===
SIXTH UNITED NATIONS CONFERENCE ON THE
STANDARDIZATION OF GEOGRAPHICAL NAMES
New York, 25 August - 3 September 1992
===

COMMITTEE II: TECHNICAL PROGRAMMES

Draft Report

(Submitted by the Secretariat)

C. <u>Committee II: Technical programmes</u>

<u>Toponymic data files (item 6)</u>

<u>Data collection procedures (6a)</u>

This subject was not addressed by the Conference.

<u>Data elements required (6b)</u>

The United States of America presented document
E/CONF.85/L.26 on behalf of the Place Name Survey of the United
States (PLANSUS). The Format and Attribution Committee of
PLANSUS had determined a specific set of required data elements
that must be utilized in carrying out toponymic research to be
sanctioned by PLANSUS. Additionally, the Committee had
identified a need for optional data elements which would be
maintained by a particular organization.

Questions were answered from various delegates on aspects of
pronunciation, location, format, the naming of extensive linear
features, and source materials. It was explained that PLANSUS
was a private organization. It was further elaborated that the
goals of PLANSUS are to encourage those doing toponymic research,
in accordance with the formatting procedures established by the
National Geographic Names Database of the U.S. Government.

Canada, introduced document E/CONF.85/L.48 which detailed
"core" data fields required for that country's federally-
maintained database. Names are recorded for physical features
and populated places, but not for buildings or streets. Some
two-thirds of the entries have been approved by the Canadian
Permanent Committee on Geographical Names (CPCGN); the remaining
names were unofficial or cross-reference entries. The history of

the development of names records in Canada was outlined. It was noted that in recent years with the introduction of automation, there was sometimes a divergence in approach to aims of toponymic data storage by the federal government and the provincial/ territorial authorities. The need to define such core data fields had, therefore, arisen.

Automated data processing (ADP) systems (c)

Austria referred to document E/CONF.85/L.15, and commented that a data file had been developed containing entries for settlements, physical features and administrative units of the Danubian countries. The data file made use of both Roman and Cyrillic script, and included variant names.

The United States of America presented a document E/CONF.85/L.25 which was an information sheet describing the U.S. Geographical Names Information Systems (GNIS). It was explained that all entries and their geographical coordinates were based on 1:24,000 or 1:25,000 scale mapping, and that each name entry in the database was official.

Document E/CONF.85/L.51 was presented by the Commission de toponymie du Quebec, and described the new, updated, toponymic data-processing system (TOPOS) which contained more than thirty data elements. The principal users of this data-processing system were government departments.

Document E.CONF.85/L.73 was presented by Japan. The status of automated data processing in this country was outlined, and attention was drawn to the geographical names data file which

2

would be completed by the end of 1992 by the Geographical Survey Institute. The file would contain 120,000 geographical names, based on the 1:200,000 scale regional maps.

An overview of the national Canadian Geographical Names Data Base (CGNDB) (document E/CONF.85/L.49) was presented. The hardware and software capabilities of the system were described and the ensuing studies of compatibility with provincial databases was described. It was indicated that the CGNDB was developed primarily for cartographic application and gazetteer production, but that it was not scale-dependent.

Venezuela introduced Working Paper No. 7, and described the design and development of that country's national toponymic database. This included procedures for formatting gazetteers and geographical names dictionaries.

The United States of America described the Geographic Names Processing System (GNPS) (Working Paper No. 8), which was designed for foreign names application in the United States of America. It would contain 4.5 - 5 million entries, generally based on a map scale of 1:250,000, and was intended to be operational in early 1993. The hardware and software requirements were described and it was indicated that the system would make full use of diacritical marks. The system was designed to support map and gazetteer production and future development would include the ability to handle non-Roman writing systems. It was stated that the database would contain exonyms, mainly names of countries, and that the geographical coordinates

for each data entry would represent the location of the feature, rather than the name placement on the map.

Compatibility and structure of systems (d)

Document E/CONF.85/L.50 which described the long-term vision and development plan for a Canadian digital toponymic service was presented. The aim of this plan was to reflect the mandates of provincial, territorial and federal names authorities, and allow the broadest possible dissemination and use of geographical names information. It was indicated that there could be opportunities in the future to link government databases with academic databases for research purposes.

Austria indicated in document (E/CONF.85/L.16) that two national digital files were available: a gazetteer of inhabited places and a register of buildings and street names. An additional file had been developed containing geographical names based on 1:50,000 scale mapping, and collected during the most recent census (1991). Procedures for updating and the nature of the various data elements were described. It was reported that approximately 60,000 names had been collected, and a further 60,000 were estimated to be collected by 1995.

The United States of America presented E.CONF.85/L.29 and outlined general procedures for developing an automated toponymic database. Specific topics to be addressed were the determination of products and application, as well as database content and requirements. Further attention was given to

hardware and software requirements together with methodology for data collection, data entry, quality control and maintenance.

Canada presented E/CONF.85/L.52 dealing with the recording of aboriginal names using syllabics and modified Roman alphabet characters. It was stated that Canada has ascribed to ASCII code and ISO standards, but that no standard existed for encoding such writing systems. It was further indicated that Canada did not wish to invent such standards but stressed an immediate need for an interim solution. The United Kingdom, recommended patience in dealing with modified alphabets, and remarked that ISO was considering such topics.

National gazetteers (e)

An update was provided on the data compilation and gazetteer programme of the United States of America (E/CONF.85/L.30). It was indicated that approximately 75 percent of the 25 year project had been completed and it was estimated that the project would take a further ten years. The period of compilation was about 3 to 4 years per state, and a graphic indication of the status of compilation of each state gazetteer was provided. States in active compilation could add thousands of records to the database on a monthly basis. Publication of conventional gazetteers was a slow process and generally required about ten months of editing and processing.

Germany reported E/CONF.85/L.94 that the gazetteer of the Federal Republic of Germany was published in 1981. A supplementary volume to cover the former German Democratic

Republic had been prepared. The differences between the two systems of data collection and representation would require additional time for analysis, and publication of a volume for a unified Germany is anticipated in 1993.

The United States of America presented Working Paper No. 9 to describe its publication of foreign gazetteers since the Fifth Conference. These gazetteers were generally based on 1:50,000 scale mapping and the categories of information were generally locative. There were two approaches to gazetteer revision: either a total or a limited revision based upon a survey of source materials. Delegates were informed of the increase in revision of gazetteers covering some Latin American countries, a response to the need for information in the anti-drug field. A revised gazetteer of undersea feature names had also been published. It was stated that the inclusion in foreign gazetteers of pronunciation guides was not practical, but that spelling is of paramount importance.

In thanking the United States of America, for the production of the "Gazetteer of South Africa", South Africa offered its services in dealing with orthographical problems related to the country.

Other publications (f)

Austria introduced document E/CONF.85/L.17 as the index to E/CONF.85/L.15, and described the content and categories of Austria's comparative multilingual gazetteer of the geographical

names of Danubian countries. Attention was drawn to the fact that the transliteration systems used were those of the ISO.

Romania introduced document E/CONF.85/L.20 describing the national gazetteer of geographical names based upon a map scale of 1:500,000, and published in accordance with resolutions of UN Conferences. The gazetteer is divided into four sections. The first contains the names of departments, with their short forms and principal towns; the second and the third municipality names. with their geographical coordinates; and the fourth physical features with geographical coordinates. The procedures followed in the drafting of this gazetteer are those found in toponymic guidelines of Romania.

Canada presented E/CONF.85/L.53 by the Commission de toponymie du Quebec and provided an illustrated page from the forthcoming dictionary of geographical names in Quebec. This dictionary details cultural and historical information for some 6,000 names, and is to be published in 1993 with maps and illustrations.

Thailand presented Working Paper No. 32, "A list of country names and their capitals".

Canada presented the Report of the Working Group on Toponymic Data Files (E.CONF.85/L.40). It was identified that the primary tasks of the group are:

1) to collect information on toponymic databases (whether automated or manual) through a questionnaire (E/CONF.85/INF/7).

2)   the need to recommend that guidelines on toponymic
      databases be made more widely available and that
      software, free of copyright restrictions, be
      disseminated

3)   the need to include a basic workshop on automated
      databases in the various training courses.

Sweden elaborated on the results of the above mentioned
questionnaire. So far there were 33 replies from 28 countries,
so too little information was yet available from which to draw
any meaningful conclusions.

The questionnaire did not concentrate only upon
national toponymic databases and, therefore, the answers dealt
with databases covering special categories of names as well the
names in several nations.

Canada, Denmark, Hong Kong, Hungary, Malaysia, Norway,
Sweden and United States of America have completed digital
national databases where coverage is uniform and names are
treated by a national names authority.   Israel mentioned that
its national digital database is both multilingual and
multiscriptual.

The United Kingdom and Switzerland have toponymic databases
in the same sense as above, but the names are not treated as
official by a national names authority. There are advanced plans
for digital toponymic databases in Ethiopia, the Philippines and
Thailand.

All nations having toponymic databases are collecting their names from official maps and/or from fieldwork. Historical material and gazetteers are sources used by several nations.

The most common map scale is 1:50,000 used by 15 nations, however several databases are based on various scales. The list of data elements or fields of information shows fourteen major elements. The main fields in existing databases seem to be type of feature, map reference, coordinates, administrative units, source of name, status of name and variant names.

Most of the responses indicate continuous updating, the most common sources of updates are decisions made by national names authorities, published maps and fieldwork.

The most common reasons given for producing toponymic databases are the creation of a national gazetteer and cartographic applications. The questions regarding the Roman alphabet, diacritics and other languages should be dealt with separately in the future.

Almost every nation having a toponymic database also has documentation available. The documentation is usually user manuals in hardcopy format, but a few have digital documentation.

Several nations are asking for technical and financial support. Specifically there are questions about the role of the United Nations and the role of the linguistic divisions of UNGEGN. Also decisions on appropriate international data element and data exchange standards have been requested.

7. <u>Terminology in the standardization of geographical names</u>

Israel presented the report of the UNGEGN Working Group on Terminology. It outlined the history of the Working Group, and presented the new version (1.2) of the Glossary (E/CONF.85/CRP.1). It was indicated that comments received subsequent to the 31 January 1992 deadline had been evaluated and processed accordingly. It was further proposed that the Conference accept the Glossary and present it to the United Nations for translation into the other official languages. Also an appropriate member of the Working Group should be appointed by UNGEGN to supervise the translation into each language.

The United States of America inquired as to the extent of collaboration by the UN terminology services in the preparation of the document, and wondered if there had been any duplication of effort by the two bodies. It was replied that the Working Group was autonomous, but that the United Nations would be involved in the production of the Glossary.

It was reported on behalf of the Place Name Survey of the United States (E/CONF.85/L.27) that PLANSUS would await publication of the UN Glossary in order to avoid duplication of effort and discrepancies in the production of its own dictionary of toponymic terms. The United States and Canada felt that any glossary should be specialized and aimed at the research community.

Hungary presented document E/CONF.85/L.70, offering remarks on definitions. It was recognized that the document may already

be out of date, as the comments included were based on the previous version of the glossary published in May 1991; many terms had already been corrected. Czechoslovakia supported certain corrections of terms included in the paper. Israel thought it necessary to add the term "minority name", and commented that UNGEGN had decided in the past that the absence of diacritics did not turn a name into an exonym. The new Glossary had been used for both lectures and exercises at the recent training course held in Pretoria.

The United States of America presented Working Paper No. 19 on linguistic terminology in toponymy, and emphasized that approximately one third of the terms in the glossary were derived from the science of linguistics. It was felt that terminology was needed for purposes of communication and it was recommended that definitions be clear and concise. The delegate elaborated upon the high level of interrelationship among terms and highlighted the consequent danger of arbitrary additions, deletions, and corrections.