

RE-USE OF EXISTING DATA IN RESEARCH

Researchers have a responsibility to protect participants from any potential harm or distress that may arise from their participation in a research project. Therefore, researchers wishing to use existing datasets for a new research project (whether the original data were collected for research, clinical or other purposes) need to consider the dignity, rights, safety and well-being of those who provided the data, including whether information may need to be provided to those individuals about the new project, and what kind of ethics approval and/or consent/permissions they may need to obtain.

There is likely to be minimal harm to participants if their data has been truly anonymised, via the removal of any identifying data (not just names but dates of birth, addresses, post codes, phone numbers, user IDs, IP addresses etc.). However, consideration should still be given to any new research purpose, in terms of whether the original participants (or relevant groups of individuals) would be likely to object should they become aware of the project (this would need to be considered on a case by case basis).

Ethical approval is therefore NOT required for research that only involves existing data that has been *robustly* anonymised, such that the original providers of the data cannot be identified, directly or indirectly, by anyone (i.e. it does not involve personal data). In such cases, researchers are encouraged to use the self-declaration process available via the online Ethics Application System, to ensure that they have covered all relevant considerations in using existing data as part of their project, and to ensure that this process has been appropriately documented.

Informed consent is not a legal requirement for truly anonymised data, although from an ethical stand point, the researcher should seek informed consent where possible for the re-use of data for a new research purpose (either by contacting the participants directly, or by requesting evidence from the original researcher/data provider to confirm that consent for the data to be used for secondary research purposes has been obtained, along with a copy of the terms of the original consent so that the data can be used in line with the original consent).

If this is not possible, then in general, providing the data has been robustly anonymised, then it would be acceptable for the data to be used for a secondary research purpose, even if consent for secondary research (or primary research in the case of clinical/other data) was not originally sought. However, if consent had been sought for secondary research, but not been granted by a participant, then that participant's data can never be used.

Researchers should be aware that even when they have sought to anonymise data for secondary analysis, there is still a risk that the original participants could become identifiable, even within large scale data sets - perhaps because they have distinctive characteristics (e.g. families with large numbers of children may stand out in cohort studies) or because a method of analysis combines variables in ways that identify small groups within a larger sample. In such cases, the data should be considered to be pseudonomised, and would still be classed as personal data, thus requiring ethical approval, and requiring compliance with data protection legislation. In particular, the General Data Protection Regulation (GDPR) sets out specific obligations relating to the information that should be provided to the original data subjects when using re-data for a new purpose, unless certain circumstances apply – for example if re-contacting the participants is impossible or would involve disproportionate effort. For further

guidance, refer to the Specialist Research Ethics Guidance paper 'Principles of Anonymity, Confidentiality and Data Protection'.

Researchers should also be aware that where datasets containing personal data are obtained from an external company or organisation, data may not have been 'provided' by people directly and with their knowledge (e.g. mobile phone data, loyalty card data, location data, internet activity logs). Researchers may gain access to such data to analyse it on the external organisation's behalf, and in some cases the analysis might be research-led, whilst in other cases it may be driven by the needs of the organisation (e.g. where the researcher is acting in a consultancy role).

Ethics approval would be required for any work using personal data obtained from an external organisation that falls under the definition of research set out in the Ethics Policy Governing Research Involving Human Participants, Personal Data and Human Tissue. Additionally, the researcher must consider the requirements of data protection legislation, as mentioned in the above paragraphs. As part of the ethical review of such research, the applicant and reviewers should consider the ethical implications of how the data was generated (e.g. participants' potential lack of awareness of their data being used for research), as well as the use to which the analysis is to be put by the external organisation. The researcher should also check whether the external organisation is complying with relevant data protection legislation in collecting, processing and sharing the data.

In addition, it should be noted that even if data from an external organisation has been de-identified when passed to the researchers, the results of the researchers' analysis might be re-identifiable by the organisation (e.g. via the use of a unique identifier), and may be used directly to do things that might be deemed unethical by many people (e.g. the identifiable results could potentially be sold on to other companies). If it is likely that the external organisation will be able to re-identify participants from the analysis, then ethical approval should be obtained, even if the researchers will not have access to the personal data themselves.

Finally, all researchers are strongly encouraged to consider the possibility of secondary research and data sharing at the outset, before the primary data collection begins, and to build this in to the informed consent process. As such, where a researcher plans to use the data for secondary research (or to share the data) they should include details of this in the information given to potential participants, and include an appropriate section on the consent form.

UREC-approved providers of research datasets

A number of organisations provide access to datasets for research purposes. The UREC has approved a number of these providers, meaning that data obtained from them can be used for secondary research purposes without explicit informed consent from the participants, even if the dataset contains personal data (NB. it should be noted that ethics approval should still be obtained if personal data will be accessed). This is due to the fact that they require the researcher to follow a series of robust procedures to gain access to the data, and often require the researcher to comply with a number of specific requirements (e.g. following the terms of any original consent).

A list of UREC-approved organisations can be accessed here:

http://www.sheffield.ac.uk/polopoly_fs/1.670012!/file/URECApprovedDataProviders.docx

The UREC considers the merits of such arrangements on a case-by-case basis; researchers wishing to establish whether data obtained from a particular provider, but not already on the

above list, may be used without informed consent, should provide details to the Minute Secretary to the UREC.

Governing Principles and Procedural Steps for the Transfer of Research Data which relates to human participants between Principal Investigators within The University of Sheffield

The University has developed guidance for those wishing to share research data with other researchers internally, to ensure that ethical and legal requirements are met. This guidance can be found here:

http://www.sheffield.ac.uk/polopoly_fs/1.670014!/file/RDMTransferSENATEapprovJun16.doc