National Institute of Environmental Health Sciences
Your Environment. Your Health.

Standard Operating Procedure (SOP) for Protection of Personally Identifiable Information (PII), Protected Health Information (PHI) and Genomic Data in Epidemiology Branch supported studies.

Summary

This document defines the National Institutes of Health (NIH) National Institute of Environmental Health Sciences (NIEHS) Epidemiology Branch (EB) standard operating procedures to protect research participant Personally Identifiable Information (PII), Protected Health Information (PHI) and controlled-access human genomic and phenotypic data from loss of control, compromise, and/or unauthorized use. This SOP applies to all PII, PHI and controlled-access genomic and phenotypic data (defined below) in any form whether written, spoken, recorded electronically, or printed:

Personally Identifiable Information (PII): Any information about a human subject including, but not limited to, medical, criminal or employment history, and information which can be used to distinguish or trace an individual's identity, such as their name, Social Security Number (SSN), address, date and place of birth, mother's maiden name, biometric records, etc., including any other personal information which is linked or linkable to an individual. PII, which if lost, compromised, or disclosed without authorization, could result in substantial embarrassment, inconvenience, or unfairness to an individual. SSN, driver's license number, citizenship, or medical information, in conjunction with the identity of an individual are considered Sensitive PII and require stricter handling controls.

Protected Health Information (PHI): PHI is health information, including demographic information which relates to the past, present, or future physical or mental health or condition of an individual, or the provision of health care to an individual.

Controlled-Access Data: For this purpose, human genomic and phenotypic data defined by NIH and addressed in the this purpose, controlled-access data is human genomic and phenotypic data defined by NIH and addressed in the NIH Genomic Data Sharing (GDS) Policy (https://sharing.nih.gov/) as large-scale human data from genome-wide association studies (GWAS), single nucleotide polymorphisms (SNP) arrays, and genome sequence, transcriptomic, metagenomic, epigenomic, and gene expression data.

Data Collection and Storage

EB researchers (investigators, trainees, guests and special volunteers) conduct research through secondary data analysis of external datasets or primary data and samples collected and controlled by government Contractors. Contractors are required to deliver annual Information System Security packages and to follow information security standards for all electronic data access, handling, transmission, and storage, as defined in the Federal Information Security Management Act (FISMA; https://olao.od.nih.gov/content/what-fisma) and NIST 800-53 Rev5 (Security and Privacy Controls for Federal Information Systems and Organizations; https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r5.pdf). Contractors securely store original hard-copy study materials, abstracted and redacted medical records, and biological and environmental samples according to the requirements specified in their prime contracts with the government.

Security Training

All EB researchers, staff and contractors complete all required NIH research, human subjects, ethics and security training (https://junction.niehs.nih.gov/training/index.htm -NIH internal link only) and follow NIH Information Security Policies and Standards (https://ocio.nih.gov/ITGovPolicy/Pages/spec_policy.aspx).

Data Requests and Access

Requests for data must be submitted formally through the appropriate portal[1] or channel following these steps:

1. The researcher submits a research proposal/concept (with hypothesis) for review and approval.
2. Once the proposal/concept is approved, the researcher completes a data request and submits via the appropriate portal or channel for review and approval by the Principal Investigator (PI) or governing body (e.g., Executive Committee). Access via the portal is secure and adheres to all encryption and security protocols. For researchers/collaborators from external institutions, IRB approvals are confirmed and data transfer agreements are obtained before a request is approved.
3. All users of data are required to complete and return a **DATA SHARING AGREEMENT (DSA)** acknowledging receipt of instructions for data security, permitted use, and disposition of data.
4. The Contractor makes the requested dataset available to the researcher only through a secure method (e.g., secure shared drive, Secure File Transfer Protocol (SFTP), or other encrypted format). Passwords are provided separately (by telephone, separate email, or in person).
5. Researchers are required to store datasets on a secure Network drive, a government furnished equipment (GFE) computer (encrypted and with 2-factor authentication), or encrypted on a non-GFE computer.
6. Some large genetic and phenotypic datasets with coded IDs are maintained on a server supported and monitored by the NIEHS Office of Scientific Computing (OSC). OSC grants access to the server upon approval by the study PI. **It is the responsibility of the study PI** to obtain a completed **DATA SHARING AGREEMENT** for any additional users granted access to genetic and phenotypic data on the OSC server(s). Servers can only be accessed via GFE, either onsite or through VPN.

Data Use

1. Data users agree not to distribute, duplicate, or use data, in whole or in part, other than as permitted by this Data Sharing Agreement, or as otherwise required by law or regulation.
2. Recipient shall not link the data with other datasets unless explicit approval, in writing, has been given by an authorized study representative and only as required to address a specific approved study aim.
3. Secondary uses of the data provided (e.g., for a related publication or for publication when data were shared for validation purposes only) require a new concept/data request submission and approval from the study PI.
4. All study participant data made available to approved users are de-identified with only a uniquely coded ID. The link between the coded ID and participant name is kept by the Contractor and is never available to users but is maintained in the event of questions about the data, requests for additional variables, or return of processed data to the study. The Contractor securely maintains contact information for all participants. Users (researchers) do not have access to participant contact information. Recipient shall make no attempt to identify or contact individuals to whom the data pertains. If warranted, fully anonymized datasets may be provided.
5. Datasets are de-identified and/or anonymized as much as possible; users receive only the data necessary for their approved analyses. In situations where it is necessary to grant access to data that contain some identifying information (e.g., addresses required for geocoding), other non-essential identifying data are removed from the dataset to limit the risk of identification of a participant.
6. An EB PI or designee may grant access to PHI and PII to researchers who request it on a “need to know” basis.
7. Study data (hard copy or digital) must never be left unsecured, and steps taken to secure PII/PHI should be documented. Study data must be disposed of when no longer required, consistent with the applicable study records disposition protocols. All data must be disposed of as described below (Section 9.), within 3 years of the signing of this agreement, except with prior explicit approval in writing of an authorized study representative.
8. Upon conclusion of the research project, the user shall transfer the created code, working data files, variables and manuscript(s) to the secure drive via SFTP for records retention.

9. Upon the earlier conclusion of the research project or expiration of this DSA (3 years), users are required subsequently to delete all data and sanitize the hard disk drive to ensure secure deletion of all files. Before a computer, personal electronic device, computer drive, or other electronic device is transferred to another person or disposed of, the device must be stripped of any study data that may have been stored within.
10. **If a computer/device is lost or compromised, users must report it immediately** to the NIH IT Service Desk (866-319-4357), the NIEHS ISSO (984-287-3032, or jordanm2@niehs.nih.gov), the user's supervisor and the NIEHS contractor (DLH_NIEHS_Data@dlhcorp.com).
11. If using your own equipment and data are lost or compromised, users must report it immediately to NIEHS contractor (DLH_NIEHS_Data@dlhcorp.com).

[1] Links to Portals: Agricultural Health Study, EpiShare, GuLF Study and Sister Study

National Institute of Environmental Health Sciences
Your Environment. Your Health.

NIH/NIEHS Epidemiology Branch

DATA SHARING AGREEMENT

It is required that collaborators understand and sign the following pledge of confidentiality prior to receiving data from any NIH/NIEHS Epidemiology Branch research project.

I hereby certify that I will keep completely confidential all Epidemiology Branch study information, to which I gain access, concerning individual respondents. I also certify that I will abide by all requirements of the NIH Intramural Institutional Review Board (IRB) and other applicable IRBs.

Beyond the research team, I will not discuss, disclose, disseminate, or provide access to study data and/or identifiers except as authorized in writing by the NIH/NIEHS Epidemiology Branch study Principal or Lead Investigator. I shall use the study data only for approved purposes and shall follow instructions for disposition. I am also aware that I am responsible for the compliance of all other personnel under my supervision who have access to the data provided to me by the research study team. I will maintain any transferred information in a secure manner that restricts access by any individual not involved in the research project (e.g., for paper records—locked file cabinets or continual physical presence in a room that locks; for electronic records—encrypted with password protection). I agree to report any breaches in confidentiality to the NIH/NIEHS Epidemiology Branch study Principal or Lead Investigator within 24 hours of discovery.

I give my personal pledge that I shall abide by this assurance of confidentiality.

REQUESTOR NAME (SIGNATURE): _____

REQUESTOR NAME (PRINT): _____

DATE: _____

PROJECT TITLE: _____

SUPERVISOR NAME (FOR STUDENTS/POSTDOCS ONLY): _____