of approved task group reports, and other types of information of interest to the world community of compilers. The distribution list will include CODATA members, national committee members, data project directors, and other interested persons and groups. CODATA Newsletter No. 1 was distributed in October 1968 from the ICSU-CODATA Central Office, Westendstrasse 19, 6 Frankfurt/Main, Germany-BRD.

The foregoing review tells us that CODATA is a going concern; that it is playing a significant role in giving coherence to the needed worldwide effort to compress, evaluate, and compile the numerical data of science and technology; and that it is taking appropriate steps to improve the quality of data so compiled.

LITERATURE CITED

(1) National Research Council, "International Critical Tables of Numerical Data: Physics, Chemistry and Technology," McGraw-Hill, New York, N. Y., 1926–1933.
(2) Landolt-Börnstein, "Zahlenwerte und Funktionen aus Physik-Chemie-Astronomie-Geophysik und Technik," 6th ed. 1950–).
(3) "Tables de Constantes et Données Numériques," fondées par Ch. Marie (1909): "Constantes Sélectionnées," Pergamon, Paris, current series (1947–).
(4) International Compendium of Numerical Data Projects, Springer-Verlag, Berlin-Göttingin-Heidelberg, 1969.

A Biologically Oriented Data Retrieval System*

THEODORE LEGATT, ROBERT P. GRANDY, and SAMUEL X. DELORENZO
Schering Corp., Bloomfield, N. J.

Received November 21, 1968

A computer-based storage and retrieval system for biological data has been in use at Schering for several years. The system is intended primarily for storage and retrieval of biological data of interest to managers and laboratory scientists, and permits select printouts of cumulative information on aspects of the research program.

Before the computerized file of biological data was developed, a unit record system had been used for search requests. Although this system had been adequate for many queries, it was limited by hardware and input-data restrictions.[1] The need for a comprehensive and accessible information system led to a study to define objectives, examine available hardware systems, and reconsider input data requirements. The conclusion of the study pointed towards a computer system and a reorganization of input data as the most feasible way to accomplish the objectives.

OBJECTIVES

1. To provide means to handle effectively a growing file of biological and chemical information.
2. To provide a foundation for centralizing all information on preclinical studies.
3. To permit the evaluation of current work and assist research management in long-range planning.
4. To provide monthly and quarterly summary reports on the progress of compounds through testing.
5. To identify compounds submitted for screening in which test results had not been reported. This would permit rapid determination of the progress of the compound in the screening stage.

OPERATIONAL FLOW

Samples of compounds synthesized in the laboratories are submitted for testing through the chemical distribution center. This center acts as a central repository for samples of all intermediate and final products synthesized. Compounds are indexed and filed and are readily available. Sample loss has been virtually eliminated by this method.

The distribution center weighs out samples for specific screens and forwards them to the technical information center for processing. The technical information center prepares a "Request for Laboratory Investigation" form (RLI) which accompanies the compound to the screening area. After biological screening has been completed, each investigator forwards a completed test report to the technical information center, where it is duplicated and distributed to research personnel. All sample delivery and test report data are incorporated in the central file system.

INPUT

The sample submission form (RLI) contains standard information that identifies the sample source, the compound, and the tests requested. The chemical structure on this form is not stored on magnetic tape; instead, it is coded and incorporated in a separate optical coincidence file of structures to be reported on later.

The test report returning from the laboratories contains a maximum amount of information in a format that facilitates key punching. The information scientist and laboratory investigator collaborate in the test report design.

Memoranda, letters, and other pertinent reports are also included in the input information. They are indexed in a format similar to that used for the basic test report and can be retrieved by compound number, author, organization, date, or subject.

TECHNICAL CENTER DATA FLOW

The flow of data within the technical information center is shown in Figure 1. The various laboratory forms and

* Presented before the Division of Chemical Literature, 156th Meeting, ACS, Atlantic City, N. J., September 13, 1968.

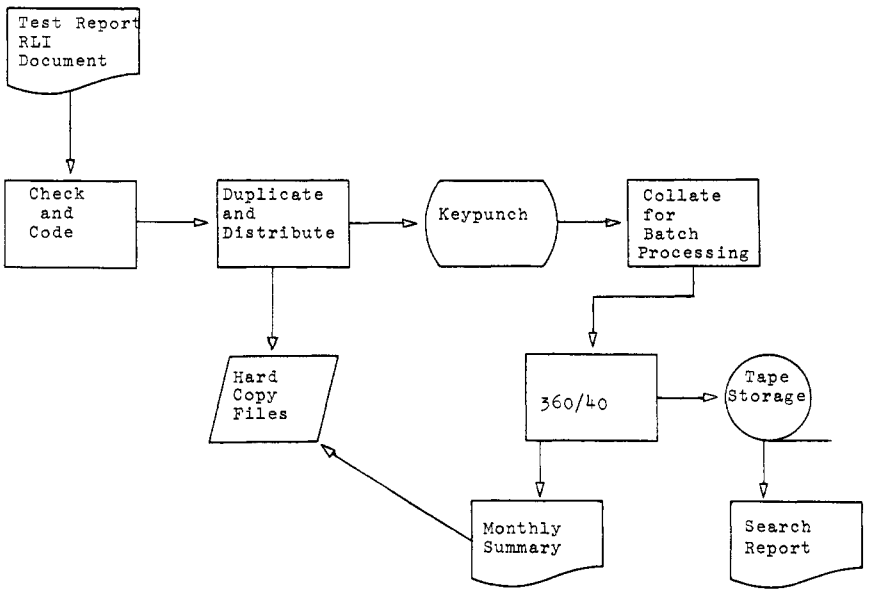

Figure 1. Data flow within technical information center

research reports are coded and keypunched for input into the system. Although some of these reports require technical interpretation and coding, most incoming reports are handled routinely by a clerical staff. Punched card data is verified, collated, and submitted to the computer monthly. An IBM 360 Model 40 processes the data and stores it on magnetic tape. A new master tape file is generated at every up-date and retained in the current tape library. Updated information is printed out in a monthly summary report and filed by compound number for subsequent use.

INPUT CARD FORMAT

Three types of data cards are used for the input to the general storage program. The three card formats are shown in Figures 2, 3, and 4. Type 1 contains chemical data; Type 2 contains scheduling and routing information; and Type 3 contains test report information.

Type 1 cards, or chemical cards, are subdivided into Type 1a and 1b. Data put on these cards are obtained from the RLI form. Type 1a data include compound identification number, requisition number, submission date, notebook page, chemist number, source of compound (if external), and a code to differentiate steroids from other compounds (Figure 2). Type 1b card data include compound identification number, requisition number, and chemical name. A maximum of three Type 1b cards can accomodate long chemical names. Chemical cards are linked by a Code 1 in Column 80. Columns 76 to 79 are used to ensure proper sequencing. Column 75 in Type 1a is used for deletions and corrections. A 2 in Column 75 Type 1b distinguishes sub-types.

Type 2 cards (Figure 3) are "submitted to—not completed" cards. These cards contain additional information from the RLI form, including compound number, chemist number, submission date, and a five-digit code identifying the specific screen requested. Ten test numbers

TYPE 1a (1 card max.)

1-6	SCHERING NUMBER
12-16	REQUISITION NUMBER
17-21	DATE (MO./YR.)
23-26	NOTEBOOK NUMBER
29-31	PAGE NUMBER
33-36	CHEMISTS' NUMBER
37-46	SOURCE OF COMPOUND
73	CODE: 1 IF STEROID
75	CODE: 1 IF CORRECTION
76-77	ACTUAL CARD NO. IN THIS SUBSET
78-79	TOTAL CARDS IN THIS SUBSET
80	CODE: 1 FOR CHEMICAL CARD

TYPE 1b (3 cards max.)

1-6	SCHERING NUMBER
12-16	REQUISITION NUMBER
17-71	CHEMICAL NAME
75	CODE: 2 FOR TYPE 1b CARD
76-77	ACTUAL CARD NO. IN THIS SUBSET
78-79	TOTAL CARDS IN THIS SUBSET
80	CODE: 1 FOR CHEMICAL CARD

Figure 2. Type 1, chemical cards

1-6	SCHERING NUMBER
8-11	CHEMISTS' NUMBER
12-16	DATE (MO./YR.)
22-71	TEST NUMBERS
72	CODE: 1-delete entire area. put info from this card on. 2-delete entire area.
76-77	ACTUAL CARD NO. IN THIS SUBSET
78-79	TOTAL CARDS IN THIS SUBSET
80	CODE: 2 FOR SUBMITTED TO - NOT COMPLETE CARDS

Figure 3. Type 2, submitted to—not complete card

can be accomodated on a card, and as many cards as necessary can be used. A 2 in Column 80 identifies all Type 2 cards. Card Columns 76 to 79 are used to ensure proper sequencing, in the same way as Type 1 cards do. The deletion code (Column 72) permits the erasure of any part or all of card Type 2 information from the tape files. The Type 2 card is used to monitor biological tests pending for a compound. When the test report is received, a Type 3 card with corresponding test number, compound number, and RLI date goes into the file and erases the pending test information.

The biological test report information is punched on Type 3 cards (Figure 4). These cards provide the bulk of input to the files. Two Type 3 cards, 3a and 3b, are utilized for these data. Type 3a has the following information: compound number, test number, animal code, route of administration, test date, RLI date, activity code, notebook page number, requisition number, and salt designation. Type 3b has the same information up to and including the test date, but the remainder of the card contains the biologists' statement of results (Column 32 to 71). This field is used to express the conclusions of the experiment, which can be accomodated on a maximum of 15 cards. Unlike other fields, it exists in free text and cannot be searched for specific parameters. It can only be printed out in its entirety.

Although raw data are not incorporated in retrievable form, an activity scale is used to indicate results (Column 37). Other card formats were investigated in the early stages of development, but it was felt that the activity code and general comment field (Columns 32 to 71) would best reflect the test report results.[2,3] A 1 in Column 72 allows the entry of a special type report. A delete code (74/1) allows for corrections in data. A 3 in Column 80 identifies the card as a biological card. Columns 76 to 79 are used the same way that these columns are used in the chemical cards.

All preclinical research documents (memoranda, reports, letters) are handled in much the same way as test reports. Insertion of a general document number in the test number field distinguishes it from other biological test numbers. Codes describing the document are substituted in place of biological parameters. A Type 3b card describes the content of the document in abstract form. Despite shallow indexing, no difficulties in retrieval have been experienced, since most queries are accessed either by compound number, author, or date.

The new system was structured to provide monthly summary reports on the progress of compounds through

TYPE 3a (1 max.)

1-6	SCHERING NUMBER
7-11	TEST NUMBER
15-16	ANIMAL CODE
20-21	ROUTE OF ADMINISTRATION NUMBER
22-26	TEST DATE (MO./YR.)
30-31	TEST DATE (DAY)
32-36	RLI DATE (MO./YR.)
37	ACTIVITY CODE
42-46	NOTEBOOK NUMBER
47-51	PAGE NUMBER
52-61	REQUISITION NUMBER
62-71	SALT OF COMP.
72	CODE: 1 ENTERING P-REPORT
74	CODE: 1 WHEN MAKING CORRECTIONS
76-77	ACTUAL CARD NO. IN SUBSET
78-79	TOTAL CARDS IN SUBSET
80	CODE: 3 FOR BIOLOGICAL CARDS

TYPE 3b (15 max.)

1-6	SCHERING NUMBER
7-11	TEST NUMBER
15-16	ANIMAL CODE
20-21	ROUTE OF ADMINISTRATION NUMBER
22-26	TEST DATE (MO./YR.)
32-71	BIOLOGIST' STATEMENT
76-77	ACTUAL CARD NO. IN SUBSET
78-79	TOTAL CARDS IN SUBSET
80	CODE: 3 FOR BIOLOGICAL CARD

Figure 4. Type 3, biological cards

```
SCH 010140                          MONTHLY SUMMARY REPORT          MAR 1,1968          SCH 010140

17A-METHYL-1,5-ANDROSTADIENE-3B,17B-DIOL-3-DECANOATE
   REQUISITION      19473     DATE 09/62     NOTEBOOK  1729     PAGE    48     CHEMIST  SHAPIRO

0  ANTIFERTILITY-MATING

     REQUISITION                          SUBMITTED  09/62     TESTED  10/62     NOTEBOOK          PAGE
        RAT          PO          13.3 MGK INACT

0  ANTIATHEROSCLEROSIS OR HYPOCHOLESTEROLEMIA

     REQUISITION                          SUBMITTED            TESTED  11/62     NOTEBOOK          PAGE
        RAT          PO          30 MGK INACT

1  ANDROGENIC-ANABOLIC

     REQUISITION                          SUBMITTED  09/62     TESTED  09/62     NOTEBOOK          PAGE
        RAT          PO          25,100MGK ANABOL ACT 1.8X 3863. ANDROG ACT .84X 3863. RATES OF ANABOL-ANDRO
                                 G ACT IS CA THE SAME AS FOR 10042

MEMO      03/63                  TOLKSDORF-CURRENT STATUS OF SEX STEROIDS + ANABOLICS

MEMO      09/64                  TOLKSDORF- BIOLOGICAL ACTIVITY OF 3- HYDROXY-1,5-ANDROSTADIENES AND 3- HYDR
                                 OXY-1,5-PREGNADIENES.

P-REPORT  07/65                  P-3522-TABACHNICK-EVALUATION OF DECA-DUXRABOLIN AND SCH 10042 AND ITS ESTER
                                  ANALOGUES FOR DURATION OF ANABOLIC ACTIVITY.

SUBMITTED TO ,NOT COMPLETED

   GENERAL HORMONE SCREEN                                                    09/62
```

Figure 5. Monthly summary report

biological testing. These summary reports offer a complete history of each compound under investigation (Figure 5). The printout includes compound identification number, chemical name, test report data with the biologists' statement, and references to research reports. The final section of the summary lists tests requested but not completed as of the report date.

The new system also eliminated previous search restrictions, making it possible to retrieve data selectively on any parameter or combination of parameters. The computer program for such retrieval enables the technical information center to answer the numerous enquiries that refer to, and cut across, the file of 15,000 compounds.

Some examples of this selective retrieval follow:

1. Search of file for all data on compounds tested in one or several screens. This search can be made more specific by any combination of parameters—for example, animal, route of administration, or degree of activity.
2. Report on the number of compounds submitted to a screen for a given time period; this emphasizes the number of compounds tested and the number that remain to be tested. These tallies are of particular interest to management for planning and budget analysis.
3. Search for specific structural types, having certain biological characteristics. This inquiry requires an initial search of the Trematrex system, a separate optical coincidence file of chemically coded structures. The code numbers of the compounds located in Trematrex provide part of the input that guides the computer search. This two-step procedure is not inconvenient because the location of structural types can be accomplished within minutes.
4. Periodic printout of chemists' submissions. Each chemist receives a printout of all compounds he submitted for testing during a given time period, along with pertinent test results. This report program simplifies record-keeping for the chemist and reduces the possibility of overlooked compounds.
5. Information needed for IND-NDA submissions is provided by the monthly summary report. This cumulative index cannot be used directly in license applications to the U.S. Food and Drug Administration, but it provides medical monitors with a format that enables them to determine easily what has been done and what remains to be done before submission can be made.

Thus, the initial objectives outlined for the system have been accomplished. Although present output is satisfactory, the program is being refined to incorporate parameters that better reflect raw data. Integration of the chemical structure file into the biological file is under study.

ACKNOWLEDGMENT

The authors express their appreciation for the technical assistance of Rita Goodemote, Walter Whitman, Elizabeth Bellamy, and the programming contributions of Andrew Tremko, James McGee, and Eileen Maier.

LITERATURE CITED

(1) Ginsberg, Helen F., and Catherine Shea, "Biological Summary Reports on Unit Record Equipment," Presented before the Division of Chemical Literature, ACS, Washington, D.C., March 1962.

(2) Elias, A. W., and M. R. Warren, "A Correlative Indexing and Retrieval System for the Screening of Biological Data," J. CHEM. DOC. 2, 185-9 (1962).

(3) Dietrich, E. V., "Machine Retrieval of Pharmacological Data," *Science* 132, 1556 (1960).