

PN-ACE-167

SECRETARIA DE AGRICULTURA Y GANADERIA¹⁰⁰¹⁰¹
UNIDAD DE PLANEAMIENTO Y EVALUACION DE GESTION (UPEG)

***FINAL REPORT: DATA BASE
MANAGEMENT PROGRAMMING
SUPPORT FOR SIMPAH***

Prepared by Bruce L. Brower Ph D

*PRODEPAH Economic Policy Analysis and Implementation Project
USAID Contract Number 522-0325-C-00-3298
February 1997*

PRODEPAH is a project of the United States Agency for International Development and the Government of Honduras, carried out by Chemonics International Incorporated and Sigma One Corporation

TABLE OF CONTENTS

SECTION I	INTRODUCTION	1
	A Purpose of the Report	1
	B Products Delivered	1
	C State of Operation of SIMPAH	2
	D Brief Overview of the CPD	2
SECTION II	WORK PERFORMED	4
	A Introduction	4
	B Work Performed	4
	B1 Data Management Systems	4
	B1a Interface With Data Input Module	4
	B1b On Demand Codebook Generation	4
	B1c Specialized Reporting Options	5
	B1d System Testing and Modification	5
	B1e System Documentation	5
	B1f Non-simpah Data Incorporation, or Data Interchange Module	6
	B1g Integration of FHIA	6
	B1h Purging and Archiving System	6
	B1i Exchange Rates	6
	B1j Report Formats	7
	B1k Program Interface	7
	B1l Menu Controlled Help System	7
	B1m Quality Control	7
	B1n Printer Control Codes	7
	B2 Operations Systems	7
	B3 Staff Training	8
	B4 System for Data Integrity Control	8
	C Problems Encountered	8
	C1 Windows 95	9
	C2 Network	9
SECTION III	SUGGESTIONS FOR THE FUTURE	10
	A Weights and Measures and Adjusted Sale Units	10
	B Training and Programming Follow-up	10
	C Marketing Assistance - Training	11

SECTION I INTRODUCTION

A. Purpose of the Report

Honduras adopted the "Ley de Modernización Agrícola" in April 1992. The PRODEPAH project has been in active support of the development of free markets in the agricultural sector in Honduras since that time. This support has included assistance in establishing SIMPAH (Sistema de Información de Mercados de Productos Agrícolas de Honduras). SIMPAH now provides public reports of the prices of products in the major markets of Honduras on a daily basis, for the markets of Tegucigalpa and San Pedro Sula, and on a weekly basis for regional markets.

This assignment was to adapt the Commodity Price Database (CPD) so as to provide SIMPAH with a computerized data management system which allows SIMPAH to maintain historical price data and generate reports from those data. It has also involved assisting SIMPAH with all aspects of making the market reporting system and office operations functional.

This report reviews the elements of the assignment and describes their completion. It describes problems that were addressed, that had not been anticipated in the assignment. Finally, it discusses issues for future consideration and management. This report does not attempt to provide a detailed explanation of the technical work accomplished. The CPD manual and codebook, also submitted with the completion of this assignment, provide those details.

B. Products Delivered

The completion of this assignment includes the delivery of a number of products, which are listed below:

- 1 Working copy of the Commodity Price Database, installed at SIMPAH offices in Tegucigalpa
- 2 Working copy of the Commodity Price Database, installed at SIMPAH offices in San Pedro Sula
- 3 Working copy of the Commodity Price Database, installed at FHIA
- 4 Documentation of the program, in Spanish
- 5 Copy of the source code, comprising over 40,000 lines of programming code. (At the rate of finished programming calculated by commercial software houses, this represents almost 10 man years of programming effort. About 60 per cent of this programming existed in the previous version of the CPD, pre-dating this assignment.)
- 6 Codebook. This book, comprising several hundred pages, gives a detailed listing of the meaning of the codes in all the fields of all the data files.
- 7 Various support documents, such as abbreviations used by the USDA in market reports.

- 8 The combination of SIMPAH generated and donated data now comprise a data set with more than 250,000 records (At the normal data entry rate, this represents more than two and one third man years of data entry)
- 9 Final report

C State of Operation of SIMPAH

SIMPAH is currently gathering and publishing Honduran price information on a daily basis. There is anecdotal evidence in that the information is having an impact. The number of users requesting daily faxes of the information has climbed. Growers are showing up in the market with the latest reports in hand to assist them in negotiating prices. The establishment and effective functioning of SIMPAH has evolved in a remarkably smooth fashion. The close attention of the PRODEPAH staff, the generally excellent team of employees at SIMPAH, the support of the Government of Honduras, the goodwill of business in the markets, and the excellent assistance of USAID have all combined to promote an environment that has allowed the rapid implementation of the project. Though there have been, and still are, major problems to resolve, none have compromised the technical functions of the office. The quality of the data is high, and has been from the start. The consistency of the service has been perfect. This is both gratifying and remarkable given the complexity of the process of establishing a new office, defining and refining the processes of gathering and reporting market information, as well as training staff. -All of the participants in this process can justifiably be proud of the results to date.

The staff of SIMPAH has been, by and large, well focused, well directed, efficient, committed, and careful. The mission statement is targeted and has been followed. The commitment to maintaining high quality control over the data management aspects of the work is particularly commendable.

The data are being input, on a daily basis into the Commodity Price Database. Staff have been trained in the process of inputting data, quality control, and report generation. However, as will be discussed below, there continues to be a significant need for follow-up training.

The products in the data base include fruits, vegetables, grains, and inputs. The CPD has over 2,200 products classified, using as its base the classification system employed by USDA. Of these, SIMPAH is inputting data for several hundred products.

D Brief Overview of the CPD

The CPD is a relational database of market information. It provides a storage system and information management structure for the system. A full explanation of the system is provided in the system documentation. By way of summary, the system includes the following characteristics:

- The USDA classification of products is used for the CPD. In addition, the market descriptors of USDA are used as the basis of the CPD structure because the USDA system is the world's most complex and complete market information reporting system for perishable agricultural products.
- The system allows for users to input data directly, according to their own need, or to import data prepared at other locations.
- Two types of reports are produced, price histories and market reports. Price histories show historical price patterns. Market reports show market activity for one point in time. In their simplest forms, price histories are reports for one product, in one market, over a range of dates. In their simplest forms, market reports are the prices for many products in one market, for one day. Both types of reports have a large number of options for customized reporting.
- The reports can be generated according to the user needs, controlling on any of the variables of the system. In this fashion, it is possible for the user to look at products in specific markets, under any combination of circumstances, such as the origin of the product and the quality.
- No matter how the product was sold in the market, the user can control the output to the sale unit of choice, or to price per pound, per kilogram, or others.
- The reports may be generated in any currency of the world.
- The system functions in English and Spanish.
- The system is menu driven and highly interactive, providing the user with continual assistance at identifying appropriate codes or querying the meaning of codes, structuring reports, and covering other needs.
- The system will generate reports to be viewed on the screen, printed, output to spreadsheets, or to databases.

The strength of the CPD is in data handling. It assures that data are adjusted to the user's reporting needs. It is not a sophisticated report formatting program. Rather, through the generation of reports as spreadsheets, it provides the user with the flexibility of arranging output as needed, such as graphing and combining reports.

SECTION II WORK PERFORMED

A Introduction

This assignment involved a number of activities, which are reviewed below. The primary focus was to adapt the existing data management program, known as the Commodity Price Database to the needs of the SIMPAH operation. During that process, a number of problems were encountered and addressed in addition to the work that was initially envisioned and assigned. The more serious of these unanticipated issues are described later in this report.

B Work Performed

B1 Data Management Systems

B1a Interface With Data Input Module

This assignment occurred in parallel to the establishment of the SIMPAH office and operations. As a result, the need to begin to input and store data preceded the completion of the work on the CPD. It was also the case that the CPD was originally designed as a single use system, yet the application with SIMPAH is a network installation. As a result, it was deemed useful to have a separate data input program that would interface with the CPD. That data input system was programed by SIMPAH staff in FoxPro.

This assignment has included close coordination to insure that the SIMPAH data input system and the CPD match data structures and coding schemes. They interchange data under the control of program modules. They have congruent structures and have complementary data integrity checks.

B1b. On Demand Codebook Generation

Modifications to the CPD for SIMPAH have included the option of adding to the code structure of any of the variables in the data base. Because of that capability, the potential exists for the codebook to become outdated and inconsistent with the actual codes used in the system.

A module was added to the CPD to allow the user to generate, interactively, any or all of the codebook. In this fashion, the user can generate replacement pages for any of the variables that are altered through coding changes.

B1c Specialized reporting options

The computerized reporting system currently allows for daily market reports which summarize the prices for all products in the selected markets for a given day. It also allows for the generation of price histories for a given commodity for a specified date range. In addition to these capabilities, SIMPAH can now generate reports with the following characteristics:

- Ability to force date inclusion in spreadsheet generated reports for days when no prices were reported so that multiple reports can be combined into one spreadsheet and have corresponding dates fall on the same lines
- Ability to select date ranges for multiple market reports, rather than one date. This will allow the user to report daily, weekly, monthly reports, or reports with any other periodicity desired. In addition, if the user enters exactly two dates, the option is given to consider those dates either as a range or as individual dates
- Ability to combine markets for reporting purposes
- Ability to generate market reports for specific products or for product groups, allowing the CPD to search the databases for those products reported for the selected market(s) for the selected day(s)
- Ability to save market report definitions so that recurring, complex reports can be called at will and adjusted for only the dates desired

B1d System testing and modification

The CPD has been tested as the programming has progressed. Through the patience of SIMPAH staff, a large number of normal programming errors have been identified and resolved. Over the last few months, the numbers of errors has declined dramatically, as the system refinement has progressed. Most importantly, no errors have been encountered, since mid-1996, which compromised data. That is, the errors that have resulted have been procedural, but have not been either destructive to data, nor compromised the integrity of the reports.

B1e System documentation

A manual for the CPD has been developed which covers both the functioning of the system according to the structure of the menus, as well as documenting the concepts and variables used in the CPD. It can be used as both a training or refresher for users as well as a reference when questions arise.

B1f Non-SIMPAH data incorporation, or data interchange module

The CPD has been programmed to accept data from five sources. It will accept direct input, spreadsheets, data files in the format of db, data files in the format of dBase (dbf), and data files from other installations of the CPD

The program provides for easy exporting of data, in several formats, as well. This makes it possible for the San Pedro Sula and Tegucigalpa offices to interchange data with no intermediate data inputting steps. It also makes it possible for SIMPAH to exchange data with FHIA, which inputs international price information

B1g Integration of FHIA

FHIA has been using an earlier copy of the CPD for several years. Toward the close of this assignment, a trip was made to San Pedro Sula. Several things were accomplished during that trip

- 1) The SIMPAH modified CPD was installed at FHIA
- 2) The FHIA data structures were modified so as to be compatible with SIMPAH
- 3) Training was provided to FHIA staff as to how to import data from SIMPAH as well as export data to SIMPAH
- 4) Training was provided regarding the modified reporting system

B1h. Purging and archiving system

The CPD was altered so that the SIMPAH staff can control the volume of data that is actively available, and thereby control the system responsiveness during report generation. The system allows for archiving any or all of the data files according to a user specified date. In other words, the user may pick a date, such as December 31, 1995, and archive all data for a given product, a range of products, or all products, prior to that date. At any time a historical report for the archived data is desired, the user may request the retrieval of the archived data for any product, range of products, or all products

Additionally, the normal backup process of the CPD has been altered so that after the addition of several thousand records, the CPD will ask the user if he wishes to archive those duplicate data, or purge them. In this way, the user can control the growth of the system size by the way in which data are archived

B1I Exchange rates.

The CPD was altered so as to allow for data output in any currency of the world. Furthermore, it was designed so that multiple currencies can be controlled for any one country, such as the currency at the official exchange rate or at the black market exchange rate

B1j Report formats.

The report formats were added to and modified This included increasing the number of output formats as well as programming a module to allow for report review, on screen, prior to printing

B1k Program interface.

The program interface was customized to the SIMPAH office

B1L. Menu controlled help system

An extensive system was added to allow the user to request assistance through the program operation for selecting codes or asking the meaning of existing codes This obviates the need for frequent inquiries of the codebook and speeds the processes of data management

B1m. Quality control

The CPD was altered so as to provide the ability to control for data reporter, inputer and inputting organization These reports can be output to spreadsheets from whence analysis of variance among reporters, data inputers or inputting sites can be performed, for the purpose of identifying error patterns

B1n Printer control codes

Printer control codes were added to allow for printing from the CPD on the SIMPAH printers These are selected by menu While this work was accomplished as envisioned, there remained some problems, discussed below, caused by the intermediate control of the network over printers

B2 Operations Systems

The operation of the SIMPAH involves the close integration of all of its staff Unlike most organizations, where job responsibilities are relatively compartmentalized, the SIMPAH requires a system in which all staff are capable of filling in for key missing staff Most of the positions of the SIMPAH operate at some point in the chain of data gathering, inputting, report generation, or dissemination All staff have been required to be sufficiently proficient in all of these positions so that the temporary absence of any staff member can be covered by coworkers Training has been given to all staff in the reporting and basic data inputting systems A system has been devised for actualizing this training by having staff accompany reporters on a recurring basis so as to understand the data gathering and dissemination processes

This assignment has included assistance to SIMPAH to develop a carefully designed and documented system of responsibilities and authorities. It details the working and reporting responsibilities of each position. It also establishes a system of hierarchies of replacement should a staff member be absent.

B3. Staff Training

Staff training for SIMPAH has included a number of aspects

- a Training in the basics of data gathering and the concepts of market information
- b Training in reporting
- c Training in data entry
- d Assistance with data dissemination, particularly with regard to formats and reporting combinations of variables
- e Training in the use of the CPD

The documentation was designed with the thought in mind that it could also be used to backstop SIMPAH staff training when new staff are hired.

B4 System for Data Integrity Control

A number of program modules were developed to test processes and insure data integrity. These included the following:

- a Extensive use of menus and pop-up selection screens to control the integrity of the codes entered
- b Controlled access to certain data input functions so that existing data could not be altered if those alterations would have unforeseeable consequences for existing data. This was particularly true for the management of sale units, where a coding change would invalidate the meaning of data previously input using the prior code with its meaning.
- c Data integrity checking of input data against existing code structures. This also involved a curilinear system for scanning newly input prices to see if they varied more than should be expected from existing data.

C Problems Encountered

It is normal for unforeseen problems to arise during an assignment such as this. However, several of these problems were particularly difficult and bear mentioning.

C1 Windows 95

SIMPAH uses Windows 95 as the operating system. Early in the development of the CPD, operating system crashes were perplexing and of frequent occurrence. The "work around" solutions to these crashes were not elegant and caused frustration among SIMPAH staff and compromised the utility of the CPD. On the surface of the problem, it appeared that the CPD caused the crashes intermittently when it made normal calls to DOS commands that underlie Windows 95.

After months of investigation and unhelpful calls to Microsoft and Borland, it was discovered, through interaction with other Paradox programmers, that the problem was common and affected a large number of programs. The difficulty had been traced to an interaction between Symantec software (including Norton Antivirus in use at SIMPAH) and Windows 95. Microsoft had violated its own published rules on the use of memory addresses. Symantec used memory addresses in accordance with Microsoft published instructions, but when third programs made certain requests of Windows 95, it attempted to use memory locations in such a way as to conflict with the Symantec software and cause the system crashes. Once the Norton Antivirus protection was removed from the SIMPAH system, these crashes stopped and program operation returned to normal.

C2 Network

Networks, by their very nature, are dramatically more complex than single user environments. The CPD was designed on a single user environment. A number of unforeseen problems arose as it was applied on the network.

Multiple access file locks caused a major reprogramming of modules in which multiple users might access the same information at the same time.

The network does not allow individual programs to directly "talk to" printers. Rather, it stands as an intermediary which intercepts information bound for the printers, which it then controls in its own way. When this control problem was identified it was manifest in the difficulty of printing accents and the letter "ñ". The printing system of the CPD was reprogrammed to correct for this problem. However, by the end of the assignment, only half of the reports had been reprogrammed to avoid this problem. All of the reports print, but some do not correctly print the Spanish characters and the printing occurs without allowing the user to preview the results.

The network speed was found to be significantly compromised at certain stages of CPD operation. These problems were overcome by both altering CPD programming as well as working with the network to optimize its method of resource control. By the end of this assignment, network slow downs of significance were no longer noticed.

SECTION III

SUGGESTIONS FOR THE FUTURE

A Weights and Measures and Adjusted Sale Units

The key to the CPD's ability to manage data in a variety of formats, sale units, currencies and others is its ability to establish definitions of equivalency among data points. The most important in the system relates to the weight of sale units.

During the course of the assignment, PRODEPAH commissioned a study of the weights associated with different products. This study was very useful. However, it was quickly recognized that the study needed to be followed up in two ways. First, the number of products needs to be extended. Second, the work needs to be done at, perhaps, three points during the year to account for the significant variations in weights associated with subjective size indicators. That is, a "large" during the height of production is often significantly different in size from a "large" during the off season.

The implication for the CPD is to add to and adjust sale units according to the information gathered. The CPD, as installed is fully capable of allowing for SIMPAH additions to the sale units. There is a difficulty with alterations. Some of the sale units in use by SIMPAH were added using the best guess of the reporters as to the weights of the sale units seen in the markets. With greater experience and with the more formal work done in the study of weights and measures, some of these units need correction. All of those identified to date have been modified. However, there is no way for SIMPAH to modified those identified in the future.

End users of the CPD are not allowed to modify sale units that have been added to the system. This is an important quality control measure. If users were allowed to alter the meaning of sale units, all data previously input using the earlier definition are compromised. In order to verify that data errors are not implied if a sale units is altered, the data bases must be searched and the data manually reviewed to insure that the proposed change will not create an error.

It would be very helpful, after several additional months of operation, to revisit the issue of modifications to existing sale units so that the SIMPAH data are correctly brought in line with the weights as verified in the markets.

B. Training and Programming Follow-up

The CPD itself, is not a difficult system to use. However, the concepts associated with market information are complex. As a result, multiple training sessions have proven useful to solidify the understanding of users. The data inputter and the reporters are the ones who have benefited from these multiple sessions. Unfortunately, the reporters in San Pedro Sula have had

only two days of training and will most certainly need follow up. Also the data analyst has not been available for the training sessions programmed to include him.

As has been the case thus far, increased use of the CPD has identified programming errors or needed improvements. These will certainly continue. It would be useful, in a few months, to revisit the issue of adjusting the programming to account for these problems.

There are several hundred products and sale units which do not have the Spanish equivalencies entered. It would be helpful to have these added, so that SIMPAH staff properly identify the products or sale units of choice.

The assignment has concluded before a stable system of data transfer using the Internet, has been fully established. It will be helpful to future SIMPAH operations if this system can be added and stabilized.

C. Marketing Assistance - Training

SIMPAH is currently doing an excellent job at its primary function of generating and disseminating information on a daily basis. However, the long range vitality of the operation is highly dependent on the ability of the office to generate saleable products.

The development of these products and their marketing have been compromised to date for three reasons:

- a. The management of SIMPAH is still in the process of stabilization so that the collection of fees can be accomplished in a reasonable way.
- b. The affiliation of the SIMPAH directly with the government has caused a naturally high concentration of staff effort, particularly with regard to data analysis, on the needs of the government, rather than the private sector.
- c. Problems with the management of the budget have compromised the ability of SIMPAH staff to travel, interact with the private sector, and market services.

It would be helpful to the long range functionality of SIMPAH to provide for assistance in report generation and analysis for private sector clients, and most especially for marketing efforts.

There should be a programmed effort to make contact with and interact with other entities in Central America, the USDA and perhaps others, with whom data interchange arrangements can be developed.