DRYAD   Human subjects data

Data archived in Dryad are publicly available, and any human subjects data must be properly anonymized and prepared under applicable legal and ethical guidelines.

Tips for preparing human subjects data

- Ensure that there are no direct identifiers.

- Limit indirect identifiers to no more than three.

- Remove any nonessential identifying details.

- Aggregate data that may be revealing. For example, group ages into ranges.

- Reduce the precision of a variable. For example, remove day and month from year of birth; use county instead of city; add or subtract a small, randomly chosen number.

- Restrict the upper or lower ranges of a continuous variable to avoid outliers.

- Provide a README file with thorough documentation.

Dryad's policies on human subjects data are in accordance with published standards such as Hrynaszkiewicz et al. (2010). As a result, **Dryad does not allow any direct identifiers** such as an individual's name, initials, email address or postal code.

Furthermore, **a dataset may contain no more than three indirect identifiers** such as demographic, biological and geographic data that could lead to identification if combined with other available data.

> *...researchers should consider removing indirect identifiers and other information that could lead to 'deductive disclosure' of participants identities. Deductive disclosure of individual subjects becomes more likely when there are unusual characteristics or the joint occurrence of several unusual variables. Samples drawn from small geographic areas, rare populations, and linked datasets can present particular challenges to the protection of subjects' identities.*

- National Institutes of Health (2003)

To minimize the risk of disclosure, work with your institutional review boards and/or directly with study participants. Whenever possible obtain consent to release participant level data. While our curation team evaluates each submission and we provide guidance, researchers are responsible for ensuring their data do not contain information that can be used alone or in conjunction with other data to personally identify any individual.

To properly de-identifying your data, you will need to consider both direct and indirect identifiers. Following are partial listings of direct and indirect identifiers.

Direct Identifiers (none allowed)

- Name

- Initials

- Address, including full or partial postal code

- Telephone or fax numbers or contact information

- Electronic mail addresses

- Unique identifying numbers (e.g., social security number)

- Vehicle identifiers

- Medical device identifiers

- Web or internet protocol addresses

- Biometric data

- Facial photograph or comparable image such as fMRI data showing facial structures

- Audiotapes of participants' voices

- Names of relatives

- Dates related to an individual (e.g., birthdate, interview date)

Indirect Identifiers (maximum of three per dataset allowed)

· Gender

· Rare disease or treatment

· Place of treatment

· Place of residence, or geographic location

· Name of health professional responsible for care

· Sensitive data such as illicit drug use or risky behavior *

· Criminal record

· Place of birth

· Socioeconomic data, such as occupation, place of work, income, or education

· Household and family composition

· Sexual attitudes, practices and orientation

· Organizational membership such as religious, political and trade

· Information regarding an individual's psychological well-being or mental health

· Anthropometric measures such as height and weight

· Multiple pregnancies

· Ethnicity, race, indigenous status

· Small denominators—population size of less than 100

· Very small numerators—event counts of less than 3

· Year of birth or age

· Verbatim responses or transcripts

Tools / resources

- Amnesia - From the Institute for the Management of Information Systems, Amnesia can be used to remove and transform identifiers in a dataset.

- pydeface and mri_defaceddress - Used to remove/obscure facial structures in fMRI data

References:

Canadian Institutes of Health Research (2005, September) CIHR Best Practices for Protecting Privacy in Health Research. Ottawa, Ontario: Public Works and Government Services Canada. Retrieved January 23, 2017 from http://www.cihr-irsc.gc.ca/e/29072.html

Hrynaszkiewicz I, Norton ML, Vickers AJ, Altman DG (2010) Preparing raw clinical data for publication: Guidance for journal editors, authors, and peer reviewers. BMJ 340: c181 http://dx.doi.org/10.1136/bmj.c181

The Inter-University Consortium for Political and Social Research (2012) Guide to social science data preparation and archiving: Best practice throughout the data life cycle, 5th ed. Ann Arbor, MI: Institute for Social Research, University of Michigan. Retrieved January 30, 2017 from http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf

National Institutes of Health (2003, March 5) NIH Data Sharing Policy and Implementation Guidance Retrieved March 13, 2018 from https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#hs

Olesen S, Australian National Data Service (2014) ANDS Guide to Publishing and sharing sensitive data. Australian National Data Service. Retrieved February 01, 2017 from http://ands.org.au/guides/sensitivedata.pdf

The UK Data Service (n.d.) Anonymisation. Retrieved February 1, 2017 from https://www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation

United States Department of Health and Human Services (2007, February 2) HIPAA privacy rule: Information for researchers. Limited data set and data use agreement. Retrieved January 20, 2017 from https://privacyruleandresearch.nih.gov/pr_08.asp#8d

Van den Eynden V, Corti L, Woollard M, Bishop L, Horton L (2011) Managing and Sharing Data: Best Practice for Researchers, 3rd ed. Essex, UK: UK Data Archive. Retrieved February 01, 2017 from http://www.data-archive.ac.uk/media/2894/managingsharing.pdf