

US Department of Health and Human Services

Privacy Impact Assessment

Date Signed:

12/09/2025

OPDIV:

NIH

Name:

NIMHD SCHARE Data Repository (SDR)

PIA Unique Identifier:

P-5864048-763858

The subject of this PIA is which of the following?

Minor Application (stand-alone)

Identify the Enterprise Performance Lifecycle Phase of the system.

Operations and Maintenance

Is this a FISMA-Reportable system?

No

Does the system include a Website or online application available to and for the use of the general public?

Yes

Identify the operator.

Agency

Is this a new or existing system?

Existing

Does the system have Security Authorization (SA)?

Yes

Indicate the following reason(s) for updating this PIA.

PIA Validation

Describe in further detail any changes to the system that have occurred since the last PIA.

This validation is intended to refresh content and account for personally identifiable information captured in the Data Use Oversight System (DUOS).

Describe the purpose of the system.

The Science Collaborative for Health Disparities and Artificial Intelligence Reduction of Errors (SCHARE) Data Repository (SDR) is a cloud-based data repository for intra- and extramural National Institute on Minority Health and Health Disparities (NIMHD) researchers. The SDR will allow researchers to host, manage, and share data from NIMHD- and National Institute of Nursing Research (NINR)-funded research programs, including data on health disparities and health care outcomes. NIH-based researchers, external researchers, and health care organizations will also be able to access data within the repository.

Requests to access the data on SDR will be submitted and approved in the Data Use Oversight

System (DUOS), a subset of the SDR and operated by Broad Institute.

Describe the type of information the system will collect, maintain (store), or share.

The primary purpose of the SDR is to store data collected from NIMHD-funded studies. There are approximately 130 funded programs with data that will be put into the SDR. Study data includes demographic characteristics and health outcomes data of study participants. To support this functionality, the SDR will store information about the users of the system.

The SDR is designed around preparing data for artificial intelligence (AI) use, but there is no AI at use in the system itself. The SCHARE Platform is focused on providing data ready for machine learning analyses and the resources to perform those analyses, thus the mention of AI in the title of the platform. The Privacy Impact Assessment (PIA) will be updated to reflect any future AI use cases that introduce new privacy risks.

The DUOS subcomponent is used to register studies with controlled-access data, receive data access requests, and allocate permissions to approved users. To support these functions, DUOS will store metadata about the studies with controlled-access data, information contained in the data access requests, and contact information for the SDR Data Access Committee. NIMHD and NIH researchers will use the SDR to host, manage, and share data from NIMHD- and NINR- funded research programs, including data on health disparities and health care outcomes. There are approximately 130 current set of NIMHD and NINR-funded programs in health outcomes research are:

The SDR will collect/maintain/share the following personally identifiable information (PII) in the uploaded datasets. The PII collected will follow the SCHARE Core Common Data Elements (CDEs) set:

Age at enrollment, Birthplace,
Zone Improvement Plan (ZIP) code,
Demographics (ethnicity, sex, sexual orientation, marital status, education level, annual household income, household size, disability status,)
English proficiency, health insurance status, employment status, Usual place of health care (e.g. doctor's office, walk-in clinic, etc.),
Social needs and economic security,
Health status (social determinants of health, self-reported health status, health conditions, and self-reported social stressors).

(CDEs are defined fields describing the data to be collected (e.g., identifying specific variables) along with how to gather the data, and how the response is represented in a dataset. CDEs are structured as indivisible units of data. This can be either an individual field (e.g., sex) or multiple fields taken together (e.g., the composite score of a scale)). The system will contain additional information specific to each study that fall under the categories of health information and mental health information.

Those requiring access to this system log in using the NIH Identity, Credential, and Access Management (IAM) Services which maintains its own unique privacy impact assessment (PIA) on record, with all legal authorities documented. The purpose of IAM Services is to authenticate and authorize all users and computers in a Windows domain type network assigning and enforcing information security policies for all computers and installing or updating software. The IAM Services collects unique user credentials and stores them in an encrypted format. The IAM Service is an essential service which facilitates and governs network access to various resources. The DUOS system will collect the following data submitter information: Name, Email and Custodian Email. DUOS will collect the name(s) of the Researcher, internal lab staff, internal collaborator, Principal Investigator (PI), Internal Collaborator, Institutional Signing Official, Institutional Information

Technology, and External Collaborator. In addition, the eRA Commons identification (ID) is collected

Provide an overview of the system and describe the information it will collect, maintain (store), or share, either permanently or temporarily.

The SDR is a cloud-based data repository for intra- and extramural NIMHD researchers. The SDR will allow researchers to host, manage, and share data from NIMHD- and NINR-funded research programs, including data on health disparities and health care outcomes. NIH-based researchers, external researchers, and health care organizations will also be able to access data within the repository. Requests to access data on the SDR will be submitted and approved in the DUOS, a subset of the SDR SDR and operated by Broad Institute.

The SDR is designed around preparing data for AI use, but there is no AI at use in the system itself. The SCHARE Platform is focused on providing data ready for machine learning analyses and the resources to perform those analyses, thus the mention of AI in the title of the platform. The PIA will be updated to reflect any future AI use cases that introduce new privacy risks.

There are approximately 130 NIMHD and NINR-funded programs in health outcomes research. The SDR will collect/maintain/share the following personally PII in the uploaded datasets. The PII collected will follow the SCHARE CDEs set:

Age at enrollment, Birthplace,

ZIP code, Demographics

English proficiency, health insurance status, employment status, Usual place of health care Social needs and economic security,

Health status

The system will contain additional information specific to each study that falls under the categories of health information and mental health information.

Those requiring access to this system log in using the NIH IAM Services which maintains its own unique PIA on record, with all legal authorities documented.

The DUOS system will collect the following data submitter information: Name, Email and Custodian Email.

DUOS will collect the name of the Researcher, internal lab staff, internal collaborator, Principal Investigator, Internal Collaborator, Institutional Signing Official, Institutional Information Technology Director, and External Collaborator. In addition, the eRA Commons ID is collected.

Does the system collect, maintain, use or share PII?

Yes

Indicate the type of PII that the system will collect or maintain.

Name

E-Mail Address

CDE set (Age at enrollment, Birthplace,

Zone Improvement Plan (ZIP) code, Demographics

English proficiency, health insurance status, employment status, Usual place of health care Social needs and economic security, Health status)

eRA Commons ID

Indicate the categories of individuals about whom PII is collected, maintained or shared.

Employees
Public Citizens
Patients

How many individuals' PII is in the system?

10,000-49,999

For what primary purpose is the PII used?

The PII is used for research purposes to study the nature and impact of social determinants of health on health outcomes.

The PII in DUOS is used to properly assign access permissions to the correct users.

Describe the secondary uses for which the PII will be used.

NA

Identify legal authorities governing information use and disclosure specific to the system and program.

5 U.S.C. 301 and 302, 44 U.S.C. 3101 and 3102,
42 U.S.C. 241, 242, 248, 281, 282, 284, 285a, 285b, 285c, 285d, 285e, 285f, 285g, 285h, 285i, 285j,
285l, 285m, 285n, 285o, 285p, 285q, 287, 287b, 287c, 289a, 289c

Are records on the system retrieved by one or more PII data elements?

Yes

Identify the number and title of the Privacy Act System of Records Notice (SORN) that is being use to cover the system or identify if a SORN is being developed.

09-25-0200 Clinical, Basic and Population-based Research Studies of the National Institutes of

Identify the sources of PII in the system.

Directly from an individual about whom the information pertains

In-Person
Government Sources

Identify the OMB information collection approval number and expiration date

Office of Management and Budget (OMB) approval is not required for the SDR, however, each Non-Governmental Source collects data from the public has to receive OMB/Paperwork Reduction Act (PRA) approval in order to do so.

OMB approval is required for the DUOS data registration form and the Data Access Request form and is in process.

Is the PII shared with other organizations?

Yes

Identify with whom the PII is shared or disclosed and for what purpose.

Describe any agreements in place that authorizes the information sharing or disclosure.

PII is shared with the Broad Institute for the functioning of the DUOS system. This sharing is governed by the Broad Institute's DUOS Membership Agreement.

Describe the procedures for accounting for disclosures.

Administrators can provide logs of all data submissions/data access requests for the SDR upon request.

Describe the process in place to notify individuals that their personal information will be collected. If no prior notice is given, explain the reason.

The data that will be uploaded to the SDR is collected by studies performed under Intramural Review Board (IRB) approval, and thus the notification of data sharing is covered in the IRB researcher participant agreement.

Is the submission of PII by individuals voluntary or mandatory?

Voluntary

Describe the method for individuals to opt-out of the collection or use of their PII. If there is no option to object to the information collection, provide a reason.

Individual opt-out of collection is done at the level of agreeing to participate in the studies whose data ultimately will be uploaded to the SDR.

For DUOS, the PII collected is needed for dataset registration and Data Access Request (DAR) review. If they don't provide the data, they can't register.

Process to notify and obtain consent from individuals whose PII is in the system when major changes occur to the system.

Each study whose data will be stored on the SDR, is responsible for receiving OMB/PRA approval. They cannot collect PII from the public without doing so.

DUOS will notify all registered users of major changes to how their information is gathered and displayed.

Describe the process in place to resolve an individual's concerns when they believe their PII has been inappropriately obtained, used, or disclosed, or that the PII is inaccurate.

Individuals can contact the Investigator who can then forward the request to the SCHARE team. The individual should identify the record and specify the information being contested, the corrective action sought, and the reasons for requesting the correction, along with supporting information to show how the record is inaccurate or incomplete.

In the case of concerns of PII disclosure on DUOS, the individual should contact the SDR Data Access Committee (DAC), who will work with the DUOS development team to remediate the concern.

Describe the process in place for periodic reviews of PII contained in the system to ensure the data's integrity, availability, accuracy and relevancy.

The SDR is not a source system and PII is not collected directly from individuals. Source system operators are responsible for putting a process in place to resolve an individual's concerns about the collection and use their PII. However, individuals may contact any Institute/Center/Office (ICO) Privacy office, or the NIH Senior Official for Privacy.

The SDR will conduct regular security and health-checks, data will be securely backed up, and security vulnerabilities will be addressed and remediated.

The SDR DAC will conduct annual audits of the PII stored on DUOS to ensure data integrity,

availability, and accuracy.

Identify who will have access to the PII in the system and the reason why they require access.

Describe the procedures in place to determine which system users (administrators, developers, contractors, etc.) may access PII.

Access to PII will be on a need-to-know basis and will be managed by the platform's permission-based system. System users are either federal employees or direct contractors who undergo the annual NIH security and privacy training.

External researchers accessing PII must sign a controlled-access data use agreement and are authenticated through the NIH IAM System.

For DUOS, only developers and approved members of the SDR DAC may view PII. Members can only be added by current members, and members must agree to DAC Terms and Conditions prior to approval.

Describe the methods in place to allow those with access to PII to only access the minimum amount of information necessary to perform their job.

Determinations are made based on role-based access controls and least privilege use. User rights are provisioned based on controls within the system supported by data use agreements, making sure that users only have access to the minimum amount of PII necessary to conduct their research. Researchers are only granted access to PII of the studies specifically requested in their Data Access Request and upon approval of said request.

For DUOS, minimal PII is requested from data submitters and data requesters in order to properly assess Data Access Requests.

Identify training and awareness provided to personnel (system owners, managers, operators, contractors and/or program managers) using the system to make them aware of their responsibilities for protecting the information being collected and maintained.

According to NIH policy, all personnel who manage or operate NIH applications must successfully complete annual Information Security and Management training. Training is completed on the <https://irtsectraining.nih.gov> with valid NIH/HHS credentials.

Developers of the SDR will complete the annual Application Security Checklist & Best Practices for Developers training.

Executives and managers, which includes the SDR program team, will complete the annual role-based training for Executives & Managers.

Describe training system users receive (above and beyond general security and privacy awareness training).

Administrators and developers will receive additional, domain specific training applicable to their role and responsibilities.

Do contracts include Federal Acquisition Regulation and other appropriate clauses ensuring adherence to privacy provisions and practices?

Yes

Describe the process and guidelines in place with regard to the retention and destruction of PII.

Records are retained and disposed of under the authority of the NIH Records Retention Schedule.

Item Number 01-003: Records of All Other Intramural Research Projects. Disposition instruction is Cut off annually at termination of project/program or when no longer needed for scientific reference, whichever is longer. Destroy 7 years after cutoff under Disposition Authority DAA-0443-2012-0007-0003.

Item 07-203 System access records. Systems not requiring special accountability for access. These records are created as part of the user identification and authorization process to gain access to systems. Records are used to monitor inappropriate systems access by users.

Includes records such as:

- user profiles
- log-in files
- password files
- audit trail files and extracts
- system usage files
- cost-back files used to assess charges for system use

Exclusion 1. Excludes records relating to electronic signatures.

Exclusion 2. Does not include monitoring for agency mission activities such as law enforcement.

Systems not requiring special accountability for access.

These are user identification records generated according to preset requirements, typically system generated. A system may, for example, prompt users for new passwords every 90 days for all users.

Disposition: Destroy when business use ceases. DAA-GRS-2013-0006-0003

Describe, briefly but with specificity, how the PII will be secured in the system using administrative, technical, and physical controls.

Physical: The SDR is deployed in the Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability (STRIDES) enterprise environment within the Google Cloud Platform (GCP) and inherits all of STRIDES' security and privacy controls. Physical access controls are implemented by the STRIDES team.

Administrative controls include system security and contingency plans. Data on the system is backed up regularly using GCP's Multi-region backup. Physical controls for access to the servers storing the data are administered by Google Cloud at the individual Google Data Centers. All technical personnel who access SDR information technology (IT) systems have met background investigation criteria for Public Trust positions. All personnel have taken mandatory information security and privacy training classes and annual refreshers. System administrator access to the platform will be through RAS and through STRIDES based GCP access control mechanisms leveraging second factor authentication.

Technical controls include role-based access controls. Only researchers submitting their data to the SDR and the Data Access Committee will have access to datasets with PII. Only the Data Access Committee will see the PII contained in Data Submission and Data Request forms on DUOS.

Identify the publicly-available URL:

Schare-repository.nimhd.nih.gov

Duos.org

Note: web address is a hyperlink.

Does the website have a posted privacy notice?

Yes

Is the privacy policy available in a machine-readable format?

Yes

Does the website use web measurement and customization technology?

Yes

Select the type of website measurement and customization technologies is in use and if it is used to collect PII.

Does the website have any information or pages directed at children under the age of thirteen?

No

Does the website contain links to non- federal government websites external to HHS?

Yes

Is a disclaimer notice provided to users that follow external links to websites not owned or operated by HHS?

Yes