



Economic and Social Council

Distr.  
LIMITED

E/CONF.74/L.58

27 August 1982

ENGLISH ONLY

---

Fourth United Nations Conference on the  
Standardization of Geographical Names  
Geneva, 24 August to 14 September 1982  
Items 7 and 9 of the provisional agenda\*

GAZETTEERS

AUTOMATED DATA PROCESSING

The Computer Processing of Geographical  
Names for Map and Gazetteer Production

Paper presented by the United Kingdom

---

\* E/CONF.74/1

## 1. Introduction

The automated processing of geographical names and related information for map and gazetteer production has long been a goal for mapping agencies in the United Kingdom. Until recently, geographical names files have been created during the production of a particular map series. Such files have included the approved name, the type size, style and colour (according to the map series specification), and the position and orientation of the name in relation to other digital map detail. The initiative now is towards the creation of a more comprehensive names file, or toponymic data base. Such a data base, with the capability to support map and gazetteer production, may be described as scale-free.

Without this development it is felt that the processing of geographical names might constrain the advance of digital cartography. The research to establish approved names, the selection and application of these names to mapping, the sorting and preparation of names lists, and the proof checking and corrections to names, together represent a considerable investment of resources in the production of maps and gazetteers. The potential benefits from the automation of this work, particularly from the point of view of maintenance and revision, are considerable.

To date, the processing of geographical names has been carried out utilising main-frame computers, or mini-computers dedicated to particular automated cartographic production systems. The advent of relatively cheap and powerful micro-computers has presented new possibilities for the processing of geographical names by researchers, editors and names authorities. The subsequent selection, sorting and application of names may then be carried out in machine-readable form, without further opportunities for the introduction of errors.

## 2. Gazetteer production

The Mapping and Charting Establishment has recently developed and implemented a production system for gazetteers based on Hewlett Packard micro-computers (HP9835A/B), plotters (HP7245B) and printers (HP9876A). The data for each main entry is recorded for six categories or fields of information: approved name, designation (letter coded), grid reference, latitude, longitude and sheet number. The data for each cross-reference entry is recorded in two fields: variant name and approved name.

Each entry has a maximum size of 62 characters. It is therefore possible to store four entries in each computer physical record (consisting of 256 characters). The data is recorded from map and document sources on to cards. Each card is allocated a data entry number which is the address for recall from the data tape. The data is input continuously in random alphabetical order as the research progresses. Alphabetical sorting and formatting for gazetteer output is carried out by the micro-computer software. Data is checked by the researchers during the input against a formatted print-out of the input. Amendments are carried out using correction software.

The final gazetteer print-out is used directly to produce the printing plates. A sample gazetteer page is shown at Annex A. Subsequent to this print-out, the software for sorting alphabetically has been further improved. To date, names data has been recorded for two areas, each with approximately 4500 entries. This data is stored on cassette tapes, each with a capacity of 250,000 characters. Eight hundred records are utilised for data, allowing 3200 name entries per tape. The data, recorded in random alphabetical order on the entry tapes, is read to the memory of the micro-computer (50,000 character capacity) in appropriately sized alphabetical groupings, each tape being read once, for sorting, formatting and output.

To enable the handling of names files for larger territories, floppy disc storage with a capacity of 1,200,000 characters per disc is being acquired. A further enhancement of this capability is planned, with the acquisition of a dedicated digitizer to enable the direct collection of grid and geographical coordinates from names research map overlay traces, and for the collection of designation codes through a menu system.

### 3. Map production

At the Mapping and Charting Establishment the production of aeronautical charts at scales of 1:500,000 and 1:1,000,000 by automated means is now well advanced. The automated type plotting of geographical names utilises software drawn fonts with character sets for extended Roman alphabets. The type size, style and colour according to the series specifications are encoded with a six character code, as are other digitized features. At present, the placement of the selected names relative to other detail is carried out manually on a plot of the other detail. The position and orientation of this manual placement is then recorded for subsequent machine plotting. A disadvantage of this method is that some processing of the name is duplicated at each scale in the creation of product specific files. A more flexible approach is possible.

#### 4 Toponymic data base design

The initial design of a file structure for an improved names record with the capability to support both map and gazetteer production has been undertaken. The data input proforma for this file is shown at Annex B. The file has been designed with 17 fields with a maximum size of 253 characters per names record to occupy one physical record of storage. The first seven fields provide main entries for gazetteer outputs. Cross-reference entries for mis-spellings, conventional names, former names and variant transliterations are recorded in fields 9, 11 and 13 and may be related to the approved name in field 1. For map production, names may be selected by any of the fields, any combination of fields, or by specified information: for example by population bands in field 16, by administrative importance in field 15, by geographicals and designation in fields 2, 3 and 4, by sheet number in field 7, by source codes keyed to source dates in fields 8, 10, 12 and 14. Field 17 is included to facilitate any work involving the encoding of characters.

The aim will be to provide the map compiler with the ability to select names from the toponymic data base. The names will be presented alongside other digitized map detail for the encoding of type characteristics and for the positioning relative to that detail, to create product files. Approved names and any variant names required for inclusion on any map may then be directly accepted without the need to retype the name. Such a toponymic data base interfaced with production work or edit graphics screens, computer typesetters and plotters would greatly enhance the production capabilities of mapping agencies. The file design is presently being evaluated.

#### 5 Conclusions

The advantages of automated names processing are seen as follows:

- (1) The reduction of name errors resulting from the multiple handling required by non-automated systems.
- (2) The avoidance of repeated names research at various scales, through the creation of product independent files.
- (3) An improved facility for revision and update of names files from new sources.
- (4) Control of the approved name remains in the hands of the names authority, or of editors and researchers working under their policy direction.
- (5) A variety of outputs, lists, and plots or graphic displays is possible, both for the specialist map compiler and for the non-specialist user.
- (6) Decreased response times for access to the names data is possible for all users.
- (7) Names records, once created in machine-readable form, may be readily adapted and interfaced with automated production systems and exchanged with other agencies.

NAME	DESG	GRID	LATITUDE	LONGITUDE	SHEET NO.
TRYPOT	AREA	7308	51.21S	58.49W	07
TRYPOT, THE	LCTY	7109	51.21S	58.50W	07
TUESDAY ISLAND	ISL	9704	51.22S	59.54W	05
TUMBLEDOWN	AREA	3472	51.41S	57.57W	15
TUMBLEDOWN MOUNTAIN	HLL	3372	51.41S	57.57W	15
TUMBLEDOWN, THE	LCTY	8989	51.41S	60.15W	10
TURKEY ISLAND	ISL	8682	51.34S	60.18W	10
TURKEY ROCKS	HLL	0991	51.29S	59.44W	05
TURKEY ROCKS VALLEY	VAL	0994	51.28S	59.44W	05
TURN ISLAND	ISL	6923	52.07S	58.54W	27
TURN ISLAND HILL	HLL	6830	52.03S	58.54W	27
TURNER'S STREAM	STM	3478	51.38S	57.57W	15
-do-	STM	3081	51.38S	58.00W	14
TURNERS STREAM	STM	9389	51.32S	58.32W	13
TUSSAC ISLAND	ISL	1863	51.46S	58.11W	22
TUSSAC ISLAND	ISL	9901	51.23S	59.53W	05
TUSSAC ISLAND		SEE CHRISTMAS ISLAND 9304			
TUSSAC ISLAND		SEE HAMMOND ISLAND			
TUSSAC ISLAND	ISL	7012	52.12S	60.29W	24
-do-	ISL	7012	52.12S	60.30W	23
TUSSAC ISLANDS	ISLS	3934	52.01S	60.58W	23
TUSSAC ISLANDS	ISLS	4975	51.40S	57.44W	15
TUSSAC POINT	HOLD	7450	51.52S	60.28W	18
TUSSAC POINT	PT	3897	52.20S	59.22W	28
TUSSAC POINT	PT	4374	51.40S	57.49W	15
TUSSAC ROCK	RKL	1906	51.23S	58.09W	08
TUSSAC ROCKS	RKSL	6911	51.20S	58.52W	07
TUSSAC ROCKS VALLEY	VAL	6911	51.19S	58.52W	07
TUSSAC VALLEY	VAL	7838	51.59S	58.46W	21
TUSSACK ISLANDS		SEE TUSSAC ISLANDS 3934			
TWEEDS VALLEY	VAL	3505	52.16S	59.24W	28
TWELVE APOSTLES HILL	HLL	9279	51.37S	58.33W	13
TWELVE O'CLOCK HILL	HLL	2296	51.28S	58.07W	08
TWELVE O'CLOCK MOUNTAIN	HLL	3980	51.37S	57.52W	15
TWIN PONDS	LK	1310	52.13S	59.44W	25
TWINS, THE	HLL	6420	50.14S	60.38W	02
TWINS, THE	ISLS	6420	50.14S	60.38W	02
TWO BOB VALLEY	VAL	0479	51.36S	59.49W	11
TWO PASS STREAM	STM	0075	51.38S	60.06W	10
TWO POND RIDGE	RDGE	5553	52.03S	60.44W	23
TWO POND VALLEY	VAL	5628	52.04S	60.42W	23
TWO PONDS	AREA	5427	52.05S	60.44W	23
TWO SISTERS	RKSL	3073	51.41S	58.00W	14
TWO SISTERS	AREA	3173	51.41S	57.59W	15
-do-	AREA	2873	51.41S	58.01W	14
TYSSSEN ISLANDS	ISLS	1848	51.53S	59.38W	19

TOPONYMIC DATA BASE PROFORMA

Input Number

1 Approved Name

\_\_\_\_\_

2 Designation

\_\_\_\_\_

3 Latitude

\_\_\_\_\_

4 Longitude

\_\_\_\_\_

5 Grid

\_\_\_\_\_

6 Administrative Area

\_\_\_\_\_

7 Series Sheet Number

\_\_\_\_\_

8 Sources for Approved Name

\_\_\_\_\_

9 Variant Name (1)

\_\_\_\_\_

10 Sources for Variant Name (1)

\_\_\_\_\_

11 Variant Name (2)

\_\_\_\_\_

\_\_\_\_\_

12 Sources for Variant Name (2)

\_\_\_\_\_

13 Variant Name (3)

\_\_\_\_\_

14 Sources for Variant Name (3)

\_\_\_\_\_

15 Administrative Information

\_\_\_\_\_

16 Population/Dimension

\_\_\_\_\_

17 Character Codes for Approved Name

\_\_\_\_\_

\_\_\_\_\_

Name

Date

Characters

Prepared

Input

Checked