M U D P I E, no. 4

Museum and University Data, Program, and Information Exchange

NEW INSTALLATIONS

Several additional time-shared installations, all of which have teletype facilities for standard telephone transmission of programs and information, have provided the following information:

University of Minnesota, St. Paul. Area Code 612; 647-3739 (Ke Chung Kim).

University of Michigan, Ann Arbor. A. C. 313; 763-3066 (A. Kluge).

AVAILABLE PROGRAMS--BASIC

9. CHI-2--Calculates chi-square using Yate's correction for 2x2 tables [Wayne Moss, Academy of Natural Sciences, Philadelphia].

10. RUTH-4--Calculates a diversity index (p_i log p_i) for two samples as well as their mean diversity [Wayne Moss, Academy of Natural Sciences, Philadelphia].

ACTIVITIES INVOLVING TIME-SHARE COMPUTING

Ke Chung Kim, Department of Entomology, University of Minnesota, St. Paul, Minnesota, 55101, is planning work on a computer program for identification of the species of Sphaerocerinae, as well as on a catalogue of the sucking lice.

Arnold Kluge, Museum of Zoology, University of Michigan, Ann Arbor, Michigan, has been using the teletype connection in the Museum both for research and for teaching. The Michigan computer does not use BASIC, but takes MAD or Fortran IV. His research involves testing and prediction of phylogenies using weighted characters, among other things, and the teaching is part of a course in evolution, where the computer is used for simulation of evolutionary situations, with each student working out individual problems.

James Ewin, Room 2C-528, Bell Telephone Laboratories, Holmdel, New Jersey, 07733, has used time-share for computer simulation of biological evolution, and recently gave a seminar and discussion of his work with members of the staff at the Smithsonian Institution. Copies of his program, called "ECSYM2" can be made available if anyone is interested.

FUNDING OF ACTIVITIES

The National Science Foundation has issued a leaflet entitled "Grants for Computing Activities," the number of which is NSF 68-4. It can be obtained from the Office of Computing Activities (Milton E. Rose), National Science Foundation, Washington, D. C. 20550.

RECENT PUBLICATIONS ON TIME-SHARE COMPUTING

Sharpe, William F. An Introduction to Computer Programming Using the BASIC Language. Free Press, New York, 1967, 148 pp. Cloth, $6.95, paper $3.95.

Uttal, William R. Real-Time Computers: Technique and Applications in the Psychological Sciences. Harper and Row, New York, 1968, 352 pp.

Higman, Bryan. A Comparative Study of Programmin g Languages. Elsevier, New York, 1967, 172 pp. $8.50.

Dowling, H. G., I. Gibson, and I. Palser. Current Herpetological Titles: 1967. Herpetological Review, no. 2, April, 1968, pp. 9-39. [A computer readout from magnetic tape...as an example of the work being done (at the American Museum of Natural History) to consolidate and make available the total output of herpetological data].

Peters, James A. A Computer Program for Calculating Degree of Biogeographical Resemblance between Areas. Systematic Zoology, 17, 1968, pp. 64-69.

Peters, J. A. and B. B. Collette. The Role of Time-share Computing in Museum Research. Curator [AMNH], 11, 1968, pp. 65-75.

CONTINUATION OF THE ABSTRACTS FROM MEXICO CITY

The abstracts from this symposium, included in part in MUDPIE 3 [of which there are still plenty of copies], are continued here.

Smithsonian Institution
June 10, 1968

14. "THE INTERNATIONAL PLANT RECORDS CENTER PROJECT"

R. D. MacDonald
University of Tennessee Arboretum
Oak Ridge, Tenn.

The arboreta and botanical gardens of the world contain fantastic collections of plants of every sort. An average number of taxa for the botanical gardens (436) which have this information available is some 3,800. The number of taxa contained in any one established garden may range from 60 to 25,000. It is reasonable to assume that a specimen of almost any cultivated plant, as well as a great number of wild plants, might be found in one or more arboretum or botanical garden.

There are about 100 different record-keeping systems now in use in North America. Yet, with one exception, for all their differences these systems are very similar in that most of the information contained is not easily retrieved or tabulated. The one exception to the general methods used for record-keeping is the system, developed at the University of Tennessee Arboretum, which utilizes electronic data processing (EDP) methods and equipment. As a result of the presentation of a paper detailing this system, an International Association of Botanic Gardens committee was formed to look into the feasibility of establishing an International Plant Records Center which would function to document and make readily available data on living collections of plant material in botanical gardens and arboreta by utilizing EDP methods and equipment.

The American Association of Botanical Gardens and Arboretums agreed to act as the sponsor for a pilot project. The purpose of the Pilot Project, was to detemine methods and costs for the proposed plant records center specifically:

1. Data Input - I. Development of standardized input formats (accession cards). 2. Data Input - II. Determination of methods, equipment, and personnal needed for recording of data presently on record at gardens.
3. Data Output. Development of information retrieval programs and output formats. 4. Data Ownership. Development of data ownership areas (as affects distribution and utilization).

Progress on the Pilot Project objectives to date is described.

15. "UNA BIBLIOGRAFIA AUTOMATIZADA PARA FLORA DE NORTEAMERICA.

S. Ahumada R.
Centro de Cálculo Electrónico.
U.N.A.M.
&
S.G. Shetler.
Smithsonian Institution,
Washington, D.C.

RESUMEN

Flora de Norteámerica será un manual conciso de diagnóstico de plantas vasculares de norteámerica al norte de México, preparado por una gran comunidad de especialistas en taxonomía en los próximos 10 a 15 años. El trabajo será coordinado y el manual compilado y editado por un Comité Editorial. Para asegurar un cubrimiento total de la literatura el Comité mantendrá una bibliografía centralizada de la cual se les pueden suministrar a los especialistas individuales listas de las referencias pertinentes antes de la preparación de tratados para la Flora. Se estima que esta bibliografía podría incluir más de un millón de entradas primariasal final del proyecto. Por lo tanto la compilación y utilización de esta bibliografía representará una tarea enorme en el almacenamiento, recuperación y diseminación de información. Para cumplir con esta tarea se está desarrollando una bibliografía automatizada la que en última instancia utilizará la tecnología de computadoras más avanzada disponible. Inicialmente, los datos se están automatizando por medio de cinta de papel la cual es leída a cinta magnética cuyo formato está preparado para búsqueda. Ha sido preparado un formato preliminar de tal cinta por el autor principal, formato que suministra campos separados para la clasificación tanto por jerarquía taxonómica como por materia. El acceso primario a los datos es por medio del nombre taxonómico y los códogos de localidad geográfico política.

Se suministra un acceso secundario para casi todos los campos de datos. En general el formato de la cinta de papel permite campos variables separados por códigos específicos. Se utilizan términos clave para clasificación de materias y se genera una tabla de referencia de términos de materias a medida que crece la bibliografía. Para la fase inicial se construirá un archivo convencional en tarjetas, con referencias cruzadas, a medida que las entradas se perforan en cinta de papel. Se anticipa que eventualmente se desarrollará un sistema para entrada y acceso en línea a la bibliografía almacenada en la computadora por medio de teletipos situados quizás en toda norteámerica. Podrán utilizarse pantallas y se espera poder eliminar el archivo en tarjetas.

16. "PREPARATION OF IDENTIFICATION KEYS BY COMPUTER FOR FLORA NORTH AMERICA"

L.E. Morse & J.H. Beaman.
Michigan State University.

And
S. G. Shetler.
Smithsonian Institution,
Washington, D. C.

Flora North America, as a manual, will have dichotomous keys for identification of all included taxa. As these keys are constructed by the individual specialists, the Editorial Committee will circulate them to taxonomists in all parts of the -- country for testing before adoption in the Flora. Thus the Committee will find it necessary to revise the keys frequently during preparation, and an efficient means is needed to accomplish this with speed and accuracy. For this purpose the senior author has designed a program for computer printing of conventional indented keys from data presented on cards in unnumbered, non-iden ted form. By adding, removing, or correcting specific cards, revisions can be made in a particular part of the key without affec ting the rest of the data. The chance for error is greatly reduced, and revised editions can be prepared very rapidly. This --- program is especially useful for abstracting smaller keys that - will cover more restricted geographic areas or taxonomic groups than the original key. As a natural outgrowth, another program - is being developed that will enable the computer to construct a useful artificial key directly from the raw descriptive data on the plants. The computer is particularly suited to this task -- since it can consider all possibilities and print sample keys - in which the most useful characters are employed in the most direct manner. The taxonomist can impose whatever criteria for --- judgment he desires. A third application of computers being studied involves direct, on-line identificatiob of specimens from - a teletype terminal connected to a central computer. Randon -- choice of characters is possible here, allowing the researcher - to use the characters that are observable on his particular specimen. The computer could either print a list of suggested identifications or request additional data to continue the process.

17. "THE USE OF DATA PROCESSING METHODS IN THE HERBARIUM"

J. Soper.
National Museum of Canada
Ottawa, Canada.

This paper describes a system developed in the herbarium of Vascular Plants (TRT) at the Botany Department, - University of Toronto, during the years 1963 - 1967 and now being introduced and exparded at the National Herbarium of Canada (CAN) in Ottawa. Much of the information has already been published (Soper & Perring, 1967) but changes are reparted in some of the procedures and formats prevlously cutlined.

Descriptions are given of the application of data-processing techiques to routine operations in a herbarium such as the preparation of (a) catalogue records or other index entries; (b) tabels for herbarium specimens; (c) - lists of exchange and loan material; (d) inventaries; (e) distribution maps. A discussion is included of some aspects of a general computerizad search program being developed - for retrieving data from the information system. Ilustra-- tions are provided to show the equipment used in the her-- barium and samples of catalogue record forms, herbarium -- labels, distribution maps and varius outnut lists.

18. "COMPUTER PROCESSING OF INFORMATION ON ECOLOGICAL SYSTEMS"

J.S. Olson & M.F. Olson
Oak Ridge National Laboratory
Oak Rdige, Tenn.

Existing computer programs for documentation and indexing of literature can be adapted for rapid dissemination of information about communities of plants, animels and environments (ecosystems) and about current research on these natural and man-modified systems. Examples include permutation indexing programs and other approaches to retrieval of scientific knowledge. Presently available programs and new ones should help in organizing existing information on the state of landscapes and waters, on their changes, and on the mutual influences between parts of any ecosystem which tend to explain stebility or instability in the system as a whole. A COMpartmental SYStem simulation program provides numerical and graphic display of the kinds of ecosystem behavior which follow from given hypotheses about the initial condition of the system and the transfers of matter or energy between its parts. It can also represent successional change of a given area, or the redistribution of area between different kinds of vegetation if the probabilities of successional change between types can be estimated.

Electronic data processing of simple kinds (punchcards, punch tapes) and more elaborate uses of computers can be very helpful in saving time and confusion about plant and animal names. One example is the "Oak Ridge, Tennessee, Flora" which was generated from punchcards and IBM 1401 to punchtape, which in turn composed the complete text and checklists in upper and lower case type. A map of trees located near the entrance of the University of Tennessee Arboretum in Oak Ridge was generated by a CALCOMP plotter. Several useful guidelines for the wider use of automation in Botanical Gardens have emerged from the International Plant Records Center Pilot Project. Many applications of similar methods can be anticipated in several sections of the International Biological Program.

19. "SIMULACION DE LOS PROCESOS EVOLUCIONARIOS Y SU APLICACION EN LA ENSEÑANZA

O. T. Sobring
Departamento de Botánica, Universidad de Michigan,
Ann Arbor, Mich. U.S.A.

Fenómenos evolucionarios contienen un elemento estadístico importante, y por lo tanto se prestan al manipuleo por medio de computadoras electrónicas. Varios modelos han sido producidos para simular las condiciones que -- producen hibridación entre dos especies, y para prede-- cir el futuro de tales fenómenos. Las variantes estudiadas son: número cromosómico, fertilidad del híbrido, - tamaño de la población y preferencias ecológicas. El - modelo ha sido adaptado para la enseñanza en el laboratorio en el curso de Evolución Orgánica en la Universidad de Michigan, usando una computadora IBM-360.

Modelos de este tipo se prestan para la interpretación de problemas taxonómicos y evolucionarios. Usando datos obtenidos a traves de la experimentación biosistemática, complementados con datos obtenidos de ejemplares de herbario, abren nuevas posibilidades para el funcionamiento del herbario.

22. " THE FUTURE TAXONOMIC BANK "

L. A. Proctor
Computer Center
Texas Technological College
Lubbock, Texas

Taxonomic revisions in the future will not be isolated afforts in time and discrete units of publication as they are today. Rether the informa - tion on a particular taxon (whether a sub-order of mites or a class of algae) will be in a state of continuous revision at some particular computer cen ter. Individuals working within the taxon will relay information (possibly via individual console units) to the "bank" and groups of taxonomists will neet for yearly or semi-yearly evaluations of progress attained and difficulties encountered, speci fic points of contention to be investigated, etc. During such sessions, distribution maps based on collection data or experimental results (i.e. cro ssing, genetic studies, measurements, etc.) will be called for and immediately displayed. Basic information on units of particular interest may be retrieved from disk-type storage units through use of a code typed on a console or with a tap from a light pen. Additional information, as required, may also be requested and retrieved.

It should be emphasized that the unit in question would be both a storehouse of information and an experimental tool storing data of value to the particular investigators and allowing comparisons and correlations now only obtained with diffi culty or not at all.

Computer usage has now reached a level of sophistication where we can begin in reality to erect such a unit. The present paper will first delineate types of information to be stored and assign priorities to this information: i.e. what is the minimum information capability required for a valuable prototype or "level one" system and by what steps would one advance to the "level five" complete taxonomic unit. Second, it will describe appropriate computing system specifications, their costs and state of development for minimal and optimal information levels.

23. "THE BRITISH BIOLOGICAL RECORDING NETWORK"

F. Perring.

Nature Conservancy Huntingdon, England

24. "AUTOMATIC DATA PROCESSING IN THE STUDY OF SEABIRD DISTRIBUTION"

W. B. King
Smithsonian Institution
Washington, D.C.

25. "A RETRIEVAL SYSTEM FOR ZOOLOGICAL COLLECTIONS"

R. G. Van Gelder.
The American Museum of Natural History
New York, New York.

A non-electronic data retrieval system for scientific study specimens of mammals has been developed and is in operation at The American Museum of Natural History. The system utilizes the optical coincidence of holes drilled in cards and is an inverted system- each card represents a characteristic, each hole represents the catalogue number of a specimen having that characteristic. The data stored in the system include identity of the specimen to species, geographyc region of collection, the sex, the month of collection, and the type of prepara - tion (skin, skull, skeleton, in spirit, etc.). Any combination of these characteristics can be retrieved, with searches being made at a rate of about 10,000 specimens per minute. The sys tem retrieves the catalogue numbers of the specimens having the characteristics sought. Input of information, including coding of taxo nomy and geography, is at a rate of about 1.5 specimens per minute , or 10,000 per work month using non-scientist help. The costs of mate - rials and machinery are considerably less than for electronic data retrieval.

28. "SOME REQUIREMENTS OF DATA PROCESSING SYSTEM FOR GEOLOGY AND PALEONTOLOGY"

J. L. Cutbill
Cambridge University
Cambridge, England

Current practice in geological research is often dictated by difficulties of communication. Methods now used in rock and time stratigraphy and in fossil classification are designed to communicate conclusions and hypotheses rather than original observations. It is seldom possible to recover original data used to reach a conclusion, or to assemble an adequate data base with which to demonstrate a major hypothesis.

Therefore new systems of data handling should not merely automate current methods but should provide effective links between conclusions and original observations. In order to design adequate systems it is necessary to know the kinds and amounts of data needed to answer major problems. Once this is known for a particular problem data collection techniques must be developed until the necessary data base can be collected within a reasonably short time and within the budget available. Systems must be developed which allow international cooperation in collection and communication of these data bases, so that existing resources can be harnessed to solve major problems. Tools must be developed to enable individuals and small teams to handle these enormous data bases during the course of research which will inevitably be unpredictable in its requirements. These systems must also enable individuals to pass easily on their data and results into the main data base.

29. "THE DATA SET FORMAT vs. I/O FORMATS FOR VERTEBRATE PALEONTOLOGY AND "INFINITELY" VARIABLE-LENGTH LOGICAL RECORDS"

J. R. MacDonald & E.D. MacDonald
Los Angeles County Museum of Nat. Hist.
Santa Monica, Calif.

Specimen records in Paleontology differ little from those of other disciplines. The temporal dimension is added as part of the locality record and often the data with older collections leave much to be desired.

Formating seems to be a general concern of curators faced with the prospect of using EDP methods. This certainly is the least of his problems. The major problem is the funding of the transfer of data from catalogues to an input medium and the actual use of a central processing unit and the necessary satellite equipment.

The Symposium is urged to recognize that EDP has matured to the level where the scientist need not concern himself with:

(1) Input format....each institution should be able to present its data to the computer's data bank in the format most convenient and inexpensive for that institution to prepare. "Control cards" preceding such data will explain the format to the input program.

(2) Output format...."Report Program Generators" are now available for virtually every make and model of computing system, and increasingly sophisticated versions of these forms of software are constantly being offered by the hardware vendors.

(3) Data set Storage format....i.e., the arrangement and coding of the items of information ("fields") within each "logical record" (e.g., the cross reference entry for one species, or the unique entry for one specimen, or the index entry for the type locality for one geologic formation) within the hardware storage device. Skilled Systems Analysts will be supplied by the hardware vendor to assist in this design; the problems of data management and maintenance are basically identical to those of inventory and production control and materials handling which industry (especially aerospace) has had fully automated for may years... including random access retrieval with and without tele-processing and/or graphics display stations, etc.

Those who will ultimately be concerned with item (3), primarily from the cost standpoint, are urged to consider the newly devised technique of the "infinitley variable-length logical record" which conserves storage space and provides incredible flexibility for the user; this software package (Data Language I - "DL/I") employs an hierarchical structure within each logical record and makes selective addition, deletion or change of fields within a record a more rapid, and therefore less expensive, process.