A public health data system

by JOHN C. PECK and FRANCIS M. CROWDER

Clemson University
Clemson, South Carolina

In mid-December, 1972 Clemson University became actively engaged in the technical development of a Public Health Data System for the Appalachian II District Health Department. Phase I of the implementation (ending June 30, 1973) has now been completed and has allowed both the Health Department and Clemson to more accurately measure development costs and benefits.

PUBLIC HEALTH PROBLEMS

Several factors which distinguish public health care from other clinic or hospital care are:
1. The size of the target population requires the necessity for extremely large volumes of data.
2. The widespread geographical distribution of the target population and public health service facilities makes record access extremely difficult.
3. The mobility of the target population is relatively high. A patient may require service in several facilities which have no records or knowledge of health care in other facilities.
4. Redundant data (name, address, sex, race, etc.) is captured not only in facilities at remote geographical locations but also within different programs within the same facility. In addition, many immunizations, lab tests and other services may be provided unnecessarily because of inadequate data access arrangements.
5. Standards for reporting of health care service are nonexistent; and because of the large number of health care providers who treat the patient, records are many times incomplete or ambiguous.
6. Manual systems typically present health care information in a source oriented sequence as opposed to a problem oriented sequence. Lab records are all stored together, immunizations together, prescribed medications together, etc. The problem oriented approach stores all information relating to each specific problem together. The results of a lab test, for example, must be interpreted in reference to the medications which are being taken. By storing information in the problem sequence, each patient complaint can be followed from beginning to end by any health care provider currently treating the patient.
7. Time series analysis of health related data becomes difficult with manual records. For example, a blood pressure of 140/80 may not be abnormal for some patients; however, if three months ago it was 120/60 then the physician may have ample reason for concern.
8. Investigation of trends in health care or disease becomes extremely laborious if not impossible with manual records.
9. Without a centralized data base drugs and/or physical conditions which conflict are difficult to discover. For example, many drugs should not be taken together or perhaps not be taken by pregnant women. If patients are being treated out of different facilities with different sets of records, conflicts of this nature may be difficult to determine.
10. Federal and state reporting requirements are becoming more time consuming and, therefore, provide less time for the health care professional to spend treating patients.

AUTOMATED HEALTH CARE

With an automated data file and remote access to the data, many of the problems inherent in manual records can be overcome. In addition, information processing and reporting procedures which were impossible to undertake with manual systems become simplified.

A brief description of the major points of implementation included in Phase I follows.

1. A master file consisting of patient related personal and socioeconomic data as well as pointers linking the patient to specific public health programs is created and maintained from cathode ray terminals (CRT's) located in each clinic. These terminals are connected to the Clemson University IBM 370/155 computer via telephone lines and can access and modify health related data under certain circumstances. As data is entered into this file it is edit checked for validity and flagged if incorrect or inconsistent so that the CRT operator may make corrections immediately. Records for approximately 100,000 persons in a two
county area are now accessible online. Expansion to a thirteen county area will be completed within the next eighteen months. (See Figure 1.)

2. A cross reference file, based upon the spelling of the last name, provides for quick and easy access to the master file data. When a patient master record is required the last name (and birthday, if known) is entered into the computer with the CRT. The computer then searches for all patient records whose last name "sounds" approximately the same as the name entered and displays them on the screen of the CRT. If the birthday was entered then only those patients with the sound-alike name and born on that date are displayed. (See Figure 2.)

3. Lab data related to a specific patient can be displayed in a time series format in which the last three lab reports are displayed side by side. Previous lab data can be displayed by backing up in the file using a "B" command on the CRT. (See Figure 3.)

4. Patient history data can also be displayed for any specific patient. This data provides a history not only for the patient but also for ancestors and siblings of the patient. (See Figure 4.)

5. Family planning regular checkup data is displayed in a time series format similar to lab data. A backup feature exists for checkups before the last three is also available. (See Figure 5.)

6. Immunization data can be displayed in a format which indicates the sequence number of each innoculation in a series along with the date of the innoculation. (See Figure 6.)

7. A "write" function is provided which allows the CRT operator to obtain a hard copy printout of any display on the CRT. In addition, special mark sense forms for the collection of additional immunization data can be prepared by the computer.

8. A "help" function is provided which lists all commands for the CRT operator in the event he cannot remember how to interact with the system. (See Figure 7.)

9. A "send" function is provided so that the CRT operator can communicate with the computer operator on the Clemson campus about any special procedures required on problems he might be experiencing.

10. Special data recovery programs have been written to guarantee integrity of data in the event of a device malfunction in which the data being collected online in the data being collected online is destroyed.

11. Special security functions ensure that unauthorized personnel cannot access patient records. In addition, several levels of security clearance are provided for special data access functions. A predefined user identification and password must be supplied before the program can be initiated. Certain users can access all patient data while other users can access only a subset of the data. A clerk in the front office of the Health Department, for example, can enter and update patient records; however, he may not access medical records for the patient.

Figure 1—Patient master

Figure 2—Cross reference display

Figure 3—Laboratory display

Figure 4—Family history display
Phase II is currently in various phases of implementation and includes the following projects:

1. Scheduling subsystem to appoint patients to various clinics or health care providers subject to various constraints. Included in this subsystem is automatic notification of pending and missed appointments along with related reports for governmental and accounting requirements.

2. Data base management graphics display subsystem. A language allowing questions to be asked about statistics related to the data base will be developed. Answers to these questions will be answered in graphical form on a graphics display terminal. A typical question might be: "What percentage of the women between the ages of 15 and 35 in the family planning program over the last six years have also been enrolled in the venereal disease program?" The answer to the question might be a graph in the form:

```
100

<table>
<thead>
<tr>
<th>Year</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1968</td>
<td>50</td>
</tr>
<tr>
<td>1969</td>
<td></td>
</tr>
<tr>
<td>1970</td>
<td></td>
</tr>
<tr>
<td>1971</td>
<td></td>
</tr>
<tr>
<td>1972</td>
<td></td>
</tr>
<tr>
<td>1973</td>
<td>50</td>
</tr>
</tbody>
</table>
```

4. Child Health subsystem.
5. Tuberculosis subsystem.
6. Veneral Disease subsystem.

APPROACHES TO IMPLEMENTATION

During the initial design phase of the Public Health Data System, several assumptions concerning the ultimate operating environment were made:

1. The program should be able to support an arbitrarily large number of terminals. Thus, modules must be designed to support re-entrant coding techniques.
2. Input/output logic should be confined to a central module so that commercially available data base management packages can be used if economically advantageous.
3. Program organization should lend itself well to conversion to a teleprocessing monitor system such as CICS (IBM), INTERCOM (GTE), or TASK MASTER (TURNKEY ASSOC.).

The development of the current operating package was carried out through several phases on Clemson University’s IBM 370/155. The time sharing option (TSO) of IBM’s operating system (OS/MVT) was used in the first phase of development for several reasons:

1. Program preparation and debugging was much easier in an interactive environment.
2. The main storage requirement for a non-time-shared implementation during program development would have been prohibitively expensive.
3. Only one data communications system was under development on the computer system so that costs for package systems could not be shared.

As developments of additional system features were begun and as the number of terminals serviced by the system increased, phase two of the technical development was initiated. The objective of this phase was to modify the system to run outside the time sharing region with a mixed...
environment message control program. Several factors prompted this change:

1. In the TSO environment each user had a separate copy of the program and was swapped between memory and disk storage as time was shared between users. With a large number of time sharing users the swap time became excessive and response time for transactions unacceptably large.

2. When an input or output operation is started by a time sharing program, no other time sharing program may be serviced until that I/O operation has been completed. Chaining through index structures and overflow areas using IBM's index sequential access method was locking out other time sharing users for the duration of the I/O.

3. Even with the carefully designed overlay structure implemented in the modular program, main memory allocated to the time sharing region became insufficient.

In order to move away from the time sharing environment, a driver program was written which provides the following functions:

1. Data blocks for each terminal required for re-entrant (multiple user) support must be maintained in memory or disk storage. A paging routine, using a least recently used algorithm, coordinates the swapping of data blocks between memory and disk storage.

2. Additional security provisions became available since the driver program controls and coordinates input and output to all terminals and users.

3. Special operation commands allow privileged users to assume the role of a master station and control the activity of other terminals and users.

The primary disadvantage of removing the system from the time sharing region was that a considerable amount of main storage must be dedicated for long periods of time throughout the day. Changes in the file organizations and access methods are now under way which will reduce the operating system data management storage requirements.

FUTURE DEVELOPMENT

Future developments depend upon many factors—the most important of which include continued acceptance by medical personnel and adequate funding. Eventual plans include the complete automation of the medical record to be stored in a centralized database with computer terminals in all health care centers in a multiple county area. Records will be cross-indexed by both problem and source so as to provide standardized and organized files which can be quickly accessed by health care providers.

Consultants in the medical profession will provide direction in the design of information displays and interpretation. This procedure will greatly help overcome many of the acceptance problems.

Additional computer hardware will eventually become a necessity if the system is to become completely operational. A central processor on the order of an IBM 370/145 will be dedicated to the application. A backup processor must be available to support the system in the event of a hardware failure on the main system. The Clemson University computer would serve as the backup computer without significantly degrading performance for campus users. Production will be restricted to the Health Department computer while development and systems support work will be restricted to the Clemson University computer.

During January 1974 Clemson University will acquire an IBM 370/158 and will run VS2. At that time all overlays and paging currently provided by the PHDS control program will be performed by the operating system. All ISAM files will be converted to VSAM which will provide faster access in the new environment.

A terminal monitor system capable of supporting multiple applications will be installed. IBM’s time sharing option will still be used for program development but not for production support.

Terminal independent input/output subroutines, which make possible logical level programming for CRT’s of arbitrary screen size and special features, will be written. Device dependent characteristics will be specified as constants and used with table driven logic keyed to terminal identification or line number.

The relationship now enjoyed by Clemson University and the Appalachian II Public Health Department will hopefully continue to benefit Clemson by providing areas for research in medical information systems and benefit the public health in South Carolina by providing the necessary technical expertise to successfully implement the automated systems.