



Economic and Social Council

Distr.
LIMITED

E/CONF.79/L.65
16 July 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: AUTOMATED DATA PROCESSING (ADP) SYSTEMS

Biscriptual vs. bilingual automated gazetteers**

Paper submitted by Israel

* E/CONF.79/1.

** Prepared by Naftali Kadmon, Professor of Cartography, Hebrew University of Jerusalem, Israel.

represented by numerical codes. Thus we know that on 1 May 1987, 18.6 per cent of toponyms in Israel were names of populated places, 15.5 per cent were antiquity sites and ruins, 17.8 per cent were perennial streams and seasonal watercourses (wadis), 15.1 per cent were names of mountains and hills, 4.1 per cent were nature reserves and national parts, and 0.3 per cent - marine features. Since each record, i.e. the entry for each toponym, carries the full gazetteer information on the respective place, each of the gazetteers or printouts mentioned above, by whatever criteria it is arranged or ordered, can carry the complete data against each name. Records in the automated file of toponyms in Israel are formatted, with each item or characteristic, whether alphabetic or numeric, whether Hebrew or Romanized, occupying a specified field of given length. This enables simple software to produce the various printouts mentioned above.

Amendments and updating: search routines

Amendments and revisions may take the form of new (additional) records, revisions to existing records, as well as deletions. In the past, when the gazetteer resided on a series of diskettes, all amendments were made there, new records being added at the end of the last diskette. The file was then read into main memory and processed for output. Today the entire file resides on main disk (with magnetic tape backup), where all amendments are introduced and integrated.

Search routines can be conducted by any keyword specified in either language of the gazetteer.

Lineprinter and fan wheel vs. dot matrix printer

The first edition of the gazetteer of Israel was produced on a lineprinter with a biscriptual Hebrew-Roman character printing chain (as used with many computers in Israel). At present a fan wheel printer with a biscriptual character "daisy" is being used. Since both types of lineprinter lack some of the diacritic signs in Romanization, the latter must be deduced by the reader from the Hebrew form. This form, on the other hand, has no vocalization (which is not usually required by the average Israeli reader, but which may be reconstructed in the main, e.g. from the Romanized form; the latter is, of course, vocalized).

Since all characters, including diacritics and vocalization in Hebrew, can be produced on a dot matrix printer with the aid of software (and most of the American and Japanese printers in Israel are supplied with standard Hebrew as well as Roman characters), the next generation of the gazetteer of Israel will be amended so as to make use of the possibilities given by the dot matrix and laser printer.

With the universal trend towards Romanization of non-Roman scripts, advocated by the United Nations Group of Experts on Geographical Names and the United Nations Conferences on Geographical Names, the problem of producing biscriptual gazetteers of place names is becoming increasingly important and pressing.

A bilingual (or multilingual) gazetteer includes toponyms of different languages written in the same script, though perhaps with different diacritic signs. Word order can be established between diacritics (such as before ä in German, e before é and è in French, l before ł in Polish, etc.); but even without such an order of precedence a fully integrated bi- or multi-lingual gazetteer can be produced, as used in most world atlases.

A biscriptual gazetteer

Biscriptual gazetteers reflect an entirely different set of conditions and present a different picture. Two different scripts cannot normally be integrated as regards their alphabetic order (even assuming alphabetic, not syllabic or logographic, scripts). While the alphabetic listing of names in the two scripts must be separated, this does not necessarily mean that the two sets of names, one in each script, cannot be produced in a single document. But the alphabetic listing must follow the letter order of either one script or the other.

Two cases can be distinguished. A biscriptual country such as Yugoslavia might wish to produce an automated gazetteer of names printed in two different scripts used in different regions of the country, e.g. Cyrillic and Roman. The second case would be a country with a non-Roman script wishing to produce a gazetteer of the original standardized as well as the Romanized form of all place names.

The automated gazetteer of Israel

A good illustration - and the first example of a biscriptual gazetteer ever produced by automation - is furnished by the gazetteer of Israel, which carries all names in both Hebrew characters and in the official transliteration into Roman script. A single integrated file includes the entire contents of the gazetteer, and each record carries the standardized name in Hebrew and in Roman characters, as well as other toponymic, toponomastic and topographic information. Separate printouts can then be obtained via suitable software, listed by any characteristic desired. Thus, a list ordered alphabetically by standardized Hebrew name can be produced, or by Romanized name form (the latter in Roman alphabet order). Secondary name forms can also serve as the main indexing characteristic. All the above are alphabetic criteria. Alternatively, numerical characteristics can be used. One of these is a listing by topographic co-ordinates, in the Israel Grid or in the UTM (Universal Transverse Mercator) Grid, e.g. from north-west to south-east. Another, quite useful one, is the date of ratification of the name by the Government Names Commission. Besides being of general or research interest, this reflects the annual rate of generation of new names. Thus, the years 1965, 1975 and 1985 (to name just three) resulted in 192, 223 and 207 new names, respectively. Altitude of the geographical feature is another listing index. But perhaps the most important one is the generic type of the feature named,