NIH National Institute of Diabetes and Digestive and Kidney Diseases | NIDDK Central Repository

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

Data Preservation and Access Practices

The NIDDK Central Repository - Resources for Research (NIDDK-CR R4R) facilitates sharing of data, biospecimens, and other resources generated from studies supported by NIDDK and within NIDDK's mission by making these resources available for meaningful reuse. Transfer of resources to and from the NIDDK-CR is governed by NIH and NIDDK data sharing polices, and NIDDK-CR assumes responsibility and 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 data in the NIDDK-CR, and the subsequent archiving of data to ensure long-term preservation.

I. Submission Requirements

Investigators submitting study research data, images, or biospecimen linking files to the NIDDK-CR must comply with the specific procedures described below.

1. Data Redaction

The NIDDK-CR has chosen to receive and distribute 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 based on 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 refer to the NIDDK Central Repository Data and Documentation Submission Guidelines when preparing study data, images, documentation, or biospecimen linking files for submission to the NIDDK-CR.

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

2. Material Transfer Agreement

The NIDDK-CR acts as an honest broker for sharing study resources with qualified investigators based on a study's sharing plan and/or informed consent language for sharing. An NIDDK Material Transfer Agreement (MTA) may be needed if a sharing plan and/or informed consent language is not available or is insufficient to determine any restrictions related to future use, or if the submitting institution (such as the study's coordinating center) requires a MTA to be executed before transferring resources to the NIDDK-CR.

The MTA assures that:

- The research materials are being provided for the purpose of the repository distributing the research materials to requestors. The provider grants the repository explicit permission to distribute the research materials to requestors as a research resource.
- The originating or submitting institution, certifies that the research materials were collected according to 45 CFR Part 46, "Protection of Human Subjects" at all the study sites and the research materials are not to be used in human subjects or for the treatment or diagnosis of human subjects.
- The provider certifies that an Institutional Review Board (IRB) has reviewed and verified that submission of research materials to the repository for subsequent sharing for research purposes is consistent with the informed consent of study participants from whom the research materials were obtained.
- The originating or submitting institution, and the NIDDK-CR (or an NIDDK approved repository) acknowledge that the biospecimens may be limited in quantity and that their distribution for research purposes will be based on the scientific merit of a proposed research project. Scientific merit of all requests for research materials will be determined by NIDDK. The originating or submitting institution acknowledges that the NIDDK-CR will periodically assess the ongoing scientific utility of biospecimens. Biospecimens determined to be of low scientific utility or non-viable may be discarded.
- The originating or submitting institution agrees to provide NIDDK a clear statement identifying all restrictions or limitations on the use or distribution of research materials (e.g., for diabetes research only) specified in the study participants' informed consent documents.

3. Data Quality and Documentation

In order to ensure that data submitted to the NIDDK-CR are complete, valid, and can be meaningfully used, NIDDK-CR Support Staff will replicate selected tables from published results. Data should be submitted in a timeframe that allows study staff to answer questions that may arise during this process.

All research data submitted to the NIDDK-CR must be accompanied by proper documentation to ensure meaningful use of the data. Study documentation should be in electronic format and comprehensive and sufficiently clear to enable investigators to understand the study and data. Investigators should refer to the NIDDK Central Repository Data and Documentation Submission Guidelines for the list of required study documentation for submission to the NIDDK-CR.

II. Archiving Procedures

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

1. Storage

The 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 biospecimen metadata, and request information, all of which are normalized into tables with powerful querying and indexing capabilities. The 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, the 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 the 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. The NIDDK-CR uses AWS EFS (Elastic File System) for object storage implementation. Object storage in the cloud supports a vast variety of file types and sizes, and is also 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

CSP (Cloud Service Provider) 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 repository.

- **Manual Processes:** Data are reviewed and curated by the NIDDK-CR Support Staff when data packages are delivered for study onboarding or study updates to the repository. 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 requesters
  - 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. Security

The 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. The NIDDK-CR system is FISMA moderate compliant and is reviewed annually by the 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 the NIDDK-CR Support Team 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 MFA (Multi Factor Authentication)
- 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 the NIDDK-CR must comply with the specific procedures described below.

1. Project Plan and Research Use Statement

Data and/or biospecimen 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/or Specimen Use Agreement

Project Plan and Research Use Statement, also known as the executive summary, will be appended to, and become part of the Data Use Agreement or the Specimen and Data Use Agreement before executing the agreement. Requesting investigators will be required to sign a NIDDK-CR Data Use Agreement for research data, or the NIDDK-CR Specimen and Data Use Agreement for requesting biospecimens and associated data, whereby they agree to the terms of access.

3. IRB Approval

For any level of access to the data, requestors must have an IRB clearance or exemption from their institution. If the requestor's institution does not have an IRB, they must use an external IRB. Since the data in NIDDK-CR are provided in Limited Datasets, redacted of direct identifiers, and there is no direct human-subject interaction, the IRB may instead make a determination of Not Human Subjects Research (NHSR) rather than provide IRB approval.

4. Data and/or Biospecimen Access Approval

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

IV. Inquiries

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