Paper 158-28

Multi-Center Study Data Management With A Distributed Application

Derek Morgan, Washington University Medical School, St. Louis, MO

Abstract

Central management and coordination of data in a multi-center clinical study presents several challenges. This is an overview of the FHS-SCAN data entry system, a SAS® System-based Windows application using SAS/FSP® for data entry, SAS/AF® for study management and system administration, and SAS/AF frames for its ad hoc reporting tool. The system is designed to be deployed to all sites in this study; any run-time customization is data-driven. The administration of the program requires little computer knowledge and is done locally. Support takes the form of documentation (web-based and hard copy), system patches developed centrally and deployed locally, and in rare cases, remote connection and control of the local machine from the coordinating center using third-party software. The system is set up to handle any facet of study data management through the menus, so that advanced knowledge of operating system commands and procedures is not necessary.

Hardware/Software Configuration

Each machine is configured identically on delivery at the field center from the coordinating center, both in terms of hardware and software. The coordinating center provides support for all applications installed centrally, including the SAS® system, third party software used for data transmission, remote access, and archival (zip and unzipping) software. Each machine is equipped with an Ethernet card, as data transmission is handled via the Internet. Each field center also has its own Postscript printer attached to the workstation via the parallel port.

Data Entry System Overview

The data entry system is a SAS System-based AF application using SAS Component Language (SCL). The application has its own user authentication facility, which will prevent unauthorized access of data, as well as providing the capability of logging who does what within the system. The authentication protocol also provides for differing levels of access for individual users. The administration of user access is done locally, by a designated user (or users), selected and trained as “Data Managers”.

The data entry system contains menus for managing subject contact prior to clinic, including the ability to print mailing labels and to add new subjects to the study. It has separate menus for data entry and browsing. The “Manager’s Menu” is comprised of data transmission, user authorization, and access to the machine registration dataset. This is also where any functions that require a manager’s approval are located. There is a main menu selection for an authorized user to change his/her password as well.

Reporting

Several pre-defined ad hoc reports are available in the system relating to recruitment and study progress. These reports are developed at the coordinating center, and then deployed to the field centers as an integrated piece of the system. True ad hoc reporting, while possible, is not used here. The knowledge level of the system users ranges from computer-savvy, but not SAS-aware, to fairly computerphobic. It is easier to design and implement a requested report with some flexibility than to support a wide range of users in building their own reports from scratch. The latter requires some knowledge of the database, (at the least variable names and/or labels), which is more technical than some users desire.

Security

As previously noted, the system has its own authentication system, and the data are encrypted using the encryption available through the SAS System. The read, write, and alter passwords are contained in macro variables which are defined upon successful authorization. The source code for the authorization module is not stored with the program, adding a degree of difficulty to obtaining the passwords. Unless someone is logged in, the macro variables cannot be obtained from a programming window[1]. In addition, each user’s userid/password combination has a security level associated with it, which the data entry system uses to determine which modules can be accessed. We have five levels of access, with room for more:

10	**Browse only**
20	**Activity Logging**
30	**Data Entry**
40	**Local Data Manager**
50	**Data Coordinating Center Only**

These levels are cumulative, i.e., each level has access to any functions that a lower level does. This way, only the data manager is allowed to add and delete users, while a recruiter would only have access to reporting and pre-clinic functions, but none to the clinic data.

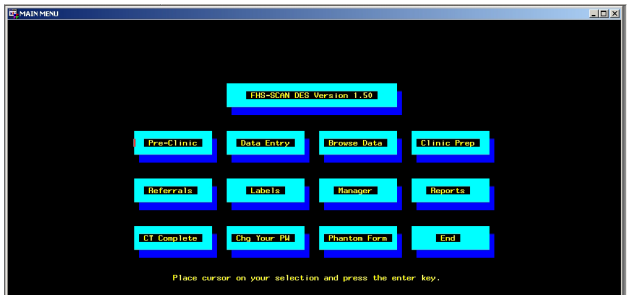
Maintenance

Maintenance of the system is handled at the coordinating center. When the system is changed, an update is released. We send the updates via Federal Express on a floppy disk. Using ZIP, a CD-ROM is rarely necessary. Any old files are overwritten with the new ones. If the database structure changes, or it is necessary to execute SAS System code within the application, a patch file with that code is also included in the zip file. This file executes the first time that a user logs into the data entry system if the data entry system has been updated. The system will also send e-mail to the programmer notifying them that

the data entry system on this particular machine has been updated[2]. Each field center computer is loaded with pcAnywhere, allowing remote control, diagnosis, and resolution of problems from the coordinating center if necessary.

The Menu System

All the system functions are accessed through AF menus.



The main menu (above) greets the user after a successful login. The title bar of the main menu displays the version number of the software, which is obtained from a SAS table. The main menu selections are:

Pre-Clinic:	Pre-clinic functions: label generation, address correction, and subject recruitment information.
Data Entry:	Entry and editing of recruitment status and clinic data.
Browse Data:	Look at data without making any modifications.
Clinic Prep:	Print forms for subject's clinic visit.
Referrals:	Enter and edit Referral Forms.
Labels:	Print laser labels for clinic forms.
Manager:	Open Manager's Menu. (Appropriate authorization required.)
Reports:	Generate field center progress reports for local use.
CT Complete:	Enter or edit CT scan completion tracking information.
Chg Your PW:	Change current user's password.
Phantom Form:	Enter tracking information for quality control samples.
End:	Exit the data entry system and go back to the login screen.

The Pre-Clinic, Manager, and Reports menus lead to submenus, while the others allow direct access to the functions noted. By organizing the main menu by task category, the user can focus on what needs to be done, rather than, "where can I find what I need to do?"

Pre-Clinic Functions

The pre-clinic functions encompass the study-specific tasks to enable, or to facilitate subject recruitment and recruitment tracking. Of these, recruitment tracking is not computerized, but this is more due to a lack of time and the fact that there are specific procedures for each field center, rather than an inability to use the SAS System successfully.

The initial subject pool was culled from a pre-existing study with contact data. However, these data are five years old, and people have moved or names have changed since then. Additionally, since this is a family-based study, most of the selections in this menu can be accessed wither via an index person for an entire family or for any individual. A menu selection leads to an address correction screen. The data in this table are used to generate labels and to correct the name-based mnemonic used by one of the associated labs in the study. Another selection in the pre-clinic menu provides bar coded mailing labels for contact letters. The field centers requested blank recruitment sheets to help them. After much discussion, (proving the value of careful systems analysis) a customized report was constructed with a summary page for the entire family and individual pages for each living member of the family. The entire family packet can be generated, or the summary sheet or an individual sheet for a given family member can be generated. The user selects the family, families, or individual(s) through an AF selection list built with an extended table. Both the mailing labels and the recruitment sheets have been built using an in-house package, the PostScript Macro Library[3] instead of ODS. Although the Macro Library is more difficult to use, it does provide the ability to produce bar codes without purchasing a commercial bar code font. Also, there is a great deal of legacy code in house using the macro library for many of these printing tasks, which reduced development time. ODS certainly does have its place in the system, as we will see later.

Finally, there is a menu selection to allow for the addition of subjects who were not in the initial subject pool. The field centers can either add someone to an existing family or you can add an entirely new family.

Data Entry

Subject information is collected and maintained in a subject folder at the field center. Since the data are collected by subject, the data entry is also organized by subject. This menu selection opens list of all subjects in an AF window. This list displays the recruitment status of each subject and is where the user selects the subject whose data they will be entering.

When a selection is made, a confirmation window appears asking the user if this is the correct person. This is also where the recruitment status is entered or changed. If the recruitment status is valid for data entry, then selecting "yes" will open the data entry module. If the user selects "no", or the status is not valid for data entry, they are returned to the subject list.

A list of forms is given, with the date and/or time that it was last modified for that subject. Any forms that are sex-specific only appear if the subject is of the correct sex (e.g., reproductive history for females). The informed consent tracking form must be entered before any other form can be entered for a given subject.

When a form is selected, it opens an FSEDIT data screen, where the full text of each question, along with its response categories, are displayed on the screen. The subject ID is already filled on each screen, having been passed through a macro variable created when the subject's name is confirmed from the selection list. This

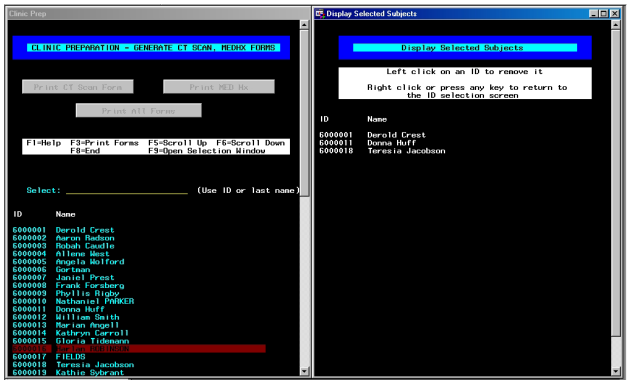
prevents mismatching ID numbers with subjects. Some of the screens use point-of-entry double data entry[4] for selected fields, but all the data entry screens are tightly controlled with a large amount of SCL programming[5].

Browse Data

This menu has browse screens to look at each dataset. The SAS DMS commands "search" and "locate" can be used to find specific subjects through the command line. In the browse screens, a function key is set that defines a WHERE clause to restrict the display to second entries of data, which are the data used for the study.

Clinic Prep

A large component of the FHS-SCAN study is the CT scan measurement for each subject, which is not performed at the field center, but at a nearby facility. The CT scan reading center is responsible for reading these tapes and providing the data to the coordinating center. The CT scan reading center requires that a scan completion form be faxed to them for each subject receiving a CT scan. This form has barcodes representing the subject's study identifier, and the mnemonic identifier used by the CT scan reading center. It provides space for patient safety information as well as important information for the CT scan technician. This form is generated using the PostScript macro library. The user selects the subjects for whom forms are to be generated through a searchable selection list. This list is displayed via an extended table that occupies the left half of the screen. When a subject is selected, another AF program screen opens on the right side, displaying the people that have been selected in an extended table that gets loaded as each person is selected.



Left-clicking on a name in the right window will remove that person from the list. That window is closed by using the right mouse button, and the user can then select another subject, or press one of the gray selection buttons in the left window to generate the CT scan form, a medical history, or both forms in one packet.

Referrals

This selection opens an FSEDIT data screen where clinical measurements and lab results for the subject are recorded. This is separated from the other data entry because not all the information to be entered is available at the end of the clinic visit, and pieces may not come available for several weeks. If the field center wants to wait until all data are in, they can enter it all at once. However, if they enter it on an as-available basis, this method gives them the most flexibility, since lab reports arrive in batches for groups of subjects. Unlike the main data entry, this is not maintained by subject. The "search" and "locate" DMS commands can be used to search for a specific subject. However, an easier way is to open a new observation and enter the ID number. If the ID number does not exist, then a new observation is ready for data entry. If it does exist, the SCL code will put the record for that ID on the screen without using a WHERE clause (and forcing the user to do a WHERE CLEAR to gain access to other subjects). In order to prevent accidental modification of an existing record, all existing records are "locked". All fields are protected against modification unless the user "unlocks" the record by pressing a specific function key.

Labels

ID numbers are printed on laser labels with bar codes. They are applied to the forms for each subject and scanned directly into the PC when necessary or convenient. Bear in mind that most data entry takes place with the ID having been selected from a selection list.

After the user selects the subjects using the paired windows previously described, the labels are generated, either as a page of 26 form labels and four laboratory labels (without names for blinding purposes), or the user can generate a single clinic or lab label for a specific person. For the latter, the number of the label where printing should start can be entered so that label sheets can be sent through the printer more than once if only some of the labels on a page have been used. Although this was a high priority item as requested by the field centers, there are no statistics on its use, nor is it apparent how many times a label sheet can survive being run through a laser printer.

Manager

The tasks in this main menu selection are restricted-use tasks. Not all data entry system users should have access to these. If the login ID does not have the appropriate permission level, the manager's tasks menu will not appear. Instead, an "access denied" window will pop up.

One of the main items in this submenu is the user password maintenance menu, where the local data manager takes care of system user login ID's, passwords, and permissions. They can change a password, look up a password, and add a user. You cannot remove a user from the database, but you can remove their authorization to access any of the functions in the data entry system.

This is also where functions that are necessary, but that need a manager's approval are located. If a subject changes his or her mind about the use of their data (rare), the data entry screen for that tracking form is accessed through this menu. Datasets can be cleaned and resorted through this menu. If any of the contact information for the field center (maintained in the register file and used on the CT scan completion form) needs to be changed, it is done through this menu.

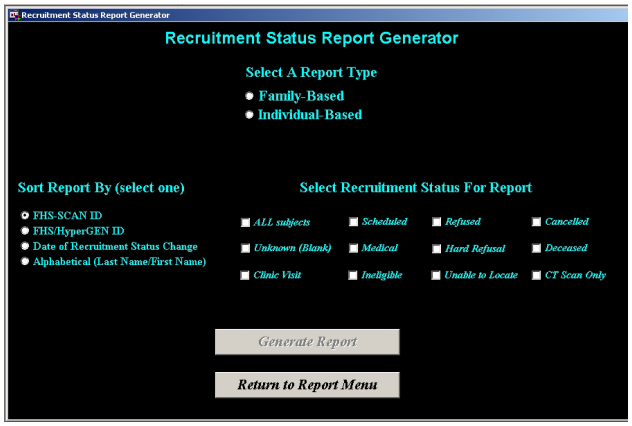
The module that transmits data to the coordinating center is located here as well. This program creates a zip file of

the data, giving it a name with a field center identifier and sequence number pulled from the field center register dataset, and then invokes an SSH file transfer program. When the data are transferred, an e-mail message is automatically sent to the coordinating center using the SAS System E-mail interface from SCL.

Finally, there is a coordinating center menu. The currently logged in user must have sufficient authorization with their ID to access the items in this menu. This is where records can be deleted, ID's changed, or specific data items in datasets changed.

Reports

This menu was developed in conjunction with the local data managers at each field center during the systems analysis phase. Using frames, an ad hoc recruitment report generator was developed. A frame was used since checkboxes and radio buttons were easier to implement, and were a natural fit for the intended design of the report generator. It allows the field centers to keep track of recruitment within families as well as overall individual recruitment. The field centers can define sort order and how to subset the data for the report.



For family reports, the user selects the families to report, while the individual report uses the entire database before any user-specified subsetting. For individual reports, a categorical summary is generated and displayed using an AF program screen, but users can also go to a list of all subjects represented in the summary. PROC REPORT and ODS provide hard copy reports of both summary and detail.

The CT scan status report is a detailed list displayed through an extended table in an AF program screen. There are four predefined subsets of the data that can be accessed by pressing a pushbutton. As in the recruitment report generator, hard copy is provided through PROC REPORT and ODS.

The alert status report gives a quick summary display of subjects with referral records through an AF program screen. It essential functions the same as the CT scan report. There are predefined subsets of data for the report that are represented by pushbuttons, and hard copy output. Unlike the CT scan report, this detail list is linked to the alert dataset and FSEDIT screen. Clicking on a subject's name will take you to the record for that subject.

CT Complete

This opens an AF program screen that allows the field centers to "close the loop" with respect to the CT scan visits. They also receive a fax copy of the finished scan completion form, and this is where they provide the coordinating center with notification that the subject has had their CT scan (or not, in the case of a refusal).

Change User Password

This is where the currently logged in user can change his or her password. This keeps the local data manager from being bothered with this simple request. The local data manager has master access to the security dataset so that he or she can change passwords for all users (except coordinating center level users) as well.

Phantom Form

This is where the record of blood samples collected for quality assurance is maintained. Existing observations are automatically protected against modification, and must be unlocked before they can be modified.

Exit

This takes the user out of the application, but it performs some housekeeping first. The datasets are sorted, any security macro variables are cleared, all work datasets are deleted, and the user is returned to a blank login screen.

System Summary

The system has been built to provide local access to the most important data and study management functions for this study, reducing the burden of support. With distributed applications to multiple sites, there is a very real danger that the time spent in supporting the application effort will approach the amount of time spent developing the application time. With the FHS-SCAN data entry system, this has not been a problem, although as with any software, the issue curve is highest at the beginning of system implementation. At the time of this paper (a year after the initial release), no issues have been reported in over a month. Given that the software is in daily use, this is most promising.

Remote Support Issues

Even though this is a distributed Windows application, no site visits have thus far been necessary to resolve issues. By using the Internet to transmit data, even data-driven errors can be quickly diagnosed at the coordinating center. The remote control software is only needed for immediate deployment of patches/transmittal of data, monitoring user keystrokes when the program will not fail at the coordinating center, and coordinating center-level tasks.

A detailed user's guide with step-by-step instructions and pictures of almost all screens has been developed, and is valuable for getting newer users up to speed. However, this is not the only solution, as people will still call for support. Again, the curve for support calls is much the same as with system issues, diminishing over time.

How It Works

Most of the system uses AF program screens. Other than the main menu, most submenus are pushbuttons or fields with colored highlighting. Since the CHOICE statement only allows for 12 characters maximum, this is a good way to provide text that is more descriptive with the same point-and-click functionality.

Selection lists are built using extended tables, as are the detail displays. Both are intelligently searchable with wild cards. This is much easier to do than it seems, and WHERE clauses speed access to the chosen observations.

SQL is used for a lot of the basic data manipulation for selection lists. It's fast, and doesn't require sorting. It's also easier to work with during debugging. The question of SQL vs. DATA step was usually answered by how much manipulation of the end result dataset was necessary. Some things are just much easier to do with a DATA step.

There is a register dataset that contains field-center and machine-specific information, which is how the program distinguishes its location. This information is used in e-mail transmissions, printed on forms, and restricts the access to data so that a field center only has access to its own data. To make maintenance easier, the application is identical when it is distributed; this includes the pre-clinic information stored in SAS datasets. No worries about sending the wrong data to the wrong place (which has happened in the past). The pre-clinic data are always subset based on the data in the register dataset.

The application is distributed with source (except for the password and login modules). Any additional overhead is negligible when you consider that you can run the debugger on the local machine through remote control software and see everything that is going on.

Rationale

Why would someone build a SAS System-based data entry system using old stale SAS System technology, on stand-alone Windows computers, when there are so many other options available, many of them much less expensive?

The answer is: legacy code, which equals fast development.

The FHS-SCAN data entry system is the most recent in a long line of personal computer-based data entry systems, beginning with the one developed for the original Family Heart Study, which started in 1993. At that time, the platform was an i486 computer, and we used version 6.06 of OS/2 PC SAS. AF frames didn't exist. ODS? Forget it. Many of the tools developed then are still in use today, with efficiency updates (SQL instead of DATA step, WHERE clauses instead of SCL searches) and increased security (encryption and passworded access to datasets).

Why haven't we updated our technology in terms of the SAS System? Given the quick concept-to-alpha time frames for these projects (as brief as 6 weeks, with a maximum of 3 months), it was a reasonable thing, given the amount of typing necessary for the questionnaire text, and the available programming staff. Using the existing framework eliminates the learning curve associated with new technology, and restricts the types of errors that can occur. As an example, the selection list technology dates back to that original system; if something fails, you know it's not the method.

FSEDIT is quick and easy for data entry. The fact that it handles the field to dataset translation automatically is worth an hour of programming time alone. With SCL and that hour, it can become a very robust data entry tool, helping provide cleaner data on the first pass. After using it in 6.06, the enhancements to SCL made it even more useful and we have many SCL tools for FSEDIT. Still, that pales in comparison to the fact that we have existing screens in FSEDIT with associated SCL that can be modified and do not have to be re-created from scratch.

Frames and ODS and PROC REPORT all have their place. They've made it possible for ad hoc queries without messy pull down lists (which may or may not display all the options without scrolling), and an easy printing interface. However, to convert all the other existing code to the new technology would not be time-efficient.

What about web data entry? Everybody who is anybody is doing it now, and all you need is a web browser. At the risk of sounding paranoid, do you know how easy it is to mess up data entry in a web browser? All it takes is a scroll button on a mouse. You have little control over the focus, and it requires that the user pay close attention to the task. Finally, there's the ergonomic concern of going from mouse to keyboard to mouse to keyboard...

Simulating the robustness of FSEDIT and being able to interactively use the tools available to you in SAS is not easy in a web environment. It has been done[6], but when you consider the investment in training and the learning curve, it won't fit the development time frame, either. The speed advantage of using legacy code has overridden all other considerations to this point. Variations on this system are currently running successfully in a number of studies, one of them a networked Windows environment using SAS/SHARE® and SAS/CONNECT®.

The Future

For many of the reasons expressed in the above rationale, this old technology still has more life left in it. However, in order to take advantage of advances in technology, we are looking to port this system to a LINUX server at the coordinating center. Field centers will access one central copy of the program and the data through an X windows client and SSH encryption.

There are new challenges here; the inherent delay of the Internet, which can range from not noticeable to, "It was faster when I had everything on my computer." There is a problem with printing, since the system no longer has guaranteed direct access to a Postscript printer. How can you simulate that connection, or at the very least, make it easy on the user? These questions are in the process of being resolved, since as you read this, the start date for the next study is but a few weeks away.

However, there is hope that the Siteman Clinical Information Portal[7] (SCIP) may provide a web-based thin client solution using SAS/IntrNet®. This technology holds the

promise of rapid application development, secure access, and the full power of the SAS System for data entry and management.

Summary

The FHS-SCAN data entry system is a mature distributed data entry/study management application, which shifts the responsibility for the data management to the distribution site. The application functions so that even with this responsibility shift, the data can be entered with reasonable quality, and certain specific aspects of study management are easy to implement. Good systems analysis practices and an understanding of workflow are critical to the success of the system.

Although it may use stagnant and old SAS System technology, the advantages of legacy code in small-to-medium-sized application in terms of rapid development cannot be understated. This is the primary reason that the system continues to be used in a number of studies. Future deployments of the system are planned to run from an X Windows server, using the SAS System under LINUX.

The promise of better web-based technologies for data entry may eventually render this system obsolete, but for now, it is a tool that has been custom-fit to its task.

Contact Information

Further inquiries are welcome to:

Derek Morgan
Division of Biostatistics
Washington University Medical School
Box 8067, 660 S. Euclid Ave.
St. Louis, MO 63110
Phone: (314) 362-3685 FAX: (314) 362-2693
E-mail: derek@wubios.wustl.edu

This and other SAS System examples and papers can be found on the World Wide Web at:

http://www.biostat.wustl.edu/~derek/sasindex.html

References:

1. Morgan D, Province, M. “Simplifying SAS® Security” *Proceedings of the Twenty-Seventh Annual SAS® Users Group International Conference.* April 2002 <http://www2.sas.com/proceedings/sugi27/p103-27.pdf>
2. Morgan D. “DISTRESS and PATCH: SCL to Support Remote Applications”, *Proceedings of the Twenty-Seventh Annual SAS® Users Group International Conference.* April 2002 <http://www2.sas.com/proceedings/sugi27/p035-27.pdf>
3. Morgan D. “A PostScript Macro Library,” *Proceedings of the Seventeenth Annual SAS® Users Group International Conference*, SAS Institute, 1992, 1049-1054
4. Morgan D, Province M. “Double Data Entry in FSEDIT Using Point-of-Entry Verification,” *Proceedings of the Twenty-Second Annual SAS® Users Group International Conference*, SAS Institute, 1997, 68-73
5. Morgan D, Province M. “Building A Better Data Entry Application Using PROC FSEDIT,” *Proceedings of the Twenty-Fourth Annual SAS® Users Group International Conference*, SAS Institute, 1999, 365-374
6. Thompson PA, Littlewood S, Adelman AJ, Miller JP. “The Web Data Entry System: Methods for Web Development and SAS Data Management” *Proceedings of the Twenty-Seventh Annual SAS® Users Group International Conference***.** April 2002 <http://www2.sas.com/proceedings/sugi27/p169-27.pdf>
7. Miller JP, Schauer, J, Baty JD, Littlewood S, Trinkaus K, Lill RM, Richards R. “Managing Clinical trials with a SAS-Based Web Portal,” *in press*.

Acknowledgements

This work has been partially funded by NIH grant U01 HL67899.

The author would like to thank the past members of the data entry system programming team: Joel Achtenberg, Harry Cheng, Sean Dyer, and Dr. Michael A. Province. Special thanks to: Greg Feitl, Nona Gallacher, Amy Haire, Emily Manders, Yvette McFarlane, Christie Oden, and Jan Skuppin, the local data managers who have survived the system’s growing pains and contributed to its utility.