|  |
| --- |
| **Overview** |

This guidance represents the Purdue HRPP’s determination that Purdue investigators’ access to specified “public use” datasets does not constitute research with human subjects. This guidance applies provided that access does not involve access to identifiable information about the persons from/about whom the data were collected. Please read the following guidance carefully before considering a dataset publically available.

|  |
| --- |
| **What is an Existing Publically Available Dataset** |

Secondary data analysis of publicly available data is a common research method. Increasingly, federal agencies supporting research require investigators to make the data they collect publicly available. Additionally, many professional organizations and journals require that research datasets of published works be made accessible to encourage scholarly interpretation and replication of research.

Data are not considered publicly available if access to the data is limited to researchers. Social media posts are not considered publically available datasets under this guidance. Public Use Datasets are datasets prepared by investigators or data suppliers with the intent of making them available for public use. The data available to the public are not individually identified or maintained in a readily identifiable form.

Restricted use data are not publicly available and are not covered under this policy. Restricted use data is defined as files distributed by federal agencies, repositories, and research organizations upon which use restrictions are imposed. These files are usually, but not always, accompanied by a data use/transfer agreement, data access agreement, or user terms and conditions that detail restrictions on use of the data. The restrictions vary, but they typically involve secure data storage, encryption, password protected computers, access limitations or destruction parameters. Researchers must not assume that a dataset not listed in this guidance is public without contacting the Purdue HRPP/IRB.

|  |
| --- |
| **Applicability** |

This guidance will be updated, as needed, to include additional datasets that are determined by the Purdue HRPP to be publically available. If at any time, the provider of the publically available dataset asserts that IRB review is required, the dataset is no longer publically available, and an application to the IRB must be submitted. Investigators who encounter new publically available datasets not covered in the list below may consult with the HRPP/IRB to have the dataset assessed for applicability to be added to this list.

The datasets to which this guidance is applicable are limited to those identified below:

* Agency for Healthcare Research and Quality (AHRQ)
* Inter-University Consortium for Political and Social Research (ICPSR) (unless otherwise specified to require IRB Review)
* Better Access to Data for Global Interdisciplinary Research (BADGIR)
* dbGAP: National Center for Biotechnology Information (those datasets specifically identified in the access agreement or data request as not requiring IRB approval)
* National Center for Health Statistics
* National Center for Education Statistics
* National Child Development Study
* National Election Studies
* Roper Center for Public Opinion Research
* University of Wisconsin-Madison Data and Information Services Center (DISC)
* U.S. Bureau of Census
* U.S. Bureau of Labor Statistics
* The University of Michigan Health and Retirement Study (HRS)
* Panel Study of Economic Dynamics (PSID)
* Survey of Consumers (SCA)
* Integrated Public Use Microdata Samples – International (IPUMS-i)
* Luxembourg Income Study Project Archive
* National Health and Nutrition Examination Survey (NHANES)
* National Health Care Survey (NHCS)
* National Immunization Survey (NIS)
* National Health Interview Survey (NHIS)
* Longitudinal Studies of Aging (LSOA)
* National Survey of Family Growth (NSFG)
* State & Local Area Integrated Telephone Survey (SLAITS)
* Vital Statistics: National Vital Statistics System
* U.S. Bureau of Labor Statistics (unless otherwise specified to require IRB Review)

|  |
| --- |
| **Data Protection** |

The applicability of this guidance does not affect any other considerations that researchers must make to ensure that proper physical and/or data security parameters are followed. Researchers must consult with their respective IT departments and/or IT Security to ensure that plans for data handling of electronic data meets all requirements beyond IRB review.

Resources from Purdue Libraries (<http://guides.lib.purdue.edu/DataStorage>) and Secure Purdue (<https://www.purdue.edu/securepurdue/>) may be useful to assist researchers with proper data protections.

|  |
| --- |
| **Identification of New Publically Available Datasets** |

Investigators may apply to the Purdue HRPP to have a dataset registered as a public use dataset under this guidance. The dataset must be a pre-existing, publicly available dataset not yet approved. Datasets that may qualify for inclusion on Purdue HRPP list of approved datasets include:

a) Public use datasets posted on the internet that include a responsible use statement or other confidentiality agreement for authors to protect human subjects

b) Public and/or published datasets, accessible without restriction (e.g., a password is not necessary) and containing readily identifiable information such that individuals can reasonably expect this information to be available to the public (e.g., letters to the editor, web logs or blogs, etc.).

c) Public and/or published datasets, with restrictions to access, that contain data that are presented in aggregate form only (e.g., zip code); thus, individuals cannot be identified.

|  |
| --- |
| **Requesting Identification to IRB for a New Publically Available Dataset** |

Investigators must submit the following information on potentially eligible datasets to the PU IRB/HRPP office prior to conducting research:

* Name of dataset
* URL of the dataset or other information on how to obtain the dataset; and
* Abstract (maximum one page) describing the content and potential use of the dataset.

For a public use dataset to be listed under this guidance by the Purdue HRPP, the dataset must be publicly available to any person through unlimited access or via a member institution or for a fee. The source must confirm that the original data collection occurred either in anonymous form or the original data collection was gathered on identified subjects but the dataset has been stripped of direct identifiers and indirect identifiers that may risk disclosure of subjects’ identity. Also, a formal disclosure analysis by the source must be performed to reasonably affirm that identification of individual subjects using variables within the dataset cannot occur.

Information for these requests should be submitted directly to the Purdue IRB by contacting [irb@purdue.edu](mailto:irb@purdue.edu).