Data consent and ethics

Codes of ethics govern all research. These are formal statements of principles relating to the moral conduct of research produced by funding bodies and institutions. Generally applied as guidelines of non-malfeasance embracing informed consent, anonymity, and confidentiality they intend to protect participants from any potential harm caused by participating in research.

Research codes of ethics should not prohibit data sharing as long as you pay early attention to consent, anonymisation, and confidentiality through access conditions of reuse.

Consent

Obtaining the freely given consent of people to participate in your research is essential. Always seek consent and never infer that consent has been given.

To obtain consent, researchers should inform participants as to the purpose of the research, what will happen to their contribution – including archiving and data sharing, indicate the steps taken to safeguard the confidentiality and anonymity of participants, and outline the right to withdraw from the research.

Researchers can be flexible as to how they approach consent. It can be negotiated and discussed at different points in the research – although obtaining retrospective consent is often time consuming and inconvenient for both researcher and participant - and where obtaining written consent is problematic, the option of verbal or recorded consent is available. However, researchers should not unnecessarily prohibit data reuse through restrictive language on consent forms, for example, by stating that the data will only be shared by the research team or only in publications. If your funding body or institution has a data archiving and reuse policy, and even after seeking specialist advice, you have serious concerns about consent for reuse and archiving compromising data collection, there is an expectation you make a case for an exemption rather than simply ignoring consent for archiving and reuse.

Anonymisation

Because social science is concerned with society and human behaviour, an anonymisation strategy to protect the identity of participants is critical to ethical research. As with consent, planning anonymisation before undertaking data collection produces both better informed consent and a less resource intensive process of data anonymisation.

CESSDA AS
CESSDA House

Parkveien 20
5007 Bergen
NORWAY

phone: +47 55 58 21 18
e-mail: cessda@cessda.net
www.cessda.net

Usually anonymisation applies to two kinds of identifiers: direct and indirect. Direct identifiers are the obvious variables like name, address, or telephone numbers that specifically highlight a participant. Indirect identifiers when pieced together could also reveal an individual, for example, by cross-referencing occupation, employer, and location.

Knowing what data you wish to collect will help guide an anonymisation strategy consistent across cases and produce ethically responsible reusable data that does not contravene data protection laws. Using meaningful pseudonyms and replacements for identifiers is a good start, meaningful in the sense of preserving the character of the identifier while concealing the identity. For example, instead of "Cologne" use "Major German metropolitan area" and instead of "Heiko" use "Kai". This is preferable to replacing identifiers with "City" or "Name" or, worst of all, "deleted".

If using pseudonyms is unworkable, could you apply restrictions on upper and lower ranges of variables? Can you remove a variable without it compromising the re-use value of the data (in which case, ask if you should you even measure that variable)? Could you apply low-level aggregation of data, like moving to a larger spatial unit or transforming age from a continuous variable into a discrete categorical one? In any case, it is best practice to create a log of anonymisation undertaken and to flag anonymised identifiers so it is clear that something is anonymised.

Of course, the principal of informed consent allows participants to waive their right to anonymity should they wish, and if in the researcher's judgment, no harm will result. In the case of oral history or elite interviews, tied to the participant's identity are their memories, perceptions and experiences. Consequently, data in these approaches is not anonymised even if it is subject to tighter access conditions or a long embargo.

Confidentiality

A commitment to making your data available for reuse does not entail uncontrolled, unrestricted public reuse. You could chose to make your data "open" - available without restriction - provided it did not contravene consent and anonymity commitments. However, the focus in social sciences on human subjects often means safeguards on access have to be in place.

If you deposit your data in a dedicated data archive, then as a condition of reuse potential users must agree to a legal agreement not to identify or re-contact participants. You could also negotiate stricter access conditions be applied, requiring the approval of potential researchers for on-site data use and prohibiting the downloading of data. However, these higher barriers of access must be justified and negotiated; they are not a strategy to deny access to other researchers because you are concerned about ensuring adequate time to publish. In scenarios such as this, you can negotiate an embargo on the data for an agreed period.

References and further reading

ABDS Guides (2012): Ethics, Consent & data Sharing. http://www.ands.org.au/guides/ethics-working-level.pdf.

Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:NOT.

Lagoze, C., Block, W.C., Williams, J., Abowd, J. and Vilhuber, L. (2013): Data Management of Confidential Data. doi:10.2218/ijdc.v8i1.259. http://www.ijdc.net/index.php/ijdc/article/download/8.1.265/311.