



U-MIC TRANSCRIPT Anonymous, Coded, and De-identified Data

We use the terms *anonymous*, *coded*, and *de-identified* to refer to three methods researchers use to protect study subjects' privacy and confidentiality.

We consider study data *anonymous* when no one—including members of the study team—can at any point link the data to a subject's identity. For example, a study in which subjects complete surveys may involve anonymous data, provided that the surveys contain no direct or indirect identifiers. If, from the time of data collection, no links exist between subjects and the data they provide, we can consider those data anonymous.

Direct identifiers include details such as a subject's name, address, Social Security number, or student or employee ID number. Not all identifiers are direct. Details that someone could combine with other information in order to deduce a subject's identity are considered *indirect* identifiers. Physical characteristics or ethnic heritage may be examples of indirect identifiers, if they enable someone to recognize a study subject. Study data that contain identifiers, whether direct or indirect, cannot be considered anonymous.

When researchers collect identifiable subject data, and then replace all identifiers with random series of numbers or letters, we refer to those data as *coded*. Coding allows researchers to discuss a study with people outside their study team without revealing their subjects' identities. Researchers should be sure to store all identifying information separately from coded study data.

Note that coded data are not anonymous, since members of the study team have access to identifiers.

We consider data *de-identified* when a researcher destroys all of the information that might link study data, either directly or indirectly, to subjects' identities. After de-identification, even members of the study team no longer possess means to link the data to study subjects. Still, since researchers had access to identifiers prior to de-identification, we cannot consider the data anonymous.

Anonymous refers to data that no one, including members of a study team, can at any point link to a subject's identity.

Coded refers to data that no one outside a study team can link to a subject's identity.

U-MIC TRANSCRIPT Anonymous, Coded, and De-identified Data

De-identified refers to data that used to be fully identifiable or coded, until the researcher destroyed all of the identifiers linking the data to study subjects.

Contact the IRB for more information about anonymous, coded, and de-identified data.

Posted: 6/27/2012