

**Twenty-eight session
28 April – 2 May 2014**

Item 9 of the Provisional Agenda

**Activities relating to the Working Group on
Toponymic Data Files and Gazetteers**

**Discussion Forum of the Working Group on
Toponymic Data Files and Gazetteers – a status report ***

* Prepared by Caroline Burgess (United Kingdom), Naima Friha (Tunisia), Teemu Leskinen (Finland), Annette Torensjö (Sweden), Laura Kostanski (Australia), Vita Strautniece (Latvia), Pier-Giorgio Zaccheddu (Germany) from the Working Group on Toponymic Data Files and Gazetteers of UNGEGN

Summary

As a result of collaboration between experts from the Working Group on Toponymic Data Files and Gazetteers (WG TDFG) of UNGEGN this document reports the achievements of the online discussion forum launched since 2012.

For the time being, the discussion forum has been utilized and investigations commenced. Some UNGEGN experts as well as other experts for the respective topics have been registered and started to comment and to extend the content of the Wiki “rooms” / pages respectively. Forum 1 of the wiki discussion pages focuses on the topic of *volunteered geographic information and crowd sourcing*. Forum 2 focused on the topic of *Definition and Content of Gazetteers* and forum 3 focused on the topic of *general feature types/categories*.

For each forum sub-sites have been created in order to structure the discussions. Each sub-site has a short introduction and some initial questions/topics to be discussed.

As the methods of accessing geographical information change and public interaction with such data becomes more dynamic, national mapping agencies and place-name organisations are looking to take advantage of the willingness of the local population to provide information on location and associated attributes (= volunteered geographic information).

Concerning gazetteers is obvious that the technical view on 'gazetteers' coming from spatial data infrastructures is different from the UNGEGN view on 'gazetteers'. Many 'gazetteers' only collect and publish official information- which is a problem for researchers and users who often want to find details of colloquial, historical or unofficial names.

For the development of global classifications for named features, amongst others, different natural, cultural and linguistic circumstances and topographical variations in different parts of the world have to be considered. Official bodies and experts involved in geographical names and spatial data collection, international and regional initiatives as well as academic and private parties should participate in the discussion on global classifications. Different kinds on end-user application needs have to be studied before setting a general global classification.

From the answers and descriptions posted it is apparent that amongst members there is a wide variety of data and technology used to create geographic names information.

1. Introduction

Within the 10th UNCSGN (conference and side events) three new tasks/issues were proposed and afterwards added to the “Scope of work” and to the “Work plan” of the Working Group on Toponymic Data Files and Gazetteers (WG TDFG) of UNGEGN:

During a lunchtime presentation about the topic “**volunteered geographic information (VGI)/crowd-sourcing**” it was recommended that investigations about VGI/crowd-sourcing shall be commenced within the WG TDFG. Within the UNGEGN session (item 9) the convenor emphasized that this can only be a starting point for some activity. After some initial technical work within the WG TDFG, at the next UNGEGN session 2014, the discussion of VGI and crowd-sourcing should be continued within the topics national policies and promotion of standardization.

During the first session of Technical Committee II (E/CONF.101/57 and Add.1) it was recommended by Australia that the WG TDFG may consider to commence investigations into defining appropriate **definitions for gazetteers and data types** as this is obviously an issue for the WG TDFG.

During a session of Technical Committee IV (E/CONF.101/56 and Add.1) it was recommended by Australia that the issue of developing **general feature types/categories** could be an item to be discussed further in the WG TDFG too. It is worth noting that different feature type classifications increase the complexity of any interpretation or searching process. Interpreting these references is a significant and pressing need for better using information to support decisions, and gazetteers were considered to be the source of the information needed to interpret these references. Obviously this topic relates with the agenda of the WG TDFG.

Regarding the methods to be used to start actions, the set up of an **online discussion forum within the website** was agreed by the WG. Three separate Wiki “rooms” / pages for these specific topics were created:

Forum 1 - discussing the topic “**volunteered geographic information (VGI)/crowd-sourcing**”

Forum 2 - discussing the topic **definitions for gazetteers and data types**

Forum 3 - discussing the topic **general feature types/categories**

The WG website is linked to the section available within the UNGEGN website. The software used for the WG website potentially allowed implementing Wiki functionality and set up a discussion forum within the website. This discussion forum should be utilized to commence investigations on these three issues mentioned before within the WG by creating separate Wiki “rooms” / pages for these specific topics and upload initial documents to it. Registered members will start to comment on the initial documents and extend the content of the Wiki “rooms” / pages respectively.

Two members of the WG per forum have been nominated to function as moderators in order to manage, monitor and contribute to the discussions. Concerning the content of each forum several sub-sites comprising selected questions have been created to help to streamline the discussions properly.

Some UNGEGN experts as well as other experts for the respective topics were registered in 2013, commented and extended the content of the Wiki “rooms” / pages respectively.

The following chapter aims at informing the UNGEGN about the outcome and findings of the discussion forum.

2. Outcome and findings of the discussion forum

2.1. Findings of Forum 1 - discussing the topic **volunteered geographic information (VGI)/crowd-sourcing**

Forum 1 of the wiki discussion pages focuses on the topic of *volunteered geographic information and crowd sourcing*. The two moderators for this topic are Ms Caroline Burgess (until 2013, United Kingdom) and Ms Annette Torensjö (Sweden).

The topic was initiated by the workshop on VGI, crowd-sourcing and web harvesting during the 10th UNCSGN in New York in August 2012. The workshop was led by the UNGEGN experts Annette C Torensjö (Sweden), Caroline Burgess (United Kingdom), Dr Laura Kostanski (Australia) and Heather Ross (Canada).

Starting point for the discussions was a presentation of different projects in the area of field collection of names from the local population and the fact that this has always been important to the process of standardization promoted by UNGEGN. It was shown that the volume of the information available is being provided in different ways and can be rather overwhelming to the names authorities that are trying to handle it. Furthermore it was discussed how the names authorities can deal with this wealth of information which is undoubtedly useful, and at the same time maintain “authoritative” data, for instance for government mapping. Other questions discussed were: If governments move away from authoritative names, would names authorities still be relevant? With reduction in government resources (in most places) can collection and maintenance of unauthorized names be viable? The workshop was well-attended and there was expressed a wish to continue this topic of discussion.

Forum 1 was divided into two sub-sites, *F1.1 General discussion* and *F1.2 Reliability*. Within the sub-site *F1.1 General discussion* it was discussed how effective and valuable the collection of names data using volunteered geographic information currently is. Two main questions were debated. Question 1: How are the crowd-sourced data being collected and processed in your institution/organization? Question 2: How willing is the public to be involved in such a project and how is data provision encouraged? Seven experts have posted advices, ideas and experiences in the topic.

Although the potential of VGI and crowd-sourcing has been recognized by organizations, as these sources provide unique elements of geospatial information, methods of processing the data are still lacking and have to be evolved. Concerning the public's involvement easy methods, i.e. the ability to be able to simply click on a link from a web page, are preferred.

Within the sub-site *F1.2 Reliability* the issues of using VGI and crowd-sourced data in authoritative products and services was discussed. Three main questions were debated. Question 1: Is the information received considered to be reliable and what are the main methods of quality control? Question 2: What types of problems does this VGI data pose to your organisation? Question 3: Have you developed any rules for using it? In this sub-site five experts took part and contributed with their expert knowledge.

The degree of reliability, sub-section 1.2-Q1, has generated the most (only) discussion so far. Contributors acknowledge that the reliability of the data naturally depends on

the source of the information and the collecting body having a disciplined approach to the analysis of the returned information. The advantages of having access to a large bank of freely-provided data need to be weighed up against the problems of managing so much information and ensuring its quality. The moderators would encourage UNGEGN experts to continue to contribute as there are clearly some interesting projects in place.

As the methods of accessing geographical information change and public interaction with such data becomes more dynamic, national mapping agencies and place-name organisations are looking to take advantage of the willingness of the local population to provide information on location and associated attributes (= Volunteered Geographic Information). The principal identifier of a location is its name and it is with this particular attribute that experts of the United Nations Group of Experts on Geographical Names (UNGEGN) are concerned.

Experts from several countries have related their experiences in collecting and utilising crowd-sourced geographical names data (supporting documentation where posted on this forum) and it is clear that there is no single established approach. For example in Sweden, the National Land Survey of Sweden – Lantmäteriet - has turned to crowd-sourcing as a new method of toponymic field collection by using a specifically-designed mobile phone application (a paper posted: "New method of field collection of Place-Names") In Great Britain, Ordnance Survey has been experimenting with several different means of making use of crowd-sourced names information. Collaboration with the English Project's Location Lingo scheme involved the use of a website to encourage public provision of unofficial or colloquial names; a separate initiative has been developed with the UK coastguard focussing on the collection of locally-used coastal names; and a further project using web-harvesting to extract information from websites through text analysis has also been investigated. Work is ongoing.

In order to assess the effectiveness and value of collecting names using volunteered geographic information, it would be useful to learn of experiences in other countries involved in such methods of geographical names collection.

2.2. Findings of Forum 2 - discussing the topic definitions for gazetteers and data types

Forum 2 of the wiki discussion pages focused on the topic of *Definition and Content of Gazetteers*. The two moderators for this topic are Dr Laura Kostanski (Australia) and Ms. Vita Strautniece (Latvia).

The inspiration for this particular forum was borne during a paper presented during the first session of Technical Committee II (E/CONF.101/57 and Add.1) wherein it was recommended that the WG TDFG commence investigations into defining appropriate definitions for gazetteers and data types.

[...] Within UNCISG resolutions and UNGEGN recommendations and policies, and indeed within the wider research literature, there does not appear to be commonly accepted definitions for the terms 'official' and 'unofficial' as they relate to gazetteer data. Rather, there seems to be a proliferation of terminology used to define both the types of data which are incorporated into gazetteers, and the gazetteers themselves- ranging from official and authorised to unofficial and informal....We strongly encourage the member states of UNGEGN to commence

the conversation on defining gazetteer and data types with the aim of developing robust definitions and increasing the relevance of the systems we currently maintain....[E_CONF.101_57_The Four Faces of Toponymic Gazetteers]

Forum 2 was divided into two sub-sites- *F2.1 Definition and content of gazetteers* and *F2.2 Differences between data types*.

Guests were informed that the term 'gazetteer' in a Spatial Data Infrastructure (SDI) context might be considered as "any geospatial dataset which contains 'spatial identifiers' ". These can be geographical names, postal codes or other indexes for indirect spatial referencing. The intended use of 'gazetteers' in the European INSPIRE initiative (using 'geographic identifiers') followed *ISO 19112*. The schema from ISO 19112 was not used as-is to correct errors in that schema and allow for a better integration in INSPIRE as a SDI. 'Gazetteers' here were simply intended as a channel to publish spatial data from the INSPIRE themes that allows others to use them in indirect spatial referencing. It is obvious that this technical SDI view on 'gazetteers' is different from the UNGEGN view on 'gazetteers':

List of toponyms arranged in alphabetic or sequential order, with an indication of their location and preferably including variant names, type of (topographic) feature and other defining or descriptive information. (UN Glossary of the Terminology, 2002)

Guests to the Forum 2 sub-sites were invited to interact with the WGTDFG by posting responses to various questions including

- F2.1-Q1: What is your understanding of the term 'gazetteer'?
- F2.1-Q2: How many faces have your 'gazetteers' got in your country/Institution? (Please provide short descriptions)
- F2.2-Q1: What indications for data types are provided through your spatial data sets?
- F2.2-Q3: What is your opinion regarding the initial definitions, placed in the table above? Have you any additions, or other proposals? Questions?
- F2.2-Q2: Have users provided you with any problems / difficulties / concerns using these data types?

Various responses were provided from UNGEGN members to these questions. From the answers and descriptions posted it is apparent that amongst members there is a wide variety of databases used to collect different types of geographic names information. The most pertinent comments included

- The term 'gazetteer' does not transpose directly to other languages- where these languages might use another term entirely.
The moderators recommend to UNGEGN that consideration be given to defining the term 'gazetteer' and offering official translations/transliterations of this term for members to utilise.
Further to this, an attempt should be made to capture the range of terms used for member's lists of geographic names- i.e. list of names, database of geographic names, dictionary of names etc.
- Many 'gazetteers' only collect and publish official information- which is a problem for researchers and users who often want to find details of colloquial, historical or unofficial names.

The moderators recommend to UNGEGN that consideration be given to defining what ‘unofficial’ names are and how they might be collected, maintained and published by names authorities.

2.3. Findings of Forum 3 - discussing the topic general feature types/categories

Forum 3 of the wiki discussion pages focused on the topic of *general feature types/categories*. The two moderators for this topic are Ms. Naima Friha (Tunisia) and Mr. Teemu Leskinen (Finland).

The topic “General feature types and categories” was initiated by the conference paper “*Feature Types for Global Gazetteers* “ (Laura Kostanski et al., Australia) submitted to the Technical Committee IV session of the 10th UNCSGN. The WG TDFG was requested to further discuss these issues and to contribute findings to the larger UN Spatial Data Infrastructure initiative (UNSDI) driven by the UN Geographic Information Working Group (UNGIWG).

Forum 3 was divided into four sub-sites – *F3.1 General discussion, F3.2 Requirements and use cases, F3.3 Best practices and proposals and F3.4 Reference materials and classifications*.

F3.1 General discussion – Concerning the biggest challenges in developing global classifications for named features, in general, following aspects were discussed and different ideas and comments were posted:

- Different natural, cultural and linguistic circumstances and topographical variations in different parts of the world;
- The terms used for the natural and cultural features in one or different languages across the world;
- How to create a harmonized global classification that different existing, often detailed national and regional classifications could be easy to map with in a rational way;
- Differences in relevant use cases and end-user application needs;
- Difficulties in expressing quantitative aspects in a common global classification for natural features, for example a “large lake” in Spain is very different from a “large lake” in Finland;
- The rationale for including a feature in a “global gazetteer service” may depend on the local status and local importance of the feature, classification by dimensions (length in kilometers, area in square-kilometers and height in meters) may not be satisfactory;
- How do named features (e.g. in gazetteers) differ from spatial objects in general (e.g. in GISs) with regard to feature identity, definition, classification, geometry and other attributes.

F3.2 Requirements and use cases – Concerning requirements and use cases for appropriate global feature classification, following aspects were discussed and different ideas and comments were posted:

- Both official bodies and any colleagues involved in geographical names data collection, GIS, SDI, international and regional initiatives (e.g. EuroGeonames) as well as academic and private parties should be invited to participate in the discussion on global classifications;

- Place names issues are part of almost any economic development projects;
- Besides the spelling and language of the name, the type of a named place is an important query term in practically any search use case or application;
- An appropriate and universally understandable feature classification helps in getting the hoped for search results, places and/or their names;
- Visualization is a prominent use case for harmonized feature types and a possible feature type hierarchy, e.g. names of different kind of features may be portrayed differently: names of cultural features in black upright typeface, names of natural terrain features in black italics, names of hydrographic features in blue italics etc.;
- A harmonized global feature classification should serve the United Nations Spatial Data Infrastructure (UNSDI) use cases as well as other, more general ones;
- The classification system should be concise enough, yet sufficiently detailed for different kinds of use cases and natural and cultural circumstances around the world as well as feasible and unambiguously mappable for any names data provider;
- It would be essential to study different kinds of end-user application needs before setting a general global classification;
- In some cases the terminology related to similar feature types differs from one country to another;
- A harmonized named feature classification should be created by geographical names experts but we should try to avoid, as far as possible, a contradiction with existing (professional) feature type classifications.

F3.3 Best practices and proposals – Concerning best practices and proposals no comments were given to this sub-site.

The moderators recommend to UNGEGN that consideration and discussions should be continued. An attempt to define a global feature classification should be based on a user case scenario, i.e. a request defined by a UN initiative or by another multi-national project or initiative.

F3.4 Reference materials and classifications – WG TDFG members were invited to introduce, upload and discuss related reference materials and classifications. Materials may be available on the web (links to documents, gazetteer services etc.) or they may be uploaded in the sub-site by WG members (e.g. national classifications).

The reference material included a variety of different international, organizational, national and regional classifications publicly available on the web, and included the following:

International/organizational classifications

- ADL Gazetteer Feature Type Thesaurus, hierarchical (up to 5 levels) classification, 6 main classes, more than 200 sub-classes;
- Getty Thesaurus of Geographical Names, Place types, a vast, detailed classification;
- EDINA, Unlock, Feature types, a five-level hierarchy with 5 main classes;
- INSPIRE Data Specification for the spatial data theme Geographical names, Feature classification. One-level classification, 9 classes, partly based on INSPIRE spatial data themes. A related discussion paper on feature classification during the INSPIRE Geographical names names (GN) data specification process;

- EuroGeoNames, Feature classification, 8 main classes and 27 sub-classes. Sheet/form for the translation of terms used in the EGN data model and the EGN feature classification comprising 6 folders including the respective information (enumerations, country names, languages, feature classes);
- Geonames.org, Feature codes, a vast hierarchical classification with 9 main classes;

National/regional classifications

- Maa-amet (Estonian Land Board), Place Names Register on-line search, in Estonian and English, a hierarchical classification with 6 main classes;
- Natural Resources Canada, Geographical Names, Feature Types, 3 main classes and some 40 sub-classes;
- U.S. Geological Survey (USGS) Geographical Names Information System (GNIS) Feature Class Definitions, some 65 classes;
- IGN, France, DB NYME data specification (2009), 2 themes, 8 main classes and more than 100 sub-classes;
- ICC, Catalonia, Topographic database 1:5000 data specification (2011). Annex 2: Named feature classification, 18 main classes and some 250 sub-classes;
- A tentative proposal (Tunisia) for a feature type classification, 7 main classes.

3. Outlook

Some UNGEGN experts as well as other experts for the respective topics have been registered and started to comment and to extend the content of the Wiki “rooms” / pages respectively. More volunteers from UNGEGN are most welcome to provide further content.

In order to assess the effectiveness and value of collecting names using volunteered geographic information, it would be useful to learn about more experiences in other countries involved in such methods of geographical names collection. Further experiences in collecting and utilising crowd-sourced geographical names data have to be made as there is currently no single established approach.

The definition of the term ‘gazetteer’ might be updated and offering official translations / transliterations of this term for members to utilise. Further to this, an attempt should be made to capture the range of terms used for member’s lists of geographic names- i.e. list of names, database of geographic names, dictionary of names etc. Additionally definitions on what ‘unofficial’ names are and how they might be collected, maintained and published by names authorities seem to be needed.

The biggest challenges in developing global classifications for named features and the requirements have been discussed intensively. Best practices and proposals are needed to continue the discussions and to find solutions. Amongst others, it would be essential to study different kinds of end-user application needs before setting a general global classification.

Further UNGEGN experts are invited to contribute to these issues. If experts are interested please contact the convenor of the WG or the moderators for the respective topics mentioned on the webpage of the discussion forum: <https://wiki.gdi-de.org/display/wgtdfg/Discussion+forum>