MODABUND—THE COMPUTERIZED MOSQUITO DATA BANK AT UNIVERSITY OF NOTRE DAME[1]

THEODORE J. CROVELLO
Department of Biology, University of Notre Dame, Notre Dame, Indiana 46556

INTRODUCTION. Many animal taxa are studied very little. This is not the case with mosquitoes. In fact, we have an embarrassment of riches because the literature is so vast that we often are not aware of all of the articles relevant to our particular research. We know that possibly 25,000 articles on mosquitoes have appeared between 1941 and 1970 in the *Mosquito News* Bibliography section. How many are not found there is uncertain.

The most common procedure by which culicidologists keep abreast of the literature is to scan the relevant journals or abstract service publications on a regular basis and record interesting articles on 3 x 5 cards, with or without notches along the edges. These notches indicate the subject matter of each article, etc. After several years of work with mosquitoes, another problem arises. It now becomes equally troublesome to determine what is present in one's personal card file, which probably numbers in the thousands. This, coupled with the positive growth of the number of new articles published each year in mosquito biology, quickly elicits a sense of frustration.

Faced with the same thing in other fields, some biologists have turned to the computer for assistance. Crovello and MacDonald (1970) list over forty such projects in systematic biology alone. Others (e.g., Arnett, 1970) ask even more basic questions; such as, should computer data banks replace formal journals. Dr. Richard Foote leads an AIBS Task Force on Information to resolve such problems. The computerized Abstracts of Entomology of BIOSIS are well known to *Mosquito News* readers.

The purpose of this paper is to describe the activity of the computerized *MO*squito *DA*ta *B*ank at the *U*niversity of *N*otre *D*ame (MODABUND). Its past, present and possible future services to you will be described. Ask yourself what you want from the ideal mosquito information retrieval system, because you must think in terms of what can be, not just what you have today.

ORIGIN. Since 1969 Notre Dame's Vector Biology Laboratory has been the recipient of an NIH Training Grant entitled, "Mosquito Biology: Genetic, Organismic, and Environmental," directed by Professor Karamjit S. Rai. Five faculty members participate, ranging from a molecular geneticist interested in the biochemistry of the female mosquito sex substance to population biologists interested in the biological control of populations of *Aedes aegypti*.

An essential part of the training grant proposal was the creation of a computerized data bank of mosquito literature references. With such a facility we felt that we could provide each trainee with a summary of the past literature in his particular area of mosquito biology, normally a very time-consuming task. MODABUND also would be used by professors in their teaching, to be more sure of not missing interesting references.

In overview, our procedure is as follows. Using a Friden 7102 typewriter that simultaneously punches a paper tape, we type (and punch) all of the references of interest to researchers in the Vector Biology Laboratory that are compiled by Mrs. Sollers-Riedel in the back of each issue of *Mosquito News* (Table 1). The

[1] This paper is respectfully dedicated to Mrs. Helen Sollers-Riedel, without whose devotion to the ordering of mosquito literature this project would not have been possible.

references chosen are transferred from the punched paper tape to a magnetic tape by computer. This amounted to more than 2,000 references in 1970 alone. Each reference is proofread and the tapes are corrected. A given decade or the entire collection then can be searched by computer for any articles of interest, e.g. on the subject filariasis and/or with the word 'chromosome' in the title. The selected references are then printed in the order desired by the mosquito worker who requested the literature search.

The next three sections describe in more detail the stages of: (1) Literature Accumulation; (2) Computerized Literature Processing; and (3) Computerized Literature Searching.

LITERATURE ACCUMULATION. Each number of *Mosquito News* is examined and all of the bibliographic information is punched for each article in the categories

TABLE 1.—Subject headings in the bibliography of *Mosquito News* from 1951–1969 and the number of references in each.

Number of references 1951–1959	Number of references 1960–1969	Subject number	Subject
*	*	1	Fumigants
*	*	2	Adulticides and larvicides
627	262	3	Various types of control
1	95	4	Sterilization methods
33	65	5	*Aedes aegypti* eradication
*	*	6	Mosquito control agency problems
36	288	7	Parasites, predators, viruses and related agents
*	*	8	Equipment
67	152	9	Attractants and repellents
252	617	10	Resistance and susceptibility
*	*	11	Insecticides and chemicals
*	*	12	Toxicology
726	1131	13	Behavior, biology and ecology
31	182	14	Genetics
236	397	15	Anatomy, morphology and physiology
305	642	16	Techniques
486	470	17	Taxonomy
424	408	18	Distribution
204	886	19	Arboviruses
240	449	20	Encephalitis
314	817	21	Filariasis
835	980	22	Malaria
*	*	23	Malaria–Cerebral
*	*	24	Malaria–Eradication
*	*	25	Malaria–Immunology
*	*	26	Malaria–Plasmodia
*	*	27	Malaria–Simian
*	*	28	Malaria–Therapeutics and antimalarials
*	*	29	Malaria–Antimalarials–resistance
43	52	30	Malaria–Vectors
320	128	31	Yellow fever
*	*	32	Reactions to bites
12	90	33	Light and other trap studies
67	126	34	Literature references and reviews
59	67	35	Biography and history
321	247	36	Subjects not covered by other headings
5639	8551	Totals for those subjects of primary interest at Notre Dame	

* Currently these subjects are being added to MODABUND. All should be available for searching by June 1973.

listed in Table 1. This includes authors, date of publication, title, citation and the subject number under which Mrs. Sollers-Riedel placed it. A Friden 7102 Communications Terminal Flexoriter was used. This is a heavy duty business machine that has a typewriter keyboard, but also records on punched paper tape whatever is typed. Table 2 provides a sample

TABLE 2.—Sample of typed references taken from the *Mosquito News* Bibliography Section, Subject 13, December 1970. Note the use of "$" and "%" as special separator symbols.

$Abul-Hab, J.% 1969.% Overwintering of *Anopheles superpictus* Grassi in Sulaimanyyah and Kirkuk Liwas in Northern Iraq (Diptera: Culicidae).% Soc. Ent. Egypte Bul. 51:243–250.% 13%

$Bond, H. A., G. B. Craih, Jr. and R. W. Fay.% 1970.% Field mating and movement of *Aedes aegypti*.% Mosquito News 30(3):394–402.% 13%

$Chao, J.% 1970.% The biological change of *Culex tarsalis* and the alteration of *Plasmodium relictum* infectivity to pigeons.% J. Parasitol. 56(4) sect. 2, pt. 1:51.% 13%

$Chauvet, G.% 1969.% Repartition et ecologie du complexe *Anopheles gambiae* à Madagascar.% Ent. Med. et Parasitol. Cah. 7(3):235–278.% 13%

$Crans, W. J.% 1970.% The blood feeding habits of *Culex territans* Walker.% Mosquito News 30 (3):445–447.% 13%

of 7102 typing from the December 1971 *Mosquito News* section on Genetics. Note that each reference begins with a special control character, here a $. This tells the computer that a new reference is beginning. Each part of the reference is separated by another control character, here a %. Thus the order of capture of a reference is: $ Author(s) % Date % Title % Citation % Subject Number %. Note that there is no fixed length to any of the fields. Also, except for the subject number from Table 1, nothing is abbreviated or coded except as was done by Mrs. Sollers-Riedel. The information is captured in upper and lower case, but since two cases usually contain no additional information and our IBM 370/155 has only one case on its printer, we only use one case in our printouts.

Unfortunately, the bibliography in *Mosquito News* was not always divided into subjects. Thus all references from *Mosquito News* volumes prior to 1951 do not contain any subject numbers. Our computer system can handle none, one, or any number of subject numbers per reference, but we have not had the manpower to assign any more than the given one. Key-punch machines producing IBM cards were *not* used because they are noisier, slower and more expensive.

COMPUTERIZED LITERATURE PROCESSING. The punched paper tapes created as a by-product of the above typing are input to the computer via a high speed optical paper tape reader (200 characters per second). The information is rearranged into "card image format" and printed out on paper and it also is "written" onto a *magnetic* tape that ultimately stores about 1600 characters per inch of tape. Thus one inch contains all of the information on about twelve references. A sample of the intermediate printout appears as Table 3. Card image format means that at this point the data are stored and printed in sets of 80 characters, from which a standard punched card deck could be generated. This arrangement serves two functions. If someone at another installation wishes a copy of all of our information, it is best to send it as card images on magnetic tape. By rough estimate, the information for all of the references from 1950–1969 could be put onto one 2400′ reel of magnetic tape. This is the equivalent of about 125,000 computer cards. But the more important reason for an initial computer arrangement as card images is for the efficiency of error correction. For example, note that there is a mistake in the spelling of the author of the second reference (Table 2, Craih instead of Craig). To correct this, only one card containing the correct information is punched. The entire reference is not reentered. The computer knows which card image to replace by the reference's inhouse serial number

TABLE 3.—A sample of the intermediate printout in card image format. Numbers at right after the year are inhouse serial numbers.

ABUL-HAB, J.	1969	MN	21882 11
OVERWINTERING OF ANOPHELES SUPERPICTUS GRASSI IN	1969	MN	21882 22
SULAIMANYYAH AND KIRKUK LIWAS IN NORTHERN IRAQ	1969	MN	21882 32
(DIPTERA: CULICIDAE).	1969	MN	21882 42
SOC. ENT. EGYPTE BUL. 51:243–250.	1969	MN	21882 53
13	1969	MN	21882 64
BOND, H. A., G. B. CRAIH, JR. AND R. W. FAY.	1970	MN	21883 11
FIELD MATING AND MOVEMENT OF AEDES AEGYPTI.	1970	MN	21883 22
MOSQUITO NEWS 30(3):394–402.	1970	MN	21883 33
13	1970	MN	21883 44
CHAO, J.	1970	MN	21884 11
THE BIOLOGICAL CHANGE OF CULEX TARSALIS AND THE	1970	MN	21884 22
ALTERATION OF PLASMODIUM RELICTUM INFECTIVITY TO	1970	MN	21884 32
PIGEONS.	1970	MN	21884 42
J. PARASITOL. 56(4) SECT. 2, PT. 1:51.	1970	MN	21884 53
13	1970	MN	21884 64
CHAUVET, G.	1969	MN	21885 11
REPARTITION ET ECOLOGIE DU COMPLEXE ANOPHELES	1969	MN	21885 22
GAMBIAE A MADAGASCAR.	1969	MN	21885 32
ENT. MED. ET PARASITOL. CAH. 7(3):235–278.	1969	MN	21885 43
13	1969	MN	21885 54

and the card number within the reference (Table 3). Once corrections have been entered, a new magnetic tape is ready for customized searching.

COMPUTERIZED LITERATURE SEARCHING. There are two possible types of printed results—those of interest to many mosquito workers (perhaps resulting in a formal publication) and those of interest to smaller groups, or even of interest to only one person (they would receive the actual computer printout). Currently we can search the references of two decades, 1951–1969, but for simplicity let me use the results of searches in just the decade of 1960–1969 as an example.

Searches of general interest might include a full list of the 8,551 articles of the decade, arranged by author. Another would be a list organized first by subject number and then alphabetical by author *within* each subject e.g. (Table 4). This would answer a question like "What has been done in mosquito genetics in the decade; please arrange it alphabetically by author." Naturally it could also be arranged by year. Table 5 describes five other actual searches. If demand warrants it, certain of these could be published, hopefully by computer typesetting.

Searches of interest to fewer people or to one worker reveal a little more computer potential. Basically each reference has the five fields given earlier—author, date, title, citation, and subject number. In the author and title field, there can be many words of interest. For example, a title like "The Genetics of Colonizing Populations of Vectors of Yellow Fever," contains words of interest to several types of mosquito biologists. Because of the cost of cross referencing, etc., the article can only appear once in *Mosquito News* and in one subject. Suppose that this one appeared under Genetics. A yellow fever epidemiologist still can find it easily using MODABUND. He can request all articles having any of a set of descriptors, e.g., subject number 31 or any of the following words or word-stems in the title; yellow fever OR epidem AND *Aedes aegypti,* etc. As another example, a researcher may wish only those articles dealing with the genetics of *Aedes.* He could ask for a search for any article with the Genetics subject number AND the word *"Aedes"* in the title PLUS any article in other subject numbers with the words "genetics" AND *"Aedes"* in the title. The arrangement of the printout can be by

TABLE 4.—Beginning and end of printout of "Hits" in Genetics, Subject 14, obtained from a computer search of the 8,551 references for 1960–1969. Number at far right is inhouse serial number.

AKSTEIN, F. RES. COUNCIL OF ISRAEL BUL. 11B(3):146–155. THE CHROMOSOMES OF AEDES AEGYPTI, AND OF SOME OTHER SPECIES OF MOSQUITOES.	1962	14	4835
ALDIGHIERI, J. SOC. DE PATH. EXOT. BUL. 54(4):712–714. CONTRIBUTION A LETUDE DE LA STRUCTURE DES CHROMOSOMES SALIVAIRES CHEZ AEDES AEGYPTI (L.) * *	1961	14	5666
WORLD HEALTH ORGANIZATION. WASH. D.C. WHO/VBC/67–47. ADD. 1. 1 PAGE. (ADDENDUM.) COORDINATION GROUP ON GENETIC CONTROL OF INSECTS OF PUBLIC HEALTH IMPORTANCE.	1967	14	455
WRIGHT, J. W. AND R. PAL. 794 PAGES. AMSTERDAM. GENETICS OF INSECT VECTORS OF DISEASE.	1967	14	96
TOTAL HITS: 182			

author, or if it is a review article, perhaps an ordering by year would be more useful.

Of interest also are our answers to the following questions: Of the 8,551 references from the decade, how many articles are in each subject (Table 1)? Table 1 also reveals which areas seem to have been studied much more intensively in the last decade than in the previous one (e.g., genetics). Other subjects appear stable while still others seem to have declined in activity. Naturally, the figures in Table 1 can not reflect such trends exactly since they are based only on a part of the total mosquito literature and because articles that may deal with two or more subjects are only assigned to one subject category.

Another question that can be answered is, what journals have the most articles in

TABLE 5.—Five recent searches of the 8,551 references from 1960–1969.

Requestor	Search request	Printout format	Number of references found
1. Prof. M. Fuchs	*Key words:* oviposition, ovary, corpus allatum, corpora allata, ovulation, follicle	First by author; then by subject number	554
2. Dr. P. Rodriguez (NIH Postdoctoral)	*Subject numbers:* 14, 21	First by subject number; then by author	999
3. Prof. K. Rai	*Key word-stems:* chromosom, karyotyp, cyto *Subject numbers:* 4, 14, 16, 34	First by subject number; then by author	1066
4. Mr. J. Petersen (NIH Trainee)	*Key word:* togoi	First by subject number; then by author	27
5. Mr. J. Petersen (NIH Trainee)	*Authors:* Haddow, Mahaffy, Gibbins, Gillett, McCrae, Gillies, Kitzmiller, Gouck *Key words:* olfactometer, simpsoni, host preference	First by author; then by subject number	196

each subject? Tables 6 and 7 provide an answer to this for two of Sollers-Riedel's subject areas. They indicate clearly to someone working in this area which journals he must look at regularly. But they also reveal that people interested in Behavior, Biology and Ecology (Table 6) will have to monitor more journals than mosquito geneticists will (Table 7). In the former, the top ten journals only account for about one third of the total references in that subject, whereas the top ten journals (and books!) in genetics contain about one half of the total mosquito genetics articles for 1960–1969. These differences may be due to differences in the total number of references, however.

Such analyses of the characteristics of the mosquito literature itself also should be of use to determine how well commercial services such as Biological Abstracts cover the mosquito literature and vice versa. By comparing our tapes with theirs, we could discover which journals each does not cover that perhaps should be covered.

PRESENT LIMITATIONS OF MODABUND. No claim is made that MODABUND is a panacea for literature retrieval in mosquito biology. Limitations are obvious at several levels. We rely completely on Mrs. Sollers Riedel's bibliographies. No matter how careful a person might be, in a field like mosquito biology not all articles will be caught. Currently with more than 2,500 references per year in the Bibliography, the question is, what percent of the total mosquito literature does this represent? We do not know.

Another problem concerns the articles that *are* in MODABUND. As discussed earlier, we have captured limited information about each one. Titles do not completely describe an article. For example, a search for all articles with *Aedes* in the title produced about 1,100 references from the 8,551. Yet probably many more involve *Aedes* mosquitoes. Also we only have one subject number per reference. Hence, the results of any current search in MODABUND must be considered a first step towards retrieving the literature of interest. But it is a huge first step that most of the time seems capable of finding the bulk of the relevant literature. Looked at from a different angle, without MODABUND a researcher faces about 25,000 references in *Mosquito News* alone. With MODABUND and the proper search request, he can let the

TABLE 6.—The ten most cited journals in Subject 13 (Behavior, Biology and Ecology) for 1960–1969.

Number of citations	Journal name
155	Mosquito News
56	Medical Parazitologie i Parazitologie Bolezni
43	WHO/VBC Circulars
42	Entomological Society of America, Annals
33	W.H.O. Bulletin
33	Bulletin of Entomological Research
32	Japanese Journal of Sanitary Zoology
30	East African Virus Research Institute Reports
30	New Jersey Mosquito Extermin. Assoc., Proceedings
28	Nature (London)
482	of the 1131 references of Subject 13.

TABLE 7.—The ten most cited journals in Subject 14 (Genetics) for 1960–1969.

Number of citations	Journal name
21	WHO Bulletin
14	Mosquito News
13	WHO/VBC Circulars
12	Articles in Wright and Pal (eds.): "Genetics of Insect Vectors of Disease"
9	Entomological Society of America, Annals
7	American Zoologist
7	Annals of Tropical Medicine and Parasitology
7	Nature (London)
5	International Congress of Entomology (London), Proceedings
4	(5 journals each had four hits each)
99	of the 182 references of Subject 14.

computer sort through this literature haystack to find the relatively small number of "needles" that are of interest to him.

PRESENT AND FUTURE USERS OF MODABUND. The primary goal of MODABUND has always been to serve the trainees and others of our NIH Training Grant. Past use has been restricted to members and visitors of Notre Dame's Vector Biology Laboratory. At present, use of MODABUND by mosquito biologists outside of Notre Dame is being encouraged on a cost basis, since we have no funds for manpower and computing time to do this. My position on this should be made clear. I am an Associate Professor at Notre Dame, a biologist whose research interests involve studies of population dynamics and of the evolution of character complexes of *Aedes aegypti,* e.g., Crovello (1970), Crovello and Hacker (1972). I have further interests in the theory of evolution and in the use of computers to enhance the value of museum and herbarium collections, e.g., Sokal and Crovello (1970), Crovello (1972). I organized MODABUND as a service to the mosquito workers at Notre Dame. I believe it is achieving this. I derive no remuneration from this project.

Although other bibliographic services exist, there seems to be a need for MODABUND by mosquito biologists. For example, recently Doctor Manabu Sasa of Tokyo visited Notre Dame. He mentioned that he was working on a review article on filariasis. The next day we gave him a printout, alphabetized by author, of over 800 references on the subject from 1960–1969 alone. Total cost of the search would have been about $15.00. We could have provided a second printout arranged by year and by author within each year for the same price. The latter might be especially useful when preparing review articles.

We at Notre Dame feel that MODABUND has value for mosquito workers. We shall try to process your search requests on a cost basis. For additional information on search capabilities and on current charges for standard searches (e.g., a printout of all filariasis references from 1951–1969, arranged by author), please contact us at Notre Dame.

ACKNOWLEDGMENTS. Beyond doubt MODABUND is a reality only because of the essential literature accumulations of Mrs. Sollers-Riedel. Given her contribution and my own time, the only other persons needed have been Mrs. Jean Winsch, who has been with MODABUND from its inception. She ably types the references, processes paper tapes, proofreads for errors, punches corrections and fills search requests for our trainees and others. She is supported by NIH Training Grant AI00378 from National Institute of Allergy and Infectious Diseases (NIAID) to Doctor Karamjit S. Rai. All computer time except $100.00 from NIH has been graciously provided by the University of Notre Dame. The computer programs were written in PLI by a senior undergraduate in Mathematics, Mr. Louis Henefeld. They are byproducts of NSF, Office of Science Information Services, Grant GN-878 to myself, "The Computerization of the Edward Lee Greene Herbarium at Notre Dame." Costs of publication were supported by NIH Grant AI 02753 to Dr. G. B. Craig, Jr.

Literature Cited

Arnett, R. H. 1970. Entomological Information Storage and Retrieval. Biorand Foundation. Baltimore. 210 pages.

Crovello, T. J. 1970. Analysis of character variation in ecology and systematics. Ann. Rev. Ecol. and Syst. 1:55–98.

Crovello, T. J. 1972. Computerization of the Edward Lee Greene Herbarium (NDG) at Notre Dame. Brittonia 24:131–141.

Crovello, T. J. and Hacker, C. S. 1972. Evolution of life table characteristics among feral and urban strains of *Aedes aegypti* (L.). Evolution 26:185–196.

Crovello, T. J. and MacDonald, R. D. 1970. Index of EDP-IR projects in systematics. Taxon 19: 63–79.

Sokal, R. R. and Crovello, T. J. 1970. The biological species concept: a critical evaluation. Amer. Nat. 104:127–153.