SURVEYS · ETHICS AND LAW

Ethical Considerations in Tracking and Geolocation Data Linked to Surveys

Location traces and app tracking can enrich survey answers, but they are among the most revealing data a researcher can hold. This article sets out the ethical issues and the safeguards that make such linkage defensible.

ModQu research article · 1,793 words · 18 references (APA 7th edition)

Smartphones, wearable devices, browser plug-ins and connected cars record where people go and what they do. Survey researchers increasingly ask respondents for permission to collect these data alongside their answers. A travel survey can be paired with GPS traces; a media survey with a record of websites visited; a health survey with step counts and the places where respondents spend their time (Harari et al., 2016; Struminskaya et al., 2020). Linking tracking data to survey answers reduces reliance on memory and adds detail that no questionnaire can capture.

The same features that make these data useful make them ethically demanding. Location traces are continuous, detailed and personal. They can reveal a person's home, workplace, religious practice, medical visits and relationships, often without the person realising how much has been recorded. This article examines the ethical issues raised by linking tracking and geolocation data to surveys and the safeguards that responsible researchers apply.

Why location data are different

Most survey answers are discrete pieces of information that the respondent chooses to give. Tracking data are different in four ways.

They are passive. Once collection begins, data accumulate without any further decision by the respondent. A person consents once but may be recorded thousands of times.

They are hard to anonymise. In a study of mobile phone records, de Montjoye et al. (2013) analysed data for 1.5 million people and found that four points in space and time were enough to identify 95 per cent of individuals uniquely. Removing names does not make location traces anonymous. Sweeney (2002) had earlier shown that combinations of ordinary attributes, such as postal code, birth date and sex, can identify individuals, and location traces offer far more such combinations.

They reveal more than was asked. A study of commuting may record visits to a clinic, a place of worship or a political meeting. The data contain sensitive information that the researcher did not intend to collect.

They affect other people. A respondent's trace shows where they live and whom they visit, so it can reveal information about household members and others who never agreed to take part.

Nissenbaum (2010) argues that privacy is best understood as **contextual integrity**: information flows are appropriate when they match the norms of the context in which the information was shared. A respondent who shares location with a mapping app expects it to be used for navigation, not for a market research report. Linking that information to survey answers creates a new flow, which needs its own justification and consent.

Consent

The principle of respect for persons requires that participants enter research voluntarily and with adequate information (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979). For tracking data, adequate information is hard to provide. Respondents need to understand what will be collected, how often, for how long, who will see it, how it will be linked to their answers and how it will be protected.

Evidence shows that respondents take these questions seriously. Keusch et al. (2019) found that willingness to take part in passive mobile data collection varied with the type of data requested and was lower among people with greater privacy concerns. Revilla et al. (2019) found that members of an online panel were much less willing to share their location than to carry out other additional tasks, such as taking photographs. Kreuter et al. (2020) reported on an app that collected survey and smartphone sensor data in Germany and found that privacy concerns and trust in the research organisation shaped who agreed to take part.

Good consent practice for tracking data includes:

- **Layered information.** A short summary of the key points, with fuller details available for those who want them.
- **Granular choices.** Separate permissions for location, app use and other sensors, so that respondents can agree to some and not others.
- **Ongoing control.** The ability to pause collection, view what has been collected, delete records and withdraw at any time.
- **Clear purposes.** A statement of what the data will be used for and a commitment not to use them for unrelated purposes.

Couper and Singer (2013) showed that even for less sensitive paradata, the way a consent request is worded affects agreement. Researchers should test consent materials as carefully as questionnaires, and should report consent rates so that readers can judge possible bias.

Legal requirements

In many jurisdictions, location data linked to an individual are personal data. Under the European Union's General Data Protection Regulation, researchers need a lawful basis for processing, must observe principles such as purpose limitation and data minimisation, and must carry out a data protection impact assessment for processing likely to result in high risk to individuals (European Parliament & Council of the European Union, 2016). Systematic monitoring and tracking of location are typical examples of high-risk processing. Similar principles apply under privacy laws elsewhere, including Canada's federal and provincial laws. Legal compliance is necessary but not sufficient: an activity can be legal and still breach respondents' reasonable expectations.

Protecting confidentiality

Because location traces are so identifying, confidentiality protection must be built into every stage, from collection to publication. Figure 1 shows the main safeguards.

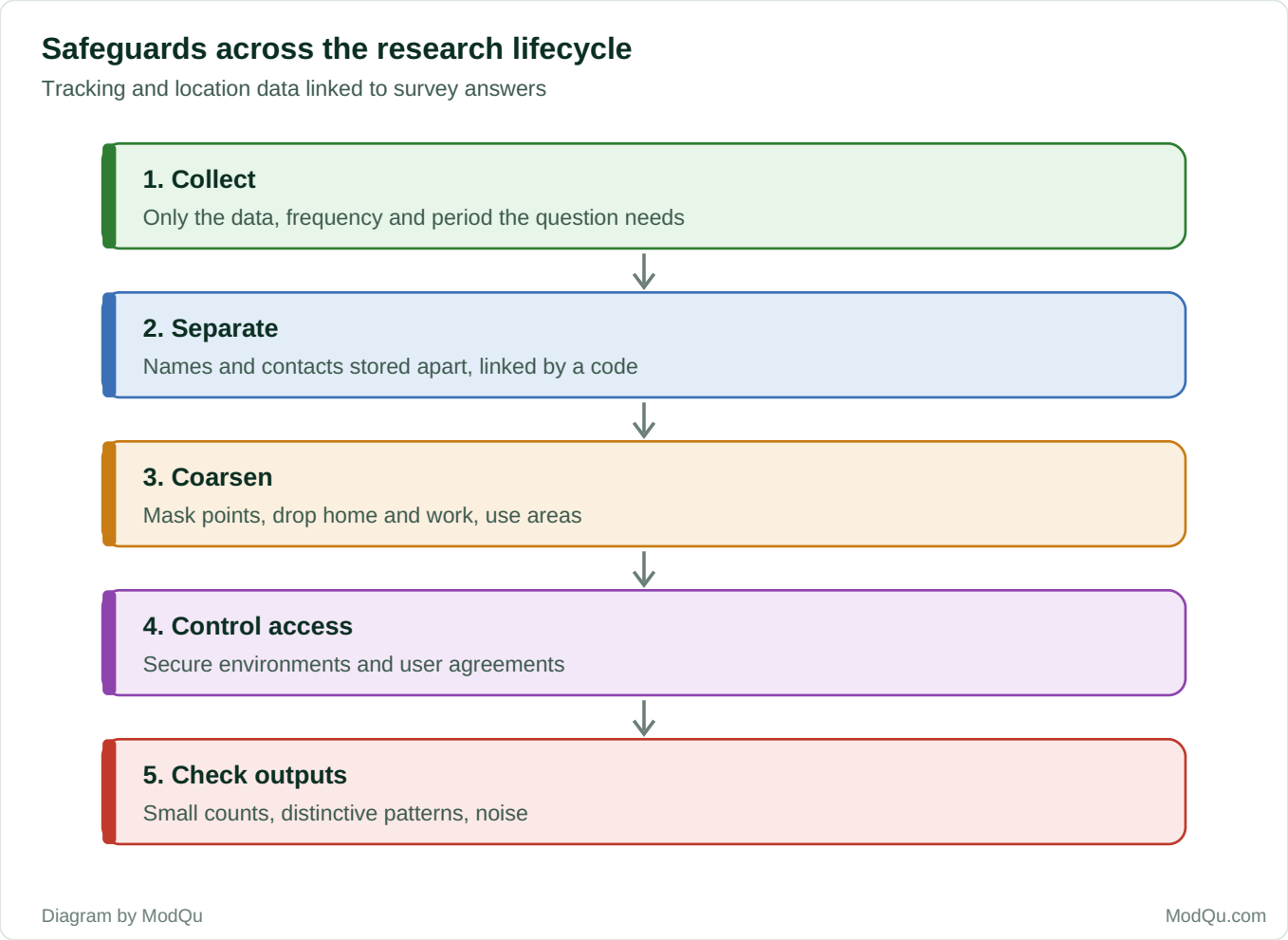


Figure 1. Safeguards for tracking and location data across the research lifecycle. Diagram by ModQu.

Collect only what is needed. If the research question concerns time spent near shops, frequent fine-grained GPS readings may be unnecessary. Lower sampling frequency, collection only during defined periods or on-device processing that sends only summaries can reduce risk.

Separate identifiers. Store names and contact details apart from the trace and survey data, linked only by a code.

Coarsen and mask. VanWey et al. (2005) reviewed the confidentiality problems raised by spatially explicit social data and argued that common approaches to anonymity fail when exact locations are retained. Armstrong et al. (1999) described geographic masking methods, such as moving points by a random distance or aggregating them to areas, which preserve broad spatial patterns while protecting individuals. Home and work locations can be removed or replaced with area codes.

Control access. The most detailed data should be analysed only in secure environments, such as protected servers or research data centres, by approved users under agreements that forbid re-identification.

Check outputs. Maps, tables and models should be checked before release for small counts or distinctive patterns that could reveal individuals. Formal methods such as **differential privacy** add calibrated noise to published statistics and provide a mathematical limit on what any output can reveal about one person (Dwork & Roth, 2014).

Harms beyond confidentiality

Confidentiality is not the only ethical concern. Metcalf and Crawford (2016) argue that research using large digital data sets can harm people even when no individual is named, for example by enabling inferences about groups, neighbourhoods or communities. A study showing where members of a minority group spend their time could be misused even if every individual is protected. Zook et al. (2017) recommend that researchers acknowledge that data are people, recognise that privacy is more than a binary value and consider the strengths and limits of the data before drawing conclusions.

There are also risks of misinterpretation. Location readings vary in accuracy; Zandbergen (2009) found that the accuracy of smartphone positioning depended on whether satellite, Wi-Fi or cellular methods were used, with errors ranging from a few metres to hundreds. People also live their lives across many places, and measuring their "environment" only around home can misrepresent the contexts that matter (Kwan, 2012). Drawing conclusions about individuals or neighbourhoods from imprecise traces can lead to unfair judgements.

Finally, there are risks of coercion and unequal burden. Incentives for sharing tracking data should not be so large that people on low incomes feel unable to refuse. Respondents who decline tracking should still be able to take part in the survey.

Retention, sharing and secondary use

Tracking data are often valuable long after the original study, and there is pressure to keep them. Each additional year of storage, and each new user, adds risk. A responsible plan states how long raw traces will be kept, when they will be reduced to derived variables, such as weekly distance travelled or the number of distinct places visited, and when the raw data will be destroyed. Derived variables usually answer the research question while carrying far less risk.

Sharing data with other researchers or clients needs the same care. Consent given for one study does not automatically cover another. Where future use is anticipated, the consent request should say so, describe the kinds of use and the protections that will apply, and allow respondents to agree to the current study without agreeing to future use. When data are shared, agreements should forbid attempts at re-identification and linkage to other sources without approval.

Transparency to respondents and readers

Respondents should be able to see what has been collected about them. Some research apps show participants a map or summary of their own data, which helps them judge whether they are comfortable continuing and builds trust. Respondents should also be told how to raise concerns and whom to contact.

Readers of the research need transparency too. Reports should state how many invited respondents agreed to tracking, how they differed from those who declined, what was collected, how it was processed and what protections were applied. Without this information, readers cannot judge either the ethics or the quality of the evidence, because the people who agree to tracking may differ from those who do not.

Commercial research

In market research, tracking and location data are used to measure store visits, exposure to outdoor advertising and media use. The same principles apply. The respondent's trust is the basis of the relationship, and a business that uses tracking data in ways respondents would not expect damages that trust for the whole industry. Professional codes require that participants be told what is being collected and why, that data be used only for research and not for direct marketing to individuals without separate consent, and that personal data be protected. Clients who receive reports should receive aggregated findings, not individual traces.

Figure 2 summarises the questions an ethics review should ask before tracking data are linked to a survey.

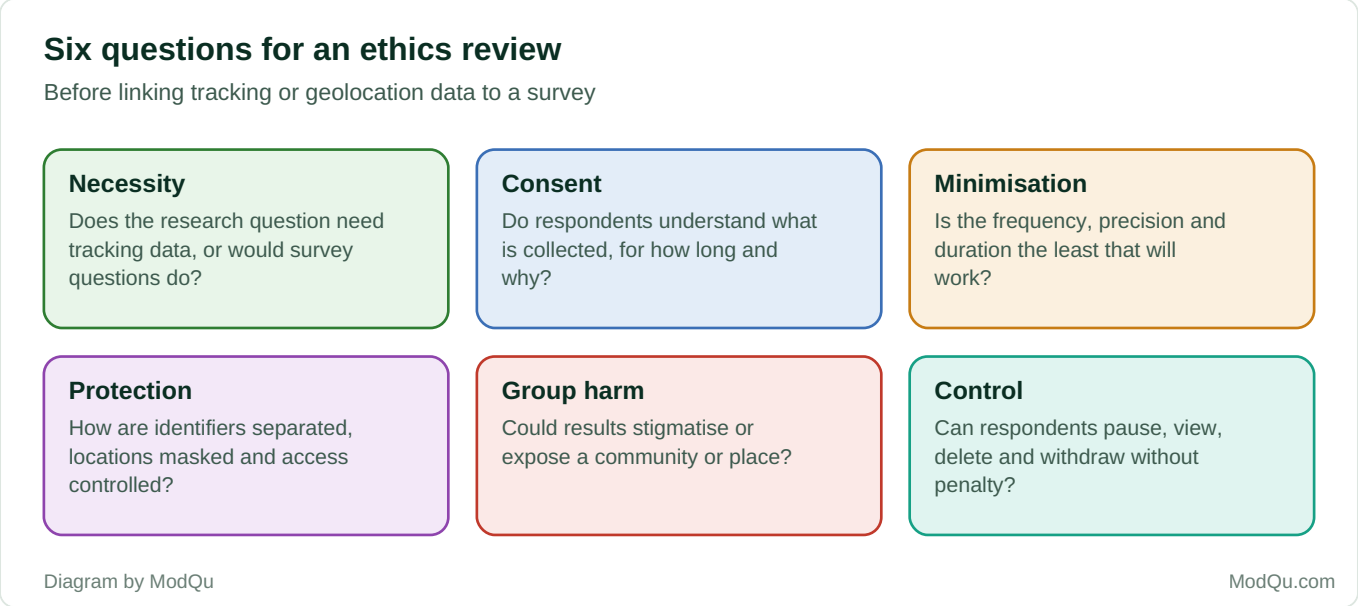


Figure 2. Questions for reviewing a study that links tracking data to survey answers. Diagram by ModQu.

Key takeaways

Linking tracking and geolocation data to surveys can produce richer and more accurate knowledge than either source alone, but the data are uniquely revealing and difficult to anonymise. Responsible use requires informed, granular and revocable consent; collection limited to what the research needs; strong separation, masking and access controls; careful checking of outputs; and attention to harms to groups as well as individuals. Researchers who build these safeguards in from the start protect both their respondents and the credibility of the knowledge they produce.

References

Armstrong, M. P., Rushton, G., & Zimmerman, D. L. (1999). Geographically masking health data to preserve confidentiality. *Statistics in Medicine, 18*(5), 497–525.

Couper, M. P., & Singer, E. (2013). Informed consent for web paradata use. *Survey Research Methods, 7*(1), 57–67. https://doi.org/10.18148/srm/2013.v7i1.5138

de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports, 3*, Article 1376. https://doi.org/10.1038/srep01376

Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science, 9*(3–4), 211–407. https://doi.org/10.1561/0400000042

European Parliament & Council of the European Union. (2016). 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 (General Data Protection Regulation). *Official Journal of the European Union, L 119*, 1–88.

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science, 11*(6), 838–854. https://doi.org/10.1177/1745691616650285

Keusch, F., Struminskaya, B., Antoun, C., Couper, M. P., & Kreuter, F. (2019). Willingness to participate in passive mobile data collection. *Public Opinion Quarterly, 83*(S1), 210–235. https://doi.org/10.1093/poq/nfz007

Kreuter, F., Haas, G.-C., Keusch, F., Bähr, S., & Trappmann, M. (2020). Collecting survey and smartphone sensor data with an app: Opportunities and challenges around privacy and informed consent. *Social Science Computer Review, 38*(5), 533–549. https://doi.org/10.1177/0894439318816389

Kwan, M.-P. (2012). The uncertain geographic context problem. *Annals of the Association of American Geographers, 102*(5), 958–968. https://doi.org/10.1080/00045608.2012.687349

Metcalf, J., & Crawford, K. (2016). Where are human subjects in big data research? The emerging ethics divide. *Big Data & Society, 3*(1). https://doi.org/10.1177/2053951716650211

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. U.S. Department of Health, Education, and Welfare.

Nissenbaum, H. (2010). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.

Revilla, M., Couper, M. P., & Ochoa, C. (2019). Willingness of online panelists to perform additional tasks. *Methods, Data, Analyses, 13*(2), 223–252.

Struminskaya, B., Lugtig, P., Keusch, F., & Höhne, J. K. (2020). Augmenting surveys with data from sensors and apps: Opportunities and challenges. *Social Science Computer Review*. Advance online publication. https://doi.org/10.1177/0894439320979951

Sweeney, L. (2002). *k*-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10*(5), 557–570. https://doi.org/10.1142/S0218488502001648

VanWey, L. K., Rindfuss, R. R., Gutmann, M. P., Entwisle, B., & Balk, D. L. (2005). Confidentiality and spatially explicit data: Concerns and challenges. *Proceedings of the National Academy of Sciences, 102*(43), 15337–15342. https://doi.org/10.1073/pnas.0507804102

Zandbergen, P. A. (2009). Accuracy of iPhone locations: A comparison of assisted GPS, WiFi and cellular positioning. *Transactions in GIS, 13*(s1), 5–25. https://doi.org/10.1111/j.1467-9671.2009.01152.x

Zook, M., Barocas, S., boyd, d., Crawford, K., Keller, E., Gangadharan, S. P., Goodman, A., Hollander, R., Koenig, B. A., Metcalf, J., Narayanan, A., Nelson, A., & Pasquale, F. (2017). Ten simple rules for responsible big data research. *PLoS Computational Biology, 13*(3), e1005399. https://doi.org/10.1371/journal.pcbi.1005399