(5) Heller, S. R., and Koniver, D. A., "Computer Generation of Wiswesser Line Notation. II. Polyfused, Perifused, and Chained Ring Systems," *Ibid.*, 12 (1), 55-9 (1972).
(6) Leiter, D. P., Morgan, H. L., and Stobaugh, R. E., "Installation and Operation of a Registry for Chemical Compounds," *Ibid.*, 5, 238-42 (1965).
(7) Oatfield, H., "The ARCS System: Ringdoc As Used with a Computer," *Ibid.*, 7, 37-43 (1967).
(8) Nubling W., and Steidle, W., "The 'Documentationsring der Chemischpharmazeutischen Industrie'; Aims and Methods," *Angew. Chem. Internat. Edit.* 9 596-8 (1970).
(9) Hyde, E., Matthews, F. W., Thomson, L. H., and Wiswesser, W. J., "Conversion of Wiswesser Notation to a Connectivity Matrix for Organic Compounds," *J. Chem. Doc.* 7, 200-4 (1967).

Conceptual Design of an Information System for Toxicological Research and a Partial Implementation*

STEPHEN J. TAUBER,** HERBERT H. DER GIRAGOSSIAN, HAROLD S. CORBIN, JASPER P. HENDREN, and WILLIAM J. LaMARSH†
Informatics Inc., 6000 Executive Blvd., Rockville, Md. 20852

Received April 10, 1972

A design is described for a system to manage the information directly related to the experimental work in a toxicological laboratory. A number of distinct input procedures provide data to a set of co-existing data bases. Periodic, event-triggered, predefined demand, and ad hoc reports are posited. Data must be collected under stringent quarantine conditions. An initial implementation follows animals during experiments for weight, feed consumption, symptoms, and mode of death. A minicomputer, programmed in assembly language, controls automated source data collection. A larger computer, programmed in MARK IV, maintains the master file and produces the reports. The reports cover current experimental events, experimental status, time trends, and detailed data.

This work results from addressing the requirements of a toxicological research laboratory intended primarily for large-scale chronic studies of the effects of very small doses of potential toxicants on experimental animals.

In the first part of this paper we address the conceptual design of an information system to support animal experiments. The implementation of a subsystem is described in the second part of the paper. The information needs which the design seeks to satisfy were presented earlier.[1]

CONCEPTUAL DESIGN

The spectrum of information requirements of the toxicological laboratory, we allocate to three groups: access to existing external technical information, management of information directly pertaining to the experimental work, and management of ancillary information.

Access to external information involves considerations such as maintenance of a technical library and use of public and proprietary search systems, extensively discussed by many authors, including very recently such as Smith *et al.*,[2] Maizell *et al.*,[3] and Reinke.[4] Also included are who's-doing-what files such as are maintained by the Smithsonian Science Information Exchange on a continuing basis[5] and were compiled ad hoc by Vasta.[6]

Ancillary information is that which is only marginally affected by the specific nature of the work—e.g., inventory control and purchasing, building and equipment maintenance records, financial records, and personnel records. Some of these areas scarcely differ from those of other enterprises comparable in size. More peculiar are the special inventory control needed over sterilized supplies and the distinctive shipping procedures for purchased animals.

We restrict our consideration here to management of the information directly related to the experimental work. This encompasses direct experimental data, animal production, colony management, sample control, and scheduling. The design philosophy is to maintain distinct data bases for these several types of information, to use concerted access to them for output, and to distribute information to them from single input procedures in the laboratory.

Input. A protocol, which describes an experiment in detail, governs the laboratory activities. Each activity consists of a specific laboratory procedure with a pre-defined schedule. Changes may be made in the experiment based on preliminary data, reflected by changes in the procedures to be performed.

Data derive from the procedures and give rise to input transactions to the files. In some instances validation of data can be in real time even with batch updating. A skeletal view of the data flow from the major procedures into the data bases is given in the upper portion of Figure 1. The "Edit and Store" function must interpret the input record to determine the files affected, must provide appropriately formatted transactions for each, and must compute values derived partly from prior data.

*Presented at Division of Chemical Literature, 163rd Meeting, ACS, Boston, Mass., April 12, 1972.
**To whom correspondence should be addressed.
†Present address: Defense Communications Agency, Reston, Va. 22070.

Table I lists the data bases affected by and the data items derived from two of these procedures.

Output Requirements. Laboratory information must be supplied primarily to the investigators, the chiefs of animal husbandry, animal care, hematology, pathology, biochemistry, and quality assurance, and the director of operations. Some interactions also exist with the purchasing agent and the chief engineer.

Periodic reports, reports triggered by specific events, pre-defined reports produced on demand, and ad hoc reports are to be produced. The first three categories differ only in the criterion for production, since the program for each such report must preexist. Explicit mention of ad hoc reports is not so much a confession of inevitable omissions as it is an assertion that hypotheses formulated upon prior data presentations trigger a desire for particular data presentations. A schematic indication of reports by content is given by the lower part of Figure 1; Table II lists a few types of reports with the anticipated users, category, and major information content.

Data Bases. The data bases are designed to be the repositories of essentially all data generated in the laboratory. If fully implemented, they would thus be not only the laboratory notebooks of the experimenters but also the quality control and maintenance logs, the vital statistics and genealogy records, and the production records. The several data bases in this design are characterized by the sources of their data and the data contents, as exemplified in Table III.

Hardware Considerations. Some rough calculations were made of transaction activity and file size for a putative 18-month study on 75,000 rodents.[1] The number of data entry transactions was estimated to vary between about 25,000 and 150,000 per week. The experiment data base for this study was estimated to reach 7.8×10^7 characters. If a remote computer is involved, the peak loads during the experiment would occupy a 2400-baud line about one-and-a-half hours per week for weighing and observation data and about three hours per week for laboratory analysis and pathology results (provided variable-

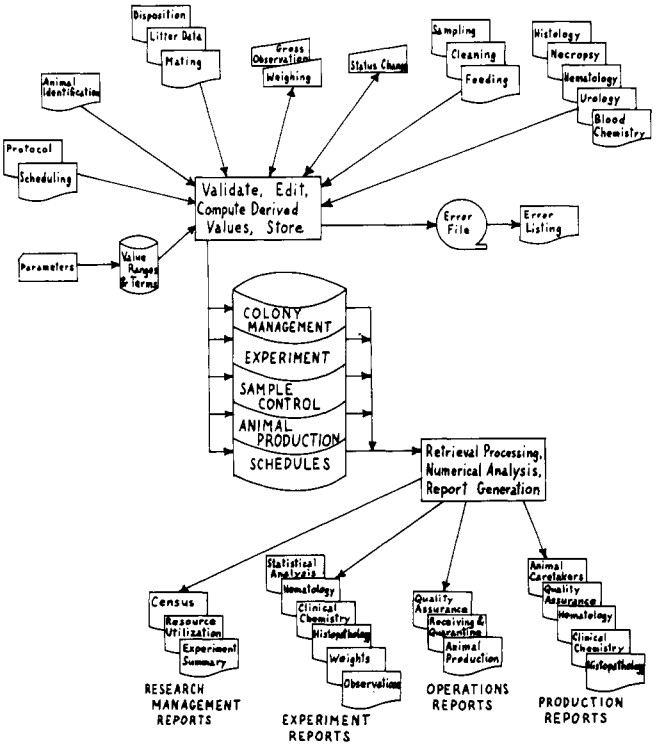

Figure 1. Data flow from input procedures to data bases to reports

Table I. Data Bases Affected by and Data Content of Selected Procedures

Procedure	Affected Data Base(s)	Data Content
Animal identification (The procedure that converts an undifferentiated animal from a litter or quarantine state to that of a usable, identified, individual animal.)	Colony Management Animal Production	Date Animal ID Source Location (barrier, room, and cage numbers) Dam ID Sire ID Date born Date weaned Weight Ear marking Sex Strain
Clinical biochemistry	Sample Control Colony Management Experiment	Date Animal ID Sample ID Sample type Procedure requested Blood sugar value Blood urea nitrogen value Total serum protein value Total serum bilirubin value Serum albumin value Serum sodium value Serum potassium value Serum chloride value Carbon dioxide value Serum calcium value Serum alkaline phosphatase value Serum glutamic-oxaloacetic transaminase value Remarks

Table II. Examples of Reports

Type of Report	Intended Users	Category	Data Content
Receiving & quarantine summary	Director of Operations Animal Husbandry Division Quality Assurance Laboratory Investigators	Periodic	Period of report Animals received, by strain Animals dispensed Total animals in R & Q Incidence of contamination, by strain Baseline blood values, by strain
Breeding production	Director of Operations Animal Husbandry Division Investigators	Periodic Pre-defined demand	Period of report Strain Number of breeders Productivity (no. weaned/no. breeders) Survival Index (no. weaned/no. born) Mean litter size Mean litter weight at weaning
Feed assay	Investigators Chief of Animal Caretaking Purchasing Agent	Triggered by completion	Feed type Lot number Nutrient contents Test compound assay

Table III. Examples of Data Bases

Data Base	Sources	Contents
Animal production	Mating Litter data collection Litter disposition Animal identification Status change Weighing	Species and strain Source of the animals Dam and sire identification Location of animals Date of mating Size of litter Numbers born and weaned, by sex Date of weaning Wean weight Weanling identification Disposition of weanling
Experiment	Protocol definition Status change Feeding Weighing Gross observation Hematology Urology Clinical chemistry Necropsy Quality assurance sampling	Experiment and investigator identification Start and finish dates Measurements and observations list and schedule Test compound name Species and strain Dose group identifiers Dosages Animal identification

length character string streams were transmitted). The high-volume routine laboratory procedures are the obvious candidates for automated digital data collection, along the lines discussed by others.[8,9]

The key characteristic distinguishing a toxicology laboratory for long-term study from many other environments is the stringent protective measures necessary to prevent infective contamination of the experimental animals.[10] This requirement applies not only to feed and bedding material coming in contact with the animals, but also to all other supplies and equipment brought into the animal rooms. Thus not only high volume, routine data collection but also the problems associated with sterilizing paper or magnetic media argue for source data automation in a mode where tape, punched cards, platen typewriters, and cassettes do not enter the barrier rooms. Similar reasoning suggests a separate data input station for each animal room even when the work load would permit use of mobile data stations.

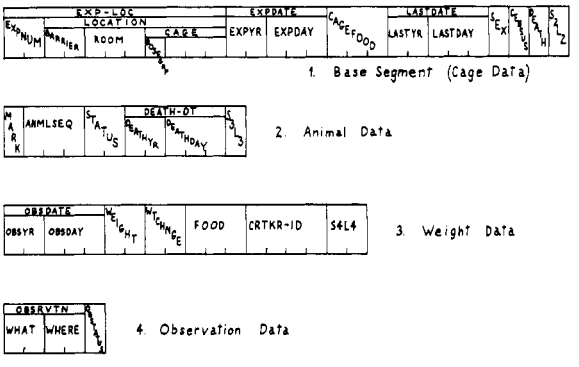

Figure 2. Format of the master file record segments

IMPLEMENTATION FOR WEIGHTS AND OBSERVATIONS

The initial implementation follows animals from the time that they are placed on experiment until they are removed, covering weighing, observation, and status change.

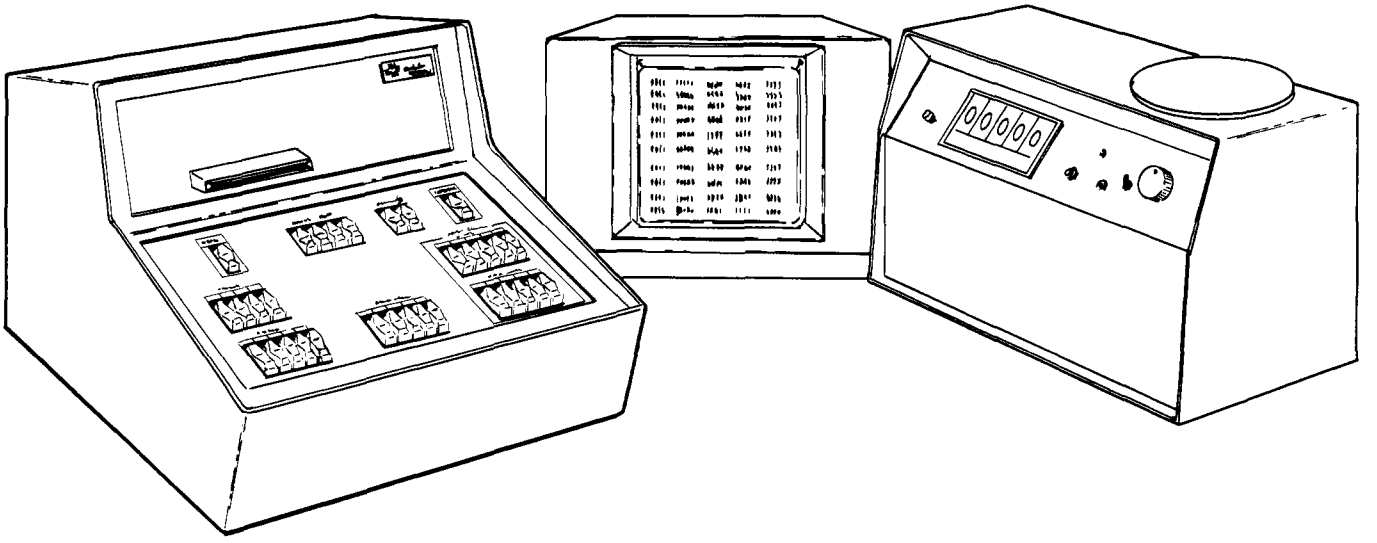

Figure 3. Data station. Badge reader and keyboard on left, cathode ray screen in center, electronic balance on right (Courtesy Modular Computer Systems)

File. The file structure is hierarchical, organized primarily by experiment number and location. Location was chosen rather than dose group and animal sequence number because all data are collected primarily by location. One record pertains to one cage. The second level contains information about an animal, the third level about one weighing, and the fourth level about the gross observations made at the time of weighing.

The layout of the specific data items included is shown in Figure 2, each space indicating one 8-bit byte. Julian dates are used. EXPDATE is the date that the animals were placed on experiment; LASTDATE is the date for which weight and observation data were last entered. Weights are stored as packed decimal numbers; all other data use 8-bit characters. The total food consumption for all of the animals in the cage determined during the most recent weighing is stored in CAGEFOOD. The average is attributed to each animal in FOOD. DEATH is a flag which is set when an animal's death is entered into the file. The computation of the average food consumption allows for the number of days the decedent partook. The flag is reset when this computation has been made. CENSUS contains the count of animals alive in the cage. S2L2, S3L3, and S4L4 are counters for the number of occurrences of the next-lower-level segments. STATUS contains a code either for "on experiment" or for the mode of the animal's death. WTCHNGE contains the difference between two successive values of WEIGHT, and it may have a negative value. OBSTATUS is a flag set for the first occurrence of an observation for a given animal or for an observation which subsequently disappeared.

Hardware. The configuration for this initial implementation consists of data stations, a 16,000-word, 16-bit Modcomp III computer, with a 1,200,000-word disc, card reader, line printer, and communications interface, and a dedicated 2400 bit-per-second line to a remote IBM 360/50 installation. The minicomputer has the task of data capture, validation, and formatting. The larger, remote computer is given the file maintenance and report generation tasks.

The data station consists of a 10-column badge reader, a custom keyboard, an electronic balance, and a 400-character CRT unit. The arrangement of the data station is shown in Figure 3. Badges are designated for use as cage cards, animal caretaker identification cards, and experiment designations. The balance is for recording the weights of feed jars and of experimental animals. The keyboard is designed for entry of gross observations, mode of death, animal identification, and animal counts. The CRT unit is for data feedback, procedural reminders, and error messages.

Although each data station is identified in the hardware the several components must be distinguished by software. Programming for the minicomputer is in assembly language.

Input. Detailed procedures have been worked out for entry of data from the input stations, such as indicated in Table IV. The meanings assigned to the keys on the data station are listed in Table V.

The assignment procedure supplies the experiment number, the locations, and the number of animals placed into each cage. The sex code for the animals is generated from the low order bit of the cage number. One C01 transaction is generated, and one C12 transaction generated per animal in the cage (cf. Figure 4). The barrier and room numbers are supplied by the data station identification. The internal computer date becomes the date that the animals are placed on experiment. The status in the C12 transaction is invariably "one experiment."

The normal weighing and observation procedure provides the weight of feed consumed, computed by difference from the standard gross weight of a feed jar, in an F11 transaction. All observations on one animal and its weight are entered in a compact group, terminated by pushing an ear mark key (cf. Table VII). The mark, weight, date, and caretaker identification (stored into the computer at the beginning of the session) go into an I23 transaction. Each observation is coded into an I34 transaction. "No feed" or "no water" is entered into an I34 transaction for each animal.

The software enforces various features of the procedure by refusing to accept data out of sequence or out of range. For example, a weight appropriate to feed jars must arrive before any observation or a weight in the mouse range can be entered. At least one observation must be entered for each animal. It is also impossible to wander from the assignment procedure into the weighing and observation procedure or vice versa.

A death is coded into an F22 transaction. Unless the mode of death is "missing animal," a weight and at least one gross observation must also be entered.

Table IV. Input Procedure for Reporting Animal Found Dead

Procedure Step	Comments
1. Bring the cage of the dead animal to the data station.	1. Reduces chance of separating cage and cage identification card.
2. Activate equipment. Insert the operator's identification card into the badge reader. Remove ID card.	2. Records the caretaker who found the dead animal.
3. Place the cage card of the dead animal's cage into the badge reader.	3. Identifies the animal's cage, barrier, and room.
4. Enter the final observations on the animal via the function codes in the keyboard.	4. One observation will be "found dead."
5. Place the animal onto the balance.	5. Records the animal weight.
6. Push ear mark key.	6. Identifies the animal.
7. Tag the dead animal for laboratory analysis.	7. Animal leaves subsystem.
8. Remove the cage card from the badge reader.	8. Signals completion of the observations.
9. Return cage with remaining live animals, or remove empty cage.	
10. If there are additional deaths to record repeat from step 3. Else log off by inserting and removing operator's identification card, then turn off equipment.	

Table V. Designations for Data Terminal Keys

No Ears Clipped *or* 1	Salivating
Both Ears Clipped *or* 2	Labored Respiration
Left Ear Clipped *or* 3	Soft Feces
Right Ear Clipped *or* 4	Nasal or Occular Discharge
Moribund Sacrifice	Rough Hair Coat
Serial Sacrifice	Sores
Terminal Sacrifice	Alopecia
Found Dead	Swelling
Missing Animal	Tissue Mass
No Water	Entire Animal
No Feed	Head
	Body
Appears Normal	Legs
Circling	Tail
No Equilibrium	
Paralysis	Corrections
Tremors	
Hunched	

Use of the correction key cancels the data for an entire cage, if the cage card is next withdrawn. The observations for an animal are rescinded by pushing the appropriate ear mark key after the correction key. An animal weight cannot be cancelled without starting afresh for the entire cage. Removal of the cage card (without use of the correction key) signals the end of data for that cage. The transactions are then finalized and stored onto disc.

Provision has been made for entry of data via punched cards. Punched card entry is intended primarily to correct the file. Delete transactions (cf. Figure 3) can be created only in this way. The transaction designations differ from one another in at least two characters in order

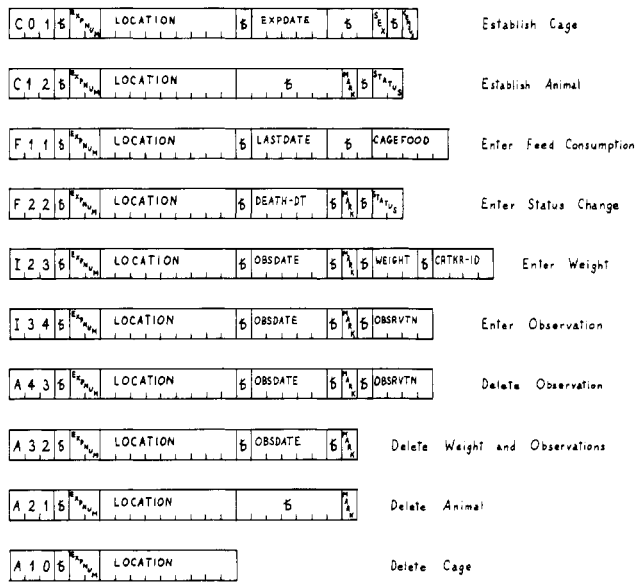

Figure 4. Transaction formats

to decrease the probability of punching a wrong but valid transaction designation.

Activating a terminal automatically opens a transaction file. Files must be closed from the computer console; a new file is immediately opened for each console active at the time. A separate file is created for transactions entered from the card reader. Transmission of the transaction files must also be initiated from the console; only the closed files are transmitted.

The transactions are processed in the remote computer by MARK IV, a generalized file management system.[11,12] The transactions for any given record are processed in alphameric sort order. The transaction designations were selected in such a way as to force delete transactions to be processed first in order to clear the way for correct data. Likewise a segment would be created before an attempt is made to update it.

Beyond prevention of updating nonexistent records or segments and of creating identical records or segments twice over, the following validity checks have been provided:

No ear mark code value will be accepted which exceeds the number of animals placed in the cage. No status other than "on experiment" will be accepted in a C12 transaction. A mode of death is accepted only if the status in the file is "on experiment." The observation "appears normal" is accepted if and only if no abnormal observation has been entered with the same date for the same animal.

Reports. Twenty-one reports have been designed to provide oversight to the experimenters and the director of research (cf. Table VI). The daily and weekly reports provide current events information, the monthly ones (actually tetraweekly) are status reports. The demand reports are intended, with one exception, to support detailed analysis.

Deaths of animals, new observations, and disappearing observation reports identify the animals by dose group and sequence number as well as by location and ear mark, and they give the sex and the dates the animals were placed on experiment. The deaths report gives the mode of death (cf. Table V) and the number of days that the newly dead animal was on experiment. The new observations report lists each observation (cf. Table V) newly made on any given animal. Observations previously

made on an animal but lacking during a subsequent weighing and observation cycle appear in the disappearing observations report. The number of days the animal has been on experiment is given; in the latter report, the date the observation was first reported and the number of days then on experiment are also given. When an abnormal observation is first reported for an animal it is listed under "appears normal" on the disappearing observations report.

If some but not all animals supposedly in a cage are weighed, the discrepancy appears on the omitted animal report. This may be the first indication of cannibalism or escape. The daily and weekly productivity reports list the number of cages and animals processed by each animal caretaker.

The missing weights and observations report lists the cages for which weights and observations were not recorded during the week, or other period specified at run time, although they supposedly contain live animals.

The weight loss and weight anomaly reports list animals for special attention because of weight data. For young animals (under about 13 weeks old) only a loss of weight is reported. An older animal is listed either if its weight underwent a large change or if its weight, weight gain, or the feed consumption averaged over the cage lies far from the mean for its contemporaries in the same experimental cell. The extent of change can be specified by run time parameters. The deviation above the mean is stated in terms of a percentage and seeks primarily weighing errors. The low-side deviation is stated in terms of standard deviations and seeks animals showing early symptoms of illness. These two reports list the animal identification, sex, weight, weight gain, attributed feed consumption, the date placed on experiment, and the number of days on experiment.

Table VI. List of Reports

Frequency	Report
Daily	Deaths
	New Observations
	Disappearing Observations
	Omitted Animals
	Productivity
Weekly	Weight Loss (Up to 8 weeks)
	Weight Anomaly (Over 8 weeks)
	Weekly Productivity
	Missing Weights and Observations
Monthly	Survivor and Loss Counts
	Modes of Death Count
	Weight Data Summary
	Observations Summary
	Facility Census
Demand	Animal Directory
	Weight Trends
	Weight Gain Trends
	Feed Consumption Trends
	Observation List
	Mode of Death List
	Individual Animal Data

Table VII. Parameters for Individual Animal Data Report

Item limit override
Experiment number (mandatory)
Dose group and animal sequence number within dose group
Location and ear mark
Dose group
Location only
Sex
Date animal was placed on experiment
Experimental status
Minimum and/or maximum body weight
Minimum and/or maximum weight gain/loss
Minimum and/or maximum feed consumption
Observations
Occurrence of disappearing observations
Name of requestor

The status of each experiment is provided in four of the "monthly" reports. Each of these is organized by the number of weeks the animals have been on experiment, sex, and dose group. The survivor and loss report states the number of animals originally assigned, the number of deaths during the current reporting period and, cumulatively, and the number of survivors. The modes of deaths count gives the breakdowns of the death counts. The weight data summary presents the maximum, minimum, and mean values of the weight, weight gain, and feed consumption as of the last weighing. The observations summary lists the numbers of animals for which each observation has been reported to date, with a distinction made for different parts of the body (cf. Table V); multiple counting occurs for an animal which shows more than one symptom.

The census report is organized by animal room and lists by number of weeks on experiment the original numbers of animals and the numbers of survivors. This provides the basis for extrapolating space availability.

The animal directory simply gives for each animal the correspondence between dose group and sequence number on the one hand and location and mark on the other. All animals assigned to an experiment still active are listed, with sex, date placed on experiment, and, where appropriate, date of death. The cage and live animal counts are listed for each experiment.

The trend reports tabulate for each of the dose groups weight, weight gain, or feed consumption data cumulated according to the number of weeks that the animals had been on experiment at the time of the data. The means, the upper and lower 2σ limits, and the number of animals providing the statistics are displayed. The observation and mode of death lists enumerate each animal tallied into the observations summary and modes of death count, identified by dose group and sequence number, with its sex and the date placed on experiment. The observations list states the date the observation was first made in each animal, the number of days it had been on experiment, and its current experimental status. If the observation has disappeared or if the animal has died the corresponding date is given. In the modes of death list the date of death and the number of days on experiment at the time are given.

The most complete report is that on individual animal data. For those animals appearing in the report all data in the file are presented except the animal caretakers' identifications. Either the number of days the animal has currently been on experiment is shown or the number of days on experiment at death. Every weight, weight gain, and feed consumption value is shown, the date of first observation for each symptom, the number of days on experiment at the time, and if appropriate the date of disappearance. Selection criteria must be provided at run time to determine which animals to include in the report (cf. Table VII). These, plus the name of the requestor, appear on the report. If individual animal identifications are used then the other selection criteria are not available. The location may be specified as one room or as one series of cages within the same room per dose

group. Minimum and maximum values act entirely independently, and only one file value need satisfy each criterion. Thus a weight from an early weighing might satisfy a low body weight maximum and a weight from a later weighing satisfy a higher minimum. For the experimental status there is the option of "on experiment," a specific mode of death, or NOT "on experiment." Similarly there is the option of no abnormal observation, or any abnormal observation. With the exceptions stated the parameter values are AND'ed in defining the selection class.

Requests for the reports must be entered from the computer console, together with any parameters. Distinct individual animal data reports, with different sets of selection criteria, are permissible. Initiation of data transmission from the console automatically closes the report request file. Normally all closed transaction files are to be transmitted with report requests. However, the possibility is designed in to restrict transmission of transactions to those entered from the card reader. This is to permit rerun of some reports based on corrections to the previous day's data. All transactions and report requests transmitted together are processed in the same computer run.

Stage of Implementation. This implementation has not yet been installed, and no actual run characteristics can be cited.

ACKNOWLEDGMENT

This paper presents results of work performed under contract FDA 71-334 with the Food and Drug Administration. The conceptual design draws on discussions with staff and observations of operations at a number of animal laboratories. We wish particularly to thank Carl Holsing and Willard Weatherholtz (Hazleton Laboratories), Ezra Berman (Environmental Protection Agency), Richard Pierson (National Institutes of Health), John Storer and Louis J. Serrano (Oak Ridge National Laboratory), H. C. Redman and A. J. Hulbert (Lovelace Foundation for Medical Education and Research), and Marvin Goldman and Jack Whitmier (University of California at Davis). We similarly wish to acknowledge the valuable discussions with A. Kennon Davis, Morris Cranmer, and Stephen Douglas of the National Center for Toxicological Research.

LITERATURE CITED

(1) Tauber, S. J., Der Giragossian, H. H., and Weatherholtz, W. C., "The Information Demands of a Toxicology Research Center," 7th Middle Atlantic Regional Meeting, ACS, Philadelphia, February 14, 1972.

(2) Smith, D. R., Beauchamp, R. O., Garber, J. L., and Daugherty, M. A., "Computerized Drug Information Services," *J. Chem. Doc.* **12,** 9-13 (1972).

(3) Maizell, R. E., Henderson, R., and Gazda, E. S., "Information Retrieval in Environmental Science," 7th Middle Atlantic Regional Meeting, ACS, Philadelphia, February 14, 1972.

(4) Reinke, R. E., "An Indexing Coverage Study of Toxicological Literature," 7th Middle Atlantic Regional Meeting, ACS, Philadelphia, February 14, 1972.

(5) Kreysa, F. J., "SSIE—An Information Center Which Stores Foresight," *J. Chem. Doc.* **12,** 19-22, (1972).

(6) Vasta, B., in B. Murray, chmn., "Report of COSATI ad hoc Task Group on Chemical Information Systems," unpublished report, September 1971. (Mrs. Barbara Murray, Bldg. 37, National Institutes of Health, Bethesda, Md. 20014.)

(7) Castleman, P. A., Briscoe, H. W., Butterfield, S. C., Frazier, F. S., Heart, F. E., Hollister, C. A., Russell, C. H., Webb, F. N., Raub, W. F., Kaufman, S. J., and Oxman, M. A., "The PROPHET System—A Final Report on Phase I of the Design Effort for the Chemical/Biological Information-Handling Program," National Institutes of Health, Bethesda, Md., 1970, p. 27 ff.

(8) Yeransian, J. A., Foltz, A. K., and Kirschman, J. C., "A Realtime Automation System—One Year Later," *Advances in Instrumentation* **26,** paper 733 (1971).

(9) Poletti, B. J., Zack, J. F., and Mueller, T. J., "Computer Control in the Clinical Laboratory," *Amer. J. Clin. Path.* **53,** 731-8, (1970).

(10) Serrano, L. J., "Defined Mice in a Radiobiological Experiment," pp. 13-43, in "Defining the Laboratory Animal," National Academy of Sciences, Washington, D. C., 1971.

(11) "MARK IV File Management System—User's Guide," Informatics Inc., Canoga Park, Calif., 1969.

(12) "MARK IV File Management System—Special Features," Informatics Inc., Canoga Park, Calif., 1971.