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

TOPONYMIC DATA FILES

Toponymic data files

Paper submitted by Sweden

* E/CONF.79/1.
TOPYNOTOMIC DATA FILES.
Paper presented by Sweden.

1. Establishing a place names data base.

In 1984 work was started on the systematic development of a place names data base for the 1:50 000 and 1:100 000 Topographic Map series. The database is being built up outside the framework of the mapping programme, but the aim is to use it in production in the future. Work with the base is scheduled to be completed by the end of 1987.

The base will contain all place names shown on the Topographic Map series i.e. settlement names - Gävle, Bäckaskog, Storlien, names of natural features - Ring-sjön (lake), Göta älv (river), Ammarfjäll (mountain) and Kolmården (forest). It is estimated that approximately 500 000 names will be captured. The names will be related to the Swedish National Grid (not to the UTM grid). A coding system will be used which will make it possible to create separate data files for specific objects such as settlement names, hydrographic features etc. The relationship between place names and administrative divisions will also be established in the base.

In its role as place name authority with the responsibility for the standardization of geographical names, the National Land Survey (NLS), aims to create a place names data bank which will be a reference source for other authorities and agencies, local government bodies, the media and the general public.

1.1. Methods.

The method which will be used to create the base is an application of the methods used for computer-assisted
photo-type setting. The software is described in two professional paper, Autoka Software System (Swedsurvey publication dated 1985) and Autoka Hardware Configuration (LMV-rapport 1983:12) both published by the NLS. The principal modification to the basic system has involved the addition of information concerning administrative affiliation.

The data is captured unchanged from the printed maps and thus includes both relational and orthographic errors. Corrections to the data are also keyed in to the base and errors are flagged with a code which will ensure that they will not be included in new map production runs.

1.1.1. Coding and listing.

Every geographic name is given a category code. The place name and its code are keyed into the base through an alpha-numeric terminal. All abbreviations are replaced by the full name eg Svarttj. is keyed in as Svarttjärn and V. Aby as Västra Aby; both parts of hyphenated names are keyed in to be written on the same line. Editing and correction is done with the help of print-outs from the base.

1.1.2. Digitizing.

Coordinates in the National Grid are captured by manual digitizing following strict routines. The map sheet is broken down into 5x5km squares and points are digitized sequentially in rows alternately east-west and west-east. Coordinates for the names of natural objects are registered for a point which is selected as close as possible to the centre of the object; for settlement names, coordinates are measured for the first or last letter depending on which lies closest to the conventional sign. The coding and listing is done using the same routines which makes the checking relatively easy.

1.1.3. Checks

The captured data can be retrieved from the base in the form of lists and plots. Points defining coordinates for settlement names are plotted as black circles and the position of the names of hydrographic features as blue circles on clear film at the map scale. The clear
film is carefully placed in position on the map sheet and a check is made that every name on the sheet is represented by a circle. The spelling of place names is checked against the print-outs at the same time. Corrections are made using the digitizing tables and graphic work stations.

1.1.4. Administrative sub-divisions.

County, commune and parish boundaries are schematically plotted with as few break points as possible on the clear film. This is done to check that the coordinated points fall within the correct administrative area. Thereafter, the administrative sub-divisions are coded according to the Central Bureau of Statistic's coding system. This is a six figure code: the first two digits are the county, the second two the commune and the fifth and sixth the parish. The break points are digitized and the code figures are keyed into the computer via a terminal.

1.2. System users.

Currently those areas of the place name data base which are complete are mainly used for producing files on a county basis for internal use and for external clients. The interest for using the place name data base to produce various types of place name registers is very considerable, despite the very limited information that the NLS has, up to now, made available. Most clients have requested paper print-outs from the base, but a few have purchased magnetic tapes to run on their own hardware. Place name registers on a county basis are used by the Swedish accident emergency rescue organisations. A register containing settlement names has been made in connection with the production of a county road atlas. Communal place name registers have been sold to taxi organisations and to the police.

When the data base is fully developed a nationwide place name file will be produced. As the data volume is very large, considerable efforts are being made to decrease access times. It is hoped that the remaining technical problems will be solved before the end of 1987 which will make it possible to produce the place name file as soon as the data base is complete.
1.3. Gazetteers.

The place names data base will be used to produce a revised edition of the Swedish place names lists (Svensk ortförteckning) which have been published by a group of Swedish authorities. The lists contain of the order of 130 000 settlement names which are of interest for the Postal and Telecommunication Administrations, the State Railways etc. The revised edition will not include more names but will include the names of both settlements and natural objects. The orthography used in the present names lists differs from that used for the national maps and tends to follow local praxis. After the revision the lists can be used as a Swedish gazetteer.

The NLS created a data base for the national 1:250 000 map series several years ago. This base contains not only place names but also elevation and depth figures and marginal text. This base contains between 30 000 and 40 000 geographic names and after modification can be used as an international gazetteer in accordance with UN recommendations. Plans for doing these modifications are being prepared.

Following the decision to replace the older cartographic orthography with the north Lappish orthography, the NLS plans to publish a special Lappish gazetteer. Experience has shown that there is a demand for a cross-reference gazetteer containing both old and new spelling and spatial information. A gazetteer of this type is planned to be produced in near future.

2. The Swedish land data bank project.

In the early 1970's work on the creation of an EDP-based information system for storing, processing and presenting real estate data was started. The computer-based system is gradually replacing the old manual systems based on the old real estate registers.

In the new register each of the 4 000 000 property units in Sweden will have a unique designation consisting of the name of the commune the area name and a register number such as Gävle Varva 17:66.

In the rural areas, the area name is often a village name although it can also the name of a farm or even a
single property. In urban areas the area name is that of a block or urban district. In small communes the area name is the commune name. This means that the the new computer register, when completed, will contain approximately 100,000 names. It will also contain the names of all of 284 communes in Sweden as well as the names of the 2,566 parishes. This makes it possible, for example, to produce lists of all names within a given commune or parish. Before the names are stored in the base they are linguistically examined.

Currently about the system contains about 50% of the real estate properties, and it will be fully completed by 1993/94.