Issues in the Design of a Clinical Microbiology Database within an Integrated Hospital Information System

Berl E. Nussbaum, M.D., M.S.
PHAMIS, Inc.

Abstract

The LASTWORD hospital information system contains a clinical microbiology database which permits both review of patient reports and retrospective data searches using clinical and/or demographic criteria. The elements supporting this database are dictionary tables of coded phrases, a general purpose query language (Tandem ENFORM), an HL7 interface to a laboratory computer system, and long-term data storage of demographic, microbiology and other clinical data in a relational database.

Overview

Storage strategy for clinical microbiology must take into account the uses, and particularly the users, it must serve. To serve physicians in patient care, it must facilitate retrieval and display or printing of results for individual patients. To serve technologists in a laboratory, it must provide a convenient means to enter the data. For microbiologists, epidemiologists and infectious disease specialists, it must make the search for discrete data elements possible retrospectively, across patients. It is this last need with which this paper is primarily concerned.

We have designed a database for microbiology within a hospital information system (HIS) which makes it possible both to do efficient review of patient data by physicians and to support retrospective microbiology data searches with convenient, standard query language, Tandem ENFORM.

These data searches may incorporate many data elements which are unavailable to a laboratory information system (LIS) alone, but which are natural components of an integrated HIS database.

Microbiology Data Formats

A microbiology result is more analogous to an x-ray report than to a CBC. It is a diagnostic statement, typically consisting of several phrases, some of which are subordinate to, or even meaningless without, the others. Unlike a pure free-text report, however, the phrases in microbiology are repeated from report to report, and as a consequence lend themselves to encoding. Data entry for microbiology in virtually all lab systems consists of some form of text encoding, be it special keys, mark-sense cards, menu selection or direct entry of codes. In a well-constructed coding scheme, there is only occasional need for supplementary entry of free text.

From the technologist's point of view, the entry of encoded data is essential. The inefficiency and inconsistency inherent in entry of free text, given particularly the typical need to update preliminary with interim and final results, renders free text an unacceptable form of data entry for microbiology. Indeed, microbiology technologists are willing to memorize rather elaborate coding schemes in order to obviate free-text data entry.

Clinicians reviewing patient microbiology results wish to see the reports in a coherent, eloquent format. Although many formats are acceptable, and they typically differ from LIS to LIS, all share the characteristics of being variable in length and content, and formatted for legibility and order, to convey the relationships among the phrases. Typically, gram stains, if present, are listed first, and antibiograms, if present, last. The important point here is that the format of a clinical display of microbiology has only an indirect relationship to the data structure underlying it. A typical communication of microbiology data from the LIS to an HIS for purposes of display could be quite adequate for clinical display without conveying any of the internal structure and relationships of the data elements.

In contrast, the production reports for epidemiologists and infectious disease specialists which include data from multiple patients requires an underlying database to permit flexible and unpredictable associations of data. Sample queries against such a database might include:

1. Plot organisms with certain antibiotic sensitivities by hospital location over time.
2. Show all patients with negative sputum cultures admitted to surgery service, irrespective of bed location.
3. Show all patients from zip code area 98115 with positive acid fast culture, with phone number of referring physician.

These requests are characterized by:

1. The variety of data elements used as search or sort criteria.
2. Data from many patients.
3. The search across extended time periods.

0195-4210/91/$5.00

4. The need for storage of data as discrete, even numeric values for analysis.
5. The need for "hierarchical" knowledge built into the data; e.g., some formal criterion by which a result description may be considered to be a "pathogen," or "positive," or "gram positive coccus," or "negative."

Clearly, these are requirements best fulfilled by a general-purpose database. In addition, such a database might be expected to support:

1. Data compression or encoding to minimize the volume of data storage, especially given the long term nature of the searches.
2. High-volume storage.
3. A general purpose query language.

We have developed a database of clinical microbiology results within the LASTWORD hospital information system. Microbiology laboratory data are received from a Sunquest clinical laboratory information system, via an interface using an HL7 applications-message protocol.

The LASTWORD database is implemented on a Tandem computer, using its ENSCRIBE file structure which permits files to be treated as relations. ENFORM, a report generator designed for the ENSCRIBE system, is a nonprocedural, high-level query language which supports relational operations such as logical joins and unions across files.

The LASTWORD HIS comprises a variety of clinical and administrative modules surrounding a core system which supports patient registration, admission-discharge-transfer, user registration and security. Clinical modules include nursing support, order communication, pharmacy, radiology, and laboratory results display.

The lab display module provides a tabular display of hematology and chemistry data, as well as text displays for microbiology and anatomic pathology. Data are received from interfaces to several different laboratory information systems at different customer sites. Data may be stored for any length of time, and are typically not purged at all, creating a long-term database.

Microbiology data have in most sites been transmitted from the laboratory systems as sixty- or eighty-character-width lines of text, and stored and displayed as such. In these settings, no database retrievals of the sort mentioned above are possible. With the newly implemented database storage format, microbiology data are now additionally retrievable as data. The elements of the design which make this an effective database are discussed below, but can be summarized briefly as follows:

1. Dictionary tables: Provide consistency, "hierarchical" knowledge; minimize storage use.
2. Microbiology data records (relations): Provide specific structure for storage of tests, specimen descriptions, organisms and susceptibilities.
3. General LASTWORD database: Provides file records (relations) containing supplementary information to be linked for complex data searches.
4. Text files: Provide storage for text version of microbiology for rapid clinical reporting.
5. ENFORM query language: Provides tool for convenient retrieval and elegant formatting of data.
6. LIS (Sunquest): Provides standard, convenient data entry.
7. Interface: Provides encoded data for compact storage.

Dictionary Tables

Each record in the LASTWORD database contains references to dictionary tables. The principal benefits of this are compact data storage and consistency. For example, rather than store "Pseudomonas Aeruginosa" in the record for each isolate of the organism, the text is stored once in the dictionary table, and individual records contain a four-byte reference to the appropriate dictionary table. This also guarantees the consistent display of the text in any circumstance.

Extended text strings may similarly be constructed from concatenated dictionary references, so that "Left arm biopsy tissue" can be represented by dictionary entries for each of "Left," "arm," "biopsy," and "tissue." Similarly, "Hemophilus influenzae, 3+, beta lactamase negative" can be represented by dictionary records for each of "Hemophilus influenzae," "3+," and "Beta lactamase negative."

Another feature of the dictionary is to represent permanent relationships among dictionary items. For example, the dictionary entry for "Bacteroides fragilis" may itself contain a reference to "Bacteroides" and/or to "Aerobic gram negative bacillus." This represents "hierarchical" knowledge in the dictionaries in a way that will guarantee that isolates of "Bacteroides fragilis" will be retrieved not only when specifically searched for, but also during searches for "Bacteroides" or for "Aerobic gram negative bacillus." Since the dictionary records are themselves relations, these associations are specifically supported by the ENFORM query language.

Schematic diagrams of dictionary tables are included in Figure 1.

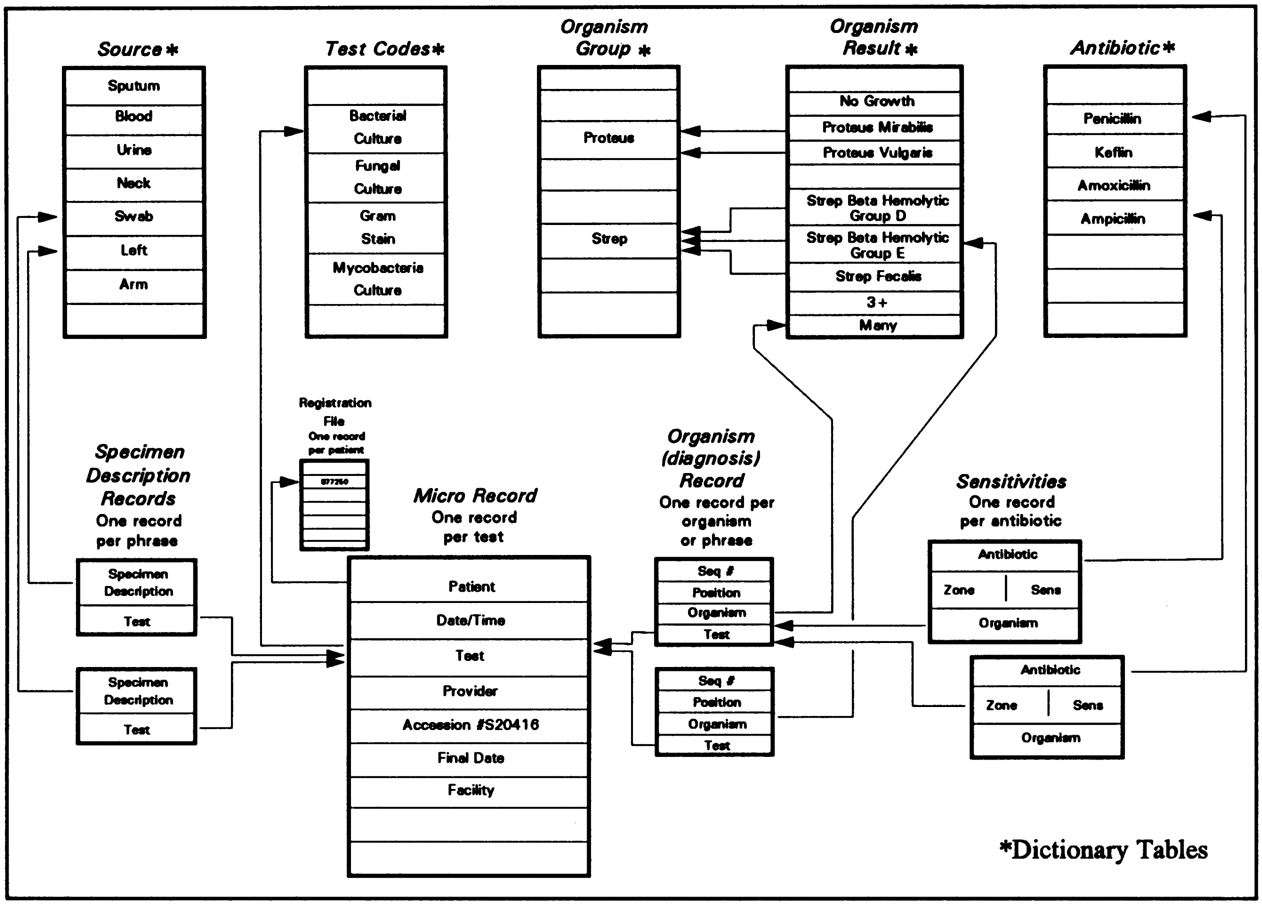


Figure 1. Microbiology Database and Dictionary Tables

ENFORM

ENFORM is a high-level query language which is used for relational searches of the ENSCRIBE database. ENFORM uses simple, English-like statements like LIST and FIND, and supports logical join and union functions among relations. Clauses may be added to provide detailed instructions about field selection, record selection, sorting and formatting. ENFORM generates count and percentage. The entire LASTWORD database is accessible to ENFORM, including the dictionary tables, which makes physical, predefined file linkages unnecessary.

ENFORM is the tool that supports convenient and flexible searches of the microbiology database. The sample queries listed above can be directly translated to ENFORM statements to create the searches and report formats desired.

General LASTWORD Database

The data contained in LASTWORD files provide a vast array of clinical, demographic and administrative information which can be linked via ENFORM to microbiology information. For example, the presence of a referring physician and patient zip code as data elements in the patient's registration-admission record, coupled with the presence of a phone number in each provider record, makes all patients from a postal zone associated with the referring physician's name and phone number a discernable subset of all patients with positive mycobacterial cultures.

It is not always the case that an ENFORM query can successfully link data elements. For example, a report of antibiotic sensitivity against actual administration of the same antibiotic is quite complicated, and might require a COBOL rather than an ENFORM report to link these data correctly. Nevertheless, the presence of all the different data elements in the LASTWORD database makes possible a range of associations not typically available in a laboratory information system.

Microbiology Data Records

The records composing the microbiology database are illustrated in Figure 1. Each record may be seen as a relation; the relations are linked by fields within the records, typically arbitrary record numbers.

In this design, each ordered test is represented by a single MICRO file record linked to the dictionary table of test names as well as to the patient registration record. The other records all have a many-to-one relationship with the MICRO record, reflecting the variability of the data. Thus, where there is a "complex" specimen description, there will be several records representing it, each of which is linked to a component phrase in the specimen table. Similarly, an organism record (also known as "result" or "diagnosis" record) exists for each fragment of each organism description. These records are each associated with a phrase in the dictionary table of organisms; the proper association and ordering of these phrases is determined by the sequence and position number in the record. This supports the possibility that multiple organisms, each with a multi-phrase description, may result from a single test.

Associated with each organism may be several sensitivity records, each of which contains a link to an antibiotic name in a dictionary table, a zone size, and sensitive/resistant flag.

Note also the "hierarchical" relationship between the organism and organism group tables. This provides the basis for retrievals like "List all Proteus isolates" across all Proteus species.

The creation of a record for each phrase preserves the normalized relational nature of the records in a setting where it is impossible to predict before the fact the number of organisms or sensitivities, or the number of elements in any compound text phrase. This normalized format permits a relational tool like ENFORM to work properly.

Laboratory Information System

The laboratory information system supplying the data plays an obvious and crucial role in creation of the database. It is the vehicle for data entry, and therefore the means by which the data are represented as individual data items rather than as simply a text buffer.

The laboratory system in this implementation is a Sunquest system. Sunquest supports several libraries of coded text phrases, which parallel the tables in LASTWORD. This is necessary to insure that the codes relayed to the LASTWORD system unambiguously convey the microbiology results as data. Any code which can be sent from the laboratory system must be known to the LASTWORD tables. This implies that the LASTWORD table must be filled with whatever phrases are present in the corresponding "tables" in the Sunquest system.

Interface

The HL7 applications protocol is a rather flexible vehicle for the conveyance of data. In this LASTWORD-Sunquest interface, we have created a message structure in which the Sunquest system can send concatenated text codes in the order in which they appear in the text display. Specimen descriptions are distinguished from results by a type code, as are antibiotic sensitivities. Where several organisms, each with an antibiogram, are cultured in a single assay, the order of the organism codes and the antibiotic codes unambiguously determines which is associated with which. The significance of the interface is that it supports the transmission of data in an encoded format which conveys sufficient information for:

1. Creation of a database; and
2. Reconstruction of the text format for physician review.

Reports

A sample report that illustrates the use of ENFORM is included in Figure 2. This report is based on the following user request: Correlation of ECG, Laboratory and Microbiology results for all patients with an ECG result of "infarct" or "ischemia"; and a potassium; and at least one positive blood culture with an organism defined as positive. The report will list the ECG, potassium and blood culture results; patient demographics including name, age, sex, clinic number and hospital location, sorted by clinic number.

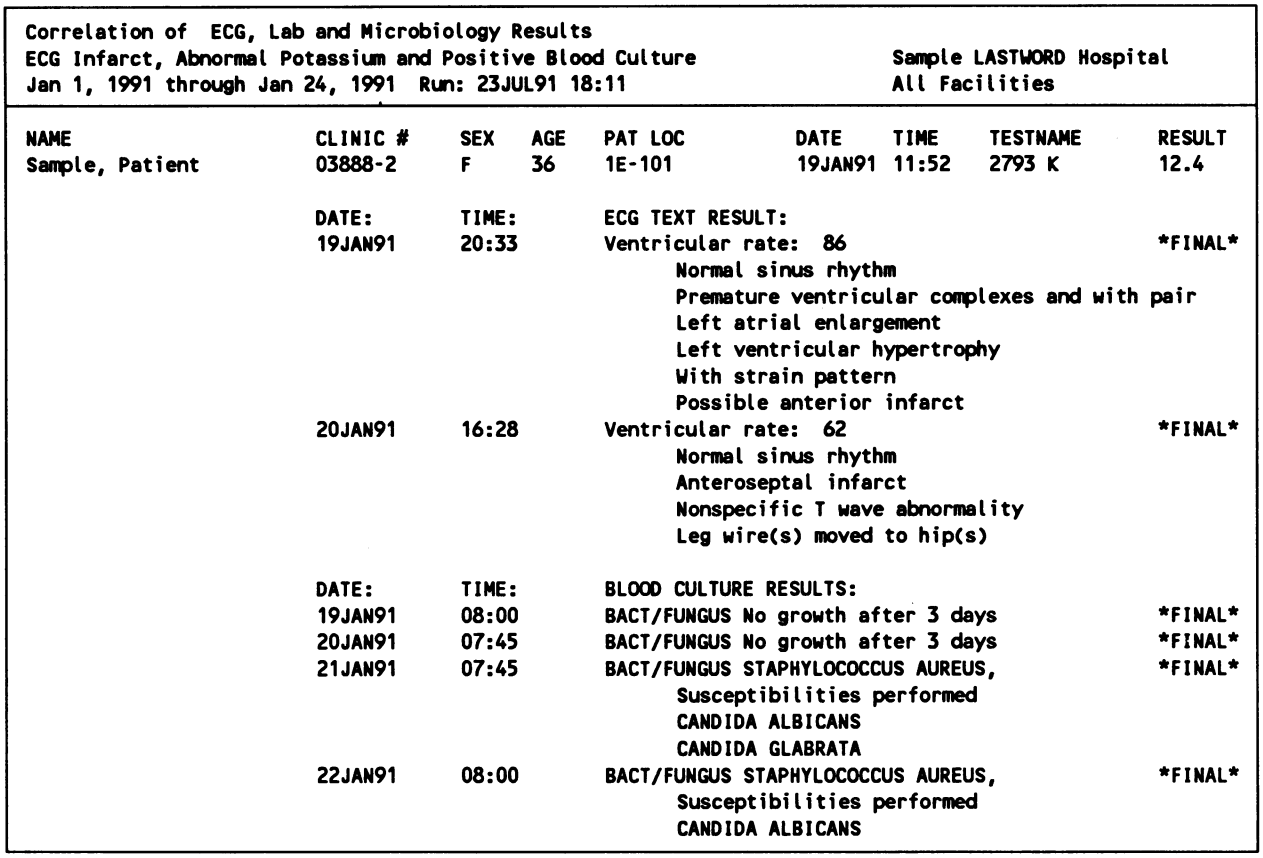
Discussion

A similar clinical database and query system, ClinQuery, has been described by Safran et. al. [1, 2]. ClinQuery, a subset database of the clinical computing system at Beth Israel Hospital in Boston, is created retrospectively from the records of patients discharged from the hospital. It offers the unique feature of a menu screen designed specifically for ease of use.

By contrast, the LASTWORD clinical database is available in its entirety for database queries; it includes all active inpatients and outpatients as well as old patient data. Old and recent data are in identical formats, lending themselves to inclusion in reports covering any time period.

The flexibility of this database derives in part from the medical event vector concept [3]. Medical events in LASTWORD share a common set of fields and keys, which insure efficiency and consistency. Thus, each medical event (an ECG, an x-ray, a chemistry test, a microbiology test) will have a key by patient and date, which optimizes searches for data for an individual patient. These patient and date keys allow clinicians to review a display, in reverse chronological order, of all microbiology (or other selected) test data for a single patient on line within seconds.

Similarly, an alternate standard key in the medical event is the date itself. This key makes it possible to review the data in which a subset date



```
Correlation of  ECG, Lab and Microbiology Results
ECG Infarct, Abnormal Potassium and Positive Blood Culture           Sample LASTWORD Hospital
Jan 1, 1991 through Jan 24, 1991  Run: 23JUL91 18:11                 All Facilities

NAME                 CLINIC #   SEX  AGE  PAT LOC       DATE    TIME   TESTNAME    RESULT
Sample, Patient      03888-2    F    36   1E-101        19JAN91 11:52  2793 K      12.4

                     DATE:      TIME:     ECG TEXT RESULT:
                     19JAN91    20:33     Ventricular rate:  86                    *FINAL*
                                               Normal sinus rhythm
                                               Premature ventricular complexes and with pair
                                               Left atrial enlargement
                                               Left ventricular hypertrophy
                                               With strain pattern
                                               Possible anterior infarct
                     20JAN91    16:28     Ventricular rate:  62                    *FINAL*
                                               Normal sinus rhythm
                                               Anteroseptal infarct
                                               Nonspecific T wave abnormality
                                               Leg wire(s) moved to hip(s)

                     DATE:      TIME:     BLOOD CULTURE RESULTS:
                     19JAN91    08:00     BACT/FUNGUS No growth after 3 days       *FINAL*
                     20JAN91    07:45     BACT/FUNGUS No growth after 3 days       *FINAL*
                     21JAN91    07:45     BACT/FUNGUS STAPHYLOCOCCUS AUREUS,       *FINAL*
                                               Susceptibilities performed
                                               CANDIDA ALBICANS
                                               CANDIDA GLABRATA
                     22JAN91    08:00     BACT/FUNGUS STAPHYLOCOCCUS AUREUS,       *FINAL*
                                               Susceptibilities performed
                                               CANDIDA ALBICANS
```

Figure 2. Sample ENFORM Report

range may be selectively searched, obviating an "expensive" review of all records to gather just those pertinent to a particular query. If several different types of events, e.g. diagnosis, x-ray and microbiology events, must be searched to generate an ENFORM report, the existence of the consistent date key restricts the search of each such event file, making it possible to create a typical retrospective report in an hour or so, obviously depending on the extent and types of data included.

Value

Although typical clinical reports of microbiology results are textual, creation of a microbiology database may be possible without the passing of text. This is so because the effort to create a convenient data entry mechanism for microbiology technologists will almost of necessity create a coding scheme, i.e., a mechanism whereby a few characters or keys are made to represent an eloquent and longer text string. The logical extension of the "tables" which such encoding requires is a condensed data storage format which can be used as a database. Typically, laboratory information systems lack two critical pieces to make such data more searchable and clinically useful:

1. A general-purpose query language; and
2. A large body of related clinical information in the same data setting.

In order to make such data retrieval possible, we have used HL7 to create an interface to send microbiology results to our HIS, but more importantly, have created both an HL7 message substructure and a database which support the storage retrieval of microbiology results as data, within a large, integrated clinical system. The result is a database in which it is possible to retrieve combinations of microbiology data with very detailed demographic or patient-location data, including, in particular, historical patient-location information. This is essential for the reports epidemiologists need--reports of associations of infectious agents with similar susceptibility profiles at a particular time and place.

References

[1] Safran, C.; Porter, D.; Rury, C.; Herrmann, F.; Lightfoot, J.; Underhill, L.; Bleich, H.; Slack, W. "ClinQuery: Searching a Large Clinical Database," *MD Computing*, 1990; 7(3):144-153.

[2] Safran, C.; Porter, D.; Rury, C.; Herrmann, F.; Lightfoot, J.; Underhill, L.; Bleich, H.; Slack, W. "ClinQuery: A System for Online Searching of Data in a Teaching Hospital," *Annals of Internal Medicine*, 1989; 111(9):751-756.

[3] Gleser, M. "The Medical Event Vector," *Methods of Information in Medicine*, 1979; 18(3):127-131.