

Policy and Procedures Manual for ICPSR Data Enclave

Data Enclave Security.....	2
Data Enclave Computing Environment	2
Data Enclave Staffing	3
Web Text re: General Information on the Use of ICPSR Data Enclave.....	3
Disclosure Review Process.	7
Appendix A	9
Appendix B	10
Appendix C	11
Appendix D	12

Data Enclave Security

A. Electronic Surveillance

The ICPSR secure data enclave is located in the basement of the Perry Building in Rooms B230 and B230A. Note that B230A is an inner office that contains the server, confidential paper files, multimedia, and other secure storage.

Electronic surveillance consists of two keycard swipe readers with punch code access, one each for the exterior and interior doors. Additionally our current building-wide camera/video monitoring system provides a camera in the hall area that shows traffic in the hall entrance closest to the enclave.

After the data enclave has been in operation for a year, security measures will be reevaluated. Possible enhancements to electronic surveillance could include video cameras in each of the two enclave rooms and in the hallway pointing to the entry door, a video camera monitor located at the monitor work station, and motion detectors in each of the two rooms. Possible enhancements to the physical security of the enclave include enclosing the monitor work space.

B. Physical layout concerns including ceiling access

The university's plant officials reviewed the ceiling for possible entry. Because of the number of beams, pipes, and other construction materials they consider it a secure ceiling.

C. Lockers

Lockers located outside the enclave will be provided for researchers to securely stow their personal belongings while they are in the enclave.

D. Housekeeping

Housekeeping will not have routine access to the data enclave. Rather, cleaning will be done by special request of the monitor and only when the monitor is present.

Data Enclave Computing Environment

The data enclave hosts two researcher workstations, one monitor workstation and one audiovisual cubicle. The server is secured in B230A. A printer is located in the monitor's work area.

The workstations are Dell GX270 and the server is a DELL PowerEdge 2600 with a tape backup system. The workstations are configured with the ICPSR baseload, which should be adequate for routine analysis. Specialized software and hardware will need to be purchased to accommodate specialized types of analysis (e.g., elaborate GIS models, video capture). There is also a battery back-up power supply for the server. USB ports and floppy drives are disabled on the researcher workstation to preclude the use of writable removable media. The monitor workstation is fully functional so that the monitor can write to removable media as required.

The audiovisual cubicle will be equipped with a VCR, DVD player, monitor, and headphones for those researchers who wish to review multimedia material.

ICPSR has established a closed network in the data enclave. All restricted data and user files will be stored on the enclave server. A unique password-protected profile is created for each user that allows the user to access the restricted files specified in the application. Monitors will have read permission to user files for the purpose of printing output or to transfer files to removable media.

Data Enclave Staffing

Initially the data enclave will be staffed by the data enclave manager. Topical and general archive staff will act as disclosure risk reviewers. The data enclave manager will meet the researcher upon his/her arrival at ICPSR, review enclave policies, demonstrate enclave computing facilities, and review what types of analysis output will be allowed to eventually leave the enclave. The data enclave manager will be present when the enclave is in use and verify that entry and exit procedures are being followed. The manager will not provide substantive consultation regarding the research being done. Other staff, however, having expertise in working with the particular study, will be available for consultation both prior to the researcher's visit and during his/her visit.

The major responsibility of the topical and general archive staff, generally at the Research Associate level, is, as mentioned, to consult with the researcher prior to his/her visit, to respond to questions during his/her visit (outside the enclave), and to meet with the researcher outside the enclave prior to his/her departure. The purpose of that final consultation is to once again review the procedures for allowing research results to leave the enclave, to learn the structure of the researcher's files, and to know exactly what output the researcher wishes to take from the enclave. No materials will leave the enclave at this time. Output materials will be sent to the researcher after a two-to-three week period of review at ICPSR.

The topical and general archive staff will have responsibility for reviewing disclosure risk in the researcher's output. A checklist should be used when reviewing the results. After reviewing the output and completing the checklist, the staff member will make a recommendation regarding whether to release the output or modify it. That recommendation will be forwarded to the manager of the relevant archive for review; the archive manager will make the final decision regarding the disclosure review. Any proposed adjustments to the output will be discussed with the researcher via telephone or e-mail by the archive manager.

(See Disclosure Review Process section for information on forms that must be completed by the researcher and by the staff and manager performing the disclosure review.)

General Information on the Use of the ICPSR Data Enclave

NOTE: The information in this section appears on our website.

ICPSR Bylaws state that "data and other material provided are to be used solely for statistical analysis and reporting of aggregated information, and not for investigation of specific individuals or organizations" and requires all users of its materials to give "assurance that such uses of statistical data will conform to widely-accepted standards of practice and legal restrictions that

are intended to protect the confidentiality of research subjects." (ICPSR Bylaws, Article I, Section 2.C)

Each data collection in the ICPSR Archive has been examined to ensure that the contents of the collection do not violate explicit or implicit pledges of confidentiality given to respondents or research subjects. Data items that could be used to identify individual respondents are typically removed, masked, or collapsed in the public-use versions of the datasets prepared by ICPSR and released to the research community.

However, not all research questions can be addressed with public-use versions of data. Those versions may not have sufficient detail to adequately answer research questions or, due to the confidential nature of the data, there may not be a public-use release version. The ICPSR Data Enclave will permit researchers access to data that are not available to the general public because of respondent confidentiality concerns, under circumstances that provide sufficient controls on the researchers' use of the data and sanctions for violation of respondent confidentiality so as to reasonably ensure the continued confidentiality of respondent identifying information.

The ICPSR Data Enclave is located at the ICPSR offices in the Perry Building, 330 Packard Street, Ann Arbor, MI 48106-1248 and provides a secure environment in which researchers can access and analyze restricted data.

A. Application Procedure

Before a researcher applicant can be granted access to restricted data, a research proposal containing the information listed below must be submitted and approved.

Cover letter

Project title

Abstract

Researcher personal and affiliation information

IRB approval from the sponsoring institution

Current resume or vitae

Dates of proposed use of the data enclave

An estimation of the total hours expected to be working in the data enclave

Funding source(s), if any, for user project

A detailed summary of the proposed research including a description of why the publicly available files (provided they exist) are not adequate for the research plan

A complete list of data enclave files being requested

A description of user supplied data, if any, to be merged with restricted data including documentation, file layout, and number of records.

A list of software requirements (application will lists those routinely provided)

A description of proposed analysis and expected analysis results, including a description of all information that is expected to be retained for subsequent use

Description of intended use of research results, including plans for public dissemination

A list of special needs requirements (if any)

Two signed copies of the Confidentiality Agreement for Use of ICPSR Enclave Data

B. Data Enclave Use Policies

Application review process

The application review process consists of three steps. The application is initially reviewed by dedicated data enclave staff to ensure all requested information has been supplied. Staff assesses any data manipulation services requested and additional software needs, along with the proposed use dates for the data enclave. Once these are addressed, the application is forwarded to the ICPSR unit manager who is most familiar with the requested restricted data. The manager will recommend approval or rejection and the recommendations will be incorporated into the application. The application is then reviewed by a committee consisting of three faculty researchers who are knowledgeable about the proposed research and appointed by the ICPSR Director. The committee will review all aspects of the proposal and communicate directly with the researcher as needed. Once the application is approved and signed, data enclave staff will contact the researcher. (Appendix A is a letter from the ICPSR Council Chair authorizing specific staff to sign restricted use data agreements on behalf of Council.) The entire review process is expected to take four to six weeks.

User charges

ICPSR's policy is not to charge for use of the data enclave, with one exception. If the researcher supplies data that must be merged with enclave data, ICPSR would assess a fee to recoup some of the cost of the ICPSR staff's time that would go into this activity.

User-supplied data

When supplying his/her own data, the user must consult with enclave staff prior to using the data enclave to ensure that their user-supplied data can be merged with restricted data. The format must be consistent with that of the restricted data (ascii ASCII files, not system files). The data must be supplied in advance of the visit and the merging completed prior to the researcher's use of the data enclave.

The user-supplied file, the merged file, and information used in linking it with the restricted data will be saved to removable media for the researcher for a period of two years solely to assist the user in recreating this information in the event it is lost or destroyed. These files will physically reside in the data enclave and may be destroyed prior to the two-year period to the extent permitted by law upon written request of the researcher or where ICPSR otherwise determines that destruction of the files is appropriate or necessary to ensure confidentiality or for other reasonable business purposes.

Role of the data enclave monitor

ICPSR must be concerned with maintaining physical security of the data enclave and with maintaining the confidentiality of respondents in the data sets being analyzed. The data enclave monitor will be present in the enclave at all times when the enclave is in use by a researcher. Researchers cannot bring any electronic devices including laptops or other portable computing devices into the data enclave. All materials brought in are subject to review by the monitor, and the decision of the monitor is final.

Disclosure review of output

Absolutely no papers, notes, printouts, computer media or other materials can leave the enclave without first being examined for disclosure risks. The researcher will work closely with a research associate to document the analyses performed and the location of work files on his/her designated directory. At a later time, all the analysis output will be reviewed by the applicable ICPSR unit archive manager for disclosure risks. Researchers can expect a recommendation on the release of their output within two to three weeks. Approved analysis output will be sent to the researcher on media supplied by ICPSR.

ICSPR will keep a copy of all output that has been reviewed for disclosure risk in a locked cabinet in the data enclave for a period of two years for the primary purpose of assisting the user in recreating this information in the event it is lost or destroyed. Output files may be destroyed prior to two years from initial creation to the extent permitted by law upon the written request of the researcher or where ICPSR otherwise determines that destruction of the files is appropriate or necessary to ensure confidentiality or for other reasonable business purposes.

C. The Secure Data Enclave Environment

The network in the data enclave is isolated from other networks which mean that the workstations do not have e-mail or Internet access. A unique password-protected profile is created for each end data enclave user that allows the user to access the restricted data files specified in the application from these workstations. Access to the data enclave printer is restricted to the data enclave monitor.

Each workstation has a high speed Pentium computer with Windows XP, Microsoft Office Suite and SPSS, SAS and Stata statistical packages. All removable media access ports have been disabled.

Researchers cannot bring anything electronic devices, including laptops or other portable computing devices into the data enclave. All materials brought in are subject to review by the data enclave monitor, and the decision of the monitor is final. Lockers are provided outside the enclave for secure storage of personal belongings.

Data enclave users will have access to a workstation outside the enclave for e-mail and telephone access. Typically, the enclave is open during normal business hours, i.e., Monday through Friday, 8:00 -6:00, although extended hours can be arranged in advance. No food or drink is allowed in the enclave.

D. Violation of Agreement

If ICPSR determines that any aspect of the Confidentiality Agreement for Use of ICPSR Data Enclave has been violated, ICPSR may invoke these sanctions as it deems appropriate:

- Denial of all future access to ICPSR Enclave Data
- Withholding of any output or related files not yet delivered to the Researcher
- Report of the violation to the Researcher's Institution's office responsible for scientific integrity and misconduct, with a request that the institution's sanctions for misconduct be imposed

- Report of the violation to appropriate federal and private agencies or foundations that fund scientific and public policy research, with a recommendation that all current research funds be terminated, and that future funding be denied to the Researcher and to all other persons involved in the violation
- Such other remedies that may be available to ICPSR under law or equity, including injunctive relief

Disclosure Review Process

Potential data enclave users will be provided with two additional documents. The first pertains to general information about documenting the statistical output that the researcher wishes to remove from the enclave and the length of time ICPSR will retain that output (see Appendix B). The second is the Request Disclosure Review Clearance Form that must be completed by the researcher and reviewed with the appropriate topical archive staff before the researcher leaves the site (see Appendix C).

Staff members reviewing output for disclosure issues and supervisors reviewing those recommendations must use the ICPSR Disclosure Review Checklist and Clearance Form (see Appendix D). It contains the disclosure risk checklist information presented below and provides a paper trail for decisions regarding release or retention of the researcher's output. If the output is to be released, the form is forwarded by the supervisor to the enclave manager for shipment of data and final sign-off.

Note that these three forms, and most importantly the disclosure risk review checklists, are initial attempts to identify areas of concern regarding disclosure risk and to document decisions made internally regarding release or retention of the output. It is expected that all forms will be expanded and modified with use.

Disclosure Risk Review Checklist.

The area of disclosure risk review is evolving. There is no one authoritative document fully describing the topic. There appears to be a short definitive listing of what does and does not pose a confidentiality risk, and a longer list about conditions/situations that should be considered when reviewing output with disclosure risks in mind.

Items to check for when reviewing output should include, but are not limited to:

1. Direct identifiers cannot be allowed in the output results. However, what constitutes a direct identifier may not be obvious. For example, in the criminal justice archive the variable ORI Code is common. Similarly, a nine-digit numeric Census code uniquely identifies local government agencies. This is why it is critical that topical archive staff do the reviews with respect to their data and that for General Archive studies, staff knowledgeable about a given study perform the review. ICPSR should compile a list of common identifiers used across all archives for use in training monitors.
2. All output tables must be completely labeled including titles, variable labels and value labels. A disclosure risk reviewer should not be expected to guess what the output means.
3. Variables used in analyses should not be renamed from their source data.

Created variables must be fully labeled.

4. No cell in a table should include maximum or minimum values.

Considerations to keep in mind when reviewing output for potential breaches of confidentiality include, but are not limited to:

1. Limitations on cell size in tables should be established.
2. Cells containing 100% responses can be as dangerous as low cell counts. They possibly indicate a rare condition.
3. A maximum number of dimensions for tables should be established. More than three dimensions should raise concerns.
4. Geographic variables pose significant risk to confidentiality. ICPSR must determine a policy regarding the lowest level of geography disclosure it will allow. Some organizations will not allow results below Census division. Others (e.g., MiCDA) restrict geography to that in public-release files.
5. Longitudinal data also poses an increased risk. The sheer number of variables recording a person's life events makes data records more unique than with cross-sectional data. It may be appropriate to increase cell size in tables that use longitudinal data.
6. Merging secure data to external sources also increases the uniqueness of the resulting records. Data that are merged prior to use at the enclave are more easily assessed for disclosure risk. Assessing the disclosure risk from potential merging after leaving the enclave is more difficult. Knowing the main identifier variables in a field will help assess this risk.
7. Sampling can create a disclosure risk. Generally simple random sampling is less risky than cluster sampling. ICPSR should determine whether to allow stratum codes to be released. Large sampling frames may effectively create a census.
8. Weights can pose a disclosure risk. Tables should be reviewed on an unweighted basis but released only on a weighted basis. Weight variables should not be released.
9. The release of individual level data is not allowed; only aggregated results can be released.
10. ICPSR should consider whether to require that all analyses be performed in syntax files so that results can be verified. However, since this may severely conflict with the work practices of data users ICPSR will formulate the policy on this practice at a later date.

MM 3/30/06

Appendix A

March 10, 2006

P.O. Box 1248
Ann Arbor, MI 48106-1248

netmail@icpsr.umich.edu
www.icpsr.umich.edu

(734) 647-5000 voice
(734) 647-8200 fax

ICPSR COUNCIL

Mark Hayward, Chair,
Pennsylvania State
University

Darren W. Davis,
Michigan State
University

Iлона Einowski,
University of
California, Berkeley

Charles H. Franklin,
University of Wisconsin

John Handy,
Morehouse College

Paula Lackie,
Carleton College

Nancy Y. McGovern,
Cornell University

Samuel L. Myers Jr.,
University of Minnesota

James Oberly,
University of
Wisconsin, Eau Claire

Ruth Peterson,
Ohio State University

Walter Piovesan,
Simon Fraser University

Ronald Rindfuss,
University of North
Carolina, Chapel Hill

Ann Green, Past
Chair, Yale University

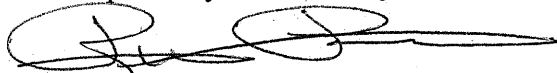
To Whom It May Concern

This letter, endorsed by the entire Council of the Inter-university Consortium for Political and Social Research (ICPSR), authorizes the individuals named below to sign data use agreements on behalf of the ICPSR Council. Those agreements (which include restricted data agreements as well as secure data enclave access agreements) describe mutually agreeable conditions under which researchers may use elements ^{from} the ICPSR data archive for their secondary data analysis projects. The ICPSR Council is a governing body elected by and representing the 550 institutional members of ICPSR, and exercises authority over the social science data archive established and maintained on behalf and for the benefit of the member institutions over the past four decades.

The individuals so authorized to sign these agreements are:

Myron P. Gutmann (Director, ICPSR) ,
Erik W. Austin (Assistant Director, ICPSR)
Christopher D. Maxwell (Associate Research Scientist, ICPSR)
Peter A. Granda (Assistant Director, Collection Development, ICPSR)

Respectfully submitted by



Ruth D. Peterson (Ohio State University)
Chair of the ICPSR Council

Cc: Office of the General Counsel, University of Michigan
James S. Jackson (Director, Institute for Social Research,
University of Michigan)

Appendix B

Documenting Your Research

You must provide documentation for any output (tabulations or models) you wish to remove from the enclave. Many forms are acceptable; most will include statistical programs, logs and listings, sometimes including spreadsheets or datasets containing tables, parameter estimates and the like. In properly documented output, all variables and variable values should be LABELED and all variables should be in a consistent format, .e.g., identical variables, such as industry (IND), should have the same name and same formats (numeric or alphanumeric) across datasets. All tables and output (research output or disclosure analysis output) should have clear TITLES indicating what the output is and indicating whether the analysis used weighted or unweighted data (and what the weight variable is).

Moreover, and very importantly, you should trace the evolution of your data sets from the initial data set(s) ICPSR provided to the data sets you used in generating your statistical output/research results. This trail should include programs and data set names and should be sufficiently detailed that ICPSR disclosure reviewer could reproduce your work. Formulas for any computed variables should also be provided.

This information will be required on the Request Disclosure Review Clearance Form and must be thoroughly reviewed when you meet with the ICPSR disclosure reviewer. The clearer and more complete the form, the easier it will be for the disclosure reviewer to understand your research and conduct his/her disclosure review.

Backing Up and Archiving Data and Programs

As mentioned in the enclave handout, any and all files that you want saved need to be copied to XXXXX. Files not copied to XXXX will be deleted at the end of your session. The disclosure reviewer will have access to the saved files for purposes of conducting the disclosure review analysis. Make sure that any files that support your output and the disclosure review are copied to this directory. The enclave system administrator will archive files copied to XXXXX for a period of two years. At the end of two years the files will be destroyed without further notification to you. Paper output that is sent to you pursuant to the disclosure review will also be kept in a locked cabinet for two years. At the end of two years it will be shredded.

Requesting Disclosure Review Clearance

When your research is completed, fill out the Request Disclosure Review Clearance Form, which can be found at XXXX, and ask to meet with the designated ICPSR staff performing disclosure analysis to discuss your clearance request.

Appendix C

Request Disclosure Review Clearance Form

This form must be completed by the researcher, printed by the monitor, and reviewed with the ICPSR Disclosure Reviewer prior to the researcher leaving the ICPSR Data Enclave. The form provides for the documentation of datasets and programs for any output the researcher wishes to remove from the ICPSR Data Enclave.

Date:

Name:

University Affiliation:

Research Project Title:

Name of Restricted Study Used:

List of research output files you want to remove from the enclave:

For each of the files listed above,

-provide documentation on how it was generated, including SAS/SPSS programs and logs, recodes, etc.

-provide description of any research sub-sample used

-provide glossary of variable names

-indicate whether a weight variable was used, and if so, the weight variable name

Appendix D

ICPSR Disclosure Reviewer Checklist and Clearance Form

Date:

Name of Disclosure Reviewer:

Name of Enclave Researcher:

Name of Enclave Research Project:

Name and Study Number of Restricted Study Used:

Checklist to use in reviewing output:

The area of disclosure review is evolving. There is no one authoritative document fully describing the topic. There appears to be a short definitive listing of what does and does not pose a confidentiality risk, and a longer list about conditions/situations that should be considered when reviewing output with disclosure risks in mind.

Items to check for when reviewing output:

1. Direct identifiers cannot be allowed in the output results. However, what is a direct identifier may not be obvious. For example, in the criminal justice archive the variable ORI Code is common. Similarly, a nine-digit numeric census code uniquely identifies local government agencies. This is why it is critical that topical archive staff do the reviews with respect to their data and that for general archive studies, staff knowledgeable about the study perform the review.
2. All output tables must be completely labeled including titles, variable labels and value labels. A disclosure reviewer should not be in the position of having to guess what output means.
5. Variables used in analyses should not be renamed from their source data. Created variables must be fully labeled.
4. No cell in a table should include maximum or minimum values.

Considerations to keep in mind when reviewing output for potential breaches of confidentiality.

1. Limitations on cell size in tables should be established.
2. Cells containing 100% responses can be as dangerous as low cell counts. They possibly indicate a rare condition.

3. A maximum number of dimensions for tables should be established. More than three dimensions should raise concerns.
4. Geographic variables pose significant risk to confidentiality. ICPSR must determine a policy regarding the lowest level of geography disclosure it will allow. Some organizations will not allow results below census division. Others (e.g., MiCDA) restrict geography to that in public release files.
5. Longitudinal data also poses an increased risk. The sheer number of variables recording a person's life events makes data records more unique than with cross-sectional data. It may be appropriate to increase cell size in tables that use longitudinal data.
6. Merging secure data to external sources also increases the uniqueness of the resulting records. Data that are merged prior to use at the enclave are more easily assessed for disclosure risk. Assessing the disclosure risk from potential merging after leaving the enclave is more difficult. Knowing the main identifier variables in a field will help assess this risk.
7. Sampling can create a disclosure risk. Generally simple random sampling is less risky than cluster sampling. ICPSR should determine whether to allow stratum codes to be released. Large sampling frames may effectively create a census.
8. Weights can pose a disclosure risk. Tables should be reviewed on a weighted and unweighted basis but released only on a weighted basis. Weight variables should not be released.
9. The release of individual level data is not allowed; only aggregated results can be released.
10. ICPSR strongly encourages all analyses be performed in syntax files so that results can be verified.
11. In no table should all case in any row or column be found in a single cell.
12. In no instance should the total number of cases be based on fewer than X cases.
13. In no instance should a quantity figure be based on fewer than X cases.
14. In no instance should a quantity figure be published if one case contributes more than 60% of the amount.
15. In no instance should data on an identifiable case, nor any of the kinds of data listed in the preceding items, be derivable through subtraction or other calculation from a combination of tables released.
16. Data released should never permit disclosure when used in combination with other known data.

Based on your disclosure review, is there any indication that respondent confidentiality is breached?

No _____

Yes _____ Give explanation and turn the matter over to your supervisor, after signing and dating the form below.

Based on your disclosure review, should the requested output files be supplied to the researcher?

Yes _____

No _____ Give explanation and turn the matter over to your supervisor, after signing and dating the form below.

Based on your disclosure review, do you recommend that your supervisor allow the requested files to be supplied to the researcher?

Yes _____

No _____

Your signature and date:

Supervisor, please review and signoff on the release of the requested output files, or indicate why the files should not be released.

Yes, release output _____

No, do not release output _____ Give explanation

Supervisor signature and date:

Supervisor: Please return this form, or a copy of it, to Mary Morris so that, if appropriate, the output can be sent to the researcher. The completed form will reside in the researcher's data enclave application folder.

Enclave administrator

____ Electronic files sent to researcher on cdrom. Itemize files copied to cd:

Administrator signature and date of shipment:

____ Paper output sent to researcher. Itemize what was sent and make a copy for our records.

Administrator signature and date of shipment: