Supplement to
TECHNICAL REPORT NO. 11

CLIS
CRIMINALISTICS LABORATORY INFORMATION SYSTEM

VOLUME 2
SYSTEMS DESIGN FOR A CONCEPTUAL MODEL

30454

SEARCH GROUP Inc.
a nonprofit corporation dedicated to the justice system . . .
SGI 1620 35th AVENUE / SUITE 200 / SACRAMENTO, CALIFORNIA 95822 / (916) 392-2550

Supplement to
TECHNICAL REPORT NO. 11
MAY 1975

CLIS
CRIMINALISTICS LABORATORY INFORMATION SYSTEM

VOLUME 2
SYSTEMS DESIGN FOR A CONCEPTUAL MODEL

Final report on work performed under Law Enforcement Assistance Administration Grant No. [illegible] SS-99-3309, awarded to the California Crime Technological Research Foundation for Project SEARCH. In 1974, Project SEARCH was incorporated as SEARCH Group, Inc., and the project was continued to completion under its guidance.

Submitted by SEARCH Group, Inc.
1620 35th Avenue, Suite 200, Sacramento, CA 958[illegible]
(916) 392-2550

TABLE OF CONTENTS

CLIS PROJECT COMMITTEE

Chairman
Thomas M. Muller
Baltimore Police Department
Baltimore, Maryland

Vice-Chairman
Fred Wynbrandt
California Department of Justice
Sacramento, California

Clyde E. Bevis
Wichita State University
Wichita, Kansas

Edward G. Bigler
Florida Department of
Criminal Law Enforcement
Tallahassee, Florida

Theodore R. Elzerman
Illinois Bureau of Identification
Joliet, Illinois

Richard H. Fox
Regional Crime Laboratory
Independence, Missouri

John W. Gunn, Jr.
Drug Enforcement Administration
U.S. Department of Justice
Washington, D.C.

Arthur F. Kivela
New Mexico State Police
Crime Laboratory
Santa Fe, New Mexico

John P. Klosterman
Eastern Ohio Forensic Laboratory
Youngstown, Ohio

Frank Madrazo
New York Division of
Criminal Justice Services
Albany, New York

Cornelius G. McWright
Federal Bureau of Investigation
Washington, D.C.

George R. Nakamura
Office of the Medical Examiner
Los Angeles, California

O. S. Neely
West Virginia State Police
South Charleston, West Virginia

Joseph D. Nicol
University of Illinois
Circle Campus
Chicago, Illinois

Carl B. Stokes
South Carolina Law
Enforcement Division
Columbia, South Carolina

S. David Kutob
Department of Public Safety
Phoenix, Arizona

Administrative support
California Crime Technological Research Foundation
Joanne Hoffmann Tony Stratton David Anderson

Technical Support
PRC Public Management Services, Inc.
William F. Cumming, Project Manager

Law Enforcement Assistance Administration Representatives

William W. Herrmann
Region IX

Joseph L. Peterson
National Institute

William H. Bailey
Systems Development
Division

PREFACE

The model for a criminalistics laboratory information system described in this report was developed by Project SEARCH (now SEARCH Group, Inc.) as part of its ongoing program of facilitating the application of advanced technology to the administration of criminal justice. The project, funded by the Law Enforcement Assistance Administration, addressed itself to three topics:

- definition of the information needs of criminalistics laboratories throughout the nation
- conceptual design of an automated information storage and retrieval system
- creation of a plan for implementing the system

Future efforts will include the detailed design, implementation, and evaluation of a pilot system and, eventually, full system implementation.

SEARCH Group, Inc. (Project SEARCH) is a private, non-profit justice research organization owned and operated by the fifty states, the District of Columbia, Puerto Rico, and the Virgin Islands, which fosters research of greater magnitude than can normally be undertaken by individual states.

Thomas M. Muller served as CLIS Project Chairman and Fred Wynbrandt as Vice-Chairman. Subcommittee Chairmen were Edward Bigler, Richard Fox, and Frank Madrazo. Administrative staff services for the project were provided by the California Crime Technological Research Foundation; technical support was provided under contract by PRC Public Management Services, Inc.

Four volumes providing detailed information about specific aspects of the project are being published.

- Volume 1 — *Identification of User Needs*
- Volume 2 — *Systems Design For a Conceptual Model*
- Volume 3 — *System and Organizational Impact*
- Volume 4 — *Implementation Plan*

Copies of these volumes are available from SEARCH Group, Inc.

GLOSSARY

DATA PROCESSING TERMS

baud Number of bits transmitted per second. (It usually requires eight bits to transmit one character.)

byte That portion of a computer word capable of containing a single character. Used synonymously with "character" in this report.

CPU Central processing unit. A computer without its data storage and other peripherals.

CRT Cathode ray tube.

hardwired Accomplished by electronics rather than programming.

I/O Input and output.

modem Device which connects a terminal or computer to a telephone line.

peripheral Device with which a computer stores data or communicates, with the outside world, such as a disk drive, card reader, or teletypewriter.

INSTITUTIONAL ABBREVIATIONS

ASTM American Society for Testing and Materials (Philadelphia, Pennsylvania).

FCIC Florida Crime Information Center (Tallahassee, Florida).

HOCRE Home Office Central Research Establishment (Aldermaston, United Kingdom).

WRAIR Walter Reed Army Institute for Research (Washington, D.C.)

GEOGRAPHIC ABBREVIATIONS

NE = NEW ENGLAND
- Connecticut
- Maine
- Massachusetts
- New Hampshire
- Rhode Island
- Vermont

MA = MIDDLE ATLANTIC
- New Jersey
- New York
- Pennsylvania

ENC = EAST NORTH CENTRAL
- Illinois
- Indiana
- Michigan
- Ohio
- Wisconsin

WNC = WEST NORTH CENTRAL
- Iowa
- Kansas
- Minnesota
- Missouri
- Nebraska
- North Dakota
- South Dakota

SA = SOUTH ATLANTIC
- Delaware
- District of Columbia
- Florida
- Georgia
- Maryland
- North Carolina
- South Carolina
- Virginia
- West Virginia

ESC = EAST SOUTH CENTRAL
- Alabama
- Kentucky
- Mississippi
- Tennessee

WSC = WEST SOUTH CENTRAL
- Arkansas
- Louisiana
- Oklahoma
- Texas

M = MOUNTAIN
- Arizona
- Colorado
- Idaho
- Montana
- Nevada
- New Mexico
- Utah
- Wyoming

P = PACIFIC
- Alaska
- California
- Hawaii
- Oregon
- Washington

PR = PUERTO RICO

CHAPTER 1. SCOPE OF VOLUME 2

INTRODUCTION

The conceptual design of a criminalistic laboratory information system (CLIS) is presented in this Volume and directly addresses those user needs presented in Volume 1. The principal emphasis of the information presented in subsequent volumes amplifies this conceptual design.

Chapter 1 includes a definition of key terms which will be used throughout the remainder of the report.

Chapter 2 profiles the potential CLIS users. For purposes of this analysis, "potential users" includes all laboratories that have returned a laboratory information form. Included is a list of respondents by geographic area (The FBI's Uniform Crime Reporting geographic area designators have been used) and an expansion of the statistical analysis of the size of responding laboratories presented originally in Volume 1.

Chapter 3 discusses each application area carried forward from Volume 1 in relation to general design considerations including:

- General operating functions
- Data base requirements
- Processing functions
- Communications
- Maintenance
- Initial implementation.

Chapter 4 summarizes the system resources needed to support current application areas and anticipated future requirements of CLIS as conceptualized.

Chapter 5 presents the design alternatives for CLIS communications, data storage, processing, and maintenance requirements.

DEFINITION OF TERMS

To lay the foundation for discussion of the conceptual design of CLIS, it is appropriate to define several key terms relative to the development of a systems design.

System Design. A general sequence of activities utilized in the design of information systems has evolved over the years. First, an idea or concept is developed suggesting that a specific information system might be appropriate to support a particular activity or solve a particular problem. The idea is either dropped, tabled or developed to the extent that a decision to pursue its possibilities is reached.

Requirements Analysis. The next step is a formal undertaking to determine what the system must do to achieve the desired results. What information will it provide? What functions will it perform? The requirements analysis answers these general questions.

Conceptual Design. A clearer picture begins to develop of how the system must perform in terms of the end products that address the requirements. But the system components, files and processing functions must all interrelate in the way that the end products are achieved in the most efficient and effective manner.

The next step is to conceptualize the entire system. Given the requirements and the available knowledge of the application area, how is the general configuration of the system going to look? The conceptual design documents the idea.

Feasibility Study. It is now known what the system has to do and how it might be structured. The question now becomes, based on what is known, is it technically practical and economically feasible to implement the system? The feasibility study addresses this final hurdle.

Detailed Design (including system specifications). If the feasibility decision is positive, all of the details that will lead to the actual implementation of the system must be developed. An information system is not usually implemented from a conceptual design, because of its general nature. Necessary implementation details are normally provided by the comprehensive system specifications included in a detailed design document.

This is admittedly a gross oversimplification of what is very often a very time-consuming, sophisticated, and comprehensive sequence of events. In some form or another, however, these basic activities are generally completed in the sequence described during the course of designing an information system, and although they usually are identified with the design of automated systems, the same general principles can be effectively applied to manual information systems or components. In summary, the steps are:

- The idea (preliminary conceptual design)
- Requirements analysis
- Conceptual design
- Feasibility study
- Detailed design.

DEVELOPMENT OF A CONCEPTUAL DESIGN

The scope of the CLIS Phase I effort includes the conceptual design of a computerized information system designed to meet the needs of the user laboratories. Some of the products of this effort are to include data element lists and coding structures, record and file layouts, and indexing and access criteria for each of the requirement areas defined in Volume 1. To some extent this level of detail is not consistent with a normal conceptual design procedure, and unless the nature of the specific requirements is kept in the proper perspective, some of these details may be considered premature at this point. Generally, these considerations are elements of a detailed systems specifications document. This volume, in conjunction with Volume 1, presents two of the basic steps of systems design — requirements analysis and conceptual design — and includes parts of another — system specifications.

Even though the scope of all project activity relates to a "conceptual design", some of these detail design elements are being included. However, a complete, detailed design of CLIS, to include all system specifications, is not a part of the Phase I effort. The conceptual design that follows in the subsequent chapters of this report, therefore, is something more than a true conceptual design in that additional considerations are addressed.

PRELIMINARY DESIGN CONSIDERATIONS

In developing a preliminary design for each of the CLIS application areas, a number of considerations will be addressed. These considerations are generally defined in the following paragraphs. In Chapter 3 they are detailed individually for each application area.

General Operating Functions. This will discuss what CLIS could do, from the users' viewpoint, for a given application. Inputs and outputs will be considered. Examples of inputs might be the punching of keys (either on a terminal keyboard or in the production of punched cards) or the acquisition of online signals from an analytical instrument. Outputs might be teletyped terminal responses, typed or printed reports, CRT displays, etc.

Data Base Requirements. Possible characteristics of a file to support this application include the following:

File Content. What will be in the files? What data elements are needed for the application? Examples of data elements are chemical names in a file of spectra or one for sources of reagents; street addresses in a file of sources, expertise or reagents; land or groove width in a rifling file; authors' names in a literature abstract or bibliographic file, etc.

File Structure. Two major types of computer files are "random" and "sequential." In a random file the computer can directly address the item sought; in a sequential file it has to search from the beginning of the file until it encounters the entry it is looking for. Magnetic tape files are generally sequential. Random files are usually on magnetic discs or drums and are more expensive; however, retrieval of data is significantly more rapid than that of sequential files.

Record Layout and Element Coding Structures. Each file will be divided into one or more kinds of records. A record will contain all the data elements for a given kind of entry. For example, a file of infrared spectra might have a record for each spectrum it contains. The data elements in each record might be number codes representing physical state and functional groups, and decimal numbers representing wavelengths at which the compound's spectrum has peaks (bands).

Access Criteria and Indexing Methodology. Since the purpose of creating these files is to retrieve information from them, they should be designed with the idea in mind of making this as easy as possible. Indexing methodology involves the gross arrangement (alphabetical? — chronological?) and hierarchical structuring (subfiles within files, such as drugs and pyrolyzates as categories within a file of infrared spectra). Access criteria concern such things as indexing by chemical compound name or by spectral pattern, inclusion of a cross-reference table for synonyms, and use of sequential file, which must be searched in its entirety, or a random-access file, which directly references items of interest via directories or dictionaries.

Estimated File Size and Growth Rate. File size is a measure of the total storage space required for all the entries or records that should be in the file. A growth rate must also be predicted (how much space will be needed for the information to be added each year or other unit of time). These parameters will impact upon the hardware chosen, if any, and indeed upon the whole implementation approach.

Processing Functions. Given the data base requirements and file structures, how are programs designed to manipulate this information and get it into and out of the files?

Data Encoding and Allocation Schemes. In either a manual or a computerized system, some sort of translation process is likely to be required to get data from diverse sources into a standardized format for filing. An example might be the conversion of a compound's chemical characteristics into numeric codes that can be stored more compactly. On a computerized system a further translation might be performed automatically by program, such as from the numeric codes into binary patterns. Also on a computerized system, programs may reallocate information from the way it is seen "outside." For instance, a long description might be broken into "standard" and "overflow" records stored in different places.

File Search and Match Techniques. For each application, means must be developed to allow the user to find what he wants in the file. He may be looking for an abstract of a specific article; he may be looking for all abstracts pertaining to a specific subject; or he wants to know which compounds in the file most closely approximate the results of an analysis he has performed in the laboratory. In each case his entry must be matched against those on file, whether it be for similarity of spectral peak patterns, closeness of spelling of chemical names, or whatever.

Applicability of a Functional Language. In some cases an operator can be "led by the hand" through a question-and-answer format so that he does not need any special knowledge to work with a computerized information system. In other cases, however, it is more appropriate that the operator learn a special set of commands tailored to his specific application and understood both by him and the computer. This latter is known as a functional language.

Concurrency of Operation. It is likely that access to the system will be desired by more than one laboratory at a time. These accesses may well involve different applications. Within a larger laboratory many examiners may want simultaneous access to the system. Ways of making this possible will be addressed.

Communications. This subject must be considered in order to determine the best means of transmitting information back and forth between the user laboratories and a computer facility or facilities.

Estimated Usage Rates. With the help of information returned on the CLIS questionnaires and other resources, estimates will be made on frequency and rate of information transmission required for each application.

Necessary Response Times. When the user asks the system a question, when does he have to have his answer? This can vary from immediately on a computer terminal to long enough to send the answer by mail.

Required User Interaction. Some information systems are implemented in conversational mode. That is, once the operator has the system's attention, he and the computer take turns asking and answering questions until they have zeroed in on the information he wants. The New York State Division of Criminal Justice Services decided that such an approach was necessary for their computerized infrared data file. However, conversational systems are more difficult and expensive to implement since they require more complex software, frequently cause longer connect times for data trannmissions, and necessitate that a live terminal be present at the user's site.

Simpler from the implementation point of view is the "remote batch terminal" approach in which the user formulates more complete questions, transmits them to the computer via his terminal, and waits for the computer to return the requested information. The whole cycle must be repeated if more information is needed. The least expensive approach, but not necessarily the most effective, would be to conduct the interaction by mail.

Anticipated Volume of Transmitted Data. For each application an estimate will be made on the length and frequency of typical messages.

Maintainability and Currency of Data. The amount of data contained in the data bases will not be static, but will continue to grow as new information becomes available. A means must be provided to allow this to take place in an orderly fashion. Conversely, a means must be provided for the removal of incorrect or obsolete data. Mistakes may get into the system, or it may be desirable to replace an existing entry with better information. Should it become necessary to limit the size of any file, the replacement of little-used entries with more needed ones might be required. Most laboratories have developed their own in-house files of spectral data. Some laboratories might be willing to share these files with other interested laboratories. Tools for maintaining accuracy and currency of all data must be part of the system design.

Sources of Updates for Maintenance. Insofar as possible at this point, these sources will be identified for each application.

Summary of Maintenance Procedures. For each application a list of procedures will be developed to ensure currency of data within the files.

Initial Implementation. This identifies some of the primary tasks that need to be accomplished in making the system a reality. In some cases implementation may by stepwise, with experience gained by both users and designers in one step being integrated into the following step.

Design and Development of Processing Functions. General programming tasks will be identified. Implementation priorities will be assigned for complex applications.

Identification of Potential Sources of Data. In most cases a considerable volume of data already exists in each application area. It is in fact this voluminous data which provides a major justification for CLIS. The sources that are known at present will be enumerated.

Encoding and Entry of Data. The development of any information system requires a step known as "putting up the data base." This requires the manual encoding or the computerized conversion (of existing machine-readable files) to the format selected for the CLIS files. This process will be detailed where known.

CHAPTER 2. LABORATORY POPULATION

List of Respondents. A list of all laboratories which had returned questionnaires (CLIS Information Forms) is appended to this Volume.

The sources of two questionnaires could not be determined. These questionnaires, logged in as numbers 20 and 67, are not included.

The following respondents, for the reasons given, did not deem it appropriate to fill out CLIS Information Forms:

- National Bureau of Standards Law Enforcement Studies Laboratory, Washington, D.C., and the University of Illinois Department of Criminal Justice, Chicago, Illinois. These two are not working criminalistic laboratories.
- City of Niagra Falls (New York) Department of Police. They do not have their own laboratory; instead, all of their work is sent to the FBI Laboratory.
- Puerto Rico Police Crime Laboratory, Ponce, Puerto Rico. They are a branch of the General Crime Laboratory in Hato Rey, Puerto Rico, and felt it would be more helpful for the information to come from the General Laboratory. That laboratory has responded and is logged in as number 161.

Duplicate returns were received from six laboratories.

Respondents by Geographic Area. The map in Figure 1 shows the geographic distribution of institutions to which information forms were mailed and indicates those which have responded. Table 1 is a tabulation of the same information by geographic divisions; the divisions are the same as those used in the FBI Uniform Crime Reports. ("Questionnaires mailed" is not a strictly accurate term. Some questionnaires are known to have been mailed to nonexistent laboratories or addresses; these are not included. Some responses are from sites other than those to which the questionnaires were originally mailed; these *have* been included.)

Respondents by Size of Laboratory. Table 2 categorizes the laboratories according to their response to question 1-h: "Number of full-time employees." Where question 1-h was not answered, the number used was the sum of the responses to the four subcategories under question 1-i: "Number of technically trained employees who perform analyses and give testimony." If this number was not available, no size data was used for that laboratory.

These data deserve special attention, since the size of the laboratories will be an important determination of the type and amount of services appropriate to each user in terms of that which is desirable versus that which is feasible.

Table 1

Respondents To CLIS Questionnaires By Geographic Area

(As of 3 June 1974)

	Questionnaires Mailed	Responses Received
NEW ENGLAND		
Connecticut	4	2
Maine	1	1
Massachusetts	2	0
New Hampshire	1	0
Rhode Island	3	0
Vermont	1	0
Total New England	12	3
MIDDLE ATLANTIC		
New Jersey	5	3
New York	12	10

(Continued on next page)

(Table 1, continued)

	Questionnaires Mailed	Responses Received
Pennsylvania	8	6
Total Middle Atlantic	25	19
EAST NORTH CENTRAL		
Illinois	12	11
Indiana	6	2
Michigan	6	6
Ohio	15	9
Wisconsin	4	2
Total East North Central	43	30
WEST NORTH CENTRAL		
Iowa	1	1
Kansas	3	3
Minnesota	2	1
Missouri	6	6
Nebraska	2	2
North Dakota	1	1
South Dakota	1	1
Total West North Central	16	15
SOUTH ATLANTIC		
Delaware	2	1
District of Columbia	12	8
Florida	12	8
Georgia	3	1
Maryland	4	2
North Carolina	3	1
South Carolina	2	0
Virginia	6	1
West Virginia	1	1
Total South Atlantic	45	23
EAST SOUTH CENTRAL		
Alabama	5	2
Kentucky	1	1
Mississippi	2	0
Tennessee	1	1
Total East South Central	9	4
WEST SOUTH CENTRAL		
Arkansas	2	0
Louisiana	6	1
Oklahoma	3	3
Texas	13	12
Total West South Central	24	16
MOUNTAIN		
Arizona	4	4
Colorado	3	2
Idaho	1	1
Montana	1	1
Nevada	4	1
New Mexico	2	2
Utah	3	3
Wyoming	1	1
Total Mountain	19	15
PACIFIC		
Alaska	1	1
California	35	30
Hawaii	1	1
Oregon	8	4
Washington	5	4
Total Pacific	50	40
Puerto Rico	4	3
Virgin Islands	1	0
Total	5	3
Unknown	0	2
GRAND TOTAL	248	170

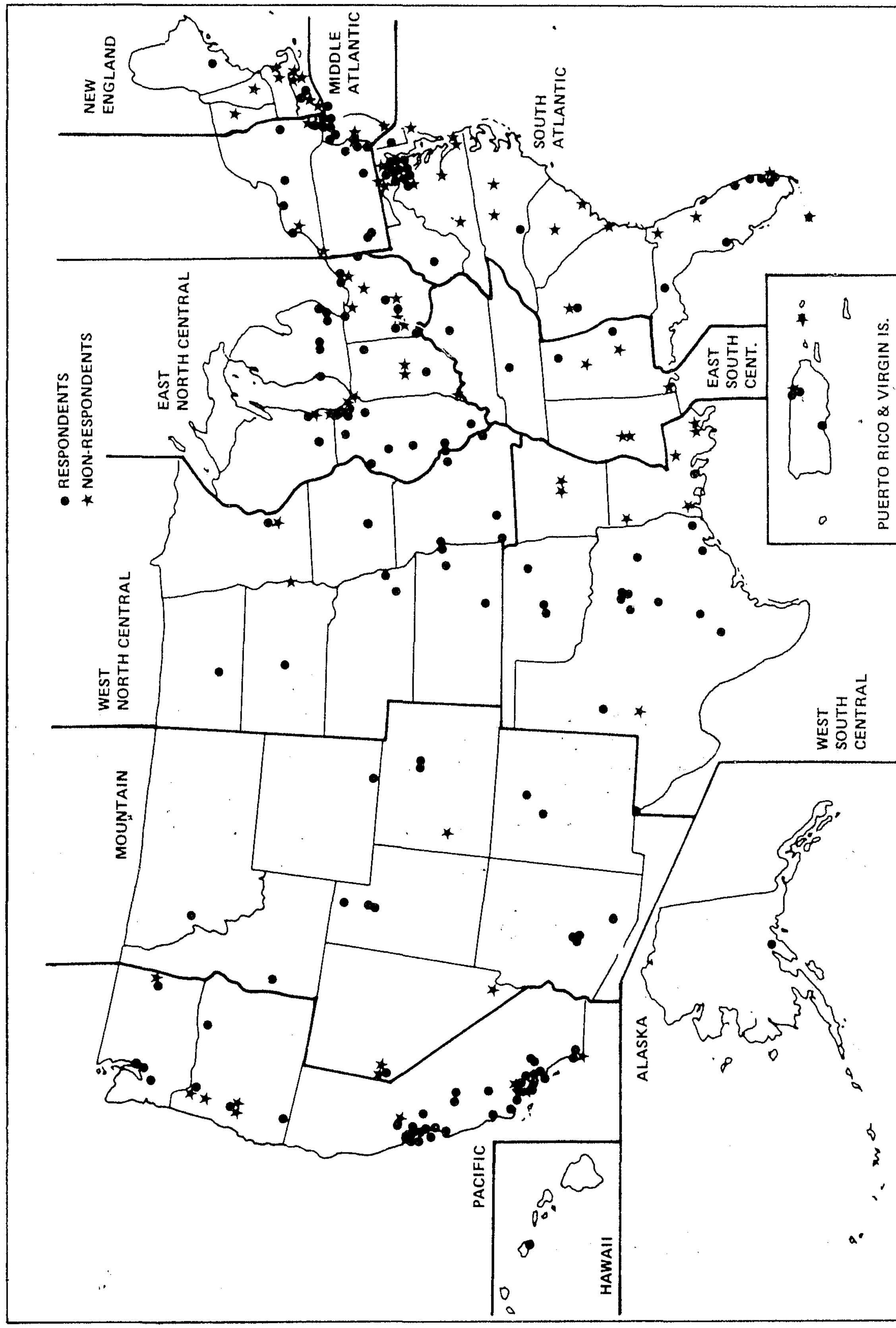


Figure 1
Geographic Distribution of Laboratories Included in CLIS Survey

Table 2

Size of Responding Laboratories

Number of Employees	Number of Laboratories In This Range
0 - 4	48
5 - 9	33
0 - 9	81
10 - 14	20
15 - 19	14
10 - 19	34
20 - 24	11
25 - 29	5
20 - 29	16
30 - 39	9
40 - 49	9
50 - 59	1
60 - 69	0
70 - 79	3
80 - 84	3
85 or more	2
1 - 100	156
101 - 200	0
201 - 300	0
301 - 400	1
401 - 500	1
	158

CHAPTER 3. GENERAL DESIGN CONSIDERATIONS

GENERAL DESIGN FOR ANALYTICAL/ IDENTIFICATION SUPPORT

Background

This application involves processing the outputs of laboratory analytical instruments. Responses to Chart 5 of the CLIS survey forms are analyzed in Table 3. It is shown that the most common types of instruments in criminalistics laboratories are:

- Infrared spectrophotometers
- Ultraviolet spectrophotometers
- Visible Spectrophotometer
- Gas chromatographs
- Emission spectrographs
- U-V fluorescence devices
- Atomic absorption spectrometers
- Electrophoresis equipment

Moreover, there is widespread prediction that the combined *gas chromatograph/mass spectrograph* instrument will become increasingly important in criminalistics analysis in the near future.

Computerization can support an analytical instrument in two ways: (1) by assisting with the data reduction required to obtain a usable result from a complex analytical technique and, once the result is obtained, (2) by assisting with the search required for identification.

In general, these instruments produce an output that represents intensity versus either extraction time or wavelength. In many cases the instrument itself produces a strip chart with intensity as the ordinate and wavelength or time as the abcissa. Some techniques, such as electrophoresis, may yield a blob of color whose intensity and position on a substrate is determined visually. Electrophoresis outputs can, however, be converted to strip charts, using a scanning densitometer.

Analytical instruments vary in the degree to which they carry out the data reduction function. Some produce an output to which considerable manual manipulation must be applied before the results are meaningful. Others, often with the help of built-in minicomputers, do the whole job on their own. Any device that produces an x-y plot has available internally an electronic signal which could be connected directly to a computer for automatic analysis or even identification by search of a reference file. This type of online lookup represents a possible CLIS function. In view of the multiplicity of instrument types and manufacturers, and the fact that special software would have to be tailored for each case, the implementation time and expense might be prohibitive. However, it is possible that certain instruments may be sufficiently common to make their interfacing worthwhile.

Whatever the instrument output, in order to determine the sample's identity, the analyst usually must compare it with outputs from previous analyses. The automation of these searches is another potential CLIS function.

In some cases, reference files are available (such as Sadtler's for I-R and U-V spectra) which may be used for these comparisons. The analyst, however, typically has to deal with samples less pure than the compounds used for the commercial reference files, and tends to lean more heavily on experience and the use of reference files compiled in his own laboratory. Unfortunately, the creation of such files for each individual laboratory represents a considerable duplication of effort which may not be entirely necessary. The sharing of in-house files among laboratories is seen as an important potential benefit of CLIS.

GENERAL OPERATION FUNCTIONS — INPUT/OUTPUT

The general operating functions, then, for the analytical/identification support application might be as follows: the acquisition of online signals from an analytical instrument and their reduction to a form useful to the criminalist, and the acceptance of an analytical result and comparison with a file of previous results in an attempt to effect its identification.

Related functions might be the following: updating of a private library of analytical results for

Table 3

Instrumentation of Responding Laboratories (1955)
(Does Not Include The FBI Laboratory)

	Total Number of Instruments	Average Number per Laboratory Reporting Instruments	Number of Laboratories Reporting This Instrument	Average Number per Laboratory Reporting This Instrument
X-Ray Diffractometers	34	0.22	29	1.0
Neutron Activation Analyzers	10	0.06	7	1.0
Gas Chromatograph/Mass Spectrographs	30	0.19	29	1.0
Differential Thermal Analyzers	16	0.10	13	1.0
Nuclear Magnetic Resonance	7	0.05	7	1.0
Scanning Electron Microscopes	7	0.05	5	1.0
Emission Spectrographs	70	0.45	65	1.0
X-Ray Fluorescence	12	0.08	9	1.0
Ultraviolet Fluorescence	70	0.45	63	1.0
Atomic Absorption Spectrometers	39	0.25	32	1.12
Mass Spectrographs (inorganic)	3	0.02	2	1.0
Electron Probes (ESCA)	3	0.02	3	1.0
Electrophoresis	94	0.61	73	1.13
Infrared Spectrophotometers	181	0.20	134	1.32
Gas Chromatographs	317	2.00	134	2.28
Ultraviolet, Visible, & Near-Infrared Spectrophotometers	245	1.55	136	1.67
Polarimeters	6	0.04	6	1.0
Flame Photometers	8	0.05	10	1.0
Liquid Chromatographs	10	0.06	10	1.0
Raman Spectrographs	0	0.01	2	1.0
Energy-Dispersive X-Ray	2	0.01	1	1.0

Total Number of Laboratories Reporting Instrumentation = 155

Average Size of Laboratories Reporting Instrumentation = 19.9 Persons

known substances; the contribution of such a private library to a larger library shared among laboratories; and hence, the availability of the library for each individual laboratory's identification purposes.

Figure 2 shows how the processing functions interrelate for this application.

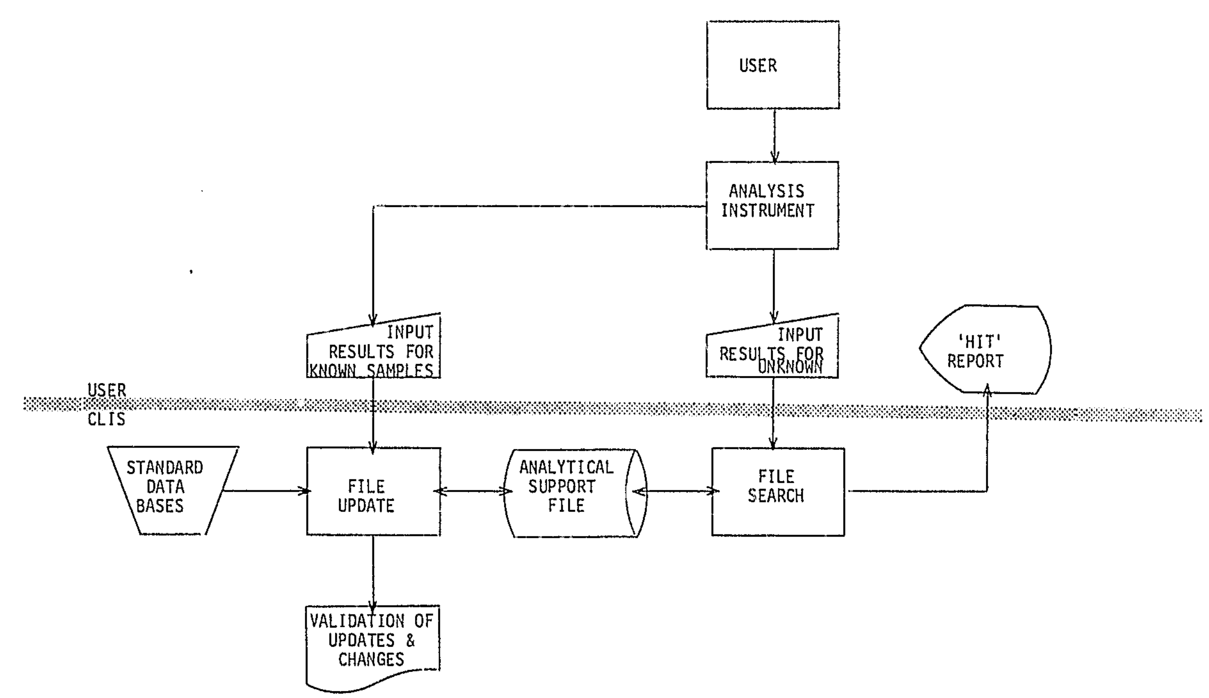
Inputs and Outputs. Since the hardware approach and configuration has not been specified at this point, it is inappropriate to do more than provide examples of how CLIS might support the analyst. The following descriptions are in order of decreasing degree of automation:

Online. Electrophoresis slides are scanned with a Beckman Micro-zone Densitometer. The densitometer is connected either to a centrally located host computer or to a minicomputer at the user's site. The electrophoresis scan is reproduced as a waveform on a cathode-ray tube. By manipulating cursors, the analyst delineates the peaks on the waveform. The computer finds the areas under the peaks so defined and types a report giving a digital value for each of the electrophoretically separated components. The computer then searches a file of previous electrophoresis determinations and reports the names of those compounds whose patterns are most similar to the result currently being analyzed.

The inputs in this case would be analog signals from the densitometer and the analyst's cursor manipulation. Outputs would be the displayed waveform and the typed report of peak values and possible identifications.

Cathode-Ray Tube (CRT) Terminal. A laboratory has obtained a mass spectrum which it wishes to identify. The three most prominent peaks have masses of 181, 221, and 323, with intensities of 40 percent, 65 percent and 100 percent, respectively. An operator sits down at the terminal and from a list of options displayed on the CRT chooses "PEAK SEARCH." The terminal then asks for mass and intensity limits; so the operator types "323," "75 percent," and "100 percent" (the maximum). The display informs him there are 307 spectra in the file which meet this criteria. At this point he has a choice of having the names, formulae, and weights of all 307 compounds displayed for him or of further narrowing his specifications. He chooses the latter and describes another peak as "181," "20 percent," "60 percent." The CRT now tells him he has reduced the list to 15 spectra. He enters "221," "50

Figure 2 Information Flow — Analytical and Identification Support



percent," and "80 percent" and discovers that there are now 12 possibilities. He elects to see the list and an identification number, name, molecular formula; and molecular weight are displayed for each. He then may choose to display any or all (individually) of the actual spectra for visual comparison with his unknown. In this manner he decides his unknown is lysergic acid diethylamide.

In this example, all inputs are keystrokes on the terminal keyboard and all outputs are CRT displays.

Teletype Terminal. A laboratory worker has obtained an unknown infrared spectrum and wishes to attempt identification. He enters a code meaning "search file" on his Teletype and begins an iterative interactive procedure similar to that described above for a mass spectra search using a CRT terminal. However, in this case the information provided by the terminal is more succinct, since it takes much longer to produce on a Teletype. Information tends to be encoded rather than written out in full. Also it is impossible to produce an actual picture of the spectrum on the Teletype.

All the inputs are entered on the Teletype keyboard and all the outputs are typed.

Remote Batch Terminal. A laboratory wishes to add its in-house file of U-V spectra to the central CLIS file. For this purpose CLIS has provided it with a FORTRAN program on punched cards. The laboratory has had its spectra keypunched after encoding them in the agreed-upon file format. The program and data cards are input through the terminal card reader, and a verification listing is produced on the terminal printer showing what data have been entered in the file. Other programs can be used to update the file or to do file searches for the identification of unknowns. Note, however, that in batch processing mode it is not practical to zero in upon identification iteratively as was described above, or interactive terminal searches. Rather, the analyst must be as specific as possible with his initial search criteria, and he may still receive a long list of possibilities to wade through. Here, input is via punched card and output via printed listing.

No Terminal. It might be impractical to provide a direct hookup to some or all laboratories. Unknowns to be searched against the main file would have to be encoded on forms and sent by messenger or mailed to the central facility. Results would be returned in the same way. Although this method would be clumsy for the identification of unknowns, it might be the method of choice for the addition of large quantities of data to the central file. Thus it might be used to add a laboratory's in-house file to the main file even though the laboratory had an on-site terminal. The input is data forms, and the output is computer printouts.

DATA BASE REQUIREMENTS

The CLIS analytical and ID application area will support the information handling requirements of many analytical methods. The following is a list of those methods most commonly used in current laboratory procedures. The CLIS approach to the data base requirements of these methods and instrumentation is decribed in the following paragraphs:

- IR Spectroscopy
- Mass Spectroscopy
- UV Spectroscopy
- X-Ray Diffraction
- Emission and Atomic Absorption Spectroscopy
- Gas Chromatography

Gas chromatography may be used to identify unique chemical mixtures or it may be used to separate the constituents of such mixtures prior to analyzing them individually. All other items in the above list are usually used to identify more or less pure chemical compounds; however, they are sometimes used in the identification of mixtures. The GC/MS provides separation followed by identification. In each technique a process is applied resulting in a data pattern which is unique for each unique compound or mixture. The files described below will contain patterns for known compounds, against which patterns resulting from the analysis of unknowns can be compared. When a match occurs, the unknown has been identified.

In these specifications, probable data elements, a provisional record layout, and estimates of file size and growth rate are given for each instrument support file. In the record layouts variable-length data element fields have been proposed wherever applicable. This will result in less wastage of file space than with fixed-length fields and overflow areas. Record layouts and file estimates are given in characters. One character is taken to equal one eight-bit byte. The following codes are used in the record specifications.

A = Alphabetic
AN = Alphanumeric
B = Binary (non-character) bytes
N = Numeric
V = Data element length or number of data elements is variable; the value given is considered representative for file estimation purposes.

Division of the record into computer words is not specified, since word length is machine-specific and unknown prior to hardware selection. Storage of one character per byte in eight-bit bytes is assumed.

Infrared Spectroscopy. In the IR spectroscope an unknown sample is subjected to infrared radiation. The wavelengths of the reflected rays are separated, and the amount of radiation reflected or absorbed at each of a series of wavelengths is measured. The resulting pattern may be related to the molecular structure of the unknown compound. The compound can be identified by comparing its pattern with ones made from known samples.

Most of the data bases for computerized spectral search systems do not contain actual spectra. A digitized representation of an actual spectrum requires that a y-displacement be stored for each increment along the graph of the spectral waveform — a hundred or more numbers for each spectrum. Instead, for UV and IR spectra a "bit-pattern" is used, with binary 1's representing important peaks and 0's representing "no-band" areas or portions of the curve without peaks. This effects a six-to-eight-fold reduction in the computer storage space required. In addition, visual examinations of this spectrum will give the experienced spectroscopist clues to the structure of the unknown, such as presence of structural components or substituents such as C=C, chlorine, OH, NO_2, etc. These and other characteristics (e.g., liquid, polymer, inorganic, etc.) are stored for the standards and increase the efficiency of the search by allowing the searching program to use information other than the bit patterns alone.

All search systems investigated use the American Society for Testing and Materials (ASTM) spectral file. To this, most have added spectra from other sources converted to the same format. When an investigator receives a list of "hits" from a computerized search, he may use the serial number identification to find, in an ASTM or other reference, a citation of where the spectrum was originally published. He may then visually compare his spectrum with the original, if that source is available to him.

File Content (Data Elements). For a file of infrared spectra:

- Compound name
- Compound molecular formula
- Most important peaks in the compound's spectrum
- Intensities of the most important peaks
- Most important "no-band" regions in the compound's spectrum
- Normal physical state of the compound
- Generalizations about the compound's structure (such as polymer, aromatic, etc.)
- Important substituents and structural elements in the compound's molecule (e.g., methyl, hydroxyl, nitor, triple bond, etc.).

Examples of other data elements in support of various instruments might be relative retention indices (for gas chromatography), molecular weight, instrument characteristics (such as temperatures and voltages for a mass spectrometer), Wiswesser Line Notation (WLN, from which the structural formula can be determined), etc.

File Structure. If the entire file were to be searched for each iteration of a spectral search, a sequential file structure would be adequate. In some cases, however, it might be better to search only a portion of the file (such as pharmaceuticals only out of a larger file). To accomplish this, a mechanism to directly access subfiles via a directory would be most efficient and would require a random-access file. (If the file were stored online, it would automatically be random-access.)

Record Layout and Element Coding Structures. Data elements (fields) which might be found in a record for an I-R spectrum are as follows. Such a record would have a length of 246 characters. For purposes of estimating file size, a fixed-length record is assumed which allows chemical compound names of up to 48 characters. Continuation records would need to be provided for those names longer than 48 characters. (In actual practice it might be more efficient to provide variable-length records which hold the entire compound name contiguously whatever its length, thus regaining the waste space created when compound names are less than 48 characters and doing away with the necessity for a continuation mechanism. Space might be gained by making the molecular formula field variable-length also.)

Data Element Name	Number of Characters	Type of Characters
Class of evidence	2	A
Serial number	6	AN
Update and access status	2	B
Compound name	48(V)	AN
Molecular formula	16(V)	AN
WLN	24(V)	AN
Molecular weight	8	N
Instrumental information	16	AN
Physical state and substituents	4	B
Spectrum	120	B

The total length of typical record, 246 characters, contains a considerable portion for the storage of peak intensities and not just their presence or absence.

Access Criteria and Indexing Methodology. The "industry standard," the ASTM IR file, is divided into subfiles on the basis of source (Sadtler, Infrared Data Committee of Japan, etc.), and of these, the Sadtler subfiles are divided into sub-subfiles by type of compound (fibers, pharmaceuticals, etc.). In the building of a CLIS data base, division of the file by type of compound and secondarily by source (each in-house file from a member laboratory would be a separate source) might be considered. If it were a random-access file, it could be searched via directories, either by type of compound or by source.

Additions to the file could be added to the end or they could be inserted in a more proper place by a periodic sort/merge operation; this is true for either random or sequential files. Corrections to a random file can be made *in situ* if the corrected record is not longer than the preexisting one.

Estimated File Size and Growth Rate. The ASTM file is expected to reach 140,000 spectra this year (1974). To contain this in 246-byte records would require 34.4 million bytes of storage. One might assume a growth rate of 25,000 spectra from all sources per year, or more than 6.1 million additional bytes per year.

To contain the Sadtler pharmaceutical file (1200 spectra) plus the Home Office Central Research Establishment (HOCRE) forensic and alkoloid files (2,300 spectra), by analogous calculation, would require about 0.8 megabytes and assuming 500 new spectra per year, would grow at a rate of 123 kilobytes per year.

Mass Spectroscopy. The Mass Spectral Search System, whose data base stores essentially complete spectra, has an input format containing up to 16,159 characters per spectrum. At one byte per character and 37,000 spectra (predicted for 1974), the file in this form would require 598 megabytes storage. A growth rate of 5,000 spectra per year would require another 80.8 megabytes each year. However, the input format allows a great deal of opportunity for data compression: Peaks with zero intensity, metastable peaks, etc., need not be stored; the 210 characters allowed for each compound name could be reduced by making the field variable length, etc. Assuming arbitrarily a reduction by 40 percent, the numbers become 354 megabytes with 48.5 megabytes per year growth.

Ultraviolet Spectroscopy. The UV spectroscope is completely analgous to the IR spectroscope in that a sample's reflection and absorption of radiation is measured at various wavelengths. The ultraviolet region of the electromagnetic spectrum, however, extends in the opposite direction from the visible wavelengths than does the infrared region. The same instrument is sometimes also used to make measurements in the visible region.

Record Layout. The same as for the IR spectroscopy data base.

Estimated File Size and Growth Rate. Using the 246-character record size, it would require about 9.10 million characters of storage for 37,000 UV spectra. A 15 percent growth rate, or 5,550 additional spectra per year, would require an annual storage increment of 1.366 million characters.

X-Ray Diffraction. X-ray diffractometry is a technique wherein an unknown substance, usually crystalline (or at least structured, such as some large

polymeric organic molecules); is subjected to X-irradiation. Each substance produces its own characteristic X-ray diffraction pattern and can be identified thereby.

Data Elements, Records, and Files. Detailed specification and estimation must await further study.

Emission and Atomic Absorption Spectroscopy. The emission spectrometer is an instrument in which a sample of an unknown is burned and in which the resulting light intensity at certain spectral deflections is proportional to the amounts of particular chemical elements present. The readout may be the light intensities corresponding to these elements or their actual concentrations calculated automatically. The atomic absorption spectrophotometer performs measurements complementary to those done by the emission spectrophotometer. Instead of measuring the amount of radiation emitted at various wavelengths, the instrument measures the amount absorbed. Some instruments are capable of performing both techniques and will sometimes also include atomic florescence.

Data Elements.

- Compound name
- Compound molecular formula
- Instrument information (e.g., preflush, preburn time, exposure time, excitation source parameters, dark current, amplifier drift, optical background, dynode voltages, etc.)
- Analysis information (e.g., standardization and background specifications, statistical procedures used, format information, etc.)
- Identity of elements measured
- Concentrations of intensities for these elements

RECORD LAYOUT

Data Element Name	Number of Characters	Type of Characters
Class of evidence	2	A
Serial number	6	AN
Update and access status	2	B
Compound name	48(V)	AN
Molecular formula	16(V)	AN
WLN	24(V)	AN
Molecular weight	8	N
Instrument parameter 1	6	N
Instrument parameter 2	6	N
Instrument parameter 3	6	N
Instrument parameter 4	6	N
Number of analyses averaged	1	N
Format of results (concentration, intensity ratio, or absolute intensity)	1	A
Element used for internal standard	2	A
Internal standard I.D.	2	A
Internal standard result	6	N
Background I.D.	2	A
Background result	6	N
Element I.D.	2	A
Element result	6 X 32(V)	N

A typical emission spectrum record would thus total 344 characters.

Estimated File Size and Growth Rate. The single reference collection of emission spectra encountered contains only 500 entries. If one assumes a criminalistic data base of 2,000 emission and atomic absorption spectra growing at a rate of 15 percent per year, a file requirement of 700,000 characters and an annual increase of 100,000 characters would result.

Gas Chromatography. Basically, this is a technique werein a gas is passed through the sample to be analyzed. Different components of the sample will be *eluted* at different rates, and this characteristic can be used to separate a pure substance from a mixture. Under a given set of chromatographic conditions, a specific compound will resist elution for a characteristic period of time. This fact makes it possible to calculate such statistics as retention times, retention indices, and capacity ratios which can be used for the identification of unknowns. Information of this latter type would be contained in a CLIS GC data base.

Data Elements.

- Compound name
- Compound molecular formula
- Instrument information (e.g., make and model, temperatures, rates, etc.)
- Analysis information (e.g., active/support/phase information flows, etc.)
- Retention data (times, indices, capacities, areas, etc.)

Estimated File Size and Growth Rate. The ASTM Gas Chromatographic Data Compilation and supplements contained 43,000 items as of 1971. Use of this in a beginning data base of 300-character records would require 12.9 million characters of storage. A 15 percent per year growth rate would require 1.94 million additional characters annually.

Total Estimated File Size and Growth Rate. Assuming 10,000, 100-byte entries in other instrument-support files and a growth rate of 1,500 entries a year, and including full UV and IR files but not a mass spectra file, CLIS would require a total file size of 58.9 million bytes growing at 10 million bytes a year. Including the compressed mass spectra file, the requirement would be 412 megabytes and 60 megabytes per year.

RECORD LAYOUT

Data Element Name	Number of Characters	Type of Characters
Class of evidence	2	A
Serial number	6	AN
Update and access status	2	B
Compound name	48(V)	AN
Molecular formula	16(V)	AN
WLN	24(V)	AN
Molecular weight	8	N
Instrument manufacturer code	2	A
Model designator	10	AN
Instrument parameter	6 X 16(V)	N
Analysis parameter code	2 X 3(V)	A
Analysis parameter	6 X 8(V)	N
Type of retention data	2	A
Retention data	6 X 4(V)	N

The typical GC record length would be 294 characters.

PROCESSSING FUNCTIONS

Data Encoding and Allocation Schemes

Data Base Establishment and Updating. It is probable that at least some of the spectra from various sources would have to be rendered into a standard format in either a manual or a computerized system. In a computer system there would be a further translation from an input format to a file format, done by program.

Entry of Individual Spectra into the System. The same sort of encoding would be required either before mailing a spectrum to be identified or before entering it into a computer terminal. Again, on the computer there would be an additional translation from input to file format.

Encoding operations required of the spectroscopist:

- Conversion of physical state, known structural components and substituents, etc., into numeric codes
- Listing of the wavelengths of the most important peaks
- Expression of the "no-band" areas as ranges of wavelengths
- Conversion of the structural formula to WLN (for a known compound).

Translation done by the computer:

- Conversion of numeric codes, wavelengths, and wavelength ranges into bit patterns
- Establishment of status masks
- Establishment of continuation codes pointing overflow records for overlength information, or other processing of variable length data.

Some data, such as compound names and molecular formulas, would be entered directly without translation.

As indicated above and especially in the case of mass spectroscopy data, file compaction through the use of variable-length records would be very helpful in reducing required storage space.

File Search and Match Techniques. As mentioned under the file structure, it would be desirable to search the file either by source of spectra or by type of compound. To do this would require having a random-access file entered via directories.

Once a spectrum is found in the file, it is compared against the spectrum being identified to determine whether they match within certain tolerances. Types of tolerances employed by existing search algorithms include the following:

- The bit pattern of the input spectrum is matched against one in the file and a certain number of the bits must be in agreement.
- A "wiggle" is applied; i.e., the input spectrum is shifted slightly up or down the wavelength scale before comparison to compensate for possible instrument or sample differences.
- A percentage variation is allowed in matching mass spec peak intensities

The development of any CLIS-designed search scheme would have to be preceded by a careful evaluation of existing methods and user opinions of each.

Applicability of a Functional Language. Implementation on a computer terminal could be accomplished by having the user learn a functional language related to spectroscopy and other laboratory techniques, or it could be done by cueing and instructing him interactively on the terminal itself. Most users prefer the latter approach once they "know their way around the application," if the instructions do not become tedious. This is less of a problem on a CRT terminal than it is on a Teletype, where the user must wait for the texts to be produced at 10 characters per second.

Either the functional language or the self-instructing approach would work for this application; the decision should be related to the type of terminal hardware selected.

Concurrency of Operation. In a manual system, simultaneity of inputs is not a real problem. If the backlog gets too large, you can hire more people.

On the other hand, depending on their design, computerized systems may or may not be able to do more than one thing at a time.

If CLIS develops into a computerized system that actually has laboratory instruments online, the probability is certainly great that two or more instruments will be running at once. In the more likely event that CLIS supports only a network of online terminals, the probability is still high that more than one will be interacting with the central host computer at a time. The New York State Criminalistic Research Bureau infrared search system encounters this problem with only three terminals. They have not attempted to provide for concurrent usage, but rather the second user gets a "busy" message until the first is done. It is, however, not too difficult to build in a "queueing," "stacking," and/or "polling" mechanism to obviate this problem. This concerns the concurrency of processing at the computer end of the hookup; concurrency must also be considered at the data communications level. This will not be a concern of the CLIS designers if existing network services are used.

If a larger laboratory needs to have more than one worker accessing the system simultaneously, then obviously they will need more than one terminal. The central system need not know whether the terminals are located in the same or different laboratories.

COMMUNICATIONS

Estimated Usage Rates. It seems likely that the average laboratory would produce several outputs a day from two or three instruments. A certain percentage of these would be so routine as not to require CLIS support. Probably several outputs for which identification assistance was desired would be collected and mailed in a batch or the laboratory analyst would sit down at his terminal and do a number of searches in a row.

Necessary Response Times and Required User Interaction. There is at least one commercial search system that operates by mail (Eastman Kodak). Others (e.g., ASTM, DNA Systems Inc.) will provide a batch processing search system. However, the most satisfactory systems are of the interactive type. Iterative searches require interactive access to the data base of spectra. Interactive access does not necessarily mean that the initial response must be immediate. It might be permissible to delay the beginning of the search for a few minutes or even hours as long as the response was interactive once the search had started.

Anticipated Volume of Transmitted Data (Message Volumes). A user at a remote site may approach his terminal either to submit an unknown spectrum for identification or to contribute a known spectrum to the data base. In either case, the amount of encoded information to be transmitted might range between 200 to 300 characters for an IR or UV spectrum and 1,000 characters for a mass spectrum. The information returned by the computer for a search interaction would be on the order of several thousand characters. A 50 to 100 percent overhead should be added for communications and header information required for interactivity. (Examples of such communications would be operator messages and cues, selection of options by the operator, and formatting information. More formatting would be required for a CRT terminal than for a terminal printer since CRT displays tend to be longer, more complete, less abbreviated than printed ones.)

Assuming the following:

- 500 characters to specify a spectrum
- 5,000 characters to inform the user of possible "hits"
- 75 percent overhead for interactivity
- Three iterations in an average search
- One search per day using the automated search facility in an average laboratory
- 100 laboratories using the system,

one arrives at an estimated average daily message volume of about 2.7 million characters.

MAINTAINABILITY

Currency of Data and Sources of Updates for Maintenance

Additions to the File. Subfiles (such as Sadtler or Coblentz) that are included in the ASTM IR file are updated periodically. CLIS could obtain the periodic updates to any commercial or institutional data bases used. However, the standard ASTM file as supplied may lag by perhaps two years the inclusion of these and other changes.

A major source for CLIS data should be the private, in-house collections from the various member laboratories. To these the individual laboratories could contribute additional standard compounds as their spectra became known.

Corrections to the File. As part of their normal updates, the compilers of standard data bases correct erroneous information, remove poorly defined entries, and sometimes delete duplicates. The same processes should be applied to user-supplied entries. The changes could be user-initiated, but should be monitored by the central CLIS facility.

Summary of Maintenance Procedures

- Inclusion of updates provided by purveyors of standard data files
- Addition of user-supplied in-house files.
- Addition of individual user-generated entries
- User-generated corrections, replacements, or deletions
- Monitoring by CLIS of user-generated alterations

INITIAL IMPLEMENTATION

In Volume I under "Priorities," it was suggested that due to its size and complexity, this application be implemented stepwise; with the first phase being the provision of infrared spectrometry support for drug analyses only. Later phases, at a lower priority than most other CLIS applications, would include UV support for drugs, GC-MS support, X-ray diffraction support, and finally IR and UV support for other types of compounds.

After reviewing the entries in Chart 5 (instrumentation) in the greater number of survey forms received, it is recommended that X-ray diffraction be dropped from the list and that other items be added so that the new list of priorities would be the following:

- IR for drugs
- UV for drugs
- Gas chromatography and mass spectroscopy (and GC/MS)
- IR and UV for nondrugs
- Emission spectroscopy, atomic absorption and electrophoresis

There are numerous implementation alternatives in this application area, for example:

- Manual versus computerized retrieval system
- Use of an existing data base and its format versus establishment of, and conversion to, a special CLIS format
- Use of an existing spectral search system versus developing a new one
- If an existing system were utilized, access to it could be gained through a commercial network or time-sharing service, or a copy could be acquired for CLIS and put up on CLIS hardware (and then possibly provide access through a commercial or government network).

Because there are so many implementation paths, it is difficult at this point to precisely define the programming tasks that would be involved. However, the following general categories are offered:

Design and Development of the Analytical/Identification Support Function. IR spectrophotometric support for drug identification is used as an example with some details applicable to other instruments noted.

Putting up the initial data base. (See discussion under "Encoding and Entry of Data" below). An appropriate beginning in this area might be provided by combining the Sadtler IR spectra for pharmaceuticals with Home Office Central Research Establishment forensic and alkaloid IR spectra. Subtasks here would be the design of the encoding procedure and the encoding process itself which might be a major clerical effort (e.g., filling out forms, keypunching or key-to-disk, etc.).

Development of a search mechanism and protocol. This might involve designing or adapting a search algorithm and working out and documenting the users' interactive options.

Design of an update capability. A mechanism and protocol would be needed to add both large and small amounts of data to the file and to correct or refine existing data. This would be needed both for the original file and for user-supplied entries.

Potential Sources of Data. The following sources of standard reference files have been identified.

Infrared Spectroscopy. American Society for Testing and Materials. The ASTM maintains encoded abstracts of all the published IR spectra they can find. It is available by subscription in hard copy or machine-readable form. It includes the files of Sadtler, the European Documentation of Molecular Spectroscopy (DMS), The American Petroleum Institute (API), Coblentz, Aldrich, etc. Actually, it does not contain the original spectra, though it references them so that they may be looked up. A problem is that the compilation process runs a couple of years behind. However, at 102,000 spectra (due to be increased to 140,000 in 1974), it is the largest data base available and is used in all computerized IR search systems.

Eastman Kodak Company has a data base (available only through their Infrared Spectral Retrieval Service) which contains 90,000 ASTM compounds plus 10,000 Eastman Organic Chemicals.

The New York State Division of Criminal Justice Services has a data base consisting of the ASTM file plus forensic files contributed by the New York State Police and the New York City Medical Examiner's Office.

Sadtler Research Laboratories, Inc. sells a large IR data base and a number of specialized subfiles. The quality of the Sadtler spectra has been unfavorably compared with others, such as those in the Coblentz file. However, our questionnaire responses show that, next to in-house files, Sadtler files enjoy the widest usage among criminalistics laboratories, perhaps because they are more actively marketed. The complete Sadtler IR file contains about 34,000 entries. Subfiles used by various crime laboratories include the following (with number of entries): Pharmaceuticals (1,200), Commonly Abused Drugs (600), Monomers and Polymers (5,100), Fats, Waxes and Derivatives (500), Plasticizers (600), Pyrolyzates of Polymers (600), Lubricants (500), and Agricultural Chemicals (500).

The United Kingdom Home Office Central Research Establishment has forensic sciences and alkaloids IR files totalling about 2,300 spectra available on microfilm. These are not included in the ASTM data base.

Other standard sources of IR spectra in use among the laboratories surveyed include Aldrich, Sunshine, E.C.G. Clarke, the Association of Official Analytical Chemists (AOAC), Hummel/Scholl, and the API.

Ultraviolet Spectroscopy. Sadtler Research Laboratories, Inc. markets a UV file of 36,000 entries. There are two subfiles in common use among crime laboratories: Pharmaceuticals (2,000 spectra) and Commonly Abused Drugs (300 spectra).

The United Kingdom Home Office Central Research Establishment provides a microfilm file of UV spectra for about 700 alkaloids.

Other standard UV sources mentioned by laboratories responding to our questionnaire include E.C.G. Clarke, Sunshine and the AOAC.

Gas Chromatography. The ASTM produces a Gas Chromatographic Data Compilation with retention indices and other information for a large number of compounds. It is available as hard copy or on magnetic tape.

Emission Spectroscopy. Sadtler is starting a collection of excitation and emission fluorescence reference spectra of pure organic compounds. It contains 500 entries to date.

Mass Spectroscopy. The Mass Spectral Search System (MSSS) which is available through GE Timesharing, but which is a public domain system developed and operated by various US and UK agencies (i.e.; National Heart and Lung Institute; National Institutes of Health (NIH); Mass Spectroscopy Data Centre (MSDC), HOCRE, England; and Environmental Protection Agency), has a data base containing the following collections: ASTM E14 Uncertified Spectra, Dow Chemical Company Spectra, American Petroleum Institute Standard Spectra, TRC Spectra, MSDC Spectra Collection, Cornell University Spectra, and NIH Spectra. When the John Wiley Registry Data Base is added in 1974, the total number of spectra in the file will be about 37,000.

Manufacturers of computerized MS and GC/MS systems may provide data bases to go with their instruments.

Other sources of MS files are the American Society of Mass Spectroscopists, Finkle and Taylor, and the Massachusetts Institute of Technology.

Miscellaneous Files. X-Ray Fluorescence. ASTM publishes *X-Ray Emission and Absorption Wavelengths and Two-Theta Tables* and *X-Ray Emission Wavelengths and KEV Tables for Nondiffractive Analysis.* Raman spectra — Sadtler's continuing collection of Raman Reference Spectra of pure compounds (2,000 entries). Nuclear Magnetic Resonance Spectra — Sadtler's collection of 20,000 NMR spectra and their collection of 2,000 Carbon — 13 NMR spectra.

In-House Files. Most laboratories have built up their own files of reference standards for various instruments. The collation and dissemination of these files is one of the most significant services CLIS can provide.

Encoding and Entry of Data. Various approaches will be necessary to establish a computerized reference file for the differing analysis methods. Among the possibilities to be considered is participation in existing computerized data bases, such as the Mass Spectral Search System which is accessible on General Electric's Mark III time-sharing service.

Some standard files are already available in machine-readable form. The Sadtler IR files can be obtained on magnetic tape. The ASTM infrared files are available as magnetic tape or disc pack. ASTM also has gas chromatography data on magnetic tape. If CLIS includes such files in its data base, either the existing format could be used, or software written to convert them to a preferred format.

For files available as hard copy only, existing in-house files from various laboratories, and files such as electrophoresis where no data base as yet exists, it would be a matter of encoding the information into a standard format, converting it to machine-readable form, and feeding it to the computer.

For a manual data base a similar process of standardization and encoding would be required, but machine-readable data and computerized conversion would not be involved.

SUMMARY OF SYSTEM REQUIREMENTS FOR ANALYTICAL/IDENTIFICATION SUPPORT

Data Storage Requirements

Manual. To store all the available standard spectra and data compilation to support criminalistics analytical techniques would require perhaps a 150-volume library (there are 45, 1,000-spectrum volumes of Sadtler IR spectra alone.

Computerized. Excluding any space to store a file of mass spectra, the amount of computer storage space required for these analytical support files appears to be on the order of 59 mega- (million) bytes. This requirement would grow at the rate of perhaps 10.6 megabytes per year.

Inclusion of the very large existing data base available from the Mass Spectral Search System would enormously inflate these estimates to 413 megabytes required with a growth rate of 59 megabytes a year.

Both these storage estimates assume the inclusion of complete files in all categories. A useful subset of the application requiring much more modest files could be implemented initially.

Greatest user satisfaction seems to require that file searches be conducted iteratively, and this in turn requires that the files involved be online and randomly accessed.

Necessary Functions or Operations

- Original establishment of the data base
- Data base updating procedure: new published spectra
- Data base updating procedure: user's private files
- Encoding of spectra to be identified
- Spectral search mechanism
- Establishment of users' private file
- Procedure for updating user's private file

Information Input and Output. Conceptually, input information in an analytical/ID support system would be the standard data bases and user's private files. The output would be the identification information provided to the users.

In a manual system both input and output would consist of data encoded on forms.

In a computerized system inputs might be analog signals for reduction, keystrokes at a terminal for spectra to be identified or input into the system, or key-to-machine-readable conversion of initial data bases or large updates. Outputs might be printouts or CRT displays at the terminals and transaction listings at the central information store.

System Availability and Response Times. A manual analytical/identification support system would be useful only if there were very rapid turn-around and even then there is some doubt that this could occur.

The most satisfactory solution would be an online system allowing iterative searches to be conducted interactively through a computer terminal. Initial availability could be postponed for as much as a few hours, as long as response was immediate after that.

Estimated Usage Rates. Analytical and identification support would be responsible for by far the largest communication volume of any CLIS application. Assuming that each laboratory would use the identification support for from one-half to two analytical results per day (depending on the content of the data base and the number of analytical methods supported), the communications traffic might amount to one to five million characters per day.

Maintainability as it Affects System Design. In an automated system a key-to-machine-readable facility should be provided for the addition of large increments to the data base. The capability for entering small amounts of information from terminals should also exist. There should also be a protocol to enable the laboratories to send in-house data bases to the central file.

For either a manual or a computerized system, there should be an individual or individuals with overall responsibility for the application or data base. A computer could assist the responsible persons with update transactions listings.

GENERAL DESIGN FOR RIFLING SPECIFICATIONS

General Operating Functions. The function of the rifling specifications application area will be to provide the firearms examiner with the capability of identifying a fired specimen with a particular caliber designation and a list of firearms that could have fired the specimen. This would be done in a one or two-step procedure depending upon the amount of information available from the fired specimen. The two-step procedure would involve the use of two separate files: an ammunition file and a rifling specification file. The function of the ammunition file will be to contain bullet characteristics in order to relate them to a caliber designation. The function of the rifling specification file will be to contain rifling specifications in order to relate them to makes and models of firearms.

In each of these two procedures, the examiner must make several measurements. In order that the data in the file be accurate and in accordance with the examiner's measurements, measurements of the fired specimen must be made in the same manner as those in the data files. This will involve the definition of standard dimensioning criteria and provide the capability of entering estimated tolerance information. The type designations of bullet base, nose and canneleur characteristics must be standardized and adhered to throughout the identification process. This standardization of measurement methodology must be used throughout the identification procedure in order to make efficient use of the data available in these two files.

In use of the rifling specification file, the single most important piece of input information will be the caliber designation. Note that this is not merely the specimen diameter dimension; it is a description of the cartridge that fired the specimen. For example, a .32 caliber auto pistol and a .32 caliber lead bullet may have the same diameter dimensions but they are actually two different caliber designations that would have chambered that cartridge. The purpose of the ammunition file is to determine the caliber designation from the physical characteristics of the specimen. In many cases it will not be necessary to use the ammunition file for this purpose, as the firearms examiner can readily identify the caliber designation of the specimen due to his extensive knowledge and experience. When this information can be readily determined, there is no need to access the ammunition file, and the inquiry can be made directly into the rifle specifications file. Thus, the ammunition file is used to verify the examiner's identification of the caliber designation and to aid him in this identification when the condition of the specimen is such that some features have been mutilated or when the specimen are beyond the scope of his experience.

The first step in the procedure is to identify the caliber designation of the fired specimen. Identification of the caliber designation will provide information on the possible cartridges that could have contained the fired specimen, which would then lead to firearms that have the capability of chambering that particular cartridge. In some cases the firearms examiner is able to quickly determine the caliber designation of the specimen just by making a few simple measurements of the fired specimen. However, quite often portions of the specimen have been altered by impact and identification must be attempted with measurement of only a few of the specimen's characteristics. These specimen characteristics would be:

- Rifling diameter
- Circumference
- Length
- Weight
- Type of nose
- Type of base
- Type of jacket (if any)
- Number and type of canneleurs
- Dimension from base to lower canneleur
- Inter-canneleur dimensions
- Dimension from nose to upper canneleur

Not all of these characteristics may be measurable due to the condition of the specimen. From these measurements it would be possible to list the caliber designation(s) of the fired specimen.

The second step of the procedure would be to identify a list of firearms that could have fired the specimen. The following measurable characteristics of the specimen would be determined:

- Caliber designation
- Number of lands and grooves
- Width of lands and grooves
- Direction of twist
- Circumference

Inclusion of circumference data is a validation of the number of lands and grooves and the summation of their respective widths.

From this information the system would provide a list of firearms by make and model that could have fired the specimen. Figure 3 depicts the flow of operations of these procedures.

Data Base Requirements. The ammunition file will contain the physical characteristics of bullets and their caliber designations. The file will be structured as a series of individual records describing physical characteristics and caliber designation. The required record elements and estimated allocations are as follows:

Element Name	Number of Characters	Type of Characters
Bullet diameter	3	N
Bullet circumference	4	N
Bullet length	4	N
Bullet weight	3	N
Type of jacket	1	A
Type of nose	1	A
Type of base	1	A
Number of canneleurs	1	N
Type of canneleurs	1	A
Width of canneleurs	3	N
Dimension base to lower canneleur	3	N
Dimension nose to upper canneleur	3	N
Inter canneleur dimension	3	N
Metallic composition	10	AN
Caliber designation	20	AN
Source of data	14	AN
Additional information	150	AN

N = numeric
A = alphabetic
AN = alphanumeric

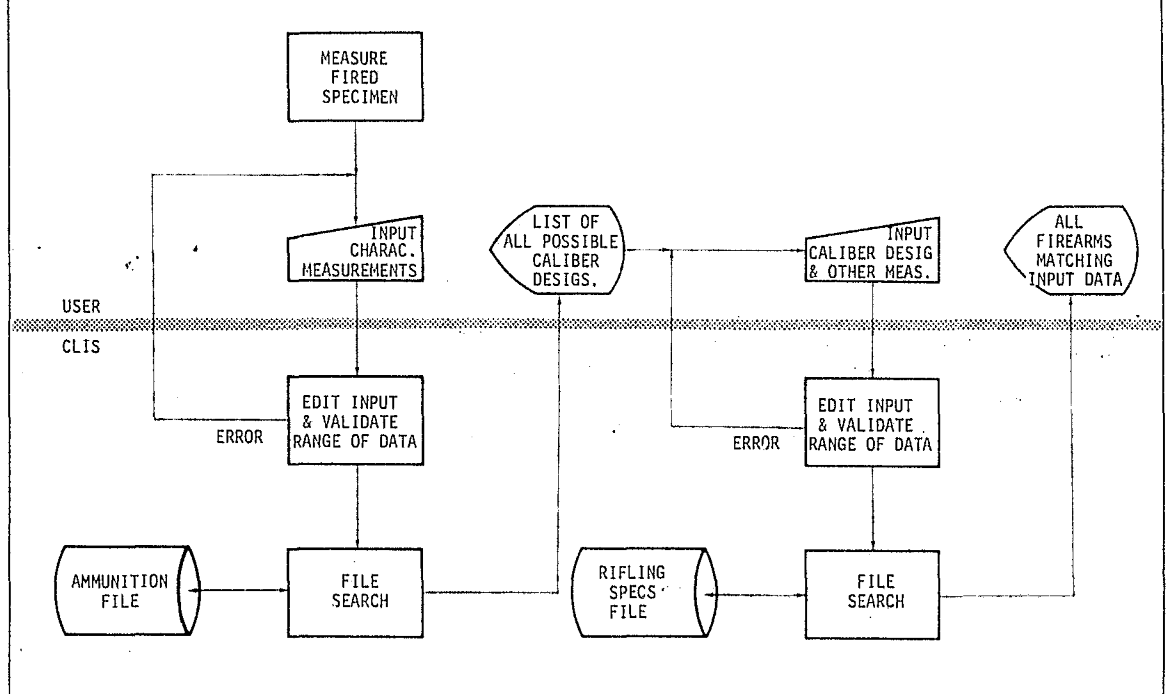


Figure 3 Information Flow — Rifling Specification

The character allocations for dimensions are based on the use of the English measurement system. Any necessary metric conversions can be accommodated in the processing function.

The file should be ordered first by jacketed and nonjacketed bullets. Within each of these categories, it should then be ordered by bullet diameter with further classification by metallic composition and type of jacket. This will result in two sequential subfiles and should be sufficient to reduce the area of search to a reasonable number of records without the need for further indexing as long as a prime index into the sequential file is available.

Based upon the size of the most extensive manual ammunition file currently available, it is estimated that the initial CLIS ammunition file would consist of approximately 7,500 records with an annual expansion rate of about 100 records per year. This would require an initial data base size of 1,687,500 characters with an anticipated annual increase of 22,500 characters.

The rifling specification file will contain the rifling specification characteristics and a description of the firearm. The file will be structured as a series of individual records describing the characteristics of a firearm. The required record elements and estimated allocation are as follows:

Element Name	Number of Characters	Type of Characters
Caliber Designation	20	AN
Number of Lands and Grooves	2	N
Direction of Rifling Twist	1	A
Land Width	3	N
Groove Width	3	N
Weapon Description	100	AN
Source of Data	14	AN
Additional Information	57	AN

As stated for the ammunition file, the character allocations are based upon use of the English measurement system.

The file should be ordered first by caliber designation, second by direction of rifling twist and then by number of lands and grooves with further breakdown by land and groove width if necessary. As in the ammunition file, this will result in a sequential file with no need for a further index or cross reference.

Based upon the size of current manual files and indexes, it is estimated that the initial CLIS rifling specification file will consist of about 3,000 records with an estimated annual increase of 500 records. This would require an initial data base size of 600,000 characters with an anticipated annual increase of 100,000 characters.

Note that the record content of both the ammunition and rifling specification file will be generally consistent from record to record, and there is no requirement for interrecord indexes or pointers to relating records or indexes. This strongly suggests that both files can be adequately accommodated with a relatively simple sequential file structure with a prime index serving as an access key.

Processing Functions. From the standpoint of effective use, the system must be simple to operate. The user should be able to obtain useful information with a minimum of input data and keystrokes. A user reference manual must be developed to define the terminology, abbreviations and methodology of system use. This will provide an easy-to-use guide to system operation.

Due to the relative simplicity of file organization, it is anticipated that a formal user language will not be required for this application area. The operational procedure should be primarily conversational, with as much system editing as possible. After program initiation, the system will query the user for data on each element necessary for the file search. After user input of each element, the system will check for data validity between ranges and ask for tolerance information, if applicable. The user has the option of skipping over those elements for which he is unable to provide information. The following illustrates how the system might interact with the user:

S = system **U = user**

U: Riflefile
S: THIS IS RIFLE FILE, ENTER CALIBER DESIGNATION
U: 25 cal
S: NUMBER OF L AND G?
U: 6
S: TWIST?
U: L
S: LAND WIDTH?
U: .069
S: TOLERANCE?
U: +.005, —.003
S: GROOVE WIDTH?
U: .95
S: UNACCEPTABLE — OUT OF RANGE
S: GROOVE WIDTH?
U: .06
S: TOLERANCE?
U: +.01, —.01
S: ANY FURTHER INFORMATION?
U: Search
S: THE FOLLOWING FIREARMS ARE LISTED AS HAVING THE SPECIFIED CHARACTERISTICS:
Colt Auto Pistol (Describe weapon and characteristics) ASTRA (Describe weapon and characteristics) and so forth as applicable.

The system processor must have the capability of editing input data to detect inadvertent keystroke errors as well as obtaining sufficient information to effect a search. When the minimum information has been solicited from the user and the user has inputted any additional information, the system will begin the search, examining each record with the prime designator (25 cal in the above example) and displaying to the user each record that matches the input data within the prescribed tolerances. This same processing function can be used for both files in this application area due to their similarity of organization and search criteria. It should also be capable of supporting a minimum of three simultaneous file search operations.

Communications. As indicated in the previous section, the interaction between the user and the system should be accomplished in a conversational mode in order to minimize the input data error rate. Typical system response to user input of data should be less than ten seconds. In the data input mode of the search process, approximately 500 characters would be interchanged between the system and the user. Since the system response could include the listing of many records depending upon the input data, it is difficult to determine exactly how many characters would be needed to complete the entire transaction. An estimated average of four responses per transaction would indicate that approximately 1,500 characters would be required.

Based upon questionnaire results, the average usage of in-house firearms and ammunition files is about 2,000 times per month. Using a 160-hour work month, the estimated usage would be about 13 times an hour. Note that this is a gross estimate and could easily double during peak hours of use. This indicates that the communications usage for this application would be approximately 25,000 characters per hour for all users of the system.

Maintainability. Since the data in both files of this application area represents the characteristics of firearms that are currently in general circulation, it will remain current for as long as those firearms are available. It is anticipated that there will be very little purging of this file. Updates, however, must be made whenever a new type of firearm or bullet is developed and made available by the manufacturers.

Addition of data to these files must be carefully controlled to prevent the admission of erroneous data. As new bullets or firearms are made available, their specifications can be obtained from the manufacturer. A much better method of acquiring these specifications would be to obtain a sample bullet or firearm from the manufacturer and have a responsible laboratory test fire and measure the bullets and fired specimens. In this way, measurement inconsistencies can be reduced, since the same organization would be making and updating all measurements.

It is recommended that a small group of experienced firearms identification and ballistics personnel comprise a review committee to supervise the inclusion of all information added to or deleted from these files. This would provide a central control for firearms information and measurement techniques. Actual updating and resequencing of the files is a relatively minor procedure and can be accomplished during nonprime time.

Initial Implementation. The initial task in the implementation of this application area is to define in detail the content of the files and the detailed functional procedure that must be used by a terminal operator to obtain information from these files. This will specify inquiry and response formats, editing criteria and error designations, and detailed file search and match operations. From this functional description, a detailed program and file allocation specification can be written. Programs are then coded and tested against the functional description.

Generating the data base will involve a large effort in obtaining the available specifications, measuring those characteristics that are unavailable, and encoding the data in machine-readable form. Wherever possible, the data should be obtained from reliable measurements, using uniform procedures. Again, it is recommended that this be accomplished under the direction of a small group of experts who will review all functional specifications and data for inclusion into the two files. Some of the sources for the data are published texts and tables as well as firearms and ammunition evidence collections maintained by some laboratories. Some of the prime sources of this data are:

- FBI Standard Ammunition File
- Matthews Text on Firearms Identification
- "Cartridges of the World," by Barnes
- "Cartridges of Rifles and Pistols," by Marshall and White
- Manufacturers' Specifications
- H. P. White's Index, currently Firearms Identification Service

Systems Requirements

Data Storage. The estimated data storage requirement is 1,687,500 characters for both files. Due to the interactive, real-time nature of the user/system interface, this storage should be online and on-direct access devices such as disks or drums.

Necessary Functions and Operations

- User reference manual
- Input editing
- User/system interaction
- File search and match within specified tolerances
- Process three searches simultaneously

Figure 3 shows the processing functions for this application.

Information Input and Output. Information input to the system describes the physical characteristics of the specimen and can be entered via a normal data entry keyboard. System outputs will consist of command inquiries for input and printouts of selected records.

System Availability and Response Times. Both files in this application area should be made available online during first and second shift working hours and extended to a 24-hour basis if possible. Total system response time (does not include operator input time) for each transaction should not total over two minutes.

Estimated Usage Rates. Based upon questionnaire responses, the estimated usage rates would be approximately 40 transactions per month per user.

Maintainability. Additions to the files should be made at least on a weekly basis. Any errors discovered in data already online should be corrected by the review committee and updated as soon as the correct information is available.

GENERAL DESIGN FOR ABSTRACT/BIBLIOGRAPHY

The scope of this section is intended to include both literature abstract indexing and bibliographic search capability. Since an abstract listing must also contain bibliographic information, all subsequent references to the bibliographic data base also include abstract indexing. Depending upon the source of bibliographic data, abstracts may or may not be available along with each citation. If an abstract is available, it will be included as part of the data base under the same subject headings as the citation. If it is not available, the citation will be included in the data base under its appropriate subject headings.

In this application, the data base is the bibliography; i.e., a list of sources of information on a given subject(s). The data base will be a collection of citations with appropriate indexes. Each citation is a direct reference to a particular article or text of interest to the forensic laboratory community.

General Operating Functions. As indicated previously, the data base is a collection of citations. It is the goal of the user of the system to obtain a list of citations (if any) that pertain to a specific set of selection criteria that defines the subject area to be searched. In order to accomplish this, there must exist a set of subject headings that define all areas of interest in forensic science and each citation must be categorized under every subject heading that the original article refers to. This lexicon of subject headings is extremely important in the overall operation to this application, as almost every search will involve references by subject heading.

A normal sequence of operation would be for the user to input a subject heading and the system would respond with the number of citations categorized under that subject heading. This would enable the user to determine whether further classification would be necessary before calling for a printout of all applicable citations. The user would have the capability of specifying boolean combinations of subject headings, authors or titles in order to reduce the number of applicable citations to his specific area of interest. The citations would then be printed at the user's command.

Data Base Requirements. Each citation in the data base would constitute a data base record. Each record would contain the following information on the referenced article:

Record Element	Estimated Number of Characters
• Title	200
• Author or Authors — list all	100
• Abstract — if available	1,000
• Source — article publication information	50
• Date article published	6
• Date citation entered into system	6
• Subject headings — list all	150

All characters would be alphanumeric. A subject heading encoding scheme could be implemented to reduce the number of characters in the subject heading record element.

Due to the flexibility of the types of information in the data base, it is appropriate that the records be organized as variable length with variable length record elements.

In addition to the file containing the subject headings, there will be a number of indexes which will reference records in the data base. The selection of data base items to be indexed will largely depend upon the anticipated user interaction with the system. The following indexes are considered to be minimal: Subject Heading, Title, Author.

All access into the data base will be by use of one or more of the indexes depending upon the search criteria specified by the user. Should the list of subject headings become too complex and interrelated, it would be possible to subclassify the subject headings into other categories such as functional or methodological or perhaps key word entries. This subclassification would necessitate a separate index into the subject heading file.

Based upon the current number of journals and periodicals in the forensic area, there would be about 1,200 citations per year available for entry into the CLIS. Assuming that all articles published within the past five years would be available for review and data base operation, the initial data base size would be approximately 6,000 citations. An average data base record of 1,500 characters would be sufficient to include a 150-word abstract as well as consideration for indexes. This would indicate an initial data base requirement of 9,000,0000 characters with an annual growth rate of approximately 1,800,000 characters.

Processing Functions. The functional operation of a bibliographic search will involve a significant amount of interaction between the user and the system. For this reason it is recommended that a simple instructional language be designed and implemented as part of the operation of this application area. This would permit the user to use such commands as PRINT, SEARCH, LIST SUBJECT HEADING BREAKDOWN or HELP (for instructional information in problem situations) as well as implementing the boolean operations and inputting search criteria.

Implementation of a user language will require additional processing capabilities. In addition to normal data input editing, a command language interpreter will be required as well as an input subject heading string builder. These programs assist the user in building his search criteria and provide for strict adherence to system operating procedures. Input editing will consist mostly of subject heading lookups and validation, and syntax analysis. Once the search criteria has been input by the user, the processor will access the indexes and by selective inclusion and exclusion of data base records, develop a list of citations to be returned to the user. A match of the search criteria with a particular citation will not automatically cause the citation to be printed; this is accomplished only by a suitable print command at the request of the user. Figure 4 depicts a possible logical flow of operations to be performed for a typical search.

Attempting to find a reference to a specific subject area can result in a substantial amount of "browsing" through the data base. The processing functions must be able to support simultaneous use of this application by several users. The processor must also be able to provide documentary assistance to the user in the form of an online tutorial package. This will be particularly valuable in operator training and system familiarization. It should also have the capability of displaying a breakdown of the subject

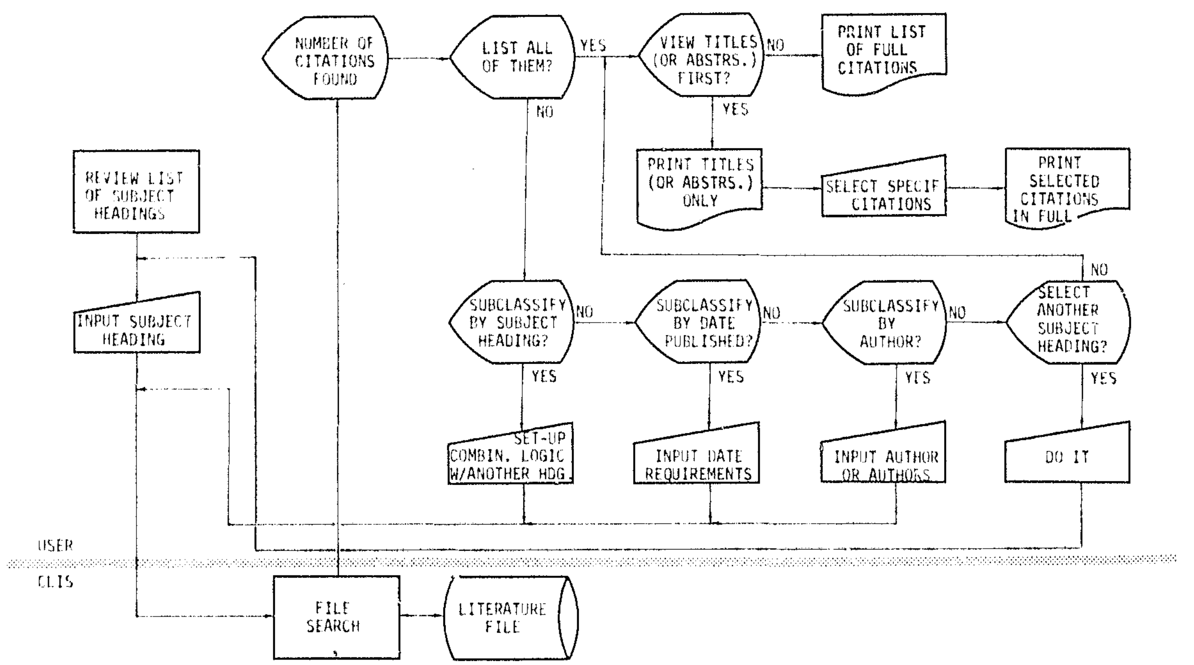


Figure 4 Information Flow — Bibliographic and Literature Abstract Services

heading list as a sort of menu available to the user.

Communications. Use of a functional application language as described in the previous section indicates that the system must interact responsively and continuously with the user. Typical system response to user command or search input should be less than ten seconds. Since the user "browsing" time could be quite extensive, and the list of citations long, it is extremely difficult to estimate the volume of transmitted data. Assuming an average search of a combination of six subject headings which list three citations, an estimated 5,000 characters would be transmitted for each search.

How often a laboratory would use a bibliographic search service such as this is difficult to determine, since laboratories were unable to estimate current activities. However, assume that each laboratory responding to the CLIS questionnaire made one search per day (eight-hour working day). This would amount to 125 searches per day or about 15 searches per hour. At 5,000 characters per search, the communications usage would be approximately 75,000 characters per hour for all users of the system.

Maintainability. Updating the bibliographic data base is a continuous task. Due to the fact that most of the citations will reference articles in periodicals, the citations must be indexed and abstracted (if applicable) on a fixed schedule. Citations to dated information must be purged according to specified time lapse criteria or the availability of more recent scientific data.

Most prominent in the maintenance process will be the selection of periodicals to be reviewed. As the field of forensic science expands in its use of available technology, the list of subject headings must be kept current with the activities and interests of the forensic laboratories. The supervision of these two critical maintenance tasks should be entrusted to a small select group of experts in the forensic sciences.

The actual periodical reviewing and data input procedures can be accomplished through the use of specialized contractual services. Commercial firms, as well as government organizations, have the capabilities and staff to provide abstracting and indexing services in specialized areas such as those in forensic science.

System Requirements

Data Storage. The initial data storage requirement of this application area is 9,000,000 characters. This does not have to be immediately, however. Conversion of data for the initial implementation will take at least several months, and this initial conversion data will be added to the file over this period of time. File growth is estimated at 1,800,000 characters per year.

Necessary Functions and Operations. The primary task in implementation will be to develop the definition of a user language and subject heading list. The subject heading list must be flexible so as to accommodate future expansion and possible subclassification of some subject areas. Concurrent searches by several users must be supported to provide browsing capability.

Information Input and Output. User provided information will provide search criteria for system matching algorithms. The system must also support various combinatorial inputs of subject headings, titles and author information. Response will be to list number of matches, partial selective record element display and full record (citation) display.

System Availability and Response Time. Due to use of a user language, response times must be rapid enough to maintain user attention. This response can be subject to degradation when the user is casually browsing through general categories. System availability could be limited to normal working hours.

Estimated Usage Rates. The estimated rate of usage is once per laboratory per day. However, this can be expected to increase if the citation data is truly responsive to user needs. Use of this capability by formal laboratory training programs will also increase usage of this application area.

Maintainability. File maintenance will be an ongoing procedure of adding citations to the data base. New citations should be placed online as soon as they are available so as to provide the user with the most current references to the selected subject headings.

Initial Implementation. The initial task in the implementation of this application area is to define the detailed protocol and functions of the user language. This should be followed by the development of a subject heading list and a detailed layout of the file-indexing structure. The functional definition of the user language will result in the development of two documents, the user's manual and the programming specifications. These programming specifications in conjunction with the file-allocation and indexing definitions will provide sufficient technical detail to code and test the computer programs.

Data conversion and data base entry is dependent upon the development of the list of subject headings. Once this list is complete and meets the satisfaction of the user community, data entry can begin. The first priority would be to organize the maintenance function such that the most current information will be available to the users. The next step would be to convert the appropriate contents of currently available sources such as:

- MEDLARS
- INFÖRM
- NCJRS
- FBI Abstracts
- AAFS What's New.

The conversion of some of this information will be somewhat difficult due to format inconsistencies from system to system. However, this method is superior to going back to the information source and developing the citation and abstract data.

GENERAL DESIGN FOR SOURCES OF SPECIALIZED KNOWLEDGE

The scope of forensic science covers a large portion of the technical spectrum. Members of a laboratory staff are frequently required to testify in court as qualified experts in the particular field or fields in which they specialize. The question of who is qualified as an expert in a specific discipline is a particularly difficult problem in forensic science because of its effects on jurisprudence. The court frequently establishes a person as an expert based upon his education, training, experience and previous court qualifications. Leaders in specific fields are often recognized by their peers as experts, although no single set of criteria can serve to define an "expert."

This application area will provide a list of sources of specialized knowledge. The fact that a person is included in this CLIS list would not be construed as an endorsement of his capabilities and experience. Inclusion in this list would indicate that the referenced person can be considered to be knowledgeable in the specified fields of interest. The CLIS specialized knowledge file should be used as a reference tool and not as a credentials validator.

General Operating Functions. A user of this application area would typically be confronted with the analysis of a specimen that is beyond his experience or scope of knowledge and training. The conventional method of solving this problem involves the laboratory examiner's personal knowledge of an acquaintance with "experts" in the specific analytical area to be analyzed. By consulting with these "experts" he is generally able to solve that particular problem with their guidance. Laboratory examiners generally do not hesitate to admit their need for assistance and will seek help when faced with problems in areas which perhaps are not a part of their normal disciplines.

If the laboratory examiner has no knowledge of where to obtain this specialized help, he must spend a significant amount of research time in qualifying and locating specific people. Use of the CLIS system will readily identify potential personnel considered to be knowledgeable in selected specialty areas. The user would merely approach the system and input the specific forensic category that he is interested in. The system would respond by supplying the user with a list of personnel within that category. The user should also have the capability of further classifying the resultant list by selection of a specific geographic area.

Important to the operation of the system is the categorizing of specialty areas of technical importance in the forensic science field. For instance, some of these specialty areas would be:

- Serology
- Odontology
- Tool Marks
- Forensic Toxicology
- Narcotic Ballistics

It is estimated that 100 to 150 of these categories would suffice to cover the forensic science field.

Data Base Requirements. The sources of the specialized knowledge file will contain names and addresses of people considered to be experts in their field. The file will be structured as a series of records, each record describing the name, address and specialty area of the person involved. The required record elements are listed as follows:

Record Element	Estimated Number of Characters
• Name	40
• Title	30
• Professional Affiliation	50
• Address	50
• Telephone Contacts	20
• Specialty Areas	10
• Fee Information	40
• Geographical Area of Residence	4

It is estimated that a fixed record size of 250 alphanumeric characters would be sufficient to contain all of the above information. In the event of record overflow, which may occur in the listing of many specialty areas, a second record could be appended to the first along with appropriate continuation flags, or an internal numeric code could be used instead of the full-text of the specialty area. The file should be ordered alphabetically by name, thus producing a sequential file. Access to the file may be made by name; however, the prime access method will be by specialty area. Therefore, an index file by specialty area must be generated and maintained with appropriate pointers to the name file.

Current studies indicate that there may be as many as 6,500 people who can be considered experienced and competent enough to be included in this file. Although the actual file size will probably be significantly less, this can be considered a maximum for file-size estimation. This would produce a file storage requirement of approximately 1,625,000 characters. The annual growth rate can be considered insignificant in comparison to this maximum file size.

Processing Functions. Access to the file will be simple and straight-forward, thus precluding the need for a user language. The user would input a specialty area and the system would respond with a list of those persons considered to be knowledgeable. This list can be further classified by use of geographic input which would reduce the response to only those records in the specified area. A complete list of the specialist areas should be available to the user upon request.

The processing function will also have the responsibility of editing user input for valid specialty area designations as well as geographic codes. Since the mode of interaction will be inquiry/response and the anticipated usage low, it is not necessary for the system to support the capability of more than one search at a time. Figure 5 depicts a possible logical sequence of operations which could meet the requirements of this application area.

Communications. At this time there is little evidence to base a usage estimate upon. Since the average priority of this application was fairly low, it is doubtful that this function will be used more than four times per day. Assuming that an average response to a specialty area inquiry would list 12 records, the total daily transmission rate would be 12,000 characters. Based upon an eight-hour day,

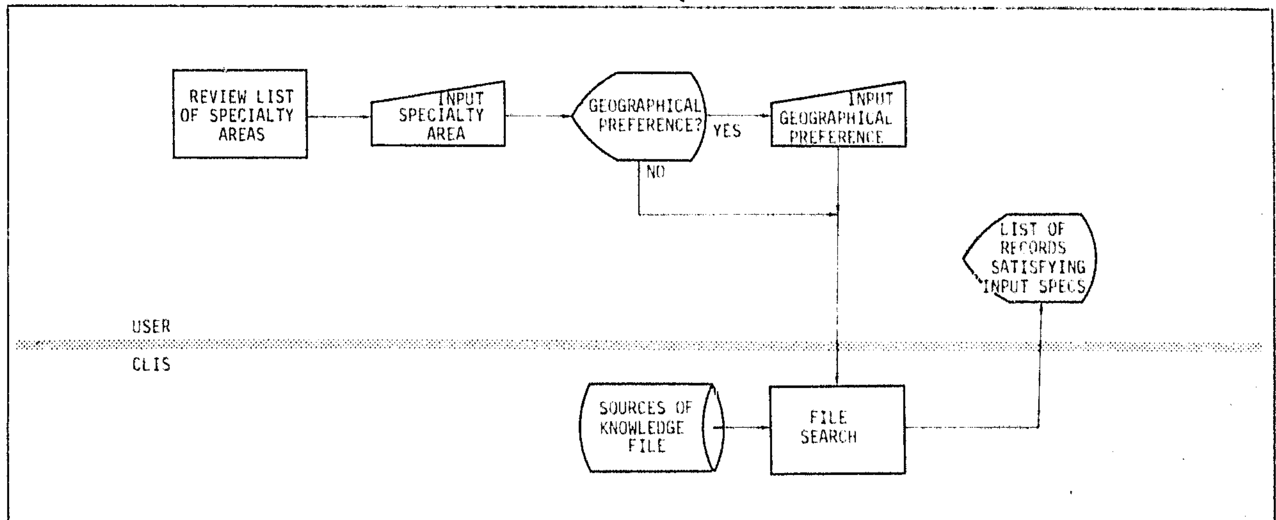


Figure 5 Information Flow — Sources of Specialized Knowledge

anticipated volume of data would be 1,500 characters per hour.

Since the mode of user interaction with the system is inquiry/response and the type of information to be provided is noncritical, the response time requirement can be expressed in minutes, such as less than five. This search function would have a low priority when contending for system resources.

Maintainability. After initial implementation, the requirement for file updating would be minimal. Additions to the file would have to be qualified by the same procedure under which the data base was generated. Purging of records would occur upon death, retirement, voluntary withdrawal or failure to maintain defined qualifications. Routine changes of address or telephone contacts would be handled in the same update cycle as additions and deletions. It is expected that each record will remain current for an extended period of time.

Initial Implementation. The functional design and detailed specifications describing the operation and implementation area will be relatively easy to develop as will the program coding and file allocation. The most critical problem will be defining the qualifications requirements of the people who will constitute the data base. A study sponsored by the National Institute of Law Enforcement and Criminal Justice is currently investigating the particular problem of definition of technical experts in the field of forensic science. The results of this study should serve as a base for determining qualification criteria for this data base file.

GENERAL DESIGN FOR SOURCES OF SPECIALIZED REAGENTS

Some reagents used in criminalistics laboratories are available only from obscure or difficult-to-locate sources. Some are rarely encountered by the laboratories because they are used in infrequently performed analyses. It may not be possible for a laboratory to do such an analysis in an urgent situation because there is no time to locate the needed reagent or reagents.

It has consequently been suggested that CLIS could serve as a store and exchange for information on sources for these reagents.

GENERAL OPERATING FUNCTIONS

Input/Output — Manual. A filing system would need to be developed. Request forms would have to be designed and protocols for their use distributed among the member laboratories. Subsequent input of a request form would cause CLIS to seek the information in its file, and if not found there, to generate a mailing to all laboratories seeking someone to meet the information request. Any response received by CLIS would be transmitted to the original requestor and entered in the main file. These contracts could be made by telephone.

Computerized. A computerized system would operate similarly, except that exchange of information would be quicker. Requests for information could be entered through a laboratory's terminal, and if not met by the main file, a call for help could go out to all terminals. Replies would be automatically entered in the main file and automatically transmitted to the other laboratories.

Figure 6 shows how information might flow in this affiliation.

DATA BASE REQUIREMENTS

File Content

- Reagent name
- Name of source
- Address of source
- Telephone number of source.

File Structure. Because of its relatively small size, there probably would be no impetus for making this file other than sequential. If it were an offline file, this would be enforced; if it were online, there might be reasons (such as continuity with other CLIS files) for making it a random file.

Record Layout and Element Coding Structures. Record length: 196 characters.

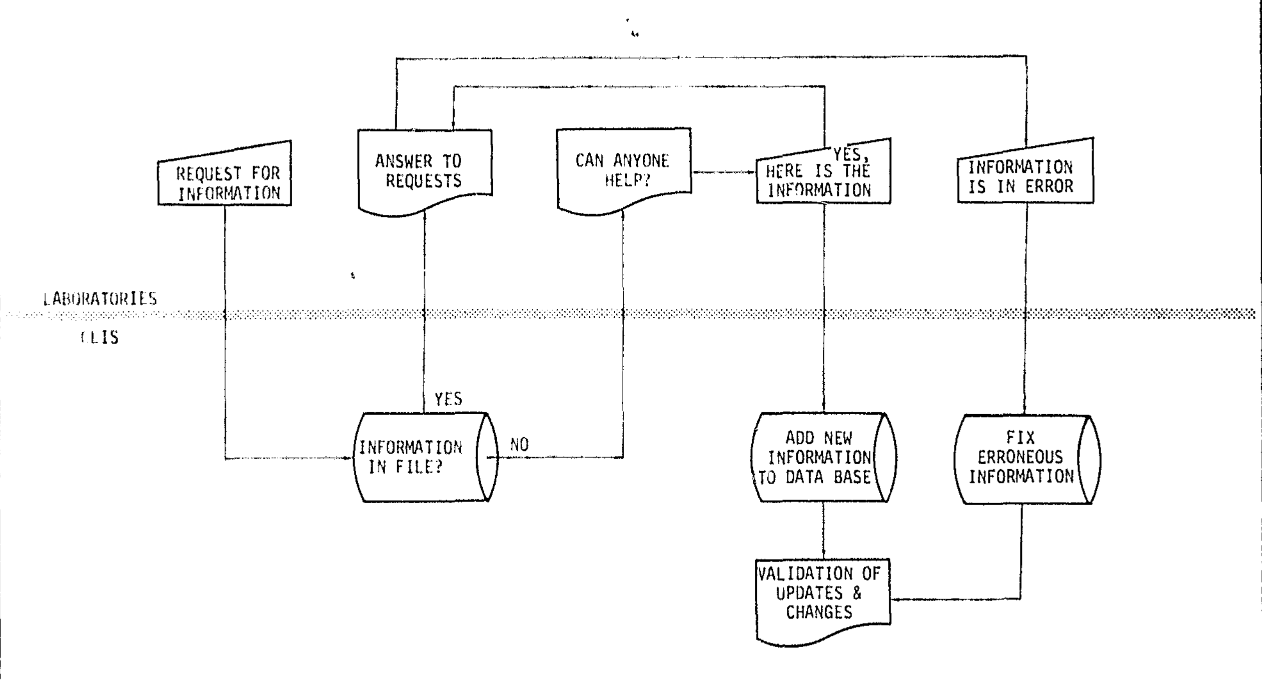


Figure 6 Information Flow — Sources of Specialized Reagents

Element Name	Number of Characters	Type of Characters
Entry Number	4	AN
Entry date	5	N
Status	1	B
Originator of entry	16	A
Reagent name	48	AN
Reagent purpose	48	A
Name of source	16	A
Address of source	48	AN
Telephone number of source	10	N

N = numeric
A = alphabetic
AN = alphanumeric
B = binary

Access Criteria and Indexing Methodology. No hierarchy or ordering is seen to be necessary for this file. The arrangement could be chronological with new entries added to the end.

Access would be by reagent name. Where reagents are known by more than one name, there could be multiple entries, or a cross-referencing system involving a synonym file could be established.

Estimated File Size and Growth Rate. Unless it were undertaken to include large lists of reagents, such as the Eastman Organic Chemicals Catalog, the file might contain a few hundred entries and might grow at the rate of 25 or so entries per year. A 500-entry file, with record sizes as estimated above, would require 98,000 bytes of storage.

PROCESSSING FUNCTIONS

Data Encoding and Allocation Schemes. Putting up the original data base, if any, in a manual system would involve only the conversion of known information to a standard format. Addition of new information would be done in the same way. In a computerized system the conversion process would require rendering the data into machine-readable form. New data could be entered at a terminal and the conversion done via program.

Since the content of this file would be largely alphanumeric, the only fields requiring encoding would be entry number, date, and status; and they would be computer-generated.

Although a fixed-length record is defined above for file estimation purposes, it would probably be more efficient in terms of storage space to allow the alphanumeric fields to be variable in length. The establishment and manipulation of these fields would be done by appropriate programming.

File Search and Match Techniques. Finding an entry in the file would be simply a matter of matching the requested reagent name against those on file. Either a manual or a computerized system would require a means for synonym checking in case the reagent might be filed under a name other than the one requested.

Applicability of a Functional Language. There appear to be only three types of user responses required:

- Request for a reagent source
- Contribution of a reagent source
- Notification that erroneous or out-of-date information has been distributed

It does not seem appropriate to develop a functional language for these few functions.

Concurrency of Operation. Concurrency is not a significant consideration in a manual system. If this application is as infrequently used as predicted, it is probably not significant in a computerized system either. However, it seems reasonable that a concurrency capability might be provided for the whole system in order to satisfy worst case needs (i.e., analytical instrument support).

COMMUNICATIONS

Estimated Usage Rates. In all probability, usage of this capability would be infrequent and unpredictable. A laboratory might go for months without having a need for special reagents and then want to request sources for a whole list of them at one time. At any rate, if implemented on a computer network the data communications usage would be so small as to probably not be justified except by being "piggy-backed" on a more frequently used application.

Necessary Response Times. This is an application that probably can be adequately performed by some laboratories themselves under routine conditions. Under emergency conditions a quickly responding system, presumably implemented on computer terminals, could make it possible for laboratories to conduct analyses they are now unable to do because they lack the time to locate the reagents. A one-day turnaround might be required to make this capability useful.

Required User Interaction. An inquiry/response facility, rather than true interaction, should be adequate for this application area. However, if implemented as part of an interactive terminal system, the inquiries and responses might themselves be interactive.

Anticipated Volume of Transmitted Data (Message Volume). The requests for information would consist mainly of the names of reagents whose source was being sought, each a 25 to 30-character message. The replies would most likely be comprised of entire file entries of about 200 characters. Added to this would be an overhead required to transmit request options, operator messages, formatting, etc., back and forth between computer and terminal.

MAINTAINABILITY

Concurrency of Data. Unless external sources of information can be found this subsystem should remain as current as the knowledge of its users.

Care should be taken to expunge from the file any sources found to be defunct or to no longer supply the reagent in question.

Sources of Updates for Maintenance. Both additions to and deletions from the file will probably generally emanate from the users. The mechanism for adding new entries to the file is described elsewhere in this section. It will also be desirable to keep channels open to learn of erroneous or outdated information and to provide a mechanism for reacting to such knowledge.

Summary of Maintenance Procedures

- Addition of user-supplied entries
- Monitoring by the CLIS staff of user-generated addition (with a transaction report if the system is computerized)
- Detection of, reaction to, outdated information

Initial Implementation. One can conceive of a service of this nature coming about in several ways:

- All the major reagent companies' catalogs could be codified with (very long) chemical names, suppliers' addresses, prices (hard to keep up to date), etc.
- The criminalistics laboratory community could be polled to determine what reagents they have had difficulty locating and where they eventually obtained them.
- The polling could be done as required; in that any laboratory having a special need could ask CLIS to spread the word among member laboratories. Any laboratory knowing how to fill the need would be referred to the original requestor. At the same time the information would be put on file in case the need arose again.

It is recommended that the last two approaches be used if this application is implemented. The survey of the laboratory community could supply the initial data base and the as-needed polling could provide updates. It is important that all exchange of information go through the central CLIS facility so that the information can be included in the data base.

SYSTEM REQUIREMENTS FOR SOURCES OF SPECIALIZED REAGENTS

Data Storage Requirements — *Manual.* All the data required for this application probably could be contained in a file drawer.

Computerized. In comparison with most other CLIS data bases, this one would not contain a large number of records, though the records might tend to be long due to the probable wordiness of the entries. A total of 100 kilobytes of storage should be adequate. If the application is to be available interactively at the terminals, an online (disc) data base is required, though batching schemes using offline storage are conceivable (if interactivity is forsworn).

Necessary Functions and Operations

- Original establishment of the data base
- User information request mechanism
- File search for requested information
- Means to notify all laboratories of information need
- Response mechanism to information need notification
- Addition of new data to data base
- Means to disseminate answers to information requests

Information Input and Output — *Manual.*

- Inputs: Information request forms and forms or telephone calls to replay to CLIS with information
- Outputs: Notification of information need forms or letters; forms or telephone calls to distribute answers to requestors.

Computerized

- Inputs: Entries on terminal keyboard by requestors and suppliers of information
- Outputs: Printouts on terminal printer or displays on terminal CRT of information requested or supplied.

System Availability and Response Times — *Manual.* There would be little limitation on availability by mail; telephone availability would be constrained by working hours at the sites involved. Response would generally be slow.

Computerized. A one-day turnaround should be adequate. Interactivity is not required; however, an online system would always be available. Response would be immediate for already stored information; for user-generated replies it would be unpredictable.

Estimated Usage Rates. Five to ten times per month.

Maintainability as It Affects System Design — *Manual.* A person or persons would be responsible for the usage and integrity of the files.

Computerized. The same sort of responsibility would be the case as for a manual system. The responsible person should be assisted by a transaction listing of any automatic updating of the file in order to eliminate or rectify any spurious or "garbage" entries.

GENERAL DESIGN FOR SOURCES OF STANDARD SAMPLES

Background

Data collected from potential user laboratories described in Volume I places this information need area third on the list of laboratory priorities. From this consensus ranking alone, one could assume that crime laboratories are now experiencing some serious problems relative to the acquisition and/or use of standard samples. An analysis of additional data tabulated from the survey information forms and discussions with laboratory personnel further substantiates this assumption.

Respondents were asked to indicate in Chart 2 of the information form the standard evidence collections maintained in their respective laboratories and the size and usage of these files. The information form defined standard evidence collections as reference files "which contain physical specimens (paint samples, firearms collections, fiber samples, etc.)." The above examples of standard samples could reasonably be extended to include:

- Drug dosage forms
- Safe insulation
- Human and other species of animal hair
- Ammunition
- Blood and other body fluids
- Paper
- Adhesives
- Inks
- Glass
- Tape
- Plastic
- Wire
- Explosives
- Wood

Use of such samples for comparative analysis varies from laboratory to laboratory. Depending upon the services provided, expertise available and cases accepted, individual laboratories will deal routinely with certain samples and infrequently with others. Primary sources for samples are the standard evidence collections maintained by laboratories. Review of the data from Chart 2 of the information form indicates some reasons why this was considered by many laboratories to be a relatively high priority item.

Table 4 presents an updated tabulation of the data collected from Chart 2 of the information form.

Table 4

Standard Evidence Collections

Type	Labs Reporting	Percent
Drug and Narcotic Dosage Forms	109	85
Hairs and Fibers	78	61
Firearms and Ammunition	75	59
Automobile Paints	49	38
Safe Insulation	28	22
Drug and Narcotics Ballistics	21	16
Miscellaneous:	85	66
Botanical	7	-
Typewriter	8	-
Explosives	10	-
Accelerants	7	-
Body Fluids	8	-
Wood	7	-
Minerals	4	-
Tobacco	3	-
Dyes	3	-
Tire Treads	2	-
Soils	3	-
Hydrocarbons	1	-
Alcohol	1	-
Paper	1	-
Plastics	2	-
Ink	2	-
Cosmetics	2	-
Gasoline	2	-
Carpeting	1	-
Feathers	1	-
Roofing	1	-
Metals	1	-
Laundry and Drycleaning	1	-
Glass	1	-
Pure Drugs	1	-
Fireworks	1	-
Headlight Filaments	1	-
Personal Protection Devices	1	-
Skeletal Remains	1	-
Palm Prints	1	-

Further analysis discloses that:

- One hundred and twenty-two (77 percent) of the laboratories reporting have at least one form of a standard evidence file.

- Thirty-seven (23 percent) laboratories do not have any form of standard evidence file at all.
- At least 50 percent of the 37 laboratories with no standard evidence files have less than ten total employees:

Size	Laboratories	Percent
1-10	21	57
11-20	6	16
21 +	3	8
Size not reported	7	19
	37	100

- Nineteen (12 percent) laboratories have only one of the standard evidence collections and at least one of the miscellaneous collections listed in Table 4.
- Twenty (13 percent) labs have two standard evidence collections.
- Twenty-eight (18 percent) labs have three standard evidence collections.
- Twenty-six (16 percent) labs have four standard evidence collections.
- Fifteen (9 percent) labs have five standard evidence collections.
- Five (3 percent) labs have six standard evidence collections.
- One (.6 percent) lab has all six of the common standard evidence collections and at least one of the miscellaneous collections listed in Table 4.

The average size and usage rates of the most common standard evidence collections presented in Table 4 are shown below:

Type of File	Average Size of File (Number of Standards)	Average Number of Times Used Per Month
Drug and Narcotic Dosage Forms	528	96
Hairs and Fibers	192	15
Firearms and Ammunition	870	40
Automobile Paints	985	9
Safe Insulation	53	15
Drug and Narcotics Ballistics	766	28
Miscellaneous	—	47

GENERAL CONCLUSIONS

Based upon the analysis of data from the survey information form, there does not appear to be a preponderance of standard evidence collections in crime laboratories throughout the country. Twenty-two percent of all laboratories responding have none at all. Almost 50 percent of the reporting laboratories have less than three standard evidence collections.

The size of a laboratory may affect its ability to acquire and maintain standard evidence collections. At least 50 percent of the laboratories without collections have less than ten total employees. If the size of all reporting laboratories was known, this figure could conceivably be increased to 75 percent. The average size of laboratories with less than three collections is 12 total employees. On the other hand, the average size of laboratories with three or more standard evidence collections is 21 total employees. The smaller the laboratory, therefore, the fewer the number of standard evidence collections.

Many laboratories do not even have the evidence collections one might expect to be found routinely in most laboratories. Table 4 lists the six most common files as reported by potential CLIS user laboratories. Three files, Drug and Narcotic Dosage Forms, Hairs and Fibers, and Firearms and Ammunition, were reported by more than half of the responding laboratories.

There are some serious questions about the currency of existing evidence collections. PMS interview teams attempted to obtain specific data relevant to the timeliness of the data in these files during visits to 17 laboratories. With the exception of information on the currency of automobile paint files, this data was not kept by the majority of the laboratories visited. The general, and almost unanimous, comments received from laboratory personnel, however, indicate that most evidence collection files are far from being current or complete.

The overall conclusion, then, is that standard evidence collections are, on the whole, scarce, incomplete and often outdated. The problem of obtaining a standard sample which is not available in a laboratory's standard evidence collection often becomes significant. The time factor is critical in two ways. First, the evidence analysis must be completed and the case returned to the contributor within a reasonable period of time. Second, the laboratory must often expend valuable analytical personnel resources to locate and obtain the sample desired.

CLIS can respond to this problem by identifying the sources of standard samples to user laboratories.

GENERAL OPERATING FUNCTIONS

CLIS could provide two basic services to users which address this general information need area. First, as indicated in Volume 1, a list of appropriate sources could be provided user laboratories seeking reliable standard samples for the purpose of completing nonroutine analyses. Requests would generally be for individual samples on an as-needed basis. To establish and maintain adequate sample files in-house can be costly and time-consuming, particularly to the smaller laboratories. This cost is increased significantly on a total crime laboratory population basis in relation to the number of times a particular sample file is duplicated around the country. It is anticipated that most requests of this type would be nonroutine in nature. Second, in those instances where it is practical for individual laboratories to have complete, up-to-date sample files in-house, a list of these sources could also be provided by CLIS. Much discretion should be exercised in this area, however, to avoid unnecessary duplication.

The primary objective of this CLIS component is to reduce the peripheral, nonanalytical activities of technical personnel and allow the time saved to be redirected to pure analytical efforts. The service performed to accomplish this objective is very simple and direct.

As an example, established laboratories could query the system to determine sources of standard samples needed for analytical comparisons but not included in in-house evidence collections. This decision can be made as soon as the evidence in question has been assigned either to the appropriate section or analyst. Even if a particular source is known to a laboratory, it may still be productive to search the CLIS file. A laboratory may be aware of one or several possible sources, but probably not all possibilities. A search request may disclose a source (or sources) which (a) is closer to the laboratory and would speed delivery, (b) is more specific to the need at hand, (c) is of higher quality than the old resource used by the laboratory, (d) is less expensive than the source normally used, or (e) can be acquired in quantities which suit either current or projected needs. Of course, if the laboratory has no knowledge of any source, then CLIS could provide a comprehensive listing of possibilities.

New laboratories could be spared the tedious task of developing such resources by being provided the capability of identifying sources of standard samples as the need arises.

Both new and established laboratories wishing to establish select in-house files could access CLIS for this information also.

A consideration which will be discussed in some detail later, is the need to ensure the data in the CLIS file represents a legitimate, reliable source which can deliver the proper sample within a reasonable period of time. This will be a continuous concern of system maintenance.

With one exception the inputs and outputs of this system component are similar to those described in the section dealing with analytical/identification support. The exception is that instrumentation is not directly involved and, therefore, would not be "online" for the purpose of satisfying this informational need area.

A cathode-ray tube (CRT) terminal with a typewriter keyboard and visual display screen could effectively be used. Through the keyboard an operator/technician could access the system by following the query format displayed on the screen and recording the response returned by the system. The request might be for the location of a sample within a certain geographic area, within a certain price range, and within a stated variance of specific sample specifications. A list of possibilities conforming to these parameters would then be displayed on the CRT.

Inputs and outputs of this component could also be processed with a teletype terminal. Inputs would be entered by means of a keyboard, but outputs typed on paper at the user's terminal.

Remote batch processing is an alternative which would require three pieces of peripheral equipment, and the requests for sources of samples would be so infrequent as to make "batching" (stockpiling requests and processing several at one time) an inefficient procedure in light of the length of time it may take to accumulate requests. Input could be either punched card or paper tape, and output would be computer printout.

This component of CLIS could also be operated in an entirely manual mode. Individual requests could be sent by user laboratories through the mail to the central CLIS facility. The reply would also be returned by mail after being processed by the central facility. Input would be a standard data request form and output would be either a computer printout or a standard data reply form.

The data elements of the standard sample source file could also be prepared in 8½x11 looseleaf or 3x5 card formats, and a copy distributed to all user laboratories. The laboratories would then perform their own search exercises with in-house, hard-copy files. If the file is small enough (several hundred entries), this approach may be practical. However, if the file is substantial (several thousand entries), the search and maintenance responsibilities of user laboratories may be too great for effective system usage. The purpose of CLIS is to relieve laboratories of some of their time and personnel burdens, not add to them.

DATA BASE REQUIREMENTS

File Content. The contents of the file which would be of specific interest to the user could include the following:

- Sample type
- Sample name
- Sample specifications
- Quantity of sample available
- Name of source
- Address of source
- Telephone number of source
- Price estimate

File Structure. If this file is ultimately implemented as a manual operation, its structure would of necessity be sequential with record entries in order according to the indexing criteria established. It would probably be sequenced alphabetically by the name of standard samples. If it were an automated file, it could also be structured sequentially, but it is anticipated that the major files of CLIS will be structured to allow random access. In the interest of system continuity, therefore, this file would probably be made a random file also.

Record Layout and Element Coding Structure.

Estimated record length: 196 bytes.

Element Name	Number of Characters	Type of Characters
Entry number	4	AN
Entry date	5	N
Status	1	B
Originator of entry	16	A
Date of last update	6	N
Originator of update	16	A
Type of sample (Code)	2	N
Specific sample name	48	AN
Name of source	16	AN
Address of source	48	AN
Telephone number of source	10	N
Availability time	4	AN
Estimated price of sample	6	AN
Date of last price update	6	N
Quantity available	8	AN

Access Criteria and Indexing Methodology. Indexing methodology could involve a two-level hierarchy: (1) the major file organization would be by type of sample (wood, glass, paint, etc.); (2) within each major organization, elements would be identified by the specific names of samples.

The file could be accessed in several ways:

- By general type only
- By specific sample name
- By source of sample

Other definitive accessing parameters could be combined with the above in any combination:

- Estimated time needed by source to process request
- Geographic location
- Currency of entry
- Quantity available
- Estimated price

Multiple entries may be required when general types or specific samples are known by several names. A cross-reference table for synonyms could also be utilized.

Estimated File Size and Growth Rate. If, as an example, the initial data base attained sources of samples suggested in Figure 8, it would include approximately 17,500 records. Since each record is estimated to include 196 bytes, the initial file requirements would be for 3.4 million bytes. It is not unreasonable to expect this file to increase by 150 percent over a two-year period — 100 percent at the end of the first year's operation and 50 percent at the end of the second year. This estimate is based upon the considerations discussed in the Initial Implementation section for this application area. Storage requirements at that time would be for 8.5 million bytes.

PROCESSING FUNCTIONS

Data Encoding and Allocation Schemes. Establishing a data base for this application area in either an automated or manual mode would be relatively simple once the data has been compiled for conversion. The file data is not overly sophisticated in nature and consists primarily of names, addresses, dates and telephone numbers. Most data elements, therefore, could be entered "as is" directly into the file. In a computerized file this data would be primarily alphanumeric. The only data elements requiring encoding (computer, not user-generated) would be the entry number, dates and status, type of sample, and geographic location. The encoding of these fields would facilitate certain of the processing functions. Any data in the encoded fields would automatically be converted to English before being included in outputs to users.

As mentioned earlier, it may be advantageous for the user to initially limit his search request to a certain geographic location. This parameter could be expanded depending upon the quality and quantity of hits. A table of users in the program could allow the system to determine, by code, first the state in which the user is located, and second, the surrounding states in the same general geographic area within which possible sources should be identified. It would also be possible for the systems, using encoded data, to prioritize the source possibilities by the average distance in miles between the state in which the user is located and the state in which the possible source is located.

File Search and Match Techniques. The primary search consideration for this file would involve a match of the requested standard sample name with all those of the same general type in the file. In this case all possible sources of the sample would be returned to the user. The user could also have the flexibility of selectively reducing his "hit" list by applying other conditions to the search request. For example:

- Sources of animal hair
- Located within a five-state area (specify)
- Which can be obtained within three days
- Whose fee will be less than $20.00 per sample

As mentioned earlier, either a manual or automated system must provide synonym checks in the event the sample might be filed under a name different than that requested.

Applicability of a Functional Language. The access modes and search techniques envisioned for this file and described earlier are relatively basic and unsophisticated. The demands upon users to retrieve data from this file, therefore, will not be complicated. The user will only be performing three basic functions: (1) requesting sources of samples, (2) contributing possible entries to the file, (3) notifying the system of information which is determined to be in error or out-of-date. These types of user interaction with the system would not require the development and use of a functional language.

Concurrency of Operation. This application area is not expected to generate a high message volume or usage rate. Therefore, operational concurrency of processing functions in either an automated or manual system will not present user/technical problems or delays. It will probably be a unique circumstance when CLIS must queue inquiries to this file. Ordinarily, a request could be entered, processed and the data returned before another request is received for entry to the file. This application area combined with other components of CLIS in a total system, however, will probably warrant the integration of concurrency considerations at the total system level.

Figure 7 shows the processing functions for this application.

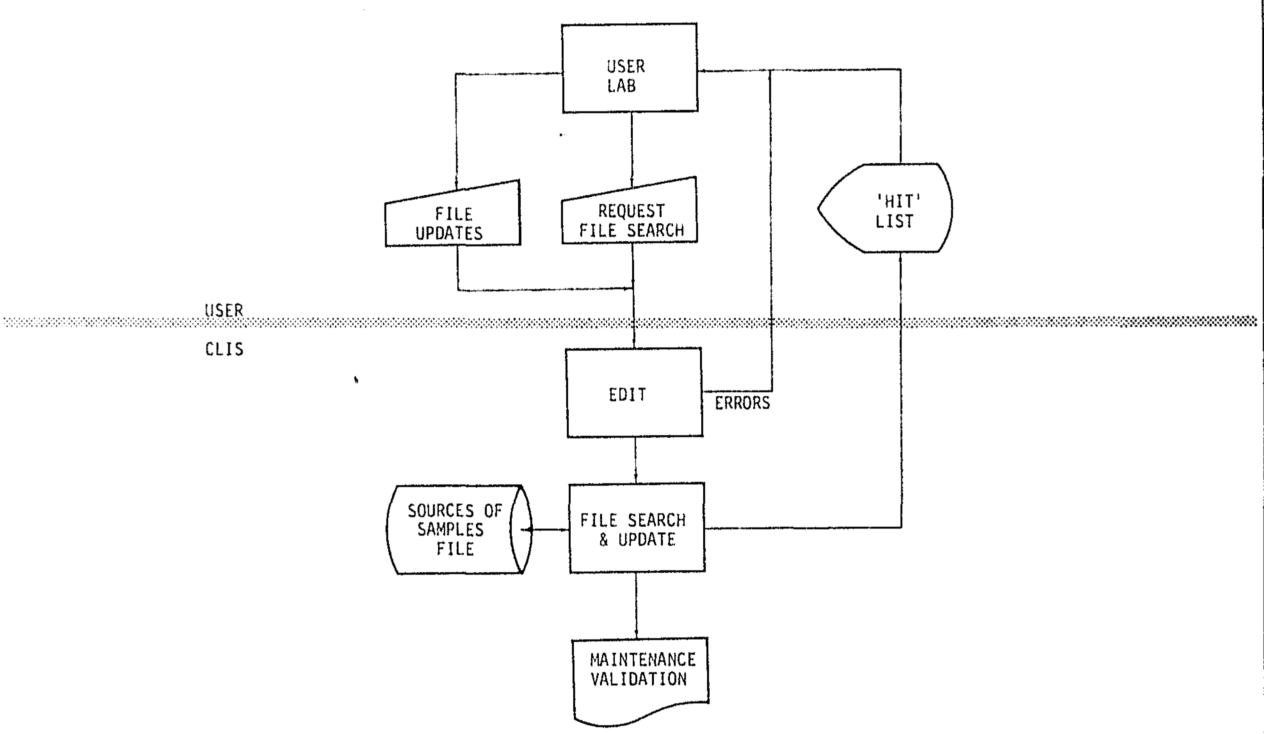


Figure 7 Information Flow — Sources of Standard Samples

COMMUNICATIONS

Estimated Usage Rates. Usage rates for this type of application area are impossible to predict with any reasonable certainty. By reviewing historical case load and activity data, most laboratories can project the volume of future work by type. This is done with reasonable certainty from available data. Most laboratories, at best, can only guess at the number of times sources of standard samples have been needed or will be needed. Records of this nature are not routinely kept by laboratories. Occasions for this type of information will be both infrequent and unpredictable. Even if reasonably accurate rates could be developed, it is almost certain that the usage of this file would be too small to justify a "stand alone" system. This file must be an important, though small, component of the larger CLIS network in order to support its automation.

Necessary Response Times. Many laboratories are currently using in-house records, usually kept on an informal basis, to locate possible sources of samples needed for comparative analyses. These lists are certainly not comprehensive, and it may take a considerable amount of time to locate a source or sources which the laboratory had never used previously. In the absence of immediately available knowledge, it is not unreasonable to expect a laboratory to take from several hours to several days or more to locate a particular sample. Crime laboratories cannot really afford to expend personnel resources for this kind of activity.

A manual system using the mail could probably provide a response to the user within 72 to 86 hours. A manual system using the telephone could probably reduce time to 24 to 48 hours. An automated system, depending upon the priorities assigned to the transactions of this and other CLIS application areas, could provide a response within several minutes and certainly several hours at the most. This is considerably better than laboratories are able to do now. The point to remember, however, is that laboratories will be getting far more comprehensive and accurate data within this time frame than they could probably develop on their own.

Required User Interaction. User interaction with the system is not a requirement for this application area. The user will know his search options and can select the most appropriate one for his request. Responses will generally satisfy the inquiry. If not, another search option can be selected and another inquiry made to the system. The search variations and file elements are not that sophisticated and should not require the user and system to "talk" to each other in order to get desired information.

Anticipated Volume of Transmitted Data. A request for information will consist primarily of the names of samples desired and require probably less than 50 bytes per message. The response would consist of the majority of the data in file for a particular source, or approximately 300 bytes per message. A 75 percent overhead required to transmit various search options, user messages and messages from others would probably be required.

MAINTAINABILITY

Currency of Data. Formal maintenance procedures should be established in accordance with the size of the file and number of individual sources on file. At least once every six months data entries for each source should be verified and the file updated accordingly. On-going maintenance will be the responsibility of users and will be performed routinely as errors are detected.

Sources of Updates for Maintenance. The sources of samples should also be a primary source of updates. They will be requested periodically to verify file data and provide current information if necessary. During the course of their use of this file, users will probably be the first to discover the need for file updates; i.e., wrong addresses or telephone numbers, change in sample specifications, price increases, samples no longer available, etc. Update information obtained by the user should be forwarded to CLIS for maintenance processing. This process would also include the identification of new sources. The currency of the file will be in direct relation to the interest and conscientiousness of the users.

Summary of Maintenance Procedures

- Periodic verification of file data
- Addition of entries supplied by users
- Verification of additions submitted by users
- Detection, submission, verification and updating of outdated information (generated by users).

INITIAL IMPLEMENTATION

An early task during this phase will be to finalize and define in detail the files, system processing functions, and procedures to be followed by users. Specific inquiry and response formats, editing criteria, error designations and search operations will be included. Following this, detailed programming and file allocation specifications can be prepared. Programs, systems and subsystems can then be completed and tested against specifications and desired outputs.

Developing the data base will probably be the single most important and difficult activity performed during initial implementation. The first problem will be to identify potential sources of standard samples. The second problem will be to determine the quality of the sources and their willingness to cooperate in the provisions of this service through CLIS.

Several studies dealing with standard reference materials or standard reference files have been completed recently. Perhaps the one most pertinent to the topic at hand is a survey conducted by the Harold L. Steinberg, Technical Analysis Division, National Bureau of Standards (NBS) for the Law Enforcement Standards Laboratory (LESL). The results of the study have been compiled in a draft document titled *"Standard Reference Collections of Forensic Science Materials: Status and Needs."* Mr. Robert Mills, LESL, and Mr. Steinberg, were most cooperative in providing a copy of this draft document for review and also taking time from their busy schedules to discuss their work in relation to CLIS. This report has not been corrected for final publication and public dissemination; however, permission has been given to quote from the draft with the understanding that some findings, conclusions and recommendations may be changed as a result of final updating.

From interview and questionnaire data, NBS identified approximately 60 Forensic Science Materials (FSM) having Standard Reference File potential. Figure 8 shows the initial NBS list. Respondents were asked to select those FSM's which satisfied their basic needs, and the original list was refined as shown in Figure 9. After further analysis, it was determined that 14 FSM's had significant potential for SRF development, and five FSM's required additional study. Both groups are presented in Figure 10.

It would be reasonable to assume that the standard samples represented in Figure 10 be the nucleus of the initial data base. Within a year of initial implementation, the data base could be expanded to include all the samples represented in Figure 9. Within two years of initial implementation the data base could take on its final form by including samples in all the classes of FSM's represented in Figure 8. Sources of file data for initial implementation, therefore, could include the following (fingerprints and firearms excluded) identified in the NBS Study:

FSM	Source
Drugs (abused, controlled, toxic)	DEA USP Pharmaceutical (manufacturers) NCDC
Solid Dosage Drugs	Manufacturers (approximately 1,250)
Alcohol	NBS
Hair (animal, feathers)	McCrone Institute Zoos Veterinary schools Schools of taxidermy Furriers
Blood (human)	American Red Cross American Association of Blood Banks Hospitals
Typewriting	USPS Zurich Police Department FBI New York Police Department
Fibers (plant)	New York Botanical Museum Harvard University U.S. Customs Paper Institute
Fibers (synthetic)	Test Fabrics, Inc. N.Y. Customs Laboratory McCrone Institute HOCRE, Great Britain Manufacturers
Glass (auto)	HOCRE, Great Britain Manufacturers Wagner Westinghouse Guide Lamp General Electric
Paint (auto)	Automobile manufacturers Paint manufacturers
Tire (tread, patterns)	Smithers Scientific Services Tire Guide (publication) Tread Design Guide (publication) Parking Dimension of Automobiles (publication) Manufacturers (approximately 25)
Explosives	ATF

DEFERRED APPLICATION AREAS

Compilation of Statistics to Determine Specimen Uniqueness

A general design of this application area is beyond the scope of this report. However, there are some ways in which CLIS would be responsive to the information handling needs and communication requirements that could be expected to evolve from such a general design. It is doubtful that such a design could be implemented before CLIS is operational.

Assume that a general design would define the following functions:

1. The selection of specific types of specimens.
2. The definition of information — both technical and demographic — for the purpose of categorizing all required characteristics of each type of specimen.
3. The definition of formal, data-gathering procedures to include required geographic distribution of each specimen.

Figure 8
List of FSM Classes Having SRF Potential

#	FSM CLASS OR FORM	#	FSM CLASS OR FORM
1.	Poisons	30.	Wood (species), Bark
2.	Drugs, Abused	31.	Soils
3.	Drugs, Toxic	32.	Wire, Cable
4.	Pill Ballistics	33.	Nails, Screws, Bolts
5.	Tobacco Products	34.	Glass, Auto
6.	Alcohol, - ic Beverages	35.	Glass, Non-Auto
7.	Lipsticks	36.	Leather, Shoe
8.	Hair Cream, Grease, Spray	37.	Shoe-, Sneaker-Prints
9.	Perfume	38.	Plastics, Synthetic Fibers, Cloths
10.	Cosmetics, Other	39.	Natural Fibers, Cloths
11.	Hair, Human	40.	Laundry Marks (visible/invisible), Dyes
12.	Hair, Animal	41.	Tire Tread Patterns
13.	Blood, Human	42.	Tire Tread Composition
14.	Blood, Animal	43.	Auto Silhouette Patterns
15.	Body Fluids, Human	44.	Paint, Auto
16.	Prints (finger-, palm-, etc.)	45.	Paint, Non-Auto
17.	Wound Ballistics	46.	Road Construction Materials
18.	Paper, Writing	47.	Ammunition
19.	Watermarks, Paper	48.	Explosives
20.	Pens, Markers, Inks	49.	Weapons, Firearms
21.	Pencils, Crayons	50.	Bullet Ballistics
22.	Typeface, Typewriters	51.	Fire Accelerants
23.	Typewriter Ribbons	52.	Safe Construction Materials
24.	Transfer Letters, Dry	53.	Known Criminal Characteristics
25.	Voiceprints	54.	Fraudulent Check, Handwriting
26.	Glues, Adhesives	55.	Pipe, Tubing
27.	Tape (black, "Scotch", etc.)	56.	Shoe Polish
28.	String, Rope	57.	Jewelry Markings
29.	Knots		

Figure 9

FSM Classes for Further Consideration

FORENSIC SCIENCE MATERIAL CLASS

1. Drugs (abused, toxicological)
2. Solid Dosage Drugs
3. Tobacco
4. Alcohol
5. Cosmetics
6. Hair, Human
7. Hair, Animal (and feathers)
8. Blood, Human
9. Blood, Animal
10. Body Fluids, Human, Animal (excluding blood)
11. Prints (finger-, palm-, etc.)
12. Voice Prints
13. Paper, Writing
14. Watermarks (paper)
15. Typewriters, Typewriting, Ribbons
16. Writing Inks, Markers
17. Dyes, Stains, Pigments
18. Fibers, Synthetic
19. Fibers, Plant
20. Woods, Barks
21. Adhesives
22. Soils
23. Glass, General
24. Glass, Auto (and plastic lenses)
25. Paints, General
26. Paints, Auto
27. Shoe, Sneaker Prints
28. Tire Tread Patterns
29. Tire Tread Composition
30. Safe Insulation
31. Explosives
32. Firearms, Weapons
33. Ammunition, Residues
34. Accelerants

Figure 10

FSM Classes Recommended For Development	FSM Classes Requiring Further Study
Tire Tread Illustrations	Adhesives
Solid Dosage Drugs	Body Fluids
Glass, Auto	Prints, Voice
Firearms	Safe Insulation
Blood, Human	Tobacco
Alcohol	
Paints, Auto	
Drugs	
Typewriting	
Fibers, Synthetic	
Prints, Finger	
Explosives	
Hair, Animal and Feathers	
Fibers, Plant	

4. The selection of laboratories or contractors (private or public) to perform the data gathering for each specimen.
5. The accumulation of specimen data in a central repository.
6. The reduction and statistical representative of the data and its dissemination to criminalistics laboratories.

CLIS could respond to these requirements in a number of ways. Since it will have established telecommunications net that would connect all major laboratories, it would be eminently suitable for gathering data and disseminating reports. If necessary, temporary communication links could be set up between CLIS and the data-gathering sites. CLIS processor capability would provide input editing and formatting functions in order to maintain strict procedural control in the data-gathering mode.

As the data passed through edit checking, it would be accumulated on CLIS data storage media in separate files. Compilation of statistical information could be accomplished on CLIS processors as low priority background work or in nonprime time.

Thus, since CLIS would already be serving many of the needs of the forensic laboratories, it is appropriate that a service which enable laboratories to determine specimen uniqueness be implemented on an operational system.

COMPUTATION CAPABILITY AND EXPLOSIVE TAGGING

The two potential application areas which were assigned the lowest priorities by laboratories responding to the information form are not being considered during the conceptual design state. Depending upon the results of studies currently underway, one of these areas, explosive tagging, may possibly be considered for inclusion in CLIS at some later date. The other area, providing users access to computational and statistical routines, does not appear to hold promise as a CLIS application area. Computation capability had an average priority ranking of 6.30 on a 1-10 scale, and explosive tagging had an average priority of 6.91. Sixty-three percent of the respondents assigned a priority of over "5" to computation capability and four percent assigned a priority of over "5" to explosive tagging. Thirty percent of the responding laboratories assigned a priority of either "9" or "10" to both of these areas.

CHAPTER 4. SUMMARY OF SYSTEM REQUIREMENTS

This chapter summarizes the requirements of the application areas that have been previously defined. The general designs developed in Chapter 3 are intended to portray an overview of the functional operation and system resource requirements that would be necessary to implement each application area. However, to be effective and responsive to the laboratory community, all of these areas should be integrated into a single CLIS system, under a central organization, which would exercise management control of day-to-day operations.

Table 5 summarizes the estimated demands that each application area would make of system resources. It is readily apparent that the number one priority area, analytical/ID support, requires far more system resources than the other application areas combined. Justification for implementation of CLIS will rest heavily upon the justification of this one particular area.

DATA STORAGE

The implementation of online data files for mass spectrophotometers will require a considerable amount of data storage. Prior to the implementation of this file, the data storage requirements will be quite minimal. The total requirement of 400 million characters is well within the scope of present equipment, and depending upon hardware specifications, would require only four or five online direct access spindles. The estimated growth rate indicates that there will be a 50 percent increase in data storage requirements in the first five years. Estimates beyond this time span would be purely conjectural, as the advance of technology will continue to impact the methodology and analytic techniques or forensic laboratories, possibly producing a substantial change in the laboratory environment and instrumentation.

Table 5 System Requirements By Application Area

APPLICATION AREA		Data Storage in Characters		Communications* Characters/Day	File Maintenance Record Updates	Processing Capabilities
		Initial	Growth Per Year			
Analytical/ID Support	Small IR File	4.0 M	60.0 K	1.4 M	Annually and as available	On-Line Files Interactive Access Extensive File Searching and Matching
	All Known Data Bases	385.0 M	50.0 M	5.4 M		
	All Data Except MS	31.0 M	2.6 M	2.7 M		
Rifling Specifications		2.3 M	123.0 K	400.0 K	Weekly	Interactive Access
Bibliography and Abstracts		9.0 M	1.8 M	600.0 K	Bi-weekly or as Citations are Developed	Language Interpreter On-line Tutorial
Sources of Standard Samples		3.4 M	2.0 M Avg. 2 yrs.	3.5 K	Semi-annual, As available	Inquiry/Response
Sources of Specialized Knowledge		1.6 M	----	12.0 K	Minimal	Inquiry/Response
Sources of Specialized Reagents		100.0 K	5.0 K	500	As available	Inquiry/Response
Summation		401.4 M	53.9 M	6.4 M		

* Ten-fold usage increase after implementation

M=Million
K=Thousand

COMMUNICATIONS

The estimates made for character traffic were generally based upon daily usage. To break these figures down to an hourly rate would depend largely upon laboratory working hours and system availability. Assuming a worst case of 6.4 million characters per eight-hour day, the bit rate requirement of any central node of the communications net would average approximately 2,000 baud. This is fully within the range of present-day technology; however, this is an average figure and peak loading could easily increase this requirement by an order of magnitude, but still within the capabilities of the technology.

Equally important will be the structure of CLIS and the organization of the communications net. It must effectively connect each user laboratory to the system and be configured to provide high capacity data paths were the volume warrants. Each laboratory must have easy access to CLIS without having to use elaborate password schemes or other sophisticated communications protocol.

FILE MAINTENANCE

Application files may be modified for a number of reasons: addition of new data, modification of data to correct errors and deletion of data which has met specified purge criteria. Depending upon the types of file management systems used, maintenance may have to be performed periodically to prevent overuse of file overflow areas and the subsequent increase in file search time. Should the CLIS configuration be organized around a central processor complex, file maintenance must be carefully scheduled so as to maximize system up-time. Errors in file data should be corrected on a daily basis, while full file updates scheduled as required on a periodic or "as the data accumulates" basis.

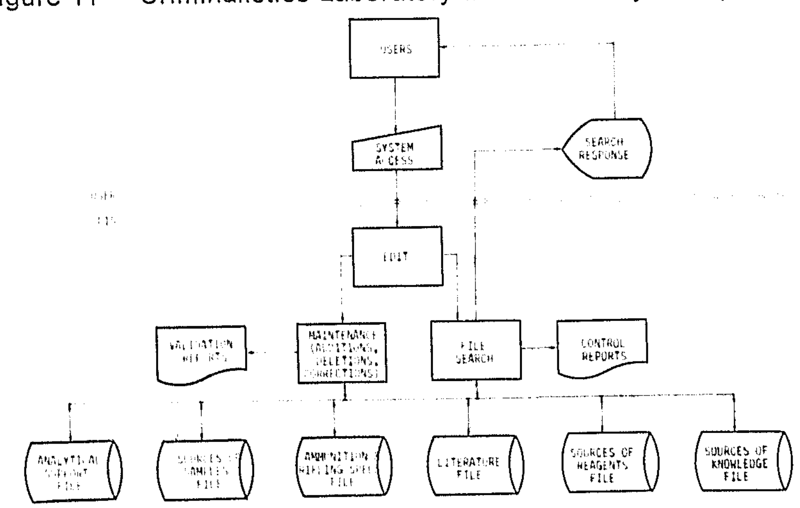
PROCESSING CAPABILITY

Processing functions will largely be the responsibility of the real-time teleprocessing monitor(s) and its subordinate application processors (computer programs). It is apparent that a single application processor could be used for the three "inquiry/response" applications while specialized applications processors would be required for the bibliography, rifling and analytical support applications. The rifling application processor would be relatively simple, and the bibliographic processor could be generated by slight modifications to currently available systems. The processing functions of the analytical processor will require a substantial effort to develop and implement on a nationwide, all-encompassing basis.

All data input will be edited as much as possible prior to being passed to an application processor so as to maximize concurrency of simultaneous operations. It will be the responsibility of the teleprocessing monitor(s) to efficiently schedule operations to be performed and allocate system resources.

Figure 11 shows the processing functions of CLIS.

Figure 11 Criminalistics Laboratory Information System (CLIS)



CLIS OPERATION AND MAINTENANCE FUNCTIONS

In order to present as general as possible an overview, this discussion of system operation somewhat anticipates future investigation of system design and staffing. Most of the discussion is applicable to either a manual or a computerized system; the philosophies expressed are thought to be applicable whether the eventual solution is a centralized computer (or computers) that communicates with its users via terminals; a computer center that communicates with its users by mail, or no computer at all.

It does seem that a noncomputerized CLIS would be adding little to already available facilities in certain application areas. For example, commercial computerized systems exist for the identification of unknown spectra of the major types. The advantage CLIS could bring to this area would be that of supporting several analytical techniques in a cogent, "single-source" manner. Analytical support comprises a sort of "worst case" component of the CLIS applications because satisfactorily meeting the requirement involves more sophisticated system functions such as the following:

- The ability to provide simultaneous access for multiple users
- Online files
- Random-access files

However, given that this level of sophistication must be supported, it becomes possible to "piggyback" other applications. An information system for sources of standard samples and specialized reagents and expertise could be given the same degree of accessibility even though the necessary facilities could not be justified for these applications.

The eventual justification for CLIS should center around its potential benefits to users rather than the number of times the system is used (usage rates) each day, week or month. If CLIS is implemented to include the application areas discussed in this report, it will provide some services to (and increase the capabilities of) laboratories which are not currently available. Some of the functions can be done manually by laboratories now, but they consume an inordinate amount of personnel resources; i.e., locating sources of samples, researching literature for specific topics, identifying an unknown spectra. Time savings to user laboratory personnel then become a significant justification for CLIS.

1. Basic Concepts and Functions

Responsibility. Systems such as CLIS will not run themselves. Decision-making responsibility may reside in the user community at large or in a central authority. As will be pointed out presently, it would seem that both should play a part.

Data Base. Two kinds are involved. There are existing data bases, such as the ASTM spectral files, and there will be user-supplied information, such as in-house spectral files. Data in some application areas (e.g., sources of specialized expertise) may be available only from the users.

Updates. Similarly, updates may be in the form of standard updates to standard files, or they may be user-initiated; for example, the contribution of a new standard spectrum or the correction of the address of a reagent supplier that has been noted to be in error. There will be general and interapplication functions — clerical in any case, and development and maintenance programming in a computerized system — which would need to be handled by a centralized authority. This authority could be in the form of a CLIS staff or of contract or consultant services.

2. Design and Organization Philosophies

The Philosophy of a User-Driven System. A system could be designed that would be useful to the working criminalist in that it would provide him, on request, with information not available to him in his own laboratory. However, in such a system his participation, in terms of contributing to, correcting, or browsing among the data bases, would be severely limited or highly formalized. Such a system might be "interactive" but still not be truly reactive. It undoubtedly would be more satisfactory to him if he had some measure of control over the information he utilized.

Because of such considerations, convenient and informal user access for contribution and correction is recommended. User control is especially pertinent in the case of in-house analytical suport files added to the shared data base. Even though they are part of a central file, a user would understandably want to retain the option of having his own file, then, should be less formal than for correcting the standard data base or the files contributed by other laboratories.

The Philosophy of a Central Overseer. This philosophy does not need to conflict with the previous one. Even in a user-driven system some authority needs to have the final responsibility for the integrity of the data base in each application area. The overseer could be part of the central CLIS staff (if there are any), or the function could be distributed among the member laboratories. The overseer could be an individual or a committee, perhaps depending on the importance and complexity of the application. (If the authority is decentralized, there must still be responsibility lodged at the CLIS data base site to coordinate the interaction of all the files and to ensure that file transactions stipulated by the overseer were acted upon.)

The overseeing individual's or committee's duties might include the following:

- Decisions as to what standard files to include in the data base
- Advice on the nature of and improvements to search algorithms and strategies
- Knowledge of sources and the acquisition of updates to the standard files
- Notification to the central facility of erroneous or duplicate entries to be corrected or removed, and following up the notifications to make certain the required actions were taken
- Decisions to remove little-used information should file space be at a premium
- Supervision of the contribution of in-house files from members laboratories; the overseers should review the content of these files and make recommendations as appropriate
- Establishment of protocol for interaction with member contributors; this would include procedures for the updates which would arrive as batches, individual entries, or corrections to existing entries.

To assist the overseers, the system, whether manual or computerized, should generate an entry in a periodically distributed document for each application area for every transaction modifying that area's file.

CHAPTER 5. DEVELOP DESIGN ALTERNATIVES

As can be seen from the intricacy and variety of its component application areas, CLIS will be a complex system. Its users will be separated not only by geography, but also by size, work load and in some by functional responsibility. The system must be responsive over a wide range of system usage and heterogeneous processing and data storage requirements.

Any CLIS system that may be implemented to meet the criminalistic laboratory information requirements that have been identified will take the form of a large-scale data base storage and retrieval system. It might be described as a "Laboratory Technical Information System" (as opposed to a laboratory management information system, which would address such tasks as evidence chain reporting and case loading for management control). The next few paragraphs will describe several CLIS alternatives from the macro system level. At this level five major functions that pertain to CLIS have been identified: the user population, the communications net, the processor function, the data storage function and the file maintenance function.

The *user population* includes all the laboratories and individual criminalists to whom CLIS might render some service.

Data Storage is taken to mean the sum of the data bases for all CLIS applications and any hardware required to store them.

File maintenance is the updating of a data storage to maintain file accuracy and data currency. This includes any control mechanisms (user committees, CLIS staff, referee labs, etc.) established to ensure the constant integrity of central file data.

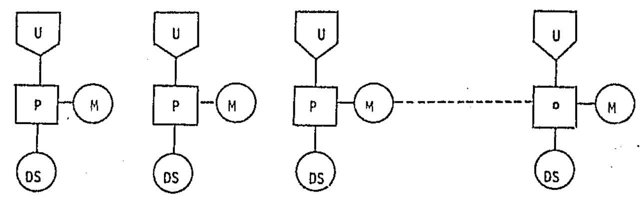
Processor function is the central computer or computers and the programming necessary to give the user population access to the data storage. Although indicated as separate functions, the processor and data storage devices should be included in a single complex due to the high rate and volume of information transfer between them.

The *communications network* which ties the other functions together requires a set of communication lines dedicated to the purpose of moving data between centralized processors and remote sites. A communications network could be specially assembled for CLIS; CLIS could utilize an existing government network such as the National Crime Information Center (NCIC) or the National Law Enforcement Telecommunications System (NLETS); or it could use a commercial system, such as the General Electric Mark III Computer Service or TYMNET.

The following symbolism is used in depicting the various alternatives of CLIS:

U = User Laboratory
NET = Communications Network and Control
P = Processor
DS = Data Storage
M = File Maintenance
— = Communications Connection

User Independent CLIS



This configuration provides each user with a complete processing and data storage capability. There is no direct communication with a central system nor other laboratories.

Advantages

— User Selectivity. The user will not only use those application areas which are most important to his operations.

— Lack of a complex communications net and its associated costs.

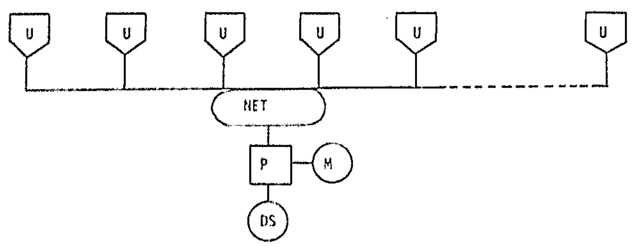
— User Control. Many computer users prefer not to be dependent upon a system controlled by an external agency.

Disadvantages

— File maintenance would be extremely difficult. Distribution of updates would be made by mail and no means to force the user to use the most current data.

— Lack of central control.
— Complete redundancy of data storage.
— The costs of duplicated hardware, software, and data bases.

Centralized CLIS

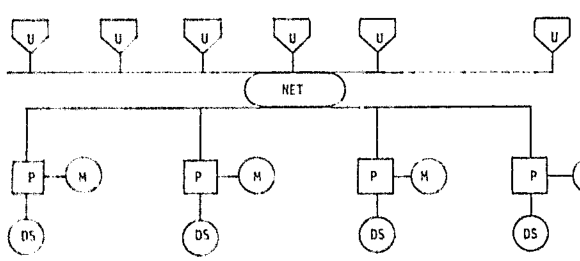


This is perhaps the simplest of CLIS configurations. All users would interface their terminals directly with a single-processor complex. This configuration also provides the greatest degree of centralized control.

Advantages
— Simplified system structure
— Centralized control.
— File maintenance can be easily coordinated.

Disadvantages
— Each application area may not be optimally configured for its specialized requirements.
— Users might compete for system resources during peak usage periods and may experience some delay in response time.

Distributed Processor CLIS

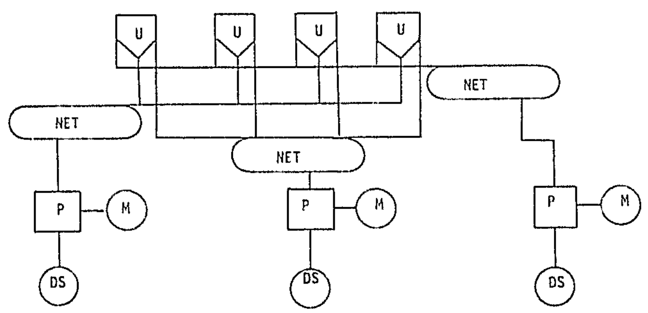


In this configuration all users have access to a common communications net and selective access to a processor complex depending upon the application area they are currently using.

Advantages
— Application areas are optimally placed in computer complexes according to their specific processor and data storage requirements.
— File maintenance is accomplished for each application area at its own location.
— File maintenance is done by persons who may have greater expertise in the specific application.

Disadvantages
— Lack of direct central control.
— Demands a high degree of coordination between processor complexes.
— Communications network must be more sophisticated to interface users to several different processor complexes.

Distributed Communication CLIS

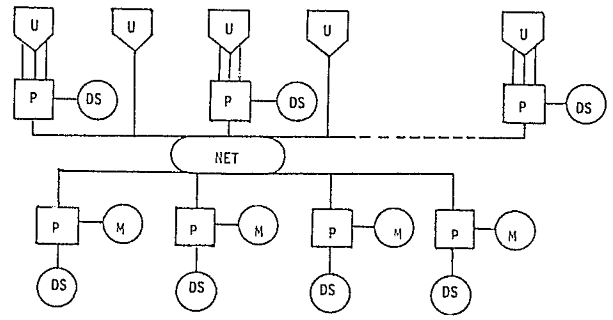


This system configuration provides for a communication network for each application area or groups of application areas. It also provides for distributed processing functions. Each user would have one or more terminals to connect to the selected application area.

Advantages
— Application areas are optimally placed in computer complexes according to their specific processor and data storage requirements.
— Application areas operate independently.

Disadvantages
— Redundancy of communication networks is difficult to justify.
— Increased cost of communications.
— Lack of direct central control.
— User interface with CLIS may be more complex.

Hierarchical CLIS

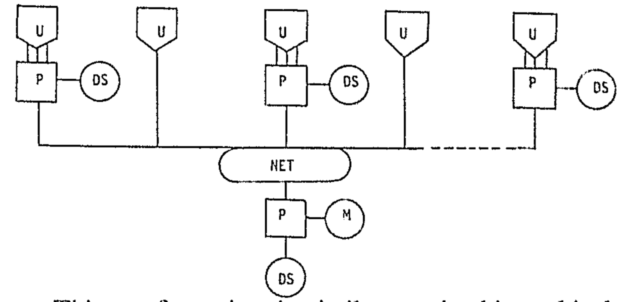


In this configuration certain high volume or specialized users will have dedicated local processor and data storage capability (using minicomputers) with multiple terminals. Other users will still be able to access the CLIS files using normal terminal operations. The communications network would become the central control point of the system.

Advantages
— Extremely flexible.
— Local processors can handle specialized data which may not be on the CLIS and are peculiar to each laboratory.
— Application areas are still optimally placed in computer complexes according to their specific processor and data storage requirements.
— Provides a multiple terminal capability for high-volume users.
— File maintenance is done by persons who may have greater expertise in specific applications.
— Potential use of handling automated data acquisition and reduction directly from instrumentation.

Disadvantages
— Increased complexity of local processor/main processor communication links.
— Potential increased cost of local processors and data storage.
— Provisions must be made for local system/programming support.

Centralized Hierarchical CLIS



This configuration is similar to the hierarchical CLIS, but it combines the advantages of a central processor complex with those of local processing capability.

Advantages
— Flexibility.
— Centralized control.
— Easy coordination of file maintenance.
— Local processors can handle specialized, non-CLIS data peculiar to each laboratory.
— Multiple terminal capability for high-volume users.
— Potential of acquiring/reducing data directly from instruments.

Disadvantages
— Configuration may not be optimal for the specialized requirements of some application areas.
— Possible peak-period competition for system resources, resulting in response delays.
— Increased complexity of local processor/main processor communication links.
— Potential increased cost of local processors and data storage.
— Local systems and programming support must be provided.

Perhaps the most attractive of these alternatives is the hierarchical configuration. This would provide local processor and data storage capability for those types of analytical data that are peculiar to individual laboratories. It will also provide multiple terminal expansion within a laboratory. A viable alternative to the distributive processor function would be a central processor and data storage which would provide a greater degree of central control.

Selection of a specific CLIS configuration cannot be made until all significant costing of major components has been accomplished. In Volume 3 cost parameters will be developed for each major functional component of the system. These cost parameters in combination with selected design alternatives will form the basis for a cost-comparison analysis. The conceptual design can then be finalized using the alternative designs and appropriate cost data.

APPENDIX A

LIST OF RESPONDENTS TO CLIS QUESTIONNAIRE

	Questionnaire log-in number
ALABAMA	
Alabama Department of Toxicology and Criminal Investigation, Auburn	85 (and 151)
Alabama Department of Toxicology and Criminal Investigation, Huntsville	9
ALASKA	
Alaska State Crime Laboratory, Anchorage	146
ARIZONA	
Crime Laboratory, Arizona Department of Public Safety, Phoenix	55
Crime Detection Laboratory, Phoenix Police Department	32
Maricopa County Medical Examiners Laboratory, Phoenix	110
City-County Crime Laboratory, Police Department, Tucson	14
CALIFORNIA	
Los Angeles Police Criminalistics Laboratory	167
Crime Laboratory, Kern County Sheriff's Office, Bakersfield	4
San Francisco Police Crime Laboratory	170
El Cajon Police Laboratory	132
Fresno County Sheriff's Laboratory, Fresno	53
Scientific Investigation Bureau, Huntington Beach Police Department	75
Criminalistics Laboratory, Los Angeles County Sheriff's Department, Los Angeles	30
South Bay Criminalistics Laboratory, Los Angeles County Sheriff's Department, Hermosa Beach	87
Laboratory of the Chief Medical Examiner, County of Los Angeles	16
Contra Costa County Sheriff's Criminalistics Laboratory, Martinez	69
San Rafael Satellite Laboratory	138
West Covina Satelli e Laboratory, West Covina, California	143
Riverside Regional Laboratory	147
Fresno Regional Laboratory, Department of Justice	155
Ventura County Sheriff's Criminalistics Laboratory	172
Criminalistics Section, Oakland Police Department	46
Institute of Forensic Sciences, Oakland	104
Laboratory Section, Alameda County Sheriff's Department, Pleasanton	131
Laboratory of Criminalistics, San Mateo County, Redwood City	60
Sacramento County Crime Laboratory, Sacramento	45
Drug Analysis Laboratory, California Department of Justice, Orange County	133
Salinas Regional Laboratory, California Department of Justice	126
San Luis Obispo Satellite Laboratory, California Department of Justice	134
Santa Barbara Regional Laboratory, California Department of Justice	125
Stockton Criminalistics Laboratory, California Department of Justice	127
Criminalistics Laboratory, San Bernardino County Sheriff's Department, San Bernardino	64
San Diego County Sheriff's Department Crime Laboratory, San Diego	100
Western Regional Laboratory, Drug Enforcement Administration, San Francisco	54
Laboratory of Criminalistics, County of Santa Clara, San Jose	35
Orange County Sheriff's Regional Criminalistics Laboratory, Santa Ana	77

Police Crime Laboratory, Columbus 5
Miami Valley Regional Crime Laboratory, Dayton 34
Cleveland Police Department, Forensic 142
Central Laboratory, Bureau of Criminal Identification & Investigation, London 112
City of Youngstown Crime Laboratory 116

OKLAHOMA

Oklahoma City Police Department 154
Tulsa Area Crime Laboratory 101
Oklahoma State Bureau of Investigation, Oklahoma City 171

OREGON

Crime Detection Laboratory, Department of State Police, Portland 108
Crime Detection Laboratory, Department of State Police, Eugene 106
Crime Detection Laboratory, Department of State Police, Medford 105
Crime Detection Laboratory, Department of State Police, Pendleton 11, 1[illegible]7

PENNSYLVANIA

Laboratory Division, Pennsylvania State Police, Harrisburg 22
Bucks County Crime Laboratory, Levittown 176
Regional Crime Laboratory, Pennsylvania State Police, Bethlehem 1
Regional Crime Laboratory, Pennsylvania State Police, Greensburg 29
Philadelphia Police Laboratory 98
Pittsburgh & Allegheny County Crime Laboratory, Pittsburgh 49

PUERTO RICO

Crime Laboratory, Puerto Rico Police Ponce. REFFERED US TO MAIN LABORATORY IN HATO REY — NO DATA 28
Crime Laboratory, Puerto Rico Police, San Juan 161
Toxicology Laboratory, Institute of Legal Medicine, University of Puerto Rico, Rio Piedras 31

SOUTH DAKOTA

South Dakota Division of Criminal Investigation 130

TENNESSEE

Nashville Toxicology Laboratory 71

TEXAS

Crime Laboratory, Texas Department of Public Safety, Austin 51
Texas Department of Public Safety, (East Texas) Tyler 145
Chemical Laboratory, Texas Department of Public Safety, Dallas 25
Criminalistics Laboratory, Texas Department of Public Safety, Lubbock 43
Field Laboratory, Texas Department of Public Safety, Waco 115
Laboratory, City of Beaumont Department of Health 99
South Central Regional Laboratory, Drug Enforcement Administration, Dallas 78, 11[illegible]
Southwestern Institute of Forensic Science, Dallas [illegible]6
Criminalistic Laboratory, Police Department, Ft. Worth 120
Houston Police Laboratory 27
Texas Department of Public Safety Laboratory, El Paso 173
Laboratory Bureau, San Antonio Police Department 121

UTAH

Cache County Sheriff's Office Laboratory, Logan 174
Bureau of Laboratories Utah Division of Health, Salt Lake City 65
Davis County Sheriff's Office 140

VIRGINIA

Special Testing & Research Laboratory, Drug Enforcement Administration, McLean 82

WASHINGTON

Joint Crime Laboratory, Tacoma/Pierce County 162
Laboratory Division, Seattle Police Department 90
King County Crime Laboratory, Seattle 91
Eastern Washington Regional Crime Laboratory, Spokane 89

WEST VIRGINIA

Criminal Identification Bureau, West Virginia State Police, South Charleston 8

WISCONSIN

State of Wisconsin Crime Laboratory 160
Bureau of Laboratories, Milwaukee Health Department 122

WYOMING

State Division of Laboratories, Laramie 79

APPENDIX B

CLASSES OF EVIDENCE EXAMINED BY CRIMINALISTICS INSTRUMENTS LABORATORY

(Compiled from response to question 6 (chart 1) of the CLIS Information Form)

X-Ray Diffraction. X-ray diffraction is used for the identification of crystalline substances. CLIS questionnaire respondents reported its use for drugs (abused, toxic, and other pharmaceuticals) and safe insulation.

GC/MS. Responses indicated its use in the identification of drugs, poisons, and trace substances.

Emission Spectroscopy and Atomic Absorption. Used for the examination of explosive traces.

IR. Infrared spectroscopy is used in the analysis of almost all classes of evidence, such as drugs and poisons, explosives, paints and pyrolosates, fibers, waxes and fats, petroleum products and lubricants, plasticizers and solvents, plastics (monomers, polymers, resins), and agricultural chemicals.

GC. Gas chromatography is a separation technique primarily used in isolating samples for further analysis and identification. It can however, be used to identify drugs and poisons and nondrug substances such as paints or hydrocarbons, flammable and accelerants.

UV. Ultraviolet spectroscopy is used for drugs, poisons, dyes and pigments, agricultural chemicals. and other organic compounds.

END