E

UNITED
NATIONS

Economic and Social Council

Distr.
GENERAL
E/CONF.91/CRP.27
14 January 1998

ENGLISH ;ONLY

SEVENTH UNITED NATIONS CONFERENCE
  ON THE STANDARDIZATION OF
  GEOGRAPHICAL NAMES
New York, 13-22 January 1998
Item 6(d) of the provisional agenda

TOPONYMIC DATA FILES: AUTOMATED DATA PROCESSING
(ADP) SYSTEMS

Development of Russia's National Catalogue of
Geographical Names

Paper submitted by Russia**

Development of Russia's National Catalogue of Geographical Names

Collection and cataloguing of geographical names is the main pledge of their storage and unified usage.

Until iecently, in this country files of geographical names were of corporate character They weie aimed at the achievement of ceitain goals within the teims of reference of ceitain agencies, and weie developed by diffeient methods, being totally incompatible. As a result, the files diffei from each othei both in theii contents and data completeness. The agencies developed specific catalogues and peiiodically issued these data in iefeience books.

After the USSR dissemination and changes in the political structure of Russia, in course of rearranging state and economy administrative agencies, the maintenance of the mentioned data bases ceased and the most their data became obsolete.

The card files of the caitographic and geodetic service of Russia that is still being maintained by CNIIGAiK, is of generic character.

These files are the entity of card records for all the countries of the World. As a result of many years work, two separate card files were formed - the card file for the USSR and the card file foi foreign countries. The structure of the files has not been recently changed. The card file of the Federal Service of Geodesy and Cartography involves the toponymic data on more than 1.5 million geographical features, including nearly 450 thousand for Russia. The file volume is constantly enhanced and updated to mainly meet the requirements of map making. However, as the card file was initially intended for manual work with cards and monitoring maps, now it is not up to modern requirements and technological level. Consequently, it cannot evolve and timely respond to users queries. This resulted in urgent necessity to use more prospective technologies.

This issue is extremely vital for Russia. And its vitality is confirmed by article 10 in the Federal Law "On naming of geographical features", that seeks to develop and maintain the National Catalogue of Geographical Features of Russia

Conversion of separate corporate files into the master state catalogue is a new task, requiring preliminary investigation. Current investigations are mainly aimed at determination of the ways to develop in Russia the national catalogue of geographical names. It involves careful investigations of the other countries experience in development, maintenance, updating and usage of national files of geographical names in different forms, including automated information systems and computer toponymic databases, fiist of all in the USA and Canada, which have achieved the best results in developing national automated information system.

The National Catalogue of Geographical Names should be constantly updated reference file of place-names and the associated information for the Russian Federation, in scope of the national topographic map adopted as the basic one. It is aimed at collection, storage, updating and processing of the information on geographical names and at providing users with the required data and materials.

The national topographic map at scale 1:100,000 is supposed to be taken as the basic. Thus, the toponymic file created on the base of the map, will comprise 2,5-3,0 million names. It is impossible to prepare, inventory. record and handle this data volume with routine manual technologies.

Progress in computers and computer-based technologies allows to provide secure storage, processing and analysis of large arrays, and gives new possibilities to maintain toponymic databases, to respond to various users calls, to release different documents (gazetteers, dictionaries and other publications).

The *goal* to create the up-to-date National Catalogue of Geographical names could be achieved by developing the automated information system With the mentioned capabilities.

The development of the Automated Information System of Russia's National Catalogue of Geographical Names (AIS RNCGN) is to be started in 1998. Eventually, the Automated Information System (AIS) will be developed, that would be a totally new technology of collection, inventory, storage, processing, output and transmission of place-name information based on modern hardware and software achievements. The AIS introduction would allow to start development, maintenance and use of the National Catalogue of Geographical Names for the Russian Federation.

Accounting for laboriousness of collecting data from separate sources to fill the Catalogue, of much time required for their accumulation, and rapid progress in hardware and software, it is planned at the beginning to develop the first series of the AIS intended to automatically process 0.7-0.9 million place-names. This would allow to create the basis for the Catalogue at minima initial expenses. As practical experience is accumulated, new hard-software and technologies appear, the first series of the AIS will be modified, its power and functionality will be increased, thus achieving 2.5-3.0 million records.

Investigation of the experience in developing automated information systems on geographical names allowed to classify, generalize and formulate basic requirements to their implementation.

AIS RNCGN will provide the following functionality:
- database maintenance;
- data processing, displaying on the screen and printing the documents at users requests;
- transmission of the required data to the users of the system.

When designing AIS RNCGN, the following will be provided:
- comprehensive approach to AIS developing (with allowance for all correlations within the system together with outer environment interactions);
- use of home and foreign experience in developing automated information systems, optimal use of off-the-shelf design solutions and program packages, in the first turn DBMS;
- multi-field retrieval and data processing based on the set-up quantitative and qualitative criteria;
- quick handling and on-line output of the required data to the users;
- flexibility of the system, that is, the capability to quickly customize it to would-be changes at minima costs.

When developing AIS RNCGN, its functionality structure (functional subsystems and tasks, their correlations) will be determined, and the executive portion of the system (information, mathematical support, soft- hardware) developed.

The database on geographical names *of* RNCGN under development, will serve as the basis for AIS performance This database meets the following requirements
- a single data input and their comprehensive usage;
- a handy and quick user s access to the data stored in the database;
- secure and long-time data storage;
- possibility to make simple and on-line changes in the database;
- database flexibility;
- data verification control;
- possibility to extract and spray various data from the database

AIS RNCGN will be implemented in accordance with current state standards for the development of automated information systerns and software documentation.

AIS and the Catalogue based on it, are intended to satisfy a wide range of information needs of state administrations, departments, agencies, state and private mapping publishers, scientific and information organisations, publishing houses and editors. mass media and individuals.
AIS performance will allow:
- to solve the tasks which could not be solved or are extremely laborious if traditional methods are used;
- to reduce labouriosness in toponymic databases maintennnce, and routine clerical work of experts engaged in it;
- to considerably increase database access speed as well as rapidity of responses to users calls;
- to provide multi-field retrieval and data processing in accordance with the set-up qualitative and quantitative criteria;
- to increase data storage security, accounting for small media size and their high resistance to outer interventions, and simplicity of making the required number of back-ups;
- to provide for unification and validity of the output data (surely, if the database is completely verified);
- to obtain the data by graphic (and combined) queries, including outputting map portions;
- to provide the spray of various data from the databases and their quick transmission to users via communication channels;
- to considerably speed up catalogue compilation, publishing of gazetteers, dictionaries and other information materials;
- to increase the efficiency of map makers work by providing them with timely and qualitative information on geographical names and capability to plot the data with a computer;
- to provide the capability to obtain handy information to carry out toponyrnic and other researches.

Information support of toponymic automated systems is a set of indices, codes, documents and arrays as well as of the methods for their arranging, storage and verification.

In course of developing AIS RNCGN information support, the following work should be implemented
- to collect and classify place-names data;
- to determine the set of the appropriate codes intended for systematized and compact data representation in an electronic form;
- to determine the database elements set and structure of their layout on data media;
- to transfer data to electronic media:
- to develop the technology of data storage within the database nnd of making the required changes;
- to determine a list of typical queries to the system;
- to determine the contents and structure of output documents and the requirements to their formation;
- to develop data verification methods and ways of data transmission to users;
- to determine information parameters of the system.

In the database, the entire information on geographical names will be stored as records comprising the name itself and the associated data. It was preliminary decided that the database of the first series of AIS will involve nearly 30 fields.

Regardless of a great diversity of the data included into the database of RNCGN, they will be split into the following groups:

1. Name (proper name of a geographical feature, generic term, name variants, reference to other names).
2. Feature location (geographical co-ordinates, rectangular co-ordinates, administrative unit, geographical reference).
3. Sources, a name is taken from, any other information (topographical map sheet number, bibliography);
4. Standardized name.
5. Historical informa'tion (name origin, its changes).
6. Additional information on a feature (height above the sea, popuiation, extension, etc.)
7. Language, structure, pronunciation (stress), peculiarities in declination. semantic information with source reference where possible.
8. Subsidiary codes (number- and various codes).

The software of AIS will be the set of programming means, providing its performance. When choosing the software, the following will be taken into consideration: wide spread in the world, openness of data formats, initial multiplatform orientation extending with each version, functional completeness.

Development of Russia's National Catalogue of Geographical Names

Summary

Collection and cataloguing of geographical names is the main pledge of their storage and unified usage. The report provides the status of this issue in Russia, and substantiates the need to develop the automated National Catalogue of Russia's geographical names.

The National Catalogue of Geographical Names should be constantly updated reference file of place-names and the associated information for the Russian Federation, in scope of the national topographic map adopted as the basic one. It is aimed at collection, storage, updating and processing of the information on geographical names, and at providing users with the required data and materials.

Selection of the state topographic map at scale 1:100,000 as the basic one for the Catalogue, is substantiated. 'I'he report seeks to prove that this goal could be achieved based on modern progress of computers and computer-based technologies, which allow to provide secure storage, processing and analysis of large data arrays.

The report describes the preliminary concept of developing the automated Catalogue of Russia's Geographical Names and the ways to implement it under the present economic circumstances.