

Guidance: NHSR – Publicly Available Data/Datasets

To qualify as human subjects research, activities must meet both the definition of research and human subjects. Projects using data that are publicly available or a publicly available dataset qualify as “Not Human Subjects Research” as they meet the definition of “research” but not “human subjects.” To meet the definition of human subjects, the project must involve interaction with participants or the use of **private** information.

Terminology:

Publicly Available Data means that the general public can obtain the data. Data may include identifiers or identifying information.

Publicly Available Datasets are those datasets shared without conditions on use. This may include datasets that require payment of a fee to gain access to the data.

Considerations:

1. Publicly available data/datasets include material that is widely available, even when:
 - Identifiers are included, so long as individuals whose data is represented would not reasonably expect data to not be made public.
 - A fee is charged for obtaining the data.
 - Access to the data is limited to researchers, if any researcher with a standard academic or research affiliation has access.
 - Social media data that is publicly available (e.g., no log in required to view information; user data whose profiles are set to public; etc.)
2. Data that is NOT publicly available includes:
 - Generally regarded assumption of private
 - Restricted Use Datasets
 - Files upon which use restrictions are imposed and access is limited to those who sign any necessary agreements to receive the data under these restrictions. These may be distributed by governmental agencies, research organizations, and others.
 - Data in the electronic medical record;
 - Social media data labeled as "private" by the data owner, or not readily available without permission of the site Owner/Administrator under the Terms of Service of the site;
 - Data protected by Copyright; and
 - Data or biospecimens that have access restrictions (e.g., are only available to clinicians or qualified researchers or may only be accessed on a secure server).
3. When seeking access to any dataset, investigators must pay close attention to how the dataset is described by the supplier. A supplier may have the ability to provide different aspects of the same general data as both a public data set (e.g., a research data set which has been stripped of

identifiers - NHSR) and a restricted use data set (e.g., that same data set is provided with identifiers – human subjects research).

4. Researchers combining multiple datasets, even if all are publicly available, may not qualify for NHSR. Please contact irb@mail.wvu.edu in these scenarios.