AUTOMATION IN SURVEY PROCESSING

Anco Hundepool, Netherlands Central Bureau of Statistics
P.O. Box 959, 2270 AZ Voorburg, The Netherlands

1. Introduction

Survey data processing can be seen as a set of activities for manipulating both data and meta-data. The data are the answers to the questions on the questionnaire form, and in the course of the process they are transformed into useful statistics (tables, data bases). The meta-data describes and documents the data. The meta-data can take various forms in the process, depending on the system or department that is working with the data.

It all starts with the design of the survey. The questionnaire form is the first specification of the meta-data. It describes variables to be measured, the domain of valid values of these variables, and conditions under which the variables are measured. During the fieldwork data is collected that (hopefully) matches this meta-data specification. After the fieldwork the data undergoes various processes, like imputations, weighting, tabulation and analysis. It is vital that in all these subsequent processes the data and meta-data remain joined together. Whatever form the data takes, there should always be a corresponding meta-data description. Although everyone involved in survey processing realises and appreciates the importance of keeping data and meta-data together, it often does not happen in practical situations. It is not uncommon that changed and undocumented copies of data files wander around in the organisation, and errors will be made if the meta-data has to be respecified for some other system or department.

To keep the survey process consistent, there is need of a system that ensures the statistician to keep data and meta-data together. If such a system is also easy to use and efficient, it can improve the quality of the results of a survey in terms of timeliness and quality. The Netherlands Central Bureau of Statistics (CBS) has developed and is still improving an integrated system of survey processing that implements these ideas. The CBS philosophy is based on three principles: concentration, standardisation, and integration.

2. Decentralisation and standardisation

New developments in hardware and software have made it possible to look for more efficient ways of processing surveys. In the traditional environment all the automation work was performed by people qualified in automation. However the now widely available microcomputers and the advent of much more user friendly software packages have made it possible for the people in the subject matter departments to build themselves the automation systems necessary for processing the surveys. Because these subject matter specialists know best what is going on in a survey, they are better qualified to solve the problems that might arise in the data processing of survey data.

However the uncontrolled introduction of microcomputers may lead into a real chaos. If each department is free to chose the hardware and software solutions it thinks is best, the outcome will be that there is no communication whatsoever possible between the departments. Also it is impossible to guarantee a reasonable level of support for all these different hardware solutions and software packages. To avoid this chaos the CBS has chosen for a strict standardisation for both hardware and software. It has proven to be a very successful decision, which only had been made possible by the development of a very powerful set of software tools for the data processing of survey data.

3. Integration

All standard software required for data processing is now part of an integrated system. The backbone is the Blaise System developed by the CBS. The basic part of Blaise is the Blaise language. In a powerful, structured language the subject-matter specialist describes the meta-information: questions, possible answers, route instructions, and consistency checks. Using this specification, Blaise automatically generates software for data processing. In the first place, a CADI-program (Computer Assisted Data Input) can be

produced. This program provides an intelligent and interactive environment for data entry and data editing of data, collected traditionally with paper forms. This way of data collection is still used very much and successfully for the processing of establishment surveys, but new variants of data collection emerge like CASI (Computer Assisted Self Interviewing), where the respondent gets a (Blaise) instrument instead of a paper questionnaire. He then uses this Blaise instrument to enter the data directly, which is then transferred to the statistical office.

Besides these possibilities mostly used for establishment surveys two other variants are available in the Blaise system. The CATI-program (Computer Assisted Telephone Interviewing) is the Blaise solution for telephone interviewing. Blaise automatically generates the software needed for carrying out the interview, including all call management. A module which resembles the CATI-program, is the CAPI-program (Computer Assisted Personal Interviewing). This program for face-to-face interviewing with laptop computers can also be generated by Blaise.

After the data collection the Blaise system plays an important role in the further processing of the survey. As the Blaise system has the knowledge of the meta-information, it is able to manipulate this information in such a way that it can make the meta-information available in all subsequent steps. We will discuss these steps in sections 5 through 7, but we will first concentrate on some special aspects of the data collection of establishment surveys.

4. Data processing in establishments surveys.

4.1. Sampling aspects.

The collecting of data for establishment surveys implies a lot of work, both for the statistical offices and for the establishments themselves. Therefore the collecting agency should pay attention to it that they reduce as much as possible the amount of work needed by the establishments to fulfil their statistical duties. If the statistical office fails, in the eyes of the establishments, to take advantage of all the new possibilities for the efficient collection of statistical information, this will lead to a high non response rate, as the establishments will cease to spend the time of their expensive employees to fill is all the paper forms of the statistical office.

Besides from the application of efficient data collection techniques, the CBS has implemented a system of spreading the response burden more evenly over all the establishments. This system (Survey Burden System) co-ordinates all the sampling based on the business register. A burden factor is assigned to each survey. For each new survey those establishments are chosen in each stratum that have the lowest total response burden from the previous surveys. This system spreads the response burden more evenly over all the establishments without violating the rules for random sampling.

4.2. CADI

Computer Assisted Data Input is the Blaise solution for entering and editing of traditional paper forms. Although we are now at the edge of a big revolution in the data collection at establishments, the traditional paper is still used quite often. Therefore the CADI module of Blaise is still playing an important role. From the Blaise questionnaire it automatically generates a data entry and editing program, that can be used for the interactive entering and editing of the forms. Especially for the more complex applications, where the amount of work for the actual entering of the data is relatively small compared to the time spend on the editing of the data, the subject matter specialists will perform both tasks, as this has proven to be the most efficient way.

However when the amount of data is large and the editing effort is relatively small compared to the entering of the data, we see more and more a three-step procedure. In the first step the data is entered at high speed by data entry typists with a Blaise instrument, which checks only for the most obvious errors (typing mistakes). The next step is that all the data is checked by another Blaise instrument performing all the checks. In the final step the (more expensive) subject matter specialists solve the remaining errors. All instruments are generated from the same meta-data specification, so consistency is ensured.

At this moment a revolution in data processing is taking place from the traditional paper questionnaires to several kinds of electronic solutions. The establishments are using the Electronic Data Interchange (EDI) for exchanging the information between the different business partners (suppliers and clients) and are expecting the statistical offices to use these new developments as well. We cannot expect that establishments will continue to spend their precious time to fill in the paper questionnaires of the statistical

offices. If we will not implement the advantages of EDI in some way for collecting the information from the establishments, we can be sure that we will lose a considerable part of the statistical information we need to do our jobs.

Some first steps in this direction have already been taken. For short and not too complex questionnaires techniques like touch-tone and voice recognition can become very useful and efficient, but for more complex questionnaires we must look for other solutions.

4.3. CASI.

CASI (Computer Assisted Self Interviewing) has been implemented quite successfully for the collection of information at the Dutch fire brigades. As (almost) all the Dutch fire brigades have PC's at their disposal, the idea emerged that we should not send them a paper questionnaire any more but a Blaise data entry instrument. With this Blaise instrument they enter all their reports. The Blaise instrument checks all the data entered and reports on any inconsistency. At a regular basis the data is send back to the CBS (on diskette). The CBS receives clean data in a format that can directly be used for further analysis.

This way of data collection has proved to be very attractive. The CBSreceives clean data on diskettes, can reduce on the work to enter the data into the computer and can produce reliable information more recently. The advantage for the fire brigade itself is that they can use their own data, just entered and edited, for making their own summaries and tables. They can e.g. make use of the tabulation package Abacus for this purpose.

Because of the advantages for both sides, this approach has proved to be successful. When this project was initiated, the diskette was chosen for simplicity reasons to exchange information between the CBS and the fire brigades, but we will change this to modem and telephone lines.

4.4. IRIS.

1992 has been a remarkable year in Europe. As the community has moved into the direction of one market, the customs control has been suspended at the borders between the member countries by the end of 1992. Of course this is an improvement for those involved in the transportation of goods between the member countries. However for the statistical offices there is the loss of the information which originated from copies of the customs declaration forms. Therefore a new system of collecting the information for the foreign trade statistics had to be developed.

From 1993 the export firms have the obligation to report their exports directly to the statistical office. With the experience of the fire brigades it was decided to try to use similar techniques to facilitate the reporting by the export firms. This has led to the IRIS (Interactive Registration of International trade Statistics) package.

The basis of the package is formed by several Blaise instruments with a powerful shell around it. It contains modules to convert data files generated by the administration systems of the export firms directly into the format required by IRIS. The converted data is then loaded into the Blaise instrument and checked for possible inconsistencies. The firms can also use the Blaise instrument to enter directly the data needed by the statistical office.

The reaction from the firms was much better than expected and within a few months up to 16.000 copies of the package have been delivered to the export firms, a number exceeding by far even our most optimistic expectations. It has proved again that these kind of data collection techniques can be applied very well at this moment.

4.5. Blaise-EDI.

Further investigations are being made into the use of EDI-techniques for the collection of information at establishments. Most establishments are quite willing to supply the necessary information to the statistical offices, but they are reluctant to spend much time on it, as time is money. It is well known that the filling in of endless forms for statistical offices is not a top priority job for establishments. The development of the Response Burden system was one approach to moderate this burden. But we should continue to investigate new ways to reduce the response burden.

Co-ordination of different statistics is one point. Nothing is more demotivating than having to give the same information on several occasions on different forms to the same statistical office. The CBS concentrates on techniques to make the reporting by the establishments more efficient. Electronic Data Interchange (EDI) will be used in the near future to collect the information.

The idea has been born that we should try to build (Blaise) modules into the administration systems of establishments. This module will be able to access the data in the administration and will convert these data into the statistical information in a format suitable for the Blaise module. This module will then check the data and transfer the information with EDI to the statistical office. Of course we should leave the control over this module in the hands of the establishments. Only if they decide to do so, the data will be converted and transferred to the statistical office.

As in the Netherlands five major software firms are responsible for the administration software of almost all establishments, the development of this Blaise module and the integration into the administration software does not seem to be an impossible problem. The first reactions from the establishments and the administration firms are encouraging. Therefore we will continue to search into this direction.

5. Weighting

The clean file with sample survey data produced by the Blaise System is usually not ready yet for making tables and other analysis about the establishments from which the sample has been drawn, because the data do not constitute a representative sample. So some adjustment procedure has to be carried out. In order to account for unequal selection probabilities and non-response, one has to compute adjustment weights.

Bascula is a general weighting program, running on microcomputers under MS-DOS. It combines several weighting techniques. In the first place, traditional post-stratification can be carried out. And if the number of empty strata is small, one can instruct the program to collapse (i.e. combine) these strata with neighbouring strata. In the case of many empty strata, or lack of sufficient population information, Bascula can carry out linear weighting (see Bethlehem and Keller, 1987), or multiplicative weighting (also called raking ratio estimation). The resulting weights can either be added to the data file, or be stored in a separate file. The setup file with meta-data for Bascula is created by Blaise. This setup also contains instructions to add variables to the data file. One of these variables must contain the input weight. The input weights are inverse proportional to the sampling probabilities. Another variable will contain the final adjustment weights computed by Bascula.

6. Tabulation and Analysis

Tabulation is one of the basic activities in the statistical production process, and it was one of the first to be automated. Many tabulation packages have already been developed, but most of them are not very user-friendly. This is partly caused by the fact that the proper generation of a complex table needs the specification of a lot of parameters controlling composition and layout. To overcome these drawbacks the CBS has developed the tabulation package Abacus. While Abacus may be seen as yet another tabulation package, it was developed with very specific design goals. In the first place, no control language is used to specify a table. The program is menu-driven instead. The user designs his table in an interactive, simple, and intuitive way, without having to know about any control language. In the second place, Abacus works in full co-operation with Blaise. Therefore can use fully the meta-data already specified in Blaise and can read all the data files produced by Blaise. However Abacus can also be used separately for any ASCII file. Thirdly, the program can produce camera-ready tables. A striking property of Abacus is its speed. On a 486-66MHz microcomputer we have reached a speed over 700.000 records per minute. The reason for this is that the Abacus program is rather small, so it can use a large part of the base memory as a working area which allows for a table of up to 100 000 potential cells.

Tables produced by Abacus can have up to three dimensions (layers, rows and columns). Every dimension can hold up to 10 variables, which may be nested or concatenated. This table can contain simple counts, but Abacus can also calculate totals, percentages, and averages. Abacus can accommodate weighted data, using the weights that are, for example, computed by Bascula. Currently we are developing a new table manipulation program. This program will make it possible to influence even further the layout of the tables produced. Also several post-calculations can be performed and the paging can be controlled, before the table is printed. An advantage is that it is no longer necessary to recalculate the whole table if one wants to influence the layout, etc. The aim of Abacus is to produce tables which are really camera-ready. If the user is still not satisfied with the resulting table and wants to do more complex calculations, he can still export the output of Abacus to a spreadsheet program like Lotus 123, and carry out further processing there. More details about Abacus can be found in Bethlehem et al. (1989a).

Tabulation may be followed by a more extensive analysis of the data. The CBS has not developed any standard software for statistical data analysis. To facilitate the use of other packages, Blaise translates the meta-information into the language of the package used. There is a growing list of packages supported by Blaise, including SPSS, SAS, Stata, Paradox, and Oracle. Translation for other packages can easily be defined using the Blaise Setup Generator. The data themselves are transferred to the package through an ASCII file.

7. Publication

The whole statistical production process is aimed at the publication of statistical information. Traditionally the statistical agencies produce large volumes of paper publications. The now widespread availability of (micro)-computers has led to the growing demand for electronic publications. People want to use the statistical information on their own computers without the necessity to key in the information from the books. Besides the production of all those books is a waste of valuable material (trees etc.).

The most easy way would have been to copy some data files on diskette or any other medium and make them available. It is however the obligation of the statistical offices to see to it that an optimal use is made of the information that has been collected with so much effort and at high costs. As data without the necessary meta-data is valueless, we should see to it that both data and meta-data is published together in a comprehensive way.

To achieve this the STATview package has been developed. The basic idea behind STATview is that it should serve as a general shell around both the data and the meta-data and should serve as an interface to the users of the information. With STATview the user can make selections from the STATview databases. He can use either the hierarchic tree structures in which all the information items are stored, or use a powerful thesaurus to find the information he is interested in. This thesaurus is based on the trigram technique. This technique has the advantage that it will find the keywords even after some mis-spellings.

When the user has made his selection the information is retrieved from the database and presented on the screen as a table using the same table manipulator as in Abacus. The user can then do some basic manipulations with the information, influence the layout etc. Besides this the information can be exported to all kind of formats, ASCII, Lotus, dBase; for the ASCII file a meta-data description can be made in a format suitable for most of the analysis packages like SPSS, SAS, Stata, etc.

With STATview the information can be published is a very user friendly way, giving the user full control over the data. The reactions of the public are very encouraging and at this moment one third of the revenues of publications comes from the STATview publications.

At this moment the information is published on diskette, but we are converting STATview into an on-line information system. This on-line system has the same capacities as the STATview system, but now the users can access the STATview databases from his own PC via a modem and the telephone line. This opens the way to access the most recent information directly, which is of course a well-known wish of many users of (economical) statistical information.

8. Conclusion

Integrated survey processing means that all the processing of a survey from the early beginning (design of the survey) until the final publication of the results are joined together in a set of homogeneous co-operating software tools. A predominant role is all these steps plays the meta-information. The basic idea is that information once given to the computer should never be entered in whatever form again into the computer; an obvious source of many errors. The backbone of the system is the Blaise language. This language serves as a knowledge base for the meta-information and from that base the software tools for the various kinds of data collection are generated and also the meta-information needed in all the other steps in the production process.

The use of Blaise for establishment surveys is not very different from the use in social surveys. It just means that different powerful possibilities of Blaise are used more during the data collection like the access of external information (e.g. from the previous years) for the data editing. The high speed tabulation package Abacus and the recent publication package STATview are very well suited for the use on establishment surveys.

References

J.G. Bethlehem, A.A.A. van Buitenen, A.J. Hundepool, M.J. Roessingh, and A.van de Wetering (1989a): Abacus 1.0, A Tabulation Package, Compact Guide. CBS report, Voorburg: Netherlands Central Bureau of Statistics.

J.G. Bethlehem, A. J. Hundepool, M.H. Schuerhoff and L.F.M. Vermeulen (1989b): Blaise 2.0/ Language Reference Manual. CBS report, Netherlands Central Bureau of Statistics, Voorburg.

J.G. Bethlehem and W.J. Keller (1987): Linear Weighting of Sample Survey Data. Journal of Offical Statistics 3, pp. 141-154.

J.G. Bethlehem, F.M. Kellenbach and W.J. Keller (1991): Computer Assisted Statistical Information Processing at the Netherlands Central Bureau of Statistics. CBS report, Netherlands Central Bureau of Statistics, Voorburg.

L.T. van Huis, C.A.J. Koeijers and S.J.M. de Ree (1993): The Response Burden System; User Manual, part A (in dutch). CBS report, Netherlands Central Bureau of Statistics, Voorburg.

PERSONAL COMPUTER ELECTRONIC DATA REPORTING OPTION (PEDRO)

Charles C. Heath, Department of Energy
Energy Information Administration (PSD, EI-42)
1000 Independence Ave., S.W., Washington, DC 20585

PEDRO'S HISTORY

Today, the Energy Information Administration (EIA) is acknowledged to be the major collector of reliable and comprehensive data on energy supply and demand in this country. During its 14-year history, EIA has tried steadily to make the data-collection process less burdensome for business, industry, and the general public. By 1989, the reporting burden on respondents to its various energy information surveys was only 18 percent of what it had been 10 years before. While minimizing respondent burden continues to be one of EIA's foremost goals, future reductions will generally come through the development of new and innovative tools and procedures rather than the elimination of reported data.

Four years ago, EIA embarked on the creation of a flexible, user friendly, portable, state-of-the-art tool that would allow those petroleum companies that submit data to EIA to save time and money in the preparation and submission of their forms. This effort was undertaken in cooperation with the petroleum industry and resulted in the development of the Petroleum Electronic Data Reporting Option (PEDRO). The initial result of these efforts was an advanced electronic data communications product available at no cost, to be used for filing the Form EIA-810, Monthly Refinery Operations Report. The new system proved to be more efficient, timely, and accurate.

PEDRO began with more questions than answers. What hardware was most readily available? What software package would most efficiently offer data entry and communications capabilities? How would survey respondents certify the accuracy of their data? How could data be protected to ensure confidentiality?

These questions and many more were answered through a systematic review process that included a survey of industry capabilities to determine what hardware/software was most prevalent, meetings with the Internal Revenue Service to hear about their electronic signature efforts to authenticate data, and the use of encryption methodologies similar to some used by the National Security Agency.

As the development of PEDRO began, specific consideration was given to the target audience to be reached. Like many survey systems elsewhere, the EIA surveys chosen for initial PEDRO development were characterized by a few respondents providing a large percentage of the volume of reported data. In other words, out of approximately 3,000 respondents reporting in the PEDRO target surveys, fewer than 600 provided 70 percent of the reported data. These companies were targeted as the respondents that would provide the best test of PEDRO's time-, energy-, and money-saving potential. To further aid PEDRO development, direct participation by industry was actively sought. Sun Oil Co., in particular, played a prominent role in the successful development of PEDRO. This company not only served as a test site for each survey but also offered many constructive suggestions that were incorporated into the final production versions.

Since PEDRO's inception in 1989, its use has expanded from a single monthly survey to 18 weekly, semimonthly, and monthly surveys in the petroleum supply and marketing areas in 1991. More than 500 respondents in the petroleum industry are currently using or have requested to use PEDRO. Out of 25 State energy offices, 23 use PEDRO to file winter fuels data submissions to EIA during the winter heating season (October 1 through March 31).

The results have been outstanding. EIA is receiving the data much faster. The time needed by respondents to complete and transmit the forms has been reduced. The placement of the data entry/editing process in the respondent's office has virtually eliminated the time-consuming and burdensome follow-up for both the respondent and EIA. For example, according to performance statistics, the error rates being experienced by EIA on the Monthly Refinery Report and the Monthly Natural Gas Liquids Report dropped by 60 percent

and 70 percent, respectively. Moreover, for all petroleum supply forms, EIA telephone calls to respondents for verification of data have fallen by more than 80 percent.

WHAT IS PEDRO?

PEDRO is an advanced electronic data communications product available to anyone upon request, at no charge to the requesting government or industry organization, and is used for filing weekly, semimonthly, and monthly petroleum supply and marketing survey data to EIA more efficiently, more rapidly, and more accurately.

PEDRO was initially developed for those companies that filed the following monthly petroleum supply forms:

EIA-810, Monthly Refinery Report
EIA-811, Monthly Bulk Terminal Report
EIA-812, Monthly Product Pipeline Report
EIA-813, Monthly Crude Oil Report
EIA-814, Monthly Imports Report
EIA-816, Monthly Natural Gas Liquids Report
EIA-817, Monthly Tanker and Barge Movement Report.

As experience was gained with the system, EIA continued to work closely with the petroleum industry to expand the number of forms to be supported by PEDRO. Today the list of forms that are filed electronically has been expanded to include:

EIA-14, Refiners' Monthly Cost Report
EIA-182, Domestic Crude Oil First Purchase Report
EIA-782A, Refiners'/Gas Plant Operators' Monthly Petroleum Product Sales Report
EIA-782B, Resellers'/Retailers' Monthly Petroleum Product Sales Report
EIA-782C, Monthly Report of Petroleum Products Sold Into States for Consumption
EIA-800, Weekly Refinery Report
EIA-801, Weekly Bulk Terminal Report
EIA-802, Weekly Product Pipeline Report
EIA-803, Weekly Crude Oil Stocks Report
EIA-804, Weekly Imports Report
EIA-818, Monthly IEA Imports/Stocks-at-Sea Report
EIA-826, Monthly Electric Utilities Sales and Revenue Report with State Distribution
EIA-877, State Energy Office Collection of Propane (Winter Fuels Report).

HOW DOES PEDRO WORK?

PEDRO operates on an IBM compatible PC with a minimum of 640K memory, 2.5 megabytes of hard drive storage, one external disk drive, and PC or MS DOS version 3.0 or higher. PEDRO is written in C and CLIPPER and uses DBASE data structures. It eliminates the need for a user to purchase any special software. PEDRO is completely self-contained and easily installed.

PEDRO allows EIA respondents to make an electronic entry on their own PC's. The data, once entered, are compared with historical data to highlight discrepancies (data that are significantly different from previous entries). Respondents are prompted to annotate or correct the discrepancies. Once the data have been completely entered and checked, they are transmitted to EIA. When data have been successfully transmitted to the EIA mainframe computer, the data from different respondents are combined and reformatted by programs for use in the presentation of analytical, statistical, and electronic reports.

The PEDRO system entails three functions: (1) data entry and editing; (2) transmitting data to EIA; and (3) EIA processing. Data can be entered for the current reporting period or previous reporting periods (resubmission) by using one of two methods via the PEDRO system. The first method is to enter data by moving the cursor to the appropriate field on the electronic image of the form and typing the entry. The data are subjected to simple error checks, and when data fail error checks, the user is prompted for an explanation.

The second method of entering data is to import an ASCII file. Data that have been entered via other software programs, such as Lotus 1-2-3 and DBASE III, can be brought into PEDRO without re-keying. Data can be imported for weekly, semimonthly, and monthly submissions and resubmissions.

Two types of data editing are performed by the PEDRO system. First, during data entry PEDRO performs simple error checks on the data, such as sums of components. Second, PEDRO checks the data against data from past periods when users invoke the range check option from the main PEDRO Menu. Possible errors from these checks are highlighted and the error check is displayed for review. Users are asked to correct these entries and

confirm them with an explanation in the comment section of the form.

Transmitting Data to EIA - Once data have been entered and edited, they can be transmitted electronically to the EIA mainframe. This electronic transfer is facilitated by a communications package provided to the user that enables the PC user to access disk space on the EIA mainframe, where data reside in a respondent's specific space called a Remote Disk Environment (RDE).

The transmission process has error-free protocol or built-in error checking to ensure that the data have been correctly and completely transmitted from the PC to the EIA mainframe.

EIA Processing - The PEDRO software organizes the electronically collected data in a format used by EIA. The essential element of this component of the PEDRO system is the consolidation of the data from the numerous respondent specific RDE's. This is done by a program residing on the EIA mainframe which combines the transmitted data. Once the available data from the respondents have been consolidated, a program is used to reformat the data into a structure compatible with EIA report generation programs.

When a respondent transmits resubmissions, EIA will transfer files for comparison purposes from the EIA mainframe to the respondent's PC. The encryption program will translate these files into clear text, make the data base comparison, and note discrepancies on the respondent's PC.

Security and Privacy - Security has been a primary concern in both the development and operation of PEDRO. Access into PEDRO requires a password and a second password is required to transmit the data. This allows for a separation of authority to enter and to transmit data. Additionally, all PEDRO data are encrypted before and during transmission and decrypted after they are received in order to prevent any unauthorized interception of readable data. PEDRO uses several levels of error checking to ensure that all data are transmitted accurately and that unauthorized access is prohibited.

RECENT MAJOR ACHIEVEMENTS

Weekly Petroleum Supply Surveys Added - In 1991, the EIA completed a year-long project with the American Petroleum Institute (API). Beginning in 1990, the EIA took the lead to adapt PEDRO so that both EIA's and API's reporting requirements could be fulfilled with a single set of forms and instructions, a single system, and a single entry of data to be reported. Two of the 5 weekly survey forms were implemented in January 1991, and the remaining 3 were implemented in the following May. This effectively cut the combined reporting burden by 50 percent for the participating companies. More than 60 companies are now participating in the weekly program.

Winter Fuels Survey Added - The EIA-877, "Winter Fuels Report," is a unique application of PEDRO. This survey was developed for the State Heating Oil and Propane Program. The survey is used to collect fuel oil and propane prices on a semimonthly basis from 25 States located in the Northeast, Mid-Atlantic, and Midwest. Each of the 25 State energy offices is provided a sample of approximately 50 fuel oil and propane dealers. The samples are loaded onto the PEDRO diskettes before they are mailed to the States. The State energy offices call each company in their respective samples to obtain retail prices of fuel oil and propane on the first and third Mondays from October through March of each year. During the 1990-1991 winter heating season, 23 out of 25 State energy offices used PEDRO to transmit their data to EIA.

The EIA-877 PEDRO survey contains many features that make it unique compared to other electronic data collection applications. First, this PEDRO survey serves as a type of Computer-Assisted Telephone Interviewing (CATI) system with specialized screens to facilitate calls to companies surveyed. Second, respondents have the capability, if directed by EIA, to change the reporting frequency from semimonthly to weekly reports to accommodate fuel emergencies. Third, the system is an integration of State and Federal energy agencies that share password-protected software and electronic transmission. In addition, respondents' PC's contain flexible data bases of heating oil and propane dealers. Lastly, this EDI application is a PC-to-PC processing system that transmits data without keying errors.

In March of 1992, the National Science Foundation (NSF) requested copies of all programs and documentation in order to adapt PEDRO for use in NSF data collection activities. It is anticipated that few modifications will be necessary for NSF to adapt PEDRO to their environment. The result should be

a significant savings to both the NSF and those that respond to NSF surveys.

WHAT IS PEDRO'S FUTURE?

As the industry response to PEDRO continues to grow, EIA will continue to enhance PEDRO's capabilities with the use of expert systems and other state-of-the-art approaches to make PEDRO more comprehensive and easier to use. It is extremely likely that PEDRO or similar electronic data transmission capabilities will be developed for other EIA surveys, such as the Form EIA-826, "Monthly Electric Utility Sales and Revenue Report With State Distributions." The Department of Interior, Nuclear Regulatory Commission, Department of Commerce, and others are also considering PEDRO or like technology for their data collection and are closely monitoring the success of the PEDRO system.[1]

1. Today, PEDRO documentation and software are available, upon request, without charge to the requesting government or industry organization. Additional Information on PEDRO can be obtained by writing to the Energy Information Administration, U.S. Department of Energy, 1000 Independence Ave., S.W., Washington, D.C. 20585, U.S.A., Attention: PEDRO Project Manager, EI-42.