Procedure Title: Public Data Sets

Associated Policy:	Human Research Protection Policy (OSA Policy 1.0)		
Responsible Unit:	Office of Scholarly Activity		
Created:	11/05/2018	**Executive Lead:**	Chief Research Officer
Effective:	12/02/2019	**Revision History:**	00 Original SOP; .01 – 04-07-2020; .02 – 2-21-2023; .03 – 08-07-2024
Approved by:	Institutional Review Board		
Procedure Number:	132.03		
Key Words:	data sets, public data sets, de-identified data sets, restricted data set, restricted use data sets, data management agreement, data use agreement		
Purpose:	To meet the responsibilities for protecting human subjects as issued by the Office for Human Research Protections (OHRP) requirement for individuals involved in the conduct or review of human subjects research at institutions holding OHRP-approved Federalwide Assurances (FWAs)		

Process:
This SOP serves to inform all agents, offices, departments, and affiliate sites of PNWU regarding how the Institutional Review Board (IRB) at Pacific Northwest University of Health Sciences (PNWU) defines a public data set and when IRB approval is necessary prior to research activities using the data set.

This SOP must be used as a guide in parallel with OSA Policy 1.0. SOPs are not intended to supersede existing institutional policies, and local, state, and federal laws and regulations.

General Information:
Public data sets are collections of data that have been made publicly available. Usually, the sets consist of medical and/or research data prepared with the intent of public use that have been de-identified and/or are available in a summary or composite format. Projects that utilize public data sets recognized by PNWU (see the list below) typically do not require an IRB Application.

Restricted-use data sets are collections of data that require the requester to obtain permissions from the owner of the data. These data sets typically include a limited HIPAA data set or other information that may compromise the confidentiality of the data. Restricted-use data sets often require a data use/data management agreement. The IRB does not consider restricted-use data sets to be publicly available. Users must not obtain access to restricted use data sets prior to IRB Approval.

Responsible Parties

The Institutional Review Board (IRB) is responsible for:

- Protecting the rights and welfare of human research participants.
- Acting with impartiality when conducting reviews of human subjects research.
- Helping determine when a project meets the definition of human subjects research.

- Remaining immune from pressure by the institution's administration, the investigators whose protocol are brought before it, or other profession and non-professional sources.

The Office of Scholarly Activity (OSA) is responsible for:

- Supporting investigators in helping determine activities subject to human subjects protection.
- Overseeing and providing the necessary support to the IRB.
- Monitoring compliance with this SOP.
- Posting this SOP for the PNWU community.
- Providing the necessary administrative support to investigators and the IRB.

The Principal Investigator is responsible for:

- Asking the OSA if their activity is subject to human research protections.
- Being a steward of a research environment that promotes the responsible conduct of research.
- Ensuring compliance with the IRB-approved protocol, federal regulations, state laws, good clinical practice, and applicable FDA guidance.
- Seeking support from OSA and the IRB when questions arise.

Definitions

Please reference the Glossary for complete definitions of the following terms and additional terms not listed.

- Data Use/Management Agreement (DUA)
- Federalwide Assurance
- Health Insurance Portability and Accountability Act (HIPAA)
- Human Subject
- Principal Investigator
- Protected Health Information (PHI)
- Personally Identifiable Information (PII)
- Publicly Available
- Public Data Sets
- Quality Improvement
- Research
- Restricted-Use Data Sets

Investigator Procedures:

1. Please review the PNWU SOP 113 – Data Management
2. Please review the List of Publicly Available Data sets (below) that PNWU considers to be public. Researchers may utilize these public data sets without submitting an IRB application as long as the research activity is the use of only **one** of these data sets. Research that combines one or more public data sets must be submitted for IRB review and approval.
3. If a data set is not featured on this list, the investigator must submit an IRB application via the electronic IRB submission system.
4. Investigators may propose that a data set be added to this list. To do so please send an email to research@pnwu.edu and include the following information.
    a. Documentation that the data set does NOT directly or indirectly identify subjects.
    b. Information on the institution and the data being requested (e.g., zip code, state, demographics).
    c. Information on access to the data set (e.g., does the user need to fill out a data use agreement or complete training before being granted access?).
5. An IRB application must be submitted when:

a. The data set is not listed on the list below as a public data set.
b. The data set is listed as a public data set, but a third party requires documentation from the IRB confirming that the use of this data does not constitute human subjects research.
c. The data set is restricted meaning there are requirements set by the owner of the research data such as training, data use or data management agreements, or non-disclosure agreements. If a data use or data management agreement is required, you must contact the Office of Scholarly Activity as only the Institutional Official may sign a data use/data management agreement.
d. Data from multiple public data sets are merged, as this may alter the nature of one or both sets (e.g., two de-identified data sets could be merged to create an identifiable data set).
e. The investigator seeks to enhance a public data set with identifiable or potentially identifiable data.
f. Data sets which are not publicly available are used alongside publicly available data sets.
g. Activities interacting or intervening with human subjects are proposed in addition to the use of publicly available data sets.
h. Using controlled-access data sets such as the NIH dbGaP data repository.
i. The investigator plans to create a public data set by preparing files with the intent of making them available for public use.

6. When the data supplier requires the researcher or researcher's institution to sign a Data Use Agreement. or explicitly requires IRB approval or a determination of exempt status please follow these steps:
    a. Complete and submit an IRB application in the IRB electronic submission system..
    b. Upload an unexecuted copy of the Data Use Agreement with the application.
    c. Data Use Agreement will be signed by the PNWU Institutional Official, once IRB approval is granted, if applicable.

Sample List of Publically Available Data sets*:

American College of Surgeons National Trauma Data Bank (NTDB)
Behavioral Risk Factor Surveillance System (BRFSS; *public data only*)
CDC Wide-ranging Online Data for Epidemiologic Research (WONDER)
Comprehensive Hospital Abstract Reporting System (CHARS) - *public use data only*
Coordination of Benefits (COB) [Centers for Medicare & Medicaid Services] – *public use data only*
Fatality Analysis Reporting System (FARS)
Global Burden of Disease (GBD)
Healthcare Cost and Utilization Project (H-CUP) healthcare databases – *public use data only*
Home Mortgage Disclosure Act (HMDA)
Hospital Compare Datasets
Household Component Event files [Agency for Healthcare Research and Quality]
Household Component Full-Year files [Agency for Healthcare Research and Quality])
Household Component Point-in-time files [Agency for Healthcare Research and Quality]
Integrated Public Use Microdata Series International (IPUMS)
Jaeb Center for Health Research Foundation
Life Tables, United States Social Security Area
LSOAs: Longitudinal Studies of Aging [Centers for Disease Control and Prevention]
Medical Expenditure Panel Survey (MEPS) (Agency for Healthcare Research and Quality)
Medical Information Mart for Intensive Care (MIMIC)-IV
Medicare Database [Data.Medicare.gov]
National Automotive Sampling System (NASS) General Estimates System (GES)
National Center for Education Statistics (NCES]
National Center for Health Statistics [Center for Disease Control and Prevention)

National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) [National Institute on Alcohol Abuse and Alcoholism Data Archive]
National Household Travel Survey [US Department of Transportation Federal Highway Administration]
NHANES: National Health and Nutrition Examination Survey (Only Datasets from 1999-2018)
NHCS: National Health Care Survey *(public use data only)*
NHIS: National Health Interview Survey [National Center for Health Statistics]
NIS: National Immunization Survey [Center for Disease Control and Prevention]
NSFG: National Survey of Family Growth 2015-2017 (*Public Use Data Only)*
Organ Procurement and Transplantation Network (OPTN) [US Department of Health and Human Services] *(Public Use Data Only)*
Pooled Linkage files [Agency for Healthcare Research and Quality]
School Health Policies and Programs Study (SHPPS) [Center for Disease Control and Prevention] (*Public Use Data Only)*
SLAITS: State & Local Area Integrated Telephone Survey [Center for Disease Control and Prevention]
Survey of Consumer Finances (SCF) [Federal Reserve Board]
U.S. ATF Firearms Trace Data [Bureau of Alcohol, Tobacco, Firearms and Explosives]
U.S. Bureau of the Census
Uniform Crime Reporting Statistics [US Department of Justice, Federal Bureau of Investigation]
Vital Statistics: National Vital Statistics System [Centers for Disease Control and Prevention] – *Public Use data only*
Web-based Injury Statistics Query and Reporting System (WISQARS) [Centers for Disease Control and Prevention]
World Health Organization Global Health Observatory Data Repository (WHO GHO Data Repository)
Youth Risk Behavior Surveillance System (YRBSS) [Centers for Disease Control and Prevention]

**Note: Some of the Public Data Sets listed above have both public data sets and other data sets which may require IRB Approval.*

References:

1. Food and Drug Administration (FDA) Federal Regulations (21 CFR 50, 54, 56, 312, 314, 600, 601, 812 and 814) https://www.ecfr.gov/cgi-bin/text-idx?SID=3ee286332416f26a91d9e6d786a604ab&mc=true&tpl=/ecfrbrowse/Title21/21tab_02.tpl
2. Department of Health and Human Services (DHHS) Regulations (45 CFR 46 Subparts A, B, C, and D) http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html
3. International Conference on Harmonization (ICH) Good Clinical Practice (GCP) guidelines https://www.fda.gov/scienceresearch/specialtopics/runningclinicaltrials/guidancesinformationsheetsandnotices/ucm219488.htm
4. PNWU SOP 113: Data Management and Disposition

Revision History:

Version/ Effective Date	Author	Section Changed & Reason for Revision
.00 /12-2-2019	M. McCarroll	Original SOP
.01 / 04-28-2020	C. Case	Added additional key words page 1; Added the General Information Section on page 1; Added items from the list of standard responsibilities for the

		IRB, OSA and PI; Added data use/data management agreement and personally identifiable information to the definitions list; Renamed procedure section to Investigator Procedures; reordered number items in the investigator procedures section and revised the information about proposing a new public data set in item 4. Added bulleted item to investigator procedures (#5) regarding data use/data management agreements and the IO signature of these agreements. Added a reference to SOP 113 Data Management and Disposition in the reference section at the end.
.02 / 3-3-2023	C. Case	Corrections of typographical error; removed IRB Manager
.03 / 8-7-2024	J. Simmons	Added the MIMIC-IV public dataset

Appendices:
None