

## See ya CLIST: Migration of SAS® Applications from the IBM® Mainframe to the Web

Joyce C. Zandee, Parke-Davis Pharmaceutical Research, Ann Arbor, MI

Paul Lewis, Parke-Davis Pharmaceutical Research, Ann Arbor, MI

### ABSTRACT

This paper discusses the conversion and consolidation of several mainframe SAS® applications to UNIX using a single Web-based front end. The Web interface enables end-users to generate raw data listings and statistical summary reports from their laboratory data with no assistance from colleagues trained in SAS or the native operating system, as previously had been required when the analysis applications resided on the mainframe computer. Laboratory technicians use the Web interface to enter study protocol information into a SAS dataset and submit a SAS job. CGI programs written in Perl then access SAS programs and data on UNIX. The Web interface provides a mechanism for colleagues to review the SAS output and create PDF renditions of the resulting reports. Features of the application for viewing the SAS protocol dataset, editing text, and administering application security will also be presented. An on-line demonstration of the interface is planned.

### INTRODUCTION

When Parke-Davis IT management informed the Pharmaceutical Research Division in 1996 that all IBM® mainframe applications would have to move to another platform before the end of 1998, initial colleague reaction was less than favorable. The mainframe had been the mainstay of computing at Parke-Davis Research for many years and colleagues were reluctant to change. Computer specialists within the scientific disciplines who had become proficient in their use of ISPF, JCL, and CLISTs to evaluate and report study data were not enthusiastic about the prospect of mastering the intricacies of submitting jobs, printing reports, and managing files on a new platform. However, mainframe computing was no longer cost-effective for our research environment. The timetable for the move was firm as the decision had been made *not* to upgrade the mainframe operating system to be Y2K compliant. Colleagues who developed and maintained SAS applications took solace in the fact that their SAS code was portable. The challenge would be to develop new front ends that would allow scientific colleagues to evaluate and interpret their data while minimizing their need to learn a new operating system.

Management in the Pathology and Experimental Toxicology (PET) department viewed the Division's directive to move off the mainframe computer as an opportunity to give scientists more direct access to analysis of their data. If the new analysis interface was straightforward and easy to use, laboratory colleagues most closely associated with the data collection could be responsible for data analysis without the need to involve a computer-savvy middleman. To this end, a

collaborative effort between PET colleagues and Scientific Information Engineering (SIE) colleagues was initiated to develop an integrated non-mainframe solution for three SAS applications initially in production on the mainframe computer for the analysis of PET study data.

### BACKGROUND

Studies conducted by the PET department at Parke-Davis are used to determine the safety profile of new chemical entities (NCEs) in development. These studies are conducted and evaluated in compliance with Good Laboratory Practice regulations and are subject to audit by the Food and Drug Administration. Data collected in toxicology studies are analyzed to determine the risks associated with acute and chronic exposure to the NCE. The SAS Reporting System (SRS) for Clinical Pathology data (SRS-CP), the SAS Reporting System for Organ Weight data (SRS-OW), and what would become the SAS Reporting system for Body Weight and Food Consumption data (SRS-BWFC) were applications developed on the mainframe to produce raw data listings and statistical summary tables for data collected on general toxicology studies. While initially conceived as a single integrated application, the SAS code on the mainframe for the three data types diverged as different support groups maintained the code. Consequently, even though each of these data types were collected on the same study, protocol information including study title, compound (NCE), dose levels, and statistical test type were entered separately for each of the mainframe applications. Not only was this a duplication of effort, but it was also a potential source of discrepancies among the resulting reports for a given study. The move off the mainframe would provide the opportunity to reunite the applications in an integrated environment.

The goals of the SRS conversion project were to:

- move the existing SRS applications off the mainframe
- minimize changes to SAS code to simplify revalidation
- provide a simple interface to be used by laboratory staff
- use a common protocol file for all data types
- make the resulting analyses available as PDF files for on-line review and electronic regulatory submission.

The PET/SIE collaboration accomplished these goals in the resulting implementation of a single, integrated Web-based user interface for the SRS trio.

## SYSTEM ARCHITECTURE

UNIX was chosen as the most suitable analysis platform when security issues were encountered with the Windows NT® platform during preliminary testing. Users could not be authenticated by the native NT security subsystem because the NT server that SAS is installed on is in a different domain than the password file.

For the most part, SAS code for the analysis and reporting of study data was moved intact to simplify the necessary revalidation on the new platform. However, minor code changes were required to input information from the integrated protocol dataset, to remove JCL and add corresponding file assignment statements, and to accommodate the SAS version change from 6.08 on the mainframe to 6.12 on UNIX.

PET colleagues access an initial SRS home page through the PET Department Home Page on the Parke-Davis Intranet. Besides the SRS home page and a series of help screens, the web interface is composed of four primary panels that provide access to the various functions of the system:

- the Study Protocol panel
- the Generate SAS Reports panel
- the View Reports panel
- the Data Sourcing panel

The panels are written in HTML and JavaScript and reside on an NT server. Each of these four panels has associated with it a CGI program on UNIX that processes the user input and responds accordingly. The CGIs are written in Perl. Apache HTTP Server is installed on the UNIX box to handle requests from the browser. User access to restricted functions is controlled by matching the entered user ID and password to the entries in the Unix */etc/passwd* file using Perl's *getpwnam* function.

### STUDY PROTOCOL PANEL

The Study Protocol panel is used to access a SAS dataset containing protocol information for all studies in the system. Study records in the SAS protocol dataset are uniquely identified by the combination of a two-character location code and the study number. The user must supply the location code and the study number for the study of interest. User ID and password fields appear on the panel but are only required for adding new studies or updating existing studies in the SAS protocol dataset.

If the user submits the panel information by selecting the option to "Add/Update Protocol Information", the CGI runs a SAS job to determine if the entered location/study number combination already exists in the protocol dataset. If no match is found, a web page is built which allows the user to enter the protocol information and add a record to the SAS protocol dataset. If a match is found, a SAS program builds a web page containing all of the existing protocol

information and displays it in the browser. The user can enter protocol information including the study title, compound, statistical methods, and group labeling information that apply to all data types for the study as well as information specific to the analysis of each data type. If no match is found, a web page is built which allows the user to enter the protocol information and add a record to the SAS protocol dataset.

**Figure 1. Study Protocol Panel**

Selecting the "View Protocol Information" option causes the SAS program to build a web page containing all of the protocol information as above. However, the pull-down boxes, input boxes, and radio buttons are disabled. If no match is found, an error message is returned to the web page.

This panel also allows the user to copy existing protocol information to a new location/study combination, provided the location/study does not already exist in the SAS dataset. Additional options allow the user to Delete or Rename records in the SAS protocol dataset.

### GENERATE REPORTS PANEL

The Generate Reports panel passes information from the protocol dataset for the specified location/study combination to the SAS programs used to generate raw data listings and statistical summary tables for the data type selected by the user, with the output files directed to UNIX. Because the SAS jobs may take a number of minutes to run (2-7 minutes depending on the size of the study and network traffic), the SAS job sends the user an e-mail message to the user upon completion. This allows the user to minimize the browser and do other work.

After the output has been reviewed for accuracy, this panel also allows the reports for the selected data type to be "locked down" using the UNIX *chmod* command so the reports can not be inadvertently overwritten. The process of locking down the reports sends an e-mail message to a distribution list of scientific colleagues responsible for the interpretation of that type of study data. If a correction to the data necessitates that the reports be regenerated, the user

will “unlock” the reports for that data type from this panel as well.

Figure 2. Generate Reports Panel

## VIEW REPORTS PANEL

After generating the reports, the responsible laboratory colleague reviews the SAS log file and SAS output file through the View Reports panel and resolves any unexpected warning and/or error messages generated either by SAS or by the logic built into the SRS programming code. Then the colleague uses the Generate Reports panel to “lock down” the reports.

Figure 3. View Reports Panel

Study directors use the View Reports panel to review the reports on-line after specifying the data type, location, study number, and the report of interest. Raw data listings, statistical summary tables, and other reports specific to the selected data type may be viewed individually or concatenated as they would appear in the research report.

Hard-copies of any report can be generated from the View Reports panel, although study directors are encouraged to review the data on-line. Reports can also be saved as PDF files, accessible in our document assembly software, Documentum. The original ASCII report is first transformed into a postscript file then

Acrobat distiller is used to create the PDF. Based on the information supplied in the SAS protocol dataset, the PDF file is automatically saved to the appropriate study folder location within Documentum.

## DATA SOURCING PANEL

The Data Sourcing panel provides the user and system administrator an interface to view and/or edit data dictionaries, create customized analysis macros, input flat files specific to a given data type, and edit the security access file which specifies the appropriate access level for authorized users of the system. A user ID and password is required to access the security access file or to edit any file. The Edit/View function combines JavaScript and Perl scripting to create an interface with the input and output files on UNIX to permit viewing/editing of these files.

Figure 4. Data Sourcing Panel

## CONCLUSION

The SRS Web interface has provided a simple, integrated method for laboratory colleagues and study directors to evaluate and report their study data and has furnished a mechanism for on-line review and electronic submission of these data—major improvements over the application predecessors on the IBM mainframe. PET management and the PET/SIE collaboration effectively transformed the initially-resisted directive to move off the IBM mainframe into an opportunity to reevaluate and improve upon the way we analyze our study data and provide access to the reports.

## ACKNOWLEDGMENTS

SAS is a registered trademark or trademark of SAS Institute, Inc. in the USA and other countries. IBM is a registered trademark or trademark of International Business Machines Corporation. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

## CONTACT INFORMATION

Joyce C. Zandee  
Pathology & Experimental Toxicology  
Parke-Davis Pharmaceutical Research  
2800 Plymouth Road  
Ann Arbor, MI 48105  
Phone: 734.622.4034  
Email: Joyce.Zandee@wl.com

Paul Lewis  
Scientific Information Engineering  
Parke-Davis Pharmaceutical Research  
2800 Plymouth Road  
Ann Arbor, MI 48105  
Phone: 734.622.7934  
Email: Paul.Lewis@wl.com