A Coordinate Indexing System for Ecosystematic Data[1, 2]

W. P. McCafferty
Department of Entomology
Purdue University, West Lafayette, Indiana 47907

Abstract

A procedure for information storage and retrieval of ecosystematic data associated with the Ephemeroptera of Indiana collections at Purdue University consists of a modified coordinate indexing system. The system presently incorporates data based on 114 species and over 11,000 specimens, representative of many water systems throughout 90 Indiana counties. Over 1,800 accessions and 880 data uniterms have been utilized. An average of 54 individual accessions are referenced per uniterm for over 47,000 reference inputs. The system is adaptable to electronic data processing.

Introduction

The need for developing modern information storage and retrieval systems for data associated with systematics collections has become imperative (7). National data standards are currently being developed through the auspices of The Association of Systematics Collections to ensure a utilitarian network for electronic data processing. The urgency for the development and incorporation of electronic data processing-information retrieval systems (EDP-IR) is based primarily on the overwhelming quantity of data, and a growing awareness of the actual and potential applications of collections data for solving environmental problems. In order to establish baselines and decipher changing environmental patterns and ecological associations of species, methods of indexing and correlating the multitude of factors involved must be available as part of the collections storage system itself. Moreover, Squires (10) has adequately argued for the necessity of data processing merely from the standpoint of collections management for systematic research. For an example of the many types of current EDP-IR projects dealing with systematic biology, see Crovello and MacDonald (5).

In 1975, a project was initiated at Purdue University by the Laboratory of Insect Diversity in conjunction with the Water Resources Research Center to implement a system for processing data relevant to the natural geographic and ecological distribution of freshwater macroinvertebrates of Indiana rivers and lakes. The system was designed to provide the needed capability for correlating available subsets of environmental data from which various degrees of biological predictiveness of environmental quality could be derived and to insure that growth and maintenance of the system would be relatively simple.

Sampling of Ephemeroptera, or mayfly species, from throughout Indiana has been underway since 1971 in order to gain an understanding of the ecosystematics of this relatively large group of ecologically im-

[1] Published with the approval of the Director of the Purdue Agricultural Experiment Station as Journal Series No. 6082.

[2] Research was supported by the Office of Water Resources Research, Purdue University, No. A-071-IND.

portant, benthic, primary consumers. The Indiana Ephemeroptera collection housed at Purdue currently includes 114 species and over 11,000 specimens. These collections are representative of 90 counties in the state, approximately 220 rivers and/or streams, over 60 lakes and/or ponds, all seasons of the year, and various degrees of water quality. The Ephemeroptera collection and associated data were therefore available as a prototype for the development of an environmental information and retrieval system. The limited Indiana Ephemeroptera housed at the Illinois Natural History Survey have also been included in the system.

The Coordinate Index

A coordinate indexing system (6, 11, and 12) was designed to manipulate taxonomic and ecological data, and also to incorporate the systematics collections with their inherent heirarchial makeup as part of the overall system. Broadhurst (3) described coordinate indexing as a convenient and useful method of referencing bibliographic data. Additional modifications and applications have provided a unique system with many practical advantages when used to encompass structured systematics collections data. The design and operation of this coordinate indexing system is detailed below in terms of three basic components.

1. Biological materials and associated data: Each vial of stored specimens (Fig. 1) represents a unique combination of a particular species taken at a particular time and place, and associated with particular ecological conditions. This automatically precludes the usage of any materials that have not been thoroughly sorted, identified, and verified by an authority. Data labels remain with the specimens at all

FIGURE 1. *Items consisting of individual vials of specimens and data.*

times. Each vial therefore represents a unique combination of data. These vials are curated (8) and catalogued taxonomically via family, genus, and species. Each vial has been assigned a consecutive accession number which ties the biological materials directly to the second major component of the system, the accession index.

2. The accession index: This is a separate serial card or master file in which all of the data associated with any one unique "item" (vial in this case) whether it be taxonomic or ecological is tabulated on a single 5″ x 8″ card (Fig. 2). This index is ordered, however, irrespective of any classification or relationship of the data represented, but only according to the sequence of accession numbers assigned to the items as they are accessioned into the system.

Pseudocloeon anoka Daggy 1708
Baetidae PU-LID

L, ML, S, F, M
Field Reared

INDIANA: Elkhart Co., Elkhart R.
1. 2 mi SE Millersburg at County
Bridge, VI-25-1973,
W. P. McCafferty, A. V. Provonsha

M. Minno

Lotic
Riffle Area
Cobble Subst.
52 F - Water
12-18 cm Depth

FIGURE 2. *A master card from the accession index.*

In Figure 2 an example is given of the way in which the various data associated with an item are presented. In the upper left hand corner is the scientific name, including the familial classification, of the item followed by various abbreviations indicating the sexes and stages involved and whether or not the adult stage if present was actually associated with the immature stages via rearing experiments. Abbreviations are as follows: S = subimago, ML = mature larvae, L = all other larvae, M = male adult, and F = female adult. The accession number is located in the upper right hand corner of each card. Below the accession number, the institutional and divisional acronyms indicate place of deposition of the material. In the center of the card the locality label data is reproduced just as it appears on the locality labels with the specimens. The identifier's name is located in the lower left. Recorded habitat data is listed in the lower right.

3. The uniterm retrieval index: All of the data which can be derived from each accessioned item are "traced." That is, appropriate terms or phrases which delimit parameters are singled out; for example, a species name, a stream, a week of the year, a substrate type, etc. These key words or phrases are called "uniterms" in the terminology of coordinate indexing (3).

Uniterms are used as individual headings to cards or sets of cards which make up the retrieval index. Under each appropriate heading the accession numbers of each item which applies are "posted" resulting in a complete system of cross indexed data. Perusal of any uniterm card provides the accession numbers of items applicable to the parameter described by the uniterm itself. For two or more uniterms, uniterm cards can be compared for coincidences of accession numbers, a procedure which in principle is similar to an "optical coincidence system" (1). Proceeding with such cross matches will thus yield reference information with various desirable degrees of complexity. For instance, one might wish to show what species occur in Indiana rivers, or in the other extreme, one might wish to know what species of *Stenonema* are known from riffle areas of Little Pine Creek as mature larvae from the month of June, etc.

If the uniterm retrieval index is to be operated without machine search (as is currently the case at Purdue) the following card design in the uniterm retrieval index is convenient. The 5" x 8" cards (Fig. 3) consist of a uniterm heading and ten columns numbered 0-9. Accession numbers are listed in the columns according to the last digit in accession number. This ensures more uniform usage across the entire card and facilitates comparison for coincidences of numbers on two or more cards. One requirement, however, is that the numbers must always be listed sequentially from smallest to largest in each column. To accomplish this all of the posting must be done at the time of accession so that interline insertions are avoided. With the present system the possible number of uniterms per item is known, and there should seldom be any need to retrace a sample for additional uniterms.

Nomenclature is subject to change. It is suggested that if such change occurs the item be reaccessioned with a new accession number. The old accession numbers can be voided or retained. If retained, the

Wabash River

0	1	2	3	4	5	6	7	8	9
170	111	412	523	164	315	16	497	168	419
260	391	672	773	524	525	36	657	258	509
420	421	772	903	544	535	526	877	658	569
560	541	882	1073	784	545	536	897	758	629
780	621	902	1173	904	885	546	1067	838	659
880	1101	1072		954	905	786	1077	1068	1069
1070	1131	1132		1074	1155	876	1147	1078	1089
1100	1141	1142		1084	1355	1066	1157	1088	1139
1130	1171	1172		1114	1455	1076	1457	1148	1149
1140	1301	1242				1106		1158	1159
1160						1456		1168	1169
1170								1348	1299

FIGURE 3. *A uniterm card with posted accession numbers.*

prior master card should be retained in the accession index with a notation that a change has occurred. Perhaps the best method of handling this would be to mark the old number appearing on uniterm cards in such a way (e.g., with a circle or an asterisk) so that it is obvious that it no longer applies but still is available if the worker wishes to refer to previous data. For an item originally indexed only under the generic name, the specific epithet can be added to the master card with a new accession number appended to replace the old. For recombined names, the generic name and accession number can be changed. Accession number changes must always be duplicated in the collection itself.

There are currently over 1,800 accessions and 880 data uniterms, the cards for which occupy two double drawer file cabinets. An average of 54 individual accessions are referenced per uniterm for over 47,000 reference inputs.

The list of categories and subcategories in Table 1 should aid in visualizing the scope of possible uniterms and data. Some subcategories are actually uniterms within themselves and are noted with an asterisk. Uniterms for the category "Week of the Year" are merely listed as 1 through 52. Under the large category of "Ecological," data may or may not always be associated with specimens. Several categories are listed under "Precision of Data." Whether or not specimens have been field or laboratory reared or stages merely associated is important in the task of describing the larvae of a great number of species which remain unknown in the literature, but are important for the study of freshwater ecology (9). "State of Specimen Preservation" involves whether or not the specimens are fluid preserved in vials, placed on pins, or prepared on microscope slides. The vast majority of Ephemeroptera in North America are fluid preserved, and so far these are the only types which have been accessioned for Indiana.

TABLE 1. *List of uniterm categories and subcategories.*

GEOGRAPHIC	ECOLOGICAL
State or Country	Water System
County or Similar Unit	Water Classification
Exact Locality	Current Range
	Substrate Type
TEMPORAL	Temperatures
Year	*Nocturnal Light Attraction
Week or Year	
	PRECISION OF DATA
TAXONOMIC	Collectors
Family	Identifiers
Genus	*Field Reared
Species	*Laboratory Reared
	*Stages Associated
BIOLOGICAL	*Identified to Species
Sex	*Not Identified to Species
Growth Stage	Place of Deposition
	State of Specimen Preservation

Adaptability

Because the system is "open-ended," new categories and uniterms can be added. For example, if a collection from a river survey were to be added with very detailed limnological data (e.g., physical and chemical parameters measured at river sites) these data could be accommodated by adding new uniterm cards.

The efficiency of the system is directly related to the comprehensiveness of the data incorporated. It is also limited by the complexities of manipulation which increase as data input increases. At present the manual search system is most applicable to collection programs of limited scope and which have limited access to EDP hardware. In time, it may become practical to employ computer methods to make searches. The question of when an EDP system is justified is discussed by Crovello (4) in terms of costs and benefits. Arnett (1) suggested that if file cards are used for more than three years, then EDP is more effective.

Although manual search appears adequate for now, the system is adaptable to a long range goal of computerization which will be necessary for increasing retrieval efficiency as the data base expands (e.g., to include the mayflies of North America). Data standards as presented by the ASC council on standards for systematics collections (2) have been met. Terminology associated with the coordinate indexing system are equivalent to the proposed "standard" EDP terms. A uniterm category (and subcategory) is equivalent to a *data element.* A uniterm is equivalent to a *data item.* A vial of specimens (in this case) containing a species from a unique space and time and all data associated with it are equivalent to an *item.*

Basic data has been defined as "data with a high probability of use and desirability of use in all of the separate discipline-wide centralized EDP systems for information retrieval" (2). The itemized standards for *basic data* have been met or are inherent in the present system. PU (Purdue University) is suggested as an institutional identifier and LID (Laboratory of Insect Diversity) an institutional subdivision; the accession number is suggested as the individual number; and the data types of taxonomic identifier, locality, and time, are self evidently included. The state of the specimen presently involves only the mode of preservation, since specimens should be assumed to be whole and in "usable" condition unless otherwise noted (i.e., no uniterm expressing partial specimen representation has been necessary so far).

Applications

Although it is not the intention of this paper to deal with data analysis per se, three examples of the types of questions that may generate useful answers follow:

1. Which species of Ephemeroptera have been taken during 1971-1975 in a two-mile stretch of Wildcat Creek, Indiana, the immediate effluent area of an impending impoundment? The pre- and post-impoundment analysis would be greatly aided by having baseline data already established and retrievable. In fact, the coordinate index system was

recently utilized by Stanley Environmental Consultants in developing an environmental impact statement for Wildcat Creek for the U. S. Army Corps of Engineers.

2. Which species of Ephemeroptera could be expected as mature larvae in early June in littoral zones of small lakes and ponds in LaGrange, Steuben, Noble, and DeKalb Counties? If such species are found to be conspicuously absent in certain of the lakes of this region and season, this empirical relationship may justify further environmental analysis of cause and effect.

3. In what streams in an area of known distribution is found species X, whose range of tolerance for dissolved oxygen or some other factor is known to be quite low? Some biological assessment or suggestion of the water quality of certain streams could be made in part from the data.

The above examples deal with ecological applications of the data. The possible systematic applications of collections data are many and these have been dealt with in the literature (e.g., 10).

Literature Cited

1. ARNETT, R. H., JR. 1970. Entomological information storage and retrieval. Bio-Rand Found., Inc., Baltimore, MD. xiii + 210 p.
2. ASSOCIATION OF SYSTEMATICS COLLECTIONS. 1975. Report of the ASC council on standards for systematics collections. *In:* ASC Newsletter. **3**:1-8.
3. BROADHURST, P. L. 1962. Coordinate indexing: a bibliographic aid. Amer. Psychol. **17**:137-142.
4. CROVELLO, T. J. 1967. Problems in the use of electronic data processing in biological collections. Taxon. **16**:481-494.
5. CROVELLO, T. J., and R. D. MACDONALD. 1970. Index of EDP-IR projects in systematics. Taxon. **19**:63-76.
6. DOCUMENTATION INC. 1955. The uniterm system of indexing: operating manual. Documentation Inc., Washington, D.C., 47 p.
7. IRWIN, H. S., W. W. PAYNE, D. M. BATES, and P. S. HUMPHREY. 1973. America's systematics collections: a national plan. Ass. Syst. Coll. xiii + 63 p.
8. MCCAFFERTY, W. P. 1974. An economical and efficient vial rack for the storage of small, fluid preserved specimens. Ann. Entomol. Soc. Amer. **67**:996-997.
9. PROVONSHA, A. V., and W. P. MCCAFFERTY. 1975. New techniques for associating the stages of aquatic insects. Gr. Lakes Entomol. **8**:105-109.
10. SQUIRES, D. F. 1971. Implications of data processing for museums. p. 235-253. *In:* Data processing in biology and geology. J. L. Cutbill, ed. Academic Press, New York.
11. TAUBE, M. 1953. Studies in coordinate indexing, Vol. 1. Documentation Inc., Washington, D.C., 110 p.
12. TAUBE, M. 1957. The mechanization of data retrieval: studies in coordinate indexing, Vol. 4. Documentation, Inc., Washington, D.C., vii + 127 p.