University of West Hungary
Faculty of Geoinformatics

Database of Hungarian Geographical Names

THESIS

Consultant:
Antal Guszlev
assistant professor

Candidate:
Ibolya Lilla Lukács
Post-graduate distance education course in Geoinformatics

Székesfehérvár
2007

Contents

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.

The Hungarian language is spoken in all countries of the Carpathian Basin 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 [Faragó, 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].

In my thesis, I would like to give a summary of the elaboration of the Database of Hungarian Geographical Names (MFNA) in a geoinformation system. This gazetteer is one of the projects of the Hungarian Toponymic Program (HTP) of the MTA-ELTE Research Group on Cartography and Geoinformatics. The framework of the MFNA will be a special geographical information system, in which Hungarian toponyms of the whole Carpathian Basin are collected, analyzed and visualized from a cartographic perspective.

I would like to express my gratitude to my consultant, Antal Guszlev for giving to me advices and moral support to complete my thesis. I thank Imre Faragó and Gábor Gercsák for sharing their databases.

2. Gazetteers

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)*.

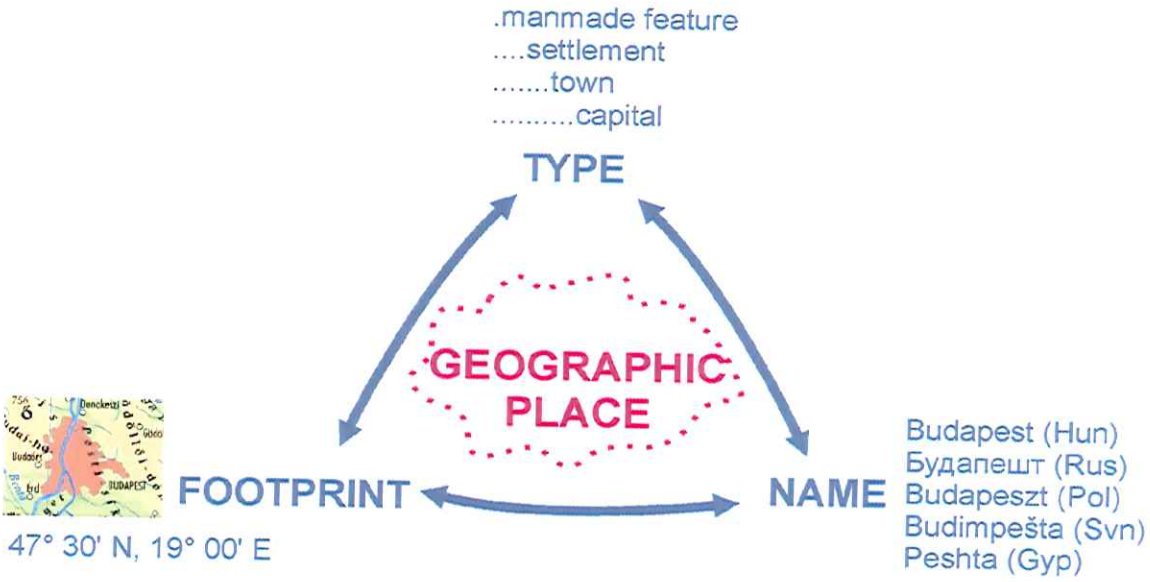


Figure 1: The object type, the name and the 'footprint' (location) as identifiers of a geographic place (based on figure of ADL Gazetteer team, 2002)

- 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, 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].

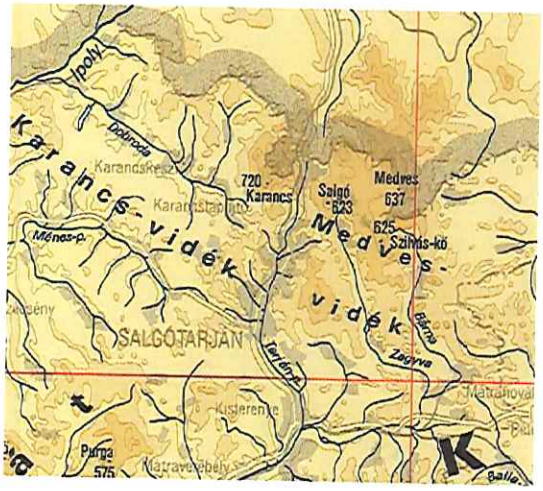
3. Traditional and digital gazetteers in Hungary

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.
Karancslapujtő; 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

Figure 2: Cartographical presentation and entries in FNT I

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 (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 EOV (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	Megjegyzés
71575	Karancs	14	13	Karancslapujtő	704514	312905	729	161	0	0	0	
36717	Karancs	16	13		696000	312000	0	0	0	0	0	Északi-középhegység része
36149	Karancsalja	33	13		702856	309839	0	0	1601	1990	0	tszh: Karancslapujtő
71576	Karancsapátfalva	2	13	Karancslapujtő	0	0	0	0	0	0	0	
36152	Karancsberény	33	13		702094	316020	0	0	1041	1990	0	tszh: Karancslapujtő
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	Karancsdűlő	20	11	Törökszentmikl	751500	206500	0	0	61	1980	0	

Figure 3: Digital FNT

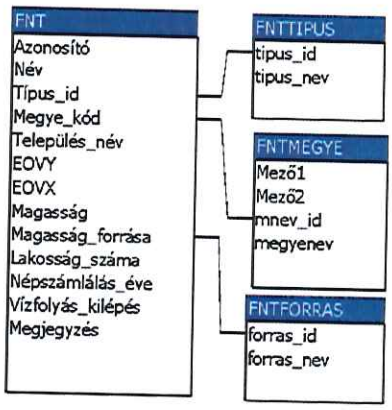


Figure 4: Relations in digital FNT

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.

XMIN	XMAX	YMIN	YMAX	ISO_3DIGIT	CNTRY_NAME	LONG_NAME	helyi_rovid	helyi_hivatalos
-70,06	-69,875	12,4111	12,6278	ABW	Aruba	Aruba	Aruba	Aruba
60,5042	74,9157	29,4061	38,472	AFG	Afghanistan	Afghanistan	Afghanestan	Da Afghanestan Eslami Jamhuriyat/Jomhuri-
11,7312	24,0844	-18,016	-4,38899	AGO	Angola	Angola	Angola	República de Angola (Angolan tasavalta)
-63,168	-62,973	18,1644	18,273	AIA	Anguilla	Anguilla	Anguilla	Anguilla
19,2885	21,0533	39,645	42,6603	ALB	Albania	Albania	Shqipëria	Republika e Shqipërisë (Albanian tasavalta)
1,42139	1,78172	42,4364	42,656	AND	Andorra	Andorra	Andorra	Principat d'Andorra (Andorran ruhtinaskunta)
-69,164	-68,193	12,0206	12,3839	ANT	Netherlands Antilles	Netherlands Antilles	Nederlandse Antillen	De Nederlandse Antillen (Alankomaiden Antil
51,5833	56,3817	22,6333	26,0839	ARE	United Arab Emirates	United Arab Emirates	Al-Imarat	Al-Imarat al-'Arabiya al-Muttahida (Yhdistyne

ISO_3DIGIT	*fnb_rovidnev*	*fnb_hosszunev*	*korabban*	*fennhatosag*	*MSZ kod*	*EN kod*
ABW	Aruba			Hollandia	AW	ABW
AFG	Afganisztán				AF	AFG
AGO	Angola	Angolai Köztársaság	Portugál Nyugat-Afrika		AO	AGO
AIA	Anguilla			Nagy-Britannia	AI	AIA
ALB	Albánia	Albán Köztársaság			AL	ALB
AND	Andorra	Andorrai Fejedelemség			AD	AND
ANT	Holland Antillák			Hollandia	AN	ANT
ARE	Arab Emírségek	Egyesült Arab Emírségek	Szerződéses Omán		AE	ARE

ISO_3DIGIT	XMIN	XMAX	YMIN	YMAX	helyi_rovid	helyi_hivatalos	*fnb_rovidnev*	*fnb_hosszune*	*korabban*	*fennhatosag*
ABW	-70,059662	-69,8749	12,4111	12,62778	Aruba	Aruba	Aruba			Hollandia
AFG	60,504166	74,91574	29,4061	38,47198	Afghanestan	Da Afghanestan E:	Afganisztán			
AGO	11,731249	24,08444	-18,0164	-4,38899	Angola	República de Ango	Angola	Angolai Köztá	Portugál Nyugat-Afrika	
AIA	-63,167778	-62,9727	18,1644	18,27298	Anguilla	Anguilla	Anguilla			Nagy-Britannia
ALB	19,28854	21,05333	39,645	42,66034	Shqipëria	Republika e Shqipi	Albánia	Albán Köztársaság		
AND	1,4213889	1,78172	42,4364	42,65597	Andorra	Principat d'Andorr	Andorra	Andorrai Fejedelemség		
ANT	-69,16362	-68,1929	12,0206	12,38389	Nederlandse	De Nederlandse A	Holland Antillák			Hollandia
ARE	51,583328	56,38166	22,6333	26,08389	Al-Imarat	Al-Imarat al-'Arabi	Arab Emírségi	Egyesült Arab	Szerződéses Omán	

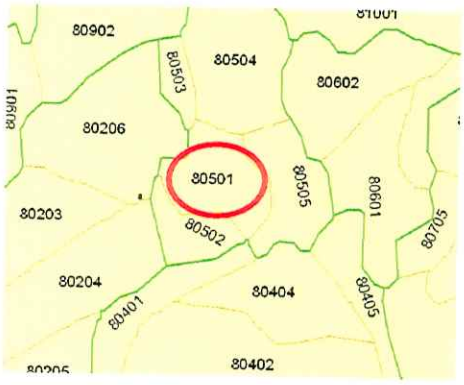
Figure 5: Joining databases by key field

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; *see Figure 5).* 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 6).*

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



tajak : tábla

KPTTTB_ID	Nag	Kö	Kis	R	Név1	Név2			Ország
80404	8	4	4		Mátrahát				H
80404	8	4	4	a	Recski-medence				H
80405	8	4	5		Tarna-völgy				H
80405	8	4	5	a	Pétervásári-medence				H
80500	8	5	0		Medvesvidék	Medvešská vrc			H/SK
80501	8	5	1		Medves-fennsík				H
80502	8	5	2		Kazári-dombság				H
80503	8	5	3		Füleki-medence	Fiľakovská brá			SK
80504	8	5	4		Ajnácskői-hegység	Hajnáčska vrch			SK

Figure 6: Joining of polygons and attribute data by ID in ArcMap

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

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).*

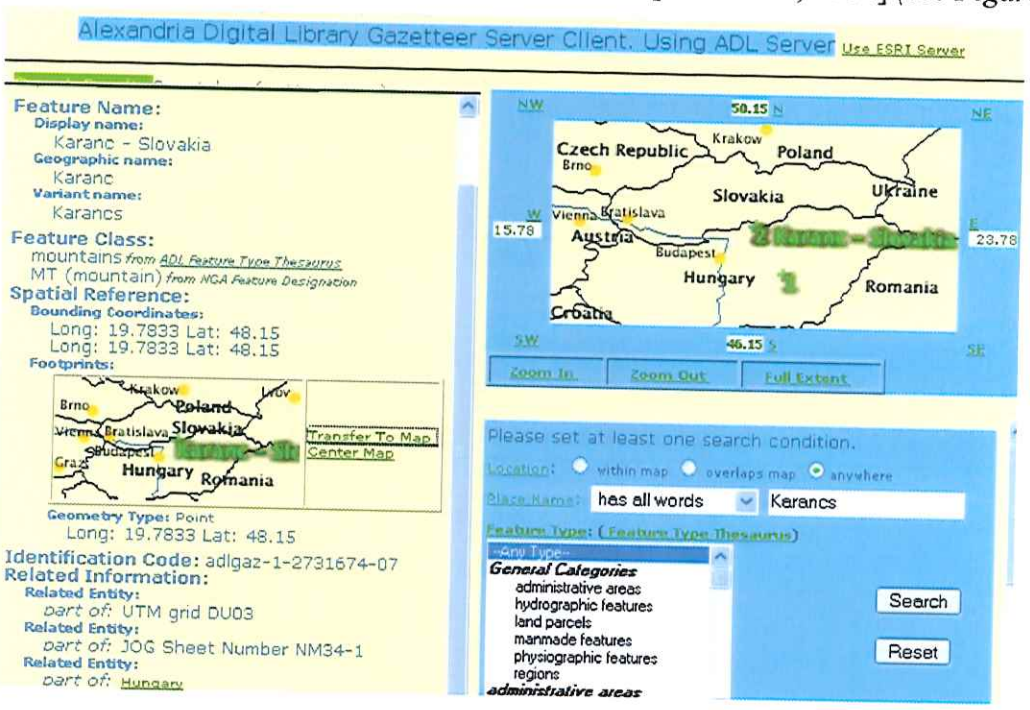


Figure 7: User interface of Alexandria Digital Library Gazetteer

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)

I was looking for 20 Hungarian toponyms in these world gazetteers and recorded the query results. Some of them are inside the current state and others are outside Hungary. I searched names of hills, mountains, plains, historical regions, lakes and rivers *(see Table 1)*.

TOPONYM	GEONAMES	GETTY	ADLG
Mátra	Mátra	Mátra	Matra
	Mátra Hegység		Matra Hegyseg
Kárpátok	Karpátok	Karpátok	Karpatok
	Carpates	Carpates	Carpates
	Carpathian Mountains	Carpathian Mountains	Carpathian mountains
	Carpathians	Carpathians	Carpathians
	Carpați, Munții	Carpati, Muntii	Carpati, Muntii
	Carpații	Karpaten	Carpatii
	Karpaten	Karpaty	Karpaten
	Karpaty		Karpaty
Göcsej	Göcsej	Göcsej	Gocsej
Karancs	—	—	Karanc (Slovakia!)
Alföld	Alföld	Alföld	Alfold
	Nagy Magyar Alföld		Nagy Magyar Alfold
	Al'fël'd		Al'fel'd
	Bol'shaya Sredne-Dunayskaya		Bol'shaya Sredne-Dunayskaya
	Great Alfold		Great Alfold
	Great Hungarian Plain		Great Hungarian Plain
	Great Plain		Great Plain
	Hungarian Plain		Hungarian Plain
Kisalföld	Little Hungarian Plain	—	Kis Alfold
	Little Plain		Little Alfold
	Upper Hungarian Plain		Little Hungarian Plain
			Little Plain
			Upper Hungarian Plain
Csallóköz	Csallóköz	—	Csallokoz
	Ostrov		Ostrov
	Great Schütt		Great Schutt
	Veľký Ostrov Žitný		Velky Ostrov Zitny
	Veľký Žitný Ostrov		Velky Zitny Ostrov
Erdély	Erdély	Erdély	Erdely
	Transilvania	Transilvania	Transilvania
	Ardeal	Siebenbürgen	Ardeal
	Siebenbürgen	Transylvania	Siebenburgen
	Transilvanie	Transylvanie	Transilvanie
	Transilvaniei	Transylvanië	Transilvaniei
	Transylvania		Transylvania
Felvidék	—	—	—
Jászság	Jazygia	—	Jazygia
Fertő	Fertő	Ferto	Ferto
	Fertő-tó	Ferto tó	Ferto-to
	Fertő Tói	Fertoto	Ferto Toi
	Fertő Tava		Ferto Tava
	Neusiedler See	Neusiedler See	Neusiedler See
	Neusiedler Lake	Neusiedler Lake	Neusiedler Lake
Ipoly	Ipoly	Ipoly	Ipoly
	Ipel	Ipel'	Ipel'
	Eipel	Ipel	Eipel
Küküllő	—	—	—

Table 1: Hungarian toponyms in world gazetteers

The table shows some of the results of my investigation. In the first column are the Hungarian toponyms and in the other columns are the name forms in the world gazetteers. I signed the Hungarian form with correct spelling and form 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, Deliblát) are missing [Guszlev and Lukács, 2006].

4. Database of Hungarian Geographical Names (MFNA)

4. 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. MFNA contains toponymic information of the Carpathian Basin, entries also have temporal relations. The gazetteer is suitable for medium and lower scales.

- 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
  - ...

4. 2. Extensions and scale

The gazetteer will cover the **Carpathian Region** which includes the Carpathians and the Carpathian Basin *(see Figure 8)*. 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.

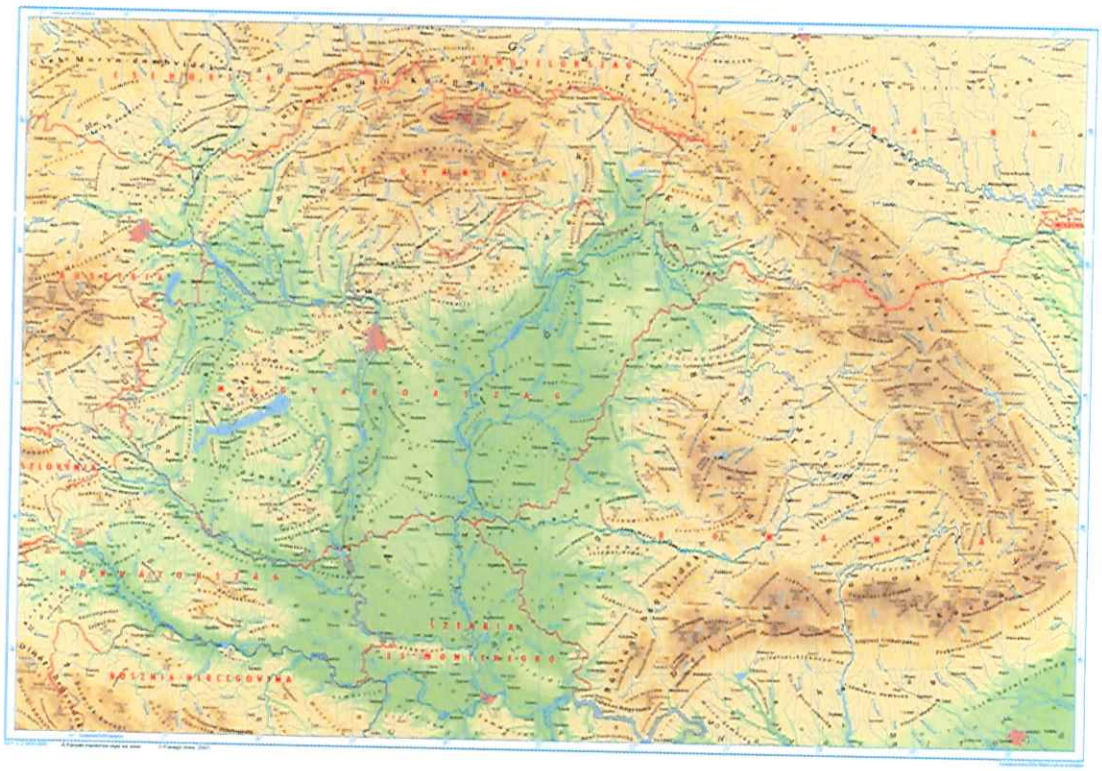


Figure 8: The Carpathian Region, area of the gazetteer [Faragó, 2005]

Temporal scope of the MFNA is from the middle ages to the present days. Analyzing historical geography is always difficult because of the lack of reliable data. Before the 19th century there are few maps or other spatial data. 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, because of political events there were significant reconstructions in the Hungarian administrative system and borderlines of units were changed largely.

Further temporal information is recorded as attribute data: changes of official names, number of population, seats of the counties, etc. The first known occurrence of landscape

region names is also stored as attribute data. The first version contains no temporal data for relief and hydrography.

4. 3. Contents and categories of names

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

4. 3. 1. Map data (in shape files) 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)

Relief is not implemented for name storing, but for example for analysis of correlation between names and altitudes. The other types of map elements (marked points, rivers, main watercourses, lakes, administrative and landscape region borders) are spatial reference objects for the toponyms.

The MFNA project relates to an other 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].

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]

Most 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. 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.) Further advantages are described in chapter 5.4.

4. 3. 2. 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. Μπελογιαννηω=Beloiannisz, a village with Greek inhabitants in Hungary). Character problems can be eliminated by using Unicode characters in the gazetteer.

Object type is the most important information of a geographic place and a required basic data for gazetteer entries. As I mentioned before, compiling 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.

geographical feature	relief, region	mountain, hill
		plain
		historical, ethnographical region
		continent
		peninsula
		island
		undersea feature
		desert
		forest
		land
		other
	hydrography	ocean
		sea
		natural watercourse
		artificial watercourse
		natural lake
		natural salt lake
		artificial lake
		glacier
		bay
		sound
		coast
		well
		other
	geographic point	peak
		pass
		volcano
		point, promontory
		cave
		other
man-made feature	administrative unit	nuts 0
		NUTS 1
		NUTS 2
		NUTS 3
		NUTS 4
		NUTS 5
		part of a settlement
		populated place in the periphery
	transportation element	
	public area	street
		square
		cemetery
		bridge
		...
		other
	man-made object	building
		historical monument
		ruin
		castle
		church
		monument
		other

Table 2: Object types of a gazetteer (MFNA types with peach color)

MFNA has just a few types because of its relatively small scale and content but it could be elaborated in the future *(see Table 2)*.

There are traditions for visualization of different **name categories** in Hungarian cartography. Each name category has unique graphical and typographical attributes in the legend (font color; typeface: fonts with or without serifs; font characteristics: Bold, Italic, etc.). Therefore, reader gets information of the object type even if there is no generic term (hill, county) in the geographical name.

The names of a gazetteer could be divided into the following main cartographical name types *(see Table 3)*:

settlement names	names of settlements with local government
	names of settlement parts
names of administrative units	names of states
	names of administrative units under state level
	names of (managed) protected areas
names of hydrographic elements	lake names
	river names
	sea names
names of geographical points	names of peaks
	names of passes
	names of points and promontories
landscape region names	names of mountains and hills
	names of plains and basins
	names of historical and ethnographical regions
	names of islands and peninsuli
	undersea feature names
land names (parts of former estates, forests etc.)	
streetnames	

Table 3: Cartographical name-types (MFNA types with peach color)

Object types and cartographical name types can be joined to each other unequivocally, because cartographic name types are just collections of graphical and typographical attributes of object types in map-making practice.

4. 4. 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.

4. 5. Database structure

MFNA has a simple **relational database** structure *(see Figure 9)*. 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.

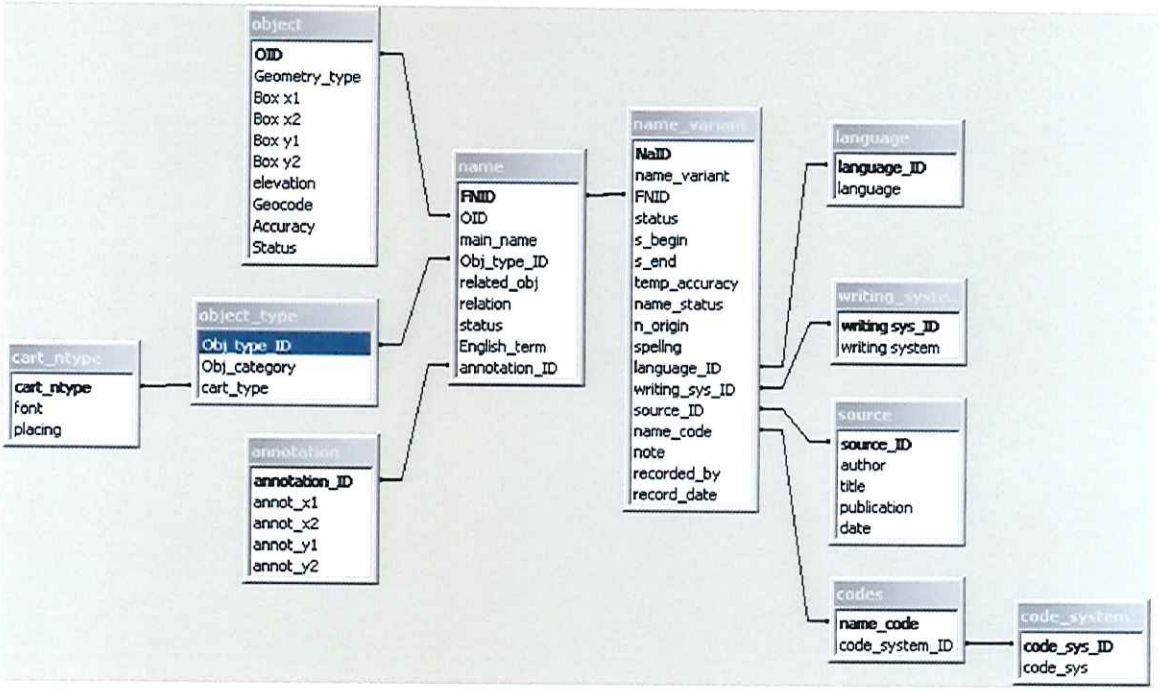


Figure 9: Data structure in the MFNA

Name variants (and not the primary names) are unique elements in the MFNA, therefore 'Name variants' is the main table of the database. The main table contains attributes applying specially to the name form: identifier (ID) of the main name, spelling, language,

origin etc. (Main name is a prior name in the name variants, it is the official name if there is any, but any other name variant also can be chosen.) 'Name variant' table is linked to the 'Name' table which contains general information of the main name of the feature: identifier of the related geometrical object, the object type, the recommended English term, etc. ' Object' table also contain identifier numbers of the map objects as key field, connecting spatial and attribute data.

MFNA is compiled primarily for cartographic purposes, therefore cartographic representation forms of names (as annotations) are included in the database. Other tables contain identifiers of the attributes of names.

4. 6. Functions

The MFNA has the tools of traditional gazetteers (data-storing, simple queries, cartographic visualization of the toponyms). Additional functions in a GIS gazetteer are the complex queries, the temporal and spatial analyses and visualization of query results (also by attributes). The map in the system could play a main role in spatial analyses of the toponymic phenomena beyond its base visualization function.

Various spatial analyses on the gazetteer elements can result new information.

- **Queries**
    - **SQL queries by attributes, e.g.**

      SELECT FROM names WHERE ('Type' = 'falunev' OR 'Type' = 'varosnev') AND 'population' >1000 and 'population' < 5000 AND 'language' = 'szlovak'

      (results of the query: names of settlements which have Slovakian name variants and have population of 1000–5000)

      Similar tasks could be for example:
        - sorting medieval names or names in foreign languages
        - sorting native and artificial toponyms
        - temporal analyses: sorting new and vanished toponyms
        - analyses and visualization of toponymic phenomena (changes of name forms, characters, problems of usage, etc.)
    - **Searching by spatial location**: retrieval of names and other attribute data of the selected map elements.

- **Overlapping**: Spatial relations between

- **point and polyline**: e.g. searching for settlement names along a river
- **point in polygon, and polyline in polygon**: e.g. searching for connections between name origin of a landscape region's peaks or watercourses
- **polygon in polygon**: connections between the names and borders of counties and landscape regions; detecting the movement of landscape region names, etc.

- **Composing buffer zones**: linguistic researches, e.g. searching for the maximum distance where names with a river name (Duna = Danube) as a correlation term occur;
  or asking query of the settlements with Slovakian variant names in 100 km from Békéscsaba (which is the center of an area of Slovakian minority group in Hungary)

- **Network analyses**: e.g. listing right and left side branches of a river; or looking for watercourses which belong to the same water network system, etc. (Obviously, for network analysis the lines of watercourses should also be processed as a network)

4. 7. 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 10)*.

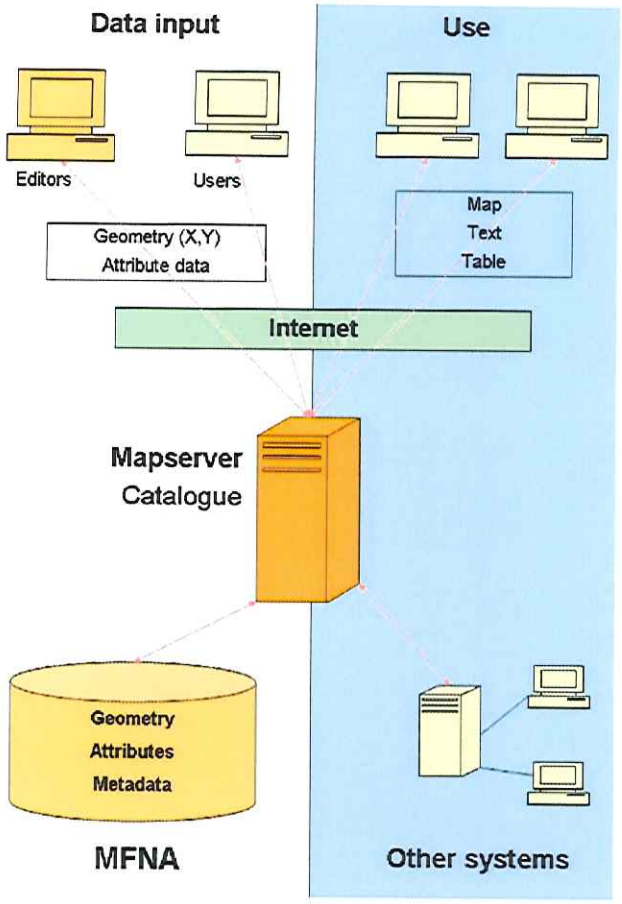


Figure 10: The MFNA system

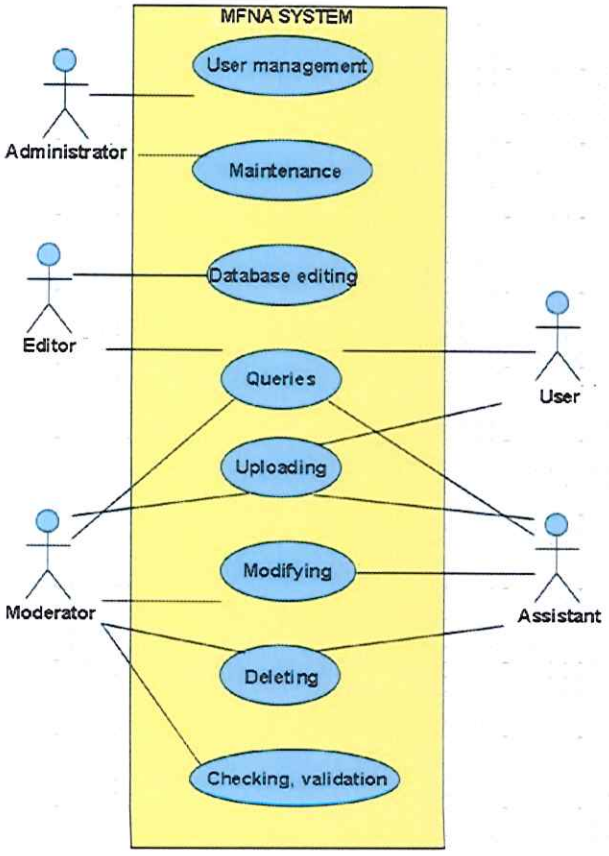


Figure 11: Use cases of MFNA system

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 11)*.

4. 8. Connections to other systems

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.

5. Pilot-project

I have compiled a sample database as a pilot-project to examine available data sources and problems during the integration process. Functions could also be tested in a sample database.

I was looking for a trans-border landscape region of medium size so I chose the northern part of Transdanubian Region: parts of the Kisalföld (Little Alföld) Region and the Transdanubian Mountains *(see Figure 12)*.

The sample gazetteer contains current administrative units (NUTS, Nomenclature des Unités Territoriales Statistiques) and settlements, hydrographic elements and simple borderlines of landscape regions of the LRSCPR. I have implemented the borders of comitats of the medieval Hungarian Kingdom as additional historical layer.

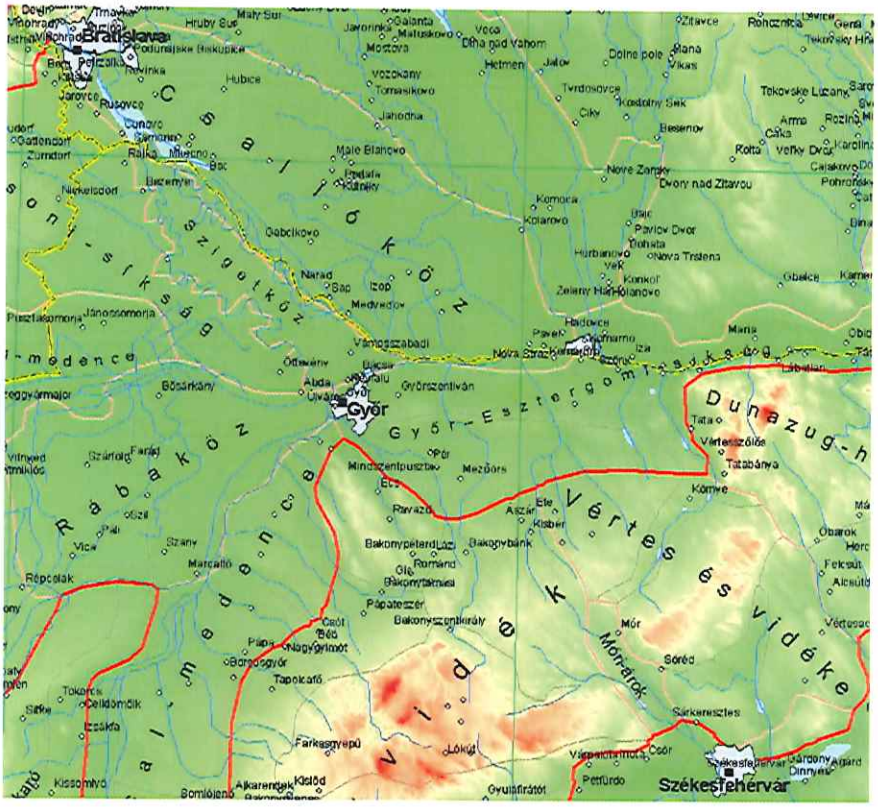


Figure 12: Sample area of the gazetteer

5. 1. Software

The name-database is developed in ArcGIS 9.1 system of ESRI. I have used ArcMap, ArcToolbox and ArcCatalog applications to integrate and process data.

For data-collecting, storing and editing additional software were also used during the pilot-project:

- Corel Photo-Paint 9.0: raster-based image editing program with crop, scale, color-adjustment, etc. tools
- Corel Draw 9.0; and
- Macromedia Freehand 10: vector-based drawing programs for general graphic design
- Able Software R2V for Windows: raster to vector conversion software
- Microsoft Excel: spreadsheet software
- Microsoft Access: database manager software
- Mapinfo Professional 7.0: GIS software

5. 2. Data sources

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. Spatial precision of a map in smaller scale is enough for a gazetteer if the topology of elements is correct. Some of the attributes e.g. name origin or recommended English terms were not available in digital format, therefore I needed to collect and digitize these data and even make agreements with special data owners.

Data policy is not very favorable in Hungary. Public data is expensive and hardly accessible. The digital FNT of FÖMI is suitable for map-designing in large scale but the gazetteer is not freely available and there is no on-line service of it. In the MFNA project the low cost was an important factor point because of financing problems. For this reason we tried to use our own maps and datasets, or free sources as far as possible. In this way, the MFNA can be the base of an on-line gazetteer service without copyright problems in the future.

5. 2.1. Relief

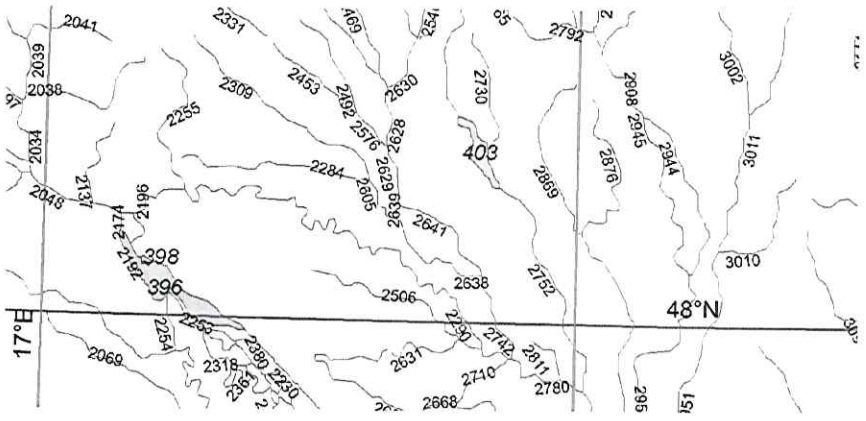
The MFNA gazetteer is 2 dimensional. Altitude data is stored as attribute data for the marked geographical points. The relief is visualized as hypsometric relief, in raster format. The source of data is the SRTM dataset. SRTM (Shuttle Radar Topography

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].

5. 2. 2. Hydrography

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), in Macromedia Freehand (FH) format. Freehand is drawing software for graphic design with general drawing tools. Freehand can handle both vector and raster graphics, objects can be separated in different layers. The problem is that georeferencing is not supported in Freehand and map elements are stored as simple vector graphics with drawing coordinates.

The names were gathered with the help of students of the Department of Cartography and Geoinformatics of Eötvös Loránd University (Budapest). Polyline and polygon objects of the watercourses and lakes were created in ArcGIS shape format by the map mentioned before, and an Object ID (OID) was assigned to the elements. Each student had a part of the printed map of the hydrographic features and their OIDs. They were examining old topographic maps, atlases and some sources on the Web for the names and they have recorded the names and OIDs in an Excel table. In this way they had to use only spreadsheet but no GIS software. We asked them to look for names in other languages also in the case of trans-border rivers *(see Figure 13)*.



térképszám	típus	folyószakasz /tó száma	magyar név	hivatalos nyelvű név	angol név	a név forrása	megjegyzés
12	folyó	2908	Nyitra			Faragó-Hibernia-Szalay-Szarvas-Tóth: Világatlasz	
12	folyó	2909	Nyitra			Faragó-Hibernia-Szalay-Szarvas-Tóth: Világatlasz	
12	folyó	2912					
12	folyó	2939	Visnya			Faragó-Hibernia-Szalay-Szarvas-Tóth: Világatlasz	
12	folyó	2944	Nyitra	Nitra		Faragó-Hibernia-Szalay-Szarvas-Tóth: Világatlasz	
12	folyó	2945	Öreg-Nyitra	Stará Nitra		Topográf: Szlovákia autóatlasz	
12	folyó	2951	Zsitva	Zitava		Topográf: Szlovákia autóatlasz	
12	folyó	2952					nem eldönthető,
12	folyó	2953					nem eldönthető,
12	folyó	2954					
12	folyó	2961		Vlára		Topográf: Szlovákia autóatlasz	
12	folyó	2967	Vág	Váh		Faragó-Hibernia-Szalay-Szarvas-Tóth: Világatlasz	
12	folyó	2976					

Figure 13: Parts of the map and the attribute table of hydrographical names

5. 2. 3. Marked geographical points

Some of peaks and passes of the mountain regions are also represented. Data source of these points was the same Freehand map file of the Carpathian Basin. The spatial location of the points has been modified manually according to the relief.

5. 2. 4. Administrative units and settlements

Current NUTS borders are processed in the gazetteer. Dataset of Hungarian NUTS is provided by FÖMI and KSH (Central Statistical Agency), and developed by the GeoX. The dataset in small scale can be downloaded in ArcGIS format from the MAKEHA site (www.makeha.hu). ESRI's sample dataset also was used as source of settlements. Settlements of Hungary are stored as polygons (NUTS 5 level) in the MAKEHA dataset and as points in the ESRI's dataset.

Borderlines of former administrative units are from a historical thematic map file of Hungary in 1910 (edited by the Institute of History of the Hungarian Academy of Sciences) *(see Figure 14)*.

This map is in CorelDraw vector graphic format (CDR) which contains map elements as drawings, similarly to Freehand. The scale of the map is 1: 400 000. Borders of the counties and districts are represented as simple curves (not as polygons) and there are some layers of additional administrative data (seats, etc.). The county names are stored in the CDR file as labels, but the names of the districts had to be looked up in the Atlas and Gazetteer of Historic Hungary 1914 [Talma, 2001].

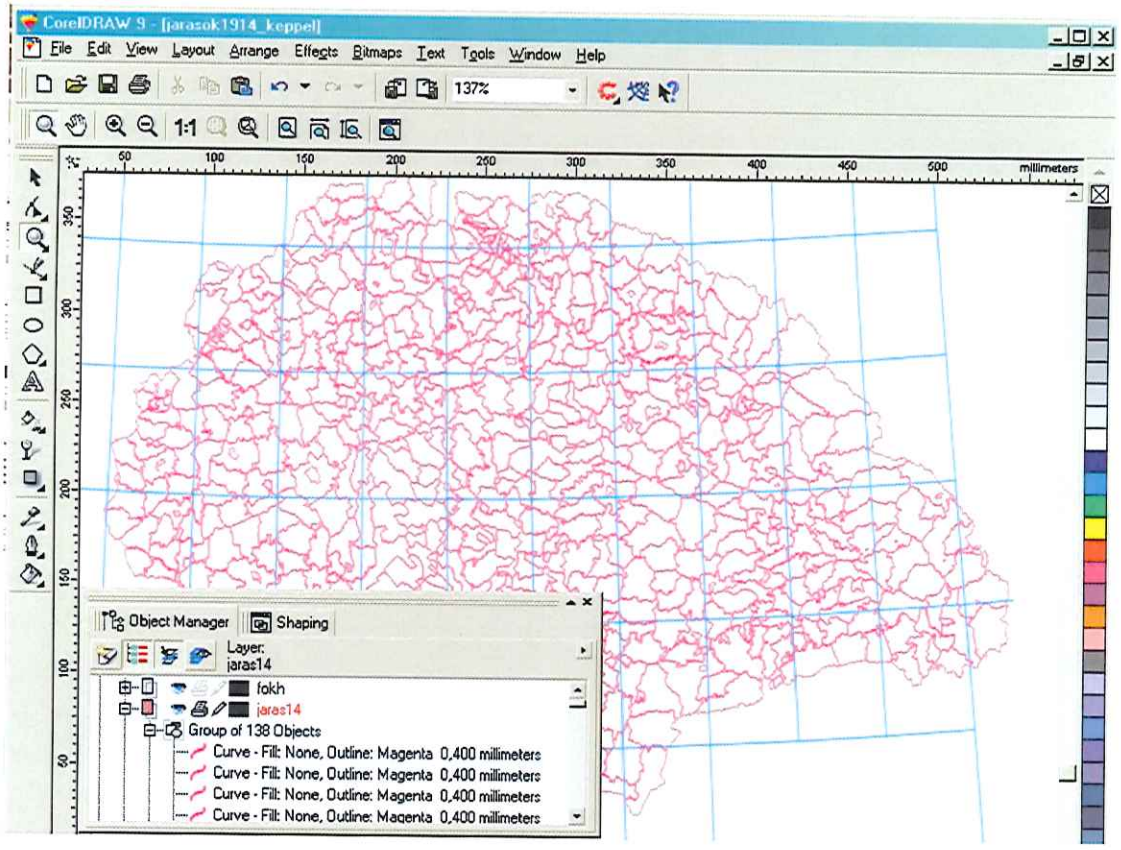


Figure 14: Source map of former administrative units (in CDR format)

5. 2. 5. Landscape regions

Modern landscape region borders are represented on maps but there is no map data of the systems made before the World War II. The borderlines of regions which are found in the LRSCPR follow the ones in the academic Hungarian landscape region system (names and hierarchy are different sometimes) inside the current state. Actually, LRSCPR is extended form of the academic system to the Carpathian region. We also have got the sketch map in CorelDraw and database of regions in Excel table format. The relatively small scale of this map causes not much impreciseness because the borders of landscape regions are not sharp by nature *(see Figure 15).*

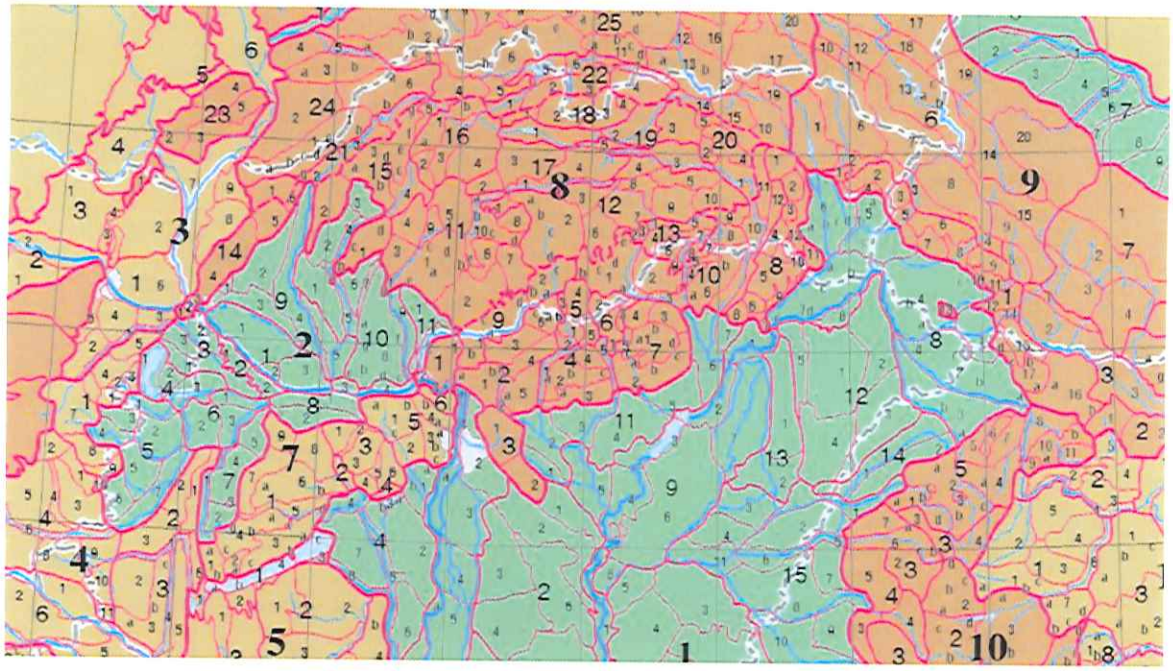


Figure 15: Source map of the landscape regions (in CDR format)

5. 2. 6. Attributes

Most attribute data of the sample gazetteer has been generated automatically during the integration process. The source maps contain different object types on different layers; multilingual name variants have been recorded in tables separately, etc.

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 database 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 simple name list with pairs of Hungarian and English terms, but without any spatial references *(see Figure 16)*.

Abaúj Abaúj
Agár-tető Agár Peak
Ágasvár Ágasvár
Aggteleki-karszt Aggtelek Karst
Alföld Great Hungarian Plain
Alpokalja Alpokalja
Alsó-Kiskunság Lower Kiskunság
Alsó-Válicka Alsó-Válicka
Általér Általér Brook
Bácska Bácska
Badacsony Badacsony
Bakony Bakony
Bakonyalja Bakonyalja
Bakonyerdő Bakonyerdő
Baláta-tó Baláta Lake
Balaton Lake Balaton
Balaton-felvidék Balaton Uplands
Balota Balota
Bán Bán
Baranya Baranya
Baranyai-dombság Baranya Hills
Baranyai-Hegyhát Baranyai-Hegyhát
Bársonyos Bársonyos
Béci-hegy Béci Hill

Figure 16: Part of the Gazetteer of English Names for Topographical Objects of Hungary

5. 3. Data integration

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.

5. 3. 1. SRTM data

Elevation model generated from SRTM dataset in raster format is downloadable from the SRTM website. The relief data can be added to ArcGIS directly *(see Figure 17)*.

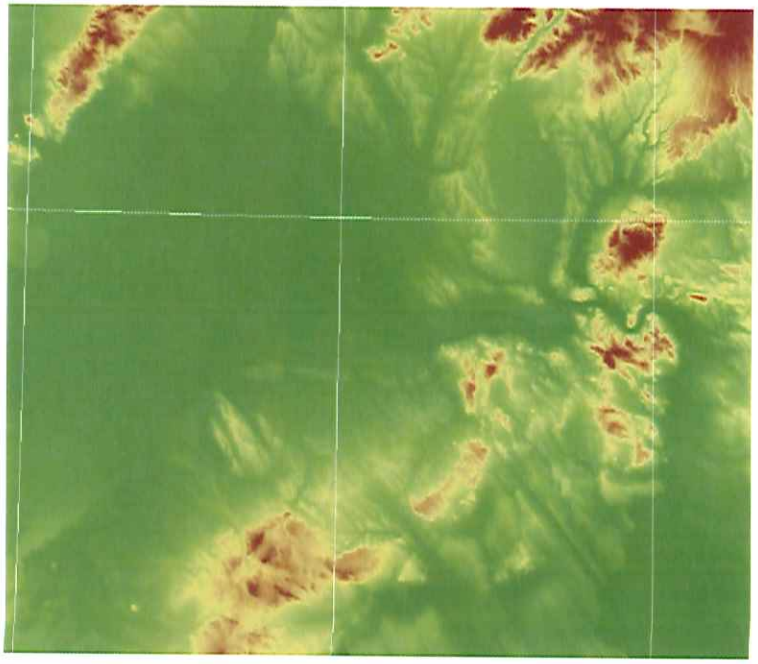


Figure 17: Relief from the SRTM dataset

5. 3. 2. Shape files

Some databases used in the sample gazetteer (polygons of the current administrative units and sample ArcGIS databases of the ESRI) are available directly in ESRI shape (SHP) format. These datasets have been added into ArcMap and transformed into EOV projection by the project-tools of ArcToolbox.

5. 3. 3. Graphic map files

Vector maps had to be converted to raster format for the further integration process. Theoretically, graphical software could export DXF format, which could be imported into ArcGIS 9.1. But in practice it did not work because ArcMap was not able to handle the large DXF file of polylines of the hydrographic elements (polylines were broken into small parts with two nodes).

For this reason additional software and raster format was used: R2V software can vectorize black&white (1-bit) raster drawings automatically and is able to export objects as shape file. The integration method consisted the following procedures:

1. Preparing the map for automatic vectorization in drawing software (CorelDraw, Freehand), and convert it to 1-bit image.
2. Exporting 1bit raster map in TIF format.
3. Vectorizing the raster image automatically in R2V software.
4. Exporting the map in shape format (with polyline objects also in case of polygons).
5. Importing polyline shape into Mapinfo software (in case of polygon objects)
6. Enclosing polylines to polygons in Mapinfo.
7. Exporting polygon shape from Mapinfo.
8. Adding the shape file to an ArcGIS workspace.
9. Georeferencing.
10. Identifying objects.

For vectorizing 1-bit raster maps by R2V software, high drawing quality was needed. It was worth to have efforts to clear line-drawing because too thin or too thick lines, opened polygons and noise caused faults in the vectorized map. Correction of polygons in

ArcMap would have been more difficult than preparing the map in drawing software *(see Figure 18)*.

Mapinfo is GIS software; it was used because of the lack of a simple tool of ArcMap: closing the polygons is not available in ArcMap, only in ArcInfo. Polyline shapes could have been added into ArcMap directly.

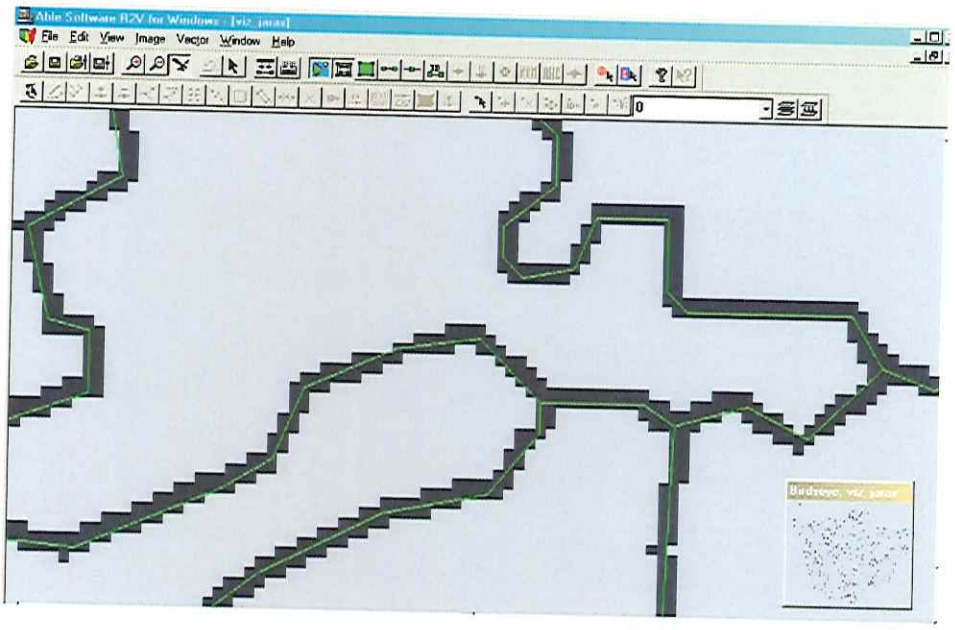


Figure 18: Vectorizing in R2V Software

Georeferencing of the simple vector maps was slow. Required spatial data is recorded as a grid in these graphic files. The simplest way of adjusting was to georeference the whole geographical source map as raster image by affine transformation with grid nodes. And then shapes (originated form the same map files) were adjusted to the georeferenced raster map. Spatial relationships between objects of different types and accuracy of the source map (CDR or FH) have been fully preserved during the integration process by this method.

ArcMap provides various spatial adjustment tools: transformations and rubber-sheet mode. I have tested different processes and I found out that the affine transformation generates the most satisfying results. Using other modes caused bigger distortions in the geometry of the map elements *(see Figure 19)*.

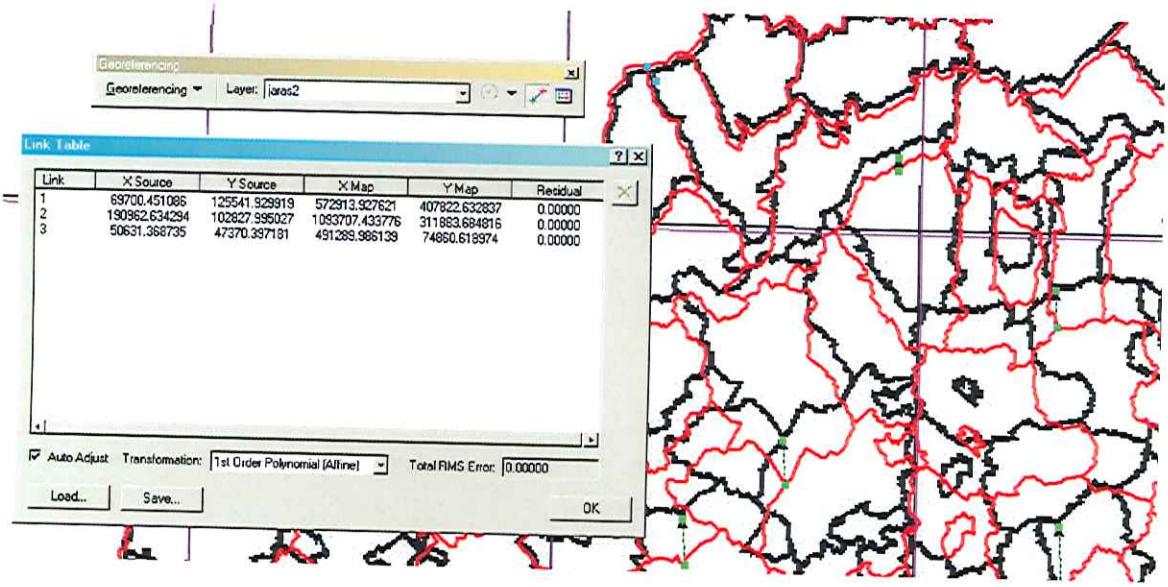


Figure 19: Georeferencing the former districts of Hungary

After implementing the lowest level of a hierarchical dataset, **Spatial Analysis** tools of ArcMap support creating the higher levels. LRSCPR consists of 3 levels: minor, medium and major regions. Child regions are stored in database as attribute data. Therefore medium and major regions were created by dissolving minor regions by attribute data *(see Table 4)*.

nagytáj id.	*középtáj id.*	*kistáj id.*	*magyar név*	*szlovák név*
8	**5**		**Medvesvidék**	**Medvešská vrchovina**
8	5	1	Medves-fennsík	
8	5	2	Kazári-dombság	
8	5	3	Füleki-medence	Fiľakovská brázda
8	5	4	Ajnácskői-hegység	Hajnáčska vrchovina
8	5	5	Zabari-dombság	Baštianska kotlina

Table 4: Dissolving medium regions by OIDs of child minor regions

Comitats of the Medieval Hungary were generated similarly by dissolving the polygons of the districts.

For these processes unique **identification** of objects was necessary. Identifying geographical features (counties, watercourses, and landscape regions) was the most time-consuming task during the gazetteer compilation. Graphic software stores no identifier (e.g. name) or attribute data for objects. For this reason objects (more than 800 regions and 600 districts in the whole Carpathian Basin) were identified one by one.

5. 3. 4. Tables

In ArcGIS 9.1 attribute data can be stored in Attribute Tables of the layers or in joined Access databases. 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 additional external databases can be joined by key fields.

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

Figure 20: Attribute table of former administrative system (MS Access database)

Most of data is stored in Excel tables but they must be imported or joined to Access database, because ArcGIS 9.1 is not be able to handle Excel tables directly (ArcGIS 9.2 can do it) *(see Figure 20)*.

External tables can be joined to the ArcMap layers as 'related' or 'joined' tables. Joining is used when tables are connected by one-to-one (1:1) relationship. In this case attributes from a table are appended to the other. Relating tables defines a relation between two tables in case of many-to-many (n:n) relations. Attributes of related tables are not appended to each other. In MFNA related tables were used because of the n:n relations.

5. 3. 5. Simple text sources

Data of name ethimology was digitized from the paper form FNESZ and recorded in Excel tables of the objects. Elaboration of the name ethimology in the gazetteer needs further sources in the future.

Gazetteer of Hungarian-English toponyms was converted to Excel table format (from simple text). Object IDs of elements stored in MFNA have been added as a key-field to this English Names table.

5. 4. Annotations

Objects could be labeled in ArcMap by any attribute data. Labeling is the simplest method to visualize geographical names of the stored features. Labels are generated from the database and a general style can be saved for them in a workspace (but not in the shape-file), but they have no individual text characteristics. Disadvantage of labeling is few possibility of placing and typography, and that labels cannot be analyzed spatially. Labeling is usable for visualizing of names with simple cartographic rules: point objects (settlements, geographical points) polylines (rivers) and some of polygons (administrative units). In cartography there are traditions for visualizing landscape region names, which are not included in ArcGIS labeling tools. Names of geographical regions are used to fitted to a curve, showing orientation and extension of the related feature (e.g. along ridge of a mountain).

To solve these problems, labels are worth to convert to **annotation features**. Annotations vary in storing and placing from simple labels. Annotations have a separate feature class in ArcMap. Each annotation has data of spatial location and attributes of visualization (font characteristics, color, etc.) *(see Figure 21)*. For creating annotation feature class, a geodatabase should have been built for the datasets of the gazetteer *(see Figure 22)*.

Dunazug-hg.	BUDAPEST	Székesfehérvár	**POZSONY**	NÓGRÁD
Spacing	Capital letters	Font size	Bold	Extended
M á t r a	Szeged	Karancslapujtő	BRATISLAVA	NÓGRÁD
Salgótarján	Duna	Alföld	Salgóbánya	**AUSZTRIA**
Underline	Color	Font type	Italic	Texture
Pásztó	Duna–Ipoly NP	Palócföld	*Inászó*	AUSZTRIA

Figure 21: Labels with various graphical and typographical characteristics

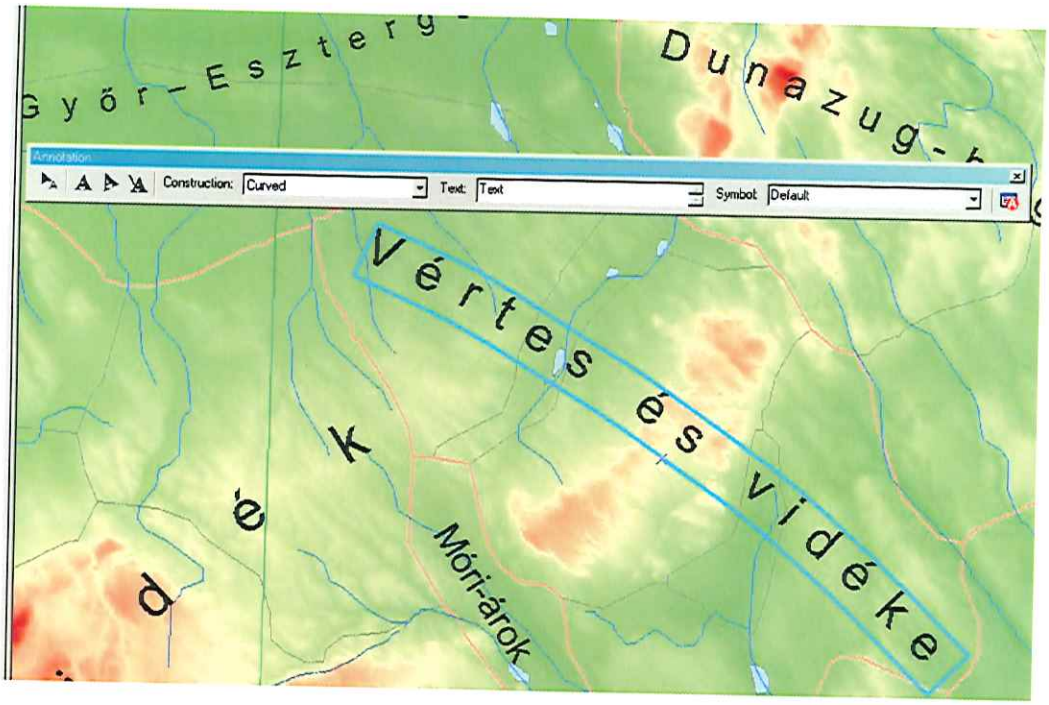


Figure 22: Formatting annotations in ArcMap

5. 5. Tests

I executed some tests to check if gazetteer functions work. The most frequent tasks in gazetteer usage are asking queries and visualizing results on map.

For example, looking for the answer of the question: 'which landscape region names contain the word "-hegység" (mountains) as a generic term?' can be a part of e.g. geographic or linguistic researches. Spatial distribution of regions with '-hegység' term can give information of geography (relief and landscape) and ethnography (present or sometime Hungarian inhabitants) of an area *(see Figure 23)*.

The MFNA contains Hungarian name variants in a separate field in the attribute table of the landscape regions. An SQL selection can be processed by the following command:

RIGHT([Név1],8) = '-hegység'

This expression selects the Hungarian names which end with '-hegység' string from the database. Selected objects can be visualized on map and can be converted into a separate layer and table.

Figure 23: Cartographical visualization of query results in ArcMap

5. 6. Metadata

Metadata gives information of spatial and temporal extensions, spatial references, sources, accuracy, usage, copyrights, etc. of the datasets. Metadata is important also in cataloguing and used in data sharing as documentation of datasets.

Metadata of the MFNA gazetteer has been created in ArcCatalog, which provides metadata forms following metadata standards (e.g. ISO, FGDC, and XML). Some of data is gathered automatically by ArcCatalog: reference system, projection, coordinates of the bounding box of the dataset, object types, and attribute fields. Other data should be given by the editor of the metadata sheet, e.g. description, purpose, originator, contacts, constraints on access, etc. Metadata can be made for any datasets (feature classes, shape files, geodatabase, etc.). I have created metadata for the whole gazetteer as an example *(see Figure 24)*.

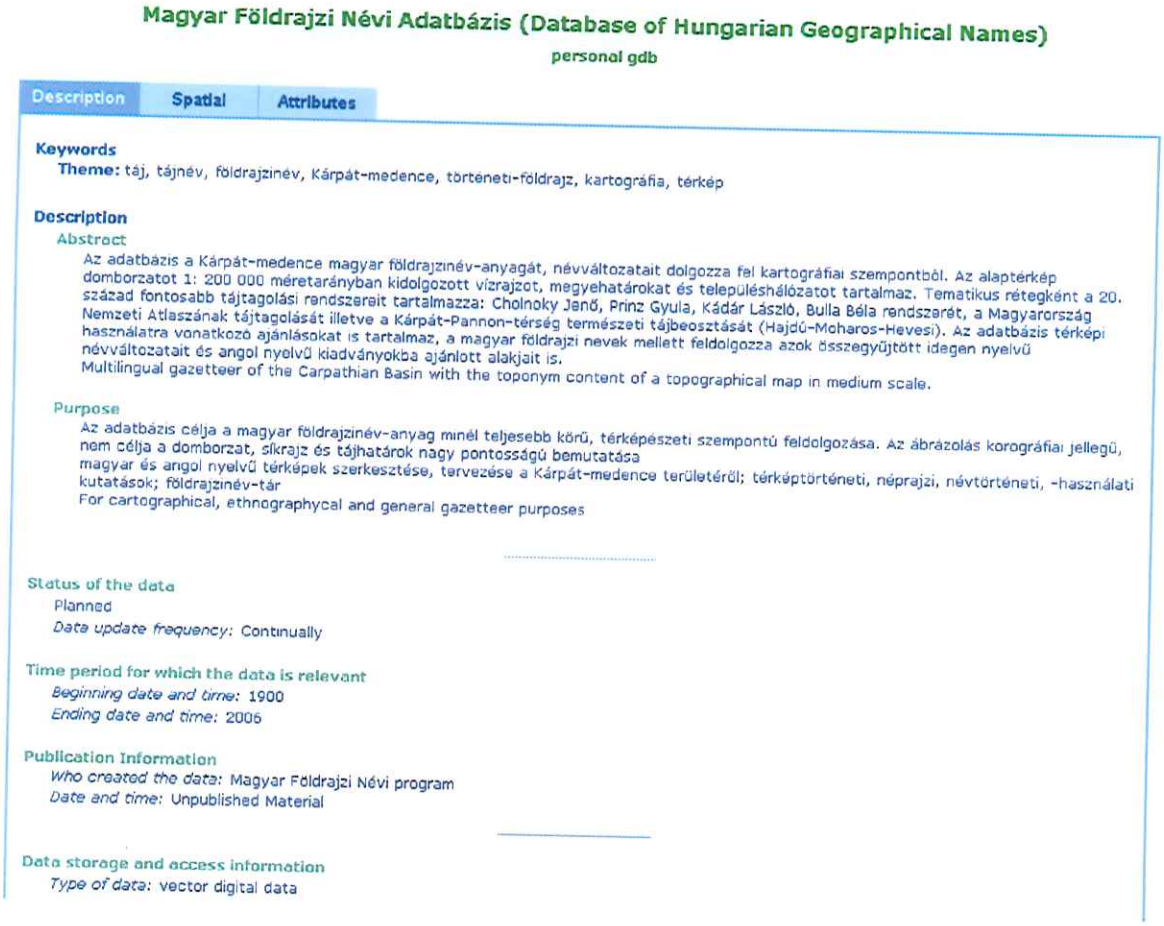


Figure 24: Metadata of MFNA

5. 7. Further improvements

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.

The sample gazetteer database can be used by the tools provided in ArcMap software: SQL queries, spatial analysis, and visualization tools. For a free gazetteer-service open-source applications should be made with a user-friend interface. System infrastructure for the service also must be established.

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. Borders of landscape regions could be refined by using buffer zones depending on geographical factors or using fuzzy logic.

Thesaurus of the object types can be more detailed. In the future, ontology will be developed for the gazetteer. Ontologies define concepts by attributes and relationships to other concepts of the system. Use of ontologies supports knowledge-sharing in semantic web.

Database structure gives an opportunity for storing links to records of other gazetteers or datasets.

6. Conclusions

In my thesis, I made an overview of the functions and characteristics of gazetteers from paper form to the GIS gazetteers. For a gazetteer research, a sample database was compiled in ArcGIS for cartographical purposes. I have collected and integrated several data sources and created database structure to support name-queries and cartographic representation. The compilation of the database gave me opportunity to study GIS tools in managing name datasets and other spatial data. I examined also the cartographical utilities of GIS software.

ArcGIS gives many tools for integrating, geoprocessing, analyzing and visualizing spatial data. Additional attribute tables can be joined to the objects from external datasets.

However, quality and structure of data sources are very different. Purposes of a database determine preciseness and structure; and there are only a few database compiled specially for gazetteer purposes. For this reason eliminating of database gaps, reclassifying data and improving spatial accuracy are needed.

During the pilot-project ArcGIS software was not enough for integrating data, some additional software were used e.g. Mapinfo and R2V, because some simple and often-used tool is missing from ArcGIS. This can raise the costs of similar GIS gazetteer compilation projects.

In Hungary there are problems with availability of data. Digital datasets are expensive also for research purposes, but the lack of digital data is also a problem. Several data-digitizing projects were executed in last years but special thematic data (e.g. origin of the names) available still in paper form only.

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. I think that MFNA project could take part in developing Hungarian SDI (Spatial Data Infrastructure). There are some similar gazetteer projects in Europe, e.g. Place Name Database (KNAB) of Estonia (http://www.eki.ee/knab), Gazetteer of Spain (http://www.idee.es), etc. The multilingual gazetteer of the Carpathian Basin also could be joined to the European SDIs (ESDI) according to the INSPIRE (Infrastructure for Spatial Information in Europe) 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.

7. References

- ADL Gazetteer Team (2002): Introduction to Digital Gazetteers and their Development Issues
- Rob Atkinson – Jens Fitzke (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
- Földi Ervin (1992): Az ENSZ földrajzinév-egységesítési szakkifejezések szótára, Névtani értesítő, 14. szám, Budapest
- Földrajzinév-tár I. Magyarország fontosabb domborzati- táj- és víznevei, Kartográfiai Vállalat, Budapest 1982.
- Földrajzinév-tár II. Veszprém megye, Kartográfiai Vállalat, Budapest, 1981.
- 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
- Klinghammer István – Papp-Váry Árpád (1997): Tematikus kartográfia, ELTE Eötvös Kiadó, Budapest
- Lilla Lukács (2005): Naming of Geographical Regions on Maps of Hungary in the 20th Century, 21st International Conference on History of Cartography, Budapest
- 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
- 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.
- Digital datasets:
  - SRTM dataset http://www2.jpl.nasa.gov/srtm/
  - Topographical map of Hungary 1: 1 000 000, edited by Imre Faragó
  - Map of administative units of Hungary in 1914., edited by Béla Nagy and Lilla Lukács

8. Appendices

Figures and tables

Abbreviations

ADLG	Alexandria Digital Library Gazetteer
CDR	Corel Draw
DEM	Digital Elevation Model
DXF	Drawing Exchange Format
ELTE	Eötvös Loránd Tudományegyetem, Eötvös Loránd University
EOV	Egységes Országos Vetület, Hungarian Uniform Projection
ESRI	Environmental Systems Research Institute
FGDC	Federal Geographic Data Committee
FH	Freehand
Fnb	Földrajzinév-bizottság, Board on Geographical Names
FNESZ	Földrajzi nevek etimológiai szótára, Ethimological Dictionary of Toponyms
FNT	Földrajzinév-tár, Gazetteer of Hungary
FÖMI	Földmérési és Távérzékelési Intézet, Institute of Geodesy, Cartography and Remote Sensing
GIS	Geoinformation System
GNS	GEOnet Names Server
GTGN	Getty Thesaurus of Geographical Names
HTP	Hungarian Toponymic Program
INSPIRE	Infrastructure for Spatial Information in Europe
ISO	International Organization for Standardization
KSH	Központi Statisztikai Hivatal, Hungarian Central Statistical Office
LRSCPR	Landscape Region System of the Carpathian-Pannonian Region
MEM	Mezőgazdasági és Élelmezésügyi Minisztérium
MFNA	Magyar Földrajzi Névi Adatbázis, Database of Hungarian Geographical Names
MTA	Magyar Tudományos Akadémia, Hungarian Academy of Sciences
NASA	National Aeronautics and Space Administration
NGA	National Geospatial-Intelligence Agency
NUTS	Nomenclature des Unités Territoriales Statistiques
OGC	Open Geospatial Consortium

OID	Object Identifier
R2V	Raster to Vector
SHP	Shape
SQL	Structured Query Language
SRTM	Shuttle Radar Topography Mission
UNGEGN	United Nations Group of Experts on Geographical Names
WGS	Web Gazetteer Service
XML	Extensible Markup Language

Content of the CD

Files of the sample gazetteer datasets are attached on CD.

NEVTAR.mxd is the ArcMap map document which includes:

- EOV grid (shp file)
- relief (raster image)
- hydrographic elements:
    - rivers (shp file)
    - lakes (shp file)
- administrative boundaries:
    - borders of NUTS in Hungary (shp file)
    - comitats and districts in former Hungary (shp file)
- settlements (shp file)
- landscape regions in 3 level (shp files):
    - major regions
    - medium regions
    - minor regions
- annotations for landscape regions
- joined attribute tables for landscape regions, hydrographic elements and administrative units