**Dickinson State University Public Data Sets Policy**

# Policy

The DSU Institutional Review Board (IRB) considers the research use of certain publicly available data sets to not involve “human subjects” as defined by federal regulations. The data contained within these specific data sets are neither identifiable nor private and thus do not meet the federal definition of “human subject.” Therefore, DSU does not require these research projects to be reviewed and approved by the IRB.

DSU does not consider controlled access data sets from the NIH dbGaP data repository to be public data sets. Prospective IRB approval or a determination of exempt status is required.

# Definitions

## Public Data Sets are data files prepared by investigators or data suppliers with the intent of making them available for public use.

* The publicly available data are not individually identified or in a readily identifiable form.
* Public data files include data files that have been reviewed under the jurisdiction of an IRB with the intent of making them available for public use. In the case of federal statistical data collections, the federal government has responsibility for that review.
* Data shared informally among colleagues do not constitute a public data set.
* Data suppliers may have (1) publicly available de-identified data as well as (2) restricted use data from the same data set.

## Restricted Use Data Sets are **not** publicly available and are not covered by this policy/procedure.

* Restricted use data are special files distributed by government agencies, research organizations, and others upon which use restrictions are imposed.
* These files often contain data such as Social Security numbers, names, or extensive life history markers that might enable an unauthorized user to identify a participant.
* The use restrictions vary, but they typically involve secure (locked) data storage and password-protected computers, and forbid the storage of data on systems that may be accessed through a computer network connection.
* The use agreements may limit the types of analyses that can be performed.

## Publicly Available means that the data are widely available. Data may still be considered publicly available when:

* 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.

## De-identified means that the recipient of the data cannot readily ascertain the identity of the subject or associate identity with the information, and that the data set does not include one of the 18 HIPAA direct identifiers. If there is a code that links the data to direct identifiers, the code: (1) may not be derived from or related to the information about the individual, and (2) could not be translated to identify the individual.

# Exceptions to this policy/procedure

## The following uses of public data sets are not covered by this policy and require prior IRB approval or a determination of exempt status:

* Merging public data sets in such a way that individuals may be identified.
* Enhancing a public data set with identifiable, or potentially identifiable, data.

## The data host may require the researcher or the researcher’s institution to sign a Data Use Agreement that explicitly requires IRB approval or a determination of exempt status. If so, then the Data Use Agreement prevails over this policy/procedure and over federal guidance about the definition of “human subjects.”

## Researchers whose research project also uses non-public data and/or interacts or intervenes with human subjects must first submit an application for IRB approval or a determination of exempt status for those activities.

# Documentation for research sponsors or funding agencies

## Research sponsors or funding agencies may require documentation that proposed research activities using public data sets do not require IRB approval or a determination of exempt status. To obtain this documentation, the researcher provides the DSU IRB with the following materials so that the IRB can provide the researcher with an appropriate Notice of Determination.

### A cover memo, requesting documentation that the proposed research activities using a public data set does not require IRB approval or a determination of exempt status. The memo should include:

#### The name, department, and contact information of the lead researcher

#### The name of the data set, which must be on the pre-approved list below

#### A description of the research activities (maximum ½ page)

### Funding information, including:

#### The type of support (grant, contract, subcontract, fellowship, gift, etc.)

#### Name of funding agency or sponsor

#### Title of award or contract

#### Name of person listed as principal investigator on the funding proposal

#### Start and end date of the funding

#### Name of the office that processed the funding proposal (example: UW Office of Sponsored Programs)

### Answers to the following questions:

#### Will the research involve merging any of the data sets in such a way that individuals might be identified?

#### Will any data sets be used that are considered restricted?

#### Does the research activity involve adding other data to the data sets?

### A complete copy of the grant application or contract

## The IRB reviews the materials and, if appropriate, provides the researcher with a Notice of Determination.

# Nomination process: Adding another data set to the pre-approved list

## Anyone may nominate a data set for inclusion on the Public Data Sets list by providing the IRB with the following information:

### Name of the data set

### Description of the data set (list of variables, or URL/web address for the list)

### Description of the identifiability of the data (are the data coded? etc.)

### Description of how the data were collected, including: from/about whom (subject populations), by whom, and when

### Description of how the data were obtained by the host (if different from above)

### Name and brief description of the host archive or institution

### Description of how the data set may be obtained by researchers

### Description of any conditions or restrictions about the use of the data

### A copy of any agreement or statement that a researcher or institution must sign in order to gain access to the data (or URL for the agreement)

### Any URLs that are relevant

## DSU researchers who propose to provide public use data files are responsible for having those files appropriately reviewed by the DSU IRB before making the data available to the public. Contact a member of the IRB for information about the appropriate application and procedures.

## The nomination is reviewed by the IRB using the **Public Data Set Nomination Review Checklist**. If a majority of the IRB agrees, the data set is added to the list at the end of this document.

## Documentation.

### The decision is documented in IRB in writing and is stored in the Office of Academic Affairs.

### Written notification of the review outcome is provided by the IRB to the person who nominated the data set.

### The IRB chair creates an office file to document the nomination, review, and outcome. The file contains:

#### The nomination (e.g., print of an email request)

#### The three review checklists

#### Copy of the written decision

#### Any other relevant materials

# List of Data Sets Considered Public

**NOTE**: Research involving only the analysis of data from the following public data sets does not require DSU IRB approval or a determination of exempt status. Researchers may conduct research involving only these data sets without submitting an application or other materials to the DSU IRB.

* National Center for Health Statistics
	+ - Life Tables
		- LSOAs: Longitudinal Studies of Aging
		- NHANES: National Health and Nutrition Examination Survey
		- NHCS: National Health Care Survey
		- NHIS: National Health Interview Survey
		- NIS: National Immunization Survey
		- NSFG: National Survey of Family Growth
		- SLAITS: State & Local Area Integrated Telephone Survey
		- Vital Statistics: National Vital Statistics System
* National Center for Education Statistics
* U.S. Bureau of the Census
* U.S. Bureau of Labor Statistics

**Regulatory and other references**

45 CFR 46.102(f)

32 CFR 219.102(f)

21 CFR 50.3(g)

OHRP Guidance on Research Involving Coded Private Information or Biological Specimens; August 10, 2004

National Human Subjects Protection Advisory Committee (NHRPAC), “Recommendations on Public Use Data Files”; January 28-29, 2002 meeting

**This policy adopted with permission from the policy of the University of Washington.**