NIDDK Central Repository (NIDDK-CR) – Resources for Research (R4R)

Data Preservation and Access Practices

NIDDK Central Repository (NIDDK-CR) Resources for Research (R4R) facilitates sharing of resources generated from studies supported by NIDDK (or materials generated by use of NIDDK-CR provided resources) within NIDDK's mission by making these resources available for secondary research. Transfer of resources to and from the NIDDK-CR is governed by NIH Data Management and Sharing (DMS) Policy, NIDDK DMS Guidance, and NIDDK-CR Resource Archival and Sharing Policy, in accordance with established NIDDK-CR practices and procedures. These resources are held under the guardianship of NIDDK-CR and the custodianship of NIDDK for a set portion of the resource life cycle per NIDDK-CR policy. NIDDK Central Repository is committed to ensuring long-term preservation and access to these resources for the broader scientific and research community.

This document describes the NIDDK-CR data preservation and access practices for submitting and accessing NIDDK-CR held resources, and the subsequent archiving of these resources to ensure long-term preservation.

I. Submission Requirements

Investigators submitting study-generated resources or specimen linking files to NIDDK-CR must comply with the specific procedures described below.

1. Data Redaction

NIDDK-CR has chosen to receive data using the Limited Data Set (LDS) method of partial de-identification to maximize the utility of the data for research purposes. LDS data must exclude all direct personal identifiers and can include indirect identifiers subject to a "minimum necessary" standard. Data submitted to the NIDDK-CR must have direct personal identifiers removed per the following criteria:

- In accordance with guidance provided for Limited Data Sets and Data Use Agreements section of the NIH HIPAA Privacy Rule summary.
- Investigators should adhere to the NIDDK Central Repository Data and Documentation Submission Guidelines when preparing study data, images, documentation, or specimen linking files.

A review will be done by NIDDK-CR Support Staff to ensure that direct personal identifiers are removed. Additional redactions will be performed as necessary to produce a data package to be shared with approved requestors through a controlled access request process.

2. Material Transfer Agreement

NIDDK-CR acts as an honest broker for sharing resources with qualified investigators based on a contributing study's sharing plan or the informed consent language relative to sharing the resources. A Material Transfer Agreement (MTA) will be needed if a sharing plan or informed consent form (ICF) language is not available, or the permitted research purpose or authorized terms of use related to future use in the ICF is insufficient or unable to be determined, or the contributing study requires a MTA to be executed before transferring resources to NIDDK-CR.

The MTA assures that:

- The resources are being provided for the purpose of the NIDDK-CR distributing these to approved requestors. The contributing study grants the NIDDK-CR explicit permission to distribute the resources to requestors as a research resource without intellectual property assertions.
- The contributing study certifies that the resources were collected according to 45 CFR Part 46, "Protection of Human Subjects" at all the study sites with the understanding that the resources are not to be used in human subjects or for the treatment or diagnosis of human subjects.
- The contributing study certifies that an Institutional Review Board (IRB) or equivalent has reviewed and verified that the submission of resources to the NIDDK-CR for subsequent sharing for research purposes is consistent with the informed consent of study participants from whom the resources were obtained.
- The contributing study and NIDDK-CR acknowledge that some resources may be limited in quantity and that their distribution for secondary research purposes will be based on the scientific merit of a proposed research project. Scientific merit of all requests for NIDDK-CR-held resources will be determined by NIDDK-CR. The contributing study acknowledges that the NIDDK-CR will periodically assess the ongoing scientific utility of resources. Options for resource management or deaccessioning will be determined for resources found to be of low utilization after an established amount of time per NIDDK-CR policy and established practices.
- The contributing study agrees to provide NIDDK-CR a clear statement identifying all permitted use and authorized terms for the distribution and future use of resources specified in the study participants' informed consent documents or indicate that there are no known restrictions or limitations for future use. NIDDK-CR agrees to provide notice to requestors.
- The contributor may request NIDDK-CR-held resources using the same procedures as other requestors.
- To respect the privacy of the human subjects, NIDDK-CR agrees not to contact or make any effort to identify individuals, families, communities, tribes, or populations that are or may be the source of the resources. In addition, NIDDK-CR Data and Resources Use Agreement (DUA) stipulates that the requestor must abide by the same.
- The contributing study is submitting the resources to the NIDDK-CR as a service to the research community. The contributing study makes no representation or warranties, express or implied, to NIDDK-CR with respect to the accuracy or completeness of the resources. However, the contributing study will use reasonable efforts to ensure that the resources are complete and accurate in all respects. Contributors will provide reasonable cooperation to NIDDK-CR to correct any failures of the resources as requested by NIDDK-CR.

3. Data Quality and Documentation

To ensure that data submitted to NIDDK-CR are complete, valid, and can be meaningfully used, NIDDK-CR support staff will replicate select tables from published results. Data should be submitted in a timeframe that allows contributing study staff to answer questions that may arise during this process.

All research data submitted to NIDDK-CR must be accompanied by proper documentation to ensure proper secondary use of the data. Study documentation should be in electronic format,

comprehensive, and sufficiently clear to enable investigators not involved with the contributing study to understand the study and data. Refer to the NIDDK Central Repository Data and Documentation Submission Guidelines for information on preparing data and the associated documentation for submission to the NIDDK-CR.

II. Archiving Procedures

The following section describes the processes and procedures for archiving resources once submitted to NIDDK-CR.

1. Storage

NIDDK-CR system utilizes two types of data storage and appropriate backups for each:

- **Relational Database:** The relational database holds all application pertinent data, such as qualified investigator information, redacted specimen metadata, and request information, all of which are normalized into tables with powerful querying and indexing capabilities. NIDDK-CR system uses a proven, robust, open-source PostgreSQL database via Amazon Web Services (AWS) RDS (Relational Database Service) as the persistent storage for application pertinent data. With this implementation, NIDDK-CR relational database is highly scalable and fully managed with security, configuration, backups, and recovery.
- **Object Storage**: The object storage system is another key data storage implementation for NIDDK-CR. It is primarily used for storing data files (i.e., data packages available for download by approved requestors) and study documentation (e.g., data dictionaries, protocols, inventory reports, and other supporting documentation) relevant to the study. NIDDK-CR uses AWS Elastic File System (EFS) for object storage implementation. Object storage in the cloud supports a vast variety of file types and sizes, and is highly scalable, durable, secure, and cost-efficient.
- **Backups:** NIDDK-CR data are backed up following the guidelines set in the NIH InfoSec Policy Handbook. Automated processes are utilized by the system to ensure that appropriate system storage procedures are implemented. The processes are implemented by managed services or by customized automated backups using a cloud-native managed service. Storage devices, databases, and servers are backed up and stored in a separate infrastructure region from the production data.

2. Updates and Migration

Cloud Service Provider (CSP) managed services are updated transparently by the cloud provider. Services managed by the team are continuously scanned by NIH-managed Tenable Nessus with automated reporting sent and reviewed daily. Migration of system deployments to development, testing, and production environments are automated via continuous deployment pipelines.

3. Data Integrity

NIDDK-CR data integrity is achieved through using multiple automatic and manual validation processes throughout the lifecycle of the data in the NIDDK-CR.

- **Manual Processes:** Data are reviewed and curated by NIDDK-CR support staff when data are submitted to the NIDDK-CR. Activities include but are not limited to:
    - Verifying that the data are free from direct personal identifiers based on HIPAA ruling 45 CFR § 164.514 for the use of limited data sets
    - Verifying that supporting documentation is consistent with data content
    - Generating Section 508-compliant versions of required study documentation
    - Organizing data and associated documentation into meaningful packages (i.e.,

zipped files) for requestors
    - Assigning data package version numbers
- **Automatic Processes**: Several data validation processes are implemented in the system including:
    - Using coded values in data entry instead of free text to reduce possible errors
    - Formatting and business rules-based validation in data entry and data loading processes
    - Checksum hash validation runs weekly to validate data package integrity in which any discrepancies are investigated and corrected

4. Data Preservation

NIDDK is committed to ensuring long-term preservation and access to resources under the custodianship of NIDDK. Resources submitted to NIDDK-CR will be preserved for use by the broader scientific and research community until these are used up or to the end of their scientific utility. NIDDK-CR will periodically assess the ongoing scientific utility of held resources:

- Resource utilization rates below five percent in a five-year consecutive period will be evaluated annually to determine continued scientific utility.

- NIDDK-CR will discuss resource management options, including conversion into renewable resources and deaccessioning of resources with low utilization and low scientific utility.

5. Security

NIDDK-CR is a federal information system. It complies with security policies and requirements established in both the NIH InfoSec Policy Handbook and the Federal and HHS Information Security Program (FISMA) moderate level categorization. NIDDK-CR system is FISMA moderate compliant and is reviewed annually by NIH Security team.

System security is achieved through compliant design and development policies and using FedRAMP compliant AWS services. Several security related processes are performed to ensure compliance including:

- Tenable Nessus vulnerability scanning reviewed by the team daily
- Netsparker web application security and penetration testing executed regularly
- Synack penetration tests performed by a third party and issues are resolved by NIDDK-CR support staff according to issue resolution policy
- Checks related to data upload process that prevents users from uploading compromised documents
- Access controls in place for cloud system components and network access including Multi Factor Authentication (MFA)
- AWS Well-Architected system best practices included to ensure confidentiality, integrity, and availability of the cloud system
- Trusted and managed security tools are deployed to team-managed system components for threat intelligence, monitoring, hardening, and reporting
- Robust and compliant user registration and login processes that require users to confirm their email address by clicking on the link sent by the system to their email address
- Captcha is implemented on the registration and login screens

III. Access Requirements

Investigators requesting access to resources from NIDDK-CR must comply with the specific procedures described below.

1. Project Plan and Research Use Statement

Data or specimen access requests should include a description of the research that justifies the use of the resources, a research objective and design that describes the hypothesis and approach, an analysis plan that includes security safeguards for the data, and a research use statement to be made publicly available.

2. Data and Resource Use Agreement

An executive summary (including the project plan and research use statement) will be appended to and become part of the DUA before executing the agreement. Requesting investigators will be required to sign a NIDDK-CR Data and Resources Use Agreement for research data, or for requesting specimens and associated data, whereby they agree to the terms of access.

3. Ethics/Institutional Review Board (IRB) Review

Access to NIDDK-CR held data, requires an Ethics/IRB review clearance or exemption from their institution. If the requestor's institution does not have an Ethics/IRB, they must use an external IRB. NIDDK-CR resources are devoid of direct identifiers, pseudo-anonymized, or anonymized and may fall under Not Human Subjects Research (NHSR).

4. Data and/or Specimen Access Approval

NIDDK has final decision authority for granting access to data and/or specimens under the guardianship of NIDDK-CR. For active studies, contributing study leadership in collaboration with NIDDK will make a determination for granting access.

IV. Inquiries

For inquiries on NIDDK-CR Data Preservation and Access Practices, please send an email to NIDDK-CRsupport@niddk.nih.gov.