Evaluation of current international toponymic databases

Ferjan Ormeling
Utrecht University

Abstract

What happens if for your country a national toponymic database is non-existant or not regularly updated? Others decide on your names! For many countries too few names are available in official databases, so those developing international toponymic databases decide on the names from those countries for themselves. National input is needed to prevent this.

Introduction

Naming behaviour is enormously diverse! Sometimes the sounds of a language, or even the words themselves, are dependent on who is speaking or who is being spoken to.The Paleo-Asiatic Chukchi language, spoken in the easternmost peninsula of the Russian Federation, contains a phoneme that is pronounced *k* by men and *ts* by women, and another phoneme pronounced *rk* by men and *tsts* by women. The word for 'walrus' is thus pronounced *kyrky* by men, and *tsytstsy* by women. Needless to say that this is a very useful piece of information for the toponymic interviewer operating in Chukchi territory. In some parts of the world males and females use different names to refer to the same topographical objects. Arnhem Land in Northern Australia is an example. North of Australia, in the isle of Java, the Javanese use different languages, depending on the fact whether they speak to a person from a higher or lower social order. And place names follow this practice also. So there would exist different versions of most place names, to be used at different occasions. Elsewhere, in many nomadic areas, the names that topographers would collect from the local population would depend on the fact whether one or another language group would be passing by. And also if people would be permanently settled in an area, minority and majority language groups might use different name versions. In many areas young people would use other geographical names than older people. Young people are more apt to use slang and older people are more apt to use formal language, and that can be reflected in the place names they use. In some parts of the world the change of the seasons has such a big impact that the topographical objects to be observed would be completely different, like in the Arctic where in winter the sea is frozen and there are names for specific parts of the pack ice. In most parts of the world names are rather permanent attributes but in some parts village names would reflect the name of those in power, and at their demise the village name would change, and reflect the successor. So we see different naming habits, and it would be difficult for outsiders to deal correctly with all those differences.

We have heard in the previous lecture by Mrs Kerfoot about the national toponymic database, and about its benefits. In Europe we are building a European database about which Mr Zaccheddu will talk later, but we developed use cases, in order to find out how gazetteers and

names databases could benefit the economy. We found that – apart from general reference applications - names databases were an essential element in information systems, such as for finding hotels, for emergency quick response applications, as for ambulances and fire brigades. But also for information systems on real estate, and for news applications when the news broadcasters need quick information on the location and pronunciation of a foreign town where an earthquake or man-made disaster happened. For translation services a names information system would also be worthwhile, for the travel branch or for historical research (geoparsing: find all the documents that have a reference to the former Bismarcksburg, the present Kasanga on Lake Tanganyika, or for the former Salisbury, the present Harare.

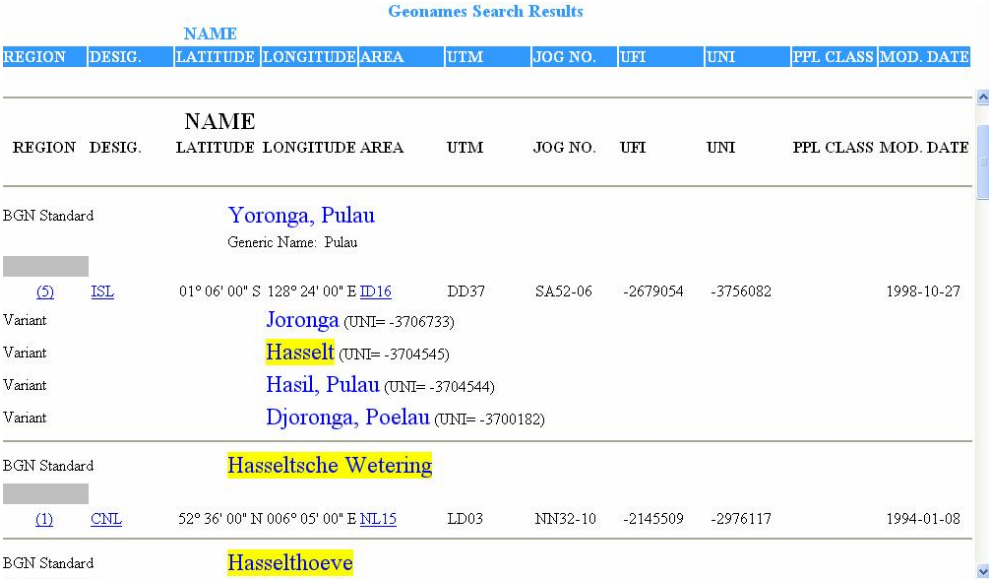
As all the world needs addresses, and wants labels for topographic features like mountains, rivers and cities, also for the new applications sketched in these use cases, these names will find their way into foreign hands anyway, and if you do not provide these names yourself the danger is great that these names would not be rendered correctly. The purpose of this lecture is to find out what international toponymic databases are currently available on the web, and to evaluate them.

We will deal with the following web name servers:

1.GEOnet Names Server (GNS)
2. Geonames
3.Getty Thesaurus
4. Fuzzy Gazetteer
5. Alexandria Digital Library Gazetteer
6. Earthsearch
7. World gazetteer
8. Global Gazetteer Version 2.1 (Fallingrain)
9. World City Name Database (Nonanet)
10. Eurogeonames
11. UN gazetteer

1. GEOnet Names Server (GNS)

The GEOnet Names Server (http://earth-info.nga.mil/gns/html/index.html)
provides access to the database of foreign geographic feature names of the National Geospatial Intelligence Agency (NGA) and the U.S.Board on Geographical Names. 20 000 of its feature names are updated monthly. The data base contains over 4 million features with over 5,5 million names. The coordinate system is WGS84, but coordinates are approximate only. The database also contains variant spellings. GNS is starting to hold native script spellings. There is also information about administrative location and quality. It is indicated on the website how many names are held for specific countries. For Kenya 30 000 names are held. This GEOnet Names Service is also the main source for most other name servers. The source of the GEOnet names server used to be the paper gazetteers the American defense establishment produced for all countries, in the 1960s. Not all of them have been updated to the same degree.



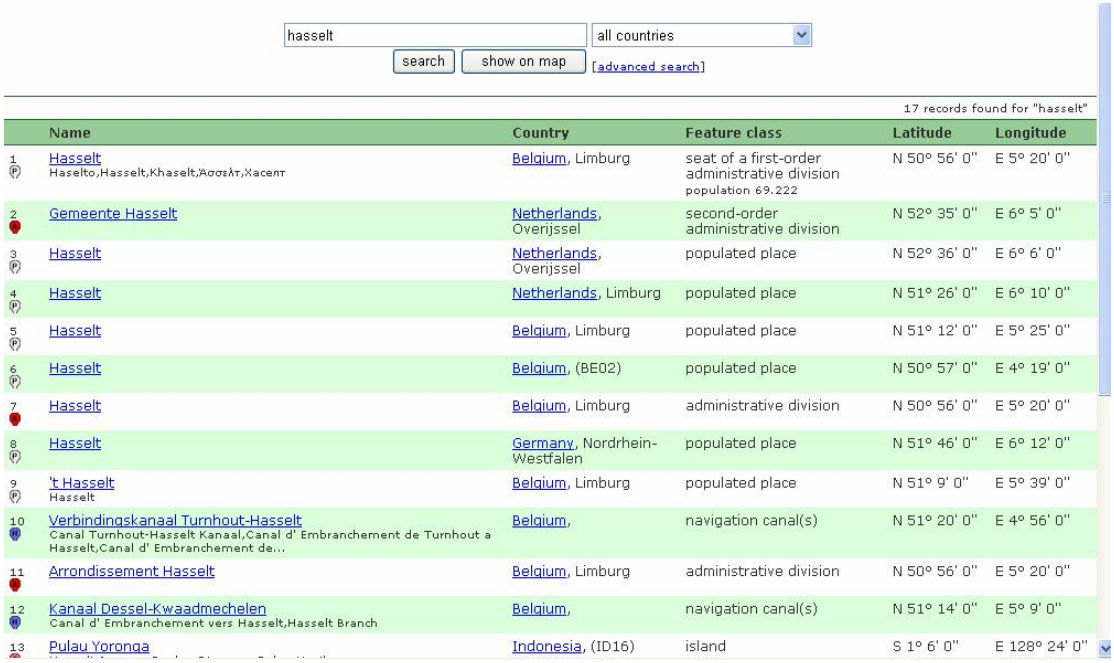
Geonames Search Results

REGION	DESIG.	NAME LATITUDE	LONGITUDE	AREA	UTM	JOG NO.	UFI	UNI	PPL CLASS	MOD. DATE
BGN Standard		Yoronga, Pulau Generic Name: Pulau								
(5)	ISL	01° 06' 00" S	128° 24' 00" E ID16		DD37	SA52-06	-2679054	-3756082		1998-10-27
Variant		Joronga (UNI= -3706733)								
Variant		Hasselt (UNI= -3704545)								
Variant		Hasil, Pulau (UNI= -3704544)								
Variant		Djoronga, Poelau (UNI= -3700182)								
BGN Standard		Hasseltsche Wetering								
(1)	CNL	52° 36' 00" N	006° 05' 00" E NL15		LD03	NN32-10	-2145509	-2976117		1994-01-08
BGN Standard		Hasselthoeve								

Figure 1 - Example from GNS: search for Hasselt

2. Geonames

The GeoNames server (http://www.geonames.org/) has as its sources the GNS server, Wikipedia and gtopo30 (elevation). There are 645 feature codes and 9 feature classes. These mostly are the same codes as those used for GNS. Geonames is a private foundation, and operates with national ambassadors helping out the coordinators.



hasselt | all countries | search | show on map | [advanced search]

17 records found for "hasselt"

	Name	Country	Feature class	Latitude	Longitude
1	Hasselt Haselto,Hasselt,Khaselt,Хасэлт,Хаселт	Belgium, Limburg	seat of a first-order administrative division population 69.222	N 50° 56' 0"	E 5° 20' 0"
2	Gemeente Hasselt	Netherlands, Overijssel	second-order administrative division	N 52° 35' 0"	E 6° 5' 0"
3	Hasselt	Netherlands, Overijssel	populated place	N 52° 36' 0"	E 6° 6' 0"
4	Hasselt	Netherlands, Limburg	populated place	N 51° 26' 0"	E 6° 10' 0"
5	Hasselt	Belgium, Limburg	populated place	N 51° 12' 0"	E 5° 25' 0"
6	Hasselt	Belgium, (BE02)	populated place	N 50° 57' 0"	E 4° 19' 0"
7	Hasselt	Belgium, Limburg	administrative division	N 50° 56' 0"	E 5° 20' 0"
8	Hasselt	Germany, Nordrhein-Westfalen	populated place	N 51° 46' 0"	E 6° 12' 0"
9	't Hasselt Hasselt	Belgium, Limburg	populated place	N 51° 9' 0"	E 5° 39' 0"
10	Verbindingskanaal Turnhout-Hasselt Canal Turnhout-Hasselt Kanaal,Canal d' Embranchement de Turnhout a Hasselt,Canal d' Embranchement de...	Belgium,	navigation canal(s)	N 51° 20' 0"	E 4° 56' 0"
11	Arrondissement Hasselt	Belgium, Limburg	administrative division	N 50° 56' 0"	E 5° 20' 0"
12	Kanaal Dessel-Kwaadmechelen Canal d' Embranchement vers Hasselt,Hasselt Branch	Belgium,	navigation canal(s)	N 51° 14' 0"	E 5° 9' 0"
13	Pulau Yoronga	Indonesia, (ID16)	island	S 1° 6' 0"	E 128° 24' 0"

Figure 2 -Example from a search (for Hasselt) on Geonames. There also is a special service which shows all the named features on imagery provided by Google Earth, with the types of features named colour coded. By comparing this image with the current large scale

topographic map, this would give a good idea of the names information density provided by Geonames.

	Name	country	feature	km to center
1	Oosterhoeve	Netherlands	farm	1.27 km
2	Steenwetering	Netherlands	canal	1.86 km
3	Rechterensgracht	Netherlands	canal	1.86 km
4	Rouveen	Netherlands	populated place	2.8 km
5	Zwartewaterskloosters	Netherlands	populated place	3.03 km
6	Streukel	Netherlands	populated place	3.03 km
7	Kievitsnest	Netherlands	populated place	3.52 km
8	Hasselt	Netherlands	populated place	3.52 km
9	Lichtmis	Netherlands	populated place	3.73 km
10	Hasseltsche Wetering	Netherlands	canal	4.65 km
11	Dedemsvaart	Netherlands	canal	4.65 km
12	Roebolligehoek	Netherlands	populated place	5 km
13	De Velde	Netherlands	populated place	5 km
14	Gemeente Hasselt	Netherlands	second-order administrative division	5.01 km

Figure 3 -Image from Geonames server, with satellite image showing the named objects represented in the database, in a search for Hasselt. On a topographic map 1:50 000 the same area would have three times as many names (Gazetteer of the Netherlands 1985). Geometry seems poor. Hasselt town (NL) is on the lower left.

3. Getty Thesaurus of geographical names (TGN) on line

This server can be found at the following location: http://www.getty.edu/research/conducting_research/vocabularies/tgn/index.html .The TGN contains about 900,000 records, including 1.1 million names. Attribute information consists of feature types, coordinates, and descriptive notes (focusing on places important for the study of art and architecture). Place names may include variant names, exonyms and historical names. Data sources are indicated. The temporal coverage of the TGN ranges from prehistory to the present and the scope is global.

ID: 1047824
Record Type: administrative

Hasselt (inhabited place)
Coordinates:
Lat: 52 36 00 N *degrees minutes* Lat: 52.6000 *decimal degrees*
Long: 006 05 00 E *degrees minutes* Long: 6.0833 *decimal degrees*

Note: Old member of the Hanseatic Leaugue, it was granted a charter in 1252. Known for the printing of the incunabula of the Italian printer Peregrinus Barmantlo in 1492-1490.

Names:

Hasselt (**preferred**, C,V,N)

Hierarchical Position:

World (facet)
.... Europe (continent)
..... Netherlands (nation)
..... Overijssel (province)
..... Hasselt (inhabited place)

Place Types:

inhabited place (**preferred**, C)
town (C)

Sources and Contributors:

Hasselt..... [VP]
..... NIMA, GEOnet Names Server (1996-1998)
..... Times Atlas of the World (1992) 81
Subject: ..... [VP]
..... NIMA, GEOnet Names Server (1996-1998)
..... Phaidon Art Guide: Holland (1987) 124
..... Times Atlas of the World (1992) 81
Note: ..... [VP]
..... Phaidon Art Guide: Holland (1987) 124

Figure 4 - Screen dump from the Getty Thesaurus of Geographical Names

4. The Fuzzy Gazetteer (European Commission/JRC Digital Map Archive)

It can be found at: http://dmaweb2.jrc.it/services/fuzzyg/default.asp . FuzzyG 1.0 is the result of research collaboration of Hof University (Germany) and the European Joint Research Center. There used to be a possibility to view the location of the searched place name on different map backgrounds, from the European Joint Research Centre Digital Atlas. These were for instance geological maps, population density maps and current weather maps. The Joint Research Centre Digital Atlas features geographic datasets with global extent. Its purpose is to better inform humanitarian and foreign affairs decision makers on the landscape and environment of places in the world. That is why there is so much emphasis on different thematic background maps. FuzzyG (or Fuzzy Gazetteer) searches for place names world-wide and can handle variations in spelling, thereby making the searches more robust. Sources are not indicated on the website, apart from the fact that most names are based on GNS.

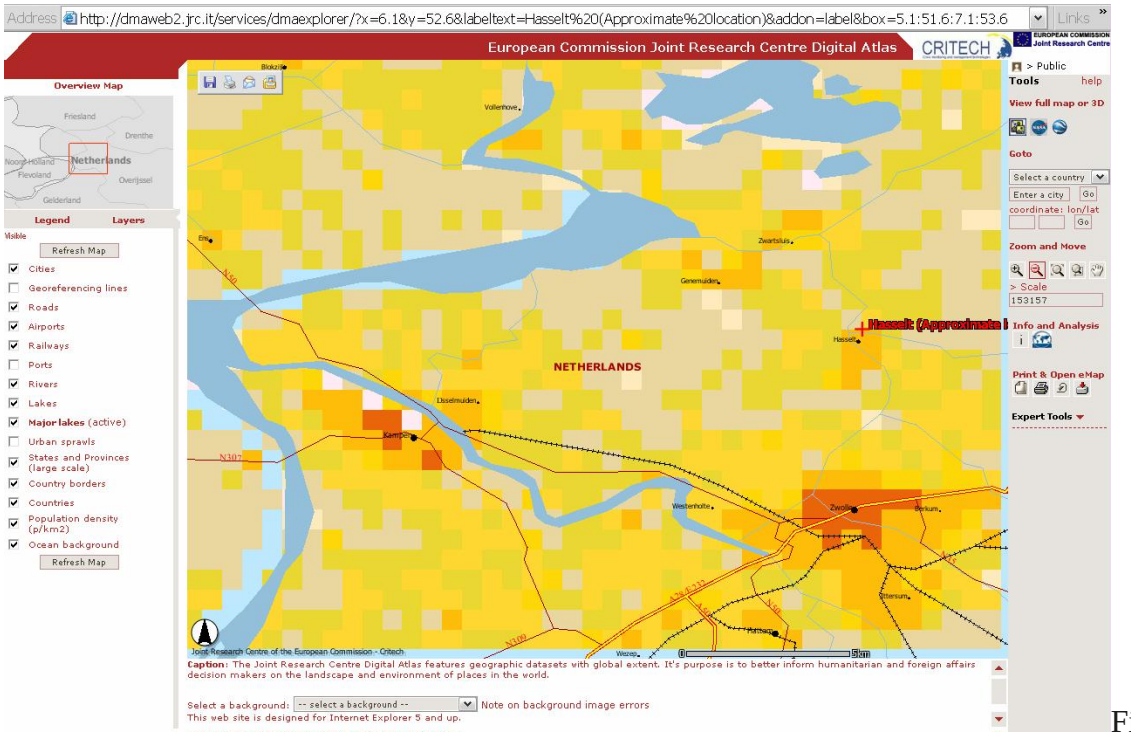


Figure 5 Screendump from FuzzyG: location of Hasselt (Netherlands) against a population density map background

5. **Alexandria Digital Library Gazetteer Server** To be found at http://webclient.alexandria.ucsb.edu/client/gaz/adl/

There are about 5,9 million entries, based on GNS. These are distributed over the following feature categories:

administrative areas	2,126,610
parks	20,408
political areas	32,623
countries	165
populated places	2,000,821
reserves and tribal areas	8,887
hydrographic features	636,564
Bays, fjords and gulfs	36,974
channels	13,874
ice masses	3,569
lakes	94,758
seas	273
Streams and rivers	480,921
land parcels	12,424
manmade features	858,145
agricultural sites	174,912
buildings, residential sites	243,448
cemeteries	64,535
historical sites	66,228

hydrographic structures	123,991
canals	21,482
dam sites and reservoirs	95,804
mine sites	24,070
monuments	8,560
recreational facilities	7,526
storage structures	11,969
towers and telecom. features	26,663
transportation features	77,933
wells	71,680
physiographic features	575,964
arroyos	45,359
bars (physiographic)	12,167
basins	8,957
capes and cliffs	25,264
dunes and beaches	8,854
mountains and mesas	367,723
plains, gaps and valleys	59,932
plateaus and flats	5,887
seafloor items and reefs	10,382
volcanic features	2,262
regions	120,132
biogeographic regions	44,782
deserts	588
forests	19,476
grasslands	4,419
wetlands	17,887
Islands	70,094

Table 1. Feature classes in ADL and their numbers

The spatial coverage (1/1/2007) of the ADL gazetteer can be seen in figure 6:

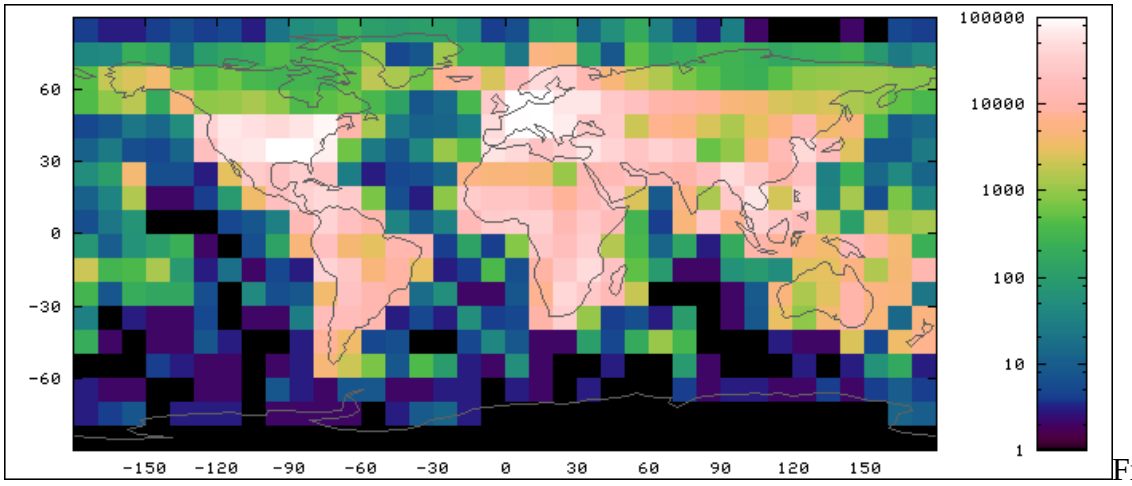


Figure 6 Spatial coverage: number of names per 10x10 degrees square in the ADL gazetteer

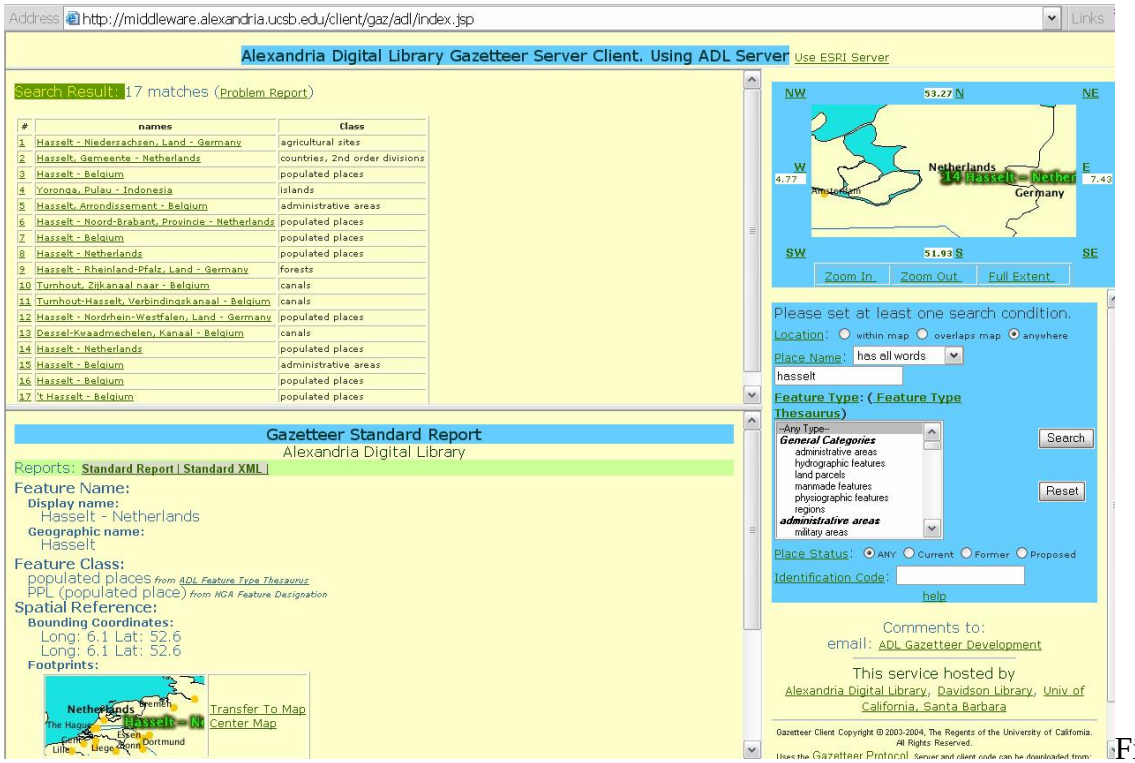
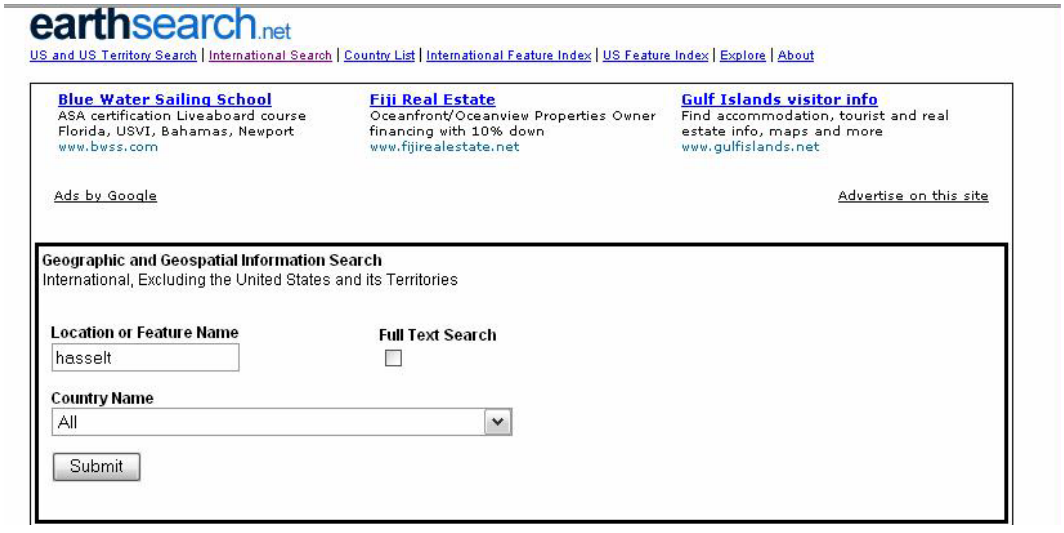


Figure 7 - Screendump from Alexandria Digital Library Gazetteer

6. Earthsearch

To be found at http://earthsearch.net/ . Based on GNS. Limited attribute information. Linked to CIA factbook information.



Search Results - first 100

Name - click for world map location	**Country**	**Type**	**Latitude**	**Longitude**
Hasselt	BE	ADMD	50.9333333	5.3333333
Hasselt	BE	PPL	50.95	4.3166667
Hasselt	BE	PPL	51.2	5.4166667
Hasselt	BE	PPL	51.15	5.65
Hasselt	GM	FRM	53.2833333	7.65
Hasselt	GM	FRST	50.0833333	6.6833333
Hasselt	GM	PPL	51.7666667	6.2
Hasselt	NL	PPL	51.4333333	6.1666667
Hasselt	NL	PPL	52.6	6.1
Hasselt	NL	PPLX	51.5833333	5.0833333
Hasselt	ID	ISL	-1.1	128.4

Figure 8 Screendump from Earthsearch

7. **World gazetteer**

To be found at http://www.world-gazetteer.com/ . The source is not clear, information is limited, the site specialises on population figures and their development. The feature can be shown on a map, coordinates are decimal.

Figure 9 World Gazetteer Screen dump

8. **Global Gazetteer Version 2.1** (www.Fallingrain.com/world)
   For Kenya this server contains 3102 names.

9. **World City Name Database (Nonanet)**
   The personal site of Alexander Mayrhofer. To be found at nona.net. Claims over 2 million place names. No metadata.

10. **Eurogeonames**
    Only provides names for Europe, built jointly by European mapping agencies. Smart software links all national databases with their different feature categories, name

models and terminology together and makes them accessible as a virtual names server for all of Europe.

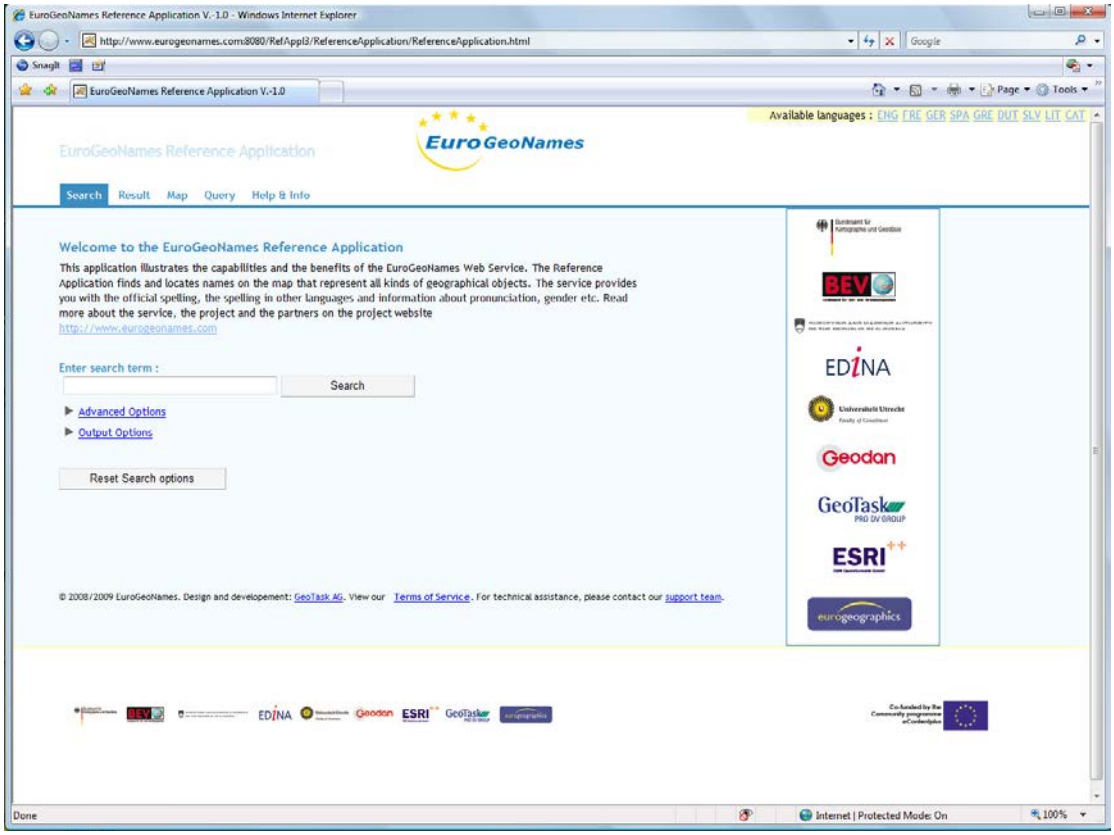


Figure 10 EuroGeoNames website

11. **UN gazetteer** (aimed at UN personnel and UN operations). Currently the short-term objective is to contain names of all places over 100 000 inhabitants and their pronunciation.

Analysis:

When we look at names of Kenyan national parks and nature reserves, it is only GNS, Geonames, Fuzzy Gazetteer and Earth Search that provide most of these names; none of them is complete; only very few refer to the actual national park or forest reserve but rather to similarly named hills and rivers nearby. As national parks and nature reserves are supposed to be important for Kenya's economy, this is a mismatch.

Name	1.GNS	2.Geonames	3.Getty	4.Fuzzy	5.ADL	6.EarthS	7.World Gaz	8.Global	9CityNames
Arabuko NP	x	x	-	x	-	-	-	-	
Bisanandi NP	-	-	-	x	-	x	-	-	-
L.Bogoria NR	x	x	x	river	-	river	-	-	-
Kakamega FR	x	x	x	x	-	x	-	-	-
Kizingo	x	x	-	x	-	x	-	-	-
Losai NR	-	-	-	x	-	x	-	-	-
Maasai MaraNR	-	Masai	x	-	-	-	-	-	-
Murera Lodge	-	x	-	x	-	x	-	-	-
Namunyak NP	-	-	-	-	-	-	-	-	-
Naro MoruLodge	-	-	-	x	-	x	-	-	-
Oldoinyo NP	x	x	x	x	-	x	-	-	-
Shimba NR	x	x	-	-	-	-	-	-	-
Ugingo I.	x	x	-	-	-	-	-	-	-

Table 2: rendering of names of Kenyan national parks in consulted gazetteers

Although sometimes some metadata are available, such as on data sources used, the attribute information available and the functionality offered, the hit-frequency would also be of interest. In order to get an idea of the completeness, we need an idea of the number of names in the data bases, and compare them to the number of names collected and processed nationally. In the Netherlands the NMCA database contained 3x as many names as were contained in the GNS database, but there is insufficient data to extrapolate that to all databases reviewed. In table 4 an overview is provided of the gazetteer server characteristics in order to be able to compare them.

From this table 4 we can see that GNS server has the most extensive attribute information (19 crosses, as compared to 14 of the runner up TGN (Getty) and 6 for the least informative (in terms of attribute information) server, Earth Search. If we take GNS as the norm, Geonames is exceptional in that it provides comments why there have been changes in names in the last month, and because they offer population numbers.

TGN is clearly marked by its art background. In the example in figure 4 one may find the name of artists from Hasselt listed. It also has data on elevation, and brings comments.

FuzzyG apparently is focused at humanitarian problems and risk management. It has most names (7.2 million in comparison with the runner-up, GNS with 5.5 million names) and is conspicuous because of its thematic maps and weather maps on which the searched features can be located. Population numbers are also provided.

ADL has average attribute information, and is well structured hierarchically. It is the only one to discern settlements that don't exist any more or that are planned.

Earth Search is only special in its comments, which have been derived from CIA handbooks.

World Gazetteer is special because of its emphasis on population data. It has not only current population numbers, but also tries to provide some population data from the past, so that trends can be discerned.

The attribute data almost all servers have in common are: name, coordinates (either decimal or lat/long) feature designation, location on map (or on Google Earth) and country codes. Only 4 servers have unique feature IDs

Almost all servers are based on GNS. The European FuzzyG has other sources apart from GNS, and the source of World Gazetteer is not indicated either. As GNS has used data from European gazetteers, it should be called a secondary source; all other GNS-based gazetteers are tertiary sources. Probably, the analog country gazetteers (produced through the use of local lists of geographical names or by harvesting names from local topographic maps) that were used by GNS for filling its database were produced in the 1960s and many of the national parks and nature reserves refered to in table 2 were only set up in the 1980s and beyond, so there might be a mismatch here because the name sets were not updated. Anyway,

it is essential that recent new geographical names be also reflected in the gazetteers or names servers, in view of the many names applications in the geo-information infrastructure.

Important Issues for providing national gazetteers:

The unique selling points of national gazetteers would be that the names data they would provide are:
-from a primary source,
-that is continuously updated,
-that is more detailed than the GNS data,
-that it is closer to the experts that collect the names data,
-that through official cooperation there is better quality control!

How do these servers that have been reviewed stand in relation to these unique selling points of national gazetteers? See table 4

Unique selling points	**National gazetteers**	**current geoname servers**
primary data	yes	no
official data	yes	?
high quality data	yes	?
up-to-date data	yes	no
complete coverage	sometimes	no
coverage	country	whole world

Table 4: Comparison of proposed EGN Names Server and current servers

See table 3 for the comparison of these name servers.

server	GNS	Geonames	TGN	FuzzyG ADL	EarthSearch	World Gazetteer	EGN
feature ID	x	-	-	-	-	-	-
name	x	x	x	x	x	x	x
adverb of name	-	-	x	-	-	-	-
decimal coordinates	x	x	x	x	-	x	x
lat/long coordinates	-	x	x	x	x	-	-
elevation	?	?	x	-	-	-	-
bounding coordinates	x	?	-	-	?	-	-
feature designation	x	x	x	x	x	-	x
unique feature/name ID	x	x	x	-	x	-	-
JOG reference	x	-	-	-	x	-	-
map scale range	x	-	-	-	-	-	-
comments:	-	x	x	-	-	x	-
search with wildcards?	-	not yet	-	-	-	-	-
show on Google Earth	-	x	-	-	-	x	-
show on map	x	-	?	x	x	x	x
fuzzy search criteria	x	-	-	x	-	-	-
language codes	x	x	x	-	-	-	-
country code	x	x	-	x	x	x	x
all diacritics	x	?	x	-	-	-	-
adm code	x	?	?	?	-	-	-
historic names	x	?	x	-	-	-	-
art connotations	-	-	x	-	-	-	-
provisional names	x	-	-	-	-	-	-
variant names	x	-	x	-	-	-	-
conventional names	x	x	x	-	-	-	-
names versions in non-Roman script	x	-	-	-	-	-	-
thematic maps	-	-	-	x	-	-	-
current weather maps	-	-	-	x	-	-	-
sources	GNS	GNS	GNS	?	GNS	GNS	?

server	GNS	Geonames	TGN	FuzzyG ADL	EarthSearch	World Gazetteer	EGN
feature search	?	?	?	-	x	-	-
place status (ruined/future)	-	-	-	-	x	-	-
population number	-	x	-	x	?	-	x
population history	-	-	-	-	-	-	x
overall number of names	5,5M	?	1,1M	7,2M	4,4	?	?

Literature:

Tichelaar, Tjeerd (2003) Field collection of names in multi-lingual and multi-cultural areas. PP79-85, proceedings Training Course on Toponymy Enschede/Frankfurt/Berlin 2002. Mitteilungen des Bundesamtes für Kartographie und Geodäsie, Band 28. Frankfurt am Main: Verlag des Bundesamtes für Kartographie und Geodäsie.