RESEARCH · RESEARCH INTEGRITY

Ethical Challenges in Secondary Use of Research Data for New Knowledge

Reusing data collected for another study saves money and extends knowledge, but it raises questions of consent, re-identification, harm to groups and fair governance. This article sets out the main challenges and how to handle them.

ModQu research article · 701 words · 5 references (APA 7th edition)

Research data are increasingly shared and reused. Funders and journals expect data to be findable, accessible, interoperable and reusable, the **FAIR** principles (Wilkinson et al., 2016). Secondary use avoids repeating costly data collection, reduces the burden on participants and allows findings to be checked and extended. Reuse is central to **open science**, and it can be an ethical obligation in its own right: data collected at participants' expense of time and privacy should not sit unused if they can answer important questions. But the people who supplied the data agreed to take part in a particular study, under particular promises. Using their data for new purposes raises several ethical challenges. Figure 1 summarises four tests that apply to every reuse.

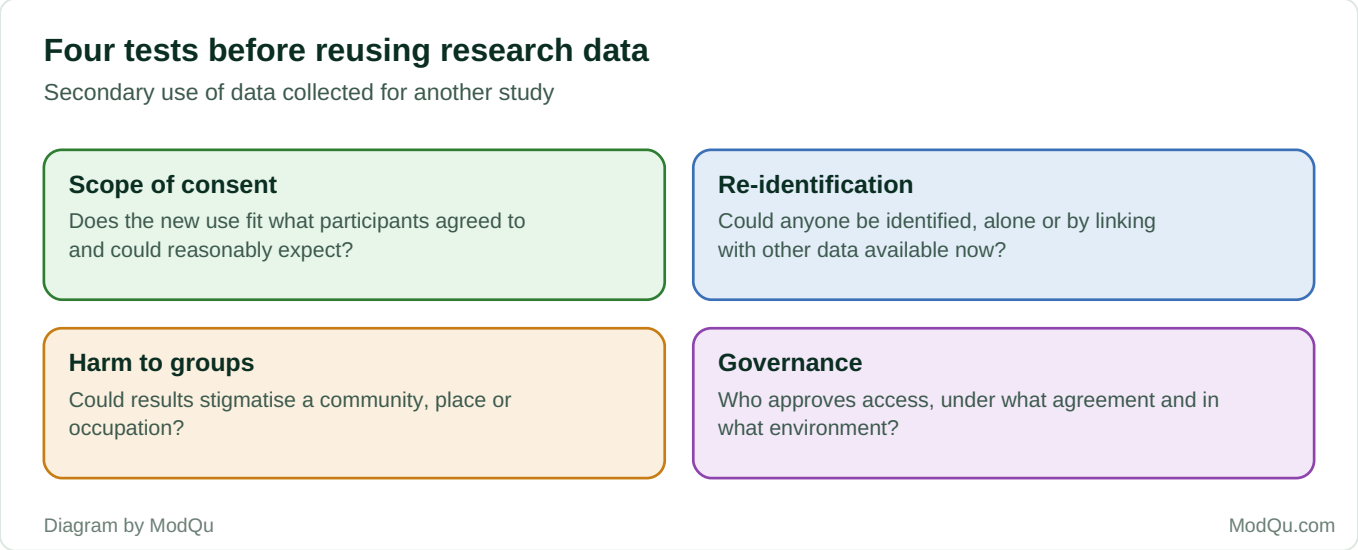


Figure 1. Four ethical tests for secondary use of research data. Diagram by ModQu.

Scope of consent

Participants usually consent to a described study. A new analysis may fall outside what they expected. Canada's Tri-Council Policy Statement allows secondary use of identifiable information without fresh consent only under conditions, such as when consent would be impracticable, privacy is protected and the use would not be contrary to participants'

known wishes; otherwise new consent or ethics review is required (Canadian Institutes of Health Research et al., 2022). Many studies now ask for broad consent to future research at the outset, which helps but does not remove the need for judgement about uses participants could not have foreseen.

The case of the Havasupai tribe shows what can go wrong. Blood samples given for research on diabetes were later used for studies of other topics, including population origins, that members of the tribe found offensive. The dispute ended in a settlement and the return of samples (Mello & Wolf, 2010). The data were not misused technically; the problem was that the new uses broke the trust on which the original participation rested.

Re-identification risk

Removing names is not enough to make data anonymous. Rocher et al. (2019) estimated that fifteen demographic attributes would be sufficient to re-identify almost all individuals in a large population data set, and that releasing only a sample of records offered little protection. Risk also grows over time, as more external data become available for linkage. Secondary users must therefore assess risk afresh, not rely on the original anonymisation.

Harm to groups

Even when no individual can be identified, analysis can stigmatise communities, neighbourhoods or occupations. A finding that a named community has high rates of a condition may affect everyone in it. Researchers should consider how results could be used and, where data concern identifiable communities, involve them in decisions.

Context and meaning

Secondary analysts were not present when the data were collected. Heaton (2008) noted that reusing qualitative data raises particular concerns because meaning depends on the relationship between researcher and participant and on the context of the original study. Misreading data out of context can produce findings participants would reject.

Commercial reuse

The same issues arise in business. Customer survey responses, transaction records and service logs are often reused for analyses that were not planned when the data were collected. Privacy law usually limits personal data to the purposes stated at collection, and customers may object to uses they did not expect, such as combining a satisfaction survey with purchase records to profile individuals. Research suppliers should agree with clients in advance which secondary uses are permitted, keep survey data used for research separate from data used for marketing, and report results only in aggregate.

Governance

Responsible reuse depends on governance: data access committees, user agreements forbidding re-identification, secure environments for sensitive data and records of who used data and why. Governance should be proportionate, with open access for low-risk data and controlled access for sensitive data. Original data collectors should also document the consent given, the collection context and known limitations, so that later users can judge what reuse is appropriate.

Key takeaways

Before reusing data, ask whether the new use fits what participants agreed to, whether anyone could be re-identified, whether groups could be harmed, whether the data can be understood in their original context and whether access is properly governed. Secondary use generates valuable knowledge when it respects the trust that made the data possible.

References

Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, & Social Sciences and Humanities Research Council of Canada. (2022). *Tri-Council policy statement: Ethical conduct for research involving humans – TCPS 2 (2022)*. Government of Canada.

Heaton, J. (2008). Secondary analysis of qualitative data: An overview. *Historical Social Research, 33*(3), 33–45. https://doi.org/10.12759/hsr.33.2008.3.33-45

Mello, M. M., & Wolf, L. E. (2010). The Havasupai Indian tribe case: Lessons for research involving stored biologic samples. *New England Journal of Medicine, 363*(3), 204–207. https://doi.org/10.1056/NEJMp1005203

Rocher, L., Hendrickx, J. M., & de Montjoye, Y.-A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications, 10*, Article 3069. https://doi.org/10.1038/s41467-019-10933-3

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. 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*, Article 160018. https://doi.org/10.1038/sdata.2016.18