

Computers in Botanical Collections

Computers in Botanical Collections

Edited by

J. P. M. Brenan

*Royal Botanical Gardens
Kew, England*

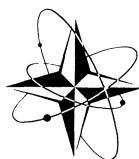
R. Ross

*British Museum (Natural History)
London, England*

and

J. T. Williams

*University of Birmingham
Birmingham, England*



Published in coordination with NATO Scientific
Affairs Division by

PLENUM PRESS • LONDON AND NEW YORK

Proceedings of an international conference, sponsored by NATO, on *The Use of Electronic Data Processing: Major Plant Taxonomic Collections*, held at Kew in October 1973

Library of Congress Catalog Card Number 75-9386

ISBN-13: 978-1-4684-2159-0 e-ISBN-13: 978-1-4684-2157-6

DOI: 10.1007/978-1-4684-2157-6

© 1975 Plenum Press, London,
Softcover reprint of the hardcover 1st edition 1975

A Division of Plenum Publishing Company, Ltd.
Davis House, 8 Scrubs Lane, Harlesden, NW10 6SE, England

U.S. edition published by Plenum Press, New York
A Division of Plenum Publishing Corporation
227 West 17th Street, New York, N.Y. 10011

All Rights Reserved

No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher

ELECTRONIC DATA PROCESSING OF HERBARIUM SPECIMENS
DATA FOR THE FLORA OF VERACRUZ PROGRAM*

A. Gómez-Pompa, J. A. Toledo and M. Soto
Department of Botany, Institute of Biology
National University of Mexico

Summary

The Flora of Veracruz program is a comprehensive study of the plant resources of this Mexican state. It involves a series of studies that go beyond the common objectives of the classical floras as it includes ecological and environmental studies and it is utilizing electronic data processing methods for the entire project.

The main objective is the production of a Flora of the same type as an ordinary good modern one. At the same time it could serve as a basis to create a data bank on the plants and the environments of Veracruz which could have many uses throughout the time of preparation of the Flora and afterwards.

A brief discussion of the kinds of data that are being introduced in the data bank will be explained here. We will present a more ample discussion of the data from herbarium specimens.

* Flora of Veracruz, contribution number 15. A joint project of the Institute of Biology of the National University of Mexico and the Department of Botany of the Field Museum of Natural History, Chicago, U. S. A. to prepare an ecological and floristic study of the state of Veracruz, Mexico.

Information about this project was published in Anal. Inst. Biol. Univ. Nat. Méx. Ser. Bot. 41: 1-2. Partially supported by N.S.F. grant GB-20267X.

Even though most of the effort has been concentrated on new collections, we have made a survey of the collections already present in the herbaria of several institutions. As an example of this effort we will present the pilot project that has been accomplished to compile the data from Veracruz specimens from the herbarium of the Royal Botanic Gardens at Kew. We will show examples of the kind of processing that we are doing with them.

A comparison of the collections of two families (Piperaceae & Araliaceae) in selected European herbaria will be presented as another example of the kinds and numbers of Veracruz collections housed in these herbaria.

A discussion on the future of herbarium data processing in future botanical research will be presented in the context of our floristic research programme.

Introduction

Herbarium specimens are one of the most important sources of information for any floristic project. They provide information about historical collections of great value for nomenclatural purposes, as well as information about areas where today we find only agricultural land or cities. Herbarium specimens also provide geographical data that are basic for protection and conservation of rare species. They supply unique information that cannot be found anywhere else about the distributions of all taxa. One of the greatest values of old collections from tropical areas is that many of them may be the only source of information about species that today may be extinct. This may be especially true for collections that came from areas now densely populated in tropical evergreen rain forest regions. There are other types of data available in herbaria that may be of interest, such as local names, plant uses and ecological information.

To capture these data the Flora of Veracruz Program has developed an information system for processing various types of data concerning the environments and plant resources of Veracruz.

One of the great problems for any newly initiated floristic study is to establish the quantity and quality of collections previously obtained in the area under investigation. This difficulty arises because collections are scattered in many different herbaria throughout the world, and herbaria are not organized in such a way that the information can be easily obtained. Another important problem is the change in scientific

names caused by revisionary and monographic work and it is also for this reason that a dynamic information system is needed.

For these and other reasons (Gomez-Pompa & Nevling, 1973) the Flora of Veracruz Program was planned from the beginning to use electronic data processing methods. One compelling reason for employing the computer in this Program was the need to have a Flora that could be useful even during its development. Normally a Flora is useful only after it has been published, because no information of a general nature can be obtained until it is completed. Another important reason for E.D.P. is that it should facilitate the rapid production of related Floras, of both more and less inclusive areas, once the initial Flora is completed. Classical Floras do not provide ways by which the information gathered for one Flora can be used for the next (Shetler 1971).

Among the sources of data for this Program, label data from herbarium specimens has had high priority. Three years ago several projects were started to capture data from specimens in selected large herbaria with the objective of evaluating the quality and quantity of Veracruz collections in those herbaria. At the same time we have captured the label data for the Flora of Veracruz data bank.

We have surveyed the following herbaria for this purpose: Arnold Arboretum, Gray, the Royal Botanic Gardens, Kew and the National Herbarium of the University of Mexico. In addition, we have made a sample survey of two families (Araliaceae & Piperaceae) in a larger number of herbaria: Leiden, Utrecht, Paris, Florence, Vienna Museum, British Museum, (Natural History), Missouri Botanical Garden, Field Museum, and Madrid.

We thought that by approaching the problem in this way we could have a good idea of what collections from Veracruz are available in the herbaria of the world. This paper presents some of the results obtained from the European collections, that will help us to explain this part of our E.D.P. system for the Flora of Veracruz.

Methods

In each herbarium in which a search was conducted the first step was to locate the specimens from Veracruz. The most serious problem encountered was with those specimens that do not have a clear indication that they come from Veracruz. In these cases the only clue was the locality name, or a combination of the collector's name and locality. All doubtful cases were recorded for future checking.

When a Veracruz specimen was found, a standard method designed to capture data was followed (Gomez-Pompa & Nevling, 1973). This consisted of filling out a pre-printed form, with a fixed format, to facilitate the preparation of punched cards for input into the data bank (Fig. 1). This process was followed with all specimens found; the only exceptions were specimens in type folders. In this case a special effort was made in some cases to verify from the literature if the specimen was in fact a type, and then a note that the type had been checked was included on the form.

The completed forms were sent to the computer laboratory of the Flora of Veracruz where the data were reviewed and prepared for entering into the data bank.

Two types of information from the labels are recognized: the first is the precise information concerned with the scientific name of the plant, the locality, and the collector's name, number and the date; the second type includes ecological information, plant uses, local names, etc. The first group is treated very rigidly for editing purposes and retrieval, whilst the second type is mainly stored in full text with very flexible format, thus allowing retrieval without standardisation of the information stored under these categories. This division of information permits us to capture all the information available on the labels without losing time in checking the reliability of the information, although this may be done later if desired.

The computerized system for the Flora of Veracruz includes an automatic method for checking and detecting errors in the information to be entered in the data bank, especially errors in the first type of information mentioned. To facilitate this process certain elements were codified, viz. scientific name and locality.

This codification of the scientific names is done using a basic master file of names in the system. This file makes it possible to work automatically and dynamically with all scientific names known for Veracruz plants. The use of this file has proved to be useful in preventing nomina nuda from entering the data bank. When possible, the locality was coded according to our system, which is based upon a latitude and longitude grid. In many cases, however, coding could not be carried out because of vague locality data or because many old place names are no longer in use and need special research for locating them.

After the foregoing steps in data preparation are completed, a set of punched cards is produced, thus converting the data into computer-readable form.

This first information input is automatically checked by the system and it only accepts that information which is error-free (according to our master files). Incomplete or erroneous information is refused with its problem diagnosis message (Fig. 2).

We want to emphasize that the data obtained from herbarium specimens is not changed or lost in this process and that we consider the information as working data to be used and evaluated by the taxonomists who will write the family treatments. Such information alone is not regarded as an end product.

The method followed in the Flora of Veracruz Program is a simple one that allows the input of information in the easiest way for future retrieval.

Once in the computer the data can be processed in many different ways. We can have a full listing of all the collections in one herbarium (Fig. 3), a list of the collections of one collector in one family (Fig. 4), a list of the localities of one collector (Fig. 5), statistical data of one collector (Fig. 6), or a list of type specimens (Fig. 7). It is important to note in relation to type specimens that a major problem results because not all specimens annotated as types are indeed types.

There are several other combinations that can be made with the information of the data bank. Figure 8 illustrates some statistical data relating to two families, and Figure 9 shows part of a chronological list arranged by collectors.

Discussion

The number of collections (2879) from Veracruz in the families surveyed up to now at Kew was lower than we optimistically estimated in the initial planning of the survey. Because of this difference a spot-checking survey was started in order to evaluate the thoroughness of the general survey. Even though the total number of specimens may be low, the value of the Kew holdings is great because of their high quality and historical importance. But the most significant aspect of this investigation in European herbaria is that we now have another basis for planning field work and herbarium studies with old collections.

Concerning the holdings of other herbaria, even though the number of taxa sampled may not be statistically representative, they can give us a clue as to total holdings by indicating the collectors represented and their

27363	NOMBRE DE ESPECIE INCORRECTO ZINGIBERACEAE	RENEALMIA MEXICANA KLOTZSCH EX PETERSEN
27363	COLECTOR AUSENTE O INCORRECTO	
27363	NUMERO YA EXISTENTE	
27364	NUMERO YA EXISTENTE	
27185	MAPA AUSENTE O INCORRECTO	
27185	FECHA AUSENTE O INCORRECTA	
27185	NUMERO YA EXISTENTE	
27187	TARJETA 1 O 3 AUSENTE	
27187	COLECTOR AUSENTE O INCORRECTO	
27187	HERBARIO AUSENTE O INCORRECTO	
27187	NUMERO YA EXISTENTE	
27367	NUMERO YA EXISTENTE	
27368	NUMERO YA EXISTENTE	
27369	NUMERO YA EXISTENTE	
27371	NUMERO YA EXISTENTE	
23662	NOMBRE DE ESPECIE INCORRECTO BIGNONIACEAE	TECUNA STANS (L.) HBK.
23662	NUMERO YA EXISTENTE	
27109	NOMBRE DE ESPECIE INCORRECTO ARACEAE	ANTHURIUM SCANDENS VAR. VIOLACEUM SCHOTT

Figure 2. A set of diagnosis errors sent by the computer

025864	GUTTIFERAE CLUSIA MEXICANA VESQUE. 28.5/28.5 VALLE DE CORDOBA. M. BOURGEAU 1957 17/02/ HERBARIO K
025865	GUTTIFERAE CLUSIA ORIZABAE HEMSL. 28.5/25.0 IZHUATLANCILLO ORIZABA. M. BOURGEAU 2967 14/08/1866 HERBARIO K *TIPO*
025866	GUTTIFERAE CLUSIA ORIZABAE HEMSL. 28.5/28.5 VALLE DE CORDOBA. H. BOURGEAU 1954 16/03/ HERBARIO K
025867	GUTTIFERAE CLUSIA PARVICAPSULA VESQUE. 28.5/29.5 PUTRERO PRES CORDOBA. M. BOURGEAU S.N. 14/01/ HERBARIO K
025868	GUTTIFERAE HYPERICUM PHILONOTIS CHAM. & SCHLECHT. 29.5/24.0 LOMA GRANDE MT. ORIZABA. E. K. BALLS 5380 27/08/1938 HERBARIO K
025869	GUTTIFERAE HYPERICUM OAXACUM R. KELLER. 40.0/25.5 BELOW LAS VIGAS PEROTE. E. K. BALLS 4790 / / HERBARIO K
025870	GUTTIFERAE HYPERICUM OAXACUM R. KELLER. 29.5/24.0 LOMA GRANDE ORIZABA. E. K. BALLS 5364 27/08/1938 HERBARIO K
025871	GUTTIFERAE HYPERICUM CONFUSUM ROSE 37.5/25.0 ACTOPAN, COFRE DE PEROTE. E. K. BALLS 5553 12/09/1938 HERBARIO K

Figure 3. Portion of the list of collections of Veracruz plants at Kew, obtained from the data bank of the Flora of Veracruz

BROTEN.		
BROMELIACEAE	1052	027085
COSSON E.		
ARACEAE	409	027186
GRAMINEAE	48	025701
COULTER DR.		
ACANTHACEAE	1209	023673
ASCLEPIADACEAE	979	023631
	980	023618
BORAGINACEAE	1062	023696
CARYOPHYLLACEAE	-	023741
	694	023719
COMPOSITAE		
	370	024069
	392	023764
	400	024053
	436	023899
	452	024166
CONVOLVULACEAE		
	1012	024510
GRAMINEAE		
	1643	025614
LILIACEAE		
	1590	027401
MELASTOMATACEAE		

Figure 4. Fragment of a list of collectors of Veracruz collections at Kew extracted from the data bank of the Flora of Veracruz which includes: names of collectors, family, collectors numbers and accession numbers to the data bank

2617 REGION DE ORIZABA, SAN CRISTOBAL.
 32.5/31.0
 10687 ZACUAPAN
 33.0/40.0
 618 VERACRUZ.
 33.5/29.0
 182 MIRADOR.
 4281 CORDILLERA VERACRUZ
 43.5/33.0
 2968 SANTA TERESA, REGION DE ORIZABA.
 44.5/29.5
 2452 MISANTLA
 BROTEN.
 28.0/26.0
 1052 ORIZABA.
 COSSON E.
 28.0/26.0
 48 ORIZABA.
 409 ORIZABA.
 COULTER DR.
 1209 SIN ONLY VER.
 33.0/40.0
 - VERACRUZ
 38.0/28.5
 - JALAPA.
 7 JALAPA
 125 JALAPA
 129 JALAPA
 133 JALAPA.
 370 JALAPA.
 392 JALAPA.
 400 JALAPA.
 436 JALAPA.
 452 JALAPA.
 694 JALAPA.
 773 XALAPA
 979 XALAPA.
 980 JALAPA.
 1012 JALAPA.
 1062 JALAPA.
 1193 JALAPA
 1238 JALAPA
 1299 XALAPA
 1397 JALAPA
 1590 JALAPA.
 1643 JALAPA.
 DE LEDN CARLOS
 32.5/28.0
 - HUATUSCO, VER.
 FINCH HUGO
 28.5/28.5

Figure 5. Partial list of collectors from specimens at Kew that includes: collection numbers, localities, locality codes and names of the collectors

	K	P	L	FI	BM	WV	Total
Number of Specimens	572	71	20	12	4	1	680
Number of Collections	542	54	20	12	4	1	633
Number of Duplicates	53	28	17	12	4	-	161
Number of Unicates	489	26	3	-	-	1	
Number of Families	64	2	2	2	2	1	65
Number of Genera	211	4	4	2	3	1	211
Number of Species	408	38	13	7	4	1	408

Fig. 6a

Figure 6, a & b. Samples of statistical data of part of the collections of M. Bourgeau.

Collection number	K	P	L	BM	FI	Total
1491	1	1	1		1	4
1492		2	1		1	3
1493	1	1			1	4
1512	2					2
1559	2					2
1569	2					2
1606		3	1			4
1677		1	1			2
1759	2					2
1804		6				6
1805	1	1				2
1806		2				2
1847	2					2
1853	2					2
1866		1		1		2
1958	1	1	1			3
1959	1	2	1		1	5
1988	1	1	1			3
1989	1	1				2
2054	1	2	1		1	5
2168	1	1	1		1	4

Fig. 6b

024576	ACANTHACEAE JACOBINIA INCANA HEMSL. 33.0/40.0 VERACRUZ. J. LINDEN 1079 / /1838 HERBARIUM K *TIPO*
025574	GRAMINEAE COELORACHIS RAMOSA (FOURN.) NASH 27.5/25.0 RIO BLANCO PDCES ORIZABA. M. BOURGEOU 2647 24/07/ HERBARIUM K *TIPO*
025603	GRAMINEAE PASPALUM MINUS FOURN. 28.5/28.5 VALLE DE CORDOBA. M. BOURGEOU 2298 24/04/ HERBARIUM K *TIPO*
025613	GRAMINEAE PASPALUM SQUAMULATUM FOURN. 27.5/25.0 RIO BLANCO PRES ORIZABA. M. BOURGEOU 2640 24/07/ HERBARIUM K *TIPO*
025660	GRAMINEAE PANICUM VISCIDELLUM SCRIBN. 38.0/28.5 GRAVELLY BANKS NEAR JALAPA. C. G. PRINGLE 8089 29/03/1899 HERBARIUM K *TIPO*
025717	GRAMINEAE MUEHLENBERGIA SETARIOIDES FOURN. 28.0/25.5 ORIZABA. BORREGO. M. BOURGEOU 3362 14/11/ HERBARIUM K *TIPO*
025738	GRAMINEAE EPICAMPES BOURGAEI FOURN. 28.0/26.5 ORIZABA. ESCAMELLA. M. BOURGEOU 2973 26/08/ HERBARIUM K *TIPO*
025787	GRAMINEAE ELEUSINE INDICA (L.) GAERTN. / VERACRUZ. DR. GOVIN S.N. / /1867 HERBARIUM K *TIPO*
025819	GRAMINEAE ARUNDINARIA ACUMINATA MUNRO / VERACRUZ. LIEBMAN. 73 / / HERBARIUM K *TIPO*

Figure 7. Portion of a list of the types from Veracruz found at Kew and entered in the data bank.

Familia Araliaceae

	K	L	P	W	WU	BM	FI	Total
Number of Specimens	20	3	14	7	1	11	4	60
Number of Collectors	8	2	6	5	1	5	2	11
Number of Collections	20	3	14	7	1	11	4	60
Number of Genera	3	2	2	3	1	2	1	4
Number of Species	5	3	4	4	1	3	1	7

Familia Piperaceae

	K	L	P	U	WU	BM	FI	Total
Number of Specimens	84	40	229	36	8	59	23	479
Number of Collectors	16	7	23	4	2	15	7	34
Numbers of Collections	84	40	205	35	8	59	23	554
Number of Genera	2	3	3	3	2	2	2	3
Number of Species	36	23	57	25	6	26	9	118

Figure 8. Statistical data of the collections of Piperaceae and Araliaceae from the survey made in several European herbaria.

GOULD FRANK W.			
17	09	1965	11710 028132
GOVIN.			
		1865	2 027891
		1866	7 027878
		1866	8 027871
		1866	9 027870
01		1866	6 028112
28	01	1866	11 028114
11	08	1866	3 027919
		1867	- 028060
		1867	- 028076
		1867	- 027729
		1867	- 028131
		1867	- 027863
		1867	- 026793
		1867	- 025787
		1867	1 027892
		1867	10 027931
		1867	12 027954
		1868	4 028125
		1868	5 028126
		1869	9 028127
GRAY A.			
05		1885	- 024077
GREENMAN J. M.			
24	06	1906	116 027379
HAHN M.			
			- 023629
			- 026570
			- 023599
30	06		- 026607
	08		- 026745
	09		- 023935
		1865	- 026634
		1865	- 026639
13	04	1865	46 028120
11	08	1865	- 027872
12	08	1865	- 027966
		1866	- 028118
		1866	- 027774
		1866	- 027932
		1866	- 024532
16	03	1866	2066 028080
16	03	1866	2066 027768
18	04	1866	46 026805
18	04	1866	59 028122
25	05	1866	- 027216
25	05	1866	189 027977

Figure 9. Portion of a list of collections arranged by collectors and also chronologically from the holdings at Kew and stored in the data bank.

collection numbers. It was evident that the herbarium at the Museum of Natural History at Paris is of high priority for our research.

One problem in maintaining an up-to-date file is that any change in the determinations of Veracruz specimens has to be communicated to us. We trust that most of the corrections will be made through specialists working for our programme. If this is carried out the problem of updating does not exist. Another possibility is that when monographers change the names of specimens, we wait until the paper (monograph or revision) is published to capture the corrections. When a curator notes a mis-determination and routinely changes the name, it will be almost impossible for us to enter this change in our files. For this reason we would greatly appreciate having sent to us information regarding any changes in the names of Veracruz specimens. Although we recognize the difficulties involved in this request, we hope that the problem will eventually be solved by common agreement among curators. However, our experience is that the name of a plant from Veracruz has little likelihood of being changed except through our own work or by monographers, so the risk of losing information is minimal.

We have demonstrated to our satisfaction that the general system designed for capturing label data for Veracruz specimens applies equally well to old herbarium specimens and new material, and the system is operational. It is important to mention that the most significant limiting factor is the personnel doing the data capture. They have to have a certain degree of taxonomic knowledge and a great amount of patience to do routine work with responsibility. It is our experience that some training and revision has to be incorporated in any survey in order to be sure of the correctness of the information gathered.

Acknowledgments

Data for this paper were obtained from the data bank of the Flora of Veracruz, those from Kew were obtained by Mrs. Laura Vit and those from other European herbaria by the senior author with the assistance of Mrs. Norma Gómez-Pompa. We wish to express our gratitude to these persons and the curators of the herbaria surveyed for their help. Also we want to thank Dr. John Beaman for his comments and corrections to the original manuscript.

REFERENCES

- Gómez-Pompa, A. & Nevling L. I. Jr. (1973). The use of electronic data processing methods in the Flora of Veracruz Program. Contr. Gray Herb. 203: 49-64.
- Shetler, S.G. (1971) Flora North America as an information system. BioScience 21:524-532.
- Vit, L. The collections of Veracruz at the Royal Botanic Gardens at Kew. Thesis, Facultad de Ciencias, UNAM, México (in preparation).