

## HUNGARIAN TOPONYMS IN GAZETTEERS

A. Guszlev<sup>1</sup>, L. Lukács<sup>2</sup>

### ABSTRACT

Gazetteers are compiled to support identification of placenames. Gazetteers are sources of name forms in map-making, administration, and other fields. Hungarian toponyms can be found also in official gazetteers of Hungary and in diverse world gazetteers on the Web. This article summarizes development of Hungarian gazetteers and shows the possibilities and problems related to on-line gazetteers.

\*

### 1. INTRODUCTION

People living in communities name geographical objects in their surroundings. Geographical names are used as identifiers of geographical features in verbal and written communication. Each community and nation has a set of toponyms in their mother tongue. Areas of different name collections overlap each other so geographical objects usually have two or more name variants in the same or also in different languages. Use of several name variants can be ambiguous, that is why international associations make an effort to standardize geographical names, determine only one recommended name for each object. On the other hand, toponyms are parts of national cultural heritage. Every nation tries to protect and use them in education, in literature and in everyday communication.

Compiling gazetteers could be a solution for the contradiction between the name standardization and the living name variants. A gazetteer can help in collecting the geographical names and their attributes, makes name variants identifiable.

Hungarian language is spoken in all countries of the Carpathian Basin (Austria, Slovakia, Ukraine, Romania, Serbia, Croatia and Slovenia), and there are populous Hungarian minorities in these countries. Hungary was extended over the whole Basin for centuries so there are a great number of Hungarian geographical names of medieval origin in this region, mostly used up to presents. The area covered by Hungarian toponyms is 3.5–4 times bigger than the current state area [Farág, 2005].

Name databases, gazetteers or detailed maps are needed for compiling the name-content of a map. There are registers, official name lists of administrative, settlement and street names and other digital sources and most of them also available on the Web, but a general gazetteer of these names is still missing [Guszlev and Lukács, 2006].

### 2. GAZETTEERS

---

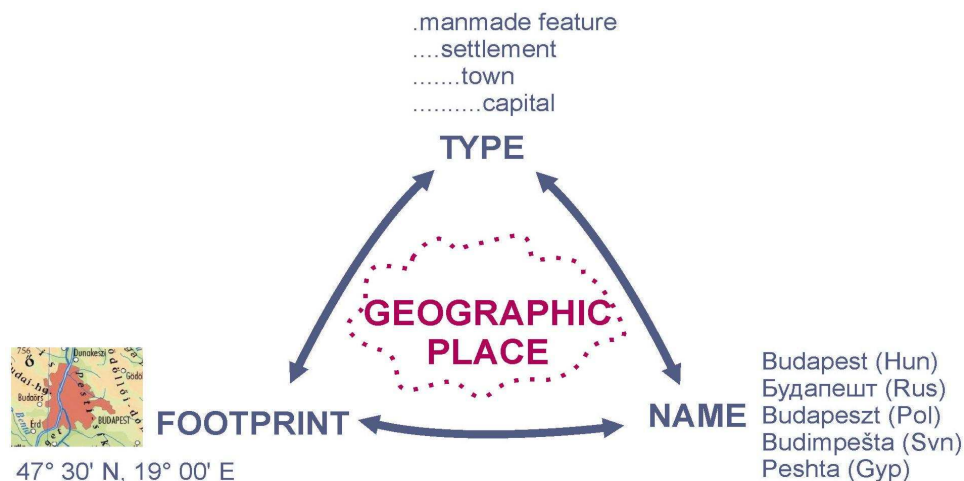
<sup>1</sup> *University of West Hungary, Székesfehérvár, Hungary*

<sup>2</sup> *Institute of History of the Hungarian Academy of Sciences, Budapest, Hungary*

Glossary of terms for Standardization of Geographical Names gives the following definition for *gazetteers* [UNGEGN, 2002]:

*List of toponyms arranged in alphabetic or other sequential order, with an indication of their location and preferably including variant names, type of topographic feature and other defining or descriptive information.*

This definition shows that a name of its own is not a unique identifier for a geographic place. The name, the object-type and spatial location are needed for unambiguous identification [ADLG, 2002] (see Figure 1).



**Fig. 1: The object type, the name and the 'footprint' (location) as identifiers of a geographic place (based on figure of ADL Gazetteer team, 2002)**

The same name can denote diverse geographical objects even at the same place (e.g. Margitsziget is a settlement part of Budapest, while Margit-sziget is an island in the Danube), so you need to know also the object type. The spelling of Hungarian toponyms sometimes gives information about the type of the object.

The same name form can denote objects of similar types at different locations (e.g. Velence is a settlement in Hungary and the name is the Hungarian exonym form of Venice in Italy). These names are called *homonyms* [UNGEGN, 2002].

Essentials of a gazetteer entry for place identification are:

- Toponym (geographical name)
- Object type and
- Spatial location (coordinates)

The aim of compiling gazetteers is to register correct forms of official names, alternate names and their location and attribute data (e.g. type, language, origin etc.). Gazetteers can be data sources for correct spelling of the names and for uniform and consistent toponym usage.

In articles and advertisements geographical names occur in various (and sometimes false) forms and there are often faults in spelling. The reason may be the lack of

professional and widely available data sources. People reckon maps as authoritative data sources, so they search for spelling of the names on maps primarily. As gazetteers are the main sources of names in mapmaking processes, cartographers have a share in compiling convenient gazetteers.

Gazetteers are used in several fields which deal with toponyms:

- Cartography
- Geoinformatics
- Geography, biology, ecology
- History, ethnography, linguistics, literature
- Education
- Press
- Tourism
- ...

### **2. 1. International organizations, standards and recommendations on gazetteers**

Relevance of toponymy and gazetteers is indicated by the fact that many international organizations deal with analysis of toponyms, and the problems of name usage. These organizations harmonize the processes of national toponymic researches, prepare standards and recommendations for compiling gazetteers, provide toponymic guidelines and tutorials for map-making.

Some of the main organizations:

*United Nations Group of Experts on Geographical Names (UNGEGN)* was established in 1967.

*Working Group on Toponymic Data Files and Gazetteers* (as a part of the UNGEGN) was established in 1998.

*ISO Technical Committee Geographic Information/ Geomatics 211 (ISO TC 211)* The standard ISO 19112 – Geographic information – Spatial referencing by geographic identifiers (compared to the ISO 19111 standard of coordinates) This is a conceptual scheme, defining structure and components of XML based name storing.

These organizations collaborate with other associations, e.g. with the *Unicode Consortium* for using uniform characters, standardized diacritical marks in digital character sets.

*Open Geospatial Consortium (OGC)* recommends specifications for protocols of Web Gazetteer Services. Web Gazetteer Service (WGS) is a specialized Web Feature Service that provides additional capabilities specific to a gazetteer-like feature collection. The WGS recommendation is based on OGC directives like Web Feature Service, Filter Encoding and Geographical Markup Language [Guszlev and Lukács, 2006].

## **3. GAZETTEERS IN HUNGARY**

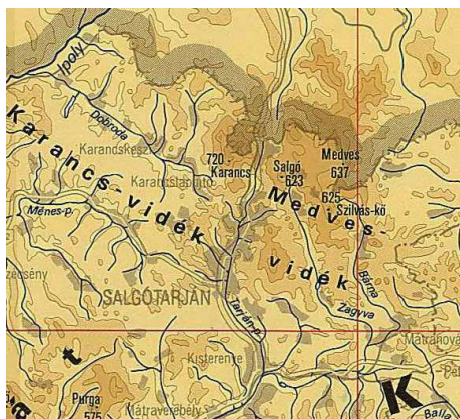
### **3. 1. Traditional gazetteers**

Traditional paper gazetteers mark geographical names with a pair of coordinates or other spatial references (e.g. administrative unit) in text or table form. There are some attribute data related to the name entries: name variants, not recommended name forms, origin etc. Paper gazetteers often do not include maps for visualization.

The needs for professional name sources necessitated the collection of Hungarian geographical names, determination of standard forms, elaboration a gazetteer of Hungarian toponyms and development of the grammatical rules. The Hungarian Board on

Geographical Names (Magyar Földrajzinév-bizottság, Fnb) was established for executing these tasks in 1964. As a result of its works, Fnb published the Gazetteer of Hungary I-II. (FNT I-II) in the 1970's. The FNT I. stored 715 of the best-known nationally spread geographical names in the form which is nearest to the local usage in the related region. According to the decree No. 10/1974.(III.31.) MEM by the Minister for Agriculture and Food, *geographical names should be used in official publications in the form given in the Gazetteer* [FNT I, 1982].

FNT has a simple data structure, gazetteer entries are in alphabetical order and, in addition, they are visualized on a map-sheet (see Figure 2).



Karancs; h; Karancs-vidék (legm. p.); Ng., CSSZ.  
Karancslapujtó; 720 m; C10

Karancs-vidék; hdt; Északi-középhegység; Ng.  
V: Karancs, Karancs-hegység, Karancs-Medves, Karancs-  
-Medves-hegység; C10

Karasica; vf; Baranyai-dombság; Bar., JUG.

\*Feked-Illocska → Duna (JUG.)

V: Karasica-vízfolyás, Karassó; Sz: Karasica-csatorna,  
Karasica-patak; J-K7

Fig. 2. : Cartographical presentation and entries in FNTI

The gazetteer was created for cartographical use. In the name entries at the first place the geographical name is given, established for official use and the name is followed by additional data:

General information:

- Object type
- The county or the region where the feature lies
- The source or respective countries, in case of water courses

Additional information:

- In case of water courses a town name is given where the source or starting point can be found.
- The name of source and the receptive branch, canal or lake
- For mountains and hills altitudes are also given.

Variant names:

- Name variants which are not suggested for use (they are included only for a more complete information, and for comparison)
- Name variants of sections in case of water courses

Spatial location: reference letter and number showing the position of the feature on the map supplement. [FNT I., 1982]

The gazetteer also has an index which contains only the variant names.

This gazetteer helps the compilation of map labels only by simple searching and gives no possibility for querying by attributes. Traditional gazetteers can be kept up-to-date hardly because of the paper form.

### 3. 2. Digital gazetteers

In the 1990's a digital name database of Hungarian toponyms was developed by the FÖMI (Földmérési és Távérzékelési Intézet, Institute of Geodesy, Cartography and Remote Sensing). This gazetteer database is called FNT like the former paper versions, but has a different content. The digital version has two forms with different scales. FNT 1 corresponds in name quantity to a topographic map in scale 1:40 000. FNT 2 corresponds to a large-scale (1:10 000) topographic map, but FNT 2 database is still not completed (60% is ready).

The digital FNT contains 41 types of geographical objects with alternate names and other attributes. The content includes the names of Hungarian settlements and parts of the settlements (with number of population and other attribute data), beyond the geographical names of the paper FNT. Digital FNT is made not for special cartographic but general use, so it gives no information about cartographical representation. For spatial reference, FNT gives a pair of coordinates for each name records in EOY (Hungarian Uniform Projection, Egységes Országos Vetület) coordinate system. In case of watercourses the reference point is the point of exit from a settlement area. In case of landscape regions or other polygon objects there is no standard point for referencing.

Azonosító	Név	Típus	Megye	Település_név	EOVY	EOVX	Magasság	Mag	Lakos	Néps	Vízfo
71575	Karancs	14	13	Karancslapujtó	704514	312905	729	161	0	0	0
36717	Karancs	16	13		696000	312000	0	0	0	0	0
36149	Karancsalja	33	13		702856	309839	0	0	1601	1990	0
71576	Karancsapátfalva	2	13	Karancslapujtó	0	0	0	0	0	0	0
36152	Karancsberény	33	13		702094	316020	0	0	1041	1990	0
71516	Karancsberényi-patak	36	13	Karancsberény	706080	316171	0	0	0	0	0
71517	Karancsberényi-patak	36	13	Karancsberény	702053	315238	0	0	0	0	5001
6279	Karancsdülő	20	11	Törökszentmiklós	751500	206500	0	0	61	1980	0

Fig. 3. Digital FNT

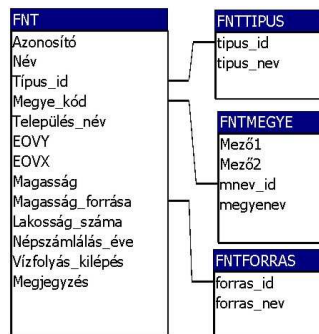


Fig. 4. Relations in digital FNT

FNT, like other digital gazetteers, is a simple relational database. Complex queries, special filtering even on parts of names or attribute data is available in this structure. A more elaborated detailed thesaurus and better structured database provides more special and faster queries (*see Figure 3 and 4*).

This structure gives more possibilities to store temporal dimensions and to analyze changes in the name status and usage. The temporal scope of the gazetteer is variable for supporting historical research or editing historical maps.

Developments and trends on the Web have an effect also on the development of digital gazetteers. Growth of web-communication and community collaboration entails compilation of new object lists, thesauri and ontologies. A semi-controlled geographical vocabulary can be base of management of the geographical names. Large on-line gazetteers have thesaurus of the object types, one of the most known example is the Alexandria Digital Library Gazetteer (ADLG) Theaurus, which contains 215 object types in 4 levels. GeoNET Names Server (GNS) includes approx. 650 object types.

Hungarian gazetteers (both paper and digital ones) also have object types (paper FNT has 19, digital FNT has 39 types) but without hierarchy.

There are projects for creating geo-ontologies as more complex conceptual systems for gazetteers. A gazetteer based on a geo-ontology could be the model of the real world by the names of the objects, considering spatial and temporal relations, topology and hierarchy.

The main role of a gazetteer is to show geographical names (and references) with correct spelling. In practice, gazetteers are made to be toponymic segment of other information systems (e.g. statistical, economic, cultural and geographical databases). For this reason, general identifiers (e.g. postal codes, NUTS codes, ISO country codes or KSH codes in Hungary) are stored in gazetteers. It should be included at the data integration. If there are no standard identifiers, names could be keys for joining more databases. In this case spelling and diacritical marks in the name must be checked.

Additionally, different digital datasets can be processed for gathering data. Process of digital datasets can be automatized if they have a formal relationship. Relation could be two harmonious data field or a normal difference (e.g. differences in diacritical marks). Datasets can join by determination of a key field. (Key field contains the same data in different databases). Integration of different sources can cause problems also with the data content, therefore automatization of the whole integration process need manual checking.

Digital gazetteers contain spatial and attribute data so it is worth to implement gazetteers also in GIS environment. Geoinformation systems are digital data-storing, data-processing and analyzing systems, which are able to process and visualize a great number of spatial data. GIS tools support asking queries both by attributes and location which are main functions of a gazetteer. In a GIS-gazetteer the map is not just a graphical method of visualization, but could take a part in analyzing of spatial phenomena of the names usage [Guszlev and Lukács, 2006].

Data contents of GIS gazetteers do not differ from traditional gazetteers. The method of gathering spatial data is the main difference. In a GIS gazetteer a geometrical object (point, polyline or polygon) belongs to each geographical name. Spatial data (usually coordinates of nodes or bounding box of the object) is joined to this reference object. In other words, the object ID of the stored geometrical object is the link between spatial data and attributes of a name entry in a GIS gazetteer (*see Figure 5*).

The large on-line gazetteers too are based on geoinformation system.

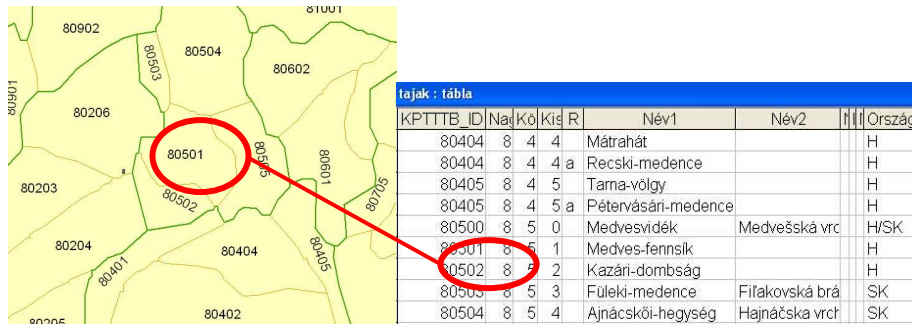


Fig. 5. Joining of polygons and attribute data by ID in ArcMap

### 3. 3. On-line world gazetteers and the Hungarian toponyms

There are several name lists and gazetteers available on the Web. Name lists are searchable but usually have no spatial references and map visualization. The multilingual place name lists on *Transindex* ([www.transindex.ro](http://www.transindex.ro)) are important sources of Hungarian names and their name variants in other languages. In practice, users should use an atlas or map to help identify the placenames spatially (see Figure 6).



Fig. 6. Place name dictionary on Transindex

World gazetteers available on the Internet contain numerous Hungarian toponyms. These gazetteers are multilingual, including variant names in Hungarian and in other languages, which facilitate identification procedure of geographical features.

Three world gazetteers were examined in the point of view of Hungarian toponyms [Guszlev and Lukács, 2006]. All of these gazetteers contain millions of names of the world.

GEOnet Names Server (GNS)

Getty Thesaurus of Geographic Names (GTGN)

Alexandria Digital Library Gazetteer (ADLG)

GNS and GTGN are based on military name-collecting. The sources of these are small scale topographical maps, world atlases and gazetteers.

The ADL Gazetteer is developed by the Alexandria Digital Library (University of California, Santa Barbara). The main point of this project is the recognition that search based on spatial reference data could be applied not only to maps but any objects having spatial reference, e.g. georeferenced photo collections, reports of specific areas, or even pieces of music compositions which contain references [Goodchild, 2004] (*see Figure 7*).

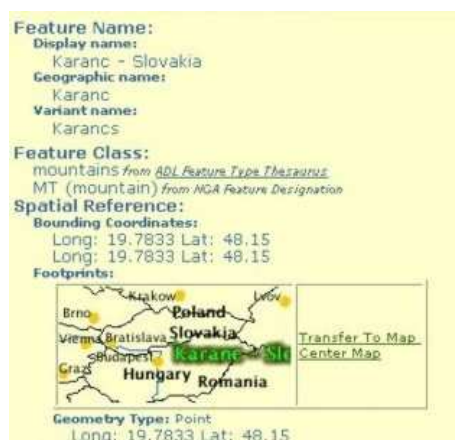


Fig. 7. User interface of Alexandria Digital Library Gazetteer

At the same time problems come forward in using these gazetteers for map-designing. Each gazetteer stores toponyms as point elements with coordinate pairs. These coordinates are determined with arc minute precision. This means that precision of data available for Hungary is 1.3–1.8 km, which is not sufficient for cartographical use under a certain scale. GNS contains 29 861 entries about Hungary (there are 105 000 entries in FNT, for comparison), in the following categories:

- Settlement names (19638)
  - Relief names (3343)
  - Landscape region names (2819)
  - Hydrological names (1962)
  - Railway station names, names of other pronounced points (1904)
  - Administrative names (119)
  - Vegetation names (76)
- 20 Hungarian toponyms were examined in these world gazetteers and query results were recorded. Some of them are inside the current state and others are outside Hungary. Names of hills, mountains, plains, historical regions, lakes and rivers were also searched (*see Table 1*).

#### Hungarian toponyms in world gazetteers

Table 1

TOPONYM	GEONAMES	GETTY	ADLG
Mátra	Mátra	Mátra	Matra
	Mátra Hegység		Matra Hegyseg
Kárpátok	Kárpátok	Kárpátok	Karpatok
	Carpates	Carpates	Carpates
	Carpathian Mountains	Carpathian	Carpathian
	Carpathians	Carpathians	Carpathians
	Carpați, Munții Carpații	Carpati, Muntii Karpaten	Carpati, Muntii Carpatii

	Karpaten Karpaty	Karpaty	Karpaten Karpaty
<b>Alföld</b>	Alföld	Alföld	Alfold
	Nagy Magyar Alföld		Nagy Magyar
	Al'fel'd		Al'fel'd
	Bol'shaya Sredne- Dunayskaya Nizmennost' Great Alföld Great Hungarian Plain Great Plain Hungarian Plain		Bol'shaya Sredne- Dunayskaya Nizmennost' Great Alföld Great Hungarian Great Plain Hungarian Plain
<b>Kisalföld</b>	Little Hungarian Plain Little Plain Upper Hungarian	—	Kis Alfold Little Alfold Little Hungarian Little Plain Upper Hungarian
	<b>Csallóköz</b>	—	Csallokoz
<b>Erdély</b>	Ostrov Great Schütt Velký Ostrov Žitný Velký Žitný Ostrov	—	Ostrov Great Schutt Velky Ostrov Zitny Velky Zitny Ostrov
	Erdély Transilvania Ardeal Siebenbürgen Transilvanie Transilvaniei Transylvania	Erdély Transilvania Siebenbürgen Transylvania Transylvanie Transylvanië	Erdelv Transilvania Ardeal Siebenburgen Transilvanie Transilvaniei Transylvania
<b>Felvidék</b>	—	—	—
<b>Jászság</b>	Jazygia	—	Jazygia
<b>Fertő</b>	Fertő	Ferto	Ferto
	Fertő-tó	Ferto tó	Ferto-to
	Fertő Tói	Fertoto	Ferto Toi
	Fertő Tava Neusiedler See Neusiedler Lake	Neusiedler See Neusiedler	Ferto Tava Neusiedler See Neusiedler Lake
<b>Küküllő</b>	—	—	—

The table shows some of the results of our investigation. In the first column are the Hungarian toponyms and in the other columns are the name forms in the world gazetteers. The Hungarian form with correct spelling and form was signed with orange color. Names with peach color are Hungarian forms with spelling faults or not used form. White ones are the name forms in foreign languages.

The Hungarian diacritical marks are treated defectively or sometimes they are missing in international digital gazetteers, for example: Karpatok instead of the correct term Kárpátok (Carpathian Mountains), Gocsej instead of Göcsej (hills). In addition, the spelling of the names does not suit Hungarian grammatical rules, Kis Alföld, instead of Kisalföld (Little Alföld).

Beyond spelling errors there are distortions in the form of Hungarian toponyms in world gazetteers (Ferto Tava instead of Fertő). Sources of names are not Hungarian but foreign databases or maps. Density of toponyms is not homogenous within a scale range,

likely because of insufficient name sources. For example some of the major regions, like Alföld and Erdély can be found in gazetteers, but other major regions, Kisalföld and Felvidék are not included. Elaboration of Hungarian toponyms for transborder regions is not uniform as well, some of the Hungarian names are stored (Csallóköz, Száva, Kárpátok) in the databases and others (Küküllő, Deliblat) are missing [Guszlev and Lukács, 2006].

## 6. CONCLUSIONS

Gazetteer of Hungary and on-line world gazetteers contain thousands of Hungarian toponyms. In our research these gazetteers were examined as place name sources for compiling toponymic content of maps. For identification of a geographical footprint data of spatial location is needed beyond the place name and feature type. Therefore georeferenced gazetteers or gazetteers with cartographical visualization are more usable than simple name lists without spatial data. The 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. In cartography, it is important to show place names for the whole map area in the language of the map-reader. The largest multilingual world gazetteers store plenty of 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.

## REFERENCES

1. ADL Gazetteer Team (2002): Introduction to Digital Gazetteers and their Development Issues
2. Faragó Imre (2005): A magyar földrajzinév-használat, Könyvtári Figyelő <http://www.ki.oszk.hu/kf/kfarchiv/2005/4/farago.html>
3. Földrajzinév-tár I. Magyarország fontosabb domborzati- táj- és víznevei, Kartográfiai Vállalat, Budapest 1982.
4. Földrajzinév-tár II. Veszprém megye, Kartográfiai Vállalat, Budapest, 1981.
5. Glossary of Terms for the Standardization of Geographical Names, United Nations, New York, 2002. <http://unstats.un.org/unsd/geoinfo/glossary.pdf>
6. Guszlev Antal – Lukács Lilla (2006): Földrajzinév-kutatás térinformatikai eszközökkel, 3. Magyar Földrajzi Konferencia, Budapest
7. ISO Geographic information – Spatial referencing by geographic identifiers, ISO, 2003. <http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=26017>
8. Michael F. Goodchild (2004): The Alexandria Digital Library Project, D-Lib Magazine May 2004, Volume 10 Number 5 <http://www.dlib.org/dlib/may04/goodchild/05goodchild.html>
9. UNGEGN (2002): Toponymic Data Files, Report of the Working Group on Toponymic Data and gazetteers for the period 1998–2002