

Protecting Confidential Data at the National Archive of Criminal Justice Data

The National Archive of Criminal Justice Data [NACJD] employs three methods to protect the confidentiality of subjects while distributing data collections for secondary research and other statistical analysis. These three methods, each more fully described below, include: (1) masking direct and indirect identifiers; (2) requiring all users to agree to follow Federal laws and regulations, and scientific standards before accessing data; and, for more sensitive data, (3) limiting the level of access. While methods 1 & 2 are applied to every NACJD data collection, the level of access restriction employed for a data collection or part of a collection is determined by NACJD's Director in cooperation with the data depositor/principal investigator, the sponsor of the data collection, and the sponsor of the data archiving. The access decision occurs after NACJD staff have performed several procedures to evaluate and reduce the risk of identifying research subjects (e.g., respondents and institutions) in the data. First, the study is assessed to determine the source and nature of the data (e.g., public use information vs. self-report; agency vs. individual). Second, all direct, personal identifiers are masked, regardless of the nature of the data, as part of the process of preparing data for release. Third, other variables such as addresses, dates or combinations of characteristics (e.g., indirect identifiers) are recoded (e.g., categories are collapsed) or masked to prevent (re)identifying the subjects. If this evaluation determines that the subject's confidentiality is adequately protected, the entire data collection is made available through the internet as *Restricted Use* data. Two alternative levels of protection are available for releasing NACJD data when reidentification is still reasonably possible or when our processing procedures will seriously reduce the analytic value of the collection. These two are *Restricted Access* and *Enclave Restricted Access*.

Restricted Use

This provides the easiest and quickest avenue for accessing most NACJD data. However, this mechanism does not allow unrestricted use (i.e., public use) by researchers or others obtaining these data. NACJD protects subject confidentiality by requiring data users to sign our online *data use agreement* (by inserting their email address) before they can download any study and most documentation. By signing the *data use agreement*, NACJD users signify their compliance with Federal laws, regulations and requirements, and give assurance that their use of NACJD data will conform to widely-accepted standards of practice and legal restrictions intended to protect the confidentiality of research subjects. The applicable laws and regulations are found in the United States Code, 42 USC Section 3789g(a), and the Code of Federal Regulations, 28 CFR 22. These laws and regulations in part require that the confidentiality of data about research subjects, including their identities, be safeguarded. Any intentional identification or disclosure of a person or establishment by a data user violates the assurances of confidentiality given to the providers of the information and can result in fines and a prison term. Prior to downloading data, NACJD data users are also instructed to: 1) make no use of the identity of any person or establishment discovered inadvertently, and advise the Director of NACJD, of any such discovery; 2) produce no links among NACJD datasets or between NACJD datasets and other datasets for the purpose of disclosing the identities of individual subjects; and 3) not to redistribute or sell NACJD data to other individuals, institutions, or organizations without the written agreement of NACJD.

Restricted Access

This level is used when an entire study, or part of a study, contain data that can not be reasonably

altered to protect subject confidentiality without significantly degrading the value of the research data. Such studies may contain self-report information from a universe of subjects within a known geographical area or multiple measurement periods of the same subjects. These studies, or parts of these studies, are not accessible from NACJD through the Internet. Access to these collections is arranged after the prospective data user completes and submits a *Restricted Data Use Agreement*. The *Restricted Data Use Agreement* requires data users to certify in writing that the NACJD data will be used only for research or statistical analysis, and that they will protect the confidentiality of subjects in data. In addition, they must describe: 1) why they need this version of the data rather than the internet accessible *Restricted Use* version, and 2) how they plan to physically secure the data collection and protect it against virtual theft. The request must be signed by a researcher with a permanent faculty appointment, and by a representative (e.g., Dean or President) of the researcher's university or other institution. The completed *Restricted Data Use Agreement* must also be accompanied by a letter from the researcher's Institutional Review Board office, which indicates that a responsible officer at the institution has reviewed the project and finds the project's terms of use for the requested data to be in compliance with their organization's human subjects and confidentiality rules. The final determination on access to these *Restricted Access* studies rests with NACJD's Director. After approval, the collection is sent on CD-ROM to the requestor via registered United States Postal Service mail. From the time of the mailing, the user is granted a one-year limited access license. After one year, they are required to destroy any copies of data and return the CD-ROM to NACJD, or reapply for a license extension.

Enclave Restricted Access

While most NACJD data are available by the previous two methods, some collections cannot be completely deidentified (e.g., video) or may still be too sensitive (e.g., narratives) to release to researchers without authorization and supervision. These data collections are only accessible in our onsite data enclave. Researchers can apply to use these data, without a fee, but are responsible for covering their travel and lodging while at The University of Michigan, the site of the NACJD data enclave. After approval, researchers are allowed access to these NACJD data on a computer that is not connected to a network. All written notes, data analysis printouts and electronic files are reviewed by NACJD staff for compliance with confidentiality standards before the researcher can take these items from the NACJD enclave. As time passes and risks are reduced, NACJD may choose to move data from the *Enclave* or from the *Restricted Access* protection level into a more accessible level.

Delayed Dissemination

When significant potential for disclosure exists or when data are still in operational use by an organization, NACJD may also elect to preserve a data collection without dissemination until human subject risks are minimized (e.g., when all human subjects are deceased). NACJD will work with the data collector and sponsors to plan a dissemination timetable. While dissemination is delayed, NACJD will still produce a copy: (1) to develop expertise with the data collection, (2) to prepare the collection while project staff are available to provide assistance, and (3) to safeguard the original data collection. Once released, NACJD will use one of the three access levels above to disseminate the processed data.