

19 July 2002

Original: English

Eighth United Nations Conference on the
Standardization of Geographical Names
Berlin, 27 August-5 September 2002
Item 12 of the provisional agenda*

TOPONYMIC DATA FILES

Report of the Working Group on Toponymic Data
Files and Gazetteers for the period 1998-2002
(Submitted by the United States)**

*E/CONF.94/1

** Prepared by Randall Flynn, Executive Secretary for Foreign Names, U.S. Board on Geographic Names, National Imagery and Mapping Agency, submitted by the United Nations Group of Experts on Geographical Names - Working Group on Toponymic Data Files and Gazetteers.

Working Group on Toponymic Data Files and Gazetteers
Eighth United Nations Conference on the Standardization of Geographical Names
Berlin. 27 August - 5 September 2002

Working Group Membership

The following experts are members of the Working Group

Mr. Milan Orožen ADAMIČ, Slovenia
Mrs. Caroline BURGESS, United Kingdom
Mr. Traian CONSTANTINESCU, Romania
Mr. Randall FLYNN, United States (convenor)
Ms. Mária KOVÁČOVÁ, Slovakia
Mr. Teemu LESKINEN, Finland
Mr. Hamid MALMIRIAN, Iran
Mr. Roger MARSDEN, United Kingdom
Mr. Jorma MARTTINEN, Finland
Mrs. Kathleen O'BRIEN, Canada
Mr. Peeter PÄLL, Estonia
Mr. Juan QUIÑE Vigil, Peru
Mr. Jörn SIEVERS, Germany
Ms. Vita STRAUTNIECE, Latvia
Mr. Lou YOST, United States

Terms of Reference

The Working Group was constituted in its present form during the nineteenth session of the United Nations Group of Experts with the subsuming of the Working Group on Toponymic Data Exchange Formats and Standards.

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 UNGEGN in New York, January 1998, the Working Group established the following outline of objectives:

1. Establish a Working Group newsletter 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

Progress Against the Work Plan Since the 7th UNCISG:

The following points summarize activities occurring during 1998-2002:

Objective 1 – Working Group Newsletter

No activity. However, the Working Group has established a web site <<http://www.zrc-sazu.si/ungegn/>> that can effectively serve as the means for publicizing the Group's activities, meetings, and reports. It is suggested that this objective should be restated as follows:

- Establish and maintain a Working Group web site to document and distribute news of events and developments relating to gazetteers and toponymic information exchange.

The Working Group expresses its appreciation to Mr. ADAMIČ for his work in establishing the web site.

Objective 2 – Maintenance of the Report of the Working Group on Toponymic Data Exchange Formats and Standards

A revision to the report has been suggested regarding the character inventories for Sámi, recommending that three extensions are required - for Northern Sámi, Inari Sámi, and Skolt Sámi - with more complete character inventories.

In addition to this suggestion, the Working Group must address two additional points in contemplating a major revision of the report

- The report was composed in 1997 using several software applications. In its present state, it cannot be reliably distributed in electronic form. A recompilation of the report is required, using a single Unicode-enabled software application, in versions that support posting at web sites and electronic distribution.
- The working party that compiled the report relied on the Unicode Standard, Version 2, for its evaluation of the Unicode standard in satisfying Roman alphabet place-name spelling requirements. The Unicode Consortium has now published Version 3.2.0 of the standard. Changes since Version 2 include the

representation of a number of characters that affect the *content* of the report. A complete review and update of Annex A and Annex B of the report in comparison with the Version 3.2.0 of the Unicode Standard is required.

It is suggested that the working party that first compiled the report in 1997 regroup to formulate a recommended strategy and schedule for completing these tasks.

Objective 3 – Liaison with the Unicode Consortium

A paper and presentation covering the special needs for text encoding in digital geographic information systems to represent correct spellings of geographic names was prepared on behalf of the Working Group. The paper was presented at the 13th International Unicode Conference in San Jose, California, September 8 – 11, 1998, and subsequently published in the proceedings of the conference. Copies are available from the Working Group convenor.

In June 2002 the Working Group collaborated with UNGEGN leadership and the UN Secretariat to establish formal liaison membership for UNGEGN with the Unicode Consortium. The Consortium consented to UNGEGN's request on 19 June 2002; UNGEGN is now listed as a liaison member at the Consortium's web site <<http://www.unicode.org/unicode/consortium/memblogo.html>>.

The Consortium describes liaison membership as follows:

Liaison Membership in the Unicode Consortium is offered (at no charge) to organizations that have interests in common with the Consortium, particularly in technical areas. The Unicode Consortium and the organization with which it has a liaison relationship each nominate a liaison representative. Liaison representatives have primary responsibility for the interchange of information between the two parties. In particular, liaison representatives are charged with alerting the other member of the liaison relationship to events materially affecting that organization's area of interest or expertise.

As a liaison member, the UNGEGN representative may participate in meetings of the Unicode Technical Committee, and is granted access to UTC working documents and the UTC private e-mail list.

In achieving UNGEGN liaison membership with Unicode, the Working Group has materially assisted in fulfilling the recommendations in Resolution 6 of the 7th UNCSGN.

Objective 4 – Liaison with ISO

In the summer and autumn of 2001, the Working Group held a series of discussions with the Outreach Committee of the ISO Technical Committee 211 - Geographic Information/Geomatics (ISO/TC 211). The discussions centered on an approach that would inject the toponymic expertise of UNGEGN into the data standards development process of ISO/TC 211. Following these discussions, in the spring of 2002, the Working Group and UNGEGN leadership corresponded with the ISO/TC 211 Secretariat.

culminating in a request from UNGEGN to ISO/TC 211 seeking category A liaison status. The Technical Committee accepted the request at its May 2002 plenary meeting in Bangkok.

ISO/TC 211 defines its scope as:

- Standardization in the field of digital geographic information
- This work aims to establish a structured set of standards for information concerning objects or phenomena that are directly or indirectly associated with a location relative to the Earth
- These standards may specify, for geographic information, methods, tools and services for data management (including definition and description), acquiring, processing, analyzing, accessing, presenting and transferring such data in digitalelectronic form between different users, systems and locations.
- The work shall link to appropriate standards for information technology and data where possible, and provide a framework for the development of sector-specific applications using geographic data.

In achieving UNGEGN's liaison status with ISO/TC 211 the Working Group has assisted the Group of Experts in making significant progress to fulfill the recommendations in Resolution 13 of the 7th UNCSGN.

In 2002 the Working Group actively participated in the work of ISO/TC 211 by conducting a review of the ISO Committee Draft Standard 19112 - "Spatial Referencing by Geographic Identifiers." Copies of the draft standard were distributed to Working Group members, and a consolidated group commentary was provided to the ISO/TC 211 Secretariat in March 2002.

Objective 5 – Promote practical programs of data exchange

The report of the UNGEGN Working Group on Toponymic Data Exchange Formats and Standards has been applied in building new data sources on the Internet, e.g. there is a Letter Database <<http://www.eki.ee/letter/>> that displays languages, their special characters and Unicode codes together with images of the letters. It is maintained at the Institute of Estonian Language. The report of the WG is mentioned as one of the sources.

Objective 6 – Toponymic data exchange workshops

Under the auspices of the Alexandria Digital Library Project at the University of California, and funded by the National Science Foundation (United States), a two-day Digital Gazetteer Information Exchange Workshop was conducted at the Smithsonian Institution, Washington, DC, on October 13 – 14, 1998. Several members of the Working Group participated in the workshop, along with several other United Nations experts. Approximately 65 people attended the workshop representing the academic community, governmental and non-governmental organizations, and the private sector. Representatives from Australia, Canada, Honduras, and the United Kingdom attended; limitations in available space and funding prevented further international participation. The workshop provided an invaluable opportunity for participants to share issues, concerns, and experience in both the practical and theoretical aspects of toponymic data

exchange. An expanded agenda for the workshop, including many of the presentations and concluding remarks, can be found at http://www.alexandria.ucsb.edu/~lhill/dgie/DGIE_agenda5.html.

A joint international meeting on geographical names was held at the Surveying and Mapping Authority of the Republic of Slovenia on 18-20 April 2001. The meeting was organised by the UNGEGN East Central and South-East Europe Regional Division in collaboration with the UNGEGN Working Group on Toponymic Data Files and Gazetteers. The central theme of the joint meeting on 18 April 2001 was the development of toponymic databases, digital gazetteers, and the exchange of digital toponymic information. Representatives from Cyprus, the Czech Republic, Finland, Germany, Hungary, Italy, Latvia, Poland, Slovenia, the Slovak Republic, and the United States made presentations on topics relating to this theme, with much useful discussion and exchange of information. The meeting also touched on recent developments at the UNGEGN secretariat, preparations for the Eighth United Nations Conference on the Standardization of Geographical Names, and the latest issue of *Onoma*, the journal of the International Council of Onomastic Sciences. Participants agreed that the meeting of an UNGEGN Regional Division and an UNGEGN Working Group to address a theme of common interest was of significant benefit in furthering the aims of the Group of Experts, and it was hoped that future joint meetings could be arranged.

The Working Group expresses its appreciation to the Surveying and Mapping Authority of Slovenia and the Chairman of the East Central and South-East Europe Regional Division for organising the meeting and for their warm hospitality.

Meetings of the Working Group

The Working Group met briefly at the 20th session of UNGEGN in January 2000 to review membership status and the group's work plan.

The Working Group next met in April 2001, in conjunction with the joint session with the East Central and South-East Europe Regional Division described above. Mrs. BURGESS prepared the following report of the meeting.

1. The Working Group met in Ljubljana on 19 April 2001 to discuss developments in the production of digital toponymic data files and gazetteers and to review its work plan.
2. Participants in the Working Group meeting included:
 - Mr. Milan Orožen Adamič, Slovenia
 - Mrs. Caroline Burgess, United Kingdom
 - Mr. Randall Flynn, United States
 - Ms. Isolde Hausner, Austria
 - Ms. Helen Kerfoot, Canada
 - Ms. Mária Kováčová, Slovak Republic
 - Mr. Teemu Leskinen, Finland
 - Mr. Blaž Mozetič, Slovenia
 - Mr. Janis Nagelis, Latvia

Mr. Giovanni Orrù, Italy
Mr. Jörn Sievers, Germany
Ms. Mara Sterna, Latvia
Ms. Vita Strautniece, Latvia
Mr. Pier-Giorgio Zaccheddu, Germany

3. Initial discussions concentrated on a first draft of a web site^{*} where information on the Working Group and its activities would be posted, along with links to other appropriate web pages. Members would be encouraged to provide material on relevant subjects which could be included in .pdf format. All submissions would be retained and an archive built for reference purposes. Initial comments concerning structure and links were to be forwarded to the Convenor by mid-May 2001, and the launch of the web site would be announced in the June 2001 UNGEGN Newsletter.

4. One of the prime objectives of the Working Group was to make contact with other bodies who were dealing with similar issues. In particular the International Organization for Standardization (ISO) should be approached by the UNGEGN Secretariat to raise awareness of the existence of the Working Group. The Convenor would also attempt to establish how the Working Group could become involved in relevant Technical Committees of ISO, including ISO/TC 211 (Geographic Information/Geomatics). The possibility of a formal association with the encoding standards consortium, UNICODE, was also to be investigated. The United Nations had recently established a Geographic Information Working Group who had requested assistance from UNGEGN with regard to the creation of a UN global gazetteer and text encoding standards for the UN GIS. The Working Group on Toponymic Data Files and Gazetteers agreed to work with the UNGEGN Vice-Chair to address these issues. A further significant body involved in digital gazetteer production was the Center for Global Georeferencing Research (CGGR) at the University of California. The Environmental Systems Research Institute (ESRI) and the US National Geographic Society were both participants. CGGR's Alexandria Digital Library project was concerned with the production of a global gazetteer and CGGR had consequently been compiling a standard for gazetteer content. Working Group Members were encouraged to study the CGGR material and consider how contact should be made.

5. Workshops or technical exhibitions which could further raise awareness of the Working Group were also discussed. It was suggested that the commercial software and digital data processing companies which were to be exhibiting at the 8th UN Conference on the Standardisation of Geographical Names (UNCSGN) in summer 2002 be approached in advance of the Conference and made aware of the concerns facing the Working Group, with particular regard to the encoding of letter/diacritic combinations and special letters. The Convenor would draft some suggested wording to be approved by the Working Group and included in a letter of invitation to the companies concerned.

6. The Working Group discussed multilateral programs of digital data exchange, such as the Scientific Committee on Antarctic Research (SCAR) *Composite Gazetteer of Antarctica* and the Nordic database. Such projects highlighted the

* The web site may be viewed at <<http://www.zrc-sazu.si/ungegn>>

importance of compatibility at all levels and the need for a standard format for databases and gazetteers

7. The question of test encoding standardis was considered by the Working Group. Members emphasized the importance of retaining special letters and diacritical marks where used in geographical names. Unfortunately many database search engines were unable to deal with such characters. The Report of the Working Group on Toponymic Data Exchange Formats and Standards (E/CONF.91/CRP.11) to the 7th UNCSGN was revisited. The Members agreed that the content needed to be updated and the lists of languages, writing systems and characters be amended as appropriate. Members would contact the Convenor with their comments and the lists would be recompiled using current technology, with a view to posting the revised document on the Working Group web site.

8. A key outcome of the Working Group discussions was the identification of a clear requirement for practical guidelines for the creation of toponymic databases and gazetteers and for digital toponymic data exchange. The Working Group would compile such a document, based on Annex C of E/CONF 91/CRP 11, which would be presented to the 8th UNCSGN

9. The Working Group would next meet during the 8th UNCSGN

Future Business

A skeleton for toponymic data exchange guidelines was compiled as Annex C to E/Conf.91/CRP.11 (7th UNCSGN). As suggested at its meeting in Ljubljana, the Working Group will undertake a review of these guidelines with the intention of publishing an expanded version as a separate document for international use. The new document will include an introduction on the work of the UNCSGN, UNGEGN and its Working Group on Toponymic Data Files and Gazetteers, and will comment in more detail on the encoding and rendering of extended Roman-alphabet and non-Roman characters.