

## **Appendix G**

# **Linkages to Extant Data Sources**

---



## G.1 Develop Linkages with Extant Data Sources

RTI will develop linkages with existing data sources to supplement the 2008 National Postsecondary Student Aid Study (NPSAS:08) interview data. RTI recognizes the great value added to the NPSAS:08 data file with the addition of data from specific extant data sources. Certain data (for example, specific financial aid amounts and associated dates) can only be accurately obtained from sources other than the student or parent. Through our experience providing data for many National Center for Education Statistics (NCES) postsecondary studies, including previous NPSAS studies, Beginning Postsecondary Student (BPS), and Baccalaureate and Beyond (B&B), we have gained considerable knowledge performing file merges with many existing sources of valuable data, including Department of Education's (ED) Central Processing (CPS) for Free Application for Federal Student Aid (FAFSA) data, the National Student Loan Data System (NSLDS), and ACT. For this study, we propose to perform file merges with the following datasets: CPS, NSLDS, American College Testing (ACT), SAT (the College Board), and National Student Clearinghouse.

The Family Educational Rights and Privacy Act (FERPA), (34 CFR Part 99) allows the disclosure of information without prior consent for the purposes of NPSAS:08 according to the following excerpts: 99.31 asks "Under what conditions is prior consent not required to disclose information?" and explains in 99.31 (a) "An educational agency or institution may disclose personally identifiable information from an education record of a student without the consent required by 99.30 if the disclosure meets one or more of the following conditions:

- (6)(i) The disclosure is to organizations conducting studies for, or on behalf of, educational agencies or institutions to:
  - (A) Develop, validate, or administer predictive tests;
  - (B) Administer student aid programs; or
  - (C) Improve instruction."

In addition, 99.31 (a)(6)(iv) specifies: "For the purposes of paragraph (a)(6) of this section, the term *organization* includes, but is not limited to, Federal, State, and local agencies, and independent organizations." All three parts of 99.31(a)(6)(i) apply to NPSAS:08, including (A), (B), and (C). The purposes of the NPSAS:08 field test is to create a research data set for a large sample of students that brings together information about federal, state, and private aid programs, and gathers additional demographic and enrollment data to establish the appropriate context. The resultant data set allows research and policy analysts to address basic issues about the affordability of postsecondary education and the effectiveness of the existing financial aid programs.

NOTE: 99.31 (a) (6) further specifies that such disclosure may only occur if the study is conducted in a manner that does not permit personal identification of parents and students by individuals other than representatives of the organization; and the information is destroyed when no longer needed for the purposes for which the study was conducted.

Additionally, the study, including the extant data linkage, qualifies for a 45 CFR 46 waiver of consent based on the following factors:

- There is minimal risk to the participants. There is no physical risk and only minimal risk associated with linkage of data to sample members. The public-use and restricted-use data, prepared as part of our contract, will not include Social Security

numbers (SSNs), even though the SSN is used for the linkage. Data will undergo disclosure avoidance analysis and disclosure treatment steps to further reduce the risk.

- The waiver will not affect the rights and welfare of the subjects. The voluntary nature of the study is emphasized to sample members. Public-use and restricted-use data are only used for research purposes and lack direct individually-identifying information. The data are further protected through disclosure avoidance procedures approved by the NCES Disclosure Review Board.
- Whenever appropriate, subjects will be provided with additional pertinent information after they have participated. For each round of the study, information about prior rounds and the nature of the study is made available to sample members.
- The study cannot be conducted practicably without the waiver. To obtain written consent from sample members, multiple forms would have to be sent to the sample members with multiple follow-up telephone and in-person visits. This process would add weeks to the data collection process and is not feasible from a time standpoint. Additionally, the value of these data would be jeopardized from a nonresponse bias perspective.
- The potential knowledge from the study is important enough to justify the waiver. These linked data for NPSAS:08 will provide invaluable data to researchers and education policy makers about the federal financial aid that students have received, as well as critical information concerning access to and persistence in postsecondary education. Rather than ask students for information about financial aid, we are getting it from the NSLDS, which is ED's system of recording federally aided student loans taken out by students and grants received. In RTI's experience, students are generally not a very reliable source of information about the amounts or timing of grants and loans they have received. This administrative record data is accurate and much easier to obtain than by administering a questionnaire.

Some information related to FERPA may be found at:

<http://www.ed.gov/policy/gen/guid/fpco/ferpa/index.html>.

As noted above, 99.31 in 34 CFR Part 99 (Family Educational Rights and Privacy, or FERPA) allows the disclosure of information without prior consent for the purposes of NPSAS:08:

(99.31) (6)(i) The disclosure is to organizations conducting studies for, or on behalf of, educational agencies or institutions to:

- (A) Develop, validate, or administer predictive tests;
- (B) Administer student aid programs; or
- (C) Improve instruction.

These sections, (6)(i) (A), (B), and (C), apply to NPSAS:08.

As part of initial sampling activities, we will ask participating institutions to provide SSNs for all students on their enrollment list. Having an initial list of all student SSNs minimizes the time and burden on both the institutions and RTI. Institutions will only have to provide one enrollment list; if the SSNs were provided only for those students selected, the institutions would

have to provide two separate enrollment lists. Immediately after the student sample is selected, the SSNs for non-selected students will be securely discarded.

**Secure Data Transfers.** NCES has set up a secure data transfer system, using their NCES member site with Secure Socket Layer (SSL) technology, described above. We will use this electronic system for submitting data containing potentially identifying information (such as SSNs, names, and dates of birth of our sample members) along with their survey ID (not the same ID that is available on the restricted-use data). As an additional precaution, we will use WinZip version 10, 256-bit AES encryption and a strong password to further protect the file. We will receive data from the NCES system as well. The system requires that both parties to the transfer be registered users of the NCES Members Site and that their Members Site privileges be set to allow use of the secure data transfer service as described above. This process will be used for all file matching procedures described below, except in instances when the vendor already has a secure data transfer system in place.

## **G.2 File Merge with ED Central Processing System (CPS)**

RTI proposes to perform file merges with the CPS data containing federal student aid application information. The merge with CPS can occur at any time for any number of cases, provided that the case has an apparently valid SSN associated with it. RTI sends a file to CPS and receives in return a large data file containing all students who applied for federal aid. We already have existing programs and procedures in place to prepare and submit files according to rigorous CPS standards. Similarly, we have already developed programs and procedures to receive and process data obtained from CPS. The CPS has recently migrated to an Internet website for file matching, and RTI has been successful sending and receiving files using the new application.

RTI will electronically upload a file on the FAFSA secure web-site for matching which contains SSN and the first 2 letters of the sample member's last name (but no other information). Access to the site for the upload is restricted to authorized users who are registered and provide identification/authentication information (SSN, DOB, and personal identification number [PIN]) to the FAFSA data site. The file is retrieved by the Central Processing System or CPS (the FAFSA contractor data system) for linkage. The linked file, containing student aid applications for matched records, is then made available to us only through a secure connection (EdConnect) which requires username and password. All CPS files will be processed, edited, and documented for inclusion on the analytic data files.

The CPS data obtained for the NPSAS:08 data collection will cover academic years beginning with 2006–2007. All CPS files will be processed, edited, and documented for inclusion in the Electronic Codebook (ECB).

## **G.3 File Merge with National Student Loan Data System Disbursement**

RTI will also conduct a file merge with the NSLDS to collect federal loan and Pell grant data for the 2006–2007 award year. The resulting file will contain cumulative amounts for each student's entire postsecondary education enrollment. NCES has set up a secure data transfer system that uses their NCES member site and Secure Sockets Layer (SSL) technology. The system requires that both parties to the transfer be registered users of the NCES Members Site and that their Members Site privileges be set to allow use of the secure data transfer service. These privileges are set up and carefully controlled by the ED's Institute of Education Sciences

(IES) NCES Chief Technology Officer (CTO). This service has been designed by ED/NCES specifically for the secure transfer of electronic files containing personally identifying information (i.e., data protected under the Privacy Act or otherwise posing risk of disclosure) and can be used for NCES-to-Contractor; Contractor-to-Subcontractor; Subcontractor-to-Contractor; and Contractor-to-Other-Agency data transfers. The party uploading the information onto the secure server at NCES is responsible for deleting the file(s) after the successful transfer has been confirmed. Data transfers using this system will include notification to the ED/IES, the NCES CTO, and the NCES Deputy Commissioner as well as the ED/NCES project officer. The notification will include the names and affiliations of the parties in the data exchange/transfer and the nature and approximate size of the data to be transferred. We have programs to create the files for the merge and also programs to read the data we receive. All matching processes are initiated by RTI staff providing a file with one record per sample member.

#### **G.4 File Merge with ACT**

To obtain valuable admissions test data, RTI will perform a file merge with American College Testing (ACT) data. RTI has the same experience file matching for ACT data as described in the sections on CPS and NSLDS file merging. We successfully used this matching process during NPSAS:96, NPSAS:2000, and NPSAS:04.

Matching of students to ACT requires Social Security number (SSN), name and date of birth to assist the data vendor in performing confirmatory data quality checks. RTI will initiate this process by providing a file with one record per SSN. We will use the same procedures described above for the NSLDS linkage: creating a password-protected, encrypted file using WinZip 10; uploading the electronic file on the NCES server for pick-up by ACT. ACT will process the data on their database and provide the matched data on the NCES server for our secure download.

#### **G.5 File Merge with the College Board**

To obtain SAT data, RTI will perform a file merge with the College Board. Matching of students to SAT data requires Social Security number (SSN), name and date of birth to assist the data vendor in performing confirmatory data quality checks. RTI will initiate this process by providing a file with one record per SSN. We will use the same procedures described above for the NSLDS linkage: creating a password-protected, encrypted file using WinZip 10; uploading the electronic file on the NCES server for pick-up by the College Board. The College Board will process the data on their database and provide the matched data on the NCES server for our secure download.

#### **G.6 File Merge with the National Student Clearinghouse**

RTI will use the National Student Clearinghouse to obtain the *Student Tracker* data on institutions attended, enrollment dates, and degree completions for the NPSAS:08 sample.

RTI will first set up an account with the Clearinghouse which will enable sending and receiving of files securely over encrypted FTPS connections. The file containing sensitive student identifiers (name, date of birth, and Social Security number) will be encrypted using WinZip 10 then submitted to the Clearinghouse using their secure FTP site. All files received by the Clearinghouse will be securely stored using FIPS 140-2 validated AES encryption, the US

federal encryption standard. Matched files, containing data on enrollment dates, institution names, and degrees completed, will be returned to RTI using the same secure FTP site.

### **G.7 Processing Extant Data**

With the exception of one initial merge with CPS data prior to computer assisted data entry (CADE), we propose to send files for matching after CADE data collection to ensure the availability of the maximum number of verified Social Security numbers and to facilitate the batch mode processing that is suitable to many of these resources. We may need to match to a source (for example, CPS or NSLDS) more than once.

The data from all of these sources, as allowed by the vendor, will be delivered in the ECB and will be useful for creating derived variables. These derived variables will also be available on the Data Analysis System (DAS). Both direct-pull and derived variables will be documented thoroughly for both the ECB and DAS.