Chapter 4

Case Studies -- I

Chair: Leda Kydoniefs, Bureau of Labor Statistics

Glen Ferri ◆ Tom Ondra

Richard J. Bennof ◆ M. Marge Machen ◆ Ronald L. Meeks

Chapter 4

Toward A Unified System of Editing International Data

Glen Ferri and Tom Ondra, U.S. Bureau of the Census

Abstract

CONCOR is the editing component of the Integrated Microcomputer Processing System (IMPS). This module was originally a stand-alone procedural language used to identify and/or correct invalid or inconsistent information. The microcomputer provided a DOS-based platform to integrate all the major tasks of survey and census data processing which IMPS accomplished. CONCOR and CENTRY, the IMPS data entry module, have been combined to provide interactive editing. IMPS and CONCOR are being redesigned to run under Windows. CONCOR will move from a procedural language used by programmers toward an edit specifier used by subject matter specialists. What parts are to be carried over from the old version and what parts need to be re-engineered?

The survey questionnaire diskette is designed from data entry into 18 tables containing Federal obligations by intramural and extramural performers, by fields of science and engineering, and by geographic distribution. ■

Chapter 4

Data Editing Software for NSF Surveys

Richard J. Bennof, M. Marge Machen, and Ronald L. Meeks, National Science Foundation

Abstract

The National Science Foundation (NSF) uses automated data entry programs to collect research and development (R&D) data for its annual national surveys. Three of these programs are the topic of this paper:

- academic expenditures program used to collect data for the Academic Science and Engineering R&D Expenditures Survey,
- FSS program used to collect data for the Federal Support to Universities, Colleges, and Nonprofit Institutions Survey, and
- FEDFUNDS program used to collect data from the Federal Funds for R&D Survey.

This presentation will describe each of these studies and demonstrate how each data entry program is suited to its users.

Data Editing Software for NSF Surveys

Richard J. Bennof, M. Marge Machen, and Ronald L. Meeks,
National Science Foundation

The National Science Foundation's (NSF) Division of Science Resources Studies (SRS) has a mission to produce and disseminate high quality data and analyses related to science, engineering, and technology. SRS is responsible for conducting surveys on a wide variety of areas. Three of those surveys that are outlined below are managed in SRS' Research and Development Statistics Program.

NSF uses automated data entry programs to collect research and development (R&D) data for its annual national surveys. Three of these programs are the topic of this paper:

- ASQ program used to collect data for the Academic Science and Engineering R&D Expenditures Survey;
- FSS program used to collect data for the Federal Support to Universities, Colleges, and Nonprofit Institutions Survey;and
- FEDFUNDS program used to collect data from the Federal Funds for R&D Survey.

ASQ

The Academic R&D Expenditures Automatic Survey Questionnaire (ASQ) is a user-friendly, PC-based program that requires at least 384K of RAM. The ASQ program provides help at all points of data entry and allows the user to make choices by selecting menus. Automatic editing will check the ASQ for arithmetic errors or inconsistencies; and, such problems will be pointed out to the respondent. Respondents have the opportunity to manually correct them or allow the ASQ to total automatically. The program allows the user to enter data in any order and does not have to be completed at the same time. This feature allows the user to compile information at different times. When the data have been entered for any item, the user will be asked whether they wish to edit or not. When all data have been entered and edited, the user is prompted to select the trend checking option for comparing the previous year's data with current year's data to identify major increases or declines. All data are stored back on the ASQ diskette as they are entered.

The ASQ program will let the user print out a facsimile of the institution's questionnaire response, both for the current and previous years.

The most recently completed survey collected data from over 500 institutions of higher education in the United States and Outlying Areas and 18 university-affiliated Federally Funded Research and Development Centers (FFRDCs).

FSS

FSS includes a PC-based program (written in Visual Basic for DOS) used as the survey instrument by many agencies in reporting its data. Since the program requires at least 490K of conventional memory, 8M of RAM, and about 2M of hard disk space, it must be loaded on the PC's hard disk in order to work. That's because there are 6,538 specific institutions eligible to provide data for in FSS, and each institution's name, code, and geographic location are stored in FSS. The program also has the capability of adding new institutions. A "CHKMEM" software feature, which allows the program to check the hardware to see if expanded memory is running, has a "soft-boot" program which allows the participant to run the FSS program without having to reconfigure the hardware. FSS is a user-friendly, menu-driven program with extensive built-in instructions for users.

The data collected from Federal agencies includes:

- total program support of both science and engineering (S&E) and non-science and engineering (non-S&E) activities to academic institutions;
- total S&E support to FFRDCs administered by academic institutions; and
- R&D and R&D plant support to nonprofit institutions and FFRDCs administered by nonprofit institutions.

The contractor has held a series of annual hands-on respondent workshops on FSS in which participants were generally enthusiastic about working with the FSS PC survey disk. FSS contains data edit checks, a function to search for an institution by name or institutional code, a convenient lookup capability for field of science and engineering detail (respondents do not have to search through a hard copy of science and engineering taxonomy), and a function for data trend analysis. An "Import" feature of the FSS PC survey disk allows agencies with a large volume of data downloaded in an already-formatted database to import that data directly into the FSS PC survey database. The data are completely edited during the import procedure. Respondents can print out a summary report which displays the total obligations of each type of institution with field of science and engineering totals, a detailed report by individual institution, and a trend report which lists individual institutions which have a large increase/decrease in obligations between the current and prior fiscal year. The prior year totals can also be displayed in summary and detailed reports.

FEDFUNDS

FEDFUNDS is a user-friendly, menu-driven system that can be used on any IBM compatible microcomputer. The entire program is written in visual basic and uses "forms" to organize and group like data for display on the monitor. The entire program is stored on a diskette to allow survey respondents to enter and edit data directly from their microcomputers. The program contains 47 "forms" that are displayed separately (each is a separate screen). The FEDFUNDS program displays a "form" containing data items form a particular questionnaire table and allows the user to enter or modify the data on display. Combined, the "forms" contain all of the data items to be maintained by the survey.

This disk based system is also equipped with extensive built-in instructions and help facilities to aid the user in completing the approximately 2,000 data fields and narrative statements. Recent efforts

to make the system more efficient included redesigning the data entry questionnaire program to build in more internal data checks (e.g., within table and cross table checks, trend analysis function). When the program finds discrepancies during data checking, an error message is displayed and identifies the exact items in question for the prompt attention of the respondent. Further advances incorporated in the data entry program include automatic totals of fields without respondent's intervention. The program also allows the respondent to combine data from several sources into a consolidated agency report.

The survey questionnaire diskette is designed for data entry into 18 tables containing Federal obligations by intramural and extramural performers, by fields of science and engineering, and by geographic distribution. The respondents are instructed to complete the tables in order to avoid table checking errors, since, for example, the detailed data requested on one table must add to the aggregated data on another table.

The respondent also has the option to run a trend report that will produce a list of all "large" differences in data from the prior survey submission and the current survey input.

The participants at the latest NSF data entry demonstration workshop for FEDFUNDS users were enthusiastic about working with the FEDFUNDS program. Several commented on the automatic totaling of subtotals and grand totals and asked if this feature could be further enhanced by eliminating the need for the survey respondent to move through the subtotal and grand total fields (currently a user must move the cursor to a total field before the program completes the automatic total). The respondents felt that since these amounts are already computed, then they should appear as the detailed level is entered. Also, it may soon be possible to transmit the FEDFUNDS program electronically to respondents. Respondents can now electronically send completed survey results via Internet.

Additional Information

For more information about the three data entry programs (and the surveys associated with them) described above, please contact the appropriate author on (703) 306-1772 or via Internet at *mmachen@nsf.gov, rbennof@nsf.gov*, or *rmeeks@nsf.gov.* ■