Policy Regarding Export of Microdata from the MiCDA Data Enclave

Purpose

By default, users of the MiCDA Data Enclave are not allowed to export respondent or establishment variables created in the course of their work with restricted data products:

> "Export of microdata files or analysis output containing information at the respondent level is not allowed unless specific permission has been obtained from the Michigan Center on the Demography of Aging and all relevant restricted data providers."[1]

The purpose of this policy is to provide:

- A framework for users wishing to export, as public data, respondent or household-level variables derived from restricted data products maintained by the Data Enclave.
- Instructions for Enclave staff members charged with reviewing user requests.
- Documentation of the review process for use by the Data Confidentiality Committee and others.

Scope

This policy covers all users of the Enclave whose research activities are governed by the *Confidentiality Agreement Restricting Disclosure and Use of Data from the Michigan Center on the Demographics of Aging Data Enclave*.

Definitions

Restricted variable: Any variable that is included in a restricted data set
Public variable: A variable included in a data set that is distributed to the general public for use without special conditions.
Candidate variable: A variable derived from restricted and/or public data that is being proposed for release as a public variable.[2]

Process

Before a variable derived from restricted data may be released from the Enclave environment, the user, Enclave staff, and restricted data provider must carry out the following steps:

1. The user carries out the data management steps necessary to create a candidate "public" variable from restricted inputs.

2. The user makes a formal request to the Enclave manager for review of the candidate variable(s) in question; this request should include:
    a. A descriptive narrative of how each variable was created.

[1] *Rules for Pre-Export Disclosure Review*
[2] Restricted data sets are "viral" in that the result of any merge between a public data set and a restricted data set is defined as restricted.

   b. The Stata do file(s), SAS command file(s), or other program(s) used to create the candidate variable(s).
   c. A data file composed of the candidate variable(s) and associated record identifier(s) to be released.

3. The Enclave Manager conducts a review of possible disclosure risks posed by the candidate variable(s) to be released; this review should include:
   a. Analysis of the type/structure of the restricted data used to create the candidate variable(s).
   b. Review of user needs; will an existing public variable serve the same purpose as the proposed variable.
   c. Review of possible disclosure scenarios, including a realistic assessment of attack feasibility.
   d. Risk assessment:
      i. Does this variable give the attacker more information/attack surface/leverage than what is already available to the public?
      ii. Is any type of personally identifiable information (education, financial, HIPAA) likely to be exposed?
      iii. Is any type of confidential proprietary or establishment information exposed?
   e. Determine necessity of obtaining data provider permission prior to release. [3]

Outcome

1. If the outcome of Step 3 is satisfactory, the Enclave Manager:
   a. Prepares a memorandum outlining the result of the disclosure risk review.
   b. Forward the user's request to the Data Confidentiality Committee (via the Working Group) for review and final approval.
   c. Informs the user that the request has been forwarded.

2. Upon receipt of DCC approval, the Enclave Manager will assist the user in exporting the candidate variable(s) as public variable(s) from the user's Enclave folder.

3. If problems are found in conjunction with the execution of Step 3, then the Enclave Manager and the user will attempt to identify disclosure limitation methods that will allow the candidate variable(s) to be released. These methods may include:
   - Rounding.
   - Top/bottom coding.
   - Global recoding.
   - Local suppression.
   - Numerical rank swapping.
   - Numerical micro aggregation.[4]

4. It is the responsibility of the user to implement the disclosure limitation method and re-submit the data set for review.

[3] For further information on risk assessment, see *Handbook on Statistical Disclosure Control v1.2*
[4] For examples of how these methods might be implemented, see the Micro-ARGUS v4.2 User's Manual