TAXON 28(1, 2/3): 197-203. APRIL 1979

A TAXIR DATA BANK OF SEED PLANT TYPES AT THE UNIVERSITY OF MICHIGAN HERBARIUM

George F. Estabrook[1]

Abstract

A computerized information management system for use with the curation of seed plant types at the University of Michigan Herbarium is now operative. Its construction, structure, and use are described and illustrated.

In his application for curatorial Grant NSF BMS-72-00071 Professor Rogers McVaugh included provision for the construction of a computerized information management system for the curation of the type specimens at the University of Michigan Herbarium. The construction of such a system is essentially complete for the seed plants (approximately 5500 types), and nearing completion for Mosses, Lichens, and Fungi. A system for Pteridophyte types has been begun. This report describes the Data Bank for seed plant types.

The antities about which information is stored are basionyms, and the type specimens to which they are attached. The basic information in the data bank is of three kinds: What *is* the name and its taxonomic placement; who *made* the name and where was the description published; and who *collected* the specimen and where was it collected.

The computer system used to make this Data Bank of seed plant types is the *TAXIR system running in the context of the MTS operating system at the University of Michigan Computing Center. Detailed and explicit instructions on how this system can be used has been published by Brill (1978). A description of the theoretical basis of its retrieval algorithm has been published by Estabrook & Brill (1969).

The descriptions for this data bank are presented below. The FREE field format option was chosen for entering data. In this convention, the specific state of each descriptor, that applies to the basionym being described is entered into a computer readable medium (such as cards).

1. GENUS
   As it appears in the basionym.
2. SPECIFIC EPITHET
   As it appears in the basionym.
3. INFRASPECIFIC EPITHET
   As it appears in the basionym, else blank.
4. FAMILY
   As recognized by Engler & Prantl, but slightly modified to reflect widely accepted modern concepts.
5. DIVISION
   No longer used since separate banks were made for each major plant group.
6. AUTHORITY
   Abbreviations as suggested by the "Author Index" of Kew, else last name followed by all initials each followed by a period. Multiple authors are connected with "&".

[1] University of Michigan Herbarium, Ann Arbor, MI 48109, U.S.A.

Instances of "in" and "ex" are also included as part of the authority.

7. KIND OF TYPE
   The commonly accepted terms: HOLOTYPE, ISOTYPE, SYNTYPE, etc. are used.
8. PLACE OF PUBLICATION
   Abbreviations for periodicals as suggested by Botanico-Periodicum-Huntianum 1968. Book titles in full, except B-P-H conventions are used for commonly abbreviated words.
9. VOLUME
   The number of the volume, else blank.
10. PAGE
    The number of the page, else blank.
11. FIGURE NUMBER
    The very alpha-numeric string of symbols used in the publication to designate the figure, plate, or etc.
12. YEAR OF PUBLICATION
    Four numerals as commonly designate a year.
13. PLACE OF PUBLICATION REMARKS
    Up to 99 alpha-numeric symbols.
14. HERBARIUM
    All specimens reside at MICH.
15. HERBARIUM NUMBER
    Blank, as MICH does not use one.
16. COUNTRY
    Country of collection, English spelling of modern name.
17. STATE
    State, province, or major political subdivision.
18. COUNTY
    County, municipality, or other minor political subdivision.
19. LOCALITY
    Any additional locality data, in less than a hundred characters.
20. FIRST COLLECTOR
    Last name, followed by initials, e.g., COWAN R.S.
21. SECOND COLLECTOR
    If only one collector then blank, else if only two collectors, then name of second collector, as in 20, else the characters "et al.".
22. OTHER COLLECTORS
    Blank. Two collectors were enough.
23. COLLECTION NUMBER
    The numerical part of the collection number. This is the entire collection number in most cases.
24. COLLECTION NUMBER PREFIX
    If a nonnumerical symbol appears near the beginning of the collection "number", then this is the string of symbols from the beginning of the "number" through the last such nonnumerical symbol, else blank.
25. COLLECTION NUMBER POSTFIX
    Same as 24 except encodes case in which a nonnumerical symbol appears near the *end* of the collection "number".
26. OTHER NUMBER
    Any other "number", if present.
27. DAY OF COLLECTION
    An integer 1 through 31 to designate day number in date of collection.
28. MONTH OF COLLECTION
    First three letters of the month name in date of collection.
29. YEAR OF COLLECTION
    Like 12.

30. COLLECTION REMARKS
    Like 13.
31. MISC. REMARKS
    Like 13.
32. RANK
    To designate the rank of the basionym use one of the following: SP, SSP, VAR, SUBVAR, FMA, SUBFMA
33. DATE RANGE
    If date of collection is given as a range then this is "yes" and 27, 28, and 29 are the earliest date given, else blank.
34. DIACRITICAL MARKS OMITTED
    If foreign spellings require diacritical marks then this is the names of the descriptors from whose states such marks have been omitted, else blank.
35. TYPE OF MINOR GEOGRAPHICAL SUBDIVISION
    The kind of subdivision, e.g, County, District, Comisaria, named in 18.
36. LOCATION OF HOLOTYPE
    Herbarium acronym.

These descriptor states can be of variable length, and are separated by commas. When data prepared in this way are entered into a *TAXIR system, the order in which the descriptor's states will be presented is specified so that the information in each such "free field" can be associated with the proper descriptor. The order in which data were usually presented to *TAXIR is 1, 2, 3, 32, 7, 4, 5, 6, 8, 9, 10, 11, 12, 14, 15, 36, 20 thru 29, 33, 30, 16, 17, 18, 35, 13, 19, 34. This order made data preparation easier by grouping related information together.

Among the states of a descriptor, each distinct string of symbols is considered a unique descriptor state (except that instances of more than one consecutive blank are reduced to only one blank). Thus, it is important to establish standard conventional ways to present information. For example, in the descriptor, Authority, McVaugh R and McVaugh R. would be considered distinct authorities. Table 1 also provides descriptions of conventions that have been used to present data.

When the basionyms have all been described to *TAXIR and the Data Bank is complete, *TAXIR can provide, for each descriptor, lists of all the descriptor states that have occurred in the data. Such lists are called the control vocabulary because these terms can be used to control for which basionyms information will be printed. The complete control vocabulary for the data bank of flowering plant types is enormous, but small portions of it are presented below, in order to illustrate the concept.

TAXIR OUTPUT: DEC 17, 1977 PAGE 1
V*
1. GENUS
TYPE: NAME NO. OF STATES: 1453 PRINT FIELD SIZE: 18
AMBIG: CODE LARGEST CODE NO.: 1618

CODE	NAME
896	ABIES
1509	ABOLBODA
1103	ABORTOPETALUM
1160	ABUTA
791	ABUTILON
647	ACACIA
667	ACACIELLA
1514	ACALYPHA
320	ACANTHOTHAMNUS
948	ACCA
1	ACER
2	ACERATES

TAXIR OUTPUT: DEC 17, 1977 PAGE 59

2. SPECIFIC EPITHET (continued)

CODE	NAME
941	MCVAUGHIANA
16	MCVAUGHII
1535	MEDIA
830	MEDULLOSA
1617	MEGACARPA
2816	MEGACARPUM
571	MEGACEPHALA
618	MEGACEPHALON
908	MEGALANTHA
1347	MEGALANTHUS

4. FAMILY

TYPE: NAME NO. OF STATES: 193 PRINT FIELD SIZE: 16

AMBIG: CODE LARGEST CODE NO.: 322

CODE	NAME
16	ACANTHACEAE
1	ACERACEAE
41	ACTINIDIACEAE
8	ALISMATACEAE
12	AMARANTHACEAE
7	AMARYLLIDACEAE
32	ANACARDIACEAE
14	ANONACEAE
11	APOCYNACEAE
39	AQUIFOLIACEAE

TAXIR OUTPUT: DEC 17, 1977 PAGE 102

6. AUTHORITY (continued)

CODE	NAME
228	BUSH
672	C. DC.
915	C. & R.
1053	CALDERON G.
1054	CAMBESSEDES
129	CAMP
876	CARLSON
763	CARTER IN BENSON & CARTER
632	CASE F.W.
699	CHASE
735	CHING
1112	CHING EX TARDIEU & CHRISTENSEN C.
270	CHODAT
502	CHODAT & HASSLER
658	CHODAT EX RUSBY

7. KIND OF TYPE

TYPE: ORDER NO. OF STATES: 7 PRINT FIELD SIZE: 12

CODE	NAME
1	HOLOTYPE
2	ISOTYPE
3	SYNTYPE
4	ISOSYNTYPE
5	LECTOTYPE
6	ISOLECTOTYPE
7	NEOTYPE
8	MEROTYPE

TAXIR OUTPUT: DEC 17, 1977 PAGE 120

11. FIGURE NUMBER (continued)

CODE	NAME
27	60

12. YEAR OF PUBLICATION

TYPE: FROM-TO NO. OF STATES: 248 PRINT FIELD SIZE: 4

FROM 1753 TO 2000 BY 1

13. PLACE OF PUBLICATION REMARKS

TYPE: NAME NO. OF STATES: 464 PRINT FIELD SIZE: 91

AMBIG: CODE LARGEST CODE NO.: 464

CODE	NAME
384	#MISPRINTED IN MADRONO AS 10462
220	#2
215	#2449 ALSO CITED. REMOVED TO E. MELANADENIA KR. & URB. 1895
366	#CITED AS PISTILLATE TYPE IN ORIG. DESCR.

23. COLLECTION NUMBER

TYPE: FROM-TO NO. OF STATES: 200000 PRINT FIELD SIZE: 6

FROM 1 to 200000 BY 1

24. COLLECTION NUMBER PREFIX

TYPE: NAME NO. OF STATES: 15 PRINT FIELD SIZE: 4

AMBIG: CODE LARGEST CODE NO.: 15

CODE	NAME
15	-S
3	A
13	A.
6	C-
1	E
7	E-
2	H
8	II

27. DAY OF COLLECTION

TYPE: FROM-TO NO. OF STATES: 31 PRINT FIELD SIZE: 2

FROM 1 TO 31 BY 1

28. MONTH OF COLLECTION

TYPE: ORDER NO. OF STATES: 12 PRINT FIELD SIZE: 3

CODE	NAME
1	JAN
2	FEB
3	MAR
4	APR

The *TAXIR system can be given instructions to print out portions of the information that is stored in the data bank of seed plants. There are two aspects in which this desired portion must be specified: for which basionyms is information desired?; and, for which descriptors is information desired?. Thus, a *TAXIR query has two parts. One part is a logical combination of properties of basionyms (structured as a Boolean expression with descriptor states as operands). This part specifies which basionyms are of interest. One simple form of query asks how many basionyms satisfy the specified condition. An example below illustrates this.

HOW MANY TYPES WITH FAMILY, MYRTACEAE AND AUTHORITY, MCVAUGH OR BERG
MORE
OR LUNDELL AND NOT COUNTRY, MEXICO*
NO. OF ITEMS IN QUERY RESPONSE: 256

NO. OF ITEMS IN DATA BANK: 5429
PERCENTAGE OF RESPONSE/TOTAL DATA BANK: 4.72%

Information about the basionyms that satisfy a specified condition can also be requested by listing the descriptors whose states are to be printed. This is illustrated in the example below.

READY
QUERY (AUTHORITY, YEAR OF PUBLICATION, PLACE OF PUBLICATION),
MORE
(GENUS, SPECIFIC EPITHET, RANK, INFRASPECIFIC EPITHET, LOCATION OF HOLOTYPE)
MORE
FOR TYPES WITH COUNTRY, MEXICO AND AUTHORITY, MCVAUGH AND FAMILY, MYRTACEAE*
NO. OF ITEMS IN QUERY RESPONSE: 5
NO. OF ITEMS IN DATA BANK: 5429
PERCENTAGE OF RESPONSE/TOTAL DATA BANK: 0.09%

MCVAUGH	1963 FIELDIANA BOT.			
CALYPTRANTHES	TENUIPES	SP	—	MICH
EUGENIA	ALNIFOLIA			
	PRINCIPIUM			
	TURNERI			
	XILITLENSIS			

For immediate reference, *TAXIR has also printed out a complete citation of each basionym. These are arranged alphabetically by Family; within Family alphabetically by Genus; within Genus alphabetically by Species; within Species alphabetically by Infraspecific epithet. A portion of this listing is given in the example below.

CAMPANULACEAE
CENTROPOGON MINIMUS MCVAUGH BRITTONIA 6: 467. 1949 ISOTYPE
EWAN J. 15699 23 MAY 1944 HOLOTYPE AT NA COLOMBIA ANTIOQUIA DEPARTMENT
PARAMO DE SONSON ABOVE SONSON; ELEV. 2743 M

CENTROPOGON MONAGENSIS MCVAUGH BRITTONIA 3: 472. 1940 ISOTYPE
STEYERMARK J. A. 61862 5 APR 1945 HOLOTYPE AT F VENEZUELA MONAGAS
CERRO NEGRO ABOVE LA SABANA DE LAS PIEDRAS NW OF CARIPE ALT. 1500 M

CENTROPOGON OAXACANUS VAR BERTERIOIDES WIMMER
E. PFLANZENREICH IV 276: 840. 1968 HOLOTYPE
HOLM RICHARD W. & ILTIS HUGH H. 120 25 JUN 1949 COSTA RICA CARTAGO
SMALL VALLEY NEAR SANTA CRUZ; SOUTH SLOPE OF VOLCAN TURRIALBA; ELEV. 1430 M

CENTROPOGON TALAMANCENSIS WILBUR BRITTONIA 21: 255. 1970 ISOTYPE
WILBUR R. L. & STONE D. E. 10053 27 JAN 1968 HOLOTYPE AT DUKE COSTA RICA CARTAGO
ALONG THE CARRETERA INTERAMERICANA AT ABOUT 97 KM; VIC. OF RESTAURANTE LA GEORGINA; ELEV. 3100 M

CENTROPOGON UNCIALIS MCVAUGH J. WASH. ACAD. SCI. 39: 159. 1949 HOLOTYPE
GARCIA-BARRIGA H. 10744 22 FEB 1942 COLOMBIA CUNDINAMARCA ENTRE PACHO Y RIO NEGRO; ELEV. 1000 A 1200 M
DATE RANGE 22-26 FEB 1942

Many ways to improve the design of a Type data bank such as this have already become apparent. Information concerning Volume, page, and figure number should all be placed in one descriptor. Detailed locality data probably should not be included, but left to reside on the label only. All the collectors' names should be put in one descriptor, just as they would appear in exsiccatae citations. Ideally, only collectors in whose number series occurs the collection number should be included. Remarks should be held to a minimum, and these in one descriptor only.

The extent and manner of use of this data bank remains to be seen. The opportunity to try and to test this modern curatorial tool, however, has been made possible by the support and help of Rogers McVaugh.

Literature Cited

Brill, R. C. 1978. *The Taxir Primer 3rd ed.* University of Michigan Computing Center, Ann Arbor, Mi. 179 pp.

Estabrook, G. F. and R. C. Brill 1969. The Theory of the Taxir Accessioner. *J. Math. Biosci-ence* 5: 327-340.