



**Economic and Social Council**

Distr.  
LIMITED

E/CONF.79/INF/34  
18 August 1987

ENGLISH ONLY

---

Fifth United Nations Conference  
on the Standardization of  
Geographical Names  
Montreal, 18-31 August 1987  
Item 6(c) of the provisional agenda\*

TOPONYMIC DATA FILES  
AUTOMATIC DATA PROCESSING (ADP) SYSTEMS

Development of a computerized gazetteer system \*\*

Paper submitted by Malaysia

---

\* E/CONF.79/1.

\*\* Prepared by Abdul Majid Mohamed, Director-General of Survey and Mapping.

DEVELOPMENT OF A  
COMPUTERISED GAZETTEER SYSTEM

1. INTRODUCTION

The Department of Survey and Mapping Malaysia, as Divisional Representative for the Division of Asia, South-East and Pacific, South-West, has been entrusted with the responsibility of publishing and updating the 1) Regional Gazetteer, 2) The Concise Gazetteer of the Division, as well as the National Gazetteer of Malaysia.

Whilst the Department has evolved a manual process to execute the above task, it has proven to be cumbersome and unweildy over the years, not only due to the volume of data involved, but also numerous and frequent changes/updates that are necessary.

In line with the Department's effort to computerise its routine processes, the decision was taken to develop an automated system for gazetteer production based on the use of a micro (personal) computer. It is envisaged that data exchange /update between member countries could be facilitated using a system that is commonly used amongst member countries.

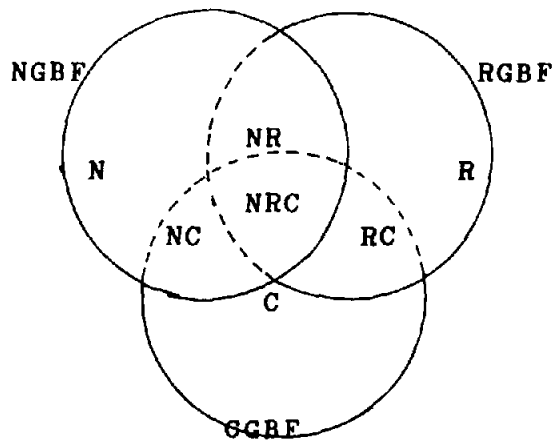
2. OVERVIEW OF THE SYSTEM

The system uses DBASE III Plus on an IBM PC/XT with 10 Megabytes of Hard disk. The system revolves around three base files namely i) The National Gazetteer Base File (NGBF), ii) The Regional Gazetteer Base File (RGBF) and iii)

The Concise Gazetteer Base File (CGBF). It is of interest to note that a place/feature is only captured once in any one of the 3 base files even though the same place/feature may appear in more than one Gazetteer. This ensures not only data integrity but also the ease of updating which by necessity has to be on a continuing basis.

### 3. THE SYSTEM BASE FILES

3.1 As mentioned earlier, 3 base files will be created and maintained. These base files can be depicted pictorially as follows:-



N = places/features in the National Gazetteer only

R = places/features in the Regional Gazetteer only

C = places/features in the Concise Gazetteer only

NR = places/features in the National Gazetteer that also appear in the Regional Gazetteer

NC = places/features in the National Gazetteer that also appear in the Concise Gazetteer

RC = places/features in the Regional Gazetteer that also appear in the Concise Gazetteer

NRC = places/features in the National Gazetteer, that also appear in both the Regional Gazetteer and the Concise Gazetteer

Diagram 1 :  
Pictorial representation of files in the Gazetteer System

### 3.2 The National Gazetteer Base File

The NGBF is made up of N+NR+NC+NRC. It is noted that places/features are adequately annotated such that their contribution to the Regional and Concise Gazetteer can be easily determined.

### 3.3 The Regional Gazetteer Base File

The RGBF is made up of R+RC only. As in the case of the NGBF, places/features that contribute to the Concise Gazetteer are adequately annotated.

### 3.4 The Concise Gazetteer Base File

The CGBF is made up of C only.

## 4. CREATION OF THE HISTORICAL BASE FILES

Action is being taken to edit the latest versions of the compiled National, Regional and Concise Gazetteers such that requirements for data are fully met. Data Capture is in progress to create the National Gazetteer base file.

## 5. UPDATES TO THE BASE FILES

The Department of Survey and Mapping Malaysia will maintain the National Gazetteer. Member countries will be responsible for the supply of new places/features and the amendments to old places/features in their respective countries with regards to the Regional Gazetteer Base File and the Concise Gazetteer Base File.

## 6. PRODUCTION OF THE GAZETTEERS

6.1 Generally, attention is given to the presentation of the Gazetteer to meet with the requirements for ease of readability and reference.

### 6.2 The National Gazetteer

The NGBF is used for the production of the National Gazetteer.

### 6.3 The Regional Gazetteer

A work file is first created by merging the RGBF containing R and RC with those places/features in the NGBF namely NR and NRC. The resultant work-file now containing R+NR+NRC+RC (see diagram 1) is then used for the production of the Regional Gazetteer.

### 6.4 The Concise Gazetteer

A work file is first created by merging the Concise Gazetteer Base File containing C with those places/features in the RGBF and the NGBF namely NC, RC, and NRC. The resultant work file now containing C+NC+RC+NRC (see diagram 1) is then used for the production of the Concise Gazetteer.

## 7. HARDWARE AND SOFTWARE USED

The system, written in dBASE III Plus programming language, can be used in either on-line mode or batch mode. Input is from any ASCII file. It is run on an IBM PC/XT under the MS-DOS operating system VERSION 2.0.

## 7.1 Hardware

This is configured around an IBM PC with minimum of 512K RAM, 2 floppy disks and a near letter quality printer.

The hardware configuration is as follows:-

1 x IBM PC/XT with 512K memory

2 x 360K Floppy Disk Drives

1 x 10Mb Hard Disk

1 x Epson LQ1500 printer

## 7.2 Software

The dBASE III Plus technical specifications are as follows.

### 7.2.1 Database File

Number of records	-	1 billion max
Number of bytes	-	2 billion max
Record size	-	4000 bytes
Fields	-	128 max

### 7.2.2 Field Sizes

Character fields	-	254 bytes max
Date fields	-	8 bytes
Logical fields	-	1 byte
Memo fields	-	5000 bytes max
Numeric fields	-	19 bytes max

### 7.2.3 File Operations

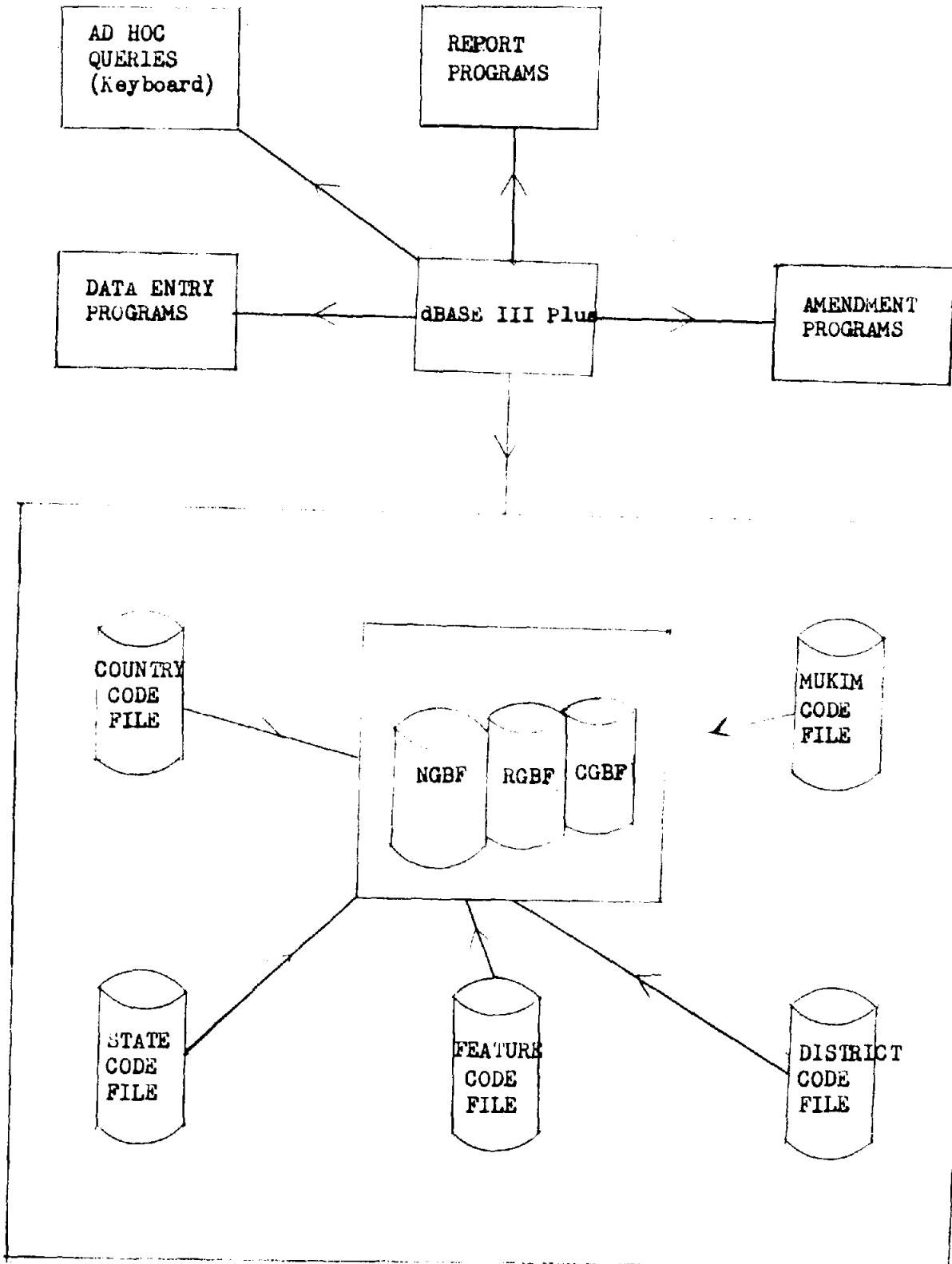
15 open files of all types

10 open database files. A database file counts as two files if memo fields are used

7 open index files per active database file

1 open formats file per active database file

8. Schematic of the Toponymic Data Base



## 9. CONTENTS OF DATA FILES

### 9.1 The National Gazetteer Base File

Data items contained in this file are :-

- a) Place name :- approved standard geographical names
- b) Feature code :- to identify feature
- c) Coordinates :- geographical coordinates
- d) Map series :- map series where this feature appears
- e) State code :- state code where this feature appears
- f) District code :- district code where this feature appears
- g) Mukim code :- mukim code where this feature appears
- h) Effective date :- effective date of changes
- i) Security classification :- to classify feature to a level of security
- j) Regional code :- to indicate whether this feature appears in Regional or not
- k) Concise code :- to indicate whether this feature appears in Concise or not

Note :- The Regional and Concise Gazetteer files contain a subset of the data items in the National Gazetteer Base File.



## 9.2 Feature Code File

Look up table for the Base Files

- a) Feature code
- b) National Gazetteer Equivalent
- c) Concise and Regional Gazetteers Equivalent

## 9.3 District Code File

Look up table for the Base Files

- a) District code
- b) District name

## 9.4 Mukim Code File

Look up table for the base files

- a) Mukim code
- b) Mukin name

## 9.5 State Code File

Look up table for the base files

- a) State code
- b) State name

## 9.6 Country Code File

Look up table for the base files

- a) Country code
- b) Country name

## 10. BENEFITS

As with any other computerised database, it became apparent that there would be obvious benefits to be accrued, such as accurate and efficient storage and retrieval of information, and an output routine that offers the user considerable flexibility.

In the input stage, several validation checks have been incorporated in the software to minimise erroneous data input. The user selects fields for input and the system does not allow a nil-return. For input of coordinates, checks are also made to ensure that the entry falls within the limits of the maximum and minimum coordinate values (or latitudes and longitudes) set to define the bounds of the geographical area (state or country).

The output format can be varied considerably, using the command language of the software. The sorting software allows for the data to be manipulated and sorted into a lexical order (our experience shows certain shortcomings in the standard computer sorting software). Outputs can therefore now be made to suit a variety of requirements.

## 11. SECURITY OF DATA

Certain security controls can be evolved, if necessary, to allow only authorised access to certain data. This is an option that has not been so far exercised.

## 12. DIFFICULTIES ENCOUNTERED

The toponymic database system was developed by using dBASE III Plus software and an IBM PC/XT with 512K RAM and 10Mb of Hard disk. Below are a list of difficulties encountered during the development process.

### 12.1 dBASE III Plus Software

Although dBASE III Plus allows working with 10 open database files at once in separate work areas, linking of files is only allowed in linear form (a parent can only have one child at a time). This is to say a file in work-area-1 can be linked to a file in work-area-2 based on a common data item and a file in work-area-2 can be linked to a file in work-area-3, again based on a common data item, and so on up to maximum of 10 files.

Unfortunately the linkage between the files in the toponymic database is non-linear. So a routine has to be written to handle this situation.

### 12.2 dBASE III Plus Language Interpreter

dBASE III Plus comes with a language interpreter. This language interpreter is slow in handling linkage of data especially in cases like the toponymic data base. This considerably slows down the program. This is particularly obvious in report-producing programs.

### 13. RECOMMENDATION

This represents a first attempt by the Department in developing a computerised gazetteer system. In order to ensure success, the scope of the system has been limited to basic requirements.

The Department is now in the final stages of evaluating a tender for a Computer Assisted Mapping System - CAMS. As map production by computer assisted means is a reality in several national mapping organisations, geographical names processed by computers is necessary to support such a production.

However, for ease of data exchangeability and updating between member countries in the Division, it is envisaged that the present micro-based system would stay for some time.

Department Of Survey and Mapping Malaysia.

August 1987.