Section 8 Websites

Chapter 21 **Evaluation of current international toponymic databases**

Ferjan Ormeling

21.1 Introduction

What happens if for your country a national toponymic database is non-existent 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.

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. That would mean that within one generation, the namescape of an area would change completely. So we see different naming habits, and it would be difficult for outsiders to deal correctly with all those differences.

In Europe while building a European names database, 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 Fort de Kock, the present Bukittingi in West Sumatra, or for the former Fort Willem I on Java island, the present Ambarawa).

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 chapter is to find out what international toponymic databases are currently available on the web, and to give an indication of their nature

We will deal with the following web name servers:
21.2 GEOnet Names Server (GNS)
21.3 Geonames
21.4 Getty Thesaurus
21.5 Fuzzy Gazetteer
21.6 Alexandria Digital Library Gazetteer
21.7 Global Gazetteer Version 2.1 (Fallingrain)
21.8 EuroGeoNames
21.9 World Geographical Names database

21.2 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 database contains over 6 million features with over 11 million feature names (2015). 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 Indonesia for example, 489 442 names were held in 2015. This GEOnet Names Service is also the main source for most other name servers. The source of the GEOnet names server

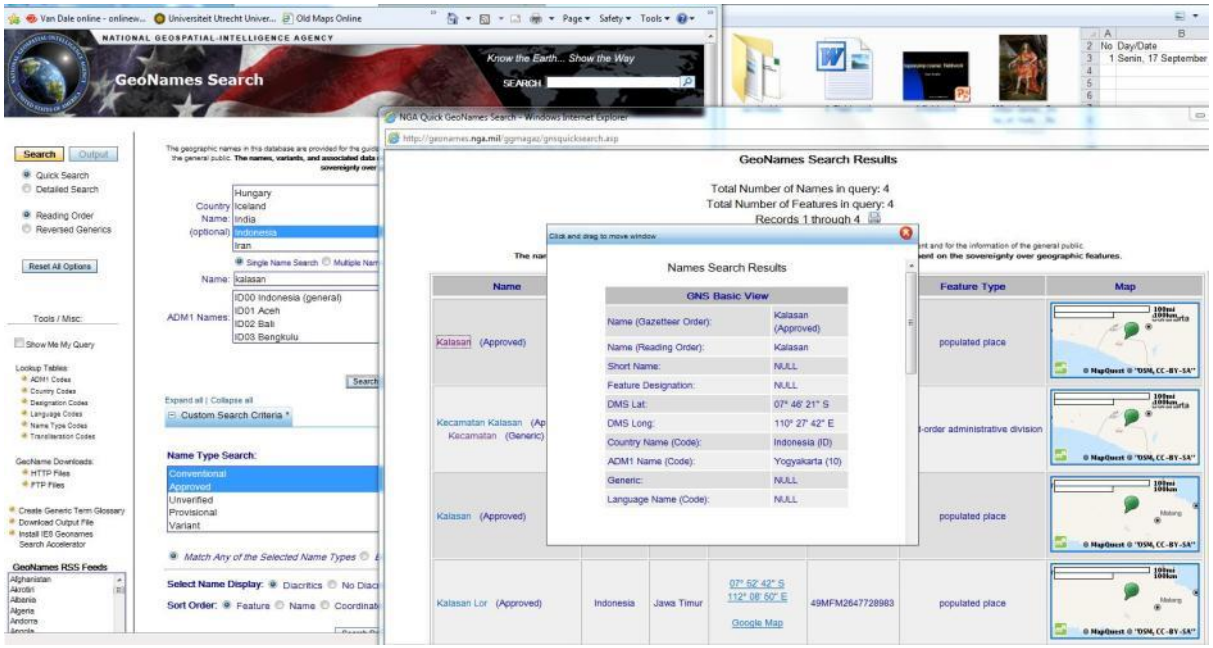

Figure 21-1 Example from GNS: looking for Kalasan, Java

used to be the paper gazetteers the American defence establishment produced for all countries, in the 1960s. Not all of them have been updated to the same degree. There are 2 search options, the Open GIS consortium viewer and the text-based viewer. In the first one, one zooms in on a world map and finds all the names in the database; clicking a name gives additional information. In the text-based viewer one enters a name, a list of matching names appears, and one can select the appropriate one; as this may need some time one can refine the search by entering countries, feature codes, and the name type wished for (conventional names, approved names, unverified, provisional, (anglicised) variant names, (unverified or variant) non-roman script names. Also, the order in with the matching names can be presented can de set (by unique feature ID, by coordinates or by feature designation). For research purposes the queries can also proceed with letter strings that form the start, end or middle of the name. The GNS database has 137 names ending in –more for Australia, and 76 starting with More-.

The maps on which the results are shown can be from Open GIS consortium, from Google Map and from Mapquest. In the latter cases, advertisements would be added to them.

There is a section on individual countries, describing the way names from each country have been dealt with.

For names in Arabic, Bengali, Chinese, Cyrillic, Georgian, Greek, Indian scripts, Ivrite, Katakana, Korean, Thai, Tigrinya, (not yet for Amharic, Armenian, Burmese, Cambodian, Lao, Singhalese) forms in local scripts are given as well.

21.3 Geonames

The GeoNames server (http://www.geonames.org/) has as its principal sources the GNS server, the GNIS server for names in the US, Ordnance Survey Open Data, www.geobase.ca for names in Canada, 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. In 2015 it had 10 million names for 9 million unique features, and 5,5 million alternate names. As does GNS, it uses the World Geodetic system WGS84. It is the user-friendliest of the global names servers.

Figure 21-2 GeoNames entry for Jaffna, Sri Lanka

GeoNames has a special service, which shows all the named features on imagery provided by Google Earth, with the types of features named colour coded (red: administrative areas; grey: populated places; terra: physical objects; yellow green: traffic infrastructure; green: vegetation names; blue: hydrography and purple: individual buildings). Clicking a coloured pin on the map leads to information on the name (also written in different scripts) feature type, administrative hierarchy, elevation, population number, coordinates, GeoNames ID, name history, and alternate names.
GeoNames now considers adding historic names and colloquial names.

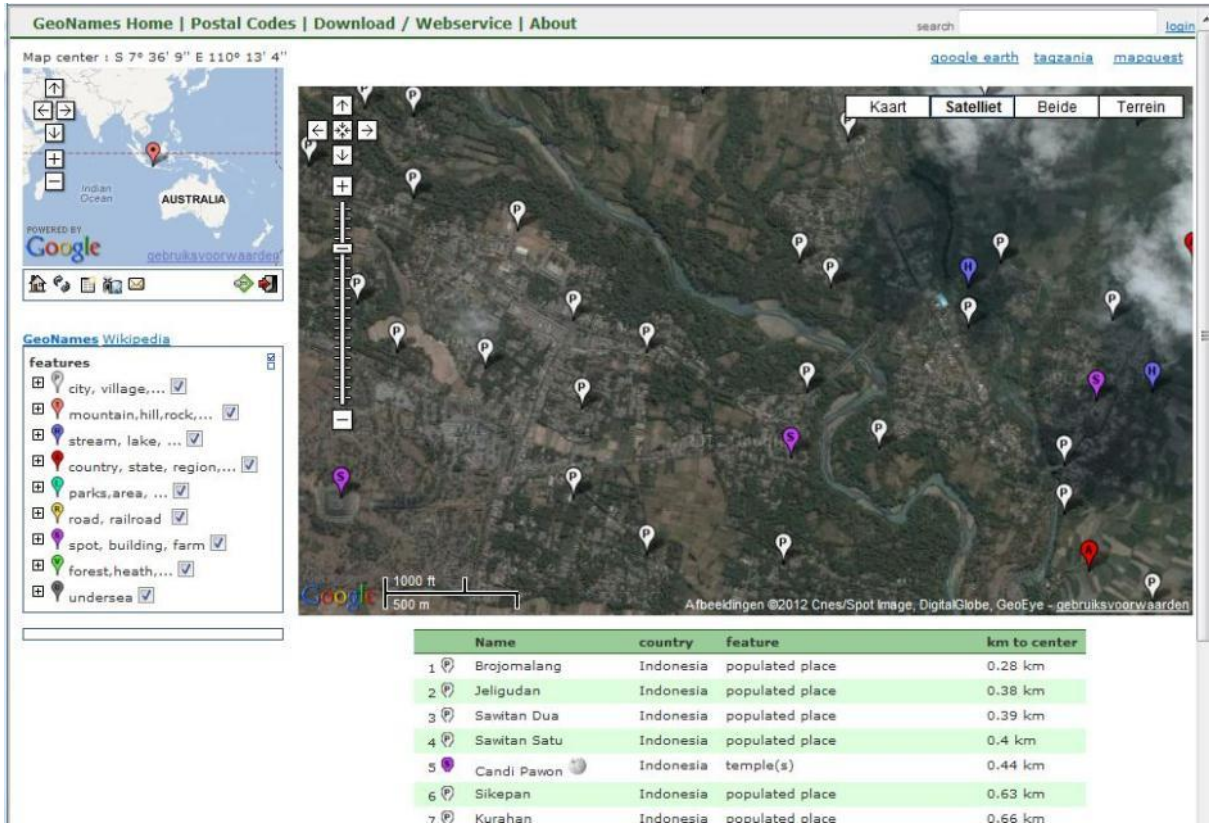

Figure 21-3 -Image from GeoNames server, with satellite image showing the named objects represented in the database, in a search for Candi Pawon, on a line with Borobudur and Mendut (purple S at right) on Java island. On the current GeoNames version (2015) the colour-coded droplets now have changed to squares.

For a simple search one can set the country in which the named object is expected to be located; for an advanced search one can also set the continent, feature class and indicate whether a fuzzy search is needed or not.
The feature codes used by GeoNames can be found at http://www.geonames.org/export/codes.html .

By clicking a bar, all the names available in the database for the area represented on the map will be presented in a list, in a numbered sequence based on their distance from the map centre, with administrative area, feature class and code added. Links to Wikipedia are incorporated for all locations mentioned there, indicated by a special W sign in a grey square

21.4 Getty Thesaurus of geographic names (TGN) on line

This server can be found at the following location: http://www.getty.edu/research/tools/vocabularies/tgn/. The TGN has been built in order to improve access to information about art, architecture and material culture. It contains (2015) about 1,400,000 records, including 2,1 million place names. Attribute information consists of feature types, coordinates (approximate, for reference use only), and descriptive notes (focusing on places important for the study of art and architecture). Place names may include variant names, exonyms and **historical names**. Start and end dates when the names were applicable are noted. Data sources are indicated. The temporal coverage of the TGN ranges from prehistory to the present and the scope is global. The data are not visualised on maps, but there are references to related places. When entering the name Mendut (a Hindu temple in Central java), one is also referred to the nearby temples Pawon and Borobudur. TGN uses the concept of homographs that is homonyms in the same script.

ID: 6003632

Majapahit (former nation/state/empire)

Note: Former state on the island of Java. Hindu-Buddhist temples c. AD 1000 to 130(

Names:
Majapahit (**preferred**,C,O)

Hierarchical Position:
World (facet)
.... Asia (continent) (P)
..... Indonesia (nation) (P)
..... Jawa (island) (P)
..... Majapahit (former nation/state/empire) (P,H)

Place Types:
former nation/state/empire (**preferred**, C)
empire (H)
state (H)

Sources and Contributors:
Majapahit..... [VP Preferred]
..... Times Atlas of World History (1993) 349
Subject: ..... [VP]
..... Times Atlas of World History (1993) 349

Figure 21-4 - Screen dump from the Getty Thesaurus of Geographical Names

21.5 The Fuzzy Gazetteer (European Commission/JRC Digital Map Archive and Cartographic Section UN)

It can be found at: http://dma.jrc.it/services/gazetteer/. FuzzyG 1.0 was the result of research collaboration of Hof University (Germany) and the European Union Joint Research Centre. 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 worldwide 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. It contains 7,2 million names (2017).

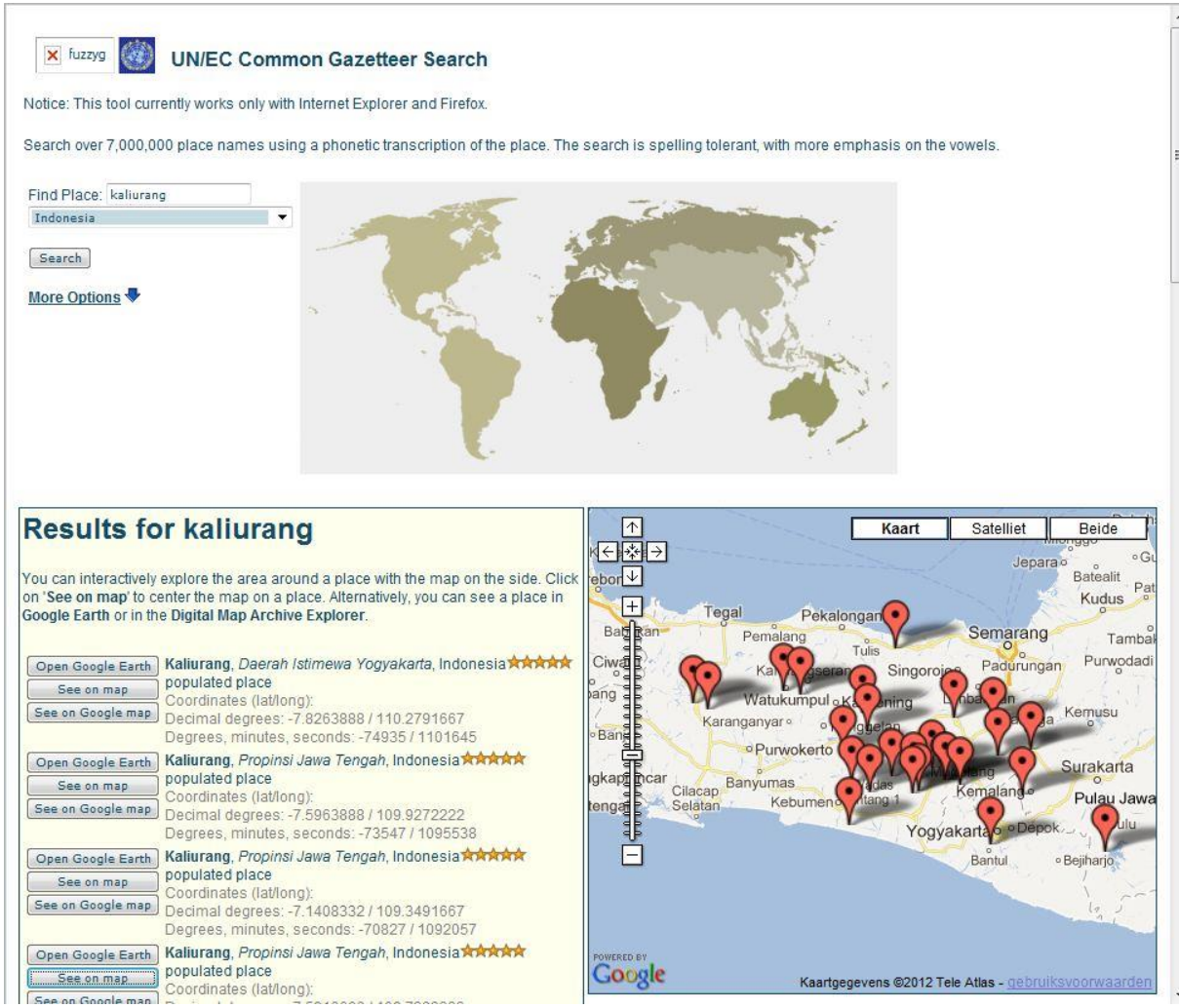

Figure 21-5 Screen dump from FuzzyG: locations of Kaliurang (Java)

21.6 Alexandria Digital Library Gazetteer Server

To be found at https://www.library.ucsb.edu/map-imagery-lab/alexandria-digital-library-gazetteer. There are about 5,9 million entries, based on GNS. These are distributed over the following feature categories (see table 21-1):

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 21- 1. Feature classes in ADL and their numbers

Presently, the site seems to be under reconstruction.

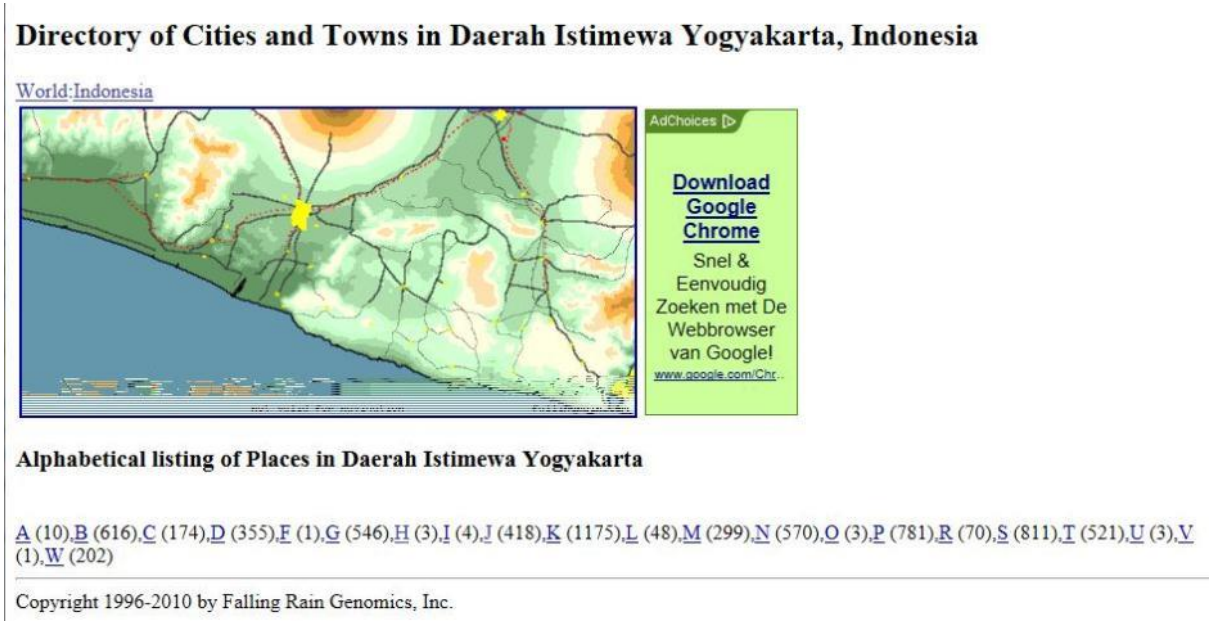

Figure 21-6 Global gazetteer screen dump

21.7 Global Gazetteer Version 2.3

(http://www.fallingrain.com/world/index.html). Here, by clicking on a country and progressively on its administrative divisions all the names in this database for that division are listed, in alphabetical groups. For all name entries, feature type, region, coordinates and population numbers are given. For D.I. Yogyakarta, a province of Indonesia, this server contains 6611 names (see figure 21-6).

21.8 EuroGeoNames

EuroGeoNames combines geographical names from the national mapping agencies in Europe. The national databases of 17 European mapping agencies, each with its own feature categories, name models and terminology, have been linked and made accessible to queries, as a virtual names server. The website has been made accessible in the languages of all participants, and also in that of the minority languages. Special attention has been focused on the exonyms, and these were stored in an additional database. The database was delivered and put on-line for a short time in 2009, but since then was discontinued in order to allow additional countries to be made participants.

The service was recently successfully reorganised by the Finnish Geodetic Institute to a cloud-based web interface. The exonym functionality still has to be integrated into it, before it is operational again.

21.9 World Geographical Names Database

is a multilingual, multiscriptual georeferenced geographical names database developed by the UN Statistics Division and UNGEGN, with input from the UN Terminology and Reference Section, the UN Cartographic Section and the UN Geographic Information Working Group (UNGIWG). Through the web, database users can access short and full names of countries (193 UN member states), their capitals, and the major cities (population over 100,000) for many countries. Authoritative city endonyms are provided mainly by national name authorities and sound files are being added to assist users with pronunciation. At this stage (2015), the database contains over 2,700 country names, some 6,100 names for 3,362 cities, with more than 970 audio files. As a useful reference tool for geo-information management, the UNGEGN World Geographical Names Database will continue to be developed and improved, and updated on an on-going basis.

temple name	Getty/TGN	Geo Names	fuzzyG	Alexan-dria DL	GNS
Borobudur	x	x	x	x	x
Gebang, Candi	-	-	-	-	-
Jawi	-	-	x	-	-
Kalasan	-	x	x	x	x
Kedulan	-	-	x	-	x
Mendut	x	x	x	x	x
Pawon	x	x	-	-	x
Penataran	-	x	x	x	-
Plaosan	-	x	x	-	x
Prambanan	x	x	x	x	x
Ratu Boko	-	X	Ratu Baka	-	x
Sari, Candi	-	X	x	-	x
Sewu, Candi	-	X	x	-	x
Sukuh	-	-	-	-	-
Trowulan	-	x	x	X	x
historic name					
Bantam, Banten	X	x	x	-	x
Galuh	-	-	-	-	-
Kahuripan	-	-	-	-	-
Majapahit	x	-	-	-	-
Mataram	-	-	-	-	-
Pajajaran	-	-	-	-	-
other names					
Demak	x	x	x	X	x
Imogiri	x	x	x	X	x
Kota Gede	-	x	x	X	x
Solo, Surakarta	x	x	x	Solo not, x	x
Yogyakarta	x	x	x	X	x

Table 21-2: Rendering of names of temples on the isle of Java and some historical names in consulted gazetteers

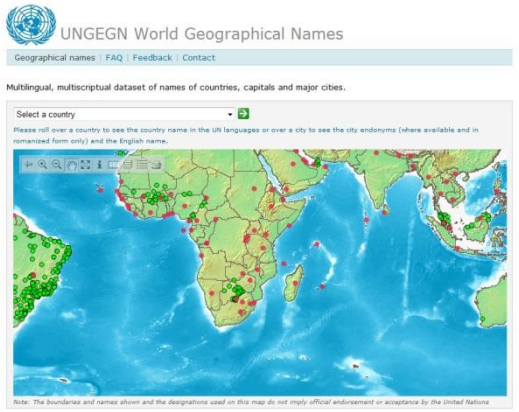

Figure 21-7 Screen dump from UNGEGN World Geographical Names Database

Currently the short-term objective is to contain names of all places over 100 000 inhabitants and their pronunciation.

21.10 Analysis:

When, as a test regarding the display of cultural/heritage items, we look at names of Indonesian temples in table 21- 2, it is only GNS, Geonames and Fuzzy Gazetteer that provide most of these names; none of them is complete; only very few (Getty) also have historical names. 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 table 21-4 an overview is provided of the gazetteer server characteristics in order to be able to compare them.

From this table 21-2 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. For some places, one may find the name of artists listed that came from that location. 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.

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 analogue 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 referred to in table 21- 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 in this chapter stand in relation to these unique selling points of national gazetteers? See tables 21-3 and (21-)4:

Table 4: Comparison of current servers

server	GNS	Geonames	TGN	FuzzyG	ADL
feature ID	x	-	-	-	-
name	x	x	x	x	x
adverb of name	-	-	x	-	-
decimal coordinates	x	x	x	x	-
lat/long coordinates	-	x	x	x	x
elevation	?	?	x	-	-
bounding coordinates	x	?	-	-	?
feature designation	x	x	x	x	x
unique feature/name ID	x	x	x	-	x
JOG reference	x	-	-	-	x
map scale range	x	-	-	-	-
comments:	-	x	x	-	-
search with wildcards	?	not yet	-	-	-
show on Google Earth	-	x	-	-	-
show on map	x	-	?	x	x
fuzzy search criteria	x	-	-	x	-
language codes	x	x	x	-	-
country code	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
server	GNS	Geonames	TGN	FuzzyG	ADL
feature search	?	?	?	-	x
place status (ruined/future)	-	-	-	-	x
population number	-	x	-	x	?
populationhistory	-	-	-	-	-
overall number of names	11M	10M	2,1M	7,2M	4,4M

Table 21-3: Unique selling points of National gazetteers as compared to current commercial geoname servers

	Nat Gaz	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

21.11 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.