# Automated Process for Fully-labeled Statistical Data Files

#### Vicky Dingler, RTI International

SAS, SPSS and Stata system files can be created automatically by associating a data file and metadata lookup files. Generic SAS, SPSS and Stata programs which call a .CSV data file and associated metadata text files containing variable and value labels, produce fully labeled system files. The researcher has only to modify the SAS, SPSS and Stata program files with the name of the .CSV file and the folder from which this and the text files reside, and the system files will be written to the same folder. The programs are run by including the reference to the metadata. The output is a fully-labelled system file with which the analyst can perform statistical procedures. The SAS and SPSS programs are modified to define the name of the metadata file. Whereas the Stata program is interactive with dialog boxes prompting file selection. This process was created as part of a National Center for Education Statistics (NCES) study for the NCES Longitudinal Studies Branch within the Sample Surveys Division.

#### Process Steps

#### Input Files

Metadata (Text file)

Format
Specs
(Text file)

ID|F6|12

PELLSEQNO|F2|12

PDATE1|F8|12

PLBR1|A2|12

PLBR2|A2|12

DISDT|F8|12

PLSCHL1|A6|12

PLSCHL2|A6|12

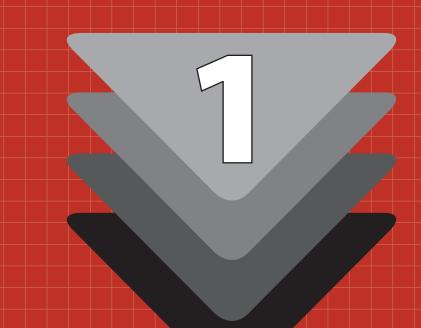
PLAMTP1|F4|12

PLYEAR|F4|12

PGM|A2|12

PLAMTSCH|F4|12

Data (CSV file)



metadata

- 0|N|id|Analysis ID|0 0|N|pellseqno|A number uniquely identifying a specific grant|0
- 0|N|pdate1|Enrollment Date|0 0|S|plbr1|Reporting Campus School Branch Code|0
- 0|S|plbr2|Attending Campus School Branch Code|0 0|S|plschl1|Reporting Campus School Code|0
- 0|S|plschl2|Attending Campus School Code|0 0|N|disdt|Date Award was Disbursed|0
- 0|N|plamtp1|Amount Paid to Date|0 0|N|plamtsch|Scheduled Federal Pell Grant Amount|0
- 0|N|plyear|School year for which a Pell Grant is to be used|0 0|S|pgm|Grant Category Type|0 2|S0|pgm|Federal Pell Grant|PE
- 2|S|pgm|Supplemental Educational Opportunity Grant|SE 2|S|pgm|State Student Incentive Grant|SS
- 2|S|pgm|Academic Competitiveness Grant|AG 2|S|pgm|National Science and Mathematics Access to Retain Talent
- Grant|SG 2|S|pgm|Teacher Education Assistance for College and Higher Education
- Grant|TG 2|S1|pgm|Iraq/Afghanistan Service Grant|IA

Users licensed to use National Center for Education Statistics (NCES) restricted use files are provided with data as commaseparated values (CSV), and a metadata file and format file in text format. Shown is an example of the metadata and format file. The user also receives generic statistical programing code (SAS, SPSS and Stata) to read the metadata and data file to produce a fully-labeled statistical data file for analysis.



### Generic Program Code SAS SPSS Stata

The statistical program (SAS code shown here) is designed to minimize the amount of editing needed to run. The user only needs to specify the name of the data and the folder location. The program parses the metadata and format file, and dynamically writes code to apply variable and value labels to the data. Shown is an excerpt of the SAS program that reads the metadata file, and the code that is generated as a result.



## Dynamically Generated Code Variable Value Labels Labels



Alphabetic List of Variables and Attributes				
#	Variable	Туре	Len	Label
8	DISDT	Num	8	Date Award was Disbursed
1	ID	Num	8	Analysis ID
3	PDATE1	Num	8	Enrollment Date
2	PELLSEQNO	Num	8	A number uniquely identifying a specific grant
12	PGM	Char	2	Grant Category Type
9	PLAMTP1	Num	8	Amount Paid to Date
10	PLAMTSCH	Num	8	Scheduled Federal Pell Grant Amount
4	PLBR1	Char	2	Reporting Campus School Branch Code
5	PLBR2	Char	2	Attending Campus School Branch Code
6	PLSCHL1	Char	6	Reporting Campus School Code
7	PLSCHL2	Char	6	Attending Campus School Code
11	PLYEAR	Num	8	School year for which a Pell Grant is to be used

The program then runs the dynamically written code to create the data file including variable and value labels.

Shown are results of a SAS Proc Contents statement at the end of the SAS program showing the variable labels.

## Labeled Analysis File SAS SPSS Stata



For Additional Information Contact Presenting author: Vicky Dingler at vdingler@rti.org

RTI International, 2150 Shattuck Ave., Suite 800, Berkeley, CA 94704 RTI International is a registered trademark and trade name of Research Triangle Institute.