LITERATURE CITED

(1) Adamson, G. W., Cowell, J., Lynch, M. F., McLure, A. H. W., Town, W. G., and Yapp, A. M., "Strategic Considerations in the Design of a Screening System for Substructure Searches of Chemical Structure Files," *J. Chem. Doc.* **13,** 133 (1973).

(2) Craig, P. N., and Ebert, H. M., "Eleven Years of Structure Searching Using the SK&F Fragment Codes, *J. Chem. Doc.* **9,** 141-6 (1969).

(3) Adamson, G. W., Bush, J. A., Lynch, M. F., and McLure, A. H. W., unpublished data.

(4) Crowe, J. E., Lynch, M. F., and Town, W. G., "Analysis of Structural Characteristics of Chemical Compounds in a Large Computer-Based File. Part I. Non-cyclic Fragments," *J. Chem. Soc. (C),* **1970,** p. 990.

(5) Adamson, G. W., Lynch, M. F., and Town, W. G., "Analysis of Structural Characteristics of Chemical Compounds in a Large Computer-Based File. Part II. Atom-Centered Fragments," *J. Chem. Soc. (C),* **1971,** p. 3702.

(6) Adamson, G. W., Cowell, J., Lynch, M. F., McLure, A. H. W., Town, W. G., and Yapp, A. M., "Analysis of Structural Characteristics of Chemical compounds in a Large Computer-Based File. Part IV. Cyclic Fragments," *J. Chem. Soc., Perkin I,* **1973,** p. 863.

(7) Adamson, G. W., Creasey, S. E., Eakins, J. P., and Lynch, M. F., "Analysis of Structural Characteristics of Chemical Compounds in a Large Computer-Based File. Part V. More Detailed Cyclic Fragments," *J. Chem. Soc., Perkin I,* in press.

(8) Adamson, G. W., Creasey, S. E., and Lynch, M. F., "Analysis of Structural Characteristics of Chemical Compounds in the Common Data Base," *J. Chem. Doc.* **13,** 158 (1973).

Computerized Monitoring of the Inventory and Distribution of Research Chemicals*

STEPHEN J. FRYCKI,** FREDERICK F. GIARRUSSO, PAUL A. ROSKOS, DOLORES E. DANCSECZ, SALVATORE J. LUCANIA, and DENNIS M. O'BRIAN***

E. R. Squibb & Sons, Inc., Worldwide Headquarters, Princeton, N. J. 08540

Received May 29, 1973

A one-time data entry system, coupled with an efficient use of the computer, which provides inventory management, distribution, and audit reporting, the ability to answer special queries, and to produce customized reports is described. Results achieved by the system have been: paperwork has been drastically reduced, streamlined, or both; clerical labor has been significantly reduced; the preparation of reports to management has been automated; and the Research Chemicals Distribution Center can devote more time and resources to its prime function, the distribution of chemicals for biological testing.

Current research efforts by the pharmaceutical industry are creating large numbers of potential drugs and great volumes of data associated with their storage, distribution, and use for biological testing.

In the past, when research teams of the pharmaceutical industry were much smaller, the bench scientist would keep track of his own inventory of chemicals and the biological data resulting from their testing. This is still an excellent practice. However, the volume of data generated by current multidisciplinary research teams and the need for information by the professionals who comprise such teams has required a greater expenditure of time and effort than is optimal for any individual whose major effort is in research.

More recently, a central depository of data, manually controlled, was in use. This method, although accurate, was hindered by the inordinate amount of manual labor necessary to search through huge volumes of data for retrieving information requested by the research staff. Additionally, the manual processing of more than 30,000 transactions of chemical compounds per year was approaching unmanageability. The manipulation of data proved to be particularly difficult when a search of the historical records or of complete inventories was needed.

The clearest way to alleviate the problem appeared to be the use of a computerized system that would make the entering of data easy and simple and would be versatile in the retrieval of data. To this end, the Chemical Audit and Distribution (CHAUD) System was designed.

SUMMARY OF MANUAL SYSTEM PROBLEMS

The difficulties associated with the manual operation of the chemical distribution function were divided into six problem areas:

1. Extraordinary volume of paper and paperwork
2. Multiple entries of the same data
3. Long search times required for any queries other than one variable lookup—e.g., difficulty in finding all chemical compounds that had been tested in a number of biological assays
4. Difficulty in maintaining adequate inventories of important compounds
5. Much manual time expended in producing regularly scheduled reports
6. No simple way of maintaining "history files"

Associated with the third problem area were the typical types of queries that were handled by the Research Chemicals Distribution Center:

1. Find all compounds sent for examination in test number NNNN

* Presented at the 8th Middle Atlantic Regional Meeting, Washington, D. C., Jan. 16, 1973.

** To whom correspondence should be addressed.

*** Present address: Johnson and Johnson, Inc., New Brunswick, N. J.

2. Find all compounds sent for examination in test number NNNN during the period MM/DD/YY to MM/DD/YY
3. Find the numbers of all biological tests for which compound NNNNNN was submitted
4. Find all compounds within a compound number range that have an inventory quantity greater than a specified amount
5. Issue a warning when the inventory quantities of specified compounds fall below a specified amount
6. Produce reports showing selected information (in any sequence requested and with any data field(s) printed or suppressed)

SYSTEM DESIGN GOALS

Awareness of the six problem areas and the types of queries handled by the Research Chemicals Distribution Center led to a formulation of system design goals to optimize the following functions:

1. Ease and simplicity of one-time data entry
2. Speed of data retrieval
3. Versatility of data retrieval
4. Generation of regular summary reports and updating of historical files
5. Updating and monitoring of inventories
6. File maintenance
7. Efficiency of computer processing (dollars)

After system specifications had been accepted by the user, the system was implemented in late 1971. Implementation required about 6 man-weeks of systems personnel time.

AUTOMATED SYSTEM OVERVIEW

The CHAUD System maintains a data base of Squibb chemical compounds, their distribution history, inventory quantities, and handling precautions. It provides all the required facilities for information retrieval, report generation, and file maintenance. The system is modular, consists of eight programs written in PL/1 and COBOL, and uses the INQUIRE software package as a core for retrieval and data-management facilities. It is currently run on an IBM 370/155 computer with 512K of core storage and 3330 magnetic disk storage. A general flowchart of the system is presented in Figure 1.

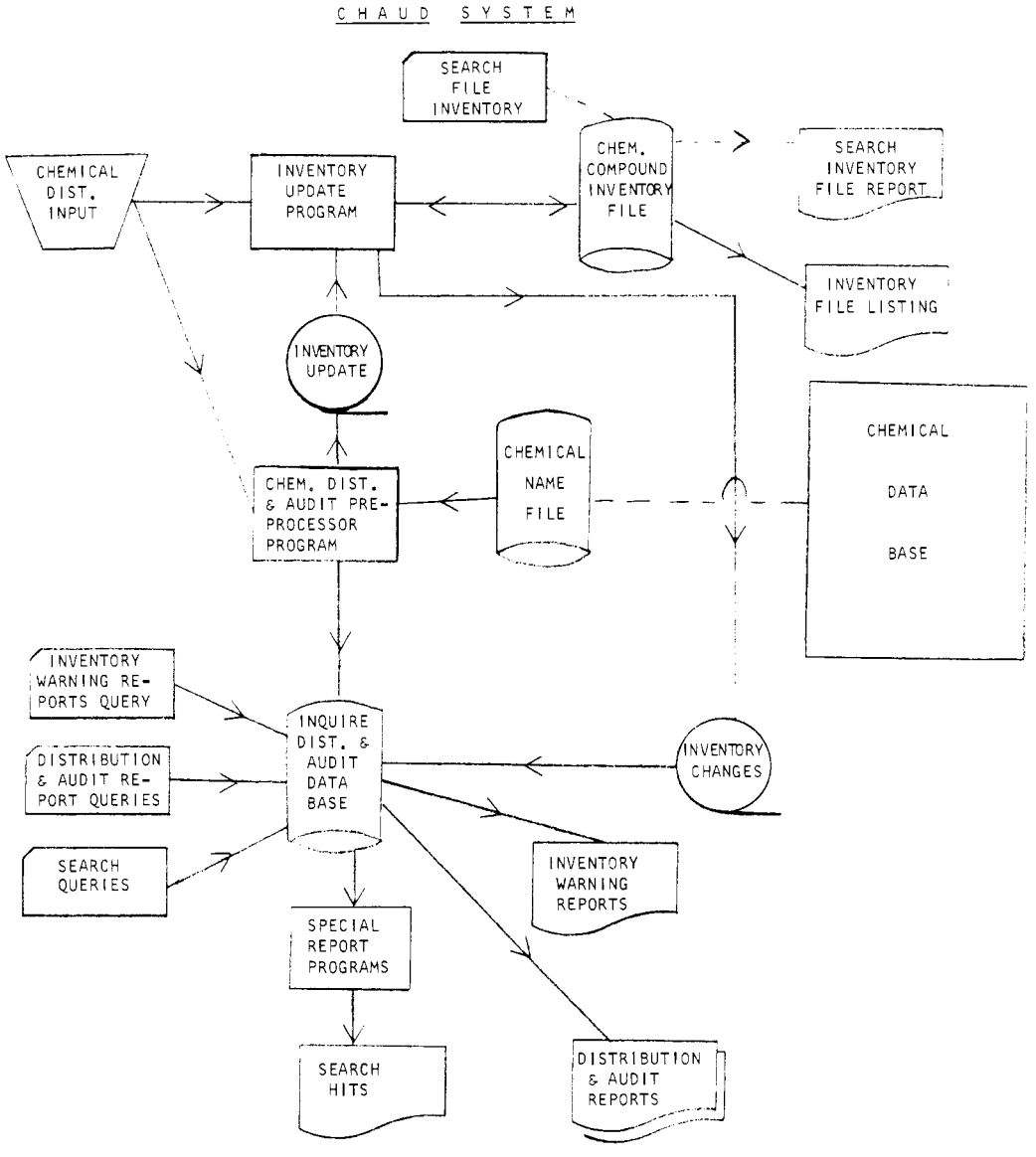


Figure 1. Chemical Audit Distribution (CHAUD)

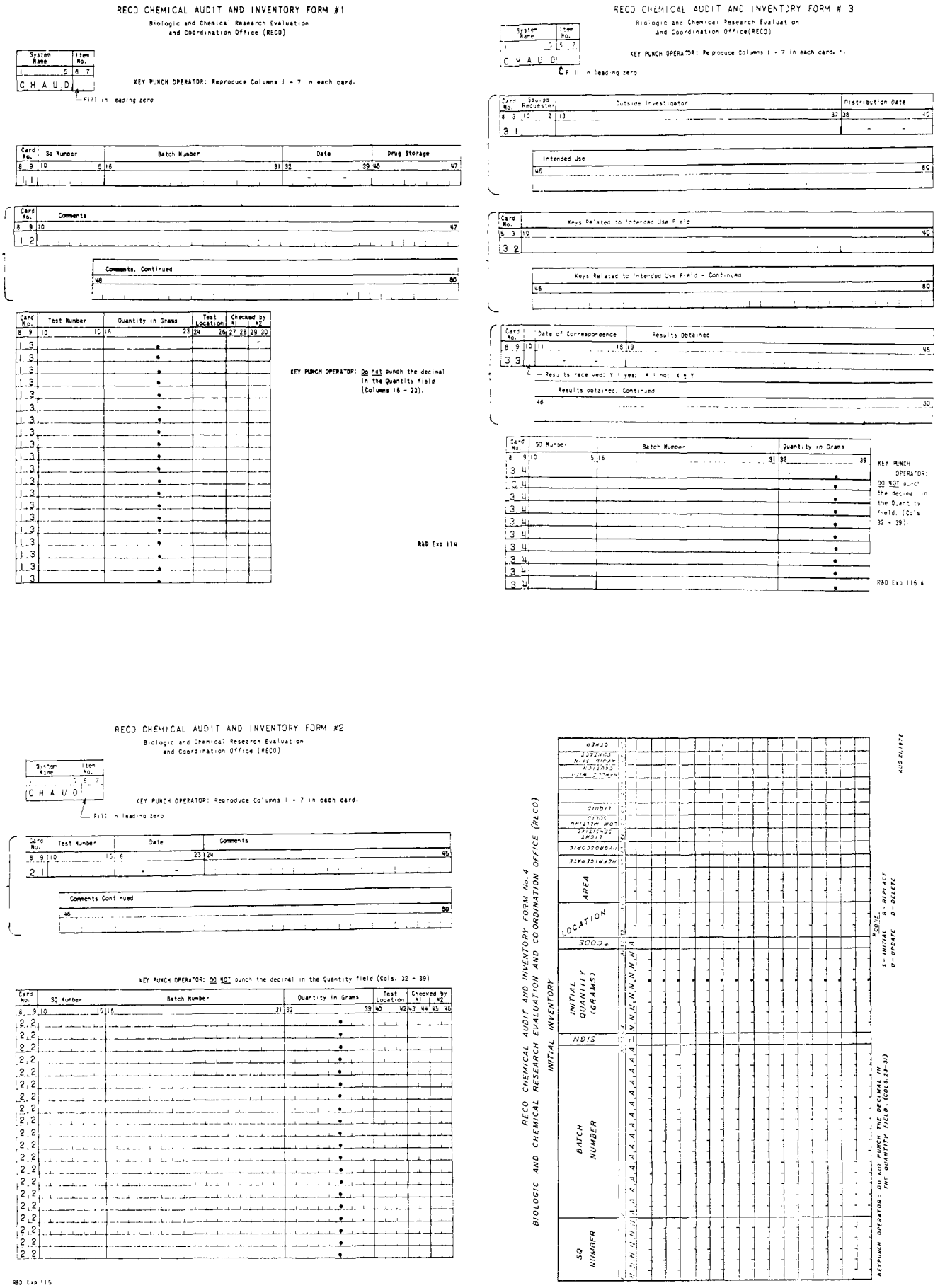


Figure 2. CHAUD input forms

CHAUD FILE STRUCTURE

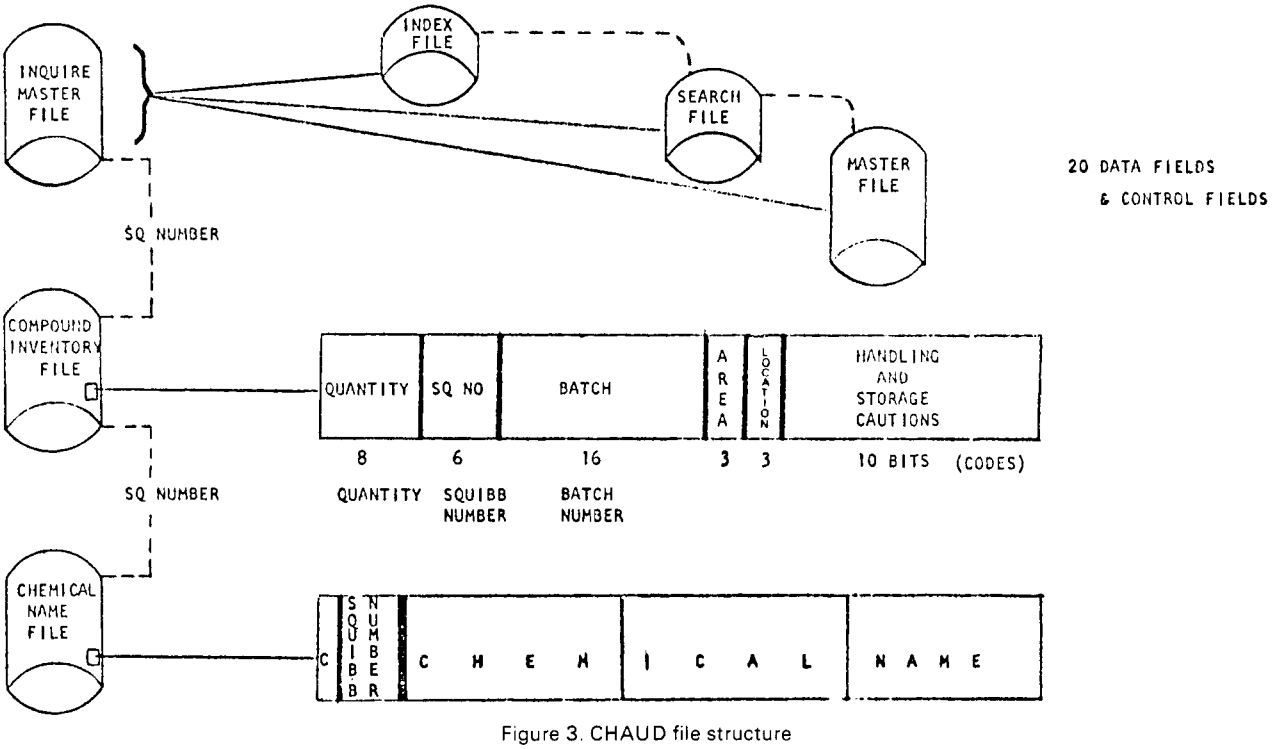
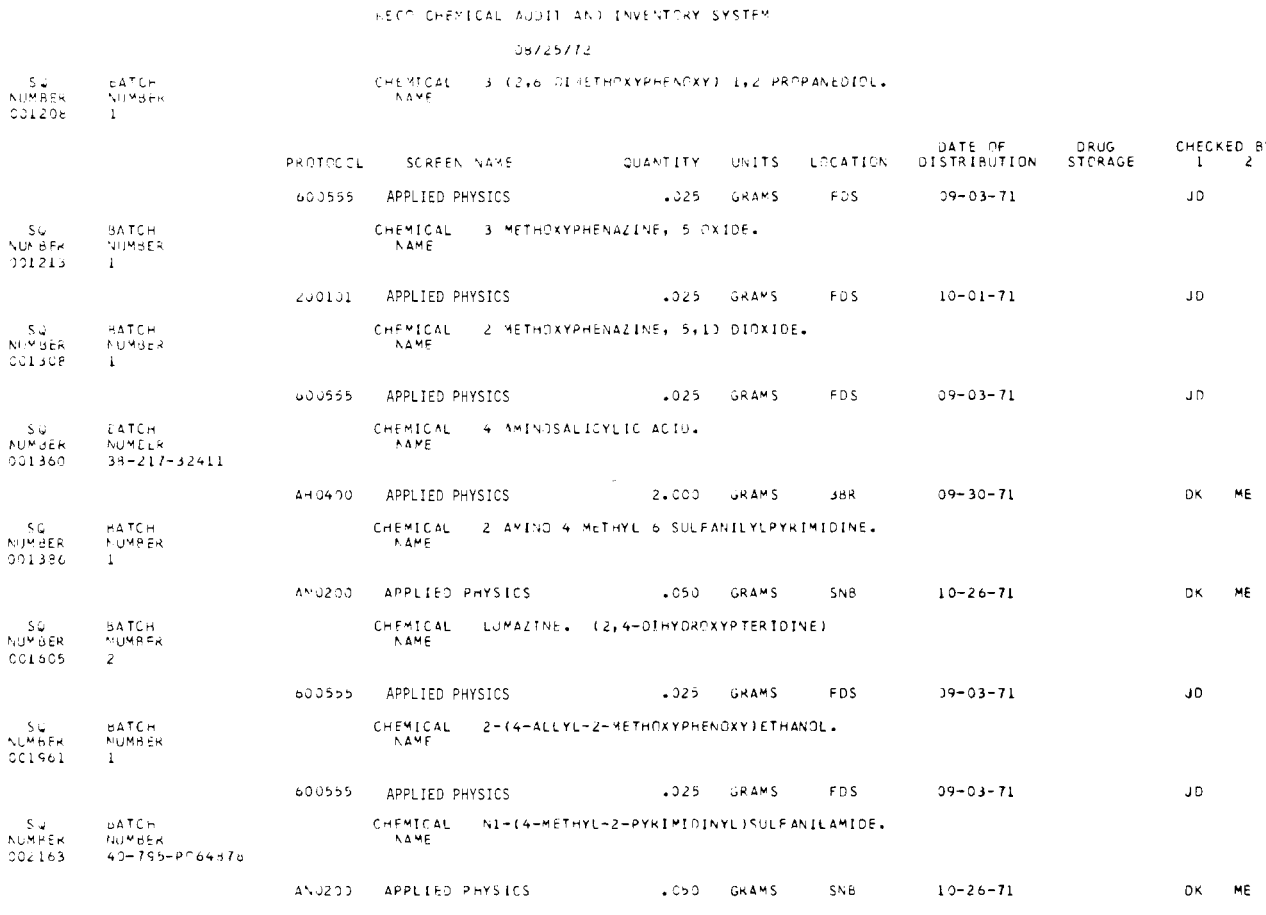


Figure 3. CHAUD file structure

KEYWORD LIST-CHEMICAL DIST. AUDIT 03-06-73 PAGE 104

SQNUMBER=080056	8	SQNUMBER=080106	10	STRUCTURE-FUNCTION	13
SQNUMBER=080057	4	SQNUMBER=080107	6	SUBCELLULAR PARTICLE	1
SQNUMBER=080058	6	SQNUMBER=080108	7	SUBSTANCE C	1
SQNUMBER=080059	5	SQNUMBER=080109	7	SUGAR TRANS	4
SQNUMBER=080060	5	SQNUMBER=080110	5	SURFACE TENSION	1
SQNUMBER=080061	7	SQNUMBER=080111	5	SUSCEPTIBILITY	8
SQNUMBER=080062	5	SQNUMBER=080112	7	SUSCEPTIBILITY DISCS	1
SQNUMBER=080063	6	SQNUMBER=080113	6	SWEETENER	2
SQNUMBER=080064	12	SQNUMBER=080114	7	SWINE	1
SQNUMBER=080065	4	SQNUMBER=080115	6	SYNAPSE	2
SQNUMBER=080066	5	SQNUMBER=080116	4	SYNAPTIC FUNCTION	10
SQNUMBER=080067	6	SQNUMBER=080117	1	SYNTHESIS	34
SQNUMBER=080068	7	SQNUMBER=080118	7	SYNTHETIC	2
SQNUMBER=080069	7	SQNUMBER=080119	5	SYNTHETIC MEMBRANE	3
SQNUMBER=080070	3	SQNUMBER=080120	5	SYSNAME=CHAUD	32260
SQNUMBER=080071	6	SQNUMBER=080121	6	T.CRUZI	1
SQNUMBER=080072	7	SQNUMBER=080122	5	T.PYRIFORMIA	1
SQNUMBER=080073	5	SQNUMBER=080123	6	T.RHODESIENSE	5
SQNUMBER=080074	6	SQNUMBER=080124	5	TABLETS	1
SQNUMBER=080075	5	SQNUMBER=080125	4	TB	1
SQNUMBER=080076	5	SQNUMBER=080126	4	TEMPERATURE	2
SQNUMBER=080077	8	SQNUMBER=080127	4	TERATOGENIC	4
SQNUMBER=080078	6	SQNUMBER=080128	4	TESTING	1
SQNUMBER=080079	4	SQNUMBER=080129	4	TESTIS	1
SQNUMBER=080080	5	SQNUMBER=080130	4	TESTLOC=AMR	603
SQNUMBER=080081	5	SQNUMBER=080131	4	TESTLOC=BEL	1
SQNUMBER=080082	6	SQNUMBER=080132	4	TESTLOC=CUL	1
SQNUMBER=080083	5	SQNUMBER=080133	4	TESTLOC=ENG	1
SQNUMBER=080084	7	SQNUMBER=080134	4	TESTLOC=FDS	5501
SQNUMBER=080085	7	SQNUMBER=080135	4	TESTLOC=FRN	1
SQNUMBER=080086	6	SQNUMBER=080136	4	TESTLOC=GRE	1
SQNUMBER=080087	7	SQNUMBER=080137	4	TESTLOC=HZL	2210
SQNUMBER=080088	4	SQNUMBER=080138	4	TESTLOC=IDL	8
SQNUMBER=080089	5	SQNUMBER=080139	4	TESTLOC=IRA	1
SQNUMBER=080090	5	SQNUMBER=080140	4	TESTLOC=JAP	2
SQNUMBER=080091	4	SQNUMBER=080141	4	TESTLOC=OUT	3301
SQNUMBER=080092	3	SQNUMBER=999999	17	TESTLOC=OVC	1
SQNUMBER=080093	10	SQUAMOUS	2	TESTLOC=PHK	267
SQNUMBER=080094	6	SQUAMOUS CARCINOMA	1	TESTLOC=PRN	4365
SQNUMBER=080095	9	STABILITY	4	TESTLOC=SAC	1048
SQNUMBER=080096	6	STANDARD	96	TESTLOC=SFR	1
SQNUMBER=080097	7	STANDARDISE	1	TESTLOC=SKF	1
SQNUMBER=080098	8	STEREOCHEMISTRY	1	TESTLOC=SNB	7554
SQNUMBER=080099	8	STEROID	20	TESTLOC=SRH	31
SQNUMBER=080100	8	STEROL	6	TESTLOC=WBR	7361
SQNUMBER=080101	6	STOMACH	4	TESTOSTERONE	2
SQNUMBER=080102	9	STREPTOCOCCUS	1	TETRACYCLINE	1
SQNUMBER=080103	6	STREPTOGRAMIN	3	TETRAHYMENA PYRIFORMIS	1
SQNUMBER=080104	7	STREPTOMYCIN	1	THIOSTREPTON	1
SQNUMBER=080105	7	STRUCTURE	28	THIRST	3

Figure 4

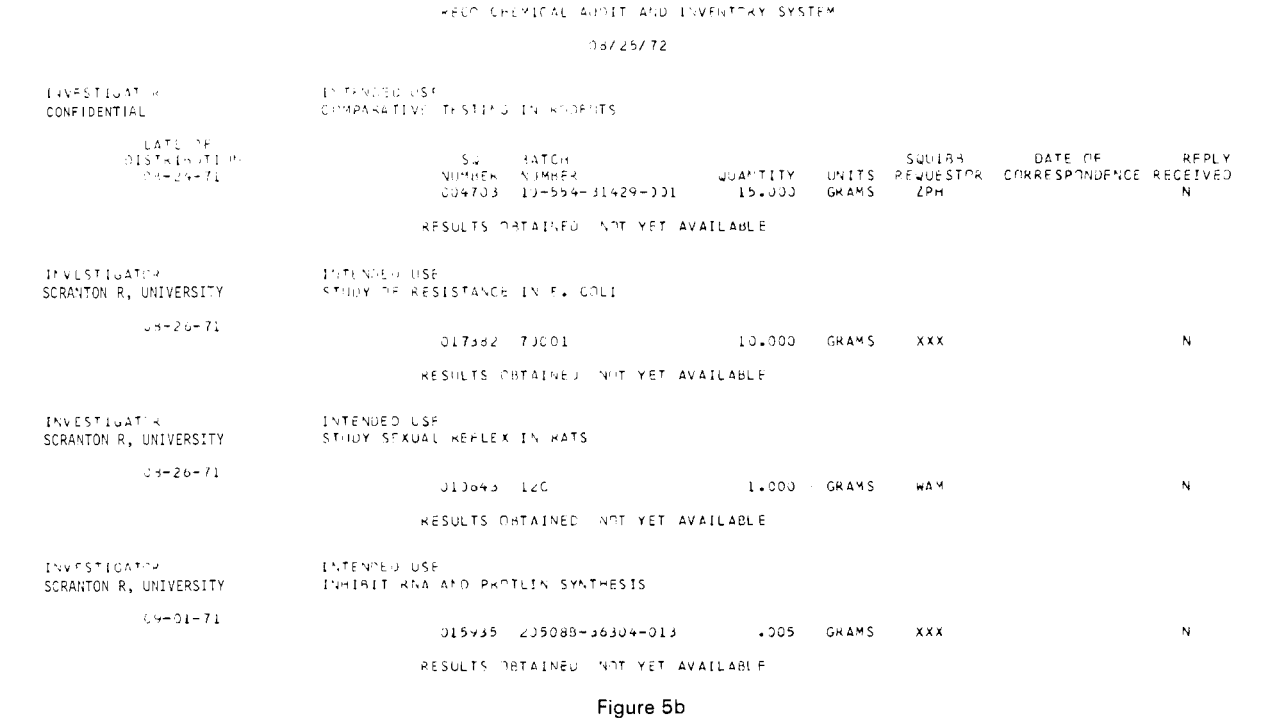


RECC CHEMICAL AUDIT AND INVENTORY SYSTEM

08/25/72

SQ NUMBER	BATCH NUMBER		PROTOCOL	SCREEN NAME	QUANTITY	UNITS	LOCATION	DATE OF DISTRIBUTION	DRUG STORAGE	CHECKED BY 1	2
001208	1	CHEMICAL NAME: 3 (2,6 DIMETHOXYPHENOXY) 1,2 PROPANEDIOL.									
			600555	APPLIED PHYSICS	.025	GRAMS	FDS	09-03-71		JD	
001213	1	CHEMICAL NAME: 3 METHOXYPHENAZINE, 5 OXIDE.									
			200101	APPLIED PHYSICS	.025	GRAMS	FDS	10-01-71		JD	
001308	1	CHEMICAL NAME: 2 METHOXYPHENAZINE, 5,10 DIOXIDE.									
			600555	APPLIED PHYSICS	.025	GRAMS	FDS	09-03-71		JD	
001360	38-217-32411	CHEMICAL NAME: 4 AMINOSALICYLIC ACID.									
			AH0400	APPLIED PHYSICS	2.000	GRAMS	BBR	09-30-71		DK	ME
001386	1	CHEMICAL NAME: 2 AMINO 4 METHYL 6 SULFANILYLPYRIMIDINE.									
			AN0200	APPLIED PHYSICS	.050	GRAMS	SNB	10-26-71		DK	ME
001605	2	CHEMICAL NAME: LUMAZINE. (2,4-DIHYDROXYPTERIDINE)									
			600555	APPLIED PHYSICS	.025	GRAMS	FDS	09-03-71		JD	
001961	1	CHEMICAL NAME: 2-(4-ALLYL-2-METHOXYPHENOXY)ETHANOL.									
			600555	APPLIED PHYSICS	.025	GRAMS	FDS	09-03-71		JD	
002163	40-795-PC64878	CHEMICAL NAME: N1-(4-METHYL-2-PYRIMIDINYL)SULFANILAMIDE.									
			AN0200	APPLIED PHYSICS	.050	GRAMS	SNB	10-26-71		DK	ME

Figure 5a



RECC CHEMICAL AUDIT AND INVENTORY SYSTEM

08/25/72

INVESTIGATOR: CONFIDENTIAL
INTENDED USE: COMPARATIVE TESTING IN RODENTS

DATE OF DISTRIBUTION	SQ NUMBER	BATCH NUMBER	QUANTITY	UNITS	SQUIBB REQUESTOR	DATE OF CORRESPONDENCE	REPLY RECEIVED
08-24-71	004703	10-554-31429-001	15.000	GRAMS	ZPH		N

RESULTS OBTAINED NOT YET AVAILABLE

INVESTIGATOR: SCRANTON R, UNIVERSITY
INTENDED USE: STUDY OF RESISTANCE IN E. COLI

08-26-71	017382	73001	10.000	GRAMS	XXX		N

RESULTS OBTAINED NOT YET AVAILABLE

INVESTIGATOR: SCRANTON R, UNIVERSITY
INTENDED USE: STUDY SEXUAL REFLEX IN RATS

08-26-71	013643	120	1.000	GRAMS	WAM		N

RESULTS OBTAINED NOT YET AVAILABLE

INVESTIGATOR: SCRANTON R, UNIVERSITY
INTENDED USE: INHIBIT RNA AND PROTEIN SYNTHESIS

09-01-71	015935	205088-36304-013	.005	GRAMS	XXX		N

RESULTS OBTAINED NOT YET AVAILABLE

Figure 5b

```
FIND SQNUMBER=009727 AND QUANTITY IS '00010000' TO '00020000',
TAB SQNUMBER 5 BATCHNO 17 QUANTITY 40 (03 9) UNITS 53,
HEADER 'INQUIRE WARNING REPORT'
'SQ NUMBER 9727  QUANTITY 10 TO 20 GRAMS'.
```

INQUIRE WARNING REPORT
SQ NUMBER 9727 QUANTITY 10 TO 20 GRAMS

PAGE 1
07/14/73

SQNUMBER	BATCHNO	QUANTITY	UNITS
009727	20-163-35749-C09	15.000	GRAMS
009727	20-163-32203-001	10.000	GRAMS
009727	20-163-38667-022	20.000	GRAMS
009727	20-163-32203-CC1	10.000	GRAMS
009727	2071A-48540-C04	15.000	GRAMS

ITEMS RETRIEVED 5

CPU TIME - 0.77 SECONDS

Figure 6

SYSTEM INPUTS

All input to the system is through the use of four forms, each regulating one of the following functions.

1. A single chemical compound distributed to many biological tests
2. Many chemical compounds distributed to one test
3. Chemical compounds distributed to outside investigators
4. Initiation of inventory for new compounds or new batches of compounds

Examples of these input forms are shown in Figure 2. They are designed to enable the chemical distribution monitor to input as little repetitive information as necessary for each compound transfer. For example, Forms 1 and 2 contain the same data fields, but Form 1 is fixed on one Squibb compound and is repetitive for investigators; whereas, Form 2 is fixed on one investigator but is repetitive for Squibb compounds. Form 3, used for distribution of Squibb compounds to outside investigators, is fixed on one investigator. Form 4 is used to initiate an inventory for a new Squibb compound or to update the inventory when more of the compound is added. Input is processed on demand as it arrives. It may consist of any number of forms in any combination. The data are keypunched, stringently edited by computer, and, when acceptable, loaded into the data base.

DATA BASE DESCRIPTION

The total data base is composed of the 21 data elements listed below:

1. compound number
2. compound batch number
3. chemical name
4. drug storage area
5. drug storage location
6. test number (protocol)
7. test name
8. comments on test
9. requester (for test)
10. test location
11. distribution date
12. intended use
13. intended-use keywords
14. quantity distributed
15. unitage
16. outside investigator
17. correspondence date
18. results obtained
19. system identification
20. handling precautions
21. on-hand inventory quantity

In addition, several control fields are used within the system for linking the modules that comprise the system and for special input/output functions.

[illegible] CHEMICAL AUDIT AND INVENTORY SYSTEM
INVENTORY ISAM FILE SEARCH 07/14/73
SEARCH CRITERIA [illegible] 10000.000

[illegible]	BATCH [illegible]	QUANTITY	UNITS
[illegible]	[illegible]	21381.560	GRAMS
[illegible]	[illegible]	31531.550	GRAMS
[illegible]	[illegible]	28552.500	GRAMS
[illegible]	[illegible]	11451.140	GRAMS
[illegible]	[illegible]	14735.000	GRAMS
[illegible]	[illegible]	10100.000	GRAMS
[illegible]	[illegible]	17559.540	GRAMS
[illegible]	[illegible]	14500.000	GRAMS
[illegible]	[illegible]	52010.000	GRAMS
[illegible]	[illegible]	11330.975	GRAMS
[illegible]	[illegible]	21820.[illegible]	GRAMS
[illegible]	[illegible]	12575.675	GRAMS
[illegible]	[illegible]	13799.950	GRAMS
[illegible]	[illegible]	10000.000	GRAMS
[illegible]	[illegible]	10000.000	GRAMS
[illegible]	[illegible]	10000.000	GRAMS
[illegible]	[illegible]	66392.000	GRAMS

[illegible] CHEMICAL AUDIT AND INVENTORY SYSTEM
INVENTORY ISAM FILE SEARCH 07/14/73
SEARCH CRITERIA [illegible] 10000.000
TOTAL RECORDS THAT MEET CRITERIA 17

Figure 7

DATA BASE ORGANIZATION

The data base consists of three modules, all of which are disk resident and linked by the Squibb compound number, a standardized unique key existing in all modules (Figure 3).

The largest module is the data base resident under INQUIRE. It is this module, currently consisting of 26,000 records, that maintains the distribution history of all compounds. Most special queries utilize this data base. As shown in Figure 3, the INQUIRE data base is composed of three linked, direct-access files: index, search, and master. After gaining access to the data base via keys in the index file, INQUIRE operates by means of search-file pointers until it is ready to enter the master file for text searching, field-value comparisons, or master-record printout. The data base is so organized that there is no reversal in the direction of the movement of the disk access arm during query processing, a feature that greatly enhances the performance of the system. Search times against this 26,000-record data base are, at most, several seconds. INQUIRE is a direct-access based information retrieval/data management software package written mostly in PL/1, using a multi-file "chaining" or "threaded list" structure. Provided that the lengths of keyword pointer chains do not become extreme, processing time is

```
FIND DISTYP=72 AND DISTMO=03,
BREAK ON PROTOCOL SKIP 2 'PROTOCOL' 1 PROTOCOL 12          TOTAL OF ITEMS
 23 (I 5),
TOTAL ITEMS 23 (I 5),
COUNT,
HEADER 'PROTOCOLS RUN DURING MARCH 1972'.
```

PROTOCOLS RUN DURING MARCH 1972 PAGE 1
07/14/73

	PROTOCOL	ITEMS
PROTOCOL	AH0000	3
PROTOCOL	AH0300	564
PROTOCOL	AH0400	1
PROTOCOL	ANC000	28
PROTOCOL	AN0100	1
PROTOCOL	AN0200	32
PROTOCOL	AN0400	3
PROTOCOL	AN0500	1
PROTOCOL	AN0600	1
PROTOCOL	BP0100	154
PROTOCOL	BP0200	193
PROTOCOL	CH0000	66
PROTOCOL	DM0000	7
PROTOCOL	FP0000	10
PROTOCOL	FP0110	11
PROTOCOL	FP0300	1
PROTOCOL	MC0000	203

Figure 8a

essentially unchanged regardless of how large the file becomes.

The second module is the data base resident under COMPOUND INVENTORY. This is an indexed sequential file, and the record (as shown in Figure 3) contains six fields:

1. compound number;
2. compound batch number;
3. quantity on hand;
4. drug storage area;
5. drug storage location (within area);
6. drug handling and storage cautions (ten 1-bit codes).

The third data-base module is the CHEMICAL NAME file. This file is part of Squibb's over-all chemical database system, but is shared with the chemical distribution system. It is an indexed sequential file consisting of a control bit, the Squibb compound number, and up to 180 characters of a chemical name. This file is used to input the standard chemical names of Squibb compounds into the INQUIRE data base, a process that significantly reduces input labor (6-character Squibb number vs. as many as 180 characters for a chemical name), and assures by computer processing that all chemical names exist in standard spelling conventions.

COMPUTER PROCESSING

Master Preprocessing Program. All input to the system, with the exception of inventory initiation and addition, enters the "MASTER" preprocessing program (written in ANS COBOL), which performs the functions of data editing, formatting for loading under INQUIRE, adding standardized chemical names to the records, and inventory updating.

Data Editing. The editing consists of checking for legal codes, naes, dates, test numbers, numerics, presence of required data, etc., and, if errors are found, printing an appropriate error message with the rejected record. At the conclusion of every update, the error messages, together with an analysis of records added, deleted, changed, etc., are provided to the Research Chemicals Distribution Center. Stringent computer editing and record accounting assures that the burden of system input rests with the Distribution Center, rather than with Computer Operations.

Formatting. For loading under the INQUIRE System, every record must be processed into a uniform format with respect to field length, content, and type. Fields may be fixed or variable and, if designated as key fields, their contents are automatically transformed into retrieval keys. Key fields may also be prefixed—e.g., all compound num-

```
FIND FORMNO=3 AND DISTYP=70 AND OUTINV CONTAINS 'UNIVERSITY      ',
SORT OUTINV,
TAB OUTINV 31 INTNOUSE 65 SQNUMBER 2  BATCHNO 18 DISTDATE 41 QUANTITY 56 (D3 9)
CHEMNAME 73,
HEADER 'SQ COMPOUNDS SENT TO  UNIVERSITY      ',
SUBHEAD '1970'.
```

SQ COMPOUNDS SENT TO UNIVERSITY
1970

PAGE 1
06/07/72

SQNUMBER	BATCHNO	OUTINV	DISTDATE	QUANTITY	INTNOUSE / CHEMNAME
		COSTANTINI B, UNIVERSITY			DENSITY GRADIENT/ MITOCHONDRIA
009410	31-770-83112		05-06-70	500.000	3,5 DIACETAMIDO 2,4,6 TRIIODOBENZOIC ACID, METHYLGLUCAMINE SALT.
		COSTANTINI B, UNIVERSITY			ANTISCHISTOSOMAL ACTIVITY
018506	11		06-10-70	100.000	1,2,4 OXADIAZOLE, 5 AMINO 3 (2 (5 NITRO 2 FURYL)VINYL) .
		COSTANTINI B, UNIVERSITY			ANTISCHISTOSOMAL ACTIVITY
065613	RR001		04-18-70	4.000	3-(2-(5-NITRO-2-FURYL)VINYL)-1,2,4-OXADIAZOLE.
		COSTANTINI B, UNIVERSITY			ANTI-SCHISTOMAL,PARTICLE SIZE
019104	NN003		07-06-70	2.000	5 ACETAMIDO 3 (2 (5 NITRO 2 FURYL)VINYL) 1,2,4 OXADIAZOLE.
		COSTANTINI B, UNIVERSITY			ANTI-SCHISTOMAL,PARTICLE SIZE
018707	RR002		07-06-70	2.000	5 (DIMETHYLAMINO) 3 (2 (5 NITRO 2 FURYL)VINYL) 1,2,4 OXA DIAZOLE.
		COSTANTINI B, UNIVERSITY			ANTI-SCHISTOSOMAL,PARTICLE SIZE
018506	RR00883		10-07-70	100.000	1,2,4 OXADIAZOLE, 5 AMINO 3 (2 (5 NITRO 2 FURYL)VINYL) .
		COSTANTINI B, UNIVERSITY			COMPARISON STUDY WITH SQ 18506
018642	RR003		11-16-70	5.000	3 (2 (5 NITRO 2 FURYL)VINYL) 5 (TRICHLOROMETHYL) 1,2,4 OXA DIAZOLE
		COSTANTINI B, UNIVERSITY			COMPARISON STUDY WITH SQ 18506
018707	RR002		11-16-70	5.000	5 (DIMETHYLAMINO) 3 (2 (5 NITRO 2 FURYL)VINYL) 1,2,4 OXA DIAZOLE.
		COSTANTINI B, UNIVERSITY			COMPARISON STUDY WITH SQ 18506
019104	NN003		11-16-70	5.000	5 ACETAMIDO 3 (2 (5 NITRO 2 FURYL)VINYL) 1,2,4 OXADIAZOLE.
		COSTANTINI B, UNIVERSITY			ANTISCHISTOSOMAL ACT
018506	11		11-24-70	2.000	1,2,4 OXADIAZOLE, 5 AMINO 3 (2 (5 NITRO 2 FURYL)VINYL) .
		COSTANTINI B, UNIVERSITY			MONKEY STUDY
018506	RR00883		12-08-70	400.000	1,2,4 OXADIAZOLE, 5 AMINO 3 (2 (5 NITRO 2 FURYL)VINYL) .
		COSTANTINI B, UNIVERSITY			ANTISCHISTOSOMAL
019104	NN003		12-28-70	3.000	5 ACETAMIDO 3 (2 (5 NITRO 2 FURYL)VINYL) 1,2,4 OXADIAZOLE.
		COSTANTINI B, UNIVERSITY			ANIMAL STUDIES
009468	22-380-91830-001		05-21-70	100.000	AMPHOTERICIN B
		COSTANTINI B, UNIVERSITY			PROSTATE GLAND/LABORATORY ANIMALS
009468	22-380-91830-001		03-25-70	100.000	AMPHOTERICIN B
		COSTANTINI B, UNIVERSITY			CLINICAL STUDY ON BPH AND CANCER
009468	22-380-91830-001		08-05-70	100.000	AMPHOTERICIN B

Figure 8b

```
FIND FORMNO=3 AND DISTYP=70 AND INTNOUSE CONTAINS 'ERYTHROCYTE GHOSTS',
TAB OUTINV 5 SQNUMBER 34 BATCHNO 44 DISTDATE 63 CHEMNAME 73,
SORT OUTINV,
HEADER 'SQ COMPOUNDS SENT OUTSIDE DURING 1970 TO STUDY  ERYTHROCYTE GHOSTS'.
```

SQ COMPOUNDS SENT OUTSIDE DURING 1970 TO STUDY ERYTHROCYTE GHOSTS

PAGE 1
06/07/72

OUTINV	SQNUMBER	BATCHNO	DISTDATE	CHEMNAME
EVANS R ,INST CANCER RES	015859	LRH-3599-90-10	02-13-70	NONACTIN.
EVANS R ,INST CANCER RES	009468	22-380-91830-001	02-13-70	AMPHOTERICIN B
EVANS R ,INST CANCER RES	009323	1898A-49921-002	02-13-70	NYSTATIN

ITEMS RETRIEVED 3

Figure 8c

bers are automatically prefixed with SQ NUMBER =— which has the effect of sorting all compound numbers in order within the keyword list. This list is a dictionary of keywords and their frequency counts produced by INQUIRE and used in query formulation. An example of the keyword list is shown in Figure 4.

Adding Chemical Name. The addition of chemical names to each record entering the system, discussed earlier, is intended to reduce input labor and the frequency of error.

Inventory Updating. Updating is accomplished by passing to the inventory 'UPDATE' program the amount of compound distributed for testing, as given on the input sheet.

Inventory Update Program. The inventory 'UPDATE' program accepts as input initial inventory amounts of compounds or additions to the inventory, and maintains, via information passed to it by the "MASTER" preprocessing program, a record of current quantities by sub-

***** CHEMICAL DISTRIBUTION AND AUDIT SYSTEM - INVENTORY FILE MASTER LISTING *****

PAGE 428
02/08/73

SQUIBB NUMBER	BATCH NUMBER	QUANTITY IN GRAMS	LOCATION	AREA	REFRIG-ERATE	HYGRO-SCOPIC	LIGHT SENSITIVE	LOW MELTING SOLID	LIQUID	NO. 6	NO. 7	HANDLE WITH CAUTION	AVOID SKIN CONTACT	OTHER
067267	RR001	0.500		PRN	X									
067268	RR001	1.000		PRN					X					
067269	RR001	2.500		PRN										
067270	RR001	1.400		PRN										
067271	RR001	1.000		PRN										
067272	RR001	1.500		PRN										
067273	RR001	0.600		PRN										
067274	RR001	2.000		PRN										
067275	RR001	1.500		PRN										
067276	RR001	3.000		PRN				X						
067277	RR001	1.000		PRN					X					
067278	RR001	2.500		PRN				X						
067279	RR001	2.200		PRN										
067280	RR002	0.100		PRN										
067281	RR001	3.770		PRN										
067282	RR001	1.525		PRN										
067287	RR002	0.125		PRN										
067289	RR002	0.240		PRN	X									
067291	RR001	0.200		PRN	X									
067292	RR001	0.350		PRN	X									
067297	RR001	19.940		PRN	X									
067298	RR001	21.940		PRN	X									
067299	RR001	2.000		PRN										
067300	RR001	1.000		PRN										
067301	RR001	0.600		PRN										
067305	RR001	1.000		PRN										
067306	RR001	1.000		PRN										
067307	RR001	0.000		PRN					X					
067308	RR001	1.000		PRN										
067309	RR001	3.000		PRN										
067310	RR001	4.000		PRN										
067311	RR001	3.000		PRN										
067313	RR001	0.550		PRN	X									
067314	RR001	1.200		PRN	X									
067315	RR001	0.500		PRN										
067317	RR001	1.000		PRN	X									
067324	RR001	0.300		PRN	X									
067325	RR001	0.240		PRN	X									
073103	EM-3103-MC	0.085												
073339	21d4A-45943-C02	118.270		PRN	X									
075094	PL4530-33A	0.355												
080000	NX001	0.610												
080000	NX002	2.090												
080001	NX001	0.320												
080001	NX002	4.950												
080002	NN007RL	0.000		SNB								X	X	X
080002	NX001	0.315												
080002	NX002	0.010												
080002	NX003	0.020												
080002	NX004	8.820												
080002	NX005	20.095												
080002	NX006	65.120												

Figure 9

tracting from the master inventory those quantities that have been distributed. In addition, every change to the inventory file results in the program producing the appropriate command statements to keep current the inventory quantities on the INQUIRE data base. This program is written in ANS COBOL.

Search and Reporting Programs. Search and reporting programs are available for all three CHAUD modules. All search and report programs are written in PL/1 Optimizer Version, except for the Inventory Search, which is written in ANS COBOL.

There are several kinds of history file reports needed by the Research Chemicals Distribution Center whose formats cannot be produced directly from INQUIRE. There reports are printouts of the history file in various sequences and are used as a desktop search tool for answering questions. They also serve as a backup for the computer system, allowing searches to be performed manually in the event of prolonged computer failure.

Figure 5a illustrates one of these reports in compound number order, along with its name, the test for which it was distributed, the quantity sent, and the date of distribution. Figure 5b illustrates a report in outside investigator sequence, showing which compounds, and in what quantities, were distributed, the intended use of each, and whether results have been received.

Inventory warning messages are also generated from the INQUIRE data base. For important compounds that require careful inventory monitoring, queries are maintained and run against the data base after each update. If the query results in a "hit" (inventory too low), an appropriate warning message is generated. An example is shown in Figure 6.

INVENTORY Data Base. The INVENTORY data base may be searched for any numeric quantity desired—i.e., $=$, $\leq$, $\geq$, and range. Search answers can be reported either as a count of the number of "hits" or as a printout of all "hits." The report produced in answer to an inventory query for all compounds that have more than 10,000 grams in stock is shown in Figure 7.

INQUIRE Data Base. The INQUIRE data base can be searched for any combination of data elements existing in the file by the use of Boolean criteria and English query statements. Searches can be based on keywords, keyword combinations, field-value comparisons, text scanning, or any combination thereof. Queries are formulated using the keyword list as a guide. INQUIRE's extensive features for generating reports allow almost all required reports to be produced directly from the data base in response to queries formulated and submitted by the Research Chemicals Distribution Center. Computer processing of such queries usually requires only a few seconds. If text scanning or field-value comparison is required, search time is increased, because the retrieved master file records must be scanned for acceptance or rejection on the basis of additional criteria.

Examples of the types of queries that are submitted by the Distribution Center are given in Figure 8. Figure 8a illustrates a straightforward query for all transactions in March 1972 (DISTYR = 72, DISTMO = 03), to be reported as a count by protocol (test). Although the query and resultant report are shown as a single unit in the illus-

tration, they are actually printed on separate pages, so that the report can be submitted directly to management.

Figure 8c illustrates a query for all compounds distributed to outside investigators in 1970 for testing "erythrocyte ghosts." Again, INQUIRE formatting instructions produce a custom report.

The inventory history file report is printed after each update and is used, as are the INQUIRE data base reports, as a desktop search tool and computer backup. An example of the report is given in Figure 9, which also shows the provision for including handling precautions within the file.

CHEMICAL NAME Data Base. This date base, which is a module of the Squibb Chemical Structure/Biological Data System, will not be discussed here, because its only use in the CHAUD System is to provide input of standardized chemical names.

FUTURE DEVELOPMENTS

It is planned that the CHAUD System will eventually function as one module of Squibb's Research and Development Information System (RADIS). RADIS was conceived as a "total" scientific information system for Squibb, and is being implemented stepwise as a series of modular data bases containing chemical, biological, and textual information for all Squibb compounds. These modules will eventually be linked via the standardized Squibb compound number, and queries will have access to all necessary data bases for providing answers.

REFERENCES

(1) Eckermann, E. H., Waters, J. F., Pick, R. O., and Shafer, J. A., "Processing Data from a Large Drug Development Program," *J. Chem. Doc.* **12**, 38-40 (1972).

(2) Frycki, S. J., "Information Transfer from Source to User Utilizing a Pharmaceutical Data Base," in *The Social Impact of Information* Retrieval, Medical Documentation Service, The College of Physicians of Philadelphia, 1970.

(3) Flores, I., "Data Structure and Management," Prentice-Hall, Inc., Englewood Cliffs, N. J., 1970.

A Computer-Based Comprehensive Bio-Data Information Retrieval System*

JEROME D. STEIN, JR.,** FRANCES M. DELANEY,*** SAL D. PELUSO, and LEE N. STARKER†
Warner-Lambert Research Institute, 170 Tabor Road, Morris Plains, N. J. 07950

Received April 27, 1973

A card-oriented biological data information system, programmed for a UNIVAC 1050 computer, suffered from a lack of flexibility. When IBM 360 equipment was acquired, it became both necessary and desirable to convert to the new equipment. Initial evaluation of requirements estimated that 3 to 4 man-years would be required to develop new systems and write necessary programs. To reach our goals at substantially less expense and at a more rapid pace, a decision was made to purchase a commercially available data management system called INQUIRE. Problems encountered in conversion from one system to the other are discussed. Data formats, input-output procedures, search strategies, etc., are described.

The Technical Information Service Group (TIS) of the Science Information Services Department at Warner-Lambert Research Institute is responsible for processing internally generated research information. The assignment is far greater than mere storage and retrieval of proprietary data. Personnel in TIS are required to coordinate and correlate various bits and pieces of intelligence received from many local and distant research facilities in all permutations and combinations that can be imagined. Such a task is impossible without the aid of a staff endowed with phenomenal memories or a most complete set of storage, search, and retrieval tools.

An early system developed for retrieval of biological data at TIS used McBee edge-notched card.[1] By its nature, the card limited the number of test procedures that could be recorded. Absolute numbers for dose and resulting raw data were not possible; card format restricted evaluative information to arbitrary ranges. Once a series of compounds meeting the requirements of a search were located, a somewhat similar card system was available for finding additional related compounds.[2] A next step was implementation of a system based on internally punched cards—the Hollerith-coded IBM card.[3]

In this system, basic 80-column tabulating cards were formatted with fixed-fields to handle three types of data. The first format handled measured animal dose-response data from fixed testing procedures. A second type of card layout rated subjective observations or scored data for behavioral effects from standard score sheets. Evaluation data from nonroutine testing procedures, devised or modified to bring out the particular properties of a compound in advanced development, were reported in technical reports.

These cards were used in and by themselves with card sorting equipment and an IBM 026 or 870 document writer. At a later stage in developing our systems approach to

* Presented before the Division of Chemical Literature, 6th Middle Atlantic Regional Meeting, ACS, Baltimore, Md., Feb. 4, 1971.

** Author to whom correspondence should be addressed.

*** Present address: G. D. Searle & Co., P.O. Box 5110, Chicago, Ill. 60680.

† Deceased.