GUIDANCE: The Meaning of Anonymous, Confidential, and De-Identified and Implications for Data Sharing or Re-use

The IRB often finds that the terms anonymous, confidential, and de-identified are used incorrectly. The following explains (1) the meanings of the terms as they relate to a person’s participation in research and the way that their data are collected and maintained for analysis, and (2) issues related to the sharing or re-use of data sets.

Participation
A person’s participation in research can be described as anonymous if it is impossible, even for the investigator, to know whether or not a given person participated in the study. For example, an internet survey that cannot be linked in any way to the participant would be considered anonymous. A study involving any face-to-face contact with a participant can never be considered anonymous.

When participation is confidential, the research team knows that a particular person has participated in the research but is obligated not to disclose that information to others outside the team, except as clearly explained in the informed consent document.

Data
Data are anonymous if no one, not even the researcher, can connect the data to the person who provided it—no identifying information is collected from the individual. Investigators must be aware, however, that even if no direct identifiers (name, address, student ID, etc.) are collected, identification of a participant may be possible from unique individual characteristics (indirect identifiers). For example, a participant who is a member of a certain ethnic group or who was studied because of distinctive personal accomplishments or medical history might be identifiable from even a large data pool.

Data are confidential when there continues to be a link between the data and the person who provided it, meaning that the person can be identified even after the research is completed. The research team is obligated to protect this data from disclosure outside of the research team, according to the terms of the research protocol and the informed consent document. Both should explain to the IRB and the prospective participant how their data will be protected against accidental disclosure and access by unauthorized persons. For example, code numbers may be used on all documents other than the consent form, the list of names and code numbers should be stored separately from other study documents and under secure conditions, the data files may be password-protected or encrypted, and the length of time the data will be stored or the destruction date should be specified. Note that coded data are not anonymous.
Data are considered *de-identified* when any direct or indirect identifiers or codes linking the data to the identities of particular persons are destroyed. This means that even the investigator cannot find data from a given person. If the investigator plans to de-identify the data at some point, include this information in the consent document. This offers the investigator the advantage of being able to use the data for some other research question.

**Sharing of Data Sets**

Sharing of data sets with another investigator or some external database can often be a gray area for investigators and IRBs. IRB has no problem with sharing of data sets with other investigators or databases or use of the data for another research question *if the data are de-identified*—that is, this step was included and approved in the original application. If this has been done, the investigator need not apply to IRB for these activities.

If the sharing of confidential research data is known to be a possibility when research participants are recruited, this possibility should be explained in the consent document and the participants should be given a separate opportunity to consent to this aspect of the study or to refuse. If this possibility was not part of the original informed consent and an occasion to share or re-use coded data for another purpose arises, the investigator must apply to IRB for permission to do so. Depending on study details, the IRB may require the investigator to re-contact participants for permission for the sharing or new use of the data, or the investigator may apply for a waiver of consent. The latter requires very strong justification, and IRB may or may not approve the request. If allowed to share non-de-identified data, the investigator must develop a data use agreement with the dataset recipient for the allowable use of the data and protection of its confidentiality.

Investigators should feel free to consult with a research compliance specialist or the Director of Research Compliance if they have questions about proper procedures for the re-use or sharing of data.

*Investigators should review the HRPP policies “Research Using a Limited Data Set” and “Research Using a Restricted Access Data Set” for more information about the nature of a data use agreement and should view themselves as the owner/source of the dataset who is responding to a request from another.*