ICPSR INSTITUTE FOR SOCIAL RESEARCH UNIVERSITY OF MICHIGAN

Restricted-use Data Deposit and Dissemination Procedures

May 2022

INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

Inter-university Consortium for Political and Social Research

Institute for Social Research
University of Michigan
P.O. Box 1248
Ann Arbor, MI 48106-1248

Voice: +1 (734) 647-2200
Fax: +1 (734) 647-8200
Email: ICPSR-help@umich.edu
Web: https://www.icpsr.umich.edu/icpsrweb/

Contents

Overview of Procedures

The transfer, processing, and dissemination of restricted-use data by the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan (U-M) typically involves several steps. When required, documents governing the transfer, storage, and use of restricted-use data are reviewed and approved by each participating party's legal counsel. All procedures and data governance documents are in place to support responsible data stewardship and to clarify and protect the interests of research subjects, depositing institutions, and ICPSR.

Once data providers decide to deposit data with ICPSR, ICPSR staff are available to provide technical assistance in determining the files, formats, and content of materials to be deposited. ICPSR staff then determine, sometimes in consultation with the data provider(s), whether the data can be released as a public-use file (PUF) or as a restricted-use file (RUF). PUFs are available to be downloaded directly from the ICPSR website and contain little to no risk of study participant re-identification. RUFs must be requested through a formal application process before access to the data is granted and may contain some risk of re-identification of study participants (e.g., indirect identifiers) or otherwise confidential or sensitive information. Data providers may place a request for ICPSR staff to review the data or documentation prior to the official deposit of these files.

If the data will be restricted-use, the data provider may choose to deposit directly under the terms of the standard ICPSR deposit agreement found in our online deposit system, or they may choose to establish a formal legal agreement between their institution and ICPSR. Establishing a formal legal agreement can give the data provider more control over dissemination of the data, makes responsibilities explicit, and allows incorporation of specific terms in cases where the data is regulated or otherwise unusually sensitive.

Release of the restricted-use data to third-party data users is governed by a Restricted Data Use Agreement (RDUA), described in greater detail further on in this document. Copies of all agreements mentioned in this document are available upon request.

About ICPSR

ICPSR's mission is to advance and expand social and behavioral research, act as a global leader in data stewardship, and provide rich data resources and responsive educational opportunities for present and future generations.

In addition to providing access to research data, ICPSR offers technical assistance and letters of collaboration for data depositors, conducts webinars and workshops on data preparation and analysis, provides other teaching and learning resources, and demonstrates the impact of research data.

Direct and Indirect Identifiers

Two kinds of information often found in individual-level human subjects research data present problems that could endanger research subjects' confidentiality: direct identifiers and indirect identifiers. With the exception of data placed in openICPSR, our self-publishing data-sharing service, ICPSR reviews data sets to assess disclosure risk. ICPSR trains data curators to apply specified procedures to protect human subject confidentiality in all of the data ICPSR curates, archives, and distributes, in part by checking each study for these identifiers.

Role of the Original Data Provider

ICPSR urges researchers to consider human subject confidentiality during the process of producing and managing their data files. Using a data management plan during data file creation allows for handling of human 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. It is also important to keep in mind that confidentiality adjustments may impose limitations on the research utility of the data. Thus, data depositors should carefully consider the analytic role that such variables play and should remove any identifiers not necessary for analysis.

Direct identifiers point explicitly to individuals or units. Information directly identifying research subjects should typically 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's license numbers, inmate identification numbers, etc.

ICPSR is able to accept data with direct identifiers for specific purposes (e.g., as a safe haven for future data linkage). Please contact us for more information.

Indirect identifiers constitute information 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

- Names of educational institutions and year(s) of attendance/graduation
- Exact occupations
- Place where respondent was born or grew up
- Dates of events (e.g., birth, death, marriage, divorce)
- Detailed income
- Offices or posts held by respondents

ICPSR staff often work closely with data depositors to resolve confidentiality issues and to determine if the data should be released as public-use or restricted-use. ICPSR staff may recommend that the data be recoded to reduce 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 ICPSR 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 understanding and use of the data.

Protection of respondent confidentiality is a core tenet of responsible research practice that begins with obtaining informed consent. Researchers are encouraged to consult ICPSR's Recommended Informed Consent Language for Data Sharing at the beginning of their study when drafting their informed consent and IRB documents. This recommendation discourages researchers from including language in the informed consent that would prohibit data sharing and encourages transparency around the plan for sharing data. When possible, ICPSR prefers to have a copy of the researcher's informed consent document—including the way(s) in which the researcher described how respondent confidentiality would be protected and how the data may be shared with the research community—in order to further assess how best to protect respondent confidentiality.

Protected Health Information (PHI)

Protected Health Information (PHI) is a special case. PHI is individually identifiable health information which comes within the regulatory scope of the Health Insurance Portability and Accountability Act of 1996 (HIPAA). HIPAA governs how PHI may be used or disclosed, and for what purposes. The rules include provisions for creation of two types of data sets with reduced disclosure risk:

- **Limited Data Sets**, in which all direct identifiers are removed, as well as certain indirect identifiers. A Limited Data Set is PHI, and has to be protected as such, but may be disclosed under a Data Use

Agreement for Research, Public Health, or Health Care Operations purposes as defined in HIPAA. ICPSR will archive and distribute Limited Data Sets, under a HIPAA-compliant DUA.

- **De-Identified Data**, in which sufficient identifiers have been removed so that there is no reasonable basis to believe that the information could be used to identify an individual. De-Identified data is no longer PHI, and is therefore not subject to HIPAA. Data users would still have to agree to the ICPSR Terms of Use, which adhere to the principles of the CoreTrustSeal Core Trustworthy Data Repositories Requirements. These require the data consumer to comply with access regulations and applicable licenses imposed both by law and by the data repository, and to conform to codes of conduct that are generally accepted in higher education and scientific research for the exchange and proper use of knowledge and information. There are two ways to create de-identified data under HIPAA:

    - *Safe Harbor De-Identification*: complete removal of 16 different identifiers and significant reduction in granularity of geographical information (retaining state and Zip-3, with a subset of Zip-3 codes being banned) and dates (retaining year only). Such a data set could be considered Public Use Data.

    - *Expert Method De-Identification*: certification by an expert that the risk that the data set, in combination with other available data sets, could be used to identify a person is very small; this involves removal of all direct identifiers as well as selective retention of a subset of indirect identifiers which can vary depending on the purpose of the data set. Recipients of data sets de-identified in this manner are often required to agree to conditions on their use of the data which are designed to further minimize re-identification risk.

ICPSR has expertise in handling PHI, and will accept it under a formal legal agreement in compliance with HIPAA. It is worth pointing out that not all individually identifiable health information is PHI, depending on its source and the circumstances of its creation, so some identifiable health data sets are not governed by HIPAA.

Data Release Options

R**esearchers should work with ICPSR staff** to determine 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 data set 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 (either virtually or in-person) to access the data under secure conditions. Below is a description of the various dissemination methods at ICPSR.

Data providers also have options to preserve a data collection without dissemination when significant potential for disclosure risk exists or when data are still in operational use by an organization. Delayed dissemination allows data to be deposited but not disseminated until an agreed-upon period of time has passed.

Public-use

Public-use files do not have direct or indirect identifiers, so disclosure risk is considered minimal. These data may be downloaded directly through the ICPSR website. Data users must agree to the ICPSR Terms of Use.

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 conditions of access. Specific information about applying for access to restricted-use data is available on the home page of each study within the Notes section. Most applications will require investigator information, research staff information, a research description, documentation of IRB approval or exemption, a Restricted Data Use Agreement (described in the following section), and a data security plan, among other possible requirements. Generally, the method of dissemination is related to the level of disclosure risk or the sensitivity of the data, and each method has a different application process and requirements.

Secure Dissemination

When a Restricted Data Use Agreement (RDUA) is approved, ICPSR staff with access to restricted-use data use ICPSR's secure server to prepare the data for the approved request. The data are encrypted and securely disseminated via a download by the researcher using a temporary link and password (provided to the researcher separately).

Virtual Data Enclave (VDE)

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. 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 all of the researcher's files and analytic output – 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. This includes being prohibited from taking photos, screenshots, transcribing results, and writing down notes. The user must request that ICPSR staff review output before it leaves the VDE to ensure the disclosure review guidelines are met before staff authorize and send the output files to the user. Users are strictly prohibited from removing or using any data that has not been authorized. Failure to follow any of the VDE protocols is considered a data breach and ICPSR will enforce the penalties described in the RDUA.

Physical Data Enclave (PDE)

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 (PDE) in Ann Arbor, Michigan. Researchers apply to use these data in the PDE. 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 reviewed by ICPSR staff for compliance with confidentiality standards. Once cleared, ICPSR staff send the materials to the researcher. Researchers are responsible for covering their travel and lodging while using the Physical Data Enclave.

Restricted Data Use Agreements

When data are determined to be restricted-use by ICPSR in consultation with the data provider, one or both of the following agreements are required to be completed, reviewed, and executed by the specified parties prior to the transfer of the data: a Restricted-Use Data Deposit and Dissemination Agreement (RUDDDA) and/or a Restricted Data Use Agreement (RDUA).

Restricted-use Data Deposit and Dissemination Agreement (RUDDDA)

This agreement is unique to the two parties engaged in the transfer of restricted-use data to ICPSR for processing and dissemination by ICPSR and is negotiated between those parties. When required, 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 to, stored at, and accessed from ICPSR. The RUDDDA denotes the conditions for release and dissemination of certain restricted-use data. The RUDDDA also determines the administration and content of the Restricted Data Use Agreement (RDUA, described below) to provide access to third-party data users. The RUDDDA also sets ICPSR's obligation to monitor and enforce all legal obligations specified in the Restricted Data Use Agreement (RDUA).

The main text of ICPSR's RUDDDA has been 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.

Restricted Data Use Agreement (RDUA)

The Restricted Data Use Agreement (RDUA) is completed by data users during the process of requesting 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 RDUA occurs. The RDUA is signed by the requesting investigator and the requestor's institutional representative. ICPSR staff make a record in ICPSR's online system when the request is approved. The RDUA is considered a unilateral agreement and considered executed once the requestor's institution signs. The main

Agreements in Brief

Restricted-Use Data Deposit and Dissemination Agreement

Allows ICPSR to receive and prepare the data for release and defines the conditions of data release; signed by the data provider's institution and the University of Michigan.

Restricted Data Use Agreement

Used by researchers to apply and obtain access to restricted-use data sets; signed by the researcher and an institutional representative of the researcher's institution.

text of the ICPSR RDUA has been reviewed by U-M's Office of Research and Sponsored Projects (ORSP).

Core Data Security Requirements

Restricted-use data require users to abide by any required data security requirements whether the data are used locally by users or in ICPSR PDE or VDE. These are generally physical, technical, and behavioral requirements around the user's access of the data. ICPSR's 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 data set, any derivatives, temporary files, and backup files
- Data storage technology (standalone computer, external hard drive, private networks, other types of networks, virtual data enclave, etc.)
- Changes in research project staff or institution
- Requests made by any outside entity to access the restricted-use data set
- How and when electronic and printed copies of the restricted-use data and any derivatives must be destroyed

Other optional data security requirements may apply depending on the specific nature of the restricted-use data.

Review, Administration, and Maintenance of RDUAs

After an application for restricted-use data is submitted, it is reviewed by ICPSR staff for approval. The amount of time it takes to review an application depends on whether – and how many – changes are required before approval. ICPSR staff will work with the requester to obtain all required information. Generally, it takes 2-4 weeks for a complete application to be approved after it is submitted.

If approved, the data are prepared for secure download by the approved Investigator or made available to be analyzed via the Virtual Data Enclave (VDE) or Physical Data Enclave (PDE) described above. 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.

After an application is approved, it is the Investigator’s responsibility to make sure the application remains up to date during the data access period. This may include adding or removing research staff, requesting additional data, uploading updated IRB documentation, submitting annual reports, and/or submitting a Data Destruction Affidavit or Request for Renewal before the data access period has ended.

ICPSR may update its Data Use Agreement templates. You can view a sample template, but please note that restricted data users must submit the specific agreement provided during the application process.

ICPSR Handling of Restricted-use Data

All ICPSR staff members who work with restricted-use data sets sign a confidentially pledge annually as a term of employment. For enhanced security, 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.

Files enter the SDE in an encrypted form through the ICPSR's Deposit Form. Only files that are authorized to be deidentified by ICPSR staff can be moved outside the SDE and this process is strictly audited. ICPSR staff must use a strong password to log on to the regular desktop and again to log on to the SDE. Desktops lock when the screen saver is activated. When ICPSR 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.

ICPSR maintains an Authority to Operate (ATO) with at least one Federal agency to run the repository securely and in compliance with government standards for security controls based on the risks of the data archived. Public-use data have "low" FISMA severity categorization while restricted-use data are "moderate." ICPSR performs annual risk assessments to ensure compliance with NIST 800-53, as well as monthly vulnerability scans of our public, private, and restricted networks and compute resources.