Economic and Social Council

Distr.: General

21 May 2012

Original: English

**Tenth United Nations Conference on the
Standardization of Geographical Names**
New York, 31 July – 9 August 2012
Item 11(e) of the provisional agenda*
Toponymic data files and gazetteers
**Data services, applications and products (for example, gazetteers and
web services).**

Toponymic databases and dictionaries in Estonia

Submitted by Estonia**

Tenth United Nations Conference on the
Standardization of Geographical Names
New York, 31 July – 9 August 2012
Item 11 of the Provisional Agenda

Toponymic databases and dictionaries in Estonia[1]

The report describes recent linguistic digital resources in Estonia covering geographical names. These resources are compiled mainly at the Institute of the Estonian Language and the Võru Institute, often with the help of onomasticians from other institutions.

The Place Names Database of the Institute of the Estonian Language (KNAB, *http://www.eki.ee/knab/knab.htm*, texts in Estonian and English) was started already at the end of the 1980's, in 1994 it was restarted on personal computers. Its main aim is to provide help in the identification and linguistic standardization of place names, both those of inside and outside Estonia. The current size of the database is ca. 141,000 records with more than 400,000 different name variants. The size of Estonia's geographical names database is ca. 44,000 records (with 111,000 different names), that of foreign countries contains ca. 97,000 records (with ca. 298,000 names). Each record contains the principal name, variant names, and data on the named feature: feature designation code, present administrative division, (historical) parish, minor area (e.g., a village or island), and geographical coordinates. Each name has also a language label + optional labels, describing the name variant's stylistic or functional value, in some cases also pronunciation. Name sources (maps, dictionaries, other databases) are also indicated.

The database on Estonia's place names may also contain data on earlier administrative divisions, data on the use of locative cases (either internal or external cases are associated in Estonian for names of populated places), etc. Of the features included, the names of populated places have the greatest coverage, with name variants from sources throughout the 20th century. Other covered features include streets (excluding the very recent ones that are in the National Place Names Register), manor houses, and natural features (rivers, lakes, islands, capes, mires). Some areas (Petserimaa, parishes in northern Estonia) are covered in even more details, including farm names in historical sources.

The database of foreign geographical names was originally compiled as part of a dictionary project that aimed at giving a list of names used in Estonian texts, both endonyms and exonyms (the dictionary was published in 1999). Additionally, names used in English, Finnish, French, German and Russian were included. Later, the database was extended to include name variants in other languages as well, particular attention was paid to multilingual areas, and those close to Estonia. Over the time, as the encoding of names became to be based on Unicode, it was possible to include various original-script forms of names, from Arabic to Chinese.

The Internet version of the database contains less data than the original database but perhaps future versions will make it possible to include most of the data in the database for public usage. (The extended version is already used in the database of language names and ethnonyms, *http://www.eki.ee/knab/linguae.htm*.) Queries in the database may be executed in Estonian, English and Russian. There is some difference between the query results, depending on the language of query. Whereas with the Estonian query the main name forms will be those used in Estonian texts, including Estonian exonyms, in the case of English and Russian query the main name form will, if possible, be the endonym.

---

[1] Prepared by Peeter Päll (Estonia).

Although the status of the database is unofficial, people use it extensively as one of the main sources, especially for foreign names. Several media organizations and e.g. EU translators regularly check their spellings through this database. More information (in English) on the database is available on the webpage indicated above.

The Place Names Database of Historical Võrumaa (AVKA, *http://www.wi.ee/index.php/ projects/linguistic-research/geographical-names,* texts in Estonian, Võru and partly English) is a project undertaken by the Võru Institute to safeguard place names in the south-western region of Võrumaa, a historic county which is larger than the present county. Particular attention is paid to the name forms in the local usage. The size of the database is ca. 24,000 features, covering in detail populated places, farms, and natural features of Võrumaa. Both KNAB and AVKA have been linked to the map interface of the Estonian Land Board, so that all features with geographical coordinates may be viewed as a layer on the Land Board's web map server (*http://geoportaal.maaamet.ee/eng/*).

The Database of the Card Index of Place Names of the Institute of the Estonian Language (*http://heli.eki.ee/kohanimed/*, texts in Estonian) is the most recent of digitized resources that has become available to the wide public, although it has not yet been officially opened. The card idex contains toponyms collected by linguists in fieldwork since the 1920's. Data on a typical card include the name's local pronunciation, indication on the usage of locative cases, description of the locality and historical or etymological background. During 2011, in seven months the 540,615 cards were scanned and organized into a database that can be viewed online. At the present stage the cards are displayed only as images, one can browse the cards parish by parish, alphabetically (to facilitate the search, cards have been provided with indicative initial letters). Work is already underway to provide each card with a searchable name form and variant names + data on the feature type, this stage will last for perhaps more than a year.

The digitization of the card index database was performed as part of the **Estonian Place Names Dictionary** project. The project started in 2009 and is due to end in 2013. The dictionary will be a one-volume reference work, intended for wide public. Previously there have been no such comprehensive works on Estonian toponyms, except regional monographs (Valdek Pall – toponyms of northern Tartumaa, 1969–1977; Marja Kallasmaa, toponyms of Saaremaa, 1996–2000, Hiiumaa 2010; Evar Saar, toponyms of Võrumaa, 2008). The only study covering nominally the toponyms of Estonia is dated 1955 (Lauri Kettunen, *Etymologische Untersuchung über estnische Ortsnamen*), some other books on Estonian toponyms in German include Paul Johansen's *Die Estlandliste des Liber Census Daniae* (1933) and *Baltisches historisches Ortslexikon*, T. I (1985). Inevitably, these sources need updating.

The dictionary aims at giving as much useful information as possible: not only possible etymologies with historical name forms, but also the names' pronunciation (prosody, palatalization), local name forms, name variants in other languages, use of locative case forms, administrative data, etc, will be provided. To increase interest towards names, also local tales connected with the place name might be added. The dictionary will contain a little more than 6,000 articles including the names of all official populated places, a selection of other types of settlements, names of administrative units and names of most important natural features.

The project is jointly undertaken by the Institute of the Estonian Language and the Võru Institute in association with other institutions, such as the National Archives, the Estonian Literary Museum, the University of Tartu, the Place Names Board, etc. Individual onomasticians may contribute to the work by compiling articles on specific toponyms. Each name article will be attributed to one or several authors. The dictionary project is funded by the Ministry of Education and Science.

The dictionary is compiled online, using the software developed at the Institute of the Estonian Language for compiling dictionaries. The application has a multi-user interface, i.e. articles may be compiled simultaneously, and every participant can see the whole dictionary text. The dictionary articles are preformatted – all the main structural elements are 'built in' in the application (including standard lists of variables), so that any compiled articles will have the same structure. The list of sources will also be compiled automatically. To avoid unauthorized use, the access is restricted by username and password, all changes are logged.

Currently, half of the articles have been compiled, so the project is behind the schedule. However, as more useful online sources become available, it is hoped that the main part of the project will be completed in time. Among the online resources, one should also mention the **Bibliography on Estonian Onomastics** (*http://www.eki.ee/nimeselts/bib/*, texts in Estonian, title page also in English) which is an annotated bibliography on general onomastics, personal and place names, other names, and applied onomastics (incl. standardization). The bibliography is searchable but it also includes listings arranged alphabetically (by authors), chronologically and by topics. Earliest bibliographical records are from the 18th century but the main part is from the 20th century. The size of the database is ca. 4,400 records (includes monographs, articles, newspaper articles, papers held at various conferences, etc.).

Summary

The paper addresses various recent linguistic digital resources on geographical names of Estonia. These are mainly compiled by the Institute of the Estonian Language, also in the Institute of Võru. The Place Names Database KNAB contains 141,000 records with more than 400,000 names, its main aim is to help in identifying and linguistically standardizing both Estonian and foreign geographical names. The database is not official but is extensively used by media. The Place Names Database of Historical Võrumaa AVKA contains ca. 24,000 records from the south-eastern county in Estonia, its main aim is to indicate names in local usage. In 2011 the 540,000 cards of the card index of toponyms in the Institute of the Estonian Language were scanned and put online, at present it is possible to browse the images but adding textual data is underway. The Estonian Place Names Dictionary project started in 2009 and is scheduled to end in 2013. It will be a one-volume reference work for wide public, containing ca. 6,000 articles on populated places, natural features, etc. Names' etymology, pronunciation, use of locative case forms, name variants in historical sources and in other languages will be given. The dictionary is compiled using an online utility used for compiling dictionaries at the Institute of the Estonian Language. The Bibliography on Estonian Onomastics contains 4,400 records.