
**UNITED NATIONS
GROUP OF EXPERTS ON
GEOGRAPHICAL NAMES**

Working Paper

No. 15

Twenty-fourth Session
New York, 20 and 31 August 2007

Item 3 of the provisional agenda
Report of the working groups

**Report of the Working Group on Toponymic Data Files
and Gazetteers for the period 2006-2007**

**Report of the Working Group on Toponymic Data Files and Gazetteers
for the period 2006-2007**

Working Group Membership

The following experts are members of, or have participated in, the activities of the Working Group:

Adela Alcázar	Teemu Leskinen
Johnny Andersen	Curt Loy
Margarite Azcárate	Jorma Marttinen
Giedre Beconyte	Juriј Mlinar
Bernd Beinstein	Blaz Mozetic
Herman Bell	Janis Nagelis
Pavel Bohač	Kathleen O'Brien
Caroline Burgess	Peeter Päll
Zane Cekula	Trent C. Palmer
Yaives Ferland	Ryan Pribadi
Randall Flynn	Agus Prijanto
Uldis Gitendorfs	Jörn Sievers
Isolde Hausner	Mara Šterna
Lan Jianxiu	Abdūlvahit Torun
Betsy Kanalley	Mimi Urbanc
Helen Kerfoot	Vita Strautniece
Mária Kováčová	Lou Yost
Glen Lauber	Pier-Giorgio Zaccheddu

Terms of Reference

The Working Group was constituted in its present form during the nineteenth session of the United Nations Group of Experts.

The Working Group endeavors to promote and further the aims of the United Nations Group of Experts on Geographical Names that relate to the collection and dissemination of geographic names information as stated in the first report of the Group of Experts in 1960:

"It is recommended that names authorities publish standardized names in gazetteer form as well as on maps since much information necessary for the proper understanding and interpretation of names cannot be included readily on maps."

This seminal recommendation has been underscored and amplified in numerous resolutions of the various United Nations Conferences on the Standardization of Geographical Names (UNCSGN).

Work Plan

At the 19th session of UNGEEN in New York, January 1998, the Working Group established the following outline of objectives (with modifications recommended at 8th UNCSGN in 2002):

1. Establish and maintain a Working Group web site to document and distribute news of events and developments relating to gazetteers and toponymic information exchange.
2. Maintain and revise the information contained in Report of the Working Group on Toponymic Data Exchange Formats and Standards to the Seventh United Nations Conference on the Standardization of Geographical Names.

3. Establish liaison with the Unicode Consortium to represent the needs of the geographic information community for digital text encoding in the context of geographic names.

4. Establish contact with relevant activities under the auspices of the International Organization for Standardization (ISO) to influence the development of international standards for the exchange of toponymic information.

5. Promote the implementation of practical programs of toponymic data exchange.

6. Promote and conduct toponymic data exchange workshops to address issues associated with data exchange

Working Group Activities since 23rd UNGEGN:

The Working Group met twice during the 23rd UNGEGN session in Vienna and discussed the following topics:

- The feasibility of establishing a standard "feature list" or thesaurus for use in compiling toponymic datasets.
- Interaction with Google Earth to promote more widespread use of national standard toponymic datasets.
- Strategies for improved liaison with the Unicode Consortium and with Technical Committee 211 of the International Organization for Standardization.
- The feasibility of using low-technology devices to store and record toponymic and other geographic information in the field in support of humanitarian assistance planning and disaster relief.
- The possibility that WG members could compile a handbook on tools, education, and data modeling in support of developing toponymic datasets and gazetteer services. An effective start to producing such a handbook would be compiling a collection of the more meaningful working papers on related topics that have been submitted to prior UNGEGN sessions and UN conferences.

In March 2007 the Working Group met a third time in Madrid, in conjunction with the meeting of the EuroGeoNames Reference Group. The following topics and recommendations were discussed:

- **WG web site:** A web site was initially established by Mr. Milan Orožen Adamič and hosted on a local Slovenian server. Recommendation: the web site should be re-hosted on a server where active web site maintenance can take place. Subsequent to the March 2007 meeting, Ms. Mimi Urbanc of Slovenia volunteered to redesign and maintain the WG web site, currently found at <http://www.zrc-sazu.si/ungegn/gazetteers.htm>.
- **Report of the Working Group on Toponymic Data Exchange Formats and Standards:** Updates to the Report have been made by Mrs. Burgess and Mr. Päll in connection with the compilation of the UNGEGN *Technical Reference Manual for the Standardization of Geographical Names* that was officially published by the United Nations in July 2007.
- **Liaison with Unicode Consortium:** In 2006 there was contact between the Consortium's Dr. Mark Davis and the UNGEGN WG on

Romanization Systems regarding application of UN recommended systems of Romanization in automated localization tools in work by Consortium members. No further information exchange has occurred. Recommendations: The WG should renew contact with the Consortium, particularly with regard to any localization and internationalization issues relating to processing & exchange of digital toponymic data.

- Liaison with International Organization for Standardization (ISO): UNGEGN established liaison status with ISO TC 211 in 2002. There has been no practical liaison activity to date between the WG and ISO TC 211, though good intentions are present on both sides. The WG convenor remains on the ISO TC 211 mailing list. Recommendations: The WG should endeavor to prepare a summary liaison report of the WG's activity & interests relative to toponymic standards that touch on the work of ISO TC 211. The liaison report would be submitted to a subsequent TC 211 plenary meeting. The WG should also investigate the current status of ISO standards development relating to toponymy. The WG should track & review ISO TC 211 information mailings, meeting notifications, and documents, and identify standards projects & activities relating to the WG's terms of reference.
- Practical programs of toponymic data exchange: With the exception of EuroGeoNames, programs of toponymic data exchange appear to be ad hoc and event-driven. Nonetheless, there is a strong perception in certain communities (e.g., humanitarian assistance planning) that a more structured cohesive program of directed toponymic collection and exchange would bring significant benefit and efficiencies to a given community of interest. Recommendation: The WG should identify the humanitarian assistance community of interest as the target community with which the WG will endeavor to facilitate a practical program of toponymic data collection & exchange. The WG should initiate coordination with relevant units in the UN Office for the Co-ordination of Humanitarian Affairs and (possibly) the Office of the UN High Commissioner for Refugees. This coordination will cover definition of toponymic requirements and standards for toponymic data collection, exchange and databases.
- Toponymic data exchange workshops: It is logical and practical to associate this element of the work plan with the previous element, linking data exchange workshops with the practical program. Recommendation: An element of the program undertaken in collaboration with the humanitarian assistance community of interest will be a data exchange workshop. Additionally, the WG discussed the possibility of conducting a brief workshop at the upcoming 9th UNCSGN. One objective of the workshop would be the review and updating of the draft toponymic data exchange standard first developed by WG members in 1997 (a copy of the latest version of this draft standard is attached to this report).

- Future of gazetteers: from publication to web services (1315-1345): Discussion led by Mr. Ferland covered the following topics:
 - o Real purposes of gazetteers;
 - o Revised, extended ontologies of generic features;
 - o Semantic & syntactic expression of "sense of place" (spatial/sensoral & linguistic);
 - o Adaptative geometries for geographical entities;
 - o "Gazetteer" as a knowledge organization system;
 - o Continuing role of "cartography".

The WG subsequently agreed to form an ad hoc subgroup to further the discussion of the future of gazetteers. It is anticipated that this sub-group will meet during the 9th UNCSGN in New York.

Future of the Working Group

The sheer number of Working Group members and associates demonstrates the high level of interest among UN experts in the concepts of toponymic data files and gazetteers. The discussions held at the three meetings of the WG since UNGEGN's 23rd session are indicative of the range of challenging tasks involving information technology, standards development, and international cooperation and collaboration to which the Working Group could meaningfully contribute. It is hoped that the Working Group can continue to find effective ways to work collaboratively to make progress in the advancement of toponymic data files and gazetteers as a means of effective communication of geographical knowledge.

Draft toponymic data exchange standard

Part 1 Data elements

The following data elements, identified in Resolution 4 of the first United Nations Conference on the Standardization of Geographical Names, comprise the minimum set of critical geographic names information for digital exchange.

1.0 GEOGRAPHIC NAME– the spelling of a standardized name referring to a feature. Spellings of variant names, if any, follow the standardized name with suitable delimiters. An explanatory note accompanying the data set must define all conventions used in the presentation of GEOGRAPHIC NAME; e.g., sorting order, reversal of generic terms, meaning of special flags such as asterisks, etc.

Data Type: text. An explanatory note accompanying the data set must define the text format and encoding; e.g.,
char, ASCII
char, ISO 8859 (2)
char, KOI 8
wchar_t, ISO 10646

2.0 FEATURE TYPE – a characterization of the kind of feature represented by GEOGRAPHIC NAME. An explanatory note accompanying the data set must define the characterization scheme employed.

Data Type: text

3.0 LOCATION – the position associated with GEOGRAPHIC NAME or FEATURE TYPE expressed in latitude and longitude (based on the Prime Meridian, Greenwich). For most toponymic data sets, the preferred format would be positions specified by degrees and minutes, rounded down to the nearest integer minute. This format allows future update to include seconds or decimal minutes. An explanatory note accompanying the data set must identify alternate formats if applicable; e.g., decimal degrees, integer degrees and minutes, integer degrees and decimal minutes, etc.

3.1 LATITUDE – the latitude value of LOCATION.

Data Type: integer
Domain: 90 00S <= LATITUDE <= 90 00N

3.2 LONGITUDE – the longitude value of LOCATION.

Data Type: integer
Domain: 180 00E <= LONGITUDE <= 180 00W

4.0 ADMINISTRATIVE AREA – an identification of the country and administrative division in which LOCATION falls.

Data Type: text
Domain: (the domain is the content of ISO/DIS 3166, Parts 1 and 2)

5.0 MAP SHEET REFERENCE – an identification of the sheet (within a standard national map series) containing LOCATION. An explanatory note accompanying the data set must provide citation for the referenced map series.

Data Type: text

Part 2 Metadata¹

The notes to Part 1 of this annex refer to additional information required by the user of a data set to interpret competently the data contained therein. This additional information is generally referred to as *metadata*.

Metadata serves to answer four questions:

- What sets of data exist for a geographic location?
- Does a set of data meet a specific need?
- How is a set of data acquired?
- What information is required to process and use a set of data?

Metadata characterizing geographic names data sets would typically comprise the following subsets:

- Identification information – basic information about the data, including publisher.
- Text encoding standard – identification of the national, international, or proprietary standard used to represent the text digitally in the data set.
- Data quality information – a general assessment of the quality (accuracy, currency) of the data set.
- Spatial data organization (if applicable) – the mechanism (text, point, vector, raster) used to represent spatial data in the data set.
- Spatial reference information (if applicable) – the description of the reference frame for coordinates in the data set.
- Entity and attribute information – the description of the content of a data set including entity types, their attributes and domains (Part 1 of this Annex qualifies as entity and attribute metadata).
- Software information – identification of the software configuration used to create the data set, including identification of operating system and application.
- Distribution information – information about the distribution of and options for obtaining the data set.

Part 15 of ISO 15046 is a draft standard for metadata pertaining to geographic information. It is based largely on the Metadata Standard developed by the United States Federal Geographic Data Committee (FGDC). While it was beyond the scope of the UNGEGN Working Group on Toponymic Data Exchange Formats and Standards to define rigorously a metadata standard specific to toponymic data, it is appropriate and necessary to emphasize the importance of describing the content of geographic names data sets in sufficient detail to ensure their utility in a program of exchange of digital information.

Part 3 Illustration

The following listing is an abbreviated example of a print-out of a digital toponymic data set adhering to the guidelines recommended by the Working Group. Two files are represented; the first is the toponymic data, the second is a text file (conventionally known as a “readme” file) containing the metadata necessary for a recipient to determine the data’s content and utility. These examples are provided for the sole purpose of illustration.

Data.txt

Aakre Peatus RSTP 5804N 2616E EE 19 NO35-05

¹ The information presented here is summarized from Content Standard for Digital Geospatial Metadata, Federal Geographic Data Committee (USA), April, 1997

Aaksi Saar → Aksi	ISL	5935N	2505E	EE	01	NO35-01
Aamse	PPLX	5855N	2342E	EE	07	NO34-06
Aamsi → Aamse	PPLX	5855N	2342E	EE	07	NO34-06
Aandu	PPL	5909N	2443E	EE	13	NO35-01
Aardla	PPL	5818N	2646E	EE	18	NO35-05

Readme.txt

This file describes the content, lineage, and utility of the toponymic data contained in the file data.txt.

Data Identification: Place name spellings for features in Estonia.

Text encoding standard: Eight-bit ASCII (ISO 8859 Latin 1). Two characters (Roman small letter c with caron; Roman small letter i with macron) are encoded in a proprietary scheme, and may be viewed using the Monotype Special G1 family of fonts.

Data quality: Name spellings are current for most features as of 1993, and have been taken from the following authoritative Estonian sources:

Eesti Topograafiline Kaart, 1:200,000, 1993.

Eesti Ülevaatekaart, 1:400,000, 1993.

Eesti Sood, 1:400,000, 1993.

Eesti Veed, 1:400,000, 1991.

Settlement names were taken from a list provided by Eesti Keele Instituut, Tallinn, in 1994.

Data organization: Data is provided alphabetically by name spelling. Data elements within a record are tab delimited (ASCII character 09). Records are delimited by carriage return (ASCII character 0D) Single geographic coordinates are provided for each record.

Data reference: Coordinates are referenced to the World Geodetic System 1984.

Entity and attribute information: Each record contains the following data elements and associated attributes:

Geographic name. Name spellings conform to standard Estonian orthography with the exception that initial letters of generic terms are capitalized. Variant names are cross referenced to standard names by use of the greater than sign (ASCII character 3E).

Feature type. A five character designation denoting the type of feature to which the name refers. The designation scheme is that employed by the U.S. Board on Geographic Names in its data base of foreign geographic names. This data element is left justified, and space filled (ASCII character 20) in cases where the designation is fewer than five characters.

Latitude. The geographic latitude of the named feature, taken at the centre of spot and areal features, and at one end of linear features. The data element is five characters in length; the first two represent integer degrees, the second two integer minutes (rounded down), and the fifth the compass point N.

Longitude. The geographic longitude of the named feature, taken at the centre of spot and areal features, and at one end of linear features. The data element is five characters in length; the first two represent integer degrees, the second two integer minutes (rounded down), and the fifth the compass point E.

Country code. The country code digraph taken from ISO 3166 “Codes for the representation of names of countries and their subdivisions.”

Administrative division code. The two character administrative division code taken from ISO 3166 “Codes for the representation of names of countries and their subdivisions.” The administrative division identified for each name is the division in which the geographic location for the record falls. Features that cross administrative or international boundaries are given the general code 00.

Map sheet reference. The map sheet number of the standard 1:250,000 map series on which the geographic coordinate of the record falls. The feature and its name may not necessarily actually appear on the referenced sheet.

Software requirements. This data set was compiled using Microsoft Word 7.0 for Windows 95. The only known special requirement is the use of the Monotype Special G1 font for display of certain characters.

Distribution information. There are no restrictions on the distribution and use of the information contained in these data sets. Additional copies of this data set may be obtained from the producer.

Producer point of contact. This data set was produced by the staff of the Foreign Names Committee of the U.S. Board on Geographic Names. Questions, corrections, and requests for additional information may be referred to:

Executive Secretary for Foreign Names
U.S. Board on Geographic Names
National Geospatial-Intelligence Agency
Political Geography Division (PRP)
4600 Sangamore Road
Bethesda, Maryland 20816-5003
USA