomly draw waves of $k = 100$ papers for an eligibility screening (see R script). For each paper, we manually check whether it was correctly included given our search limiters above. We then apply further restrictions to identify candidates for our final sample:

Personal data. We exclude any study for which no personal data, as defined in GDPR Art. 4(1), were processed.

[8] We appreciate that the UK has left the EU. However, in the period covered by the study the GDPR has either applied in the UK or the UK's national data protection law has applied, which largely constitutes an incorporation of the GDPR into national law: https://ico.org.uk/for-organisations/dp-at-the-end-of-the-transition-period/transition-period-faqs/#gdpr accessed 4 May 2021.

Research location. We examine a candidate study's methodology description for details where the study took place. If this returns ambiguous or no information, we use the affiliation(s) of the first author as an indicator of the location where data collection took place. Either way, the location is verified by correspondence with the authors (see below).

Completion date. We exclude any study for which data collection was completed before May 24th, 2016. Prior to this date, the GDPR was not a valid law. To this end, those studies are excluded for which the data collection is explicitly reported to have ended before May 24th, 2016, or where there are obvious reasons that this must have been the case (based on the context of the study). Studies for which the available information is ambiguous or clearly indicates data (or parts of the data) were collected on or after May 24th, 2016, are further considered for the final sample (see below).

Secondary data. Studies using secondary data are excluded, as the authors of these studies may not have the original consent forms or participant information.

Limited information. We exclude any study for which there was an obvious and legitimate reason that certain sorts of information would not have been provided to research subjects or consent could not have been obtained.

Multi-study papers. No more than one study per paper is considered. We always select the first study as enumerated in the paper to which the above criteria apply.

Final Sample

Using the contact information of corresponding authors of eligible papers, we ask for two things:

Confirmation of eligibility. We ask the corresponding author to verify that personal data (or parts of the personal data) were collected in the EEA, and that data collection was not completed before May 24th, 2016. If this was not the case or the response is ambiguous, the paper is no longer eligible for the final sample.

Consent forms and participant information. We ask researchers to send us the consent form, debriefing text, or any other information given to participants throughout the study that characterize the purpose of the research and the nature of the data processing.

The template for this first email can be found in the PsychArchives project folder. Note that the template is in English, but that we may send a translated version. We record the date of when we first contacted the corresponding author. If we do not hear back within two weeks, we send a reminder, also CCing the first or the senior author, respectively. Studies are no longer eligible for the final sample if a) we do not receive a response at all to our request within three weeks, b) we do not receive a sufficient response within six weeks (e.g., promises to send the materials, but no follow-up), c) the request is declined. Only those studies for which we obtain a full set (i.e., the paper, a consent form, and any other information material, if applicable) enter the final sample. Therefore, although the screening samples are drawn randomly from the population, the final sample is not random as author response determines eligibility.

The procedure of randomly drawing and screening 100 papers from the population and contacting corresponding authors of eligible papers to obtain the consent forms is repeated until there are at least $K = 100$ full sets in the final sample.

Coding

For each paper in the final sample, the coding is done in two consecutive steps. The psychologists (ME, AK) code the contents of the paper, i.e. the research itself, whereas the legal scholars (DH, FB) code the contents of the consent form and auxiliary participant information. Their synthesis is done jointly by both teams.

Research Papers

The coding of the papers includes the following information:

- Bibliographic information
- Sample size and basic sample information
- All processed data and their methodology, including
    - Demographic data
    - Self-report
    - Other-report
    - Observational data
    - Assessments or standardized tests
    - Audio data
    - Video data
    - Reaction time data
    - Physiological data
    - Brain imaging
    - Other data not covered by these categories
- Processing of sensitive data as defined in the GDPR (this includes racial or ethic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data, health data, sexual orientation, or sexual life data)
- Mode of data collection (paper-pencil, local computer, etc.)
- Data storage (local computer, cloud service, data repository, etc.)
- Data access (open access, restricted, etc.)

Consent forms and participant information

The coding of the consent form matches the coding of the paper to allow assessing whether the consent covered the actual research. The coding includes further:

- Identity and contact details of the controller and the data protection officer
- Definition of legal basis (consent, legitimate interest, public interest, etc.)
- Purpose and details of processing data
- Aims of the study

- Recipients of personal data
- Plans to transfer personal data outside the EU
- Period of data storage and period personal data remain identifiable
- Rights retained by research participants

ANALYSIS

Briefly, the goal of the analysis is to estimate the degree to which the consent forms signed by research subjects fulfil the requirements set out in EU data protection law. As such, the analyses consist of parts that are descriptive statistics (e.g., how commonly required aspects of consent forms are missing), and parts that are largely qualitative (e.g., how adequately data processing for psychological measurement is explained to research participants). No inferential significance tests are planned.

Debriefing

All corresponding authors of papers in our final sample are informed about the coding of their paper/consent forms and, if necessary and feasible, pointed towards possibilities for improvement.

REFERENCES

Regulation No 536/2014. Clinical trials on medicinal products for human use, European Parliament, Council of the European Union. https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-1/reg_2014_536/reg_2014_536_en.pdf

Foe, G., & Larson, E. L. (2016). Reading level and comprehension of research consent forms: An integrative review. *Journal of Empirical Research on Human Research Ethics*, *11*(1), 31–46. https://doi.org/10.1177/1556264616637483

German Psychological Society. (2021). *Ethics committee*. https://zwpd.transmit.de/zwpd-dienstleistungen/zwpd-ethikkommission

Hallinan, D. (2020). Broad consent under the GDPR: An optimistic perspective on a bright future. *Life Sciences, Society and Policy*, *16*(1). https://doi.org/10.1186/s40504-019-0096-3

Hallinan, D. (2021). *Protecting genetic privacy in biobanking through data protection law*. Oxford University Press.

Krause, N. M., Brossard, D., Scheufele, D. A., Xenos, M. A., & Franke, K. (2019). Americans' trust in science and scientists. *Public Opinion Quarterly*, *83*(4), Article nfz041, 817–836. https://doi.org/10.1093/poq/nfz041

McGuire, A. L., Oliver, J. M., Slashinski, M. J., Graves, J. L., Wang, T., Kelly, P. A., Fisher, W., Lau, C. C., Goss, J., Okcu, M., Treadwell-Deering, D., Goldman, A. M., Noebels, J. L., & Hilsenbeck, S. G. (2011). To share or not to share: A randomized trial of consent for data sharing in genome research. *Genetics in Medicine*, *13*(11), 948–955. https://doi.org/10.1097/GIM.0b013e3182227589

Meyer, M. N. (2018). Practical tips for ethical data sharing. *Advances in Methods and Practices in Psychological Science*, *1*(1), 131–144. https://doi.org/10.1177/2515245917747656

Mondschein, C. F., & Monda, C. (2019). The EU's General Data Protection Regulation (GDPR) in a research context. In P. Kubben, M. Dumontier, & A. Dekker (Eds.), *Fundamentals of Clinical Data Science* (pp. 55–71). https://doi.org/10.1007/978-3-319-99713-1_5

Scimago Lab. (2021). *Scimago Journal & Country Rank 2019*. https://www.scimagojr.com/countryrank.php?area=3200&year=2019

van den Eynden, V. (2008). Sharing research data and confidentiality: Restrictions caused by deficient consent forms. *Research Ethics*, *4*(1), 37–38. https://doi.org/10.1177/174701610800400111

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18