A Common and Clonable Environment to Support Research Using Patient Data

P. Van Grevenhof
C. G. Chute, M.D., Dr.P.H.
D. J. Ballard, M.D., Ph.D.

Department of Health Sciences Research
Mayo Clinic, Rochester, MN 55905

The volume and complexity of information systems for data oriented clinical research at Mayo Clinic began to exceed available database-analyst resources. A decision was made to generate a clonable, generic system for health services research; this was implemented for a medium-sized prototype project. Since 1988, it has been reconfigured to support several score data projects at less than one-tenth the cost for database analysts. Further, user involvement and empowerment has increased their satisfaction and inquiry abilities with their specific systems.

Introduction

The Mayo Clinic has a long heritage of health services research, dating to the turn of the century[1]. Applied medical informatics, albeit paper-based, has flourished at Mayo in support of this research heritage, focusing on the patient medical record and multiple indices to it[1]. Evolving from a unified record of all inpatient and outpatient health encounters[1], to punch card indices[2], to computerized databases of patient parameters[2], we have created a large body of experience about information support needs for clinical research.

Today, the Department of Health Sciences Research manages over 10,000 archived study datasets, of which 1,500 may be "active" at one time. Over 500 peer-reviewed publications have arisen from one project alone attesting to the productivity and value of our patient data as a laboratory for applied health services research and clinical epidemiology. Support of these large and disparate research databases of clinical information has been an ongoing problem. Funded projects not infrequently created analyst demand greater than the limited Information Services resources. Developed projects also experienced delays and postponements for system maintenance, mandated by changes in the research environment. Procedural inefficiencies were perpetuated by this inability for analysts to meet demand, because investigators resorted to the time-honored practice of hand-abstracting computer printouts of clinical data and key-entering from paper forms; this despite the reality that machine readable archives of this same data were accessible.

Heartened by tradition and driven by demand, Information Services analysts within the Department of Health Sciences Research collaborated with clinical investigators in the Department to invest in the development of a prototypic project support data system that could be the basis of future project specific information systems. The principal intention was to enable an "off the shelf" approach to the information support of subsequent health services research projects, requiring only minor reconfiguration of the prototype system components; thereby speeding development. Rather than treat the data support needs of every research project as a Victorian era Swiss watch (meticulously made no doubt), we would acknowledge the industrial innovations of Henry Ford and maintain "generic" modules of software parts. The issues the prototype project were to address included:

1. multiple hardware platforms;
2. data transfers to other systems;
3. interfaces to other clinical databases;
4. common user interfaces;
5. training of end users in screen design and data definitions;
6. reduction of maintenance;
7. reduction of development time; and
8. empowering user inquiry.

Methods

Overview

Analysis was done to look at creating a tool box for researchers and study coordinators that would allow them to create their own system. The goal was to have a simple system that could be used with many different projects, that would take little programmer time to setup,

0195-4210/91/$5.00

required little user training, and would provide adequate backup.

The resulting tool box now enables the user to do their own changes and enhancements without programmer involvement. The system can produce data in an easily assessable format for data analysis and reporting. The system also allows for the possibility of customizing for special needs.

Environment

The original prototype system was constructed in 1987 using the SAS® environment[4] invoking the AF and FSP design tools[5][6]. The original hardware platform was an IBM 3090 mainframe under MVS. Current hardware platforms include SUN/UNIX, VAX/VMS, and IBM PC in addition to the mainframe, with TCP/IP network connections. Modular interfaces and extensions are being created in INGRES 4GL[7], DB/2[8][9], and the Information Engineering Facility[10] CASE development tool.

The Development of the Prototype System

The Prototype system was to support a study cohort of patients with benign breast disease. The patients were identified by using a registry of over 690,000 surgical cases, including biopsies. Patients were assigned to groups based on geographic location at time of surgery, i.e. local practice versus referral practice. Patient demographics were retrieved from an online registration system of over 2.5 million persons. The study protocol specified that only patients with current breast cancer status information were eligible requiring a follow-up task. Final enrollment numbered some 8,000 patients. A two-to-one randomized control group based on estimates of breast cancer incidence determined the number of patients who would have pathologic review of their histology slides and flow cytometry on their paraffin blocks. All patients with subsequent breast cancer also underwent pathologic review of historical benign breast disease tissue that is stored in the Mayo Tissue Registry.

The system was developed to have shareable information, although most users would only see the data necessary for their function. The study coordinator developed the screens and views of the database for all of the users of the system and defined the data elements. To start the process, lists were generated to call medical charts from among 4.3 million records in storage. The records were "checked out" to the study, with their location tracked by a bar code system. Three data abstractors collected and entered over 100 epidemiologic data elements into online screen forms.

From a menu selection of "Generate Letters" the database was scanned to determine which one of five different, pre-defined letters should be sent to each patient. The system also accommodated automatic second mailing and non-responder follow-up tracking. Letters were automatically generated with a high quality, digitized signature; registration database information was downloaded to personal computer mail-merge software. Any cancer information identified by the patient during follow-up would trigger a request for slides of the cancer tissue from the institution where the surgery was performed.

Patients in the randomized sample screening progressed to the next phase by issuing an automated request to Mayo's Tissue Registry for paraffin-fixed samples. After the requested tissue arrived it was then tracked by the prototype system for the shipping of slides and blocks to the pathology and flow cytometry areas. The study pathologist entered the information on 185 data elements while he viewed the microscopic slides using a terminal next to his microscope. His view of the data was blinded to the identifying variables to preserve an independent reading. The Flow Cytometry Laboratory also had a blinded view of the information, restricted to approximately 80 attributes.

Lessons Learned from Prototype

The automatic tracking of the medical histories, slides, and letters was the most efficient way to complete the study. The monitoring of compliance of all phases was built in. The same monitoring practices were carried forward to the resulting generalized system. Over 400 data elements were created using 17 screens. With the successful training of this user group in the creation of data attributes and screen layout, this process also was made part of the generalized system. The procedures that are used to transfer information from the mainframe to the PC with an automatic download were designed and put in place. Information retrieval procedures from clinical systems to any generalized system were designed and implemented for unattended batch mode operation.

New System Cloning

Each new system developed begins as a shell in which the user scans data elements potentially appropriate to their specifications. Then a screen painter is employed to create screens suited to their environment. They identify their needs for data capture from other clinical systems and requirements for passing information to other generalized systems.

With the onset of a new project, investigators, analysts, epidemiologists, and statisticians meet for a needs assessment. Guided by the detailed research

protocol written for that study, data elements are identified which are necessary for the project to achieve its goals, including statistical modeling of the data. The analyst then determines the style of the database and what elements exist in other clinical databases that can be retrieved for study purposes. Standard monitoring elements are incorporated to evaluate study progress. Tracking and follow-up of patient participants is evaluated and mapped to further data needs. This preliminary planning is structured with Joint Application Development (JAD) methodology[11][12][13] which fosters detailed and complete problem analysis. The JAD process is not lengthy for most research projects; the scope of activities is invariably well outlined in the original project research proposal.

After the preliminary planning JAD session, the analyst initiates retrievals from the relevant clinical data systems and assigns the appropriate security authorizations for users. Standardized procedures have been developed to access the many different clinical systems that reside on different operating system and on hardware platforms. Routine backup procedures are also set up at this time.

The study coordinator and data abstractors are trained in 2-3 hour sessions on creating the database, forms layout, and screen painting. A simple training manual was developed to reinforce these sessions. An iterative process follows as the study personnel create the screens and test them for flow and content, without consuming analyst time. When the screens near completion, they are paper tested with a sample group. This testing identifies problems before the database is built. The people closest to the project will continue to work on development of the information forms and only contact the analyst when all parties have signed off on screen layouts. At this time the analyst also goes over the screens with an eye for data modelling changes.

The study personnel then use the system shell to create the data elements, define data types and lengths, and add formats and descriptions. They also determine the legal value ranges and establish color cues for screen interaction. All this activity is done without database analyst time, reducing costs, streamlining user feedback and modifications, and providing the user with confidence and understanding about the system. As changes occur in data needs, or attribute ranges change, the study team is able and qualified to make those changes when and how they are needed. Users are pleased with the responsiveness of their systems to change, and analyst demand is dramatically offloaded.

Depending on the needs of the study or on the size of the study, some customized reports may be required. These report needs are analyzed and formats designed by database analysts, then turned over to the user as an optional part of their system. The user determines when these reports will be run.

Experience

From this discussion it is evident that the analyst is involved only at project definition phase and as an occasional consultant to add customizing features if necessary. This method puts the design in the hands of the people who know what they want and how they want it.

The software used in this generalized system has not required maintenance since it was installed, four years ago. As the study coordinators and researchers complete one study and start to set up another one, the method is the same. As the user base grew in the Department of Health Sciences Research, they developed a core group of study coordinator experts and the "help calls" to analysts about the system decreased substantially. There is now a pool of trained users who help pass on their expertise to the new trainees. They offer helpful tips on screening layout and design from past experience and what they like or did not like from the systems that they had created.

The impact on analyst time and development throughput is dramatic. Table 1 depicts the analyst hours billed for a large number of systems developed before our clonable base system was created, and contrasts these with two classes of projects developed in 1990. Reviewing times for projects premised on the new clonable resource, it is readily apparent that even the most ambitious system in the larger "General Projects" category does not approach the resources consumed by the smallest project created using the older, traditional development strategies. Average and median development times have plummeted by an order of magnitude. Perhaps the most striking feature is that the resources dedicated to creating the generalizable prototype system in 1987, fell well within the range of routine projects created in the same historical time frame.

Patient numbers incorporated in these intermediate sized study databases developed using this generalized software range from several hundred to 18,000. Some of the databases have over 1,000 data elements. Size has not been a problem as network disk space is expandable nor does the commensurate complexity thwart the logical approach. Users are able to do simple queries to help manage their system independently, further reducing dependence and costs for database analysts.

Table 1
Computer Analyst Hours for Initial Research Study Setup

	Average	Median	Range Min	Range Max
Prototype*	528	528	—	—
Historically Developed*	349	272	69	787
New Systems Cloned:				
Mini-studies	10	10	2	20
General projects	19	11	2	57

*These projects were developed before 1987.

The Department of Health Sciences Research employs nurse abstractors in the roles of study coordinators and data abstractors. Most that have been trained to use this system had virtually no computer experience, yet no difficulty in learning database set up or screen design was encountered. With the more than 30 systems established using this approach, we have not found anyone that has not been able to be trained in using the system.

Future Enhancements

Putting the user in greater control of their environment by having a seamless interface to a menu-driven report writer, query manager and statistical procedures is a major goal of our ongoing development. A prototype of this is in evaluation. This will decrease cost as system customization remains the biggest expense over the basic system.

We are pursuing the capability of referencing shareable data elements directly from clinical systems rather than duplicating them as is presently done. This will improve data currency. Only a few of our clinical systems now have the capability of online shareable data.

Developing a menu selection panel where users can select views of any clinical system is underway to improve the interface. Systems that do not have the possibility of shareable information will have a retrieval initialized and data will be duplicated in the user system. This then puts additional retrieval power into the hands of the user, without interfacing with an analyst. Information about what objects are available will be in hands of the user community and not just in the Information Systems Department. End users can then select which pieces of information they need, with the system automatically creating a data view of that information for them.

Conclusion

Investment in creating a generalizable study data environment for research using clinical data has more than paid for itself in reduced database analyst cost and dependence, involved and empowered users, and improved study conduct monitoring, accuracy, and technical validity.

®SAS, SAS/FSP, and SAS/AF are registered trademarks of SAS Institute, Inc., Cary, NC, USA.

Acknowledgements

Special thanks to Jo Johnson for her aid in the development and testing of the prototype system, and to Karen Elias for manuscript preparation. This study was funded in part by NIH grants CA46332 and AR30582.

References

1. Kurland LT, Molgaard CA. The patient record in epidemiology. *Scientific American* 1981;245(4):54-63.

2. Berkson J. A punch card designed to contain written data and coding. *Journal of the American Statistical Association* 1941;36:535-538.

3. SAS Institute Inc. *SAS/AF User's Guide, Release 6.03 Edition.* Cary, NC: SAS Institute Inc., 1988.

4. SAS Institute Inc. *SAS/FSP User's Guide, Release 6.03 Edition.* Cary, NC: SAS Institute Inc., 1988.

5. *INGRES ABF/4GL Reference Manual.* Alameda, CA: Relational Technology, Inc., 1989.

6. *IBM DATABASE 2, Version 2. Application Programming and SQL Guide.* San Jose, CA: International Business Machines, Inc., 1989.

7. Brown DE. *Guide to SAS/DB2.* New York, NY: McGraw-Hill Book Co., 1989.

8. *A Guide to Information Engineering Using the IEF™: Computer-Aided Planning, Analysis, and Design.* Texas Instruments Inc., 1990.

9. Martin J. *Information Engineering: Book 1--Introduction.* Englewood, NJ: Prentice Hall, 1989.

10. Martin J. *Information Engineering: Book II--Planning and Analysis.* Englewood, NJ: Prentice Hall, 1990.

11. Martin J. *Information Engineering: Book III--Design and Construction.* Englewood, NJ: Prentice Hall, 1990.