Why is sharing data useful to me? Why should I share data that I have worked very hard to collect and analyze?

Researchers who share data benefit in the following ways:

- Citations of datasets available from NAHDAP are harvested by key social science indexes.
- Data in the public domain generate new research which cites the original research.
- Data used for secondary analysis is published more widely than data not shared, extending the research productivity of the original investigator.
- NAHDAP frees investigators from using their resources to share data.
- NAHDAP preserves the data, which original researchers can obtain if their copies are lost or destroyed.
- Archiving data helps researchers meet requirements of NIH and NSF data management plans.
- Depositors can access downloads statistics of their study files. The research community’s interest in a dataset can assist with the success of future grant funding proposals.
- The original investigator can more easily collaborate with colleagues on future research projects by using data from NAHDAP that is available in a variety of formats.

My research data and documentation are currently not in a format that can be released to secondary users. I do not have the resources or time to prepare it for broader distribution.

The National Institute on Drug Abuse (NIDA) funds NAHDAP to assist grantees in preparing data for sharing. NAHDAP staff clean and standardize data files, metadata, and documentation in consultation with grant staff. NAHDAP is built on the infrastructure of the Inter-university Consortium for Political and Social Research (ICPSR), which disseminates digitally stable data files and searchable PDF codebooks and documentation.

My data are very complicated. I am not sure users who were not affiliated with the original study will be able to use the data. Will NAHDAP staff provide user support?

ICPSR has three levels of user support. The central email and telephone help desk tracks all user support inquires. Technical questions about data downloading and software issues are answered by tier 1 support staff. NAHDAP staff who prepared the data for release provide tier 2 support for questions about specific datasets, data content, and data structure. The NAHDAP director and manager provide tier 3 support for complex technical questions. Depositors are not expected to provide ongoing user support, but rather to provide all the documentation necessary for secondary data users to make sense of the original data collection. NAHDAP’s archival holdings include many very complex data systems that have been successfully analyzed by responsible researchers.

The data from my study are on very sensitive topics such as criminal activity. I believe the risk to participants is very high should they be re-identified. How can I protect these respondents if I release data to the public?

ICPSR evaluates all data files for disclosure risk using state-of-the-art techniques developed under a grant from the National Institutes of Health. From this evaluation, staff recommend a method of data release that protects respondents from re-identification while retaining the analytic utility of the data. Release options include public release and public online analysis; restricted release with an approved user agreement; enclave-only access; and online access after disclosure.
protections are applied. A full public release is only warranted when there is little risk of re-identification or the data have been sufficiently transformed to substantially reduce that risk. NAHDAP staff can provide information to depositors about how to release the data as a restricted-use dataset.

My data are collected from very vulnerable populations. I am committed to the well-being of this population and am concerned that these data may be used to portray them in an injurious way.

ICPSR is strongly committed to protecting vulnerable individuals from being identified by data analyses. ICPSR’s policy is also to trust that responsible science, which includes appropriate analytic methods and peer reviewed venues for research results, is adequate to protect vulnerable populations from inappropriate, unfair, and inaccurate portrayals. In order to have valid scientific discussions of issues that vulnerable populations face, researchers must be willing to share the data and methods in an ethically responsible manner so other researchers can replicate or refute their findings.

In the informed consent signed by respondents, I promised that the data would only be used by an approved research team. How can I share the data when I made such promises to my respondents?

Unless the informed consent names the members of the research team specifically, an amended Institutional Review Board application that includes a plan for data protection and dissemination can be filed with the lead institution to define the research team as those persons known to the original researchers. Restrictive informed consent may prevent the release of data as public-use, but do not preclude the possibility of a research team that is defined by a group of restricted- or limited-use agreement holders. With such agreements, the researchers using the data are known to ICPSR and to the original research team.

Can I read and approve research proposals based on my data? I would like to determine the nature of research done with the data.

ICPSR’s policy is that responsible use of secondary data should be independent of the original researchers’ priorities. When data are distributed under restricted-use agreements, a research proposal is required in order to screen users for a credible research project and to ascertain whether the data are needed for that project. The proposal, however, is screened only by the designated administrator at NAHDAP.

Can my data be embargoed until I or my research team finish all our planned analyses?

ICPSR has a delayed dissemination policy that allows researchers to deposit data earlier in the research process so that they may benefit from the data and documentation preparation services offered by staff. Delayed dissemination requires depositors to commit to a timeline, which is usually two years from deposit to data release. Depositors, however, have access to ICPSR files as soon as they are prepared and need not wait for the public release.

I don’t mind depositing the baseline data from my longitudinal study, but can I hold on to the subsequent waves of data for a while?

The utility of longitudinal studies lies primarily in the follow-up embedded in the research design. While the baseline data are valuable in the short run, NAHDAP will work with depositors on a delayed dissemination plan for deposit and release of the subsequent waves of data.