Journal of Medical Systems, Vol. 12, No. 2, 1988

Postrelational Database Implementation for Newborn Screening and Tracking*

Martin Wolfson and Mei M. Wu

Development of microcomputer based systems for the management of data in newborn screening programs has posed special difficulties. This paper describes the specific requirements for implementing an effective newborn screening database system using a unique postrelational database system. This system has been developed by Neometrics Inc. for use by laboratories engaged in running high volume newborn screening programs, and is currently operational in a number of state health laboratories.

INTRODUCTION

Newborn screening for a variety of metabolic disorders has become possible over the past 20 years as a result of advances in assay methodologies, particularly radioimmunoassay, which allows for the detection of minute quantities of substances using a small dried blood sample. Due to these advances, nationwide mandatory screening programs have been set up by each state. These programs can run most efficiently on a centralized basis whereby all collected dried blood samples are sent to a single laboratory for processing. State health laboratories were suddenly faced with the problem of handling many hundreds, and in some states thousands of specimens a day, in an organized and orderly manner. Reporting out accurate results and providing coordination for rapid follow-up and treatment were of primary importance.

The problem was to set-up a cost effective computerized specimen tracking system which could handle all of the following requirements:

Rapid data entry of all pertinent demographic data associated with each specimen.
Rapid on-line search for specimen data using a number of different parameters.
Automatic identification of retest specimens.
Rapid and accurate entry of requested tests and test results.

From the Neometrics Inc., 104 Bellerose Ave., E. Northport, New York 11731.

* This paper first appeared in the Proceedings of the Twenty-First Annual Hawaii International Conference on System Sciences, Volume IV: Applications. Washington, DC, Computer Society Press, January 1988. Reprinted here (in a slightly revised and modified version) by permission of the copyright holder, IEEE.

Hierarchical storage scheme for all specimen data, including many months of on-line data and long term archival capability.

Multi-level security system with full multi-user and network capabilities.

Extensive report and automatic mailer generation features.

Extensive back-up and recovery facilities.

Various existing microcomputers and database systems were evaluated for applicability to the above requirements. Most were found lacking due to inherent deficiencies in the database managers available to run in a microprocessor network environment.

REQUIREMENTS REVIEW

As stated above, the implementation of statewide screening programs for inherited metabolic disorders established a need for a system which could reliably handle the daily large specimen volume throughout and provide appropriate test results to those who would need to ensure early treatment for diagnosed cases. The obvious features of such a system include multiple data entry stations for entering specimen data. In many states this initially included data for positive specimens only. However, as the need for more permanent record keeping grew (based on legal as well as administrative requirements) most states opted for full demographic collection for all specimens received. In the larger states volume can be over 2000 specimens per day.

Laboratory test management is another required feature. This includes the segregation of specimens by type and test requested into batches appropriate for the various test(s) being performed. Since most states are running multiple tests on each specimen (in some, up to seven different tests are run) test batching has to be flexible enough to handle the specific requirements of each. Work lists by batch have to be produced in hard copy format for each test to be run. These of course must be available in a timely fashion so that the daily laboratory work load can be processed without delay. The work list for the batch must have the accession number for each specimen, its physical location and a place to enter test results. A pictorial representation of the batch which corresponds to the array of specimens is very helpful. Therefore, a batch work list for a test that utilizes an array of test tubes in a rack should look like the tube rack with the appropriate lab number placed adjacent to each tube's representation.

In some cases test results can be automatically transferred into the database. This is being accomplished with T4 and TSH tests for thyroid function which use computer stored algorithms to process gamma counter data into numerical results. Neometrics' CEM system provides identification of first level or presumptive positives through the use of user defined criteria. This includes the specification of a fixed value cut-off, a high or low percentage cut-off, or a cut-off determined by being outside a specified number of standard deviations from the mean. All specimens falling outside of the defined limits are marked as presumptive positives. Care must be taken to assure that appropriate laboratory management review takes place for all such classifications.

Manual entry of test results has to be performed rapidly for those tests for which an instrument interface to a computer is nonexistent. After results have been entered, the system needs to store and track data on those specimens which have been adjudged presumptive positives. This often requires the obtaining of a second specimen for confirmatory testing and the laboratory needs to know if a requested repeat specimen has arrived or is overdue.

The automatic generation of mailers with the newborn's test results is required for all specimens. These have to be generated rapidly and accurately. Also required are the generation of various letters which are to be sent to the submitting hospital or physician, the parent and in some cases to the Maternal and Child Health organization within the health department who typically is responsible for coordinating treatment and follow-up. In some states, automatic transfer of positive infants' data records to a separate database managed by the M and CH organization was required. This provided a vehicle for monitoring the effectiveness of a long-term treatment program.

The system must of course be available for on-line inquiry from sources both within and external to the health department relative to the status of any specimen received. Doctors or other health officials will often call the state health laboratory to inquire about test results on a specimen, often with only partial information about the newborn. The ability to perform rapid on-line searches on a variety of key parameters was very important. Periodic statistical reports are also required by laboratory administration to measure specimen volume, costs, and other pertinent data including disease incidence rates for various categories of specimens. For example, it would be important to be able to distinguish between the incidence rate of a particular disorder for low birth weight versus normal birth weight infants.

Archival storage of complete test results and demographic data for many years are also required. Laboratories must be able to prove that appropriate testing and notification was provided to a sick infant. Law suits resulting in multimillion dollar settlements have resulted from missed cases during screening for inherited disorders, and in many states these suits can be brought until the child reaches the age of majority.

The system must also include the ability to achieve rapid recovery in case of a database crash. Therefore, back-up, recovery facilities for the system must be extensive, rapid, and reliable.

POSTRELATIONAL MODELING

A database management system, MDBSIII, developed by Micro Data Base Systems in Lafayette, Indiana, was found to satisfy all of the criteria for the newborn screening application. It can support the 40 plus megabyte data files which had to be available on-line and had no limitations as to the number of records per file. It has a wide range of applications development tools and utilities and can work with many different operating system and hardware environments.

A major feature of this system is its ability to use and model relationships between data elements as they really exist and without the constraints required by many other database systems. The system data model provides logical data structuring facilities that are not available in traditional hierarchical, relational, or inverted data models.

The primary difference among the various approaches to logical data structuring is the methods employed for depicting a relationship between two record types. While this paper is not intended as a tutorial in database design concepts, suffice it to say that using a postrelational model, schema design could be easily and effectively accomplished. This

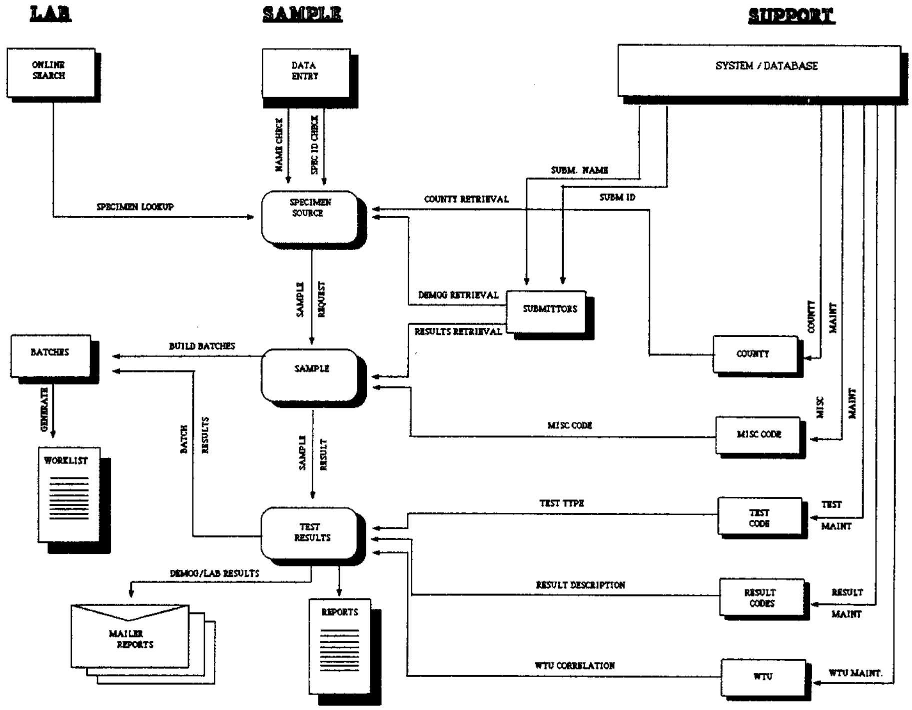

Figure 1. DATA MODEL

data model, an example of which is shown in Fig. 1, allows the modeling of complex relationships without restrictions.

APPLICATIONS DEVELOPMENT

Functional specifications for the system were created after a thorough and comprehensive understanding of the end user's requirements was established. A schema was then developed which minimizes the redundancy of data and relationships, and which attempts to minimize the database storage required. The data model, or schema once defined pictorially, is formally specified by use of a Data Description Language (DDL). Through this language, the applications programmer issues commands to the database management software which are used interactively to handle all of the physical details involved in creating, modifying, deleting, or retrieving data. The DDL also includes a facility to determine if any inconsistencies exist in the specifications, as well as generating the data dictionary and initializing the database. Modifications and expansions to the database can be accomplished through a redefinition of the DDL.

Applications programs written in Pascal (other host languages can be chosen) are then used to provide the mechanism for entering or extracting data from the database or modifying data that already exists. These programs are menu-driven and serve primarily to generate calls to the database management system for some form of data manipulation. These database management routines use a language called DML (data manipulation language) which is independent of the host language, operating system, and hardware employed. A form of DML called IDML allows for interactive use of DML commands.

Use of the DML in developing applications programs depends only on detailed knowledge of the database's logical structure. Applications programmers need not be concerned with using pointers, searching indices, disk I/O, file handling, free space management, etc. Automatic record locking is provided, thereby preventing problems associated with a user attempting to modify a record that another user is reading and thus providing true multiuser capabilities. Further, all DML operations are subject to the security constraints as defined in the DDL specifications.

Security can be provided via passwords, encryption of data items, and access codes which restrict a user's access (either read or write) to certain parts of a database. Every access to the database requires the use of a password. In addition, the applications developers can assign a combination of access codes to any user to establish that user's read and write access authorization.

Encryption allows any data item defined in the DDL specification to be encrypted. This protects the database from unauthorized viewing that can occur though a sequential scan of the database. For example, if name and test results information are stored in the same record, access to the database via the operating system could compromise security if the data are not encrypted.

SYSTEM IMPLEMENTATION

Once the database management system was decided upon, an operating system and microcomputer hardware were selected. This system, in which each user as a dedicated slave processor is a true multiuser system. The operating system also supports networking so that multiple CPU's can be linked together. Various network implementations include automatic transfer of chemistry data from the Neometrics' CEM system to the central database system, and the uploading of selected data to the main frame within the health department. By using a standard interconnect scheme, a tailored set of cost-effective peripherals could be selected for each state lab, thereby providing a high degree of flexibility.

Implementation of the MDBS database system required the development of host applications programs in Pascal as previously described. These applications routines served to generate the menu-driven features of the system. An example of the host language routines and the menu hierarchy is shown in Fig. 2. Each routine is activated by the user using the menu select features of the system. As can be seen from the figure, frequently required search formats and reports are hard coded in the host language. Less frequently required reports can be obtained using the query language facility provided by MDBS.

The database system contains facilities for recovery. These include transaction logging facilities in which all transactions since the last database back-up are automatically logged onto the log file. In case of a crash, these transactions can be applied to the previous back-up copy of the database thereby producing an up-to-date database. Also

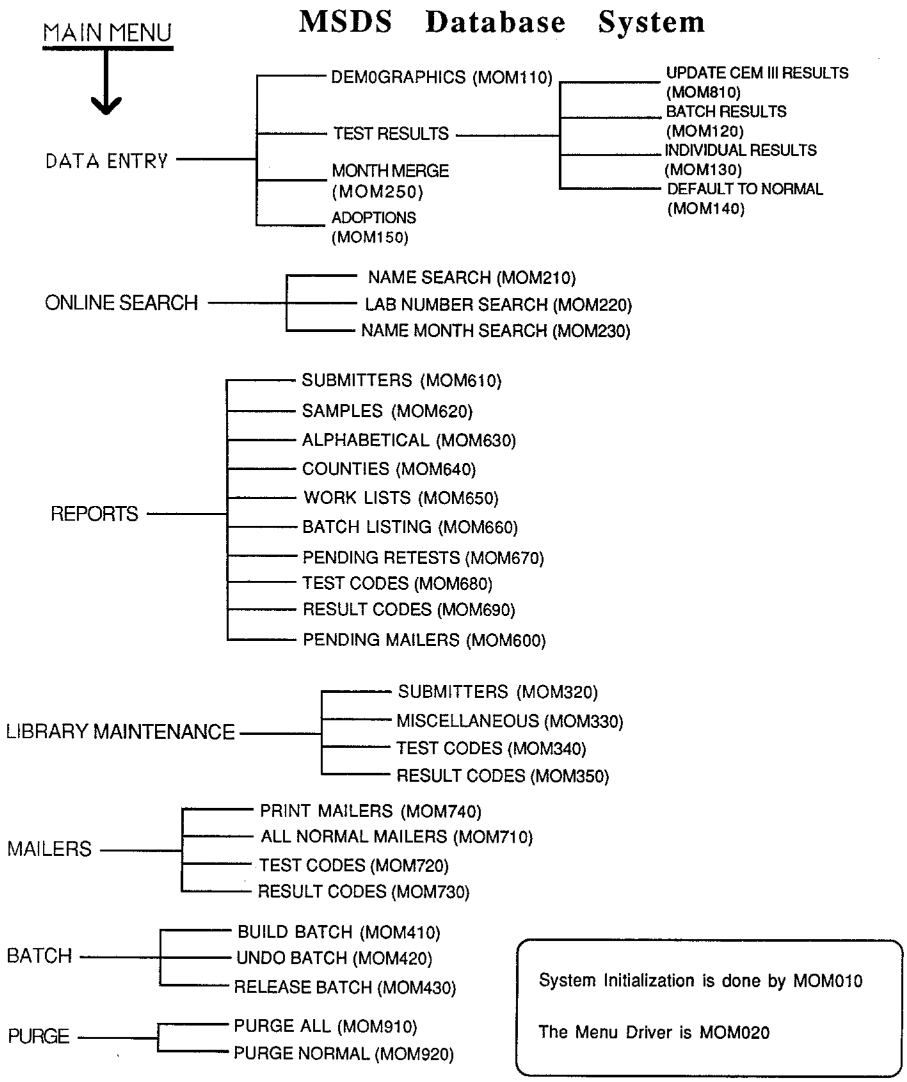

Figure 2. Menu Hierarchy

available are page image posting of complex transactions. An abnormal termination of the program in the middle of a complex transaction will not leave the database inconsistent.

CONCLUSION

Excellent performance has been obtained for this newborn screening application, through use of the MDBS III database system. Its ability to allow postrelational modeling

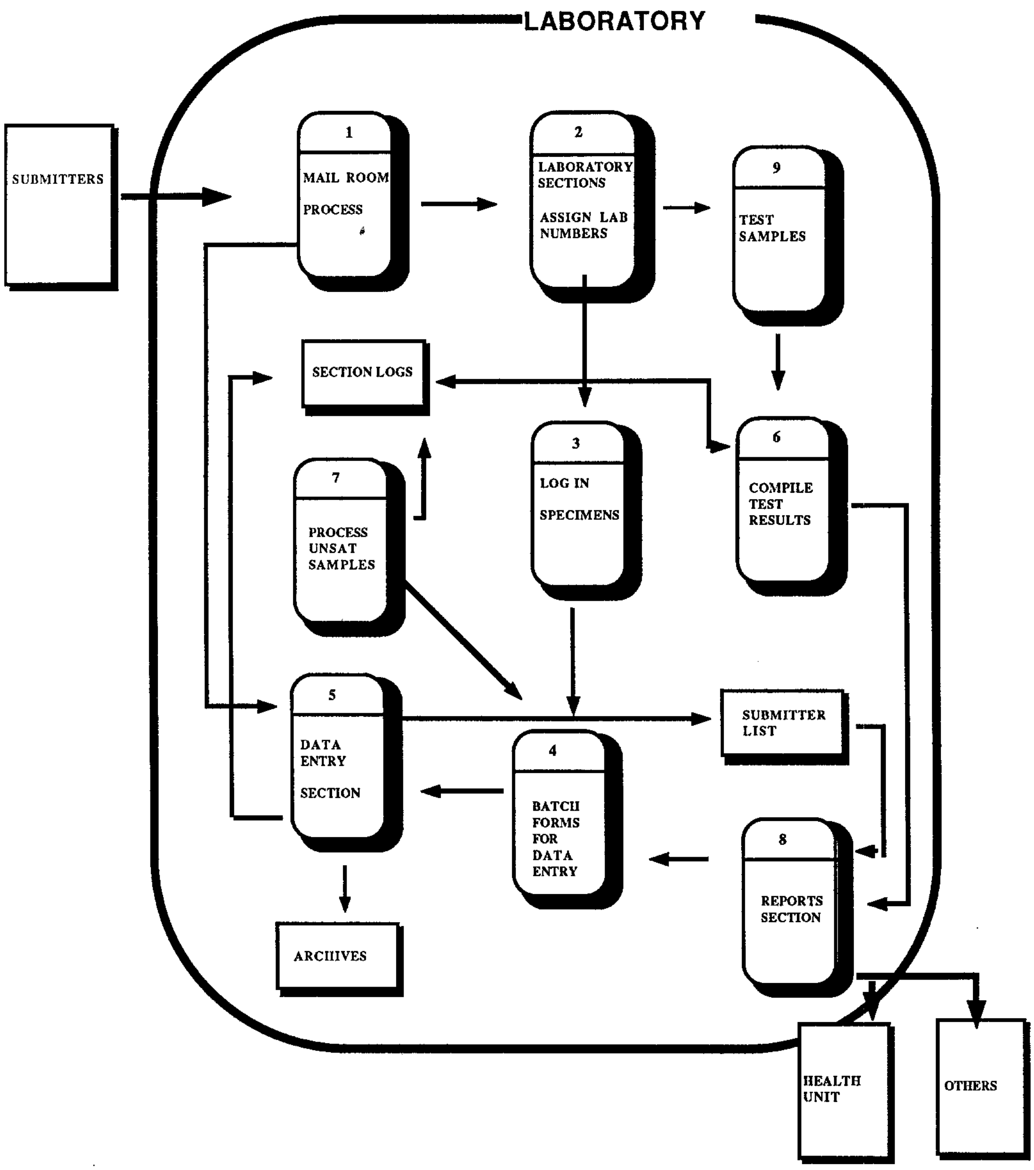

Figure 3A. Health Department Interface Model.

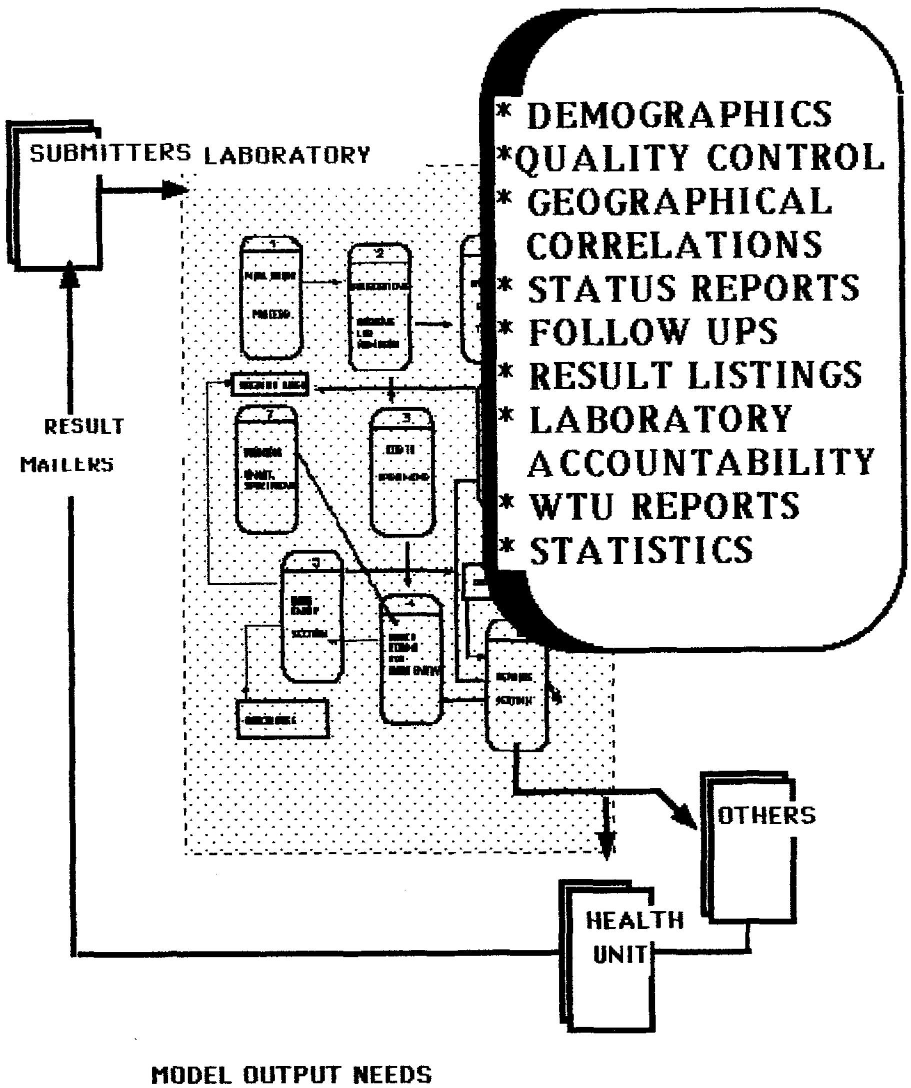

Figure 3B. Output Reports.

of all data elements, its hardware and operating system independence, and the extensive support utilities which are available have allowed the development of extremely complex and sophisticated applications in a timely and cost-effective manner.

At present, MDBS based laboratory tracking systems developed by Neometrics have been installed in Virginia, Louisiana, Missouri, South Carolina, Alabama, and Georgia. These systems have been modeled to produce similar performance to the Texas Newborn Screening Management System developed by Neometrics several years ago. The Texas system, although employing less sophisticated data management techniques, effectively handles a workload of over 2000 specimens per day with virtually no down-time.

A typical installation would include three or four data entry stations which would be used for demographic and test results entry. Demographic data entry for a workload of 500 to 800 specimens per day takes 2 to 3 hr which includes checking of the incoming specimens against a retest pending file. It takes an additional hour to build batches and generate work list for all the requested tests. Test results that are not automatically merged into the database are entered by work-list batch. Only the positives are entered individually, the remainder being defaulted to normal. This procedure is rapid, requiring 30 min for the daily workload.

Merging results from CEM requires an additional 30 to 40 min. Test results mailers and various letters are batched and usually spooled to a print file overnight. This procedure takes from 2 to 8 hr depending on specimen volume (2500 mailers in 8 hr), and provides sufficient time to generate back-ups and perform other system maintenance functions during the day.

Many of the concepts employed in the development of the newborn screening system are now being applied to a broader range of laboratory applications, including an integrated multilaboratory installation which ties many diverse laboratory operations together. Figures 3A and 3B depict the interface model used for the implementation of the application modules. These applications prove how effective microprocessor based systems can be when used with the right tools.