AD646666

[redacted]

[redacted]

THE MAPPING OF DISEASE (MOD) PROJECT

THE GEOGRAPHIC DISTRIBUTION OF INFECTIOUS DISEASES

DA 49-092-ARO-130

FIRST ANNUAL REPORT
(15 NOV 1965–14 NOV 1966)

BEST AVAILABLE COPY

15 DECEMBER 1966

20040901152

MOD

COMPUTERIZED MAPPING OF DISEASE

UAREP / AFIP (Geographic Pathology Division)

[redacted]

COMPUTER MAPPING OF DISEASE - PROJECT MOD

THE GEOGRAPHIC DISTRIBUTION OF INFECTIOUS DISEASE

First Annual Report (15 Nov. 65 - 14 Nov. 66)

15 Dec. 1966

Order Number: (761)

Program Code No.: (6G20) (25)

Name of Contractor: U.A.R.E.P.

Date of Contract: 15 Nov. 1965

Amount of Contract: $83,734

Contract Number: DA 49-092-ARO-130

Contract Expiration Date: 15 Nov. 1966

Project Scientist: Harlan I. Firminger, M. D.

Associate Scientist: Howard C. Hopps, M. D.

Title of Work: Geographic Distribution of Infectious Diseases

The work reported herein was performed as a cooperative venture between the Universities Associated for Research and Education in Pathology, Inc. (UAREP), Bethesda, Md. and the Armed Forces Institute of Pathology (AFIP), Division of Geographic Pathology, Washington, D. C. The former organization (UAREP) operated under a contract from the Advanced Research Projects Agency (ARPA) through the Army Research Office (ARO).

TABLE OF CONTENTS

Summary of Progress	3
List of Figures	4
Introduction	5
Acknowledgements	6
Objectives, Background, and Scope of the Project	7
The Total Program Divided into Stages	14
Data-Management Considerations:	17
- collection (sources)	17
- extraction	18
- preprocessing	21
Data-processing Considerations	23
- computer equipment requirements	23
- computer program requirements	24
- personnel requirements	29
- plotter equipment considerations	31
- output considerations	33
Conclusions and Recommendations	39
Bibliography	40
Fiscal Data	41
Appendix 1: Organizations consulted during first year of MOD Project	1A-1 (44)
Appendix 2: General data-analysis vocabulary	2A-1 (48)
Appendix 3: Factor catalog	3A-1 (57)
Appendix 4: Plotter survey	4A-1 (66)

SUMMARY OF PROGRESS

The accomplishments of e first year's work of the three year program, Computerized Mapping of Disease (MOD), can be summarized as follows:

(1) The broad outlines of the MOD computerized system have been clearly defined, along with requirements of equipment, programs, and personnel.

(2) A general data-analyses vocabulary, a detailed factor catalog, and a preliminary data-extraction/collection form have been developed.

(3) A large file of disease (leptospirosis and hemorrhagic fevers) and related environmental data has been collected for use in the data-extraction/processing efforts.

(4) Prototype disease-distribution maps have been produced, both manually and by a computer/plotter system.

* * * * *

We believe that the project is progressing well and that efforts to complete it should follow the recommendations included in this report.

LIST OF FIGURES

Fig. 1. - Types of relationships among disease and environmental factors. 8

Fig. 2. - Basic parts of the MOD Project (system). 9

Fig. 3. - Personnel involved in and internal organization of the MOD Project, Nov. 1965 - Nov. 1966. 13+

Fig. 4. - Schedule of major phases and tasks comprising the MOD Project, Nov. 1965 - Nov. 1968. 14

Fig. 5. - Projected personnel and internal organization requirements for successful completion of MOD Project. 16

Fig. 6. - A simple disease-data form for use in extracting data on leptospirosis from appropriate data sources. 21+

Fig. 7.- Data-collection form for use in locating already-existing maps which show the distribution of various environmental factors. 21+

Fig. 8. - Possible sequence of events for providing data for the MOD computer system. 23

Fig. 9. - Organization of MOD computerized disease-mapping system, as presently conceived. 24+

Fig.10. - Possible MOD computer-system applications. 29

Fig.11. - Possible graphic display which might be generated, using the MOD system's data file(s). 34

Fig.12. - Some actual disease data presented as a dot-type map, drawn manually. 36+

+ indicates that the figure follows the indicated page.

Fig. 13. - Examples of computer-produced dot-type maps. 36+

Fig. 14. - The same disease data as Fig. 12, but presented as a shading-type map, drawn manually. 36+

Fig. 15. - An example of a shading-type map produced by a computer. 36+

Fig. 16. - The same disease data as Fig. 12, but presented as a contour-type map, drawn manually. 36+

Fig. 17. - Examples of computer-produced contour-type maps. 36+

Fig. 18. - The same disease data as in Fig. 12, but presented as a manually-drawn map utilizing dot-, shading-, and contour-mapping techniques. 36+

Fig. 19. - Example of a computer-produced map using shading and contour techniques. 36+

Fig. 20. - The same data as in Fig. 12, presented as a contour-type map, drawn by a computer/plotter, using the CDC gridding/contouring program with a coarse grid, including both reported and estimated data points 36+

Fig. 21. - Similar to Fig. 20, but using a fine grid, and including only reported data points. 36+

Fig. 22. - Similar to Fig. 20, but using a fine grid, and including both reported and oceanic data points 36+

Fig. 23. - Similar to Fig. 20, but using a fine grid, and including both reported and estimated and both coastal, and oceanic data points. 36+

INTRODUCTION

During the period 15 November 1965 - 14 November 1966, the Universities Associated for Research and Education in Pathology, Inc., (UAREP) and The Armed Forces Institute of Pathology (AFIP) completed their first year's effort on the Mapping of Disease (MOD) project entitled, The Geographic Distribution of Infectious Diseases. This project was developed and programmed as a three-year effort (Nov. 1965 - Nov. 1968). Consequently, the present report, dealing with the first year's accomplishments, concentrates on development of concepts, methods of approach, and specific software/hardware requirements. There have been important "output" achievements, but these are of primary interest because they represent prototypes rather than finished products.

In addition, the report includes detailed plans and recommendations for work during the next two years which should lead to successful completion of the MOD project.

ACKNOWLEDGEMENTS

The Mapping of Disease Project, since it has broad multi-disciplinary coverage, is dependent upon and grateful for aid received from many, many persons and organizations. The following groups deserve special mention:

First, UAREP acknowledges with gratitude the financial grant, administered through the Army Research Office, from the Advanced Research Projects Agency, which made possible the project.

Second, those of us directly working on the MOD Project acknowledge a deep debt of gratitude to the administrative and professional staffs of both UAREP and AFIP. The resources of these organizations have been given without stint and have been invaluable.

Third, the Mapping of Disease Project has been immeasurably aided by efforts of the Planning Research Corporation (PRC). In particular, we thank Dr. J. Morenoff, Mr. J. Ferguson, and Mr. W. Richmond, all of PRC's Washington Office.

Fourth, we express our great appreciation of the interest and co-operation of the Center for Zoonoses Research and the Department of Geography, both of the University of Illinois (Urbana), of the Walter Reed Army Institute of Research, and of the Communicable Disease Center, USPHS (Chambley, Ga.).

Finally, we wish to thank the various members of the many other organizations - governmental, educational, and industrial - from whom we have received advice and counsel. (See Appendix for specific names.)

OBJECTIVES, BACKGROUND, AND SCOPE OF THE PROJECT

The Computerized Mapping of Disease Project has two principal objectives:

The ultimate objective of the program is to develop research techniques by means of which the occurrence of a particular disease may be correlated with a variety of sociological, physical and environmental factors such as population density, races, ethnic groups, altitude, temperature, humidity, character of the soil, agricultural products, possible insect vectors and animal reservoirs of disease, to give new insight into cause/effect relationships and to suggest new methods of disease control. An important potential application of the technique is in predicting the likelihood that a given disease will develop in a particular area under specific conditions of ecologic change, also in predicting major variations in prevalence, e.g., anticipating an epidemic.

The immediate objective of the program is to provide data in the form of disease distribution maps and atlases, showing prevalence, incidence, and severity of specific infectious diseases throughout the world along with the distribution of actual and potential causally related factors.

By using a computerized system of analysis and output, it will be possible to produce distribution maps in a matter of minutes rather than months, as has previously been the case. This will allow up-dating whenever required. Furthermore, such a system will permit the production of many more maps than would otherwise be practical, covering a wide range of ecologic factors. As desired, these could be printed on transparent stock suitable for overlay assembly in order to compare one pattern of distribution with another, etc.

Data on the geographic distribution of infectious disease are of obvious importance in evaluating the disease risk for groups of persons assigned to foreign posts and in any detailed planning that involves the socio-economic problems of a particular area. There have been only two major contributions in this field and these are now seriously out dated. They are: (a) Geographic Atlas of Disease, prepared by the American Geographical Socity, published during 1950-55. (b) World-Atlas of Epidemic Diseases, edited by Professor Ernst Rodenwaldt (Heidelberg), published in 1952 but reflecting data gathered years before. Neither of these efforts involved modern data storage/retrieval/processing methods, nor did either of them attempt to relate ecologic factors to disease in a manner that would allow detailed cause/effect analysis.

The Division of Geographic Pathology and the Registry of Geographic Pathology, both of the Armed Forces Institute of Pathology have an intense interest in the geographic distribution and manifestations of disease and have had much experience in this field. Until 1 April 1965, the National Academy of Sciences - National Research Council was responsible for administrative and fiscal matters pertaining to the conduct of the American Registry of Pathology. Although these responsibilities have been transferred to a non-profit organization known as Universities Associated for Research and Education in Pathology, Inc., the Chief of the Division of Geographic Pathology at the Institute (H.C.H., Associate Scientist of the MOD Project) continues to act as Registrar of the Registry of Geographic Pathology of the American Registry of Pathology. This allows access to a great deal of disease information from world-wide sources. The Division has also had considerable experience in gathering, storing, and retrieving medical data.

The MOD Project represents the first serious effort (to our knowledge) to develop a computerized disease-mapping system coupled with a comprehensive data-file of ecologic factors. If successful, the system will provide an important research tool to determine complex cause/effect relationships among disease and environmental factors.

The kinds of relationships between disease and environmental factors with which this project deals: causal, associative, or accidental, are shown in figure 1.

Et ⟶ Dis	Direct cause/<u>effect</u> relationship
Et ⟶ Et' ↓ Dis	Indirect cause/<u>effect</u> relationship
Et ⟶ Et, Et ⟶ Dis	Constant <u>association</u> of Et/Dis because of an underlying COMMON cause
Et ---- Dis	Accidental (inconstant) <u>association</u>

((Et = ecologic factor; Dis = disease))

FIGURE 1.

There are three basic parts to the MOD Project (system) and these intimately related to each other in the sequence shown below.

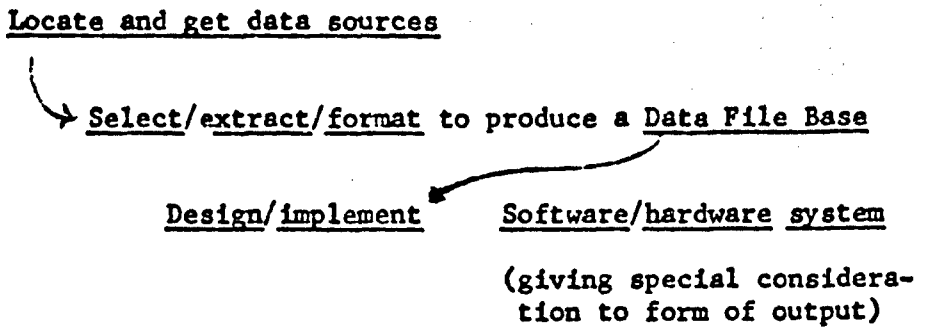


FIGURE 2

The Data File Base is, obviously, an essential (key) ingredient of the system since it provides the substance upon which the software/hardware components act. We realize full well the difficulties in getting adequate (comprehensive and reliable) basic data and know that we can never achieve perfection here. However, our own knowledge and experience, supplemented by information obtained from personal contacts with many experts throughout the world, coupled with that information to be found in scientific papers and reports will, we believe, give us the most effective data base which has yet been elaborated.

A fundamental and essential consideration in developing the MOD system is the conversion of narrative or tabular data to a form in which it can be computer processed. This requires rigid specification of the form in which the data is to be input. Not only must the data be processable, it must be mappable. Further, the data must be selected/extracted so that it is relevant to and significant for the desired output.

Compromise is inevitable in selecting the form of the input. A natural or problem oriented language would be easier for the data processor to use whereas a fixed-form input format would (probably) be easier for the computer to handle. A proper compromise is one in which the kind of language input format developed best suits the total procedure. Arriving at this proper compromise represents a critical step since an inappropriate selection would lead to much delay and costly duplication of effort.

One of the most pressing problems to be solved at this time concerns input and has to do with quantitative aspects: the measure of the disease in relation to the size and selectivity of the population sample. There are many other essential factors, of course, but the measure of disease is primary.

Assuming that an operable data processing system is developed as a result of our efforts, the true evaluation of its potential as a technique will be dependent upon and limited by the quality of the data which is input for processing. Furthermore, achievement of our immediate objective, "mapping of disease", is as dependent upon the data as it is upon the processing system.

We emphasize the importance and difficulty of the research effort necessary to select/extract/format "raw" data in order that it can be computer processed and output in the form of distribution maps. Without adequate and properly formated data, significant output is impossible. The old term GIGO expresses the situation well: garbage in/garbage out.

This aspect of the problem will be considered in detail under "Data-management Considerations.

* * * * * * *

In order to accomplish the goals of MOD it was obviously necessary that we restrict our efforts, and we have assumed certain self-imposed limitations:

(1) Of many many possible diseases for study we selected two: leptospirosis and the hemorrhagic fevers. These were carefully chosen for several reasons: (a) they are important diseases; (b) we (A.F.I.P.) know a good deal about them; (c) high-reliability laboratory diagnosis is possible; (d) they are wide-spread in distribution, but not completely diffuse; (e) more reliable distribution maps are badly needed; (f) each of the two diseases poses specific data processing challenges in relation to important ecologic factors, examples of which include --

Leptospirosis: Involves many mammalian reservoirs, both domestic and wild.

Is greatly influenced by the amount and nature of surface water, including pH, mineral content, rate of evaporation, etc.

Prevalence is greatly influenced by occupational and/or recreational habits of human beings.

Severity varies markedly, depending upon serotype (and many other factors).

Hemorrhagic Fevers: Are often (some types) sharply limited by highly restrictive ecologic factors.

Are arthropod borne for the most part

Manifestations are greatly influenced by age and by race and by the specific causal virus.

(2) Of the enormous number of ecologic factors which could be studied, we are limiting our studies (data collection) to the more current reports (where feasible) and to the kind of data which experience and reason indicates will probably be most significant. Obviously, data collection/analysis can be extended into new areas or different time periods should the need arise.

In general, four different types of data sources are available to the MOD Project:

1. Published prose summaries: monographs, books, journals, technicial notes.

2. Unpublished prose summaries: progress reports, laboratory reports, letters, oral communications.

3. Unpublished raw data: IBM cards, field notes, various completed data-collection forms filled out by other (non-AFIP-UAREP) organizations.

4. Published and unpublished maps.

These represent an extremely large quantity of potentially useful data - much more than we can hope to assimilate. We believe that we have been realistic in limiting our data collection to that which seems most pertinent, concentrating on the most recent. We have not forgotten that the primary goal of the MOD Project is the development of a computerized disease-mapping system rather than a comprehensive collection of data. But the system must have substance to work upon. The computer processing activity is but one side of the coin; an adequate data file base is the other.

(3) The geographic areas selected for mapping have been (tentatively) limited to three distinctly different scales.

(a) the world, per se, - presenting a small-scale over-all view;

(b) Thailand - presenting a medium-scale view; and

(c) a portion of the "Quadri-county" area of Southern Illinois - presenting a large-scale detailed view.

This latter region was chosen because of the intensive ecologic and zoonotic studies which have been going on there for the past several years as part of a major interdepartmental research program of the

University of Illinois. Not only will a great deal of detailed data (much of it unpublished) be available to us, but several of the key scientists working in the area have expressed their willingness to generate specific data if necessary to "fill out" certain information areas of our study.

Although we plan to limit our output maps to three size-scales, all of the data which we accumulate for computer retrieval/processing will have geographic factors specified in sufficient detail that they can be used in the production of a wide range of map scales. A more detailed discussion of this aspect of the problem is included in "Data-processing Considerations".

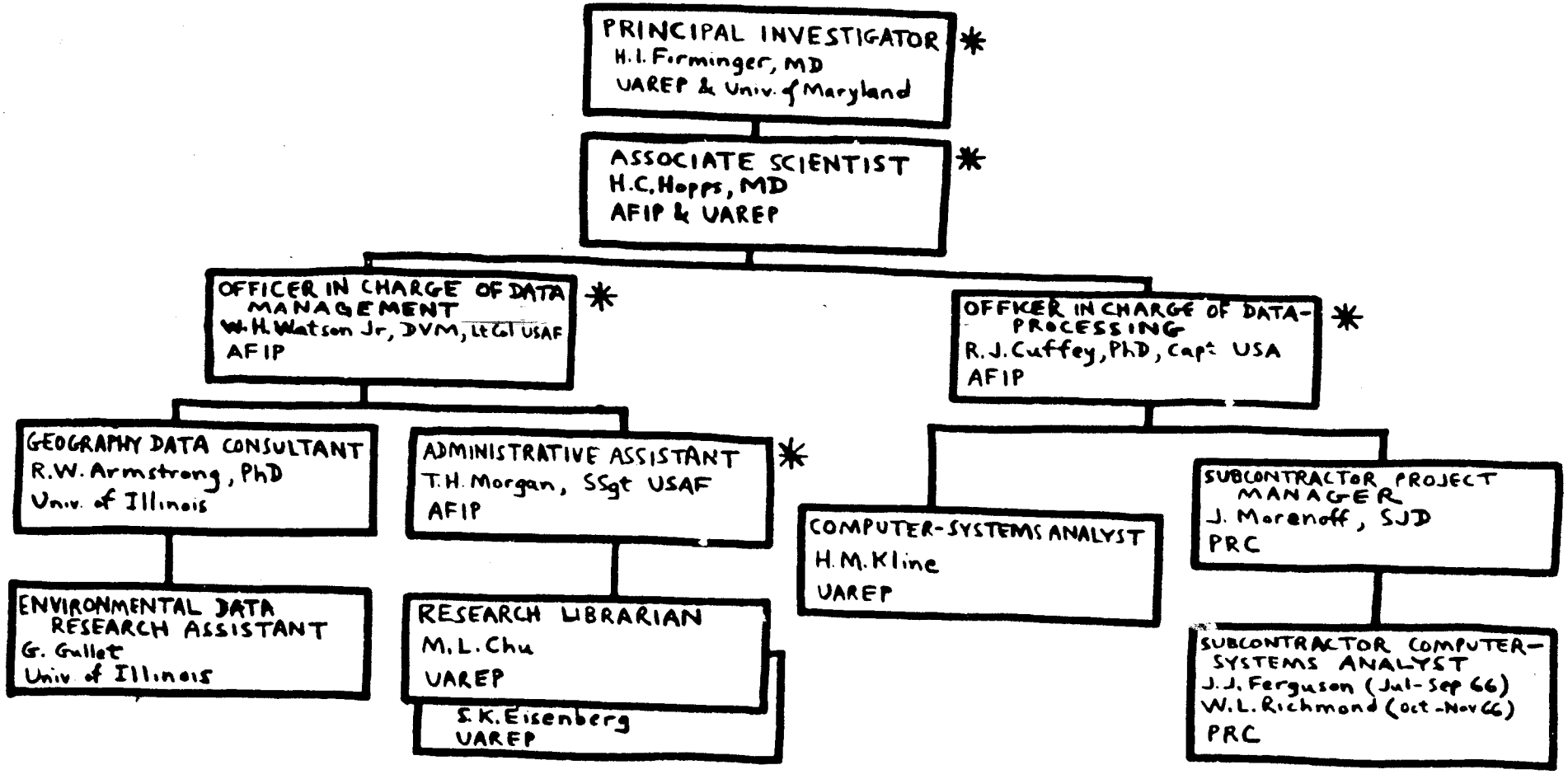
(4) Many kinds of map projections could be used. We have selected an equirectangular projection because it will provide economy of computer time/effort. Although this kind of projection leads to much distortion in the polar areas, distortion is at a minimum in the tropical zone, the region of greater interest.

Once the data are processed for plotting subprograms can be introduced so that the data could be plotted in accordance with virtually any type of projection desired, e.g. Mercator's, Goode's homolosine equal area, Azimuthal equidistant, etc.

It is appropriate once again to emphasize that the major objective of the MOD Project is to develop a system whereby narrative and tabular data can be collected and preprocessed (formulated)to a form suitable for subsequent computer processing and output in the form of distribution maps, graphs (e.g., the n-factorial three dimensional representation illustrated in figure 11), tables, and narrative. Although the self-imposed limitations described above narrow the limits of output we will seek, they do not narrow the potential limits of the system. The system is being designed to meet certain needs for information dealing with infectious disease, however, the same system could be used, with little modification, to analyze the ecologic factors which influence efficient stockpiling of corn or aluminium, or the ecologic factors which influence efficient forest preservation or development of recreational facilities, or the ecologic factors which influence efficient development and location of community blood banks or Medicare treatment centers, etc. etc.

* * * * * *

We have divided the MOD Project into six successive phases, related to the kind of effort required. Figure four shows these phases, the different tasks included in each, and the time-effort-personnel which each will require. A detailed consideration of the specific tasks comes later, however the major phases can be described here.

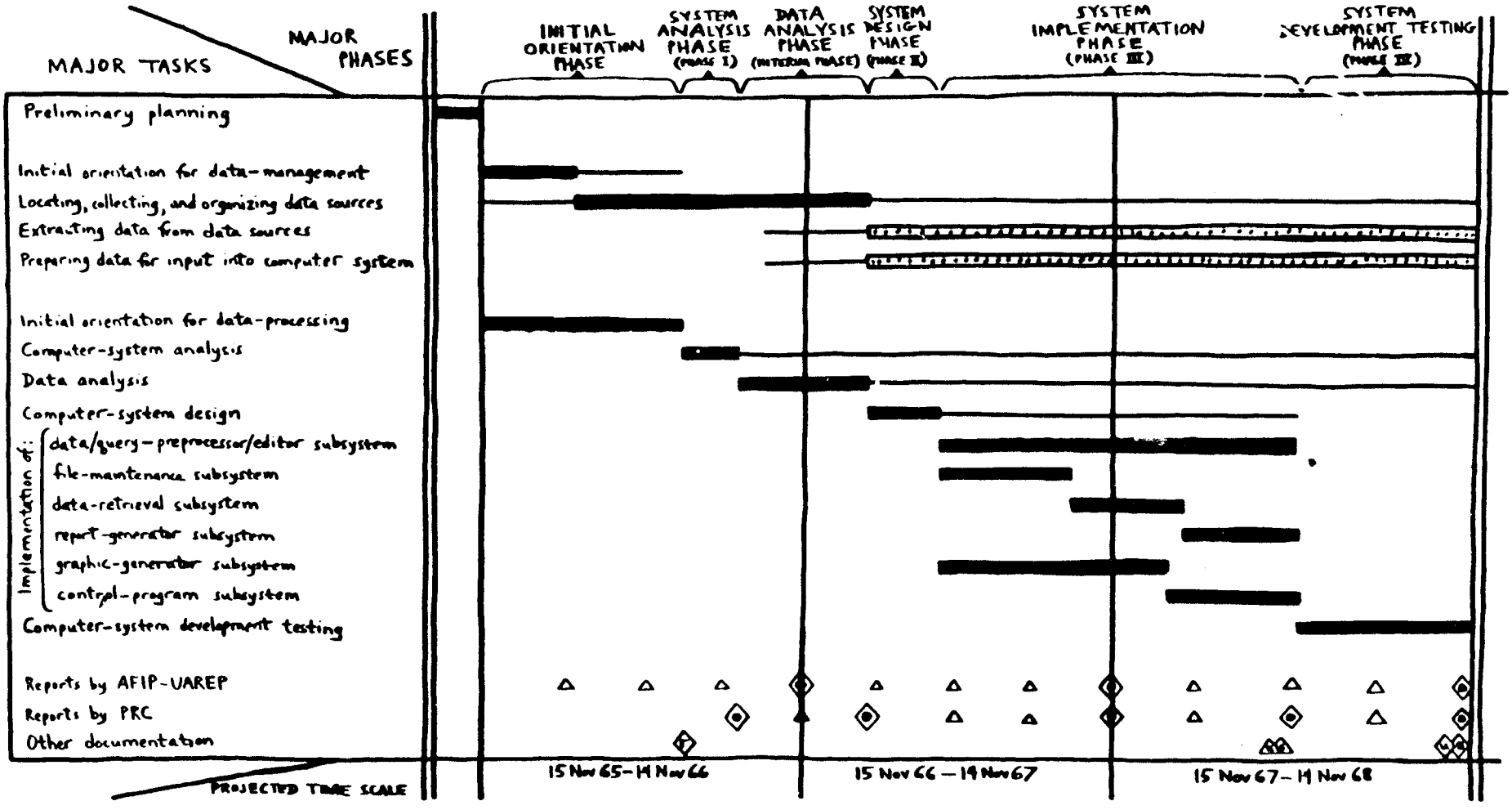


*Salaries not derived from this contract.

AFIP = Armed Forces Institute of Pathology; PRC = Planning Research Corp.; UAREP = Universities Associated for Research and Education in Pathology, Inc.

FIGURE 3.— Personnel involved in and internal organization of MOD Project, November 1965 through November 1966.

FIGURE 3. - Schedule of major phases and tasks comprising MOD Project, Nov 1965 - Nov 1968. The chart indicates what has been carried out or is currently underway (Initial Orientation, System Analysis, and Data Analysis Phases). It shows also what has been tentatively planned in order to bring the MOD Project to a successful conclusion within the three year time limit (System Design, System Implementation, and System Development Testing Phases).


MAJOR TASKS
MAJOR PHASES
INITIAL ORIENTATION PHASE
SYSTEM ANALYSIS PHASE (PHASE I)
DATA ANALYSIS PHASE (INTERIM PHASE)
SYSTEM DESIGN PHASE (PHASE II)
SYSTEM IMPLEMENTATION PHASE (PHASE III)
SYSTEM DEVELOPMENT TESTING PHASE (PHASE IV)
Preliminary planning
Initial orientation for data-management
Locating, collecting, and organizing data sources
Extracting data from data sources
Preparing data for input into computer system
Initial orientation for data-processing
Computer-system analysis
Data analysis
Computer-system design
Implementation of:
data/query-preprocessor/editor subsystem
file-maintenance subsystem
data-retrieval subsystem
report-generator subsystem
graphic-generator subsystem
control-program subsystem
Computer-system development testing
Reports by AFIP-UAREP
Reports by PRC
Other documentation
PROJECTED TIME SCALE
15 Nov 65–14 Nov 66
15 Nov 66 – 14 Nov 67
15 Nov 67 – 14 Nov 68
= Major effort essential for completion of MOD Project
= Major effort highly desirable for completion of MOD Project
= Minor or continuing effort necessary for completion of MOD Project
= progress, status, or preliminary report
= comprehensive, final, detailed, or technical report
r = resumé of orientation
s = system documentation
u = user's manual

Initially, we envisioned the MOD Project to be divided into three simple phases: (1) System Analysis, (2) System Design, and (3) System Implementation. However, as we grew more knowledgeable, it became evident that this division was an over-simplification. The six successive stages which we have settled on are:

Stage one, the Initial Orientation Phase, a time during which we created a large edge-notched file system covering many aspects of leptospirosis and associated ecologic factors. In addition we examined intensively the published literature dealing with information technology, computers and computer processing and automated mapping methods and had detailed discussions about these matters with many groups.

Stage two, System Analysis, was effectively performed with the essential help of personell of the Planning Research Corporation (P.R.C.), especially Dr. Jerome Morenoff, Senior Associate, and Joseph L. Ferguson, Associate*, through subcontract, UAREP 66.1. Work done during Stage two brought into clear focus the importance, complexity, and variety of the problems concerned with analyzing and preprocessing disease/environmental data so that it could be effectively computer processed. It was this insight which led to Stage three.

Stage three, an interim phase (between the initially visualized Phases I and II), is concentrating on data analysis. Many of the problems that we are meeting in this area have not been effectively dealt with before, and more research effort is, of necessity, going into this aspect of the study than originally planned. Obviously, this is one of the basic factors in developing an effective program. To solve these data analysis problems, one is required, among other things, to define terms very specifically, in the course of which it is necessary to develop a glossary or data vocabulary. One of the reasons why this is such a difficult task is that the terms which we must consider are derived from many different disciplines, and the same word often has a significantly different meaning when used by the geographer, the cultural anthropologist, the political economist, the epidemiologist, etc.

* Subsequently, Wayne L. Richmond replaced Joseph Ferguson.

Furthermore, our work on data analysis has forced a detailed evaluation of the essential features required of data in order that they can be meaningfully mapped. This led to the development of a data structure system to handle effectively various orders of qualitative/quantitative descriptors.

Again, we called on PRC for professional data-processing assistance, and this was accomplished through subcontract UAREP 66.2. Currently, we are in this phase of the program. A principal PRC consultant during the previous contract (Ferguson) was replaced by Mr. Wayne L. Richmond who, together with Dr. Morenoff (also of PRC), and Mr. Harry Kline (a programer with systems analysis capabilities, employed directly by UAREP) form a very effective and highly productive computer oriented team.

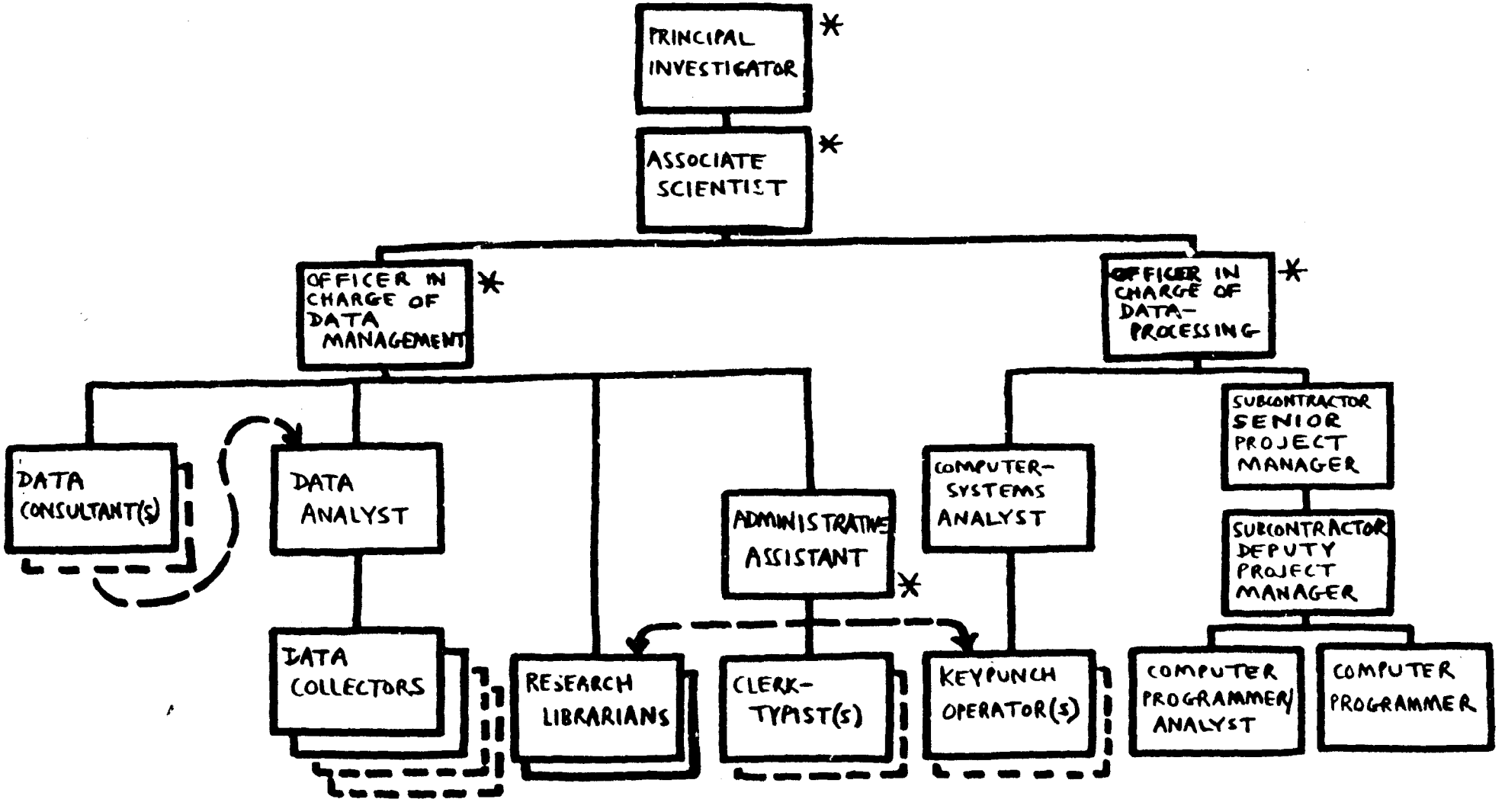
A detailed description of the data structure system, a list of many disease/environmental factors, and preliminary forms for data extraction/recording is presented in "Data-management Considerations".

After the Data Analysis Phase is completed (January 1967), we expect to know enough about the characteristics of disease/environmental data that we can proceed with developing a computer system which will process these data effectively.

Stage four, the System Design Phase (our initial Phase two), is the period during which detailed plans for the system will be made, the exact equipment needs determined, and the programs necessary to accomplish the data-processing outlined. We expect to continue our fruitful collaboration with PRC during this and the remaining phases.

Stage five, the System Implementation Phase (our initial Phase three), should begin during the mid-part of our second year and continue well into the third year of the MOD Project. During this time, equipment will be obtained and made operational, and the many required computer programs will be written.

Stage six, the System Development Testing Phase, is expected to occupy the last months of the three year MOD Project. The various programs of the MOD computer system will be integrated and tested with significant volumes of actual disease/environmental data to produce distribution maps and other forms of print-out. Numerous minor errors/defects will become apparent and require correction as we demonstrate and evaluate the system. We also anticipate the need (desirability) to add new features and capabilities during this last phase of the project as the potential of the system becomes more clearly evident, and as the answers to some questions indicate the desirability of answers to new kinds of questions. If all goes well, Phase six will culminate in a fully



* Salaries not supported by this contract.

FIGURE 5.— Personnel and internal organization projected for successful completion of MOD Project; actual numbers of certain kinds of personnel can be varied as necessary.

operational system consisting of two major components:

First, a sub-system for the selection/extraction/preprocessing of "raw" narrative, tabular, and graphic data necessary to produce a data file base.

Second, a sub-system of software-hardware which will incorporate the following into an integrated functional unit, data file base/editor and file maintenance programs/retrieval mechanism/processor/report generator (including map plotting).

After the MOD system is developed and operational, we anticipate that its many potential applications, medical and otherwise, will be readily apparent and fully exploited. We anticipate also that a number of the disease/environmental distribution maps produced by the system will be suitably reproduced for publication and wide distribution.

* * * * * *

Most of the work of the Mapping of Disease Project is being carried out at the Annex of the Armed Forces Institute of Pathology,(7th and Independence Avenue, S.W., Washington, D. C.). An area of approximately 650 square feet has been assigned for use by the Project (not including space to be occupied by card punch equipment, sorters, the computer, etc.). Internal organization of the MOD Project and personnel involved in work during the past year are shown in Fig. three. Obviously, additional personnel will be required to complete the MOD Project within the time allotted and Fig. five presents the projected internal organization/personnel requirements. Specific functions visualized for these new persons will be discussed under "Data-processing Considerations".

DATA - MANAGEMENT CONSIDERATIONS

An introductory discussion of the general nature and importa of the data-management aspects of MOD are given on page 9.

With our limited facilities, it is manifestly impossible for us to generate even a significant part of the data which we wish to process, and we have fixed on literature-search, compiled with information collected by word of mouth, letter, and "private" (unpublished) report as our primary sources of information. LTC Watson (D.V.M.), has concentrated on this, assisted by SSgt. Thomas H. Morgan, Mrs. Chu, and Mrs. Eisenberg. Over 3,500 selected references dealing with leptospirosis have been collecte approximately 1800 of which have been abstracted.

In addition, important continuing contacts have been made wi eminent leptospiral research workers, including Dr.A.D.Alexander (WRAIR), Mrs. Mildred M. Galton (CDC), and Dr. Lyle Hanson (Ill. Center for Zoonoses Research, U. of Ill.). These persons have been and will continue to be important sources of unpublished data and will continue to be very helpful in their critical analy of our work.

Data collection dealing with the hemorrhagic fevers is immediately available to us through the files of LTC Proctor Chil of the Geographic Pathology Division. Other important sources of compiled information readily available to us include MARU (Dr. Karl Johnson), the U. S. Component SEATO Laboratories at Bangkok (Col James L. Hansen, MC, USA former Director, Dr. Sylvanus W. Nye, formerly assigned there, and Capt. Will Blackburn, presently assigned there, from the Geographic Division, AFIP), the San Lazarus Hospital, Manila (Dr. Reyes), and the 5th Epi Flight, Clarke AFB (Col Kremers).

As our coverage of the disease-oriented literature becomes m nearly complete, we will be able to devote progressively more ene to the collection of data concerning those environmental factors which are likely to influence the occurrence or manifestations of the diseases under study. To this end we have begun a file of references to published maps which present the distribution of various environmental factors. We are being assisted in this effo by our consultant in medical geography, Dr. Warwick Armstrong, and a graduate student assistant, both of the Department of Geography University of Illinois. In addition, we have explored the possibility of subcontracts to: the Biological Sciences Communication Project (of George Washington University), the BioSciences Information Service of Biological Abstracts, and the American Geographic Society, in order to expand our data collectio capability.

Once the data sources have been collected and organized, the task of extracting the relevant data from the sources begins. When the MOD Project began, we did not realize that this would pose unusual problems. However, the further our work progressed, the more apparent it became that data extraction/preprocessing presented some very serious problems, far more complex than could have been anticipated in the beginning. Some of these problems have been encountered by other groups (we learned) but, often, they were avoided rather than solved, and, in some instances, this actually prevented the effective use of much potentially available data. Some of the important problems in this area have not been encountered by other groups, so far as we can determine, probably because no one else has attempted to develop the kind of system which MOD represents.

Our basic approach to solving these data-extraction problems has been through a group effort, involving both data-processing and data-collecting personnel. Attempts to extract and put into consistent form the data on disease and environmental factors contained in selected representative data sources were continued until the data was extracted meaningfully mappable and in a form acceptable to the data processors as well as the data collectors/analysts. General requirements for data content/format were formulated as an indirect result of these efforts, and this is one of our most important accomplishments to date (largely due to the efforts of Dr. Cuffey).

The first major problem encountered was that no generic terms existed which encompassed disease/environmental data. Thus it became necessary to construct a general data-analysis vocabulary before we could communicate effectively in relation to the disease/environmental data which we were attempting to extract. This data-analysis vocabulary includes definitions for and discussion of the interrelationships among such vitally important terms as "factor", "common elements", "value", "data point", "map", and "narrative". The data-analysis vocabulary, in its present (preliminary) form, is included as Appendix two.

The second major problem involved specifying precisely what items of disease/environmental information were pertinent to our major objective: the production of distribution maps. This led us to develop a catalog of disease/environmental factors which could be used by the MOD computer system in producing disease/environmental factor - distribution maps. This factor catalog, in its preliminary form, is included as Appendix three.

We have found that many of the data available for processing are incomplete in one way or another and, often, professional judgement/interpretation (sometimes extrapolation) must be carried out if the data is to be usable. Narrative print-out, to

accompany the computer maps, will note these interpretations and source document numbers will be available upon request should the user wish to consult the data report from which interpretations were made. Some of the data will be of very limited use because essential factors (which must have been known to the author) simply aren't recorded. These problems are numerous and serious, but our objective is to do the best we can with the information available. An important by product of the MOD system will be to emphasize the factors which must be specified in order for data reports to be most useful. An example of the kinds of conclusion-statements that are frequently found (and often very important) is given below. The limitations of data given in the apparently (at first glance) simple straightforward statement are indicated by the questions that need answers.

"Four percent of cattle in Southern Illinois have leptospirosis".

1. Over what time period were the data collected?
2. When were the data collected?
3. If this is a conclusion from a composite of different studies are we sure that there is no overlapping?
4. What was the size of the sample(s)?
5. What are cattle? (all bovids? a limited number of species of bovids? a limited number of breeds within one species? just cows? just mature animals? etc.?)
6. What was the nature of the sample(s) of "cattle"?
   - sick cattle?
   - cattle selected because of the State Health Department's interest in certain regions?
   - cattle selected because of University studies being carried out at specific chosen (cooperative) farms?
7. Is it likely that the prevalence was uniform throughout Southern Illinois?
8. What are the precise geographic limits of "Southern Illinois"?
9. What is "leptospirosis"?
   - disease in terms of clinical illness?
   - detectable antibodies?
   - recoverable organisms?
10. What was the inherent accuracy of the diagnostic procedure(s)?
11. What was the inherent sensitivity of the diagnostic procedure(s)?

Continued -

12. How reliable was the laboratory which performed the analyses?
13. Were the samples for analyses entirely adequate?

14. Is this report completely honest(i.e, is there intent to mislead)?
15. Were the studies which led to this conclusion well planned (i.e., was the experimental design good)?
16. If this report is a summary-analysis of a collection,is it correct? (i.e., was there an error in transcription or in mathematical manipulation of data)?

In retrospect, it seems that there were two reasons why it was necessary to develop these two conceptually important documents: First, no attempts nearly so comprehensive as ours have ever been made (or at least reported) to convert highly varying, narrative data to a consistent, geographically oriented (and therefore mappable) format. Previous disease-distribution maps have merely indicated presence or absence of particular diseases at particular times in particular locations. What we hope to do is much more involved than that.

Second, many different academic disciplines have been independently interested in various aspects of diseases or environments for a number of years. As a result, many of the terms which the different disciplines use in discussing specific data have quite different contexts. These have never been adequately correlated or even precisely enough defined for use in computer processing. Epidemiologists, agronomists, geodologists, and geologists, each with their own bias, all touch upon various characteristics of the soil. Epidemiologists, geochemists, limnologists, geomorphologists, ecologists, and recreation-oriented sociologists and anthropologists all discuss surface waters (lakes and streams) from various points of view. However, in many instances, these different workers do not use mutually intelligible terms. The factor catalog which we have developed represents an attempt to overcome these communications gaps.

Another difficult problem involves assessment of data reliability. Obviously this requires professional judgement, based upon a variety of factors not susceptible to rigid formulation. It has proved unrealistic to break this factor down into more than three categories: "more reliable", "less reliable" and "unknown". Even so, such a limited classification will allow useful separation of "good" data from questionable data. For example, data of a specific kind could be separated into three categories of reliability and each category plotted (mapped) on a separate transparent overlay sheet. Then, by proper arrangement of the sheets, the different categories could be viewed individually or together in any combination.

Turning to the mechanics of data extraction, the development of suitable data-extraction forms on which pertinent data can be effectively recorded has proved to be a very difficult task because of the highly varying content of the data sources, coupled with the requirement that the data must relate to a consistent, geographically-oriented format. Our tentative scheme for handling these matters (of Fig. 5) involves a three-stage process: (1) Data extractors (necessarily with biomedical background since value judgements are required) will fill in relatively simple data-extraction forms. These forms will be submitted to a data-analyst(s), who (with the help of data-consultants as necessary) will check (edit) the forms; (2) transcribe the data from them to a more rigidly formatted form. These latter forms will then be (3) converted to punched cards for input into the computer system. We have developed a series of data-extraction forms during the past year, each new form being an improvement over the previous one, incorporating refinements suggested by the data collectors. We are now using a data form designed for the particular purpose of collecting leptospirosis data in Southern Illinois. Only disease data MOFs (Middle Order Factors - see Appendix 2) appear on this form (Fig. 6); other forms will include spaces for both disease and environmental data. However, this form (Fig. 6) is allowing us to accumulate a significant amount of consistently-formatted disease data for the first time - data which is of great value at this stage of our computer designing efforts.

Another data-collection form (Fig. 7) is being used (tested) to collect information about environmental factors. Our data collection forms represent a "natural" language form of communication, designed to conserve the time and patience of our professional data extractors. The data analyst(s) serves as an intermediary between the data extractors and the computer. He will check/edit the data collection forms, noting frank errors, discrepancies, or omissions, and translate the data to a more rigidly formulated form from which the punch card operator can, without further interpretation, convert it to standard 80 column punched cards, subsequently to be put onto magnetic tape. As necessary, the data analyst will seek help from a professional data extractor/consultant to make certain that "translations" are accurate and that apparent discrepancies or omissions are real.

Figure 8 presents, in summary, our present concept of the sequence of events which will provide for input of data into the MOD system.

Appendices 2, General Data-analyses Vocabulary, and 3, Factor Catalog, are directly pertinent to Data Management considerations. These important documents represent a collaborative effort among all of the professional MOD workers, but Dr. Cuffey was the primary mover.

GRAPHIC NOT REPRODUCIBLE

DATA POINT NO. _____ (S. Ill. lepto.)

Geographic location by (LØ, LA) =

Geographic location by political unit =

Geographic location by UTM grid =

Manner of grouping data points =

Time period for which data applies =

Source document identification =

Disease measure. = value =

Specific disease agent =

Animal host infected: humans _____
domesticated animals (specify: _____)
wild animals (specify: _____)

Animal reservoir: humans _____
domesticated animals (specify: _____)
wild animals (specify: _____)

Method of transmission of disease to animal host =

Epidemiological status = nature of occurrence =

Average fatality in present outbreak =

Average severity in present outbreak =

Method of (continuing) detecting disease =
Type of sample examined for disease =
Size of sample examined for disease =
Type of subpopulation sampled =
Size of subpopulation sampled =
Type of total population sampled =
Size of total population sampled =

other LOA/MOA's
-over-

FIGURE 6. - A simple disease-data form for use in extracting data on leptospirosis from appropriate data sources.

FIGURE 7. - Data-collection form for use in locating and characterizing already-existing maps which show the distribution of various environmental factors.

GRAPHIC NOT REPRODUCIBLE

ENVIRONMENTAL—FACTOR MAP INFORMATION FORM

(mark all applicable boxes)

1. Scope of map: ☐ World (entire) ☐ World (major section; continent, ocean; specify: _____) ☐ SE Asia (entire) ☐ Thailand ☐ Malaya ☐ other (specify: _____) ☐ Mid-Western U.S. ☐ Illinois (entire) ☐ Southern Illinois (including Quadri-County area)

2. EXACT title of map: (i.e. name of environmental factor mapped)

3. Break-down of values of environmental factor mapped — XEROX LEGEND OF MAP AND STAPLE TO THIS SHEET

4. Method of representing data on map:
☐ dot-type symbols ☐ alphabetic/numeric symbols ☐ shading/patterns (black-white) ☐ shading/patterns (colors) ☐ contour-type lines ☐ other (specify: _____)

5. Projection used:
☐ equirectangular (plane-chart) ☐ Miller cylindrical ☐ mercator ☐ homolosine ☐ other (specify: _____)
(also note whether:
longitude meridians are ☐ straight or ☐ curved lines
latitude parallels are ☐ straight or ☐ curved lines
projection is ☐ interrupted, ☐ interrupted and condensed, ☐ condensed, or ☐ non-interrupted and non-condensed)

6. Scale of map: 1/_____

7. Dimensions of map: (round out non-applicable units) _____ (width) cm. in. X _____ (height) cm. in.

8. Date of publication of map: _____

9. Date(s) of data mapped: _____ (earliest) — _____ (latest)

10. Data compiled by (if different from Item 11): _____

11. Bibliographic reference for map: (book — author, date, title of book; publisher and city; page
journal — author, date, title of article: name of journal, volume:(number): page)

12. Call number of source containing map: _____

13. Call number of map (if different from Item 12): _____

14. Physical location of map (or source containing map):
☐ Univ. of Illinois main library
☐ Univ. of Illinois map library
☐ faculty member's personal library (specify whose: _____)
☐ AFIP Ash Library
☐ AFIP Geographic Path.—Geographic Zoon. library
☐ staff member's personal library (specify whose: _____)
☐ other location (specify: _____)

15. Additional remarks (use reverse side if necessary):

16. This form was completed on: by:

Date: _____ Name: _____

Our experiences with data-management in the context of this comprehensive program have exposed complexities which we could not have anticipated. It has become clearly evident that the most critical factor limiting meaningful computer output in our proposed system is the content/format of input data. The sources of the data are readily available, but there are major difficulties in extracting/formatting these data. Summarizing, the MOD project data extraction and file generation problems may be categorized as follows:

(1) Highly varying source document content (requiring development of a data-analysis vocabulary and a factor catalog which will establish common denominators).

(2) Highly varying reliability of raw data (requiring a system for defining reliability and, on occasion, validating data).

(3) Continual changes/additions in the data base file (making unusual requirements for editing/updating).

(4) Lack of a generic vocabulary encompassing medical/disease/environmental situations (related to item #1).

(5) Inherent complexities in the data which make it difficult to specify feasible procedure(s) for the extraction, editing, structuring, and storing of the data prior to computer input.

(6) Data file design problems due to complexities of the data in general, its very large potential volume, and the great number of interrelationships among the specific data and among descriptions associated with vocabulary/definitions/volume after computer input.

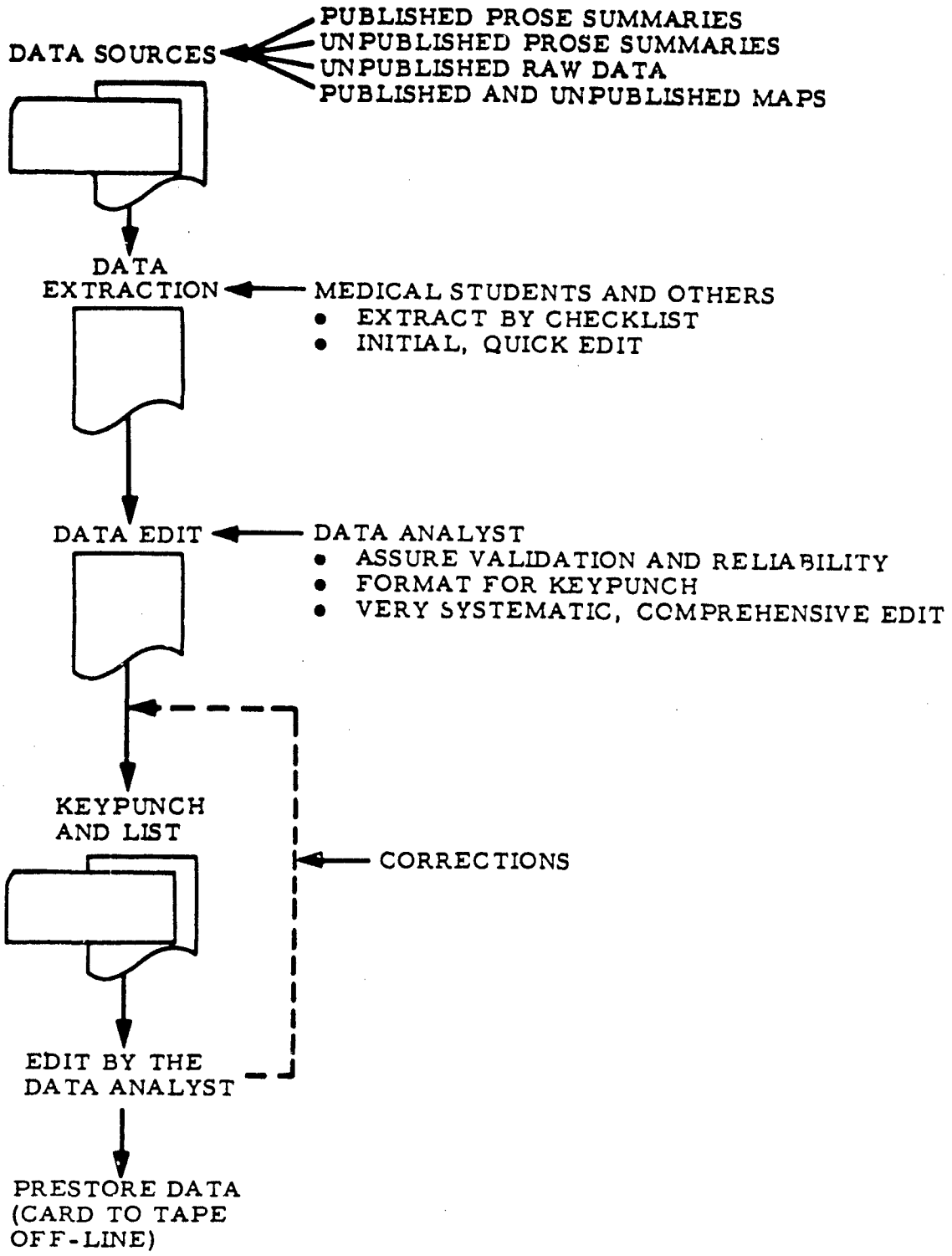


FIGURE 8.--Possible sequence of events for providing data for the MOD computer system.

DATA - PROCESSING CONSIDERATIONS

An introductory discussion of this phase of the MOD project appears on page 13.

Capt. Roger Cuffey has concentrated on this aspect of the study, working in collaboration with all members of the group, but particularly the PRC consultants (Dr. Morenoff, Mr. Ferguson, and Mr. Richmond) and Mr. Kline. In addition, Capt. Cuffey has effectively bridged the gap between data-management and data-processing and has made the major contribution in our developing data format system.

Computer Equipment Requirements: preliminary system analysis indicates that the MOD computer system should be capable of performing the following functions:

1. Input and edit data; display* data, including source.
2. Generate and/or augment/modify (including updating and general maintenance functions) data file(s), employing the input data.
3. Input and edit queries**; display queries.
4. Retrieve disease/environmental information from the data file(s) based on the query set.
5. Perform high-speed sorts.
6. Perform mathematical manipulations and transformations.
7. Generate drive tape for automatic data-plotting ("mapping") device.
8. Generate auxiliary hard-copy (printed) reports.
9. Display contents of any portion of the data file(s).

Of course, facility should be built into the system to allow any logical combination of the above functions. Also, all displays and reports should be query-controlled.

* The word "display" is used here to mean printed output of information in a suitable (easily understood) format.

** The above-mentioned queries (query set) refer to staff-specified inputs, including geographic coordinates/political units, conversion factors (map scale, map projection, etc.), and instructions as to what disease/environmental information is to be plotted (mapped).

A medium - to large-scale digital computer will be required, e.g., a CDC 3200/3300, an IBM 7090/7094, an IBM 360, or larger - scale units. Equipment should include at least four tape drives plus a disc module with capacity of 33 million ± alphanumeric characters per module. However, a limited system could probably be designed for such a configuration without the disc module, if an additional tape drive were available. Of course, usual peripheral equipment, e.g., card readers, high-speed printers, etc., will be required, in addition to a plotter, specification of which will be considered later.

The Armed Forces Institute of Pathology is scheduled for installation of an IBM 360 Model 30 medium-scale computer with 5 tape drives early in 1967, and this will satisfy a major part of our hardware requirements. Arrangements are in process which (hopefully) will allow us to use other, larger, computers in the Washington area, including an IBM 7090/7094. This will allow us to develop and demonstrate capabilities of the MOD system which are beyond the limits of the 360 unit.

Computer Program Requirements

Figure nine (next page) presents the functional components of the MOD computer system and shows their interrelationships. The extent and complexity of the programs involved in this operation are obvious. Turning back to Figure four (Schedule of Major Phases ...) page14, the System Design and System Implementation Phases are those in which there is to be major effort to write programs comprising the major subsystems: (1) Data/Query - Preprocessor/Editor, (2) File-maintenance, (3) Data-retrieval, (4) Report-generator, (5) Graphic-generator, and (6) Control-program.

As is evident from Figure three, extracting data from data sources and preparing these data for input will go along with the design of programs to carry out data/query - preprocessor/editor and file-maintenance subsystems. This is why our major effort at this time is on mechanisms of data extraction/formatting, representing a collaborative study that involves both bio-professional data-management personnel and computer-professional data-processing personnel; such collaboration is essential to develop a compatible system. As we approach solution of the data extraction/formatting problem, programming for data/query - preprocessing/editing, file-maintenance, and data-retrieval will begin. As the programs are written, they will be checked out against the data base file (in relation to hypothetical queries). As this phase of programming approaches completion, programs will be developed for the report-generator, graphic-generator, and control-program.

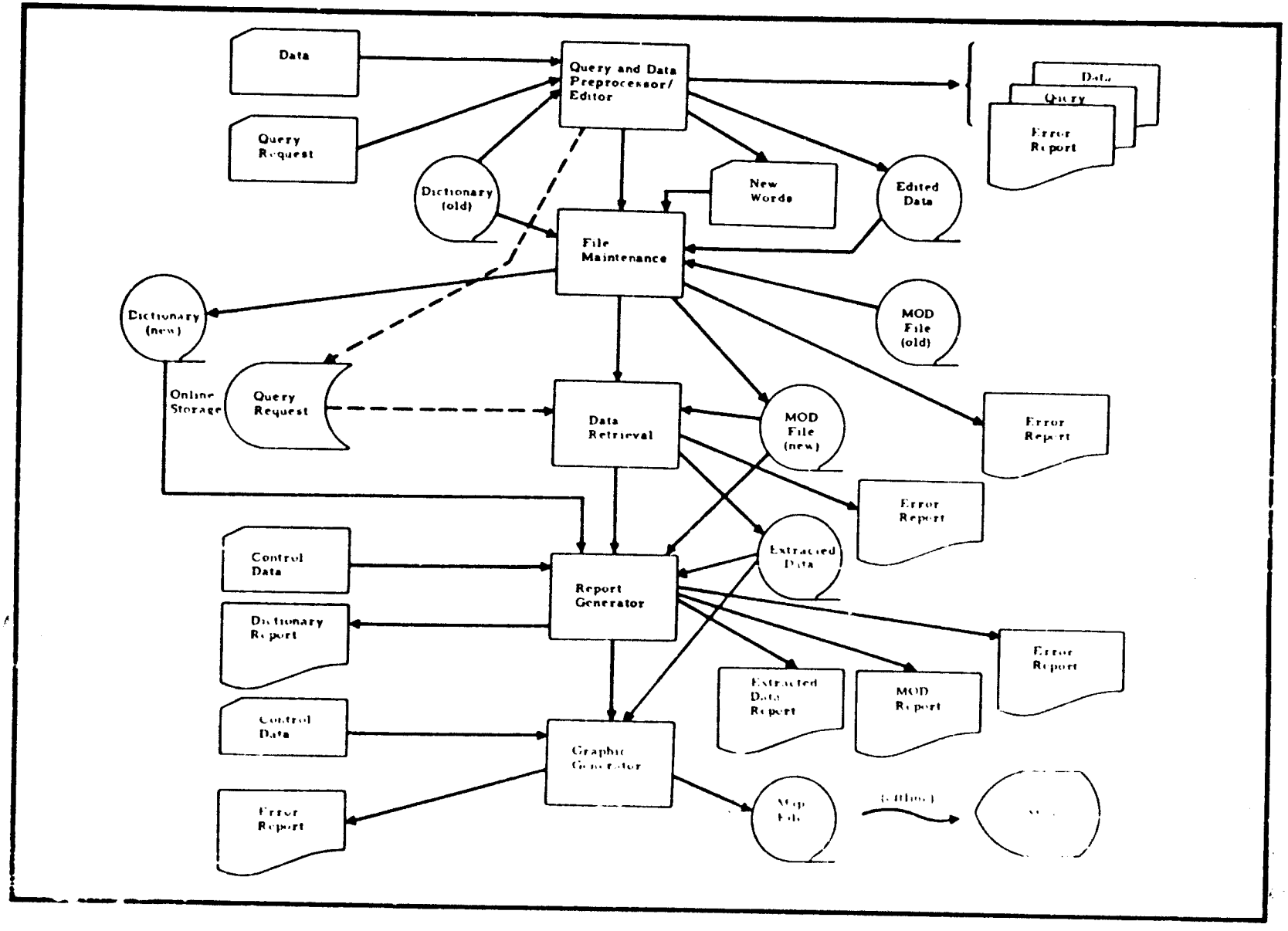

Data
Query Request
Query and Data Preprocessor/ Editor
Data
Query
Error Report
Dictionary (old)
New Words
Edited Data
File Maintenance
Dictionary (new)
MOD File (old)
Online Storage
Query Request
Data Retrieval
MOD File (new)
Error Report
Error Report
Control Data
Extracted Data
Report Generator
Dictionary Report
Error Report
Extracted Data Report
MOD Report
Control Data
Graphic Generator
Error Report

FIGURE 9. - Organization of MOD computerized disease-mapping system, as presently conceived.

These units pertain to the computer system, per se, and do not include essential staff activities. For example, no feed back is indicated in relation to the several error reports. These would be evaluated by the staff and, as necessary, correction of "File Maintenance", "Data Retrieval", or "Report Generator" would be initiated by the staff.

The problems involved in computer plotting are greater than we had originally anticipated, before extensive consultation with PRC. The problems have proved to be particularly difficult in the area of contouring because, often, contour lines cannot be drawn between points of similar arithmetic values, but have to reflect things of equal significance, and here, there are difficulties in determining actual meanings and in deriving valid common denominators. This problem requires development of new criteria, as is apparent from considering the General Data-Analysis Vocabulary (Appendix 2) and the Factor Catalog (Appendix 3).

The concept of program checkout will be to proceed from the smallest subprogram checkout to the complete system operation. Each subsystem will be checked out as completed. System design will be such that checkout can proceed on each subsystem independently. When all subsystems are checked out, they can be combined under the control program into the final system.

The final phase, System Development Testing, has been separated from the Implementation Phase because it represents a change in emphasis. At the beginning of this phase the data files will be filled with actual disease data (as opposed to "manufactured" test data during Phase III) and queries will be attempted for the first time. This will allow evaluation of the total system and any "final" changes before the system becomes fully operational.

Final program documentation to be prepared for delivery at the conclusion of the System Development Testing Phase, will consist of program listings and, for each, a functional description, logical diagram, and detailed flow chart. In addition, since the MOD computer system will provide a capability which is new, a user manual will be prepared describing the methods, procedures and language with which the potential user will operate the system.

Some characteristics of the several files and subsystems which will comprise the MOD computerized disease-mapping system are already clear, and are described in the following paragraphs.

1. Data File:

The MOD system's objectives are to categorize AND FILE disease/environmental data in such a manner that the data base can be queried to retrieve that portion desired which can then be plotted (as a map) or printed (as a hard-copy report) for detailed examination and analysis.

From previous discussion, it is evident that design of the focal point of the MOD system, the data file, is still under development. However, tentatively, each record (representing one data point) in the data file is expected to consist of four sections:

(a) common elements (CEL's), identifying the data point's geographic location, time frame, source of the data, etc. (See Appendix 3).

(b) the disease/environmental factor (a POF or HOF) studied at that point.

(c) the value (a FAV) determined for that factor at that location and time.

(d) supporting narrative material (NAR) as necessary. (See Appendix 2 in relation to b, c, and d.)

The common elements (CEL's), factor (POF or HOF), and value (FAV) are the items to be queried. The supporting narrative material (NAR) is included only to assist in understanding of retrieved data. It will consist of data which is associated with the data point but which cannot be meaningfully mapped or queried.

The data file(s) will probably be maintained on magnetic tape, but with a flexibility allowing use of disk (random access) storage if such proves desirable.

2. <u>Dictionary File</u>:

A dictionary file will be required, containing the vocabulary of words, a set of synonyms, and a tree-structured list of terms. While the words can be put into the dictionary by the computer program, the synonyms and structured list must be made up manually. The need for a synonym set was clear at the beginning of our study; the requirement of a tree-structured list of terms (carefully defined) was not. Fulfillment of this latter requirement has required a major effort.

3. Data/Query - Preprocessor/Editor Subsystem:

This program will function to insure that the data file(s) and the query-requests contain "clean" information. Data from the input forms presently being designed will be read by the Data/Query-Preprocessor/Editor. Each data point will be separated into the categories which match the file design. It will be necessary for this program to convert the CEL for geographic location from longitude-latitude, UTM Grid Coordinates, or Political Unit into an acceptable computerized form. Other CEL's may be handled in a semi free-form manner. The FAV will be checked for legality and acceptability, and each LOF will be individually edited against a continually expanding dictionary to guarantee its validity. The edited data will be written onto a magnetic tape for later use in updating the files.

Editing of the query-request information will be done in much the same way. The only major difference will be that the program will completely define a HOF or POF in such a way as to allow retrieval of a data point which is associated with any or all of the associated LOFs, in any desired combinations. The edited query-request information will then be saved, internally, for later use in the data retrieval subsystem. In addition to its other tasks, the data/query-preprocessor/editor program will produce a hard-copy report (optional) on the new data, the query, and any error messages which may have been generated.

4. File - Maintenance Subsystem:

This program will perform the updating of the MOD file(s). The edited data file will contain data points to be added to the MOD file(s) and data which modifies existing data points. Three operations are required: first, simply to add new data points to the file; second, to modify information for a data point without otherwise changing the contents of the data point record; third, to add new words, synonyms and tree structures to the "dictionary" file.

5. Data - Retrieval Subsystem:

This program will perform the retrieval of information from the MOD file(s) in accordance with queries input by users of the system. The program will use the query after it has been preprocessed and structured in the Data/Query-Preprocessor/Editor. The MOD file(s) will be read and checked against the query requests, and those data points which match the requests will be retrieved and written onto magnetic tape for final processing.

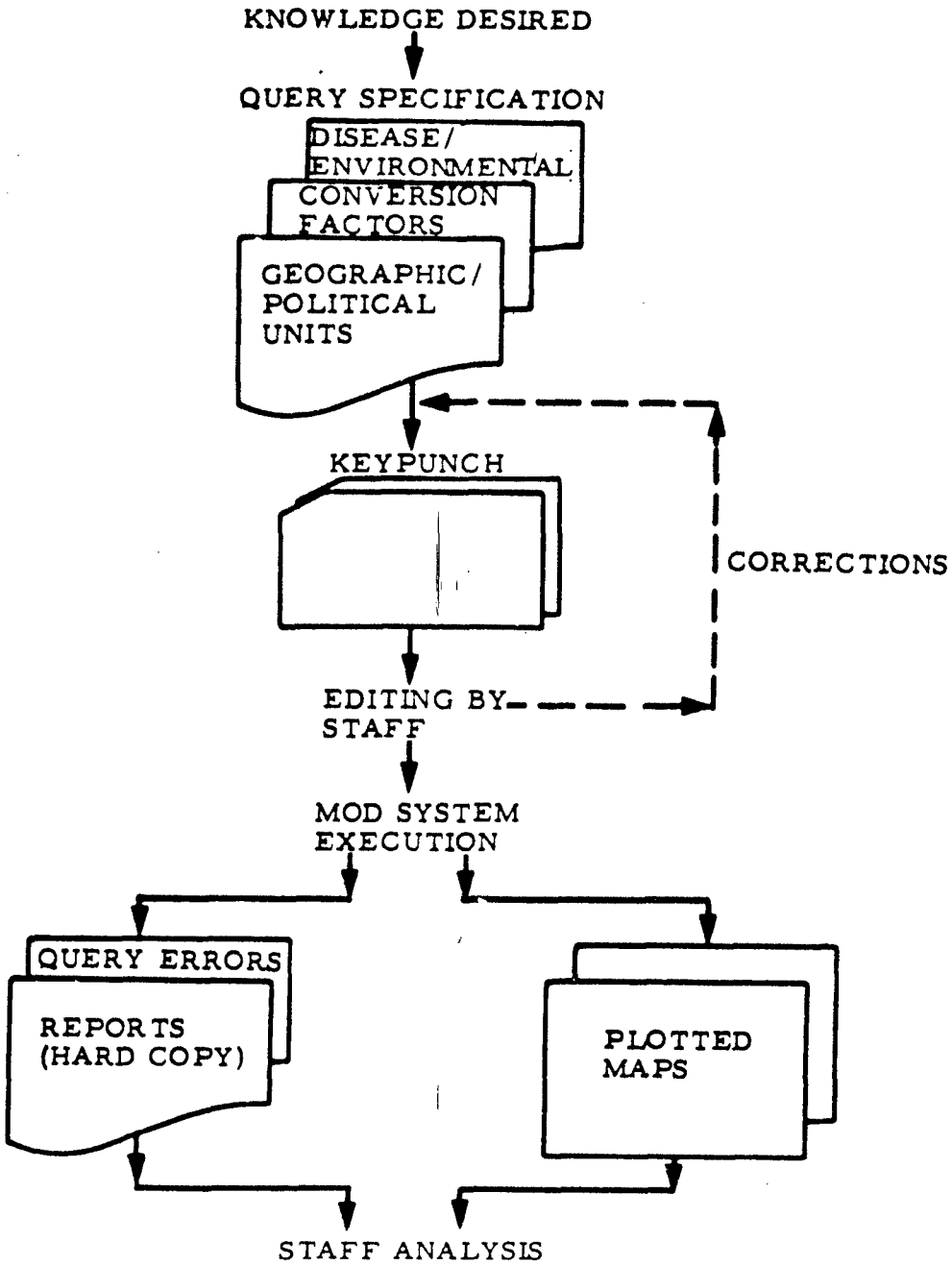
6. Report - Generator Subsystem:

This program will provide the capability for all printed reports desired; it will also catalog or list the data files and dictionary file. Reports will be controlled by the user via report request cards. Capabilities will be provided to display portions of the files in addition to cataloging the information (as, for example, the latest additions to any of the files) to supplement plotted data.

7. Graphic-Generator Subsystem:

This subsystem will extract the latitude, longitude, and factor-value from the retrieved-data file, and convert it to rectangular (x,y,z) coordinates. It is our present plan to have the program construct a rectangular grid, interpolating to fill in z - values at all (x,y) grid-line-intersections which lack observed z - values. After gridding is completed, additional control information would be read in (via the Control Program) to define further the maps requested. This control information would include: (a) name of region mapped, (b) scale and size of map, (c) boundaries of the region mapped, (d) type of map projection, (e) such "legend" data as date prepared, requestor's name, security classification, general description (title) of data, etc., (f) method of representing data on map, e.g., dots, shadings, or contours. Then, any of several existing contouring routines could be used to calculate the contours, and the selected routine incorporated as additional control information. Further processing of the data could be performed at this stage before requesting the subsystem to produce a magnetic tape, containing appropriate instructions to drive an automatic plotter (ordinarily an off-line activity).

A survey of plotters potentially useful in the MOD system is included as Appendix 4 of this report.



- DETERMINE SALIENT DISEASE/ENVIRONMENT RELATIONSHIPS
- DOCUMENT POSSIBLE MEDICALLY ORIENTED COGNATE DISCOVERIES/RESULTS AND RAMIFICATIONS
- POSSIBLE DATA BASE RELIABILITY/VALIDITY ANALYSIS
- POSSIBLE QUANTIZATION OF ANALYSIS TECHNIQUES FOR FUTURE ANALOG PURPOSES
- GAIN INSIGHT FOR FURTHER MAP GENERATION
- ERROR ANALYSIS - QUERY CORRECTION AND APPROPRIATE ACTION

↓

RECOMMENDATIONS AND CONCLUSIONS

FIGURE 10.--Possible MOD computer-system applications.

8. Control Subsystem:

In a system as complex as this, a control program is virtually a requirement. This subsystem will function to coordinate the operation of all the other subsystems. It will read all of the control information and determine the proper subsystem to be called in at the appropriate time. Such an operation will minimize procedural errors and the necessity for computer operator intervention, and will speed computer processing. Thus it will maximize efficiency of total system operation.

Data-Processing Personnel Requirements:- Several additional data-processing-oriented persons will be required to design, implement, and develop/test the MOD computer system, as indicated in Figure five, page 16.It is virtually impossible to hire high quality personnel of this sort on short notice for work of relatively short duration. Consequently, the realistic approach is to get them by subcontracting. PRC, in view of its very effective previous work on the MOD Project (and the understanding which it has developed), is the obviously preferred concern for future subcontracts.

In view of the scheduling requirements necessary to complete the MOD Project by 15 November 1968, the following persons are considered to represent a minimal requirement:

1. Computer-systems analyst (in-house,UAREP) - will assist in the system analysis and design, and will implement the graphic-generator and control subsystems.

2. Subcontractor Senior Project Manager (PRC) - will perform overall technical and administrative supervision of various phases of the project, prepare and deliver briefings as required by UAREP-AFIP, maintain cognizance of related efforts performed by other Governmental Agencies, perform systems analysis, and aid in the systems design effort.

Continued -

3. Subcontractor Deputy Project Manager (PRC) - will be responsible for technical and administrative supervision on a day-to-day basis for the various phases of the project. He will also prepare and deliver briefings as required by UAREP-AFIP, maintain daily liaison with the client to insure that all needs are satisfied, inspect the progress of programmers to assure that contractual demands and deadlines are being met, perform systems analysis, and aid in the systems design effort.

4. Computer Programmer/Analyst (PRC) - will assist in the systems analysis and design and data file design. He will also implement the data/query-preprocessor/editor. He will assist in integrating this with the control program. (This person will be more experienced than the next-described programmer, inasmuch as his duties are more demanding.)

5. Computer Programmer (PRC <u>or</u> UAREP) - will implement the file-maintenance, data-retrieval, and report-generator subsystems. He will also assist in integrating these with the control subsystem.

6. Automatic Data-processing Support personnel, e.g., keypunch operators and computer/plotter operators - will be required to supplement the work of the previously listed personnel

The UAREP (in-house) computer-systems analyst and also, possibly, the subcontractor deputy project manager will be involved with maintenance and other functions to be performed during the final System Development Testing Phase.

Plotter Equipment Considerations: Computerized maps* can be produced in several ways:

(1) by high-speed (line-) printers.

(2) by ink-on-paper-type plotters; either flatbed or drum, and either analog or digital.

(3) by cathode-ray-tube (CRT) devices either with direct recording on micro film or visual display on a screen

Either high-speed printers or cathode-ray-tube (CRT) display screens could be used to obtain a quick, overall,low-resolution view of what a particular map would look like. Then, if the user decided that the particular map would be useful to him, he could request that the map be plotted in detailed, high-resolution fashion by an ink-on-paper plotter.

At the present state of their development, CRTs could yield only low resolution maps (not more than 1000 points per 15 linear inches). High speed line printers, in addition to producing high resolution maps, are able to provide much the same "oversight" information quickly. The advantages of the 'quick look" would be that one could determine whether the data under consideration was actually worth the cost in time and effort of high resolution mapping.

We have not dropped altogether our interest in CRTs because developments within the next two years could lead to marked improvement in their resolution. The advantages of high-resolution cathode ray tube imagery are great: they could provide maps very quickly (seconds). In addition,they could provide a continuous range of scales, allowing the user to zoom in, so to speak, on a geographic area of particular interest.

* In the context of this program, a map is considered to be a graphic representation of data distributed meaningfully in relation to geographic coordinates. More often than not, the significance of the data which we produce by computer will not become apparent unless it is plotted on a base map or used as an overlay on an existing map.

A preliminary survey of automatic plotting devices potentially available to the MOD Project has been carried out jointly by UAREP-AFIP and PRC personnel. The approach was twofold: (1) several agencies in the Washington, D. C., area were visited by the study team, and the capabilities of the agencies were discussed (with their personnel), taking into consideration the problem areas and areas of specific correlation to MOD goals; (2) the various computers/plotters available were considered in terms of software requirement and plotter speed/accuracy.

In addition, a detailed survey/analysis of existing plotters was carried out by Mr. Kline, results of which are presented as Appendix 4.

At the present time, it appears that either a medium size Gerber (series 600-1000) or Calcomp (series 600-800) plotter would best suit the needs of the MOD system. It is recommended also that a "flatbed" type of plotter be employed (as opposed to a "drum" type). This is because of the increased versatility of reproduction techniques that a flatbed allows, e.g., direct photographic recording, etching, etc. which could be an important advantage if a large number of copies of a particular map were to be produced in a short time.

Turning to software requirements, Calcomp plotters come equipped with a macro-type language program which facilitates fast and efficient programming of a "drive tape". The actual advantage of this is not so great as it first appears, however, since Gerber's instruction set (micro) is less complicated than Calcomp's. Furthermore, Gerber flatbed plotters are generally more accurate than Calcomp plotters. However, highly accurate machines may not be essential for the MOD system.

With regard to plotter availability in Washington, the study team came to the conclusion there should be no significant problems. Several of the agencies visited have plotter time which would be available to the MOD project at low cost. Plotter availability will be considered further at the time detailed system design gets under way.

Output Considerations - The primary output from the MOD computerized system will consist of computer/plotter-drawn maps showing the geographic distribution of various disease/environmental situations. Before discussing map-outputs in detail, however, it is appropriate to mention several other kinds of output which will also be desirable - and possible:

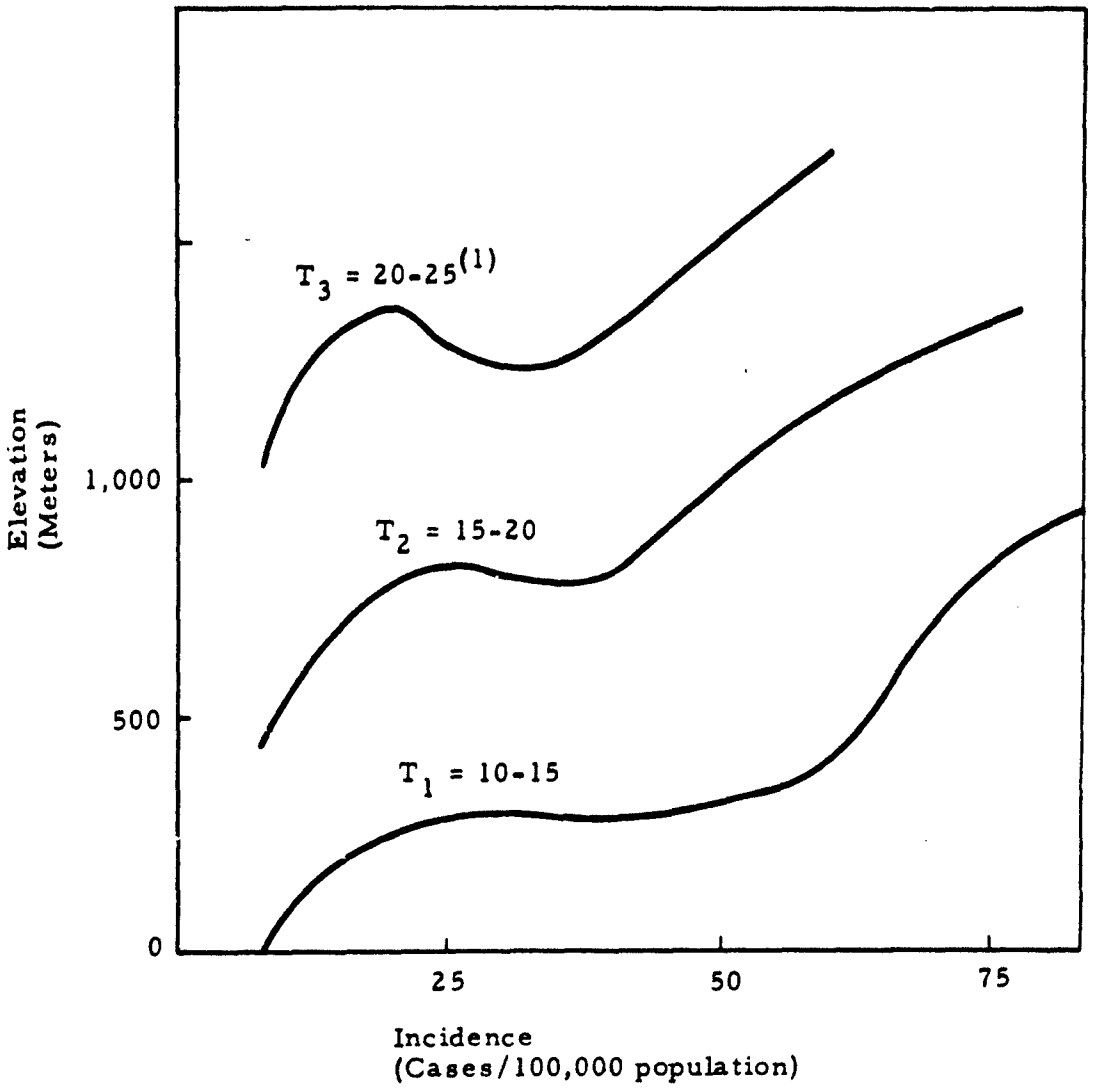
First, for purposes both of computer-system development and check-out, we should have an ability to print out, in readily-understood format, the contents of any given part of the MOD data and dictionary files.

Second, for similar reasons, we have need to print out the query set being used to generate other (visual map-form) outputs.

Third, printed listings of disease/environmental data which display certain significant interrelationships, would be of potential importance since these would suggest particular map output. Also, results of statistical correlation tests made between various disease/environmental data should be printed out.

The Fourth kind of output is rather complex. When a person compares and analyzes the data patterns displayed on several maps (in order to determine relationships among the disease/environmental factors mapped), he goes through a set of procedures which might be approximated/imitated by the computer system. If these procedures could be described by a suitable analog, the computer system itself could compare and analyze the data which, otherwise, it would have output as several different maps. This would allow it to output, in some fashion, a description of those relationships. Such a capability, even if very elementary, could be useful in deciding what particular disease/environmental factors should be subsequently mapped. Also, it would allow changes (increases or decreas in the disease/environmental situation to be readily displayed.

The Fifth kind of output concerns an area of medical interest that may have important ramifications, an area discovered during the System Analysis Phase. The anticipated data file base, including data relevant to a wide spectrum of disease/environmental situations, allows generation of graphic displays which might yield

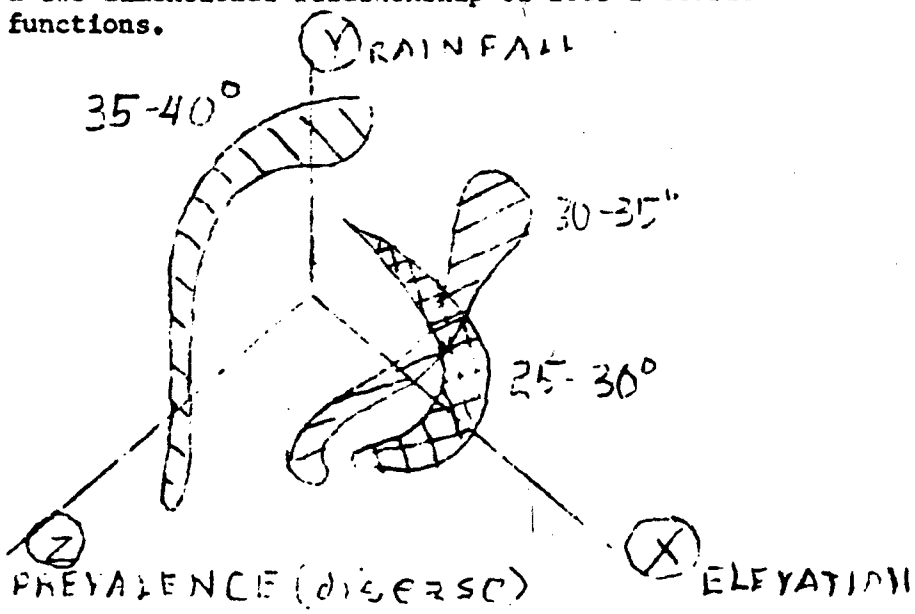


Note: (1) T = Mean July noon temperature (°C).

FIGURE 11.--Possible graphic display which might be able to be generated by employing the MOD system's data files.

important etiological patterns and/or mathematical relationships. Figure 11 illustrates a hypothetical relationship of this type: elevation versus period prevalence plotted over a range of five degree temperature spans. (This can be looked at, mathematically, as a family of curves). There is the possibility of producing virtually an infinite variety of such three factorial relationships. Extensions of this concept could, conceivably, allow predictions as to the probability that a specific geographic area would have a specific disease incidence over a specific time, suggesting specific preventative measures.

Carrying this concept a step further, one could achieve a two dimensional relationship of four-factorial functions.



Associated with this concept is the idea of overlaying on the same basic grid (using transparent material), several disease/environmental data elements. This would yield intersections of curves, the intersection points of which would probably have important significance.

Furthermore, in considering causal factors of disease, a mathematical relationship might be approximated among the variable items. If this were accomplished, altering these (in succession) might well give new insight into the precise effects of the altered variables, exposing the critically important ones.

Sixth, and finally, we may wish to compare the disease data in the MOD computer-system files with environmental data already presented in map form and this could be done in one of four ways: (a) the latter data could be digitized and input to the MOD system; (b) the already-existing map could be redrawn (manually) to a different (appropriate) scale and projection, and used as a base map; (c) the map could be photographically reduced or expanded to a different (appropriate) scale and used as a base map; (d) the map might be used directly, without change, as a base map. Decision as to the most appropriate choice will depend a good deal upon how flexible the output capabilities of the MOD system are.

Returning to the major output requirement of the MOD system - to produce (by computer/plotter techniques) maps which display the geographic distribution of certain disease/environmental situations - this has been approached in three ways:

(a) by a general investigation of presently existing cartographic techniques, both manual and automated.

(b) by a manual plotting of actual (and manufactured) disease/environmental data

(c) by computerized plotting of actual (and manufactured) disease data using a commercially available computer routine.

(Approaches 2 and 3, above, represent pilot studies involving the kind of data which will eventually be put into the MOD system's data files for use in generating maps.)

There are three general categories of problems which must be successfully handled by the MOD computer/plotter system if it is to function effectively in its "mapping of disease". These problem categories, common to fabricating any map, involve: (1) map size and scale; (2) map projection; (3) method of representing data on map.

The size and scale of any given map to be produced by the MOD computerized mapping system will be determined by two considerations:

First, the size of the map will be limited by plotter capacity (in terms of paper size); second, the scale of the map will depend upon the actual size of the region to be mapped, i.e., whether that region is the entire world, south-east Asia, or southern Illinois, in relation to the size of the maps.

The map projection to be used for a particular map can be varied by introducing appropriate (relatively simple) programs which will alter the grid pattern onto which specific data will be plotted. Although any mathematically-definable projection could be used by the MMD system, four projections have been tentatively selected as standard: equirectangular (the standard), Goode's homolosine, Miller's cylindrical, and Mercator's projections. The equirectangular projection has the advantage that it is equivalent to a simple rectangular coordinate (x,y) grid, and this allows relatively simple programming. Goode's homolosine projection has the significant advantage of being an equal-area projection. Mercator's and Miller's cylindrical projections have the advantage of wide familiarity. Furthermore, many already-existing maps that show distributions of disease/environmental factors are drawn according to these projections.

There are, for our purposes, three basic methods of representing disease/environmental data on distribution maps. One may use dots, shading, or contours - singly or in various combinations.

Dot-type maps are illustrated in figures 12 and 13. This type of map can be made quite simply by computerized techniques, and very effectively presents some kinds of disease/environmental data.

Shading-type maps are shown in figures 14 and 15. Although this type of map is quite easy to produce manually, it presents serious technical problems from a computer viewpoint; we are still working to overcome these problems.

Contour-type mapping is illustrated in figures 16 and 17. If the data presented have a fairly uniform distribution, falling into a regular, rectangular grid pattern, they can be contoured by existing computer techniques without much difficulty. If the data are randomly distributed (as is usually the case), these same computerized contouring techniques can be utilized, but the data must first be fitted into a rectangular grid by means of various averaging and interpolating techniques. (Obviously, a computerized contour map could easily and quickly be converted to a shaded-type map, manually.) Combinations of dot-type, shading-type, and contour-type mapping are illustrated in figures 18 and 19.

Figures 12,14,16,18, and 20-23 were drawn/plotted using data which we extracted and formatted for the specific purpose. The resulting maps, representing the first effort of this sort, are, admittedly, far from perfect. But they are prototypes and, as such, represent an important achievement since they demonstrate that meaningful disease-distribution maps can be produced by the

FIGURE 12 - Some actual disease data presented as a dot-type map drawn manually; these data were extracted, prepared, and mapped as part of the MOD Project effort.

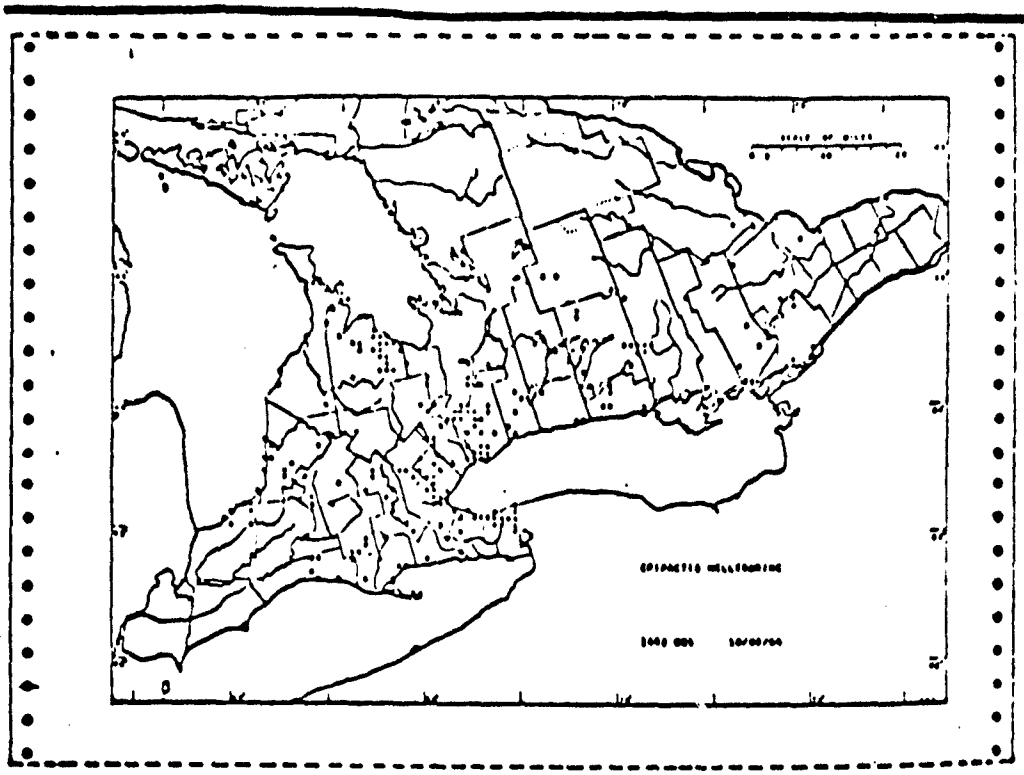
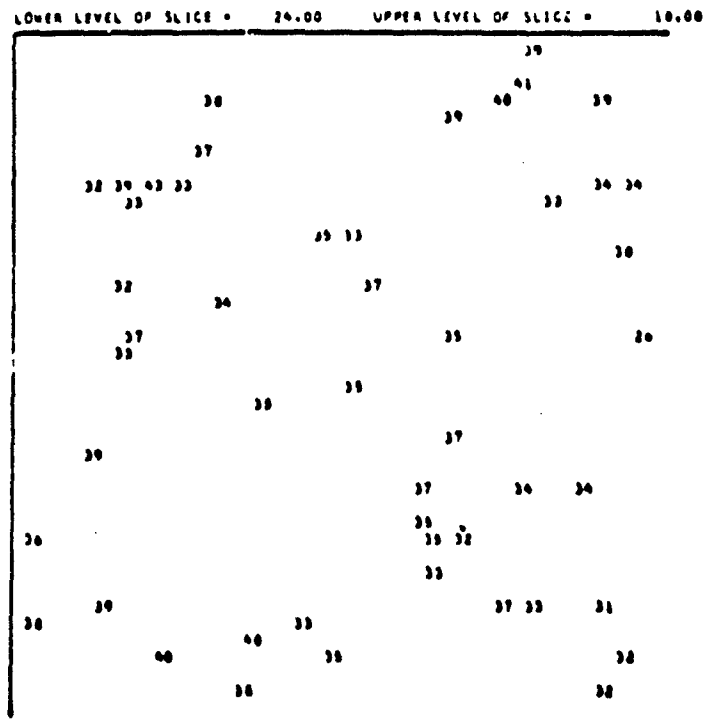


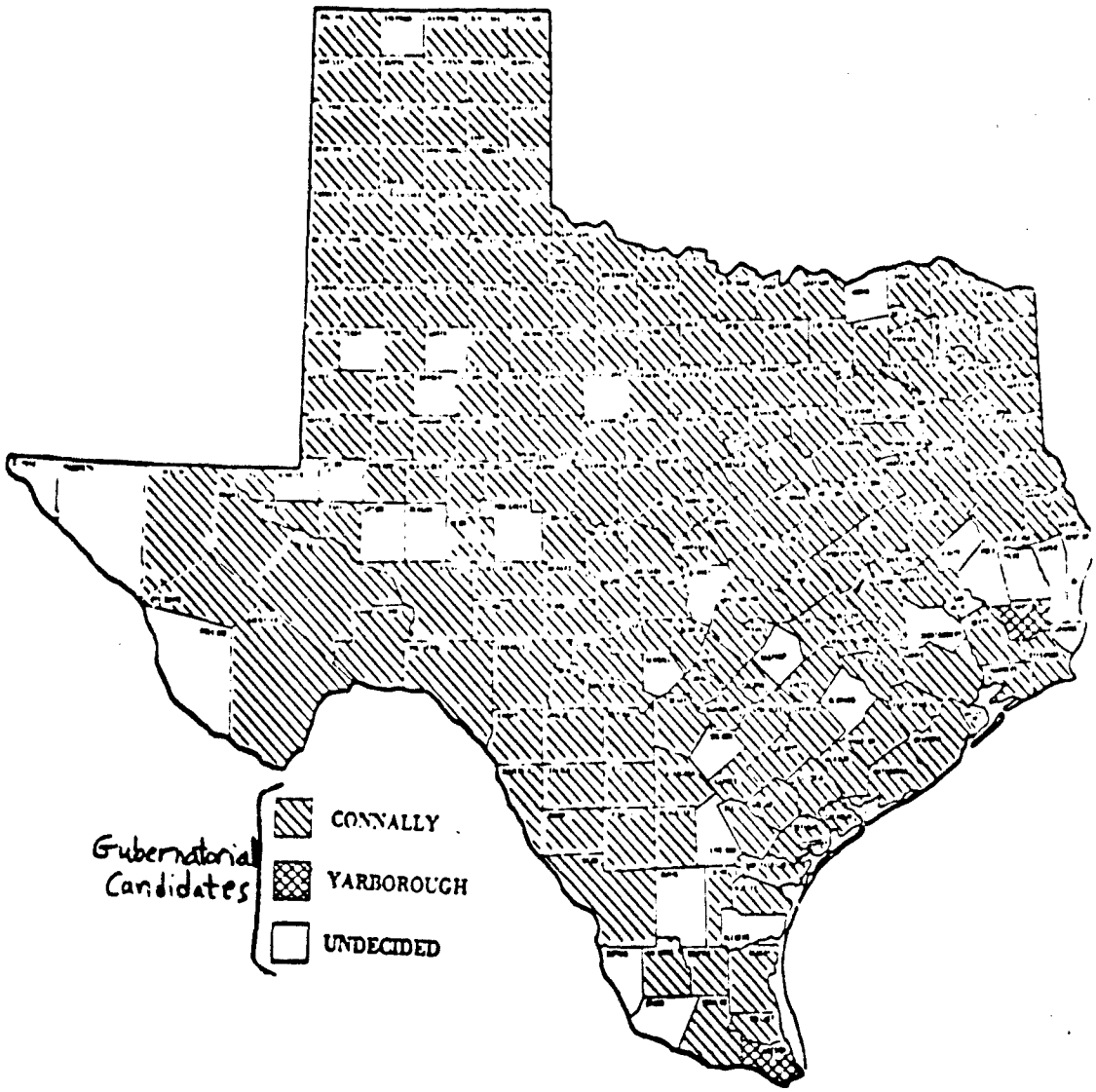
FIG. 7. Records of the occurrence of *Epipactis helleborine* (L.) Crantz in southern Ontario plotted on an I.B.M. 407 Accounting Machine. (Soper, 1964, p.1096)



(Harbaugh, 1964, p.57)

FIGURE 13. - Examples of computer-produced dot-type maps.

FIGURE 14 - The same disease data as Fig. 12, but presented as a shading-type map drawn manually.



(Digital Plotting Newsletter, June 1964, p. 1)

FIGURE 15. - An example of a shading-type map produced by computer.

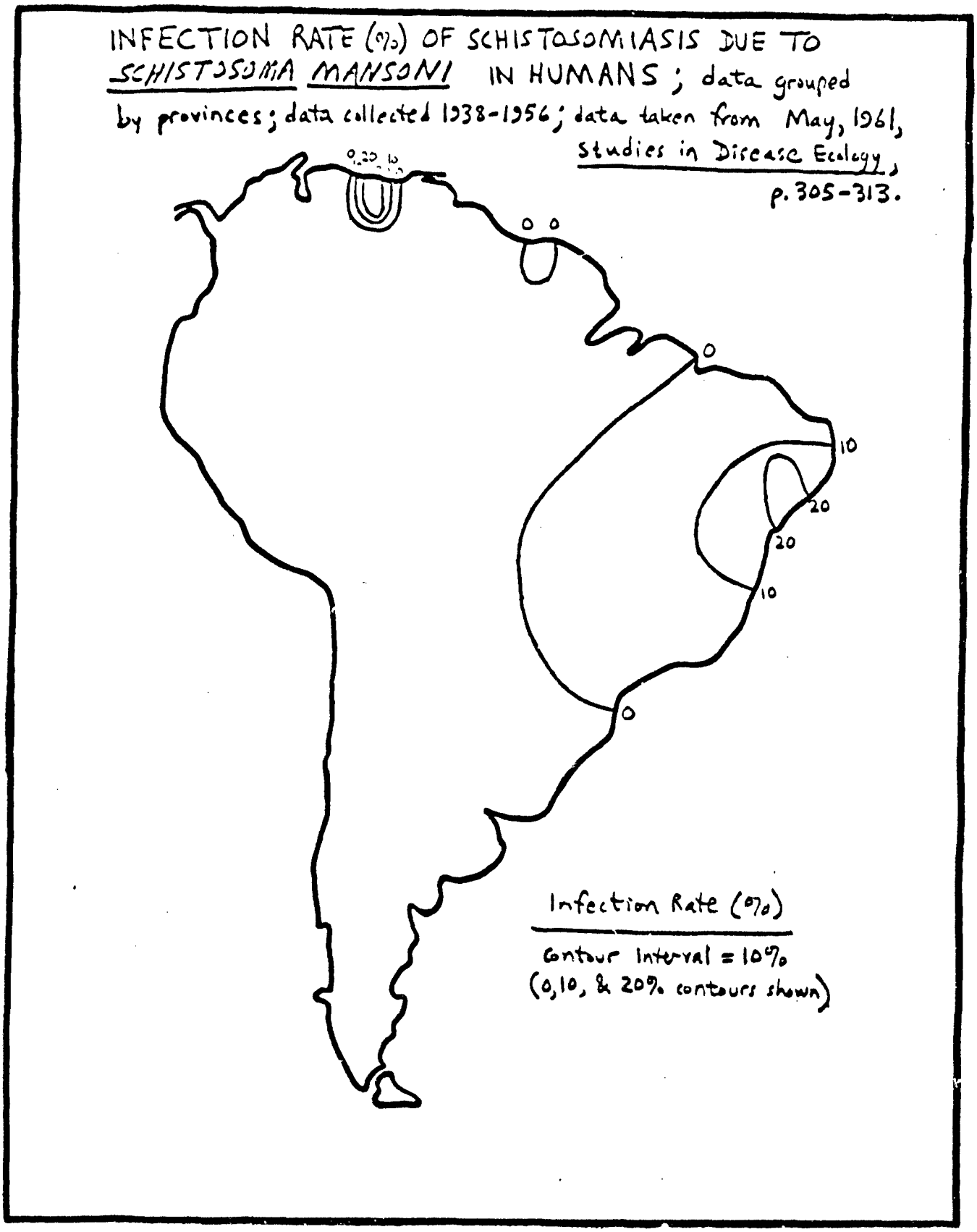


FIGURE 16 - The same disease data as Fig. 12, but presented as a contour-type map drawn manually.

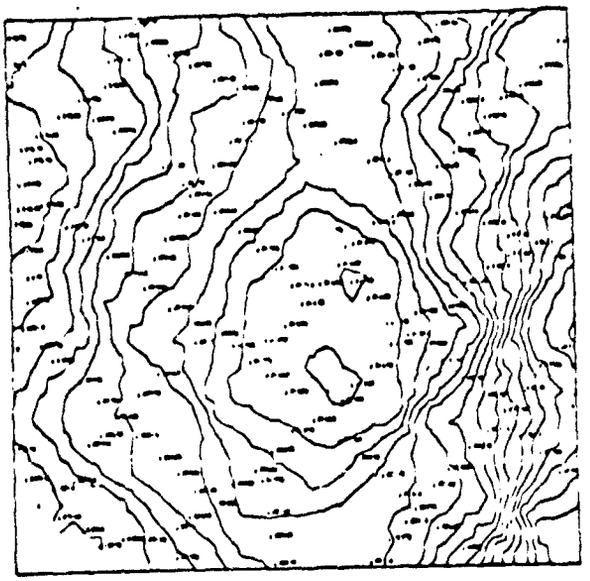
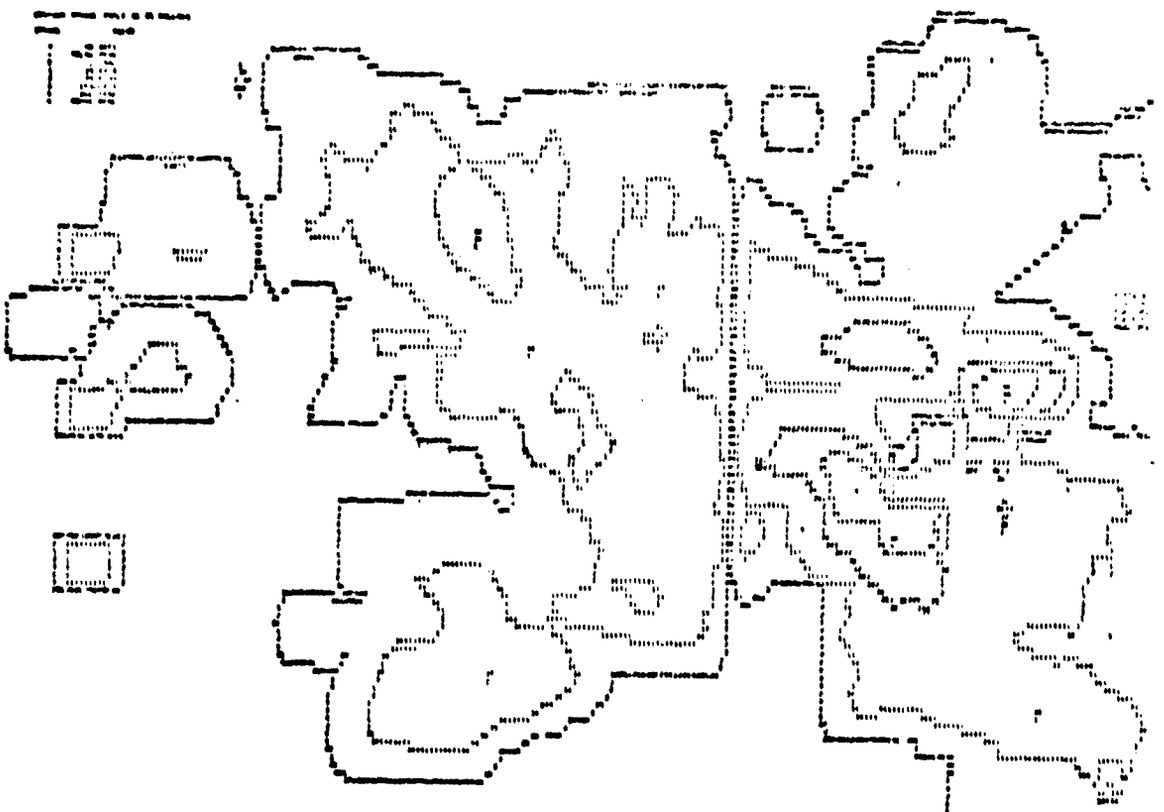


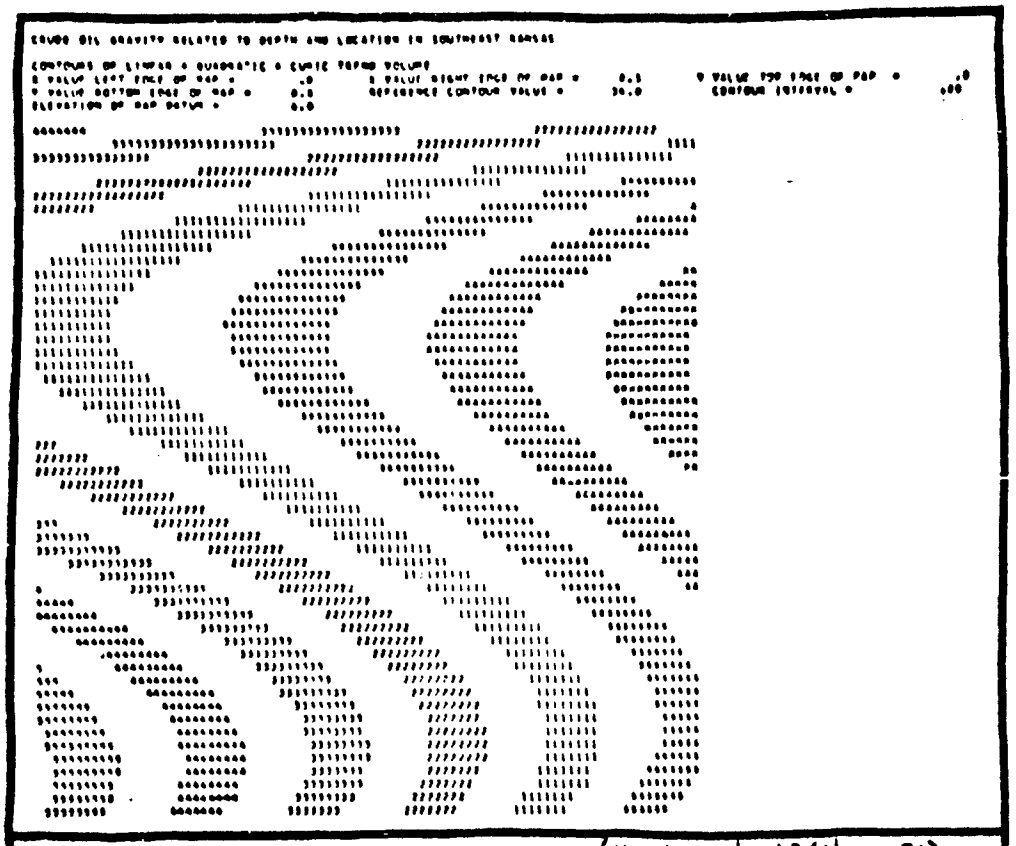
Fig. 4 Contour map produced by a digital co-ordinate plotter
Courtesy of California Computer Products, Inc. (Tobler, 1964, p.4)



(Tobler, 1966, p.7)

FIGURE 17. - Examples of computer-produced contour-type maps.

FIGURE 18. - The same disease data as Fig. 12, but presented as a manually-drawn map utilizing dot-, shading-, and contour-mapping techniques.



(Harbaugh, 1964, p.56)

FIGURE 19. - Example of a computer-produced map using shading (the alternating bands of white and black figures) and contour (the boundaries between adjacent black and white bands) technique.

INFECTION RATE (%) OF SCHISTOSOMIASIS DUE TO *SCHISTOSOMA MANSONI* IN HUMANS; data grouped by provinces; data collected 1938-1956; data taken from May, 1961, *Studies in Disease Ecology*, p. 305-313.

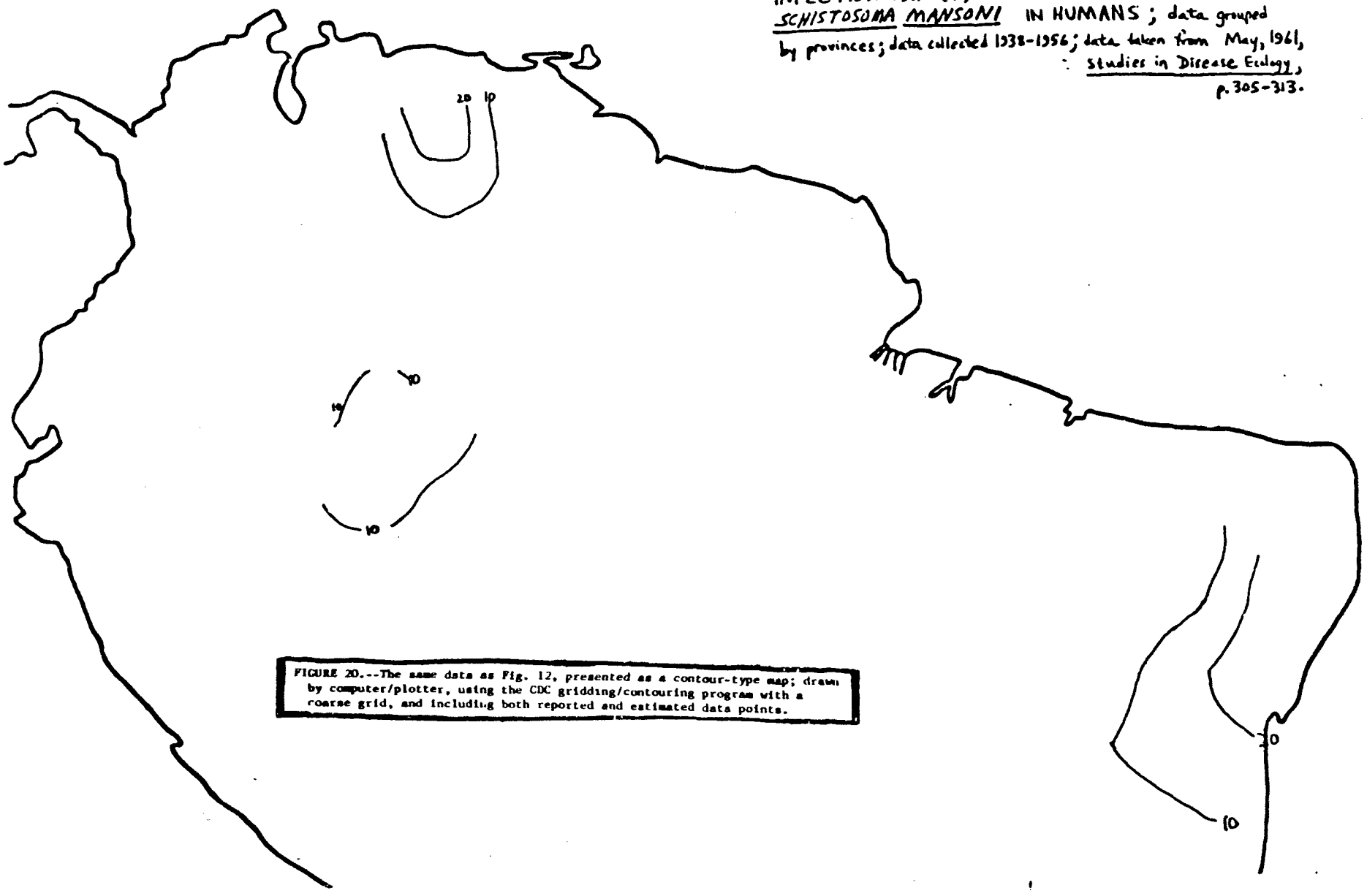


FIGURE 20.--The same data as Fig. 12, presented as a contour-type map; drawn by computer/plotter, using the CDC gridding/contouring program with a coarse grid, and including both reported and estimated data points.

INFECTION RATE (%) OF SCHISTOSOMIASIS DUE TO *SCHISTOSOMA MANSONI* IN HUMANS; data grouped by provinces; data collected 1938-1956; data taken from May, 1961, *Studies in Disease Ecology*, p. 305-313.

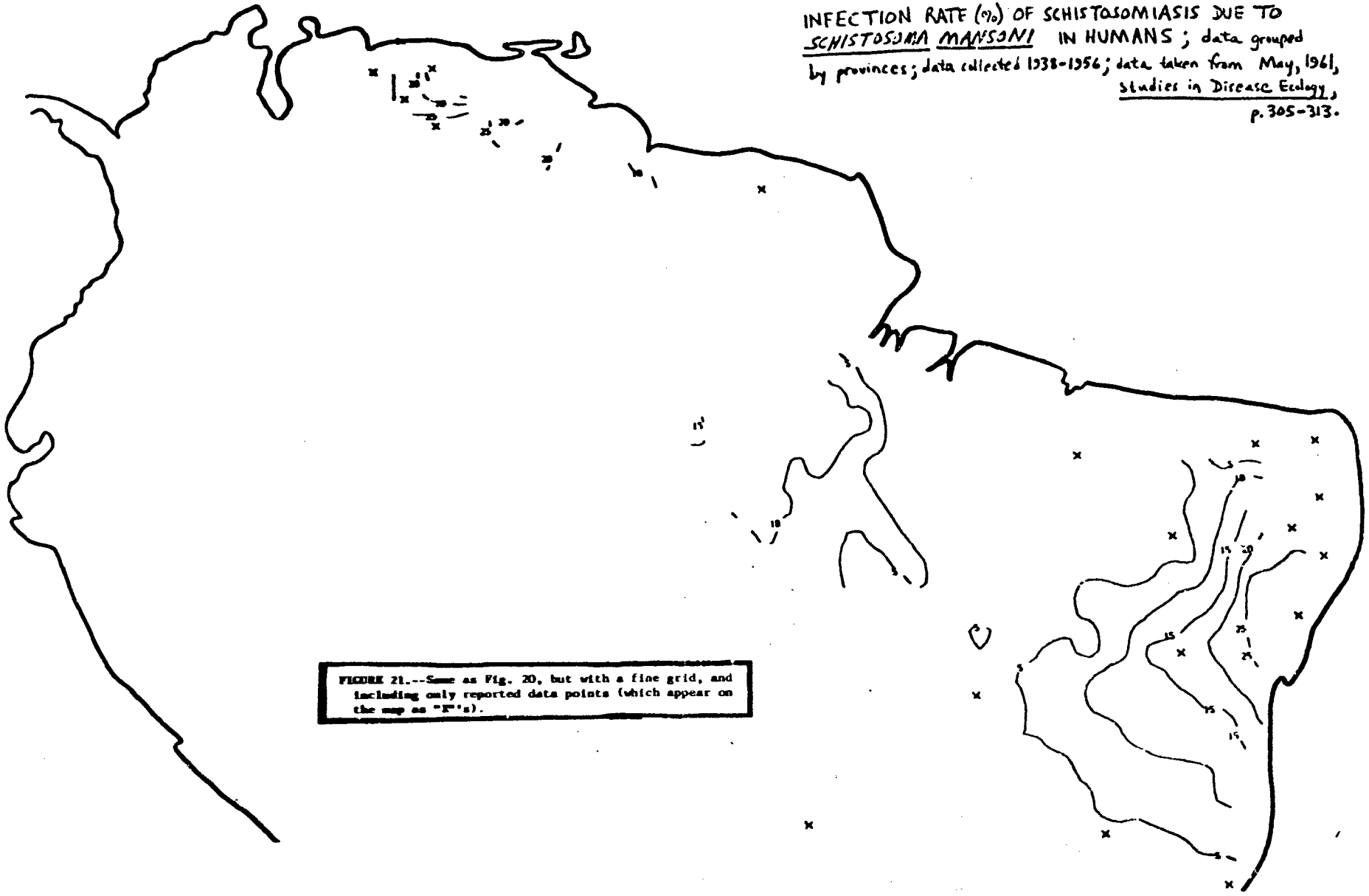


FIGURE 21.--Same as Fig. 20, but with a fine grid, and including only reported data points (which appear on the map as "X"'s).

INFECTION RATE (%) OF SCHISTOSOMIASIS DUE TO *SCHISTOSOMA MANSONI* IN HUMANS; data grouped by provinces; data collected 1938-1956; data taken from May, 1961, *Studies in Disease Ecology*, p. 305-313.

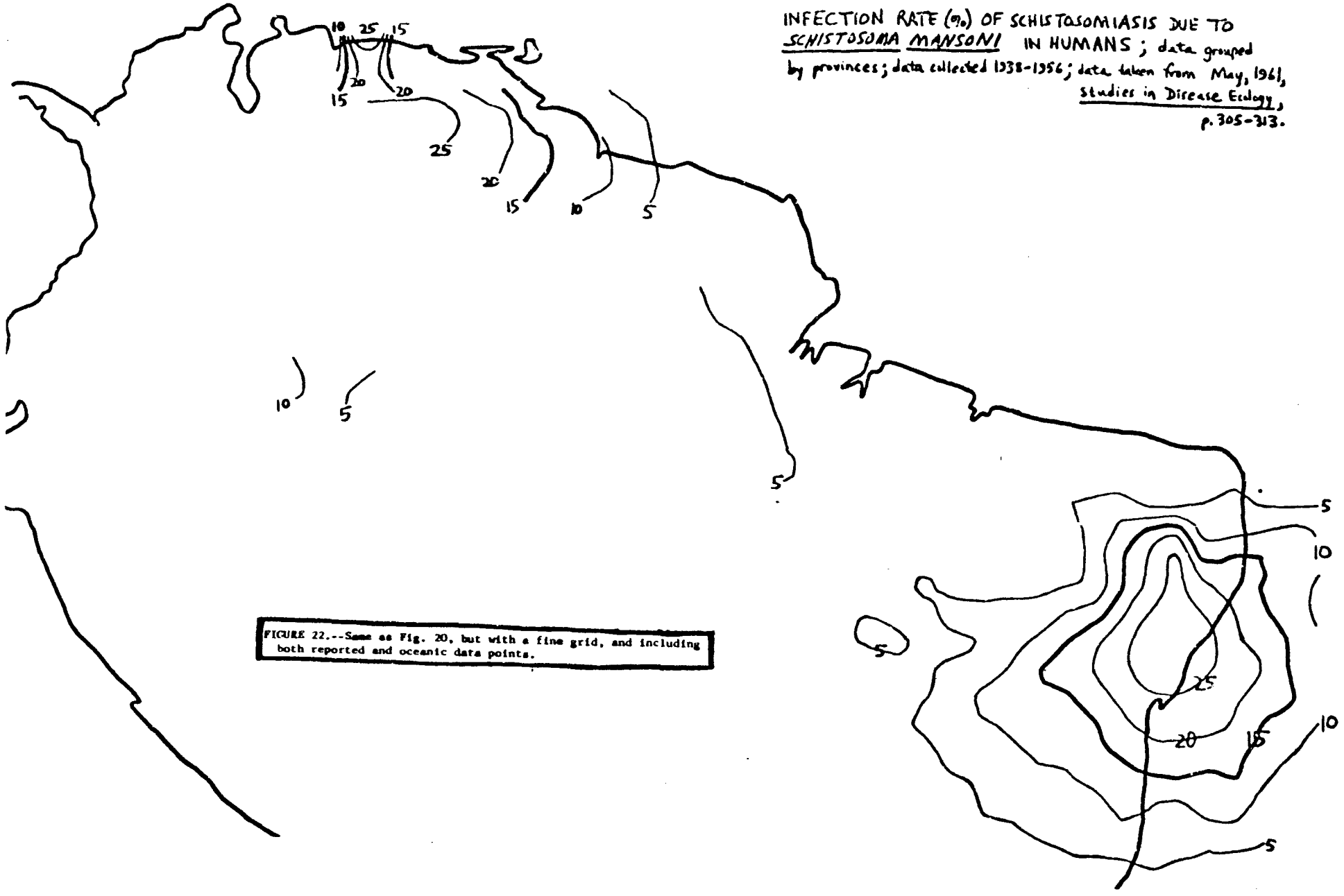


FIGURE 22.--Same as Fig. 20, but with a fine grid, and including both reported and oceanic data points.

FIGURE 23.--Same as Fig. 20, but with a fine grid, and including reported, estimated, coastal, and oceanic data points.

[illegible] computerized mapping system*.

These disease data were contoured manually (figures 16 and 18) by different members of the AFIP-USREP-FRC study team. The data were also contoured automatically (figures 20-23) at the Rockville Data Center of the Control Data Corporation, using a commercially available computer program: processing by a CDC 3600 computer followed by off-line plotting on a Calcomp drum-type plotter. The reported (actual) disease data was combined with a variety of estimated data in order to explore various techniques for presenting the data and to evaluate the effects of various limitations. The data actually reported covered only provinces in Brazil, Surinam, and parts of Venezuela. Additional points were estimated for all other South American provinces (based on actual data for neighboring regions - an interpolative process, in a sense). A set of coastal points and a set of oceanic points (each with a zero value) were also prepared.

* * * * * *

Essentially, three tasks are performed by the CDC programs. The first is to add points in areas of sparse distribution; a linear interpolation is made between two adjacent data points to add control points at levels which do not exist in the original data. The second task is to calculate grid or mesh point values for a rectangular grid. The method used in determining each grid value is to find the nearest known control points which surround the grid point and then to calculate the value by an inverse distance function. Calculation of grid values is done only when control is present. The third task is to contour automatically the gridded data. These data are expressed in x, y coordinates with a z value

* The word, system, used in this context is by no means a synonym for software/hardware components; data management aspects also form an essential "component". As a matter of fact, our most significant achievements toward producing computer/plotter output maps relate to the General Data-Analyses Vocabulary and the Factor Catalog presented as Appendix 2 and 3, respectively.

for contouring. The control values are stored within a matrix through which the program traces, interpolating to find the points through which the contour lines pass. The contouring is performed in strips of two adjacent rows of the matrix. Contouring is not performed unless control points exist. The results of two parabolic interpolations are traced to compute the path of each contour line. As positions are calculated, plotter commands are stored in an internal array, and output onto the plotter drive tape each time the array is filled. Optionally, the location of the data points, values and the grid lines may be plotted.

The size of the grid is a variable which can be entered as an additional requirement, supplementing the program. When a coarse (large-mesh) grid size was specified, the data was averaged together, and significant points were lost. However, the broad trends in the data remained clearly evident, particularly when both the reported disease-data points and the estimated-to-be-zero data points (for the other provinces) were processed (Fig. 20). When a fine (small-mesh) grid was specified, resulting in no averaging or loss of data points, the contours produced depended upon the choice of data points processed. Reasonable maps were obtained with reported data points alone (Fig. 21), and with reported plus oceanic data points (Fig. 22). However, when reported, estimated, coastal, and oceanic data points were all included, very unrealistic contours resulted (Fig. 23), perhaps due to the very large number of data points which would have had to be generated in order to fill the grid.

Our analysis of these maps indicates that the CDC method of computing the "fill-in" data points is unsatisfactory for the MOD project. This problem is still under investigation, and we plan to use other available contouring routines, e.g., the program available at the Naval Oceanographic Office. We plan also to investigate programs utilizing Tobler's method of completing the grid. Tobler's method, simply stated, is to use only the three closest points which surround the grid point in question through which to fit a plane. The grid point therefore lies on this plane and its value may be computed.

In conclusion, the actual mapping efforts which have been described above show that computer contouring of disease/environmental data <u>can</u> be successfully done <u>when</u> properly selected and characterized data can be given quantitative values. This important restriction brings us back once again to the importance of data management considerations.

CONCLUSIONS AND RECOMMENDATIONS

The foregoing pages detail the accomplishments of the first year's work on the computerized Mapping of Disease (MOD) Project.

These accomplishments are summarized on Page three of this report.

In addition to presenting accomplishments of the past year, the report discusses, in detail, the equipment, programs, and personnel which will be required to develop the MOD system to a fully operational capability during the next two years, and considers potential output of the fully operational system.

We sincerely believe that the project is progressing well, that the likelihood of success is great, and that efforts to complete it should be made along the lines indicated in this report.

BIBLIOGRAPHY

_____, June 1964, Digital Plotting Newsletter, California Computer Products, Inc., p. 1.

Ferguson, J., & Morenoff, J., 22 Aug 1966, Mapping of Disease (MOD) Project, Phase I: Final Report prepared for Universities Associated for Research and Education in Pathology, Inc., by Planning Research Corp., Washington, D. C.; 38 p.

Harbaugh, J. W., 1964, A computer method for four-variable trend analysis illustrated by a study of oil-gravity variations in southeastern Kansas: State Geological Survey of Kansas, Bulletin 171, 58 p.

May, J., ed., 1950-55, Atlas of Diseases; American Geographical Society, N. Y.

Rodenwaldt, E., 1952, World-Atlas of Epidemic Disease; Heidelberg.

Soper, H. H., 1964, Mapping the distribution of plants by machine: Canadian Journal of Botany, V.42, p.1087-1100.

Tobler, W. R., 1964, Automation in the preparation of thematic maps: reprinted from Journal of the British Cartographic Society, June 1964, 7 p.

_____, Jan. 1966, Notes on the analysis of geographical distributions: Michigan Inter-University Community of Mathematical Geographers, Discussion Paper No. 8, pt. 2, p. 1-16.

GEOGRAPHIC DISTRIBUTION OF INFECTIOUS DISEASES

EXPENDITURES REPORT
November 15, 1965 - November 14, 1966

DIRECT COSTS:

Direct Salaries	$ 8,267.45	( )
Social Security, Group Insurance and Other Fringe Benefits	527.60	(1)
Travel	130.44	
Equipment	4,607.65	(2)
Communications	43.08	
Consulting Fees	50.00	
Supplies and Services	950.82	
Books and Periodicals	414.49	
Subcontract - Planning Research Corporation	24,769.25	(3)
TOTAL DIRECT COSTS	$39,760.78	

INDIRECT COSTS:

Indirect Charges	3,140.28	(4)
TOTAL COSTS	$42,901.06	

NOTES:

1. Schedule of Personnel attached (Schedule 1)

2. Equipment:

1 Flexowriter with Desk Assembly ...	$ 2,895.00
1 Separator	1,600.00
1 Single Faced Filing Unit	112.65
	$ 4,607.65

3. Planning Research Corporation - Subcontract:

UAREP - 66.1 - $12,555.25 (7/5/66 - [illegible]/25/66)

UAREP - 66.2 - $12,214.00 (8/26/66 - 11/14/66)

$24,769.25

EXPENDITURES REPORT

NOTES: (Continued)

4. Audit performed by DCAA which established overhead rate of 37.79% of Salaries and Wages for the period ending December 31, 1965. Present billable provisional overhead rate is 38% of Salaries and Wages.

5. This report includes all costs incident to performance of the contract for the period November 15, 1965 through November 14, 1966.

6. Supporting vouchers and other documents on file are available for audit.

* *

(Schedule 1)
PERSONNEL
November 15, 1965 - November 15, 1966

NAME	POSITION	SALARY	FRINGE BENEFITS
Margaret L. Chu	Research Librarian	$5,175.00	$387.40
Shirley K. Eisenberg	Asst.Res.Librarian	1,015.05	52.95
Gary G. Gullet	Research Assistant	514.90	21.63
Harold M. Kline	Systems Analyst	1,562.50	65.62
	TOTALS	$8,267.45	$527.60

APPENDIX 1

ORGANIZATIONS CONSULTED DURING

FIRST YEAR OF MOD PROJECT

IN RELATION TO DATA AND/OR DATA PROCESSING

American Geographical Society

American University, Dept. of Geography

Arco Corp.

Atlantic Research Corp.

Auerbach Corp.

Benson-Lehner, Inc. (B-L Plotters)

BioSciences Information Service of Biological Abstracts (BIOSIS)

Bowman-Gray School of Medicine, Pathology Records Retrieval Program

Bunker-Ramo Corp.

California Computer Products, Inc. (CalComp Plotters)

Catholic University of America, Dept. of Geography

Chemical Abstracts Service

Control Data Corp. (CDC)

Electronic Associates, Inc. (EAI Plotters)

FMA, Inc.

General Motors Corp., Allison Div.

George Washington University, Dept. of Geography

Georgetown University, School of Medicine

Geo-Space Corp.

Gerber Scientific Instrument Co. (Gerber Plotters)

Harvard University, Laboratory for Computer Graphics

Howard University, Dept. of Geology and Geography

Illinois Natural History Survey

Illinois State Geological Survey

Indiana University, Dept. of Astronomy and Dept. of Geology

International Business Machines Corp. (IBM)

London School of Hygiene and Tropical Medicine,
Dept. of Parasitology

McLean Paleontological Laboratory

Planning Research Corp. (PRC)

RAND Corp.

System Development Corp. (SDC)

Systems Research Group, Inc. (SRG)

Thailand Govt., Royal Thai Army, Medical Service

United Kingdom Govt., Ministry of Overseas Development,
Dept. of Technical Cooperation

U. S. Govt.:

- Dept. of Agriculture, Washington Computer Center
- Bur. of the Census, Computer and Data-Processing Dept.
- Central Intelligence Agency (CIA), Medical Division
- Clearinghouse for Federal Scientific and Technical Information (CFSTI)
- Dept. of Defense:
  - Aeronautical Chart and Information Center (ACIC)
  - Air Force Technical Applications Center (AFTAC)
  - Armed Forces Institute of Pathology (AFIP), Automatic Data Processing Section

Armed Forces Pest Control Board (AFPCB)

Army Map Service (AMS)

Army Materiel Command (AMC), Foreign Science and Technology Section

Army Materiel Command (AMC), Systems Development and Design Division

Army Natick Laboratories, Earth Sciences Division

Army Research Office (ARO)

Defense Documentation Center (DDC)

Defense Intelligence Agency (DIA)

Military Entomological Information Service (MEIS)

Naval Command Systems Support Activity (NAVCOSSACT)

Naval Oceanographic Office (NAVOCEANO)

Naval Weapons Laboratory (NWL)

Walter Reed Army Institute of Research (WRAIR)

Geological Survey

Library of Congress, Map Division

Library of Congress, National Referral Center for Science and Technology

National Aeronautics and Space Administration (NASA), Goddard Space Flight Center

National Bureau of Standards (NBS), Center for Computer Science and Technology and Computer Sharing Exchange

National Institutes of Health (NIH), Division of Computer Research & Technology, Environmental Health Division, National Cancer Institute, and National Institute of Allergy and Infectious Diseases.

National Library of Medicine (NLM)

National Oceanographic Data Center (NODC)

Public Health Service (PHS), Communicable Disease Center (CDC)

Smithsonian Institution, Natural History Museum, Dept. of Invertebrate Paleontology and Dept. of Vertebrate Zoology

Weather Bureau

UNIVAC Division (of Sperry-Rand Corp.)

University of Buffalo, School of Medicine, Computer Center

University of Illinois, Dept. of Computer Science, Dept. of Forestry, Dept. of Geography, Division of Human Etiology, School of Veterinary Medicine, Center for Zoonoses Research

University of Kansas, State Geological Survey of Kansas

University of Maryland, Dept. of Geography and School of Medicine

University of Michigan, Dept. of Geography

University of Missouri, College of Medicine, Computer Center

Woodard Research Corp.

APPENDIX 2

GENERAL DATA-ANALYSIS VOCABULARY

LOF (Low-Order Factor). . . Definition: The most specific name or description of a particular disease/environmental situation.

Examples: Point prevalence, period prevalence, incidence, leptospirosis, schistosomiasis, *L. pomona*, *L. canicola*, *L. hardjo*, *S. mansoni*, *S. japonicum*, raccoons, skunks, foxes, isolated from urine, isolated from blood, serologic tests, etc.

MOF (Middle-order Factor) ... Definition: The set of all LOFs which describe the same aspect of disease/environmental situations.

Examples:

MOF		LOFs Making Up the MOF
Kind of epidemiologic index	...	Point prevalence, period prevalence, incidence
General kind of disease	...	Leptospirosis, schistosomiasis
Specific disease agent	...	*L. pomona*, *L. canicola*, *L. hardjo*, *S. mansoni*, *S. japonicum*
Animal hosts involved	...	Raccoons, skunks, foxes
Method of diagnosis	...	Isolated from urine, isolated from blood, serologic tests

HOF (High-Order Factor) ... Definition: A specific combination of LOFs, no two LOFs being drawn from (belonging to) the same HOF.

Examples:

Specific HOF	Kind of Epidemiologic Index*	General Kind of Disease*	Specific Disease Agent*	Animal Host Involved*	Method of Diagnosis*
HOF 1	Incidence of	leptospirosis	due to L.pomona	in skunks	as determined by isolation from urine
HOF 2	Incidence of	leptospirosis	due to L.canicola	in skunks	as determined by isolation from urine
HOF 3	Point prevalence of	leptospirosis	due to L.hardjo	in raccoons	as determined by serologic tests
HOF 4	Period prevalence of	schistosomiasis	due to S.mansoni	in foxes	as determined by isolation from blood

Using only the LOFs and HOFs listed on the preceeding page, we can construct

3 x 2 x 5 x 3 x 3 = 270 HOFx.

*Specific HOFs

POF (Poly-Order Factor) ... Definition: A specific combination of LOFs, to which at least one MOF has contributed more than one LOF.

Examples:

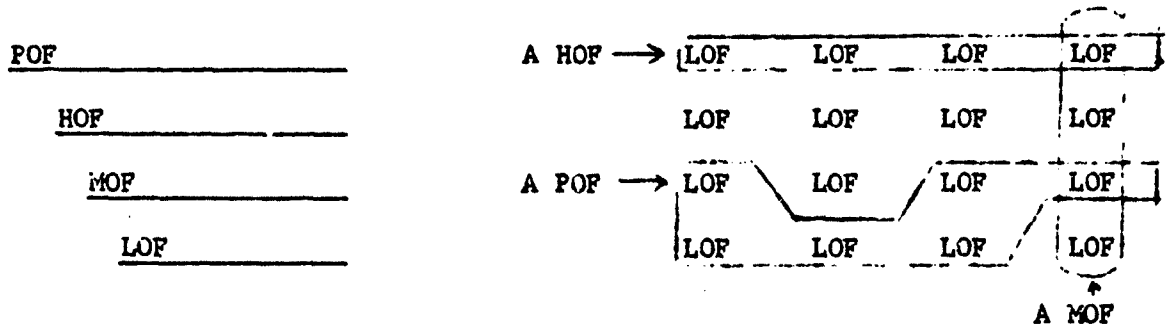
Specific	Kind of Epidemiologic Index	General Kind of Disease	Specific Disease Agent	Animal Host Involved	Method of Diagnosis
POF 1	Incidence of	leptospirosis	due to L.pomona and L.canicola	in foxes	as determined by isolation from blood
POF 2	Incidence of	leptospirosis	due to L.pomona, L.canicola, L.hardjo, and all other L.	in skunks	as determined by isolation from urine, isolation from blood, and serologic tests
POF 3	Point prevalence of	leptospirosis and/or schistosomiasis	due to L.pomona, L.canicola, L.hardjo, and all other L., and/or to S.mansoni, S.japonicum, and all other S.	in animals i.e., raccoons, skunks, foxes, and all other animal hosts	as determined by isolation from urine, isolation from blood, and serologic tests

Using only the LOFs and MOFs listed on page 2A-1 we can construct

7 x 3 x 27 x 7 x 7 = 27,783 POFs

POF/HOF/MOF/LOF together can be viewed as a kind of hierarchy or a kind of matrix.

Example:



With regard to disease data, LOFs and MOFs in general cannot be meaningfully mapped because, by themselves, they do not convey enough information. However, HOFs and POFs can be meaningfully mapped, with each HOF or POF yielding one map. Sometimes a HOF can consist of only one MOF, which in turn can consist of only one LOF. Thus, it is possible for a HOF/MOF/LOF structure to consist of a single description or name. This situation can be viewed as a uni-LOF uni-MOF HOF, or as a LOF which is also a MOF which, in turn, is also a (mappable) HOF.

Example:

Multi-LOF HOF 1 - Point prevalence of leptospirosis due to *L. pomona* in foxes

Uni-LOF uni-MOF HOF 1 - Type of bedrock

Uni-LOF uni-MOF HOF 2 - Total annual rainfall

In order to map HOFs and POFs, CELs and FAVs must be coupled with them.

CEL (Common Elements) . . . Definition: The set of items which are necessary to describe every bit of data. These items do not fall into an order or hierarchical relationship and are therefore different from the POF/HOF/MOF/LOF structure. The five following CEL items will suffice for illustration (see Appendix 3).

Item	Example
1. Geographic location by (λ, L) or (LO,LA)	W088°31', N37°29'; E179°01', S17°09'.
2. Geographic location by political unit	North America, U.S.A., Illinois, Pope County; Europe, France, Dordogne, Les Eyzies.
3. Time period for which the data applies	1901-1910; January 1961.
4. Reliability of the data	More reliable; less reliable.
5. Source document number	00087; 00243.

FAV (Factor-Value) . . . Definition: An alphabetic and/or numeric symbol expressing one member of the set of all possible results, the result-set describing, in effect, the functional relationship between a specific HOF or POF and a specific CEL. FAVs, like CELs, do not fall into a hierarchical type of order or relationship and are therefore different from the POF/HOF/MOF/LOF structure.

FAV (Factor-Value) continued — Examples:

0, 1, 2, 3, ..., ∞;

0, 0.01, 0.02, 0.07, ...,∞ ;

0-10, 10-20, 20-30, ...;

Absent, present;

Absent, rare, common, abundant;

Shale, limestone, sandstone, granite.

Factor . . . . . — Definition: A general term including POFs, HOFs, MOFs, and/or LOFs; i.e., in essence, a name or description of some aspect(s) of disease/environmental situation(s).

Value . . . . . — Definition: A general term including FAVs (and some other items not yet precisely defined).

DATA POINT . . . . . — Definition: The combination of a specific CEL, a specific POF or HOF, and a specific FAV; in essence, a specific geographic locality (CEL - geographic location) where, for a given time point or interval (CEL - time frame), some person(s) (CEL - source document) determined (CEL - reliability) (or observed or measured) a specific value (FAV) for a specific factor (HOF or POF).

Examples:

*Data Point n	Geographic location by (λ, L) / Geographic location by political unit / time period / reliability / source document // factor (POF or HOF) // value (FAV).
Data Point 1	W088°31' N37°29' / North America, U.S.A., Illinois, Pope County, Dixon Springs Experimental Station / 14 June 1962 / more reliable data / 00087 // point prevalence of leptospirosis due to L. pomona in foxes as determined by isolation from urine (a HOF) // = 1/13.
Data Point 2	W091°36' N40°33' / North America, U.S.A., Iowa, Burlington County, Keokuk / 13 January 1966 - 27 July 1969 / less reliable data / 00107 // period prevalence of leptospirosis due to L. pomona and L. canicola and L. hardjo in raccoons and foxes as determined by isolation from urine and isolation from blood and isolation from tissue (a POF) // = 11/4337.
Data Point 3	W090°00' N40°00' / North America, U.S.A., Missouri, St. Louis County, Crystal City / 1951 / reliability unknown / 17734 // qualitative estimate of leptospirosis due to L. pomona and L. canicola in skunks as determined by isolation from urine, isolation from blood, isolation from tissue, and serologic tests (a POF) // = Rare.
Data Point 4	W093°15' N40°01' / North America, U.S.A., Kansas, Riley County, Manhattan / 1962 / more reliable / 01072 // lithologic type of bedrock (a uni-LOF uni-HOF HOF) // = Limestone interbedded with shale.

* All data points follow this format.

NAR (Narrative) . . . . . <u>Definition</u>: Supporting nonmappable, narrative or textual information or data associated with a specific data point - information useful to the person examining the map in order to increase his understanding of the data mapped.

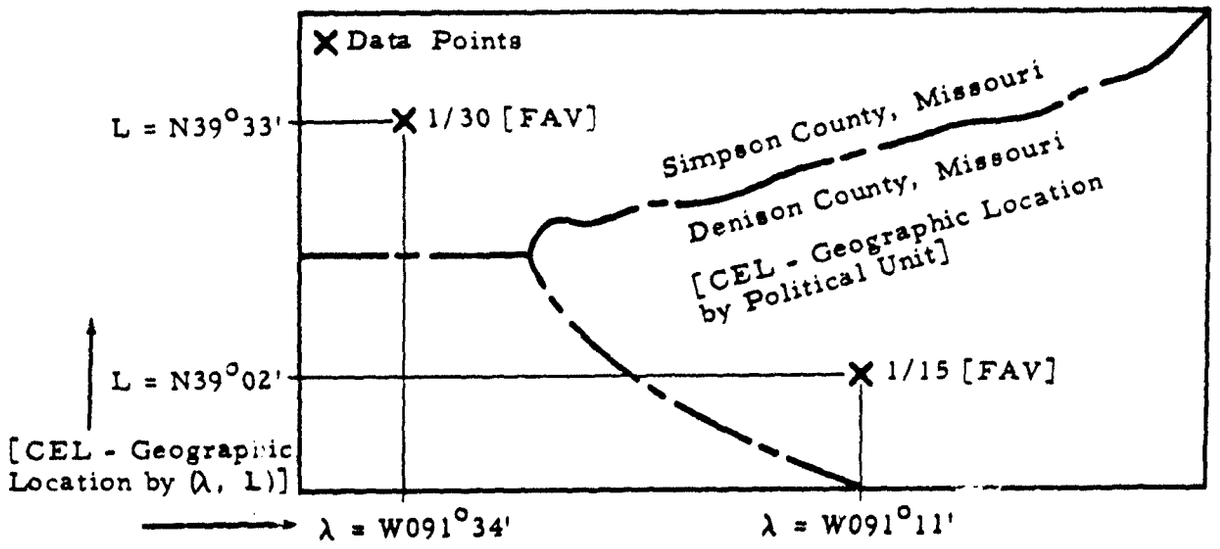
Thus, a complete record for a particular data point would include these items:

CEL // POF or HOF (MOFs and LOFs) // FAV /// NAR (+)

The specific examples presented schematically as Map 1 and Map 2 illustrate some of the difficulties in mapping so relatively simple a factor as disease (leptospirosis) incidence. Note that "Incidence of leptospirosis" and "incidence of leptospirosis due to all leptospires" must both be treated as a logical sum: "Incidence of leptospirosis due to L. pomona and L. canicola and ... (all other LOFs specifying L. species)."

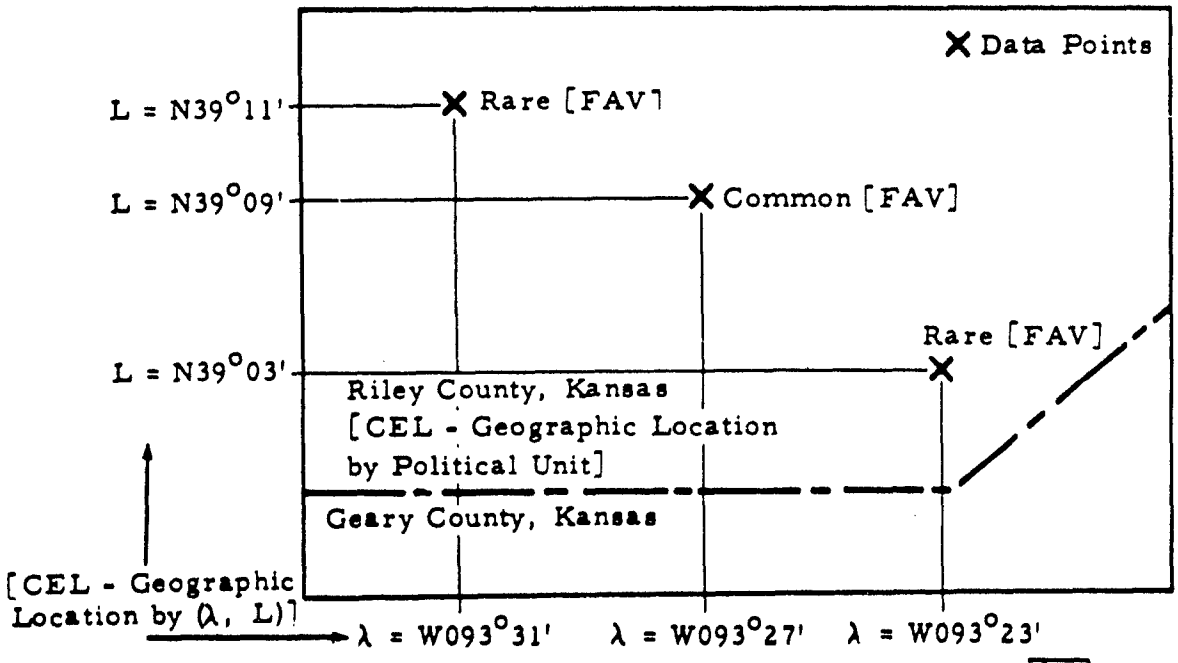
In relation to this presentation, it is appropriate to define map, in the context of the MOD program. A map is considered to be a graphic representation of data distributed meaningfully in relation to geographic coordinates.

Map 1 - Point prevalence of leptospirosis due to L. pomona in foxes as determined by isolation from urine [a HOF], based on more reliable data [CEL - reliability] for 1960-1965 [CEL - time period] taken from all source documents (i.e., from documents 00001, 00002, ...) [CEL - source document].



Note: (This information can be displayed as dots, shadings, or contours.)

Map 2 - Qualitative estimate of leptospirosis due to L. pomona and L. canicola and L. hardjo in raccoons and skunks and foxes as determined by serologic tests [a POF], based on more reliable and less reliable data [CEL - reliability] for 1950-1960 [CEL - time] taken from all source documents (i.e., from documents 00001, 00002, ...) [CEL - source document].



Note: (This information can be displayed as dots, shadings, or contours.)

APPENDIX 3

FACTOR CATALOG: A List of the Disease/Environmental Factors to be used in the Mapping of Disease (MOD) Project.

FACTOR CATALOG, Part I: Common Elements (CELs)

CEL - Geographic location by (, L) or (LO, LA):
Ex: (W118°37', N22°21'),
(E089°15', S00°37'), etc.

CEL - Geographic location by (political unit):
Ex: (NAmer., USAmer., Ind., Monroe Co., Bloomington),
(Afr., Ghana, Accra), etc.

CEL - Geographic location by (UTM military grid coordinates):
Ex: (37041973), etc.

CEL - Manner of reporting data:
Ex: Data reported as individual cases
Data reported grouped for city/town/village
Data reported grouped for state/province
Data reported grouped for country/large colony

CEL - Security classification of data:
Ex: Top secret
Secret
Confidential
Restricted - For official scientific use only
Unclassified

CEL - Time period for which data applies:
Ex: 1966,
1963-1965,
Mar 1964,
Jan 1950-Jun 1962,
17 Nov 58,
13 Jan 63-21 Aug 64, etc.

CEL - Reliability of the data:
Ex: Highly reliable
Not highly reliable
Undetermined

CEL - Source of the data:
Ex: 00123 p 1097,
00087 p 83-91, etc.

* * * * *

FACTOR CATALOG, Part II: Disease Factors (MOFs/LOFs)

MOF - Method of indicating extent of disease within population:

LOFs: Occurrence
Abundance
Number of cases existing at given point in time
Number of cases beginning during given time interval
Number of cases existing anytime within given time interval
Number of deaths during given time interval
Point prevalence
Incidence
Period prevalence
Mortality

MOF - General kind of disease:

LOFs: Leptospirosis (=Weil's disease = 7-day fever = etc.)
Hemorrhagic fever
Dengue

MOF - Specific disease agent or specific disease type:

LOFs: *Leptospira canicola*,
Leptospira pomona,
Omsk hemorrhagic fever,
Crimean hemorrhagic fever, etc.

MOF - Broad category of primary host infected:

LOFs: Human beings
Domesticated mammals or birds
Wild mammals or birds
Other vertebrates
Arthropods
Other invertebrates
Plants
Protists

MOF - Specific primary host infected:

LOFs: Homo sapiens,
female Australoid human beings 20-30 years old,
raccoons,
stink-pot turtle,
Chihuahua dogs, etc.

MOF - Broad category of intermediate host infected:
LOFs: (Same as for primary host)

MOF - Specific intermediate host infected:
LOFs: (Same as for primary host)

MOF - Broad category of reservoir infected:
LOFs: (Same as for primary host)

MOF - Specific reservoir infected:
LOFs: (Same as for primary host)

MOF - Broad category of carrier infected:
LOFs: (Same as for primary host)

MOF - Specific carrier infected:
LOFs: (Same as for primary host)

MOF - Broad category of vector infected:
LOFs: (Same as for primary host)

MOF - Specific vector infected:
LOFs: (Same as for primary host)

MOF - Method of transmission to primary host:
LOFs: Direct contact with living infected animals
Direct contact with dead tissue or blood
Direct contact with excreta
Indirect occupational contact with water
Indirect recreational contact with water
Indirect domestic contact with water
Indirect occupational contact with soil
Indirect recreational contact with soil
Indirect domestic contact with soil
Bite of vector or carrier

MOF - Epidemiologic state of disease within population:
LOFs: endemic or enzootic
hyperendemic or hyperenzootic
sporadic
epidemic or epizootic
pandemic or panzootic

MOF - Kind of outbreak reported:
LOFs: isolated case; one case
smaller group of cases; 2-29 cases
larger group of cases; 30 or more cases

MOF - Duration of outbreak reported:
LOFs: 10 days,
7 weeks, etc.

MOF - Type of medical facilities involved in treatments:
LOFs: Military hospital or clinic,
University/academic hospital or clinic
Large/urban hospital or clinic
Small/rural hospital or clinic
Individual doctor
Nurse, paramedical person, e.g., aid or "dresser"
Folk or witch doctor
None

MOF - Lethality of disease in outbreak reported:
LOFs: Always fatal
Often fatal
Seldom fatal
Rarely fatal
Never fatal

MOF - Average severity of disease in outbreak reported:
LOFs: Fatal
Severe clinical
Moderate clinical
Mild clinical
Subclinical or asymptomatic

MOF - Average course of disease in outbreak reported:
LOFs: Acute
Subacute
Subchronic
Chronic

MOF - Immunity (relative) of hosts infected:
LOFs: Susceptible or not immune
Naturally immune
Artificially immunized

MOF - Type of medical facilities involved in diagnosis:
LOFs: (Same as medical facilities involved in treatments)

MOF - Method of diagnosis:
LOFs: Clinical observation
Isolation of organism from water
Isolation of organism from soil
Isolation of organism from wine
Isolation of organism from blood

Isolation of organism from other body fluid or excretory product
Isolation of organism from tissue
Serologic tests
Xerodiagnosis
Biopsy
Autopsy

MOF - Type of sample diagnosed:
LOFs: Randomly selected individuals
Individuals selected because of their sickness/health
Individuals selected because of their occupation
Individuals selected because of their recreational habits
Individuals selected because of their domestic habits
Individuals selected because of other characteristics (including grouping based on social structure,e.g., family)

MOF - Size of sample diagnosed:
LOFs: 100 individuals; etc.

MOF - Type of subpopulation sampled for diagnosis:
LOFs: Natives examined during visit/expedition,
Patients treated as outpatients by clinic or hospital,
Patients with suspected leptospirosis,
Sewer workers,
Military draftees,
College or University students,
Residents of odd-numbered street addresses,
Live-trapped animals,
Dairy herd, etc.

MOF - Size of subpopulation sampled for diagnosis:
LOFs: 1,000 individuals, etc.

MOF - Type of total population sampled for diagnosis:
LOFs: Urban or larger city (human beings)
Suburban or smaller town
Densely settled rural
Sparsely settled rural
Concentrated animals (humans)

MOF - Size of total population sampled for diagnosis:
LOFs: 10,000 individuals, etc.

* * * * * * *

FACTOR CATALOG, Part III: Environmental Factors (MOFs/LOFs)

This is still under development. It will include MOFs/LOFs dealing with the following kinds of factors. Obviously, emphasis will be on those factors considered to be most pertinent.

Soils: Types
Temperature
Moisture content
Chemical-mineral content (including trace elements)

Bedrock: Types
Structure
Chemical-mineral content (including trace elements)

Topography, relief, elevation, or altitude

Landforms

Water: Types (soil, surface, ground)
Temperature
Chemical analyses (including pH, salinity)
Pollution or sewage?

Evapotranspiration, pan evaporation

Climate types

Weather types

Temperature: High, low, mean, ranges, etc.
Monthly, annual, average annual, etc.

Precipitation: Total amount (monthly, annual, average annual, etc.)
Seasonal distribution
Types (rain, snow, etc.)
Frequency and duration of dew formation

Humidity

Clouds and fog, clarity or transparency of atmosphere

Illumination, days of sunshine, insolation

Winds: Direction
Frequency
Severity or force

Barometric pressure

Atmospheric pollution

Natural disasters (hurricane, flood, dust storm, drought, etc.)

Magnetism (terrestrial)

Lightning (static electricity)

Solar radiation

Cosmic-ray radiation

Organisms (pertinent) occurring in same area as the disease under consideration, e.g., names of taxonomic groups, including wild and domesticated species, and including vertebrate animals, invertebrate animals, plants, and protists)

Distributions (geographic) of all such organisms, and their abundances

Degree of concentration versus dispersal of animal populations in the area

Disease or health conditions of such organisms

Special attention will be given to known or potential:

- Intermediate hosts
- Reservoirs
- Vectors
- Accidential hosts
- Artificial or experimental hosts

Insecticide or drug resistance among such organisms

Local habitats (grassland, swamp, desert, forest, etc.)

Biomes (tropical rain forest, temperate forest, northern coniferous forest)

Biogeographic region

Population:

- Total
- Density
- Settlement patterns, or type of settled area (large city, small city, camp, barracks, family group, etc.)

Age distribution of population (or average year of birth)

Sex distribution of population (or average year of birth)

Types of family groupings

Average size of family groupings

Racial groups within population

Ethnic or nationality groups within population

Language groups within population

Socio-economic (including caste) groups within population

Blood-group distribution

Distribution of other human hereditary or genetic factors

Medical facilities available
- Type (large hospital, small hospital, clinic, mobile aid staition, etc.)
- Sponsorship (government, military, missionary, industriel, private, etc.)
- Ease of access to facilities
- Numbers

Medical personnel available
- Type (doctors, nurses, veterinarians, etc.)
- Numbers

Treatment of water supply

Treatment of sewage

Public health service facilities and expenditures

Other pertinent diseases-conditions, e.g., drug addiction, alcoholism, malnutrition, tuberculosis, mental disorders, etc., in the population
- At present
- In the recent past

Average person's medical, hygenic, and sanitary practices and habits

Average person's dietetic and nutritional habits

Average person's clothing habits

Average person's housing preferences and habits

Educational level of population
- Literacy
- Number of college, high school, and elementary school graduates

Land use

Type of economy (hunting-gathering, farming, machine civilization, etc.)

Basis of economy (fishing, forestry, farming, mining, manufacturing, etc.)

Occupations
- Types present
- Relative proportions (i.e., predominant jobs)
- Those involving special risk of exposure to disease in which interested

Economic levels, distribution of income, standard of living index

Communications available, degree to which used

Transportation available, degree to which used
- Types
- Average mobility of population (including migrations, travel patterns, troop movements)
- Kinds involving special risk of exposure (such as walking through jungle, fording streams, etc.)

Political movements, political views

Military organization of population (none, militia, away-from-home active duty, etc.)

Religions

Superstitions

Other customs

Artistic, literary, or musical customs and activities

Recreational and entertainment habits
- Kinds
- Frequency indulged in
- Special-risk types (water sports, hiking in jungle, etc.)

Crime statistics

APPENDIX 4

This survey was designed to evaluate off-line plotters potentially useful in the MOD Project. The plotters described here have one basic common characteristic: they all accept magnetic tape as a source of input.

The maximum speeds given for contouring are those specified by the respective manufacturer, and the accuracy/reproducibility figures are for those maximum speeds.

The plotters considered here are listed in alphabetical order, by manufacturer.

I. The BENSON-LEHNER COMPANY produces six sophisticated plotter systems. Upon request by a Benson-Lehner plotter owner, one may obtain their contouring programs, and these are presently being evaluated by the MOD Project team. The STE-LTE unit has, as a special feature, the ability to change pens (four) by program command.

A. Benson-Lehner, STE

type: flat bed
size: 30" x 30"
input: 7-track, 200/500 bits per inch (bpi) magnetic tape
accuracy: ± 0.015"
repeatability: ± 0.005"
max. speed: 2400 points per minute
price: $47,700
options
9-track, 800 bpi input ($1000)
48 character alphanumeric printer ($4,500)

B. Benson-Lehner, LTE

type: flat bed
size: 42" x 58"
input: 7-track, 200/500 bpi or 556/800 bpi
accuracy: ± .015"
repeatability: ± .005"
max. speed: 4500 points per minute
price: $52,500
options
9-track, 800 bpi input ($1000)
48 character alphanumeric printer ($4,500)

C. Benson-Lehner, MTD-105

type: drum
size: 12"
input: 7-track, 556/800 bpi magnetic tape
max. speed: 1.5" per second
price: $27,000
options
9-track 800 bpi input ($2,000)

D. Benson-Lehner, MTD-110

type: drum
size: 12"
input: 7-track, 556/800 bpi magnetic tape
max. speed: 3" per second
price: $27,000
options
9-track, 800 bpi input ($2,000)

E. Benson-Lehner, MTD-305

type: drum
size: 30"
input: 7-track, 556/800 bpi magnetic tape
max. speed: 1.5" per second
price: $29,000
options
9-track, 800 bpi input ($2,000)

F. Benson-Lehner, MTD-310

type: drum
size: 30"
input: 7-track 556/800 bpi magnetic tape
max. speed: 3" per second
price: $29,000
options
9-track, 800 bpi input ($2,000)

* * * * * * *

II The CALIFORNIA COMPUTER PRODUCTS, INC. (CALCOMP) produces eight plotters and five control/tape units. These, in combination with compatible units, result in a total of twenty plotter systems. For the purposes of this survey, the plotters and associated control units were combined and will be discussed in that fashion. The first number

will be that of the control unit; the second number will be that of the plotter.

All Calcomp plotter systems have an accuracy and repeatability of better than ± 0.05 inch.

A. Calcomp, 470-502

type: flat bed
size: 31" x 34"
input: 7/9-track, 200/556/800 bpi magnetic tape
max. speed: 3" per second
price: $32,100

B. Calcomp, 470-518

type: flat bed
size: 48" x 72"
input: 7/9-track, 200/556/800 bpi magnetic tape
max. speed: 1.5" per second
price: $50,100

C. Calcomp, 470-563

type: drum
size: 30"
input: 7/9-track, 200/556/800 bpi magnetic tape
max. speed: 2" per second
price: $23,100

D. Calcomp, 470-565

type: drum
size: 11"
input: 7/9-track, 200/556/800 bpi magnetic tape
max. speed: 3" per second
price: $19,650

E. Calcomp, 750-502

type: flat bed
size: 31" x 34"
input: 7/9-track, 200/556/800 bpi magnetic tape
max. speed: 3" per second
price: $38,200

F. Calcomp, 750-518

type: flat bed
size: 48" x 72"
input: 7/9-track, 200/556/800 bpi magnetic tape
max. speed: 1.5" per second
price: $56,200

G. Calcomp, 750-563

type: drum
size: 30"
input: 7/9-track, 200/556/800 bpi magnetic tape
max. speed: 2" per second
price: $29,200

H. Calcomp, 750-565

type: drum
size: 11"
input: 7/9-track, 200/556/800 bpi magnetic tape
max. speed: 3" per second
price: $25,750

I. Calcomp, 760-502

type: flat bed
size: 31" x 34"
input: 7/9-track, 200/556/800 bpi magnetic tape
max. speed: 3" per second
price: $45,500

J. Calcomp, 760-518

type: flat bed
size: 48" x 72"
input: 7/9-track, 200/556/800 bpi magnetic tape
max. speed: 1.5" per second
price: $63,500

K. Calcomp, 760-563

type: drum
size: 30"
input: 7/9-track, 200/556/800 bpi magnetic tape
max. speed: 2" per second
price: $36,500

L. Calcomp, 760-565

type: drum
size: 11"
input: 7/9-track, 200/556/800 bpi magnetic tape
max. speed: 3" per second
price: $33,050

M. Calcomp, 770-702

type: flat bed
size: 31" x 34"
input: 7/9-track, 200/556/800 bpi magnetic tape
max. speed: 2.25" per second (8.5" per second in the ZIP mode)
price: $60,500

N. Calcomp, 770-718

type: flat bed
size: 48" x 72"
input: 7/9-track, 200/556/800 bpi magnetic tape
max. speed: .9" per second,(3.4" per second in the ZIP mode)
price: $79,500

O. Calcomp, 770-763

type: drum
size: 30"
input: 7/9-track, 200/556/800 bpi magnetic tape
max. speed: 3.5" per second (13" per second in the ZIP mode)
price: $51,500

P. Calcomp, 770-765

type: drum
size: 11"
input: 7/9-track, 200/556/800 bpi magnetic tape
max. speed: 4.5" per second (16.9" per second in ZIP mode)
price: $47,500

Q. Calcomp, 780-702

type: flat bed
size: 31" x 34"
input: 7/9-track, 200/556/800 bpi magnetic tape
max. speed: 2.25" per second (8.5" per second in ZIP mode)
price: $65,500

R. Calcomp, 780-718

type: flat bed
size: 48" x 72"
input: 7/9-track, 200/556/800 bpi magnetic tape
max. speed: .9" per second,(3.4" per second in ZIP mode)
price: $84,500

S. Calcomp, 780-763

type: drum
size: 30"
input: 7/9-track, 200/556/800 bpi magnetic tape
max. speed: 3.5" per second (13" per second in ZIP mode)
price: $56,500

T. Calcomp, 780-765

type: drum
size: 11"
input: 7/9-track, 200/556/800 bpi magnetic tape
max. speed: 4.5" per second (16.9" per second in ZIP mode)
price: $52,500

* * * * * *

III The ELECTRONIC ASSOCIATES, INC. (EAI), produces one plotter system which will accept a magnetic tape input.

A. EAI, 3500

type: flat bed
size: 30" x 30"
input: 7-track, 200/556 bpi magnetic tape
max. speed: no speed given for a magnetic tape setup
price: $46,750

* * * * * *

IV The GERBER SCIENTIFIC INSTRUMENT CO. produces a number of plotters that come within the basic requirements of this study. The first one or two digits (from the left), a Gerber's reference number, is the control unit designation, whereas the two right-most digits specify the particular plotter.

A. Gerber, 622

type: flat bed
size: 50" x 60"
input: 7-track 200/556/800 bpi magnetic tape
max. speed: 10" per second
accuracy: ± .009
repeatability: ± .0045
price: $50,000
options
72 character print wheel ($4,800)

B. Gerber, 632

type: flat bed
size: 4' x 5'
input: 7-track, 200/556/800 bpi magnetic tape
max. speed: 33" per second
accuracy: ± .0025
repeatability: ± .0013
price: $55,000
options
72 character print wheel ($4,800)

C. Gerber, 675

type: flat bed
size: 5' x 8'
input: 7-track, 200/556/800 bpi magnetic tape
max. speed: 10" per second
accuracy: ± 0.009
repeatability: ± 0.005
price: $68,000
options
a) 5' x 12' plotter size ($2,000)
b) 5' x 20' plotter size ($4,000)
c) 72 character print wheel ($4,800)

D. Gerber, 822

type: flat bed
size: 50" x 60"
input: 7-track, 200/556/800 bpi magnetic tape
max. speed: 10" per second
accuracy: ± 0.009
repeatability: ± 0.0045
price: $86,000
options
72 character print wheel ($4,800)

E. Gerber, 832

type: flat bed
size: 4' x 5'
input: 7-track, 200/556/800 bpi magnetic tape
max. speed: 3.75" per second
accuracy: ± 0.0025
repeatability: ± 0.0015
price: $91,000
option
72 character print wheel ($4,800)

F. Gerber, 875

type: flat bed
size: 5' x 8'
input: 7-track, 200/556/800 bpi magnetic tape
max. speed: 10" per second
accuracy: ± 0.009
repeatability: ± 0.005
price: $104,000
options (price)
a) 5' x 12' plotter size ($2,000)
b) 5' x 20' plotter size ($4,000)
c) 72 character print wheel ($4,800)

G. Gerber, 1022

type: flat bed
size: 50" x 60"
input: 7-track, 200/556/800 bpi magnetic tape
max. speed: 10" per second
accuracy: ± 0.009
repeatability: ± 0.0045
price: $109,000
option (price)
72 character print wheel ($4,800)

H. Gerber, 1032

type: flat bed
size: 4' x 5'
input: 7-track, 200/556/800 bpi magnetic tape
max. speed: 3.75" per second
accuracy: ± 0.0025
repeatability: ± 0.0013
price: $114,000
option (price)
72 character print wheel ($4,800)

I. Gerber, 1075

type: flat bed
size: 5' x 8'
input: 7-track, 200/556/800 bpi magnetic tape
max. speed: 10" per second
accuracy: ± 0.009
repeatability: ± 0.005
price: $127,000
options (price)
a) 5' x 12' plotter size ($2,000)
b) 5' x 20' plotter size ($4,000)
c) 72 character print wheel ($4,800)

J. Gerber, 2022

type: flat bed
size: 50" x 60"
input: 7-track, 200/556/800 bpi magnetic tape
max. speed: 10" per second
accuracy: ± 0.009
repeatability: ± 0.0045
price: $118,000
option (price)
72 character print wheel ($4,800)

K. Gerber, 2032

type: flat bed
size: 4' x 5'
input: 7-track, 200/556/800 bpi magnetic tape
max. speed: 3.75" per second
accuracy: ± 0.0025
repeatability: ± 0.0013
price: $123,000
option (price)
72 character print wheel ($4,800)

L. Gerber, 2075

type: flat bed
size: 5' x 8'
input: 7-track, 200/556/800 bpi magnetic tape
max. speed: 12.5" per second
accuracy: ± 0.009
repeatability: ± 0.005
price: $136,000
options (price)
a) 5' x 12' plotter size ($2,000)
b) 5' x 20' plotter size ($4,000)
c) 72 character print wheel ($4,800)

V. Conclusions:

Since the accuracy and repeatability of all the plotters described herein meet our requirements, these factors will not limit our selection. Neither is speed a limiting factor among the plotters described above. However, plotter size is. We will require*a plotter area of 30" x 60" or greater. The Benson-Lehner LTE, MTD-305, and MTD-310, all of the Calcomp control units with a 518, 563, 718, or 763 plotter, and all of the Gerber plotters meet this size-requirement.

If the plotter is to be used with the projected AFIP IBM 360-Model 30 computer, it must be compatible with that computer. With this further limitation, acceptable plotters are only those which can utilize a 9-track magnetic tape as input. Only the Calcomp series mentioned above meets this last requirement along with the other requirements. However, were an intermediate tape-copy device-process to be added to the system, this would permit use of several of the other plotters.

* strongly desire