

[Page One](#)
[Campus
Computing
News](#)
[Holiday Hours](#)
[Computing
Outage](#)
[Notification
Mailing List Now
Available](#)
[Computer-
Based Training
at UNT: Aargh,
I'm so confused!](#)
[Student
Computing
Services Survey
to be Launched
Online](#)
[EDUCAUSE
Reloaded](#)
[Today's Cartoon](#)
[RSS Matters](#)
[The Network
Connection](#)
[Link of the
Month](#)
[WWW@UNT.EDU](#)
[Short Courses](#)
[IRC News](#)
[Staff Activities](#)
[Subscribe to
Benchmarks
Online](#)

Research and Statistical Support University of North Texas

RSS Matters

Link to the last RSS article here: [Out With the Old, In With the New...Format!: UNT's New Faculty Evaluation Reports.](#) -- Ed.

Got EBCDIC? Take This PROC and Call Me in the Morning

By [Patrick McLeod](#), Research and Statistical Support Services Consultant

With the [decommissioning](#) of the academic mainframe, the UNT research community moved into a brave new world of computing. This transition means that certain mainframe standby formats for data are now not so easy to use any longer. As I recently found out, these older mainframe formats are also out in the wild in some of the data banks that researchers commonly use, including the [ICPSR](#) (Inter-university Consortium for Political and Social Research) at the University of Michigan. Since UNT is an institutional member of ICPSR, any faculty, staff, or student can access all of ICPSR's data holdings from any computer within the UNT subnet (any machine with a 129.120. IP address (UNT's subnet)). But what happens when you find your data and it isn't in ASCII text format or some common statistical platform format?

Are You Down With EBCDIC?

Certain institutions that contribute data to the ICPSR provide this data in formats associated with mainframe data. One of these formats is EBCDIC (Extended Binary Coding Decimal Interchange Code). If you are really interested in comparing the nitty gritty details of ASCII to EBCDIC, check out the Natural Innovations web page complete with side-by-side chart:

<http://www.natural-innovations.com/computing/asciiebdic.html>.

EBCDIC is not often found in the PC world, but one place where you will still encounter EBCDIC data is, you guessed it, the ICPSR! The International Monetary Fund's direction of trade data is available in its most current format only in EBCDIC data format. In order to use EBCDIC data for this particular data set in SPSS, S Plus, SAS, Stata, Eviews, or Lisrel, we first need to "translate" the data from EBCDIC format to ASCII format. In the case of this

particular data set, I was only able to find one data management program or routine that could accomplish this task while keeping variable names and formats intact: The PROC DATASOURCE procedure in SAS 8.2.

PROC DATASOURCE

PROC DATASOURCE uses a specific handling statement for this particular type of data (from SAS 8.2 System Help):

PROC DATASOURCE: FILETYPE=IMFDOTSP--Direction of Trade Statistics, Packed Format

The DOTS files contain time series on the distribution of exports and imports for about 160 countries and country groups by partner country and areas.

Data Files	Database is stored in a single file.	
INTERVAL=	YEAR (default), QUARTER, MONTH	
BY variables	COUNTRY	Country Code (character, three-digits)
	CSC	Control Source Code (character)
	PARTNER	Partner Country Code (character, three-digits)
	VERSION	Version Code (character)
Sorting Order	BY COUNTRY CSC PARTNER VERSION	
Series Variables	Series variable names are the same as series codes reported in <i>IMF Documentation</i> prefixed by D for data and F_D for footnote indicators.	
Default KEEP List	By default all the footnote indicators will be dropped.	

After downloading this particular EBCDIC formatted data set to my hard drive, I assembled the following SAS file using PROC DATASOURCE to open the IMF direction of trade data in SAS 8.2:

```
FILENAME TRADE 'C:\Data\directionoftrade\da7628o.ebcdic';

PROC DATASOURCE FILETYPE=IMFDOTSP

INFILE=( TRADE )

INTERVAL=YEAR

OUTSELECT=OFF

OUT=TRADEVAR

OUTBY=TRADEBY

OUTCONT=TRADECONT

OUTEVENT=TRADEEV;
```

```
KEEP _ALL_;
KEEPEVENT _ALL_;

RUN;
```

FILENAME tells SAS where to access the data file; the remainder of the SAS code is simply setting some of the options within PROC DATASOURCE for this particular file. Below is a list of all the possible options within PROC DATASOURCE (from SAS 8.2 System Help):

PROC DATASOURCE: Options

PROC DATASOURCE *options;*

The following options can be used in the PROC DATASOURCE statement:

ALIGN= <i>option</i>	ASCII	DBNAME= ' <i>database name</i> '
EBCDIC	FAMEPRINT	FILETYPE= <i>entry</i>
INDEX	INFILE= <i>fileref</i>	LRECL= <i>lrecl</i>
RECFM= <i>recfm</i>	INTERVAL= <i>interval</i>	OUT= <i>SAS-data-set</i>
OUTALL= <i>SAS-data-set</i>	OUTBY= <i>SAS-data-set</i>	OUTCONT= <i>SAS-data-set</i>
OUTEVENT= <i>SAS-data-set</i>	OUTSELECT= ON OFF	

For a complete listing of the supported data types that PROC DATASOURCE can work with, open SAS 8.2, click on the System Help drop-down menu at the top of the screen, select SAS System Help (which should be the first option on the menu), click on the Search tab, type PROC DATASOURCE into the keyword field, click on the List Topics button, select PROC DATASOURCE: Supported File Types from the Select Topic to Display menu, and click on the Display button to read through the many supported file types for this powerful SAS procedure.

Conclusions

Outside of political science, economics, and possibly finance, most researchers at UNT will not have a particular cause to access the IMF direction of trade data from the ICPSR. However, since PROC DATASOURCE offers the researcher a multitude of options for moving data into SAS (Bureau of Economic Analysis, Bureau of Labor Statistics, Center for Research in Security Prices, and the Organization for Economic Cooperation and Development to name a few), application of this procedure is not limited to these fields.

Until next time, happy computing trails and best wishes in your research endeavors!

