TAIS, A DATABASE SYSTEM FOR THE TAI HERBARIUM, NATIONAL TAIWAN UNIVERSITY

CHANG-FU HSIEH[1], TSENG-CHIENG HUANG[1]
and SHING-FAN HUANG[1]

(Manuscript received 22 June, 1991; revised version accepted 22 July, 1991)

Abstract: The newly implemented database system, called TAIS, for the TAI Herbarium is outlined. TAIS is a versatile and changing system used for recording and manipulating information about herbarium specimens. It was designed for run on microcomputers, and will eventually allow encoding of the 220,000 specimens housed at the TAI. The system consists of a main SPECIMEN base and two accessory bases containing species names and geographic maps. The incorporation of map images into application made it more powerful. Three different ways of localizing collection sites is described. Grid references and latitude/longitude can be generated automatically. Outputs of the system include specimen label, specimen citation, distribution map, and user defined information.

INTRODUCTION

A computerized data system, called TAIS, has been developed by the TAI Herbarium and is being used currently to facilitate data storage and retrieval. The system was a refined version of the second part of the PRIS system (Hsieh and Su 1990; Hsieh 1991), and was particularly useful for routine herbarium tasks.

Since its founding in 1928, over 220,000 specimens have been deposited at the TAI Herbarium. The specimens comprised an extensive representation of Taiwan flora, and have served as an indispensable aid for taxonomic, floristic, ecological, or many other related researches in this region. The richness of the flora represented in the TAI and the need for data which are retrieved more efficiently were the base reasons for the implantation of the system in the herbarium.

Developments in the field of data processing in herbaria have often been mentioned (Forero and Pereira 1976; Vitt *et al.* 1977; Sweet and Poppleton 1977; Morris and Glen 1978; Morris 1980; Morse *et al.* 1981; Jones *et al.* 1983; Gomez-Pompa *et al.* 1985). More recent systems designed for such herbarium procedures include TROPICOS (Crosby and Magill 1989), the AAU-Flora of Ecuador Information system (Frost-Olsen & Holm-Nielsen, 1986), etc.

The first attempt towards the implementation of the data system at TAI began in 1986 (Hsieh 1986). However, it was not till 1990 when the system was made fully operation. The first step taken by the TAI was the encoding of the backlog specimens. At the same time, a geographical grid reference system was initiated to simplify and standardize input of localities. Moreover, a nomenclature file was set up to facilitate the handling of species names. Now good progress

(1) 謝長富、黃增泉、黃星凡, Department of Botany, National Taiwan University, Taipei, Taiwan 10764, Republic of China.

is been made with both encoding and the suite of programs necessary for the storage and retrieval of information.

SYSTEM REQUIREMENTS

1. Hardware requirement

(1) An IBM PC/XT/AT, or compatible computer with at least 1024 KB of memory.
(2) A hard disk or laser disk.
(3) A VGA compatible display monitor.
(4) An analog multisyn RGB monitor.
(5) A Texnai full-color graphic adapter (with a $1024 \times 1024 \times 24$ bit frame buffer).
(6) A Texnai TX 200 color image scanner and a handy scanner.
(7) A mouse system.
(8) A line printer.
(9) A digitizer.
(10) A Roland XY-Plotter.

2. Software requirements

Microsoft Quick Basic 2.0/4.0, dBASE III, AutoCAD, ETen Chinese System, FBX full-color and SkySCAN software programs.

FILE STRUCTURE

Fig. 1 shows the main interrelationships between the more important parts of the system. Briefly the sysem is divided into three subsystems: TAXA, SPECIMEN, and GRAPHICS (IMAGE) files.

1. TAXA Subsystem

The TAXA subsystem maintains a main taxonomic name file which includes family, species, Chinese common names, and habit for each one of the species in Taiwan. Most of this taxonomic information came from Taiwan Flora (Li *et al.* 1976-1979). Currently, there are 4,300 native and cultivated plant species entered in the system. All taxa were arranged alphabetically and numbered in that sequence. Then, these numbers were used in the SPECIMEN file to assure consistently spelled names. In addition to the data file, several program files were used to build, correct, sort and manipulate data file of the names of the specimens.

2. SPECIMEN Subsystem

The second subsytem is the SPECIMEN file which contains information, coded where necessary, in more than 500 fields on each of the specimen. The categories encoded include family, species, collectors, collector's number, date, state of the specimen, grid, altitude, physiographic type, habitat type, specimen number, country, major locality (county), precise locality, on-loan, type status, damage, identifier, duplicate, image, and notes. Remarks on the specimens are stored as separate file. For a multistate category (e.g., habitat type), a hierarchical format of the

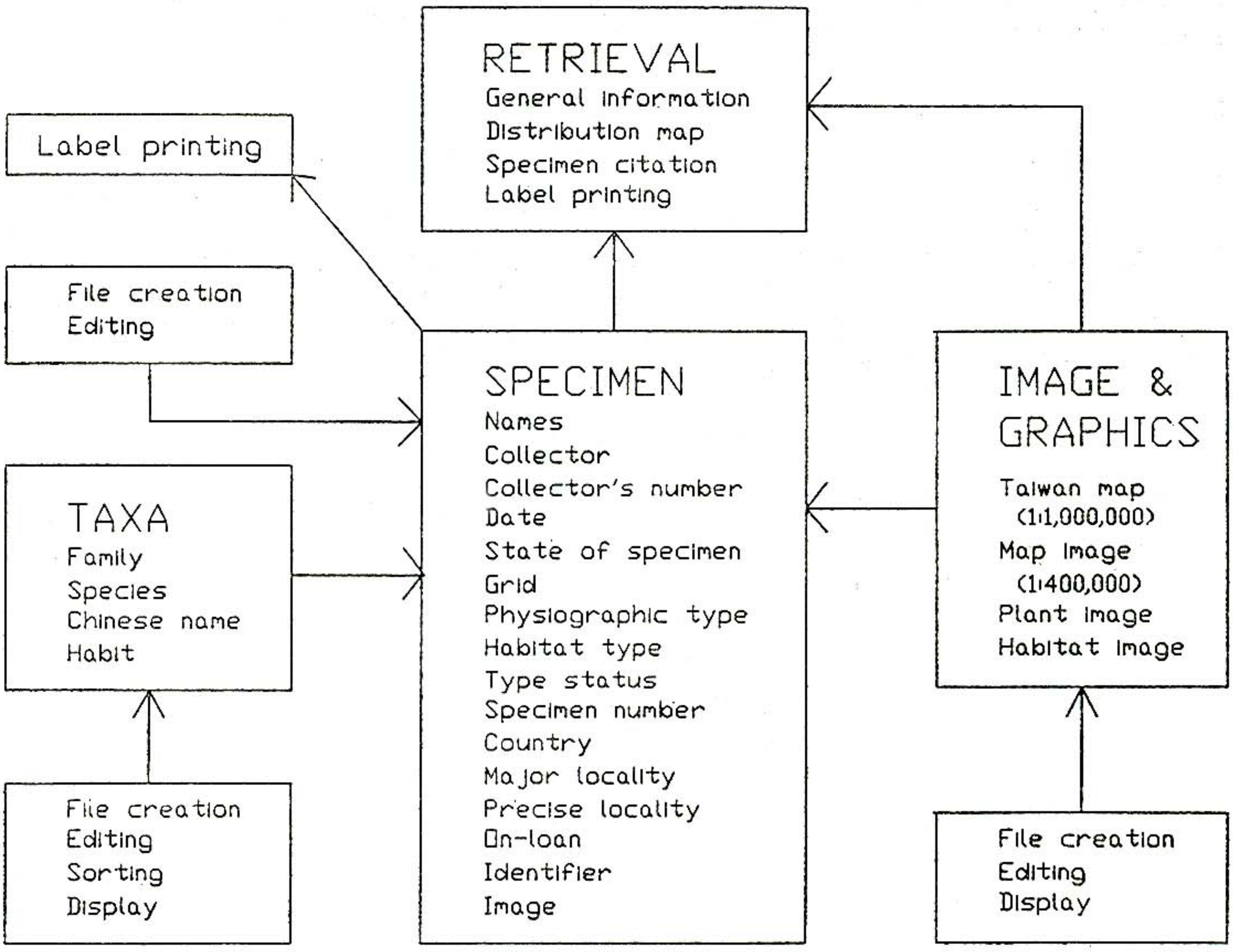

Fig. 1. Structure of the TAIS database system.

items included was provided. They can be called for selection during data encoding. The grid field refers to geographical grid-area bounded by the lines of latitute and longitude. To use this kind of grid-area is because it is given on all better-quality maps and will last longer than the district name. Three different sizes could be selected. The variable size used could better reflect the information provided by the original specimen data. The major grid-unit is the six-by-six minute area of latitude and longitude (X-Y coordinates). A total of 972 main grids (27 * 36) were set up to cover the whole Taiwan Area. Within each main grid-unit, thirty six (6 * 6) one-by-one minute grid-units (XX-YY coordinates) could be divided. The most precise locality was given by intersection of lines of latitude and longitude. For coding, each locality was assigned a number of 20 digits (eg., 10203202134301222030). The first four digits represent coordinates (X and Y) of the six-minute grid, then the following two digits refer coordinates (XX and YY) of the one-minute grid. Latitude and longitude occupy the remaining digits. Up to three of these grid criteria can be used to meet the requirement of collection along a route, such as along Cross-Island Highway from Taroko to Tienhsiung.

3. GRAPHICS and IMAGE Subsystem

Initially, a 1:1,000,000 contour map of Taiwan at 1,000 meter interval was

digitized. Then the major grid-units were generated. Map creation can be accomplished by the drawing function of TAIS or by the commercially available software AutoCAD (then a conversion should be made to transfer the file format into that adopted by the TAIS). For every pair of right-left grids, a correspod-ing map-image file was created by the use of handy scanner. The base map was 1:400,000 contour map at 200 meter interval (Council of Agriculture, 1988). A total of 207 image files were created to cover the whole terrestrial area of Taiwan. Map at lower scale acquires greater accuracy, however it requires much more disk space for storage. In addition to the map-image files, full-color image files could be created to incorporate images of the type specimens, species of particular interest, or habitats into the TAIS system.

DATA PROCESSING

After the program SPECIMEN is run, the TAIS Main Menu appears on the text display screeen. The menu provides access to various parts of TAIS. The menu display looks like this:

Main Menu
0. Exit TAIS
1. Create a new file
2. Enter data
3. Edit data
4. Print labels

As soon as a task number is selected, TAIS data screen appears as shown in Fig. 2. It is now in a full screen edit mode. By using one of the shortcut keys, one can carry out several functions as follows:

Key(s)	Function
Direction	Up and down or left and right.
PgUp	Move to the last record.
PgDn	Move to the next record.
Esc	Exit to TAIS main menu.
End	Save current data.
F1	To show subitems for a multistate category, or to show digitized map and then map-image for entering grid coordinates or latitude/longitude.
F2	To show collection localities of the current record if grid information is available.

Family and species name can be retrieved from the TAXA file by entering the numeric code, or directly entered as characters. Grid information are auto-matically generated. After the grid field is selected, press F1 key to show Taiwan map and six-by-six minute grids on the screen. Then move digitizer stylus to collection area on the map (1:1,000,000 scale) attached on the digitizer. Push stylus button to replace the previous map by the one-by-one minute map-image. At the same time, the one-by-one minute grids are visible. Move cursor to the area or spot where the collection was made. After the button is pushed, three kind of previously mentioned coordinate system (X-Y, XX-YY, and latitude/longitude) are provided for selection. As a response, the TAIS will automatically generate a

```
Family: Leguminosae                                    Record: 15
Species: Abrus precatorius L.

Country: Taiwan
Major locality: Pingtung
Precise locality: Banana Bay
Collector's name: T.C. Huang & C.L. Huang
Collector's number: 8348
Date collected: 04/01/1980
State of the specimen: Fruiting phase
Coordinates: 351652021554312049ll
Altitude:
Physiographic types: Coral reef
Habitat type: Broad-leaved forest
Type status:                          On-loan:
Specimen number: 178496               Deposition: TAI
Damage:                               Image: 1783
Det.: S.H. Huang                      Duplicates: 3
Notes:
〈F1: to enter coordinates; F2: to show localities〉
```

Fig. 2. Screen form for entering and editing specimen data.

number of 20 digits for grid field. If only latitude/longitude is specified, the X-Y and XX-YY coordinates are also presented.

Label printing is one of the basic and salient work at herbarium. The fourth function in the TAIS Main Menu can be called to print out labels with all relevant information of the current record.

INFORMATION RETRIEVAL

Both dBase program and Quick BASIC programs were used for information retrieval. The principal data fields to be indexed would be scientific name, major locality, precise lacality, collector, data, state of the specimen, habitat type, and physiographic type. The query system is designed in such a way that gerneral information to be queried may be initiated on the screen form shown in Fig. 2. The user just position the cursor to the proper items, and then enter the values or combination of values for searches. Other outputs of the query system include species citation, label printing, and distribution map. Distribution records of species onto six-by-six minute map may be produced on a plotter, however the one-by-one minute distribution map-image can only be produced through hard copy on a lineprinter or laser printer. The system also offers a way to show color images of some representative and rare plants with their habitats.

DISCUSSION

TAIS is a versatile, changing database system used for recording and manipulating information about herbarium specimens. The field name and field length can be changed anytime as desired. The incorporation of the map-images into application made it more powerful. Besides, it avoids many tedious tasks to

digitize all the detailed maps. For general uses, a monochrome handy scanner seems quite satisfactory. Color or full-color scanner produces best result, however it needs additional devices such as fullcolor interface card, analog multi RGB monitor, etc. The grid reference system permits three increasing ranges of accuracy for localizing the collection site. This especially useful when encoding backlog specimens, where locations are always non-specific.

In general, the system is operational and functioning. We are at present considering the addition of another file which will be especially useful for the revision works of Taiwan Flora. It will be linked to the TAXA subsystem and will contain synonyms from the literature. Reference to which each synonym is attributed will also be included.

ACKNOWLEDGMENT

This study was supported in part by grant NSC-79-0201-B002-13 the National Science Council, ROC.

LITERATURE CITED

COUNCIL OF AGRICULTURE, ROC, 1988. Map of Republic of China on Taiwan, scale 1:400,000.

CROSBY, M.R. and R.E. MAGILL, 1989. *TROPICOS, a Botanical database System at the Missouri Botanical Garden.*

FORERO, E. and F.J. PEREIRA, 1976. EDP-IR in the National Herbarium of Columbia (COL). Taxon **25**: 85-94.

FROST-OLSEN, P. L.B. HOLM-NIELSEN, 1986. *The AAU-Flora of Ecuador Information System.* The Botanical Institute, University of Aarhus, Report No. 14.

GÓMEZ-POMPA, A., N.P. MORENO, V. SOSA, L. GIDDINGS and M. SOTO, 1985. Flora of Veracruz project: an update on database management of collections and related information. Taxon **34**: 645-653.

HSIEH, C.F., 1986. A data system for Taiwanese grass species (part 1): specimen label data storage and retrieval. Taiwania **31**: 143-150.

HSIEH, C.F. and M.H. SU, 1990. Using PRIS Ver. 1.0, a plant resources information system, Part II. Plant information handling. Ann. Taiwan Museum **33**: 147-167 (in Chinese with English summary).

HSIEH, C.F., 1991. PRIS, a software for vegetation survey and plant information. Taiwania **36**: 136-158.

JONES, S.B., N.C. COILE and R. MARTIN, 1983. The University of Georgia Herbarium (GA) optical-scan data encoding system. Taxon **32**: 47-50.

LI, H.L., T.S. LIU, T.C. HUANG, T. KOYAMA and C.E. DEVOL (eds.), 1975-1979. *Flora of Taiwan*, Vol. 1-6, Epoch Publ. Co., Taipei.

MORRIS, J.W. and H.F. GLEN, 1978. PRÉCIS, The National Herbarium of South Africa (PRE) computerized information system. Taxon **27**: 449-462.

MORSE, L.E., M. LAMSON, A.F. TRYON and R. WALTON, 1981. Specimen locality indexing by the New England Botanical Club. *In* L.E. Morse and M.S. Henifin (eds.) *Rare Plant Conservation: Geographic Data Organization*, pp. 185-191. The New York Botanical Garden.

MORRIS, J.W., 1990. Encoding the National Herbarium (PRE) for computerized information retrieval. Bothalia **13**: 149-160.

SWEET, H.C. and J.E. POPPLETON, 1977. An EDP technique designed for the study of a local flora. Taxon **26**: 181-190.

VITT, D.H., D.G. HORTON and P. JOHNSTON, 1977. A computer program for printing herbarium labels. Taxon **16**: 425-428.

國立臺灣大學植物標本館(TAI)所使用之資料管理系統TAIS

謝長富 黃增泉 黃星凡

摘 要

本文介紹臺灣大學植物標本館（TAI）所採用之資料管理系統 TAIS。該系統適用於標本資料之儲存及運用，所使用設備爲 PC 個人電腦。預期在未來數年中將完成 220,000 份標本之建檔工作。該系統主要由標本檔所組成，附屬檔案尚包含學名檔及圖形影像檔。將地形圖影像納入系統之內，更加强其功能。爲順應標本之記載，TAIS 提供三種標定採集地之方式，卽六分網格、一分網格及經緯度，網格座標及經緯度均自動產生。此外特殊植物及其生育地之影像亦可併入運用。系統之輸出格式包含標籤、標本目錄、分佈圖形、一般資訊以及使用者所設定之特殊查詢項目。