**Central Connecticut State University**

***Institutional Review Board***

**Guidance Related to Data Sharing**

**What is data sharing?**

Increasingly, journals, conferences, grant funders, professional organizations, and other entities are requesting or requiring researchers to share the data sets from their research studies in a publicly available repository. The reasons for such sharing include: increased confidence in the research community; replication and verification of findings; secondary data analysis for new research questions; encouraging transparency and accountability; improving confidence in conclusions drawn; among others. Shared data refers to information on individuals and might include all primary data used to produce findings in a research study, which could be in the form of tables, graphs, spreadsheets, code books, transcripts, artifacts, among other data formats.

**Will you be sharing data?**

As noted above, some journals, conferences, grant funders, professional organizations, and other entities require data sharing.

Other organizations, such as the ones where research data are collected, might put specific restrictions on data sharing, even disallowing it entirely.

In many cases, the researcher can choose whether they wish to share data sets, such as in the case of the many journals that offer this option, but do not require it. *If you reasonably suspect that you might be sharing your data, then best ethical practice dictates that you should notify participants prior to data collection.*

**Considerations for Data Sharing**

Identifiable Information: Data that are collected with identifiable information should be deidentified prior to data sharing. Deidentification might include:

* removing identifiable fields from the data set including, but not limited to: names (or replacing with pseudonyms), addresses, email addresses, etc.;
* redacting information from transcripts and artifacts that could be used to identify participants, places, or other individuals; and/or
* removing demographic information when a small number of individuals are represented from a particular group, making their identification possible.

Proprietary Information: In general, proprietary information should be omitted or redacted to ensure that the data set does not share proprietary information.

Anonymously Collected Data: If the data set provides no means of identifying a research setting, participants, or any other individuals (via participants’ responses), the data set should not need to be altered prior to sharing. However, as best practice, participants should still be informed about the possibility of data sharing through the informed consent process (see next section).

**Sample Language for Informed Consent**

Informed consent procedures and documentation should include all anticipated procedures for participants and their data, as well as how participants’ confidentiality will be maintained. This includes data sharing. *If you reasonably suspect that you might be sharing your data, then best ethical practice dictates that you should notify participants prior to data collection.*

*Sample language about data sharing for use in informed consent forms/documentation*:

Sample 1: We will make our best effort to protect your statements and answers, so that no one will be able to connect them with you. These records will remain confidential. Federal or state laws may require us to show information to university or government officials [or sponsors], who are responsible for monitoring the safety of this study. Any personal information that could identify you will be removed, redacted, or changed before files are shared with other researchers or results are made public.

Sample 2: We may use the information you provide for future research studies. We will not ask for your additional informed consent for these studies. We will store the research data at a research repository that allows sharing of data with interested parties. Before being made available to other researchers, the data will be reviewed to make sure it cannot be used to identify you in any way.

Whereas, IRB norms and samples have traditionally included statements in informed consent forms about the *destruction of data* after a period of time (ex. 3 years), such a statement can still apply to the *identifiable* versions of data sets, though the *deidentified* data sets could be shared publicly and in perpetuity.

**Sources Consulted and for Further Information:**

American Psychological Association. (2024). *Data sharing*. Retrieved September 24, 2024, from <https://www.apa.org/pubs/journals/resources/data-sharing>

Center for Open Science. (n.d.) *Tools*. Retrieved September 24, 2024, from <https://www.cos.io/tools>

Inter-University Consortium for Political and Social Research. (n.d.). *Guide to social science data preparation and archiving: Best practice throughout the data life cycle* (6th ed.). Author. Retrieved September 24, 2024, from <https://www.icpsr.umich.edu/web/pages/deposit/guide/index.html>

International Committee of Medical Journal Editors. (2024). Clinical trials. Retrieved on October 24, 2024, from [https://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html](https://nam10.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.icmje.org%2Frecommendations%2Fbrowse%2Fpublishing-and-editorial-issues%2Fclinical-trial-registration.html&data=05%7C02%7CMalleys%40ccsu.edu%7C50e887ebc27341157bde08dcf45ed8bc%7C2329c570b5804223803b427d800e81b6%7C0%7C0%7C638653935594575746%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C0%7C%7C%7C&sdata=K%2F0bR4CqegMt8gWiSStYRiW%2FsLY89HEn7ijv6%2BkytJ8%3D&reserved=0)