Restricted-Use Data Deposit and Dissemination Procedures

October 2015
Restricted-Use Data Deposit and Dissemination Procedures

National Addiction & HIV Data Program
Inter-university Consortium for Political and Social Research

Contents

Overview of Procedures

Direct and Indirect Identifiers
  a. Social Science Research Data
  b. Protected Health Information (PHI)

Description of Agreements
  a. Restricted-Use Data Deposit and Dissemination Agreement (RUDDDA)
  b. Restricted Data Use Agreement (contained in the RUDDDA)
     1. Core Data Security Requirements
     2. Data Security Requirement Optional or with Options
  d. Materials Transfer Form (MTF)

Flow of Restricted-Use Data Deposit and Dissemination Agreement Between Parties

ICPSR Handling of Restricted-Use Data

ICPSR Review and Administration of Restricted Data Use Agreements

Data Release Options
  a. Public-Use
  b. Restricted-Use
  c. Virtual Data Enclave
  d. Physical Data Enclave
  e. Delayed Dissemination

Revised October 2015
Overview of Procedures

The transfer, processing, and dissemination of restricted-use data by the National Addiction & HIV Data Program (NAHDAP) in the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan (U-M) typically involves several steps. All steps regarding the legal agreements require the review of the documents by the legal counsel’s office of the participating parties. Although currently no covering legislation or requirement demand that these procedures be followed, they protect the original owners or providers of the data, ICPSR, and the University of Michigan.

Once data owners or data providers decide to deposit data with NAHDAP, NAHDAP staff provide technical assistance in determining the files, formats, and content of materials to deposit. The data providers and NAHDAP staff determine whether the data can be released as a public-use file (PUF) (i.e., downloadable from a web site) or if the data need to be designated as a restricted-use file (RUF) due to confidential content (i.e., indirect identifiers) in the data. A test deposit can be made so that NAHDAP staff can view the data or documentation prior to the official transfer of data or documentation files. Technical assistance can be provided by email, conference calls, or in-person as needed.

If the data will be restricted-use, the data provider will need to identify the legal representative at their institution and/or funding agency that will review the legal agreements. They may want to contact the legal representative early on to inform them of the plans for data transfer and their role in the process.

NAHDAP staff typically start the process by preparing a customized version of the legal agreement with information supplied by the data provider (e.g., the institution’s legal name and address, preference for electronic or hard copy legal agreements). After the fully executed agreement is in place, the data provider can officially transfer the data to NAHDAP staff so that they can begin their work. Release of the restricted-use data is governed by the conditions specified and agreed to by all parties. A flow diagram of this process is described later in this document.

Copies of the all agreements mentioned in this document are available upon request to NAHDAP staff.

About NAHDAP

NAHDAP’s mission is to facilitate research on drug addiction and HIV infection by enhancing, sharing, and preserving data produced by research funded by the National Institute on Drug Abuse. In addition to providing access to research data, NAHDAP offers technical assistance for data depositors and e-workshops on data preparation and data systems of interest to the addiction and HIV research community.

NAHDAP has an Authority to Operate (ATO) and is certified as FISMA Moderate for its secure computing environment used for restricted-use data.
Direct and Indirect Identifiers

a. Social Science Research Data

NAHDAP urges researchers to consider subject confidentiality during the process of producing their data files. Using a confidentiality plan during data file construction allows for handling of subject confidentiality in a systematic manner and makes masking or removing information that could be used to identify research participants easier before submitting the data to an archive. Confidentiality adjustments, however, may impose limitations on the research utility of the data. Data depositors should carefully consider the analytic role that such variables play and should remove any identifiers not necessary for analysis.

Direct identifiers are variables that point explicitly to particular individuals or units. All variables directly identifying research subjects must be removed or masked prior to deposit. Examples include:

- Names
- Addresses, including ZIP codes
- Telephone numbers, including area codes
- Social Security numbers
- Other linkable numbers, including driver license numbers, inmate identification numbers, etc.

Indirect identifiers are variables that may be used in conjunction with other information to identify individual respondents. Indirect identifiers may remain in the deposited data if needed to reproduce original research findings or if removing them significantly degrades the analytic value of the data. Examples include:

- Detailed geographic information (e.g., state, county, or census tract of residence)
- Organizations (to which the respondent belongs)
- Educational institutions (from which the respondent graduated and year of graduation)
- Exact occupations
- Place where respondent was born or grew up
- Exact dates of events (birth, death, marriage, divorce)
- Detailed income
- Offices or posts held by respondent

NAHDAP staff work closely with data depositors to resolve confidentiality issues and to determine if the data should be released as public-use or restricted-use. NAHDAP staff may recode data to address the threat of disclosure. Recoding can include converting dates to time intervals, exact dates of birth to age groups, state of residence to regional codes, and income to income ranges or categories.

Data depositors should document any modifications they have done to the data to address confidentiality. This will ensure that NAHDAP staff do not make unnecessary changes to the researcher’s modifications when performing their confidentiality review. Such information can also be made available to secondary users of the data to assist them with their use of the data.
b. Protected Health Information (PHI)

Protected Health Information (PHI) is collected by a covered entity as defined in the Health Insurance Portability and Accountability Act (HIPPA). Quoting 45 C.F.R. § 164.514(b), health information is not individually identifiable health information if the following identifiers of the individual or of relatives, employers, or household members of the individual, are removed:

(A) Names [including initials];

(B) All geographic subdivisions smaller than a State, including street address, city, county, precinct, zip code, and their equivalent geocodes, except for the initial three digits of a zip code if, according to the current publicly available data from the Bureau of the Census:
   (1) The geographic unit formed by combining all zip codes with the same three initial digits contains more than 20,000 people; and
   (2) The initial three digits of a zip code for all such geographic units containing 20,000 or fewer people are changed to 000.

(C) All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death; and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older;

(D) Telephone numbers;

(E) Fax numbers;

(F) Electronic mail addresses;

(G) Social security numbers;

(H) Medical record numbers;

(I) Health plan beneficiary numbers;

(J) Account numbers;

(K) Certificate/license numbers;

(L) Vehicle identifiers and serial numbers, including license plate numbers;

(M) Device identifiers and serial numbers;

(N) Web Universal Resource Locators (URLs);

(O) Internet Protocol (IP) address numbers;

(P) Biometric identifiers, including finger and voice prints;

(Q) Full face photographic images and any comparable images; and

(R) Any other unique identifying number, characteristic, or code except as permitted by §164.502(c).

The covered entity (i.e., a health care provider, health care clearinghouse, or health plan) must not have actual knowledge that the information could be used alone or in combination with other information to identify an individual who is a subject of the information.


2 Alternatively, a “person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable, determines the risk is very small that the information could be used by the recipient, alone or in combination with other reasonably available information, to identify an individual (section 164.514(b)(1)), and documents the basis for such determination” (see http://privacyruleandresearch.nih.gov/research_repositories.asp and the above referenced document.
A de-identified data set created by using the safe harbor method as described above is not protected by the Privacy Rule and may be used and disclosed without restriction\(^3\). Such a data file is considered a public-use data file and may be accessed directly through the ICPSR Web site or the NAHDAP Web site. Data users must agree to the ICPSR Terms of Use.

A limited data set excludes the direct identifiers listed above (see also 164.514(e)(2)). In contrast to a de-identified data set, a limited data set is Protected Health Information (PHI) since data may retain:

- Dates
- Five-digit or nine-digit ZIP codes plus any other geographic subdivision, such as state, county, city, precinct, and their equivalent geocodes, except for street name or street address or post office box.
- Other unique identifying codes or characteristics not listed as direct identifiers above.

In other words, in order to qualify as a limited data set, all of the following direct identifiers of the individual or of relatives, employers, or household members of the individual must be removed from the PHI:

- Name
- Street name or street address or post office box (i.e., postal address information other than town or city, State, and zip code)
- Telephone and fax numbers
- Email address
- Social security number
- Certificate/license numbers
- Vehicle identifiers and serial numbers
- URLs and IP addresses
- Full-face photos and other comparable images
- Medical record numbers, health plan beneficiary numbers, and other account numbers
- Device identifiers and serial numbers
- Biometric identifiers, including finger and voice prints

A limited data set may be used for research purposes in accordance with 164.512(e), only if the covered entity and the recipient of the limited or restricted-use data set enter into a Data Use Agreement\(^4\). Under the Privacy Rule, a valid Data Use Agreement must establish ways in which information in the limited data set can be used and how the information will be protected by the recipient throughout the research project.

Note: PHI is only collected by a covered entity. Data collected by a social science researcher using a survey or other collection method not involving a covered entity is not considered PHI.

\(^3\) However, this statement does not apply if the covered entity has actual knowledge that the remaining information alone or in combination with other information can be used to possibly identify the subject.

\(^4\) More information on limited data sets and Data Use Agreements is available in Protecting Personal Health Information in Research: Understanding the HIPAA Privacy Rule.
Description of the Agreements

The following three documents are completed, reviewed, and executed by the specified parties.

a. **Restricted-Use Data Deposit and Dissemination Agreement (RUDDDDA)**

This agreement is unique to the two parties engaged in the transfer of restricted-use data to NAHDAP/ICPSR for processing and dissemination by NAHDAP/ICPSR and is negotiated between those parties. Restricted-use data are usually transferred only under a use agreement, which is not required for public-use data. The RUDDDA protects the data while housed at ICPSR during processing as it extends the standard ICPSR deposit terms to specify the conditions under which the data will be transferred, stored, and handled at ICPSR. The RUDDDA also denotes the conditions for release and dissemination of the restricted-use data and the administration and content of the Restricted Data Use Agreement (described below).

The RUDDDA absolves the data provider’s institution from liability for security or disclosure breaches (intentional or inadvertent) if they occur while the data are housed at ICPSR. Two versions exist: one between a university or institution and NAHDAP/ICPSR and the second between NIDA and NAHDAP/ICPSR.

The Data Security Plan is incorporated into the Restricted Data Use Agreement as specified within the RUDDDA (see Appendix B). Any role that the sending institution wants in the administration and/or enforcement of Restricted Data Use Agreement can be described in this document. The RUDDDA also sets NAHDAP/ICPSR’s obligation to monitor and enforce all legal obligations specified in the Restricted Data Use Agreement.

The main text of the RUDDDA was reviewed by the University of Michigan’s Office of Research and Sponsored Projects (ORSP). Some sections can be modified under negotiation by both parties. The RUDDDA is signed by the data provider’s institution and U-M’s ORSP.
b. Restricted Data Use Agreement (RDUA)

The Restricted Data Use Agreement is completed and signed by data users during the process to request access to the restricted-use data. The document specifies the terms of access to the restricted-use data, data confidentiality and data security requirements, and consequences if a violation of the Agreement occurs. The RDUA is signed by the requesting investigator and the requestor’s institutional representative. NAHDAP/ICPSR staff record in ICPSR’s online system that the request was approved. The RDUA is considered a unilateral agreement and is no longer signed by a representative of U-M. The main text of the NAHDAP/ICPSR Restricted Data Use Agreement was reviewed by the U-M’s Office of Research and Sponsored Projects. Some sections (e.g., Section I.G.) can be modified under negotiation by both parties to set specific conditions on data user access and release of information based on the restricted-use data. Modifications of the standard RDUA require approval from legal representatives of the requestor’s institution and NAHDAP/ICPSR.

The Restricted-Use Data Deposit and Dissemination Agreement identifies data security requirements for the RDUA. Some data security requirements apply to all Restricted Data Use Agreements and are considered core data security requirements. Other requirements can be selected to apply to the Data Security Plan for a particular restricted-use dataset. NAHDAP/ICPSR indicates in the RUDDDA often-used optional data security requirements. These options may be modified by the data provider.

Core Data Security Requirements cover:
- ICPSR specifications on passwords and encryption
- Restrictions on access to the restricted-use data and any derivatives
- Electronic and physical storage of the restricted-use dataset, any derivatives, temporary files, and backup files
- Changes in research project staff or institution
- Requests made by any outside entity to access the restricted-use dataset
- How and when electronic and printed copies of the restricted-use data and any derivatives must be destroyed

Data Security Requirements Optional or With Options cover:
- Data storage technology (standalone computer, external hard drive, private networks, other types of networks, virtual data enclave, etc.)
- Disclosure review of output if using the virtual data enclave
- Student requestors with a faculty member as a sponsor
- Number of backup copies of the restricted-use data allowed
- Time frame of the RDUA and access to the restricted-use data

The standard NAHDAP RDUA is available upon request to NAHDAP staff and is posted under Quick Links on the NAHDAP Web site (www.icpsr.umich.edu/NAHDAP).
c. The Material Transfer Agreement (MTA) or Analogous Form

The University of Michigan requires an agreement when no research sponsorship exists that sets the conditions under which materials (in this case, data) are transferred by another university or organization to the University of Michigan. If the data provider and/or the provider’s institution elects to execute a RUDDDA rather than use the ICPSR online deposit form and standard Deposit Terms, the MTA assures that the exchange and use of the material is in compliance with U-M policies, other sponsorship requirements (for NAHDAP, this is NIDA), informed consent, and other overlapping agreements or licenses. The MTA allows the University of Michigan to insure that NAHDAP’s receipt and use of the materials do not violate the intent of the originating institution or the University of Michigan. The incoming U-M MTA is prepared by NAHDAP staff, signed by NAHDAP/ICPSR and is transmitted through an online U-M portal with the RUDDDA to the U-M Office of Research and Sponsored Projects.

However, some institutions require their own version of an outgoing Material Transfer Agreement in order to transfer the data to NAHDAP. If the data provider’s institution requires such an agreement, the data provider completes their information for the form, signs it, and sends it to NAHDAP staff. NAHDAP staff add information for U-M and forwards the MTA to the U-M’s Office of Research and Sponsored Projects for their signature.

Flow of Restricted-Use Data Deposit and Dissemination Agreement Between Parties

The data provider may be at a university, funding agency, or other institution or organization. The provider’s legal representative is the person that can sign binding legal agreements on behalf of the provider’s institution.

Executing the RUDDDA between the data provider and ICPSR/NAHDAP involves the following steps.

1. ICPSR staff customizes the generic Restricted-Use Data Deposit and Dissemination Agreement to specify the data provider’s institution as party to the RUDDDA with ICPSR using information supplied by the data provider.

2. ICPSR staff send, typically via email, the customized RUDDDA to the data provider or designated contact (e.g., principal investigator, project manager, etc.).
3. Provider reviews the RUDDDA to understand the terms as it applies to provider’s role in the agreement pertaining to preparing and transferring the restricted-use data to NAHDAP/ICPSR and data security options for the RDU, if desired.

4. Provider forwards the RUDDDA to their legal office for review, modification, and/or signature.

5. Provider’s legal representative reviews the RUDDDA and signs it if no modifications need to be negotiated.

6. Provider’s legal representative returns the partially-executed RUDDDA to Provider.

7. Provider forwards the partially-executed RUDDDA to NAHDAP/ICPSR staff for their review.

8. NAHDAP/ICPSR staff review the partially-executed RUDDDA and, if no changes need to be addressed, compiles documents (PDF of partially-executed RUDDDA, Word version of unsigned RUDDDA), prepares the online incoming MTA, and uploads these documents to the U-M system. The ICPSR/NAHDAP Archive Director signs the MTA.

9. NAHDAP/ICPSR staff submits the documents to U-M’s Office of Research and Sponsored Projects (ORSP), aka U-M legal for signature.

10. U-M legal representative reviews the partially-executed RUDDDA and signs it if no modifications need to be negotiated.

11. U-M legal representative sends, typically via email, a copy of the fully-executed RUDDDA to provider’s legal representative and NAHDAP/ICPSR staff.

12. NAHDAP/ICPSR staff forward a copy of the fully-executed RUDDDA to Provider (upon request).

The red arrows in the flow diagram indicate the return of the RUDDDA due to a request for modification(s) to the RUDDDA that need to be negotiated between the parties. Although not common, a request for modification(s) can be initiated by either party.

- Provider or provider’s legal representative requests modification(s). The modified RUDDDA may be sent to NAHDAP/ICPSR staff or U-M legal depending on the stage at which the modification is initiated.

- NAHDAP/ICPSR staff or U-M legal representative may notice a discrepancy from the generic legal language. NAHDAP/ICPSR staff or U-M legal representative will send a modified version of the (partially-executed) RUDDDA to the provider or the provider’s legal representative, again depending on the stage at which the change is initiated.

A modified RUDDDA unsigned by the provider’s legal representative, if approved, is signed by U-M legal and sent to provider’s legal representative for review, approval, and signature. The legal representatives correspond between themselves to achieve a full-executed RUDDDA with
copies for both parties. A modified RUDDDA that was initialed and signed by the provider’s legal representative, if approved, is fully executed by U-M legal.

The U-M legal representative sends the provider’s legal representative and NAHDAP/ICPSR staff with a copy of the fully-executed RUDDDA. A copy can also be sent to the Provider upon request.

**ICPSR Handling of Restricted-Use Data**

All NAHDAP/ICPSR staff members who work with restricted-use datasets signed a confidentiality pledge as a term of employment. For enhanced security, NAHDAP/ICPSR staff perform their data processing tasks within ICPSR’s Secure Data Environment (SDE). Staff use a “virtual desktop interface” that is created anew with each login and is isolated from Internet access. The virtual desktop is isolated from staff’s local computer. Staff cannot cut-and-paste between the virtual desktop and the local computer or email files in the SDE to others. Staff can print files only locally and access Internet applications using connections provided within the SDE.

Files enter the SDE in an encrypted form through the ICPSR’s Deposit Form. Movement of files to outside the SDE other than through the ICPSR data processing system is infrequent and strictly audited. Any removable media (e.g., CDs, DVDs) containing restricted-use data are kept in locked cabinets in the NAHDAP office or work area until completion of the data processing. NAHDAP/ICPSR staff must use a strong password to log on to the regular desktop and again to log on to the SDE. Desks lock when the screen saver is activated. When NAHDAP staff are working with restricted-use data and associated materials (e.g., documents or printouts with confidential information), desktops and office doors are locked when the processor is not present. ICPSR, located within the Institute for Social Research (ISR), backs up SDE content to a secure tape library located within an ISR machine room, and occasionally rotates tapes to a second secure location within ISR for disaster preparedness. ICPSR produces archival copies of all content, but encrypts any copies located off-site.

**NAHDAP/ICPSR Review and Administration of Restricted Data Use Agreements**

ICPSR developed online systems for receiving, processing, and administering RDUA to access restricted-use data. Separate online request systems are used by a prospective data user (i.e., Investigator) to request access to restricted-use data available to be analyzed in the Investigator’s local work environment and for access to restricted-use data only able to be analyzed in ICPSR’s Virtual Data Enclave (VDE). The Access Notes on the study download page provides prospective investigators with a link to the request system.

Requests are reviewed by NAHDAP staff for approval. If approved, the data are copied to a secure place for secure download by the approved Investigator or sent to the Investigator on an encrypted compact disk through the postal service with signature required. NAHDAP/ICPSR staff track the status of a RDUA and alert Investigators for updates to the online request.
information as needed (e.g., annual report, updated IRB document, agreement period ending soon). Investigators and named research project staff electronically sign pledges of confidentiality. Electronic signatures are not used for the affidavit of the destruction of the data at the end of the period as Investigators must provide a notarized affidavit.


Data Release Options

Researchers should consider whether the data should ultimately be made available in a public-use or restricted-use file. In some instances, the same data can be provided in both forms. A restricted-use version that includes confidential information is available to approved researchers under a set of controlled conditions. The restricted-use dataset approach is an effective way to permit access to confidential data and has proven acceptable to researchers. For an even greater level of security, some archives provide confidential data through data enclaves, which require that researchers visit the enclave to access the data under secure conditions.

a. Public-Use

Public-use files do not have direct and indirect identifiers and so disclosure risk is considered minimal. These data may be downloaded directly through the ICPSR Web site or the NAHDAP Web site. Data users must agree to the ICPSR Terms of Use. Since NAHDAP is sponsored by the National Institute on Drug Abuse (NIDA), a federal agency, users obtain the data at no charge.

b. Restricted-Use

Restricted-use data are distributed in cases when removing potentially identifying information would significantly impair the analytic potential of the data. Users (and their institutions) must apply for access to restricted-use data, use data security plans, and agree to other access controls.

When a Restricted Data Use Agreement (RUDA) is approved, ICPSR Release Management staff with access to restricted-use data use ICPSR’s secure server to prepare the data for the approved request. Some data are encrypted and securely disseminated via a download by the researcher using a temporary link and password (provided to the researcher separately). In some instances restricted-use data are sent on encrypted CD to the Investigator by registered mail that requires the Investigator to sign for the package to insure receipt. The CD includes an identifier so that when the data are returned by the Investigator upon termination of the agreement period, ICPSR can confirm the original data were returned.
c. Virtual Data Enclave

ICPSR’s Virtual Data Enclave (VDE) allows remote access to restricted-use data by approved researchers from their own desktop but operating on ICPSR’s servers. The VDE uses the same technology used for ICPSR Secure Data Environment (SDE) for NAHDAP/ICPSR data processing.

To access the VDE, researchers install client software on their computer to open a secure portal to the data servers at ICPSR. All restricted-use data and the researcher’s analytic output or other files remain on the ICPSR servers within the VDE. The researcher’s virtual machine in the VDE is created anew with each login and is isolated from their physical desktop computer, restricting the researcher from downloading files or parts of files to their physical computer. The virtual machine is also restricted in its external access, preventing users from emailing, copying, or otherwise moving files outside of the secure environment, either accidentally or intentionally. A review for risk of disclosure of confidential information is typically conducted by NAHDAP/ICPSR staff on any files that the Investigator wants to receive outside the VDE.

d. Physical Data Enclave

Some data cannot be sufficiently de-identified to allow researchers access to the data without authorization and supervision. Such restricted-use data are only accessible in ICPSR’s Physical Data Enclave in Ann Arbor, Michigan. Researchers apply to use these data. After approval, researchers are allowed access to restricted-use data on a computer in the physical enclave that is not connected to a network. Any written notes, data analysis printouts, and electronic files that the researcher wants to have outside the enclave are subject to review by NAHDAP/ICPSR staff for compliance with confidentiality standards. Once cleared, NAHDAP/ICPSR staff send the materials to the researcher. Researchers are responsible for covering their travel and lodging while using the data enclave.

e. Delayed Dissemination

When significant potential for disclosure risk exists or when data are still in operational use by an organization, NAHDAP/ICPSR may also elect to preserve a data collection without dissemination until human subject risks are minimized (e.g., when all human subjects are deceased). Delayed Dissemination allows data to be deposited but not disseminated until an agreed-upon period of time has passed (typically one year). During the time between deposit and dissemination, NAHDAP/ICPSR staff develop expertise with the data collection, review the data and ask questions while project staff are available to provide assistance, and preserve the original and processed versions of the data. One or more of the data release options listed above will be used to disseminate the processed data once the embargo period is over.