

GIS BASED GAZETTER OF HUNGARIAN TOPONYMS

A. Guszlev¹, L. Lukács²

ABSTRACT

The Hungarian Toponymic Program (HTP, MTA-ELTE Research Group on Cartography and Geoinformatics, Budapest, Hungary) is dealing with research on place name usage in Hungary, and compiling databases and guidelines for supporting the elaboration process of toponymic contents on maps. The main purpose of HTP is collecting Hungarian place names of the Carpathian region and establishing a spatial database and on-line gazetteer of them.

*

1. INTRODUCTION

Gazetteers are compiled to support identification of place names. Gazetteers are sources of name forms in map-making, administration, and other fields. Use of gazetteers is not widespread among professional cartographers, geographers, experts of geoinformatics and non-professional readers in Hungary. The most possible reason for that is the difficult availability of data suitable for cartographical use.

Thousands of Hungarian toponyms can be found also in official gazetteers of Hungary and in diverse world gazetteers on the Web. The official Gazetteer of Hungary contains correct name forms, coordinates and useful attribute data, but does not embrace the whole Hungarian naming area, only the current state. The largest multilingual world gazetteers store plenty of past and current Hungarian name forms for the Carpathian Basin, but names in them are often not correct in spelling. For these reasons it could be useful to compile an on-line gazetteer with visualization possibilities for the whole Carpathian Region.

2. DATABASE OF HUNGARIAN GEOGRAPHICAL NAMES

2. 1. Purposes

Gazetteers are rarely used in Hungary. Their use is not widespread among professional cartographers, geographers, geoinformaticians and non-professional readers. The most possible reason for that is the difficult availability of data suitable for cartographical use (Guszlev and Lukács, 2006).

The main goal of HTP is to develop an organized **Database of Hungarian Geographical Names (Magyar Földrajzi Névi Adatbázis, MFNA)**, a user-friendly Hungarian name database and publish it on the Internet according to international standards and recommendations. The gazetteer is suitable for medium and lower scales (*Figure 1*).

¹ University of West Hungary, Székesfehérvár, Hungary, ga@geo.info.hu

² Institute of History of the Hungarian Academy of Sciences, Budapest, Hungary, llilla@map.elte.hu

- MFNA takes into consideration aspects of a *gazetteer for cartographical use*:
 - Contains geographical names with object-type and location information
 - Overview of current and past name versions
 - Ethimological references
 - Multilingual identification (official names for trans-border Hungarian names, names in minority languages in Hungary, and recommended English terms for Hungarian features.
 - In case of landscape region names the division-system and its hierarchy are represented.
 - Information for cartographical representation
- The framework of the MFNA is a *geoinformation system*, including facilities:
 - different accessing levels for using, editing or supplementing the gazetteer,
 - visualization of gazetteer entries by tables and maps
 - name search, selecting objects by attributes, spatial query
 - temporal analysis of areal references
 - comparison of landscape region delineation systems

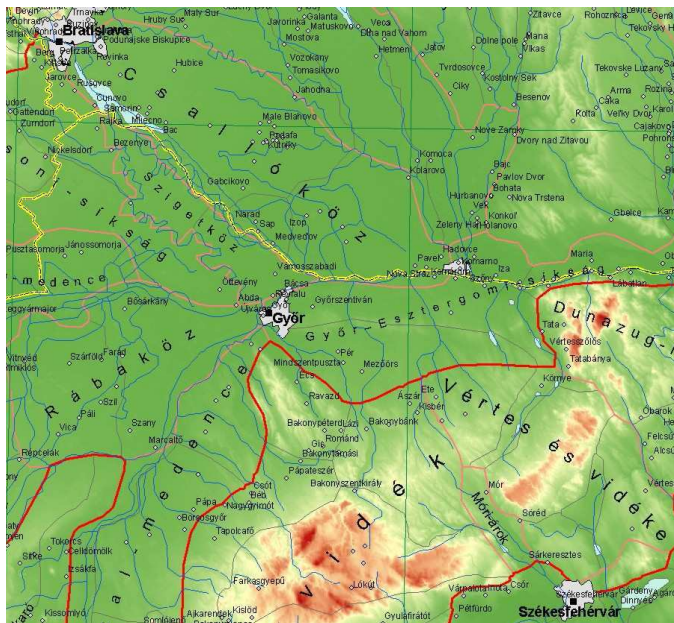


Fig. 1. Part of the MFNA gazetteer

2. 2. Contents and data sources

The gazetteer will cover the Carpathian Region (see Figure 2). For the first step, the database is according to a medium scale (1:250000) map, with approximately 50 000 names. In the future, names of larger scales (e.g. street names and land names) can be integrated into the database.

Temporal scope of the MFNA is from the Middle Ages to the present days. For visualizing historical events or phenomena cartographers usually use historical milestones on the background map: maps of some well-documented states and changes of administrative units, data and names from census documents. In this gazetteer some of these milestones are elaborated: borders of administrative units before the World War I, in 1950, in 1990 and in 2007. In these years, there were significant reconstructions in the Hungarian administrative system and borderlines of units were changed largely because of political events.

For compiling a multilingual GIS gazetteer various map and attribute data are used. Base map of relief, hydrography, borders and settlements is required with detailed geometry. For implementing the base map of MFNA, several sources were used with different scale, project and precision. The data integration was complicated because different integration methods had to be chosen for each type of data source.

Entries in the gazetteer are geographical names. Their spatial references are recorded with geographical features as map elements, points, polylines or polygons.

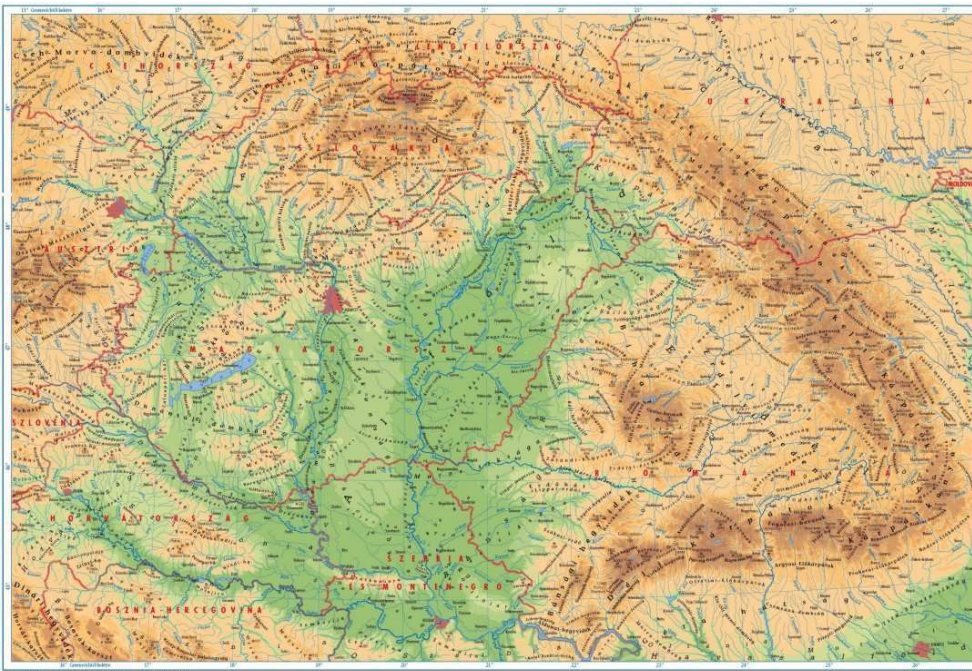


Fig. 2. The Carpathian Region, area of the gazetteer (Farágó, 2005)

Map data in the MFNA

- Relief (raster format)
- Marked geographical points: main peaks, passes (point geometry)
- Watercourses (polyline geometry)
- Lakes (polygon geometry)
- Administrative units and settlements (polygon geometry)
- Landscape regions (polygon geometry)
- Geographical names as map labels (annotations)

The source of relief data is the SRTM dataset. SRTM (Shuttle Radar Topography Mission) is a project of the NASA (National Aeronautics and Space Administration) and the NGA (National Geospatial-Intelligence Agency) to create a near-global DEM (digital elevation model) of the Earth using radar-interferometry (SRTM, 2005).

Current NUTS borders (nomenclature d'unités territoriales statistiques) are processed in the gazetteer. Dataset of Hungarian NUTS is provided by FÖMI (Hungarian Institute of Geodesy, Cartography and Remote Sensing) and KSH (Central Statistical Agency), and developed by the GeoX Ltd.

Borderlines of former administrative units are collected from historical thematic map files of Hungary in 1910 (edited by the Institute of History of the Hungarian Academy of Sciences).

The gazetteer contains main rivers, watercourses, canals and lakes. Spatial data source is a geographical map of the Carpathian Basin in scale 1: 600 000 (edited by Imre Faragó). The names were gathered with the help of students of the Department of Cartography and Geoinformatics of Eötvös Loránd University (Budapest) (see Figure 3).

The MFNA project relates to another project, a research on **landscape region names**. Therefore landscape region systems are emphasized in the gazetteer. The base system of the gazetteer is the Landscape Region System of the Carpathian-Pannonian Region (LRSCPR), established by Hungarian geographers, József Hajdú-Moharos and Attila Hevesi in 1990. In this system transborder regions are specially elaborated, and even the region names in different languages are harmonized (Hajdú-Moharos, Hevesi, 1996) (see Figure 4).

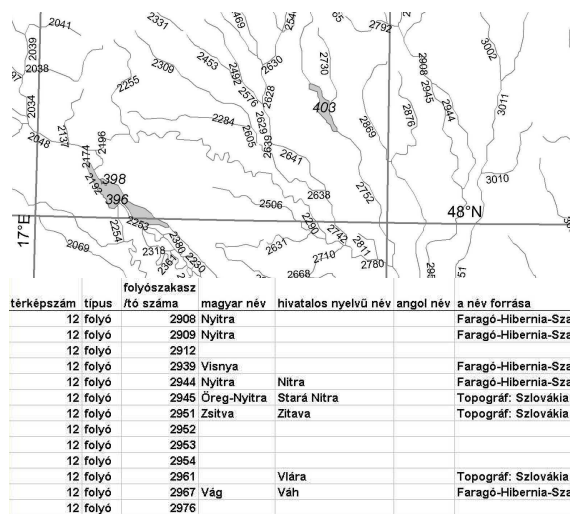


Fig. 3. Parts of the map and the attribute table of hydrographical names

The second represented system is a complex geographical system of Gyula Prinz from 1936. Paper source of this was only in text originally but the system has been reconstructed and plotted on a map in a previous research work (Lukács, 2003)

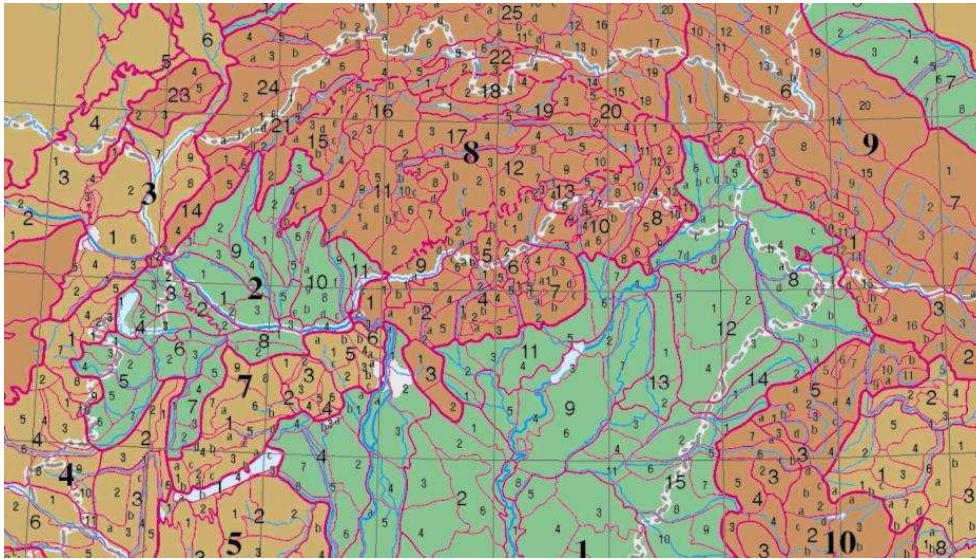


Fig. 4. The Landscape Region System of the Carpathian-Pannonian Region

Most of GIS name-databases store names as attributes. In a gazetteer for cartographical purpose recording and representing labels on maps may come forward, so texts become map objects instead of being simply attributes. The labels can be located automatically in various orientations according to the reference point. Positions of the labels could be specified one by one; they could be stored and treated as **annotations** in separate map layers (see Figure 5). Advantages of using annotations are the higher cartographic quality and that topology of the labels can be also analyzed. Analyzing of the labels can be interesting because they contain additional information about the objects and the names by graphical and typographical attributes (placing, font type, orientation of labels, etc.).

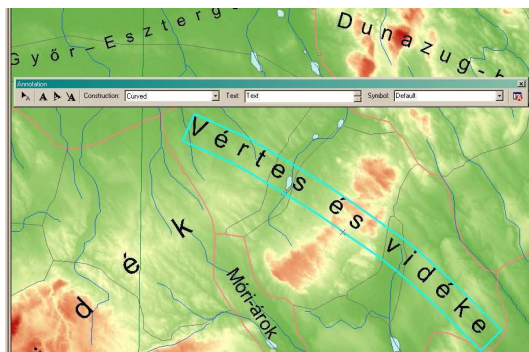


Fig. 5 Annotations in ArcGIS

Attribute data in the MFNA

- Name
- Status
- Temporal validity
- Origin
- Source
- Language
- Writing system
- Object type
- Cartographic name-type

Each **name variant** is a separate record in the database. The **status information** shows that the name is *official*, *alternate* or *former* name. Geographical names are continuously changing also in Hungarian language, new names come to existence and sometimes former and rarely used names vanish. The gazetteer is multilingual for identification of

geographical features laying in overlapping name areas. Most of the toponyms used in the Carpathian Region are in Latin writing system, but there are lots of names in Cyrillic (Ukrainian, Serbian and Bosnian names) and some native names also in Greek letters (e.g. Μπελογιαννίω=Beloianisz, a village with Greek inhabitants in Hungary). Problems related to characters can be eliminated by applying Unicode character sets.

Some attribute data were collected from academic publications, textbooks on geography and history, name-lists and several maps.

Name origin was the only attribute, which has been collected from special source. The most complete Hungarian database of the origin of toponyms is the Ethimological Dictionary of Toponyms (FNESZ, Földrajzi nevek etimológiai szótára; edited by Lajos Kiss). This dictionary contains ethimological data (and further name variants) of 13330 toponyms, most of them are Hungarian. This dataset is in paper form and moreover the data content is much more detailed than required in MFNA, so some base data of origin (first occurrence and the meaning of the name) was digitized for the gazetteer.

The source of recommended English terms was the Gazetteer of English Names for Topographical Objects of Hungary (compiled by Gábor Gercsák, 2002.) This gazetteer is a

Abauj Abauj	Álta-ér Által Brook	Balota Balota
Agár-tető Agár Peak	Bácska Bácska	Bán Bán
Agasvár Agasvár	Badacsony Badacsony	Baranya Baranya
Aggteleki-karszt Aggtelek	Bakony Bakony	Baranyai-dombság Baranya
Karst	Bakonyalja Bakonyalja	Hills
Álfföld Great Hungarian Plain	Bakonyerdő Bakonyerdő	Baranyai-Hegyhát Baranyai-
Alpokalja Alpokalja	Baláta-tó Baláta Lake	Hegyhát
Alsó-Kiskunság Lower	Balaton Lake Balaton	Bársonyos Bársonyos
Kiskunság	Balaton-felvidék Balaton	Bécsi-hegy Béci Hill
Alsó-Válicka Alsó-Válicka	Uplands	

simple name list with pairs of Hungarian and English terms, but without any spatial references (see Figure 6).

Fig. 6. Part of the Gazetteer of English Names for Topographical Objects of Hungary

Object type (e.g. hill, pass, settlement, etc.) is the most important information of a geographic place and a required basic data for gazetteer entries. For this reason, a controlled vocabulary or a thesaurus is necessary for a gazetteer database. Controlled vocabulary is a collection of generic terms used in a gazetteer; thesaurus is a vocabulary which contains also the hierarchy and relations between vocabulary elements. Vocabularies or thesauri could contain hundreds of object types in the huge world gazetteers. MFNA has just a few types because of its relatively small scale and content but it could be elaborated in the future.

2. 3. GIS framework

Gazetteer is compiled in GIS environment, in ESRI ArcGIS 9.1 software. Spatial data of the gazetteer (points, polylines and polygons) are stored as OGC simple features. Excel tables of attribute data are joined into an Access database, and this Access database is linked into the ArcGIS system.

MFNA has a simple **relational database** structure (see Figure 7). The attribute data are recorded in relational database tables. Elements of the database are rows of a table (*records*), and the attributes are recorded in the columns (*fields*). Data can be stored in a lot of tables which are linked to each other by a *key*. Key fields contain the same data in the different tables. Using more tables can result in faster queries in the database and helps to avoid storing redundant data. This way of data storing results in smaller database size and eases modification and broadening of attribute tables.

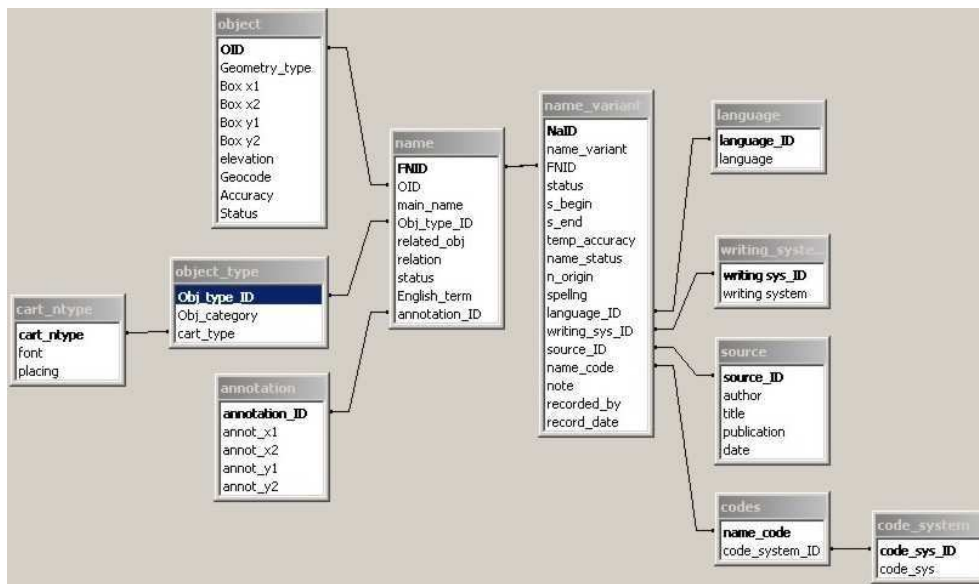


Fig. 7. Data structure in the MFNA

Name variants (and not the primary names) are unique elements in the MFNA, therefore 'Name variants' is the main table of the database. 'Object' table also contain identifier numbers of the map objects as key field, connecting spatial and attribute data.

Attribute data can be stored in Attribute Tables of the layers or in joined Access databases (see Figure 8). In the MFNA data is stored in Access databases (except spatial data and Object IDs), because data records in Excel and Access can easily be modified and updated and also key fields can join additional external databases.

obj_id	jaras_varos	jszekhely	jogallas	mszekhely	Megye
390	Esztergom		rendezett tanácsú város	m	Esztergom
402	Párkányi	Párkány			Esztergom
388	Esztergomi	Esztergom			Esztergom
384	Váli	Vál			Fejér
375	Váli	Vál			Fejér
374	Adonyi	Adony			Fejér
383	Móri	Mór			Fejér
372	Sárbogárdi	Sárbogárd			Fejér
376	Székesfehérvári	Székesfehérvár			Fejér
513	Székesfehérvár		törvényhatósági jogú város	m	Fejér

Fig. 8. Attribute table of former administrative system (MS Access database)

2. 4. MFNA system

MFNA will be stored and published on a map server in the future, which performs queries, data processing and provides data and information about the toponyms. The system will have a **catalogue** service managing metadata of the gazetteer (see Figure 9).

Both *editors* and on-line *users* will be able to add new records to the MFNA or attribute data to existing records. In addition, editors will be able to create new fields and new tables in the database, or edit existing content. The database is controlled, a *moderator* checks on new data before entering them into the gazetteer (see Figure 10).

MFNA considers international standards and recommendations. Standards for gazetteers can be found at International Standards Organization. Open Geospatial Consortium (OGC) has been important for years in developing and publishing open and interoperable geoinformation systems.

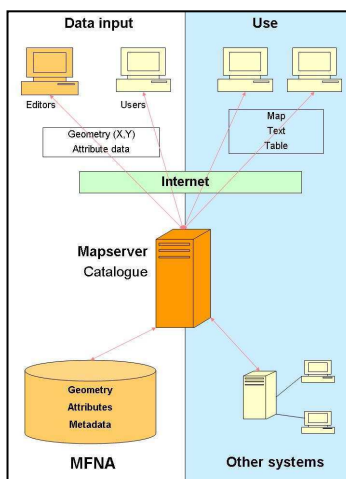


Fig. 9. The MFNA system

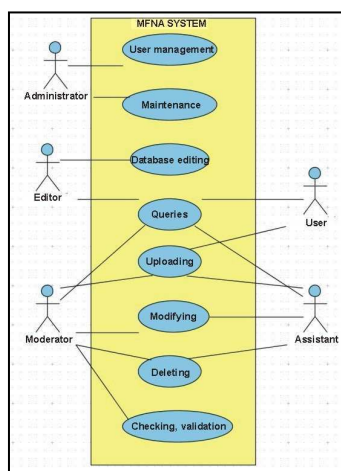


Fig. 10. Use cases of MFNA system

MFNA considers international standards and recommendations. Standards for gazetteers can be found at International Standards Organization. Open Geospatial Consortium (OGC) has been important for years in developing and publishing open and interoperable geoinformation systems.

Advantages of the WGS standards:

- It provides an opportunity of integrating the MFNA into the Internet-based geoinformation systems, therefore makes it accessible in service.
- The application is supported by well-documented open source software
- Extendable and dynamic database

The open database format provides the possibility for transforming geographical names for one's own interest. Names can be easily visualized and used in map or database applications. As an example, the labels provided by Google Earth can be changed to the own selected and more correct version (Guszlev and Lukács, 2006).

The database will be published under **Creative Commons** license (creativecommons.org). It can be downloaded for free, for any non-profit usage or modification, while mentioning the source and property rights.

3. CONCLUSIONS

The aim of MFNA project of HTP is to elaborate an organic, multilingual, open name database by collecting and integrating available spatial and name sources of Hungarian toponyms. However, quality and structure of these data sources are very different. Therefore, eliminating of database gaps, reclassifying data and improving spatial accuracy are needed.

In the MFNA-project our research group tries to support and enable community collaboration in gazetteer development. The aim of our work is to elaborate an organic, multilingual, open database supplemented with up-to-date information and news.

Recent MFNA is suitable for designing toponymy content for maps in medium and lower scale. The database structure allows the enrichment of name-content and attributes for supporting toponym usage of larger scales. In the future, the gazetteer could be extended also spatially. Using buffer zones depending on geographical factors or using fuzzy logic could refine borders of landscape regions.

MFNA gazetteer would embrace the toponym content of several scientific and cultural fields in the Carpathian Basin and can support identification of spatial locations in international communication. MFNA project could take part in developing Hungarian Spatial Data Infrastructure, and could be joined to the European SDIs according to the Infrastructure for Spatial Information in Europe (INSPIRE) Directive.

For completing the MFNA gazetteer and for further developments more financial and human resources are needed. For further research project applications have to be prepared and submitted.

REFERENCES

- Atkinson Rob, Fitzke Jens, (2002), *Gazetteer Service Profile of the Web Feature Service Implementation Specification*, OpenGIS Discussion Paper
- Faragó Imre, (2005), *A magyar földrajzinév-használat, Könyvtári Figyelő*
<http://www.ki.oszk.hu/kf/kfarchiv/2005/4/farago.html>
- Gercsák Gábor, *Gazetteer of English Names for Topographical Objects of Hungary*, in manuscript
- Glossary of Terms for the Standardization of Geographical Names, United Nations, New York, 2002.
<http://unstats.un.org/unsd/geoinfo/glossary.pdf>
- Guszlev Antal, Lukács Lilla, (2006), *Földrajzinév-kutatás térinformatikai eszközökkel*, 3. Magyar Földrajzi Konferencia, Budapest
- Hajdú-Moharos József, Hevesi Attila, (1996), *A Kárpát-Pannon térség természeti tájbeosztása*, Pannon enciklopédia, Magyarország földje, Kertek 2000 Kiadó
- ISO Geographic information – Spatial referencing by geographic identifiers, ISO, 2003.
<http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=26017>
- Evan Prodromou, *Ten Web 2.0 APIs you can really use*, LinuxWorld.com, 12/18/06
- UNGEGN, (2002), *Toponymic Data Files, Report of the Working Group on Toponymic Data and gazetteers for the period 1998–2002*
- Zentai László, Guszlev Antal, (2006), *Web2 és Térképészet*, Geodézia és Kartográfia 2006/11.