A Lockheed Computer-aided Data Verification System
John Lau, Lockheed EMSCO
Greg Silva, Lockheed EMSCO

Abstract

Lockheed Engineering and Management Services Company (Lockheed-EMSCO) is under contract with the Environment Protection Agency (EPA) to provide data verification for portions of the EPA - National Acid Precipitation Assessment Program. This work involves Quality Assurance / Quality Control (QA/QC) of data from different laboratories. Originally the process was performed using a number of computer programs, as well as some manual operations.

Using SAS* products, a system has been developed to automate most of the process. The system allows users to create, modify, and verify acid deposition survey data. This system has been used successfully, and because of its general structure, can be adapted to future surveys.

Introduction

Lockheed Engineering & Management Services Company is currently involved in numerous EPA projects. Many involve the study of acid deposition, in which Lockheed performs the QA/QC. Although the basic fundamentals of the studies remain the same, the methods from sampling to analysis of data differ for each survey, so that the data vary in structure as well as values.

In early projects, scientists had to perform a large number of manual calculations with data from several sources in a large database. This method was inefficient and prone to mistakes. A more reliable system was needed to analyze and edit the database, as well as keep audit trails of data changes for later reviews. The system would also have to be adaptable to data from future surveys.

Using SAS Products

The SAS language was already in use at Lockheed when the system development started. Programmers and scientists had been using the language because of its powerful and easy-to-use data manipulation and statistical capabilities.

Although numerous SAS applications were developed, they were not integrated into a single system. When it was decided that a system should be developed, the SAS language was considered at first for the following reasons:

- SAS products are capable of processing a large number of complex records.
- The Fourth generation language features of SAS products allow programmers to enter pseudo-code in many instances. This means more time is spent to develop the overall system rather than to program details. As a result, development time is reduced.
- An editing facility already exists with SAS/FSP*.
- Many programs which help scientists review the data had already been developed using the SAS system.

When selected programmers returned from the SAS/MACRO course offered at the SAS Institute, it became clear that the SAS language would be the appropriate choice for the Lockheed system, since a modularized environment can be set up using the autocall library and SAS/MACRO*. Once the initial system was developed, enhancements and modifications were added.

In the current system, Lockheed library datasets are concatenated to SAS libraries during initialization. The datasets consist of members,
each a module whose name is that of
the macro defined within. The
macros become subroutines or proce­dures, and parameters can be passed.
From the user's point of view, the
above facilities combined with ISPF*
and CLISTs provide an efficient,
user-friendly system.

The System

The resulting product is a menu-driven Computer-Aided Data VERification System (CADAVERS). ISPF routines call CLIST subroutines which in turn call SAS macros. Menus are developed using the ISPF facilities available on the MVS/TSO operating system. ISPF panels (the menus) are fast and checks can be performed to obtain a valid user response. A series of panels are provided to determine which SAS routine to execute, and with which available options. A CLIST routine is called to invoke the proper macro with user selected options. That macro calls other macros defined in the Lockheed library as needed.

The Lockheed library consists of two sets of programs for each study. The first set integrates features that are common among all studies such as editing and browsing a database. The other set contains facilities specific to the each study. These include programs which analyze certain aspects of the data. There are also CLIST routines which perform operating system related tasks such as allocating and creating files, and submitting a printer job.

The SAS system is used for processing the study data. The SAS/FSP product allows us to create customized edit screens and simplify the entry/edit sessions. Edits are not updated to the database directly. Instead, they are stored as "transactions". These are printed and reviewed before being applied to the database. The SAS UPDATE command simplifies the update process. SORT and MERGE commands are used extensively for analysis and reports of large data.

Finally, SAS MACRO facilities allow the development of flexible programs which vary according to the study because of their dependence on the data structure.

Some of the functions available, all of which were developed using the SAS system, include:

- Edit / Browse the database.
- Browse / Remove transactions.
- Analyze the data. This step consists of about a dozen SAS programs which vary for each study.
- Update the database with a user's transactions. This function can only be performed by the database administrator.
- Review past records used for the update.

Advantages

There are many reasons why this system is better than those used in earlier projects. It is consistent throughout studies, faster, and keeps records of past changes. Because it is menu driven, it is simple to use. Perhaps more important, because the system is modularized, it can evolve to meet the needs of future projects.

There are also advantages in using the autocall facility. In addition to a well-defined library, the following are provided:

- Conditionally executed library macros are compiled only when invoked.
- The SAS system includes the macro code automatically at the call of a macro so the user does not have to define the macro.

These imply a more efficient tool than the past methods of submitting macro definitions.

Finally, there are the well-
known advantages of using the SAS system: faster development time because of less time spent in coding, powerful data manipulation capabilities, and compatibility throughout a wide range of computers.

Conclusion

The Lockheed data verification system has proved to be better than past methods used for QA/QC of acid deposition studies. The system is more efficient, consistent, accurate, and is flexible enough for future growth.

SAS products have played a major role in making this possible with its simple yet powerful features. In addition, with its other facilities such as SAS/GRAPH* and SAS/AF*, there is always room to grow and improve in the future without worrying about incompatibility or lack of user support.

Notice

Although the research described in this article has been supported by the United States Environmental Protection Agency through contract number 68-03-3249 to Lockheed Engineering and Management Services Company, it has not been subjected to Agency review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

* SAS, SAS/FSP, SAS/MACRO, SAS/GRAPH, and SAS/AF are registered trademarks of the SAS Institute Inc., Cary, NC, USA.

MVS, ISPF and IBM are trademarks of the International Business Machines Corp.