Collaborative Research Support Program (CRSP)

NUTRITION INTAKE AND FUNCTION

931-1309

TRIP REPORT

MEXICO CITY, MEXICO

May 27 - September 7, 1985

Mr. Jeffrey R. Backstrand
Data Manager - University of Connecticut
Mexico Nutrition CRSP

TRIP REPORT OF JEFFREY R. BACKSTRAND
May 27 to September 7, 1985

Overview

With the exception of two brief trips to the fieldsite at Solis, most work was carried out in Mexico City. Work was done either at the **Instituto Nacional de la Nutricion** (INN) or at the IBM **Centro Cientifico** (see description below). My main activities were aimed at increasing the flow of data to Connecticut as well as providing aid and advice in solidifying and systematizing the data management system in Mexico.

Available Computer Resources in Mexico

The data management system in Mexico was considerably different in May, 1985 that the one in place during my previous visit in November, 1984. During the previous six months, a number of important changes had taken place with respect to data management resources available to the Mexico Project in Mexico:

(a) In early 1985, the Mexico project had purchased a Columbia IBM-compatible microcomputer with hard disk. This machine is and was capable of running the sophisticated data management programs available for the IBM PC/XT. The Columbia is also capable of using SPSS-PC (and the time, the future SAS for IBM compatibles). Moreover, the Columbia had the potential for being a smart terminal linked to INN's Hewlett Packard 3000.

(b) INN had recently installed a new Hewlett Packard 3000 minicomputer. This machine had been in the anticipated for many years, and was purchased for use in patient billing and records. However, the research divisions of INN were to be allowed access to the HP 3000 as lower priority users. The Hewlett Packard as of May had **SPSS** 9.0 and the relational data base management system, **Oasis**. INN staff were critical of the Hewlett Packard's slow operations. Nevertheless, the HP-3000 provided the Mexico project with an important new data management resource. In May of 1985, the Mexico Project's primary use of the machine was for data entry and editing.

(c) The Mexico Project had reestablished relations with IBM-Mexico's **Centro Cientifico**, and a contract was being negotiated between INN, IBM, and the Mexican Government. **Centro Cientifico**'s mainframe computer ran SAS, and was capable of processing large datasets with reasonable speed. **Centro Cientifico** is and was a free facility providing service by IBM-Mexico to research projects considered by the Mexican Government to be in

the public interest. In addition to its computer hardware and software, **Centro Cientifico** employees statisticians and other professional staff whose job is to provide advice the various client research projects.

(d) The Mexico Project had reestablished a working relationship with the Mexican Ministry of Public Health (**Salud Publica**), allowing use of their data entry system and personnel. **Salud Publica** maintains a large staff for the entry of administrative data. **Salud Publica** is capable of entering large amounts of data within a very short period of time. However, **Salud Publica** requires that data be entered in large batches, and only during slack times.

Prior to these developments, the Mexico Project was restricted to the use of Apple microcomputers for data entry and data management. At the time of my arrival, two Apple microcomputers were under repair, no doubt due to the stresses of field usage.

Although the Mexico project had recent access to relatively powerful computer resources, these resources had not been fully integrated into the data management system. The Columbia microcomputer had not been used because of a lack of adequate training in the use of the machine, and because of a lack of useful software. Moreover, because of the new computer resources, the future role of microcomputers in the project was unclear.

In the Spring of 1985, SAS had begun to be used by the Mexico Project for some data management purposes. Ramon Lira had begun to use SAS at IBM's **Centro Cientifico**. However, the most recent release of SAS available at **Centro Cientifico** was SAS 76, the 1976 release of SAS. The 1976 version of SAS is very limited in its capabilities, when compared with more recent versions of SAS. In addition, the documentation for SAS 76 is especially poor. No one on the staff at **Centro Cientifico** had any extensive experience with SAS, especially in the use of its data management features. As a result, Sr. Lira (a very competent programmer in advanced languages such as APL, and Fortran) was doing his best to learn SAS on his own. By May, 1985, Sr. Lira was doing basic SAS programming and was anxious to learn the advanced SAS programming necessary in order to work with complex datasets such as those collected by the Mexico Project.

As of May, the Mexico Project had used **Salud Publica** to enter data. The Mexico Project had also used **Centro Cientifico**'s mainframe and INN's Hewlett Packard (a) to edit files originally entered at **Salud Publica**, and (b) to create and edit the Sample File (a file containing data on the entry and exit of each person in the project sample).

Data Management In Mexico at Arrival: SOLIS

Early in my trip to Mexico I visited the fieldsite, Solis, in order to access the current data management system. I talked extensively with Elsa Molino (field administrator), Jorge Alba (Head of Data Entry in Solis), and Jorge's assistant, Francisco. In the time since my last visit to Solis, great improvements had been made in the tracking of data collection. A system had been developed in which the scheduling date and completion date of each test was recorded. As each questionnaire or test was completed, a record was made of its completion. Similarly, the each time a interview or test was scheduled or rescheduled, this was recorded. As each questionnaire came from the field it was checked by a special team for the correctness of the data on its face sheet (ID's, dates, names, etc.). Another team checked the face sheet of older questionnaires for correctness and completeness. By the mid-summer, the face sheets of all questionnaires had been checked. Any errors were noted and returned to the appropriate person for correction. The dataset containing the schedule and completion dates for each test was referred to as the "Archivo" or Archive File.

In the spring of 1985, a great deal of effort was expended by the Solis staff in double checking the Basal Census for correctness. This file had originally existed in Mexico solely on Apple diskettes. Connecticut also had a copy of these diskettes. The Basal Census data had recently been re-entered onto the Hewlett Packard mainframe computer at INN. The Mexico staff had engaged in a complex checking and rechecking, and correcting and recorrecting of both the Apple and Hewlett Packard files. These efforts left many of the Solis staff disenchanted with the data management efforts of the Mexico project, and resistent to any efforts at further data quality control, especially with respect to the Basal Census. It became immediately clear that much of the work concerning the Basal Census would have been unnecessary had the Apple file been directly transferred to the mainframe computer rather than reentered. Given the current access to both microcomputers and mainframe computers by the Mexico project in Mexico, a major need was the integration of the existing constellation of microcomputers, minicomputers and mainframes into one system in which data would be transferrable to all nodes in the system.

Data Management in Mexico: MEXICO CITY

Throughout my stay in Mexico, I spent a great deal of time with Ramon Lira (Data Manager/Programmer) and his staff. In May, staff were primarily involved in the correction of the Basal Census and the Archive File. Both files had been visually checked for accuracy by staff in Solis. Corrections were then made on the mainframe files, and new listings made and sent to Solis. By this iterative process, the Mexicans hoped to provide

as clean datasets as possible. There were, however, several costs associated with this time consuming process: (a) manpower was taken away from other data management activities, and (b) Solis staff were increasingly impatient with INN staff. It became clear that there was very little communication between Mexico City and Solis-based data management staff, and that this was non-productive in a number of ways. Little or no attempt had been made to explain to the field staff why they were being asked to certain things. No Solis staff had seen the data activities in Mexico City. In addition, Solis staff were correctly irritated by seemingly endless requests for a new iteration through the data. As a result, Solis staff would often not respond to requests from INN. The INN data management staff appeared to have stalled on just one facet of data management. A top priority for my visit became the establishment of a process by which priorities for data activities would be defined, and schedules for data entry, correction and flow established.

Accomplishments of the Trip

As a result of my observations, I established several goals for my stay in Mexico:

(1) The establishment of data flow between the various microcomputers and mainframe computers in the computer system. The following were accomplished by the end of my trip:

a) the linking via dedicated line of the Columbia microcomputer (marketed in Mexico under the name of Printaform) to the INN Hewlett Packard HP-3000 computer. As a result, the Columbia could be used as a terminal and for transferring data between the microcomputer and the mainframe and 9600 baud.

b) the linking of an Apple II and the Columbia microcomputer Printaform) by serial ports.

Achieving these two tasks allowed data to be transferred between all nodes in the computer system. Data could be entered using the Apple computers in Solis. The data contained on these diskettes could be transported to INN by car, and the data then transferred by serial port to the Columbia microcomputer using the Apple II at INN. This data could then be transferred to the HP-3000 for editing and transfer to tape. Once data were transferred to tape, they could be sent to any of the mainframes within the system: Centro Cientifico's IBM mainframe, INN's HP-3000, and UConn's IBM mainframe.

(2) Increasing communications between staff at INN and Solis with respect to data management:

a) Several meetings were organized involving Solis and INN staff regarding issues related to data management: coding strategies, scheduling, etc.

b) It was recommended that regular meetings between INN and Solis staff be initiated. This recommendation was later adopted.

(3) Instructing INN staff in the use of the Columbia (Printaform) IBM-compatible microcomputer.

a) I established a menu driven system for accessing programs on the Columbia's hard disk, making use of the machine easier for novices.

b) I installed the editor **KEDIT** on the microcomputer. **KEDIT** is a **XEDIT** look-alike, **XEDIT** is the editor used on most IBM mainframes. As a result, INN data management staff were able to use similar text editors at **Centro Cientifico** and on the Columbia in the Mexico office at INN. Early in the summer, data management staff began using the Columbia for preparing documentation.

(4) Increase the frequency of use and sophistication of use of SAS at **Centro Cientifico.**

a) I insisted in installing and testing the SAS82 version of SAS/CMS and SAS Graph on the **Centro Cientifico** mainframe. I also wrote a number of "EXEC"s in order to make use of the system as "user friendly" as possible.

b) when time permitted, I instructed INN staff as well as **Centro Cientifico** staff in the use of SAS. Unfortunately, the work demands of both INN and IBM staff did not permit as complete instruction in SAS as might have been possible. Nevertheless, the use of SAS at IBM **Centro Cientifico** increased markedly during the summer of 1985, both by Mexico project staff and by other researchers making use of **Centro Cientifico**'s facilities.

c) I wrote a number of complex SAS programs which allowed in country analysis of data, data management, and data quality control. The programs specifically dealt with the project areas of Morbidity, Anthropology, Diet, Socioeconomic data, Sociocultural data, the Basal Census, as well as the Archive file. Many of these programs dealt specifically with the area of data quality control. During the summer of 1985, the prevailing opinion of INN administrative staff was that all entered data should be checked against the

original data forms in Solis. I was opposed to what seemed to be an extremely labor consuming method of data quality control, and attempted to demonstrate the utility of computer-aided data quality control. By the end of the summer, the hand-checking method of data quality control had prevailed, with computer-aided data quality control adopted as a second level check for data quality. Although the hand-checking of data continued, the writing of these SAS programs proved beneficial in providing programs (i) which were used by INN and UConn for data management, and (ii) which provided models which could be used by INN staff for increasing their SAS skills.

(5) Develop protocols for the entry data. Codebooks were developed along with INN-based staff and Solis-base staff in a number of areas of the study. When possible, methods for double checking critical data were built into the data structure (e.g. in the area of diet, a Food Name was entered along with the Food Code, and a Recipe Names was entered with the Recipe Code).

(6) Develop a schedule for the flow of data to Connecticut. A schedule was developed for the entry of data, the checking of entered data against the original questionnaires, the correction of data files, and the shipment of corrected files to Connecticut. INN-based staff, Solis-based staff, and Connecticut-based staff all participated in the development of the schedule. This initial schedule was subsequently maintained. This first comprehensive schedule became the model for all subsequent data management schedules.

Summary

During this trip to Mexico (May 27 - Sept 7, 1985), I was involved in a wide variety of activities. Most activities were aimed at increasing the flow of data to Connecticut as well as providing aid and advice in solidifying and systematizing the data management system in Mexico.