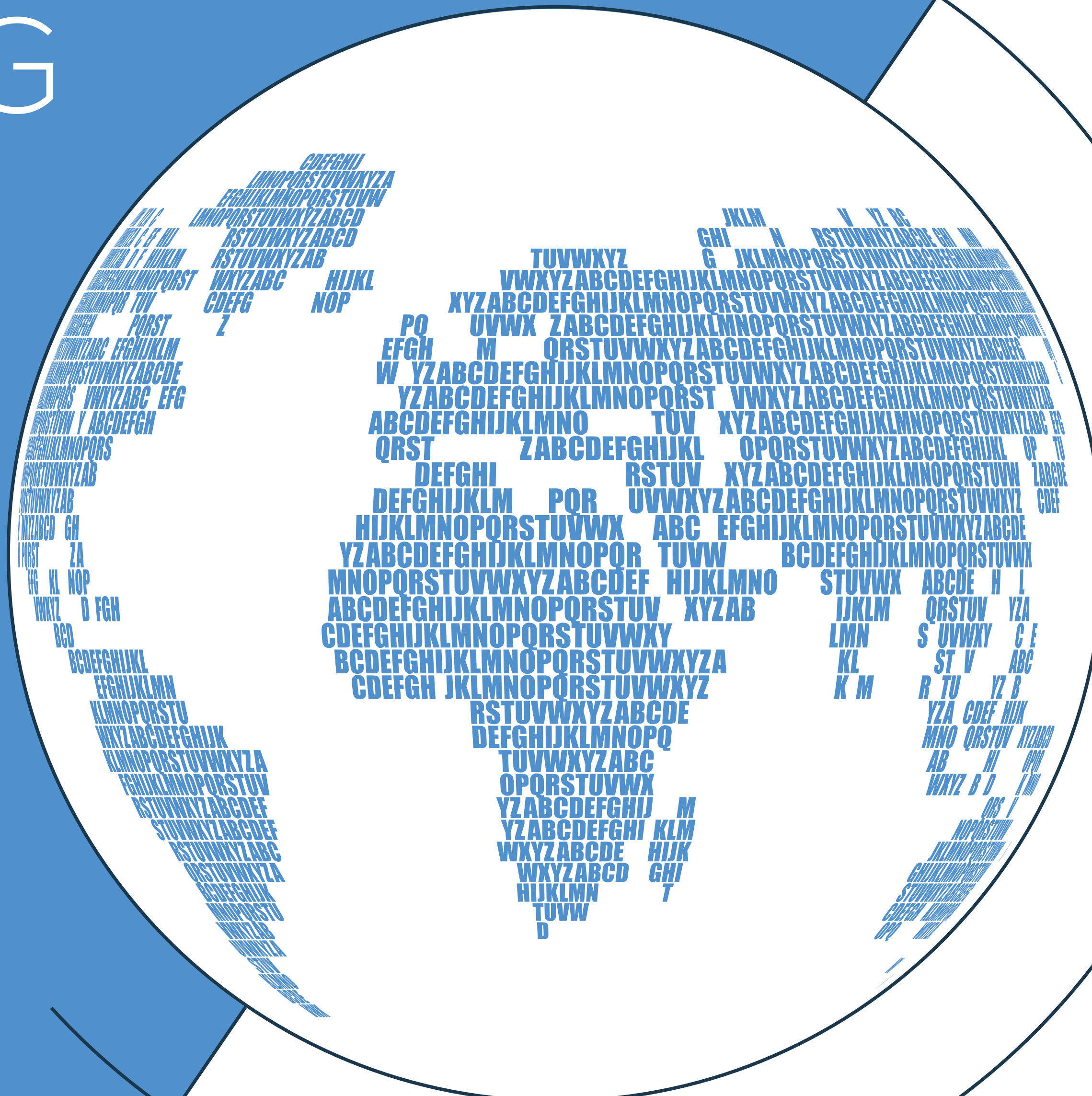
TOPONYMY TRAINING MANUAL



UNITED NATIONS
GROUP OF EXPERTS
ON GEOGRAPHICAL NAMES

DEPARTMENT OF ECONOMIC
AND SOCIAL AFFAIRS

CONTENTS OF UNGEGN'S *TOPONYMY TRAINING MANUAL*

Introduction: About the Manual

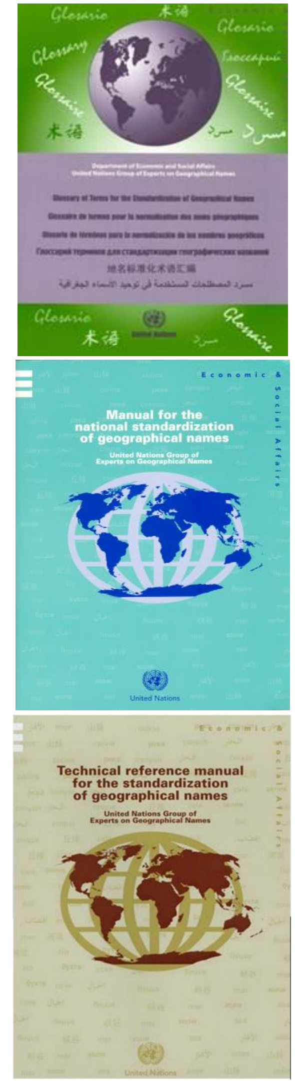
UNGEGN has offered training courses in the administration of geographical names since 1982. These courses have taken place on every continent and over the years the content has developed to address advances in technology and changing attitudes to the standardization of geographical names. As well individual courses have taken into account the particular interests and needs of the host country and associated UNGEGN divisions.

A second step was taken in meeting the needs of those countries looking for training opportunities to develop the skills and knowledge of staff responsible for the toponymic processes and records. A web course was developed in conjunction with the International Cartographic Association (ICA), and later with financial support from UN Statistics Division. The course, containing 20 modules of lessons and exercises, is designed for independent study and is freely available through the UNGEGN website (http://unstats.un.org/unsd/geoinfo/UNGEGN/docs/_data_ICAcourses/index.html). It was first released in 2012 after several years of development.

Following the 10th UN Conference on the Standardization of Geographical Names in 2012, a decision was taken by the Convenor and members of the UNGEGN Working Group on Training Courses in Toponymy to follow a third approach to disseminating materials for training and capacity building purposes. This new venture was designed to gather together a collection of articles on a variety to toponymic issues – some of a more advanced nature, others covering subjects that receive but a brief mention in a basic training course. And so the UNGEGN Toponymy Training Manual was started! Articles have been volunteered by individuals – working with geographical names within their governments, using the results of names programmes in geospatial information, or teaching the concepts of toponymy in higher education institutions. Some articles were completed by 2015, while others have been written as late as June 2017; consequently, portrayal of web pages, for example, may show differences.

The target audience for this new manual includes those working for national names boards and in names collecting and processing programmes. We have tried to make the chapters "friendly" for the reader. For example, the texts will not be too scientific, footnotes and endnotes are used sparingly. Some chapters start with objectives and the skills the reader may expect when they have made themselves familiar the chapter's contents. We hope that the necessary theoretical background and practical approaches to realise particular processes has been provided. Realizing that it may be difficult for many readers to access libraries where toponymic literature can be accessed in person, references, preferably online, have been included in each chapter.

The manual has some 30 chapters, numbered sequentially in such a way that we start with national names programmes and their benefits, and their management, and then zoom out to regional and global toponymic activities and initiatives, technical aspects, cultural aspects, research, mapping, assessing the results through auditing, and finally communicating geographical names to the media.



UNGEGN publications

We realise that the team of authors that voluntarily produced contributions for this book did not have the benefit of knowing in advance about the contributions of other authors – consequently, there might be some overlap, but the reviewers have made sure that there are no contradictory passages. There are large differences in background of authors, some more technical, other more theoretical or practical, and the readers will have to accustom themselves to the different styles of approach.

The range of the chapters is vast and we doubt that all readers will be interested in all topics. But the reviewers are most grateful for the efforts made to provide the wealth of material that is offered here.

All chapters can be freely downloaded from the UNGEGN website, as the authors have transferred the copyright of their work to UNGEGN. We hope that these chapters will be used, copied, and translated into other languages as well, with appropriate copyright credits given to UNGEGN.

The chapters have been organized into the following sections:

Section 1 General /strategic issues
In this first section, we highlight the eminence of the national toponymic database, and its benefits for administration, development, emergency mapping and preserving the cultural heritage, and we refer to the starting level of this manual: knowledge of the matter dealt with in the basic toponymy webcourse: fieldwork, names processing, role of UNGEGN, etc.

Section 2 Management of a national names programme
This section deals with the finances, administration, legal issues and human resources necessary for maintaining and publishing the geographical names database, with discussion on logistics, production time needed, necessary upkeep costs.

Section 3 Regional activities and applications
This section covers the collection of names for different purposes, such as for hydrographic or for tourist maps and also deals with organizing urban street naming programmes.

Section 4 Regional and global initiatives
The working of both institutional and commercial global name servers. We hope to extend this section in future with an example of a regional small islands naming project.

Section 5 Progress of toponymic names databases
Standardization of names for statistical enumeration areas and smoothing differences between statistical and topographical areal object names are dealt with here, as is the integration of crowd-sourced or volunteered geographic information, Open Street mapping and geographical names. A chapter on Open Street mapping is still expected.

Section 6 Technical issues: **databases and pronunciation**
Preliminary discussion on the structure and content of toponymic databases and gazetteers, with coverage of database design & management. A commercial example will show the requirements for school atlas names databases. As databases might also incorporate sound recordings, best practices for recording names pronunciation will be dealt with here as well.

Section 7 Technical issues: internet, web services infrastructure and applications
Here the issue of open source options vs commercial options will be dealt with, as will criteria for software selection.

Section 8 Websites
Here the functional evaluation of current national place name web services and of commercial name servers and their applications is dealt with.

Section 9 Cultural aspects
This section deals with place names as cultural heritage and includes the history of place names, and – in the near future - the collection of indigenous names.

Section 10 Toponymic research and documentation
This section introduces the science of onomastics and its relevant terminology, etymology and the study of linguistic and toponymic sources.

Section 11 Cartographic aspects
This section covers the updating of names sheets and the assessment of the optimal names density for map sheets, as well as multiple naming. In future we hope to extend this with a chapter on digital place name labelling.

Section 12 Auditing place name records
Auditing place name records in order to coordinate toponymic information from different parts of the country, from different survey periods or in order to coordinate names record from different national institutions is essential for quality control of the place name records.

Section 13 Media training
This section deals with the media as well as with the public (both as informants and as customers)

Hopefully the range of chapters will be extended over time, also leading to a process of renewal of the contents of this manual.

When the UNGEGN Working Group on Training Courses in Toponymy, at the 10th UNCSGN conference in 2012, launched the proposal to produce this manual, and defined the topics to be included, Helen Kerfoot, Pier-Giorgio Zaccheddu and myself were asked to act as reviewers for the incoming chapters. At BKG, Sabine Afflerbach applied the final layout. The UNGEGN Secretariat with Cecille Blake saw to the incorporation of the manual in the UNGEGN website, and had the UN design section develop the most appropriate cover, which will be displayed in the publications section of the UNGEGN website. The Working Group owes the team many thanks for the time and effort invested in this project!

Ferjan Ormeling
Convenor, UNGEGN Working Group on Training Courses in Toponymy

Section 1: General/strategic issues

Chapter 1 **Geographical names, their standardization, and their value as part of geospatial data infrastructures**

Helen Kerfoot

1.1 Overview

In this introductory chapter, we will look at an overview of geographical names – their usefulness to us, their standardization and the benefits from this, together with today's need for Romanised forms. We will see how geographical names have been recorded and how today the geographical names database is a powerful tool for integration with other spatial data. Whereas standardized geographical names are essential for georeferencing, they are also significant to history, culture and identity of the peoples of the world. Geographical names or toponyms are at the meeting point of history, geography, linguistics, psychology, etc. and UNGEGN plays its part in promoting their standardization to benefit world users.

1.2 Our everyday relationship with geographical names

Can we pass a day without using geographical names (or toponyms)? Probably not! They are part of our everyday vocabulary, to say where we are going, to find our way, or to plan ahead. They are all around us! We see them as signs on streets (Figures 1-1, 1-2 and 1-3) and on highways (Figures 1-4 and 1-5), on panels announcing towns (Figures 1-6 and 1-7), on maps and in atlases, and as part of the crucial information provided through different media. Geographical names are expressed in different languages and in different scripts.

Figure 1-1 Street sign in "Chinatown", Ottawa, Canada

Figure 1-2 Street sign in Seoul, Republic of Korea

Figure 1-3 Street sign in Tunis, Tunisia

Figure 1-4 Highway sign, Iceland

Figure 1-5 Highway sign, outside Gaborone, Botswana

Figure 1-6 Town sign, Slovenia (double name in Slovenian and Italian)

Figure 1-7 Town sign, south of Strasbourg, France

Geographical names are one of our primary points of query in searching data on the web and in relating elements from different spatially-referenced data sets (for example, villages liable to flooding and hydrographic data). They are the link between men and women and the land on which they live. They are guardians of our history, our language, our culture and our aspirations, and some names have been in use for many centuries. The misuse or incorrect representation of toponyms (for example, on maps) can be an emotional source of local or national concern.

Geographical names may be quite clear or may pose us problems! Are we looking for Sydney or Sidney? Is it Dallas or Dulles? Are Bécs and Vienne the same as Wien (Figure 1-8)? Why is there a Swedish name (Helsingfors) for Helsinki? Should I use Falkland Islands or Las Malvinas on my map? Toponyms can be the source of many interesting and complex questions!

Figure 1-8 Commercially sold bag showing the endonym Wien, as well as various exonyms for the city

1.3 Guardians of a country's toponymy

Over many centuries, explorers kept their route logs and often recorded names for the landscape features they saw and perhaps mapped. Although some of these names may have been in local use, many were created by those explorers sailing for distant lands and using the naming process as a way of claiming sovereignty and of commemorating those who had provided financial backing for their expeditions. Hence, one finds the scattering of names of royalty, financiers, industrialists, friends and family on features "discovered" in the "new world".

Lists may have been filed and used by cartographers to produce local, regional or world maps. Countries (for instance, Norway) had produced names for official maps and passed legislation to standardize their spelling in the mid-1800s. However, it was not until the late nineteenth century that any country established a national authority that would be the guardian of the nation's toponyms, to avoid ambiguity in spelling or in application of the names. The first country to establish a board for this purpose was the United States of America. In 1890, the United States Board on Geographic Names (US BGN) was created by Executive Order of the President, Benjamin Harrison. This was their time of mapping associated with exploration, mining and settlement of the west, and this federal body was given authority to resolve questions concerning geographic names. Decisions of the Board became binding for all departments and agencies of the Federal Government.

Similar reasoning regarding westward expansion of settlement was behind the establishment of the Geographic Board of Canada in 1897. During the first half of the twentieth century, Denmark (1910), Iceland (1935), New Zealand (1946) and Ireland (1946) recognized the benefits of standardizing their geographical names and established national names authorities. These numbers have grown to around 80 countries having national authorities – influenced by the increased number of UN member states, with the independence of many countries in Africa in the 1960s, and the break-up of the Soviet Union and Yugoslavia in 1991. In recent years: Afghanistan, Brazil, Mozambique and Serbia created names boards in 2009; Sri Lanka and the Føroyar (Faroe Islands) region of Denmark established their boards in 2010; Saudi Arabia did likewise in 2013, while in the same year Burkina Faso and Tunisia re-established their national toponymic committees with new legislation. Figure 1-9 shows the creation of national names authorities over time, whereas Figure 1-10 shows the global coverage of geographical names authorities as self-determined by the countries prior to January 2015.

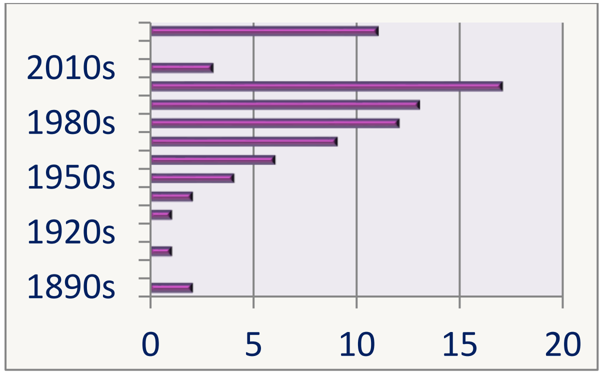


Figure 1-9 Growth of national names authorities – by date of first establishment

Figure 1-10 Geographical names authorities (as of January 2015). Green countries have national names authorities.

As recommended by resolution 4 of the First UN Conference on the Standardization of Geographical Names in 1967, the basis for national and hence international standardization rests with the countries of the world – that have the responsibility to make decisions on toponyms in their own jurisdictions. There is not just one model that can be followed by all countries in establishing a national names board! The nature of the government of the country (centralized or decentralized), the distribution of population, the number of recognized languages, and the cultural background may be among the factors that will influence the establishment of a toponymic authority. In basic terms, the names board is likely to be (a) a centralized government led names board, (possibly with advisors, or sub- or advisory-committees) and which takes decisions through this one group, or (b) a decentralized process, where decisions on names are made at the individual state/province level and accepted for national use, with an umbrella national group existing to make policy frameworks and provide international representation.

Just as the structure of a national names board may vary around the world (perhaps Madagascar is one of the largest boards, with 44 members), so too will their mandates. Most boards will have responsibility for names of topographic and hydrographic features and unincorporated smaller communities; in many cases, names of incorporated towns and administrative areas will be established by law, rather than directly by the board. Some authorities may have a mandate to decide on names of undersea and maritime features, ephemeral features (such as seasonal routes), and a few may have some jurisdiction over street, building and local park names (a topic more commonly left to municipalities for decision-making).

1.4 How are the names recorded and stored?

In the past, drawers of file cards carrying detailed records of each toponym and stored alphabetically were a common feature.

- A name card might include variant names in addition to the approved form; coordinates; administrative area where the feature is located; the history of the name; its language; the source of the name; office details of the development; date of Board approval, etc. (see Figure 1-11)

- In some offices, other sets of card records may have been kept, for example, cards with information gathered during field recording (Figure 1-12) or cards used in the preparation of gazetteer entries.

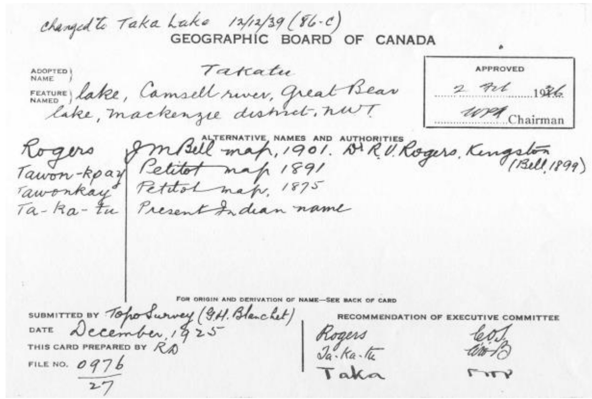


Figure 1-11 Old name record card of the Geographic Board of Canada (1926), showing variant forms found on historical maps, date of Board approval and other details.

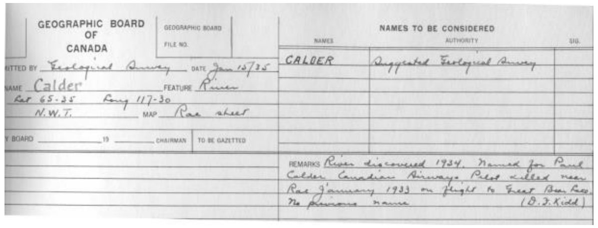

GEOGRAPHIC BOARD OF CANADA
GEOGRAPHIC BOARD FILE NO.
NAMES TO BE CONSIDERED
NAMES | AUTHORITY | SIG.
CALDER | Suggested Geological Survey
...ITTED BY Geological Survey DATE Jan 15/35
NAME Calder FEATURE River
Lat 65-35 Long 117-30
N.W.T. MAP Rae sheet
...Y BOARD 19 CHAIRMAN TO BE GAZETTED
REMARKS River discovered 1934. Named for Paul Calder Canadian Airways Pilot killed near Rae January 1933 on flight to Great Bear Lake. No previous name (D.J. Kidd)

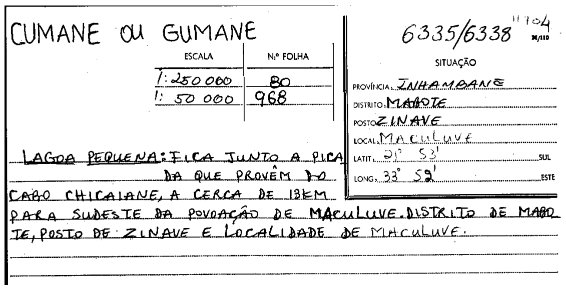
Figure 1-12 Example of a field record card of the Geological Survey of Canada, collected for consideration by the Geographic Board of Canada, 1935.

1.5 The development of digital toponymic databases

Today many national names authorities have moved to the use of digital toponymic databases, often of a sophisticated nature and integrated, or at least with the possibility of interoperability, with other place-related databases.

The data that formed the basis for today's databases, may have been collected from a variety of sources. In some cases, for instance in the United States, the names were captured directly by digitizing names data on existing published map sheets.

In other cases, for example Mozambique, existing names information was stored on cards filling many drawers. A systematic approach was needed to fit this data (often stored on handwritten or typewritten cards) to the fields established for the digital database. Questions arising through interpretation of handwriting needed to be solved, information needed to be entered in upper/lower case (not just capital letters), administrative units to be used for location had to be rationalized, coordinates had to be improved wherever possible, and reference to maps at various scales had to be consistently represented. These are just some of the details to be resolved in transferring rather "free-wheeling" early cards to the more rigorous structure of a database with specifications for each field of name attribute information. Figure 1-13 provides a card from existing Mozambique records that was entered into the national toponymic database.


CUMANE OU GUMANE
6335/6338 11704
SITUAÇÃO

ESCALA	N.º FOLHA
1:250 000	80
1: 50 000	968

PROVÍNCIA INHAMBANE
DISTRITO MABOTE
POSTO ZINAVE
LOCAL MACULUVE
LATIT. 21° 53' SUL
LONG. 33° 52' ESTE
LAGOA PEQUENA: FICA JUNTO A PICADA QUE PROVÊM DO CABO CHICAIANE, A CERCA DE 13KM PARA SUDESTE DA POVOAÇÃO DE MACULUVE. DISTRITO DE MABOTE, POSTO DE ZINAVE E LOCALIDADE DE MACULUVE.

Figure 1-13 Card record of Cumane or Gumane, Mozambique, as used in building a national toponymic database. Note that Cumane and Gumane were shown as alternative forms on the card; two records were created on the database and were given the identity numbers 6335 and 6338. These numbers were added to the card, for future follow up or possibly for the addition of a scanned version of the card to the digital records.

Figure 1-14 illustrates the record created for Cumane on the national database ("toponymic management system") showing the English-language interface (as opposed to the Portuguese one) indicating the data fields included. (Gumane was shown as a variant for Cumane and also has its own separate record.)

In some databases, toponyms are attributes of different layers of data (e.g. hydrography, land features, buildings, etc.), but generally today, a toponymic database can be a collection, or layer, of data, with each name record (with its own name id) linked to other georeferenced data, using a unique feature (object) id and supported by its coordinates.

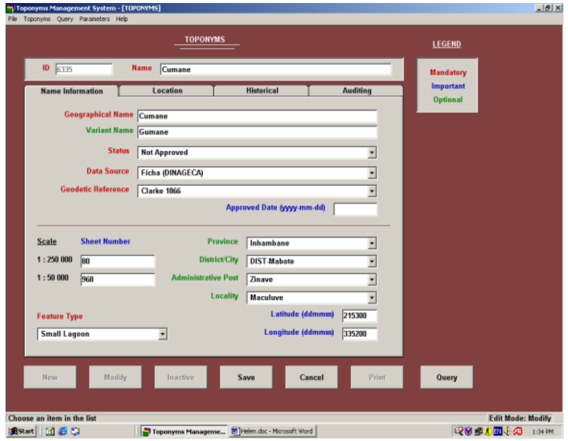


Figure 1-14 Database record created for Cumane (with variant Gumane), Mozambique.

Authorized geographical names should be easily available and accessible for public and private use – not set aside as a "best kept secret". Since the mid-1990s countries have been aiming to have their toponymic databases available through the Internet, so that the nationally approved forms of names can be queried, downloaded and used by governments and the general public. In some cases, the files may be gazetteer files, prepared at a particular point in time and/or for a particular map scale, and made available as Word or Excel files. In other cases, the public may have direct access to query a national geographical names database (likely a recent copy of the main toponymic database).

A number of these databases are described in documents presented to the UN Conferences on the Standardization of Geographical Names and at the intervening UNGEGN sessions. To single out a particular example, I would refer you to the database of Finland, in which names are related to topographic information and also to the process of names selection, font size, etc. for use in by the National Mapping Agency. Mr. Teemu Leskinnen presented the most recent version of this document to the ICA in Rio de Janeiro in 2015 (icaci.org/files/documents/ICC_proceedings/ICC2015/papers/21/fullpaper/T21-516_1430431470.pdf).

1.6 What can a toponymic database achieve?

The use of a toponymic database has a wider application than just names for topographic maps, and the use of *standardized* written forms (spelling, diacritics, hyphens, capitalization, etc.) can provide a number advantages, to both individuals and to a country, over ambiguous situations where it is unclear which form(s) of a toponym are authorized.

Apparently approximately 80% of information affecting our daily lives has a spatial dimension. In all "place" referenced databases, geographical names are an important layer of the spatial data and a natural way of querying to find other data, so becoming vital keys to access the digital world. As a common field of information they help enable the integration of digital data sets, which become essential and powerful decision-making tools for policy makers and managers in business and government. At both national and local levels, accurate and consistent geographical names are vital for the optimal use of geo-information in such fields as:

- sustainable national planning strategies,
- environmental management,
- police quick response,
- emergency services coordination,
- trade and delivery systems,
- utility infrastructure development,
- tourism,
- communication and media services,
- cultural heritage promotion,
- registration of land holdings,
- census and statistical social surveys,
- signposting on local roads and major routeways,
- developing on-board navigation systems.

In cases, such as tsunami warnings, dangers from forest fires, climate change vulnerability, geographical names are essential to the planning and provision of emergency aid.

By way of example, low-lying areas of the west coast of Vancouver Island, Canada, are vulnerable to tsunami perils. Carefully developed map and GIS plots of the terrain with the standardized microtoponyms (small localities, street names) have been created to indicate flood levels, and road signs have been posted on the few existing routes to indicate the necessary direction of travel to attain higher ground (see Figure 1-15).

Figure 1-15 Tsunami evacuation route indicators on western Vancouver Island, Canada.

We can see that databases of clear, unambiguous (i.e. standardized) toponyms can be of benefit to us in all the situations mentioned. The benefits can be considered to cover four major areas - although with overlapping boundaries - namely: technical, economic, social, and cultural benefits, as shown with a few examples in Figure 1-16.

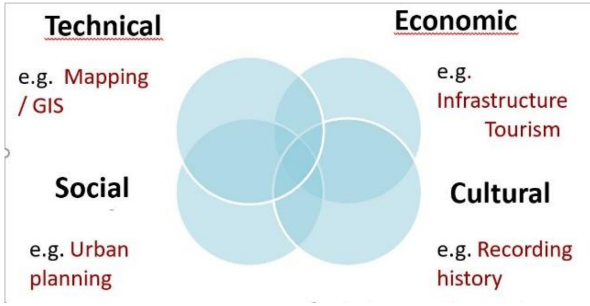


Figure 1-16 Technical, economic, social and cultural benefits of accessible standardized toponyms

n addition, today the use of crowd-sourced data has been noted as valuable to aid response in critical situations. Although the names may have no official status, their utility as unofficial (or perhaps variant names) in a database has been seen as advantageous, for instance, in serving police responses (as in the local knowledge collected by the Ordnance Survey in the United Kingdom in the Location Lingo Project) and in guiding emergency response (as in the collection of street and building names immediately following the earthquake in Haiti in 2010).

1.7 Problems from the lack of standardized toponyms

Two examples can be used to indicate the problems of lack of available standardized toponyms and associated data in times of crisis.

- In its coordination of relief efforts, the United Nations requires up-to-date toponymic data. In 2005, the severe earthquake near Balakot in a remote valley of the Himalayas in northern Pakistan resulted in over 88,000 losing their lives. UN OCHA reported that among the difficulties that made the humanitarian aid more difficult and caused delays in providing assistance, were the lack of standardized names and available gazetteers, as well as lack of access to maps, population statistics, and locations of the villages. Such data is needed for input into GIS to enable rapid response.

- UNECA (2014) reported on the situation in Somalia where lack of standardization of toponyms, repetition and duplication of toponyms, and incomplete data were problematic. For instance, Xuddur, Oddur, Xudur, Xuudur, Huudur were written forms for the same place. Baydhaba, Baydhabo, and Baidoa were equally ambiguous, as were Beledweyne, Belet Weyne, Beled Weyn, Balet Wayn. Such inconsistencies across sources lead to loss of resources, possible compromise of security, confusion and often bad decisions being made.

Street naming and addressing

Another type of situation was pointed out to UNGEGN at its session in Nairobi in 2009. In the slum areas of the city, the absence of street names and any addressing system contributed to the lack of personal identity. This included problems for an individual to receive delivered goods or to open an account that requires a residential address.

A report of the World Bank published in 2005, indicated that over 50% of urban centres in sub-Saharan Africa lacked an addressing system. Even if the city core originally had such a system, densification of the city centres and urban expansion had taken place without the assignment of names keeping pace. As this is essential for town planning, taxation systems, plotting the spread of epidemics and so on, the World Bank provided financial support for countries to develop systems of addressing and the assignment of building numbers, together with appropriate database initiatives. As street numbering is less controversial than the assignment of street names (odonyms), this was considered the first step. The World Bank recommended that streets for priority naming should be identified, lists of potential names should be created, the residents should be able to consider choices and the municipality should approve the name.

In Yaoundé, Cameroon, for instance, the urban area was divided into six zones, with the 1670 streets numbered within each zone, and buildings numbered according to odd and even sides of the street. Gradually as names are decided they can be added to the street signs as decisions are taken (see Figure 1-17).

Figure 1-17 Yaoundé, Cameroon: Street 74 in Zone 1; Street 29 in Zone 2, also referred to as rue Malam Ibrahim

The city of Ouagadougou in Burkina Faso has set up a municipal Commission de toponymie, developed principles of naming and the process of approving street names and storing these odonyms, as well as creating a detailed street map of the city (sample in Figure 1-18). A list of potential names was compiled – including mainly names of people; by 2005, 2000 streets were named compared to only 70 (of 4910) in 1997.

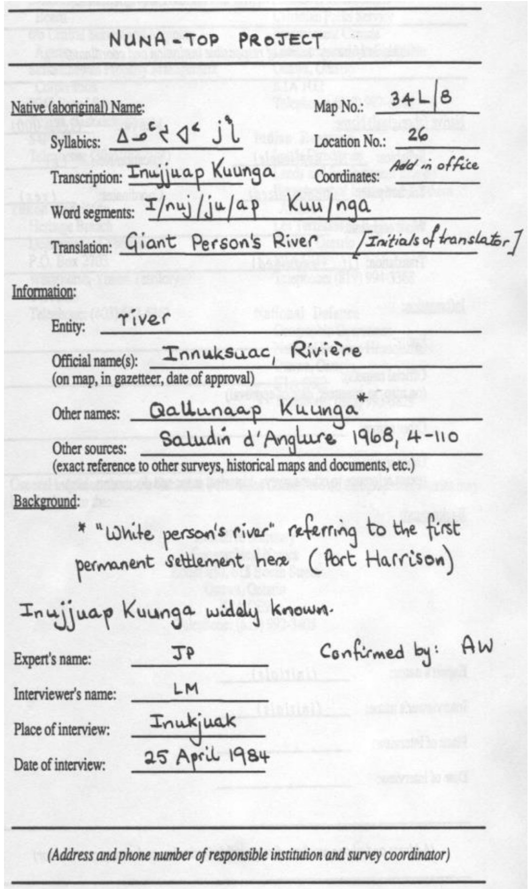
Figure 1-18 Ouagadougou, Burkina Faso: part of the street map created for the city; street sign in the Ouaga 2000 sector of Bogodogo

1.8 Romanization

Toponymic databases may be maintained in any language and writing system. China, for example, has reported to UNGEGN on their authorized geographical names data being available and searchable through publicly accessible computer terminals. Other non-romanized writing systems may provide rich toponymic databases, as for instance in Arabic, Greek, Korean or Russian Cyrillic characters. For international use, UNGEGN favours the use of single scientifically-based romanization systems as a means of converting toponyms into the Roman (Latin) alphabet. As of the 10th UN Conference on the Standardization of Geographical Names in 2012, 30 romanization systems (with conversion tables) had been adopted for use through UN resolutions. Of these 30, some are still under review for their continued suitability or their need for modification. Single romanization systems for scripts of other languages continue to be discussed and refined before their authorization and adoption.

Products from these toponymic databases may include both the original script and romanized forms. Inuktitut (or Inuttitut) as used in Northern Quebec (and Eastern Nunavut) in Canada is generally written in syllabics (i.e. with symbols for syllable sounds, as for example: pi ᐱ, pu ᐳ, pa ᐸ). Syllabics are recorded during field collection and included in toponymic databases for the area. For Nunavik (=Northern Quebec) a multilingual gazetteer was produced from the toponyms collected by surveys in the Inuit communities. The gazetteer used Inuktitut syllabics and romanized forms of the toponyms, with the addition of French and English for the explanatory texts (see Figures 1-19 and 1-20).

Difficulties can arise with data compatibility and character display of various scripts, hindering the necessary exchange of data between systems. UNGEGN addressed this issue in 1998 and concluded that the use of the Unicode standard (ISO/IEC 10646) best addressed these potential difficulties.



NUNA-TOP PROJECT

Native (aboriginal) Name:
Syllabics: ᐃᓄᑦᔪᐊᑉ ᑰᖓ — Map No.: 34L/8
Transcription: Inujjuap Kuunga — Location No.: 26
Word segments: I/nuj/ju/ap Kuu/nga — Coordinates: Add in office
Translation: Giant Person's River [Initials of translator]

Information:
Entity: river
Official name(s): Innuksuac, Rivière
(on map, in gazetteer, date of approval)
Other names: Qallunaap Kuunga*
Other sources: Saladin d'Anglure 1968, 4-110
(exact reference to other surveys, historical maps and documents, etc.)

Background:
* "White person's river" referring to the first permanent settlement here (Port Harrison)
Inujjuap Kuunga widely known.

Expert's name: JP — Confirmed by: AW
Interviewer's name: LM
Place of interview: Inukjuak
Date of interview: 25 April 1984

(Address and phone number of responsible institution and survey coordinator)

Figure 1-19 Field collection form for the recording of Inuit place names in Inukjuak, Quebec

Figure 1-20 The Gazetteer of Inuit Place Names in Nunavik (Quebec, Canada), published from the database of names collected in the field.

1.9 Geographical names and cultural heritage

The value of geographical names as part of our cultural heritage has had considerable discussion at recent UNGEGN sessions, special regional meetings and among members of the UNGEGN Working Group on Geographical Names and Cultural Heritage. Of the tens of thousands of geographical that have been used and recorded over many generations (and sometimes many centuries), some refer to major features, others may be for local community features, such as farm and field names (for example in Figure 1-21), and yet more may record names of heritage structures (for example in Figure 1-22). All are an important part of our web of knowledge and of the languages, history and customs that are the framework for the toponyms we use today and so constitute a significant element of our identity and our association with "place".

Figure 1-21 Local names of holdings in southern Iceland

Figure 1-22 Preservation of cultural heritage in the vicinity of Yogyakarta, Indonesia

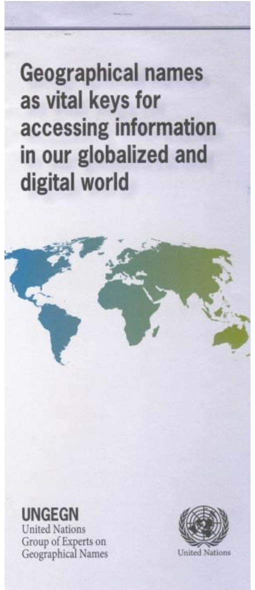
In reference to odonyms, the President of the municipal Commission de toponymie in Ouagadougou, Burkina Faso, indicated that: "Paths, spaces and boulevards shape life from afar and leave their mark on movements of thought and activity. As a result, streets, squares and buildings define and reflect the essence of the nation, in its sovereignty, its history and its culture." [translation]

Many toponymic records have been gathered (from archives, from current materials, and from oral tradition) to create collections of culturally significant names and their attributes. In the past, linguists and others involved with such collections have not necessarily recorded locational data for the named places (e.g. coordinates) in enough detail to allow these valuable records to be integrated easily and accurately into today's digital toponymic/topographical databases.

As data is more rigorously gathered and as the links between databases become better developed, it will be easier to roll in these elements of our collective memory and heritage. Being able to query (often by toponym) across a variety of databases will add considerably to the robustness of accessible data, and will help protect and preserve our toponymic heritage.

1.10 Conclusion

UNGEGN has since the 1960s been promoting national and international standardization of geographical names. On the UNGEGN website, you will find documents presented on this subject over the span of more than 50 years, as well as links to websites of the UNGEGN working groups, UNGEGN divisions, and individual countries. Manuals and pamphlets produced by UNGEGN and published by the United Nations are available to view and download for your use (see examples in Figures 1-23 and 1-24).



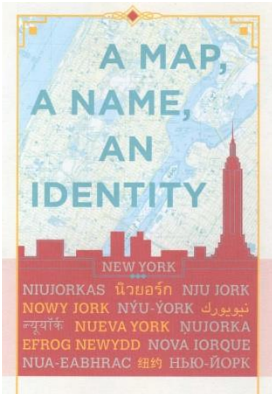


Figure 1-23 Example of a publicity brochure produced by UNGEGN

Figure 1-24 Promotional postcard, showing various exonyms used for New York, published on the occasion of the 10th UN Conference on the Standardization of Geographical Names held in New York in 2012.

In this manual, you will find a series of articles prepared by experts in various aspects of toponymy and national standardization. The material is offered to you as a follow up to the *Manual for the national standardization of geographical names* and the Toponymic web course (see Chapter 2) prepared in cooperation with the International Cartographic Association.

1.11 References

Farvacque-Vitkovic, Catherine, Lucien Godin, Hugues Leroux, Florence Verdet and Roberto Chavez (2005): *Street Addressing and the Management of Cities*. World Bank, Washington.

Kadmon, Naftali (2000). *Toponymy: the Lore, Laws and Language of Geographical Names.* New York: Vantage Press.

Kerfoot, Helen (2011). *Toponymy and administration: from plazas to planets … with particular reference to United Nations' work.* Keynote address at the International Congress of Onomastic Sciences, Barcelona.

Parker, John (2001). *The importance of geographic names in a spatial data infrastructure.* E/CONF.93/B.P.1. Invited paper at the Seventh UN Regional Cartographic Conference of the Americas, New York. (http://unstats.un.org/unsd/geoinfo/ungegn/default/other Material.html)

Ulgen, Suha and Williams, Craig (2007): *Standardization of geographic names in humanitarian information management.* Presentation to the Ninth UN Conference on Geographical Names (unstats.un.org/unsd/geoinfo/Secial-Presentations_Docs/Presentation-OCHA.pdf)

UNGEGN (2012). *Resolutions adopted at the ten United Nations Conferences on the Standardization of Geographical Names (1967-2012).* Available as a document on the UNGEGN website and as a searchable database. Both forms are in English, French, and Spanish versions.

UNGEGN (2006). *Manual for the national standardization of geographical names.* New York: United Nations. (Available in pdf format under "Publications" on the UNGEGN website.)

UNGEGN website: http://unstats.un.org/unsd/geoinfo/UNGEGN conference and session documents, brochures, bulletins, resolutions, media kit, longer UNGEGN publications – all available to download; links to websites of UNGEGN Working Groups, Divisions, and individual country databases and toponymic materials.

Section 1 General/strategic issues

Chapter 2 **Toponymy web course**

Ferjan Ormeling

2.1 Introduction

Just to show what we regard as the knowledge possessed by all those to whom the UNGEGN toponymy manual is directed, we will deal with the content of the toponymy web course developed by the International Cartographic Association and the UNGEGN. It can be accessed from the UNGEGN website (see figure 2-1) as well as the ICA Commission on Education and Training website (see http://lazarus.elte.hu/cet/ , under 'Internet Cartography Teaching Courses').

Figure 2-1 UNGEGN website

The UNGEGN website, accessed through http://unstats.un.org/unsd/geoinfo/UNGEGN leads through the UNGEGN Working Group on Training Courses in Toponymy and its link to the English language toponymy web course (see figure 2-2) to the actual web course site at http://unstats.un.org/unsd/geoinfo/UNGEGN/docs/_data_ICAcourses/index.html . It consists of 20 modules, each dealing with a specific aspect of toponymy, and consisting of the actual lessons, exercises (and the answers to them), relevant documents (full text) and links to further literature or other relevant websites.

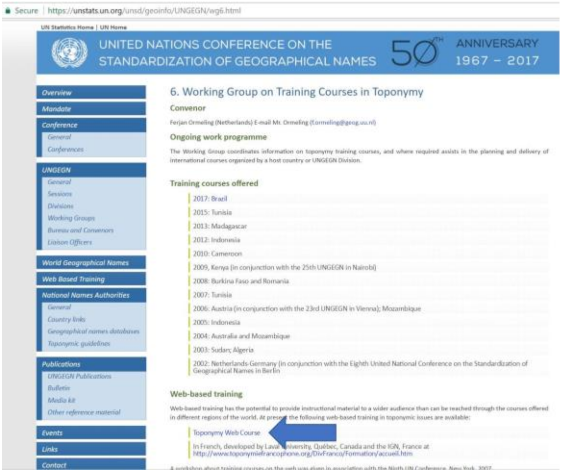


Figure 2-2 WG Training Courses in Toponymy

This web course (see figure 2-3) has been structured in such a way that it is meant to be self-explanatory, and that one can follow it independently, with the help of the exercises and documents provided online.

2.2 Web course contents and structure

After an introductory module (see figure 2-4 for the modules that together compose the course), explaining the need for standardization of geographical names, the

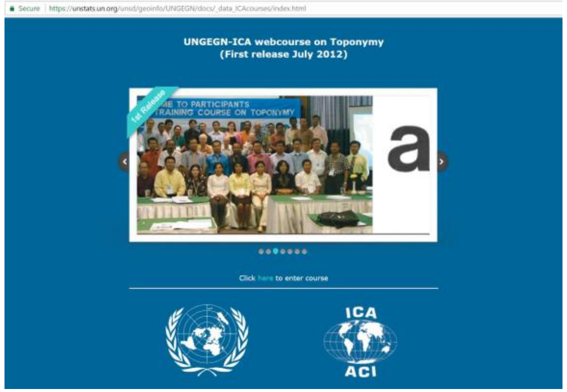


Figure 2-3 UNGEGN/ICA Toponymy web course

course first deals with the naming process (module 2), explaining why humankind gives names to the spatial objects in its environment, what kind of names are given, and how these names develop over time. Once given, these geographical names tend to get other functions (module 3), next to their primary orientation function: they become links in data exchange, but also serve as symbols, are used as brand names and as educational concepts.

The course then follows the institutional process developed by national spatial data agencies (module 4), by describing how names are collected during field work (module 5), necessitating not only the recording of their correct spelling and pronunciation, but also their georeferencing (module 6).

Once collected and georeferenced in the field, the geographical names are being processed at the office, prior to their standardization (module 7), and this processing requires quite a number of considerations, depending on the actual situation: there might be different language groups, each with its own set of geographical names, living within the area for which the names have been collected (module 8, multilingual areas), and this in turn necessitates dealing with languages (module 9), writing systems (module 10) and the conversion of one writing system to another (module 11). These are complex issues, and as the UNGEGN has described and documented a number of best practices in its publications, these are described in module 12.

Once geographical names have been collected and standardised nationally, they are used in school and reference atlases, in the news media, by press agencies, etc., and in this international environment it becomes necessary to deal with the issue of exonyms (module13). To help with these international issues, UNGEGN has induced individual nations to each publish its toponymical guidelines (module 14), to help map and news editors from other countries to deal with the nation's geographical names.

UNGEGN-ICA webcourse on Toponymy - PROGRAM

COURSE MODULES	SELF STUDY	EXERCISE(S)	DOCUMENT(S)
1 Introduction	S01	E01	D01
2 The naming Process	S02	E02	D02
3 Functions of geographical names	S03	E03	D03
4 National Agencies	S04	E04	D04
5 Field collection systems	S05	E05	D05
6 Reference systems	S06	E06	D06
7 Office processing, standardization	S07	E07	D07
8 Multilingual areas, minority names	S08	E08	D08
9 Languages	S09	E09	D09
10 Writing systems	S10	E10	D10
11 Conversion systems	S11	E11	D11
12 UN Role	S12	E12	D12
13 Exonyms	S13	E13	D13
14 Toponymical Guidelines	S14	E14	D14
15 Toponymical data files	S15	E15	D15
16 Names placement	S16	E16	D16
17 Legal status of names	S17	E17	D17
18 Editorial issues (atlases)	S18	E18	D18
19 Toponymical Planning	S19	E19	D19
20 Names as cultural heritage	S20	E20	D20

Figure 2-4 Web course modules contents

Back again to the institutions that have standardised the names, their work isn't done as yet, as they have to make the standardised geographical names available to the public at large, and this calls for the production of gazetteers, names data bases and names servers (module 15). The names now can be entered on maps, and module 16 shows how that is done. All the other editorial aspects of applying the geographical names to atlases, and how to proceed consistently here, are dealt with in module 18.

In the description of the course contents, I have forgotten module 17, which is status-related and shows how names are made official, for example by place name acts. When changes in place names are foreseen, to avoid ambiguity, this requires careful planning and intermediate periods in which both the old and new names are valid, a process described in module 19.

Finally, module 20 deals with geographical names as our cultural heritage, something of value to be preserved, even if no longer referred to in our day-to-day usage, as humankind is becoming increasingly urbanised.

2.3 Concepts dealt with in the web course

In the course of the 20 modules, a number of concepts are introduced, explained and defined. The primary concepts, of course, are *geographical names* (this term will be alternated with *toponyms* and/or *geonames)*, and *standardization*. *International standardization of geonames* can only occur when these names have been standardized on a national basis first. Here we have the concept of *univocity*, the ideal of having only one unique, specific name for every geographical object. When studying geonames, immediately the concepts of *variant* *names* /*allonym*s (different name versions for the same geographical object) and *homonyms* (similar names for different objects) will crop up, as reasons, why standardization is necessary in order to decrease the ambiguity in dealing with geonames.

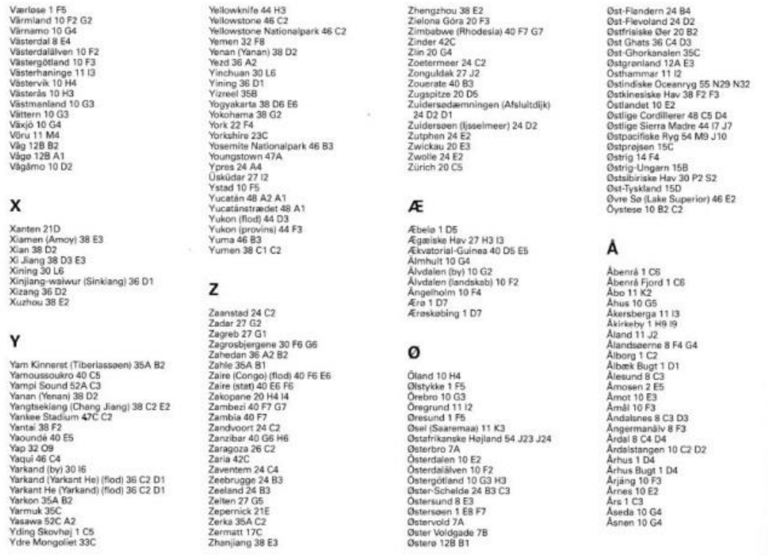
Geonames are differentiated according to categories, such as *oronyms* (mountain names), *odonyms* (road names), *choronyms* (area names), *hydronyms* (water names), and may further be differentiated according to: *anthro(to)ponyms* (objects named after people) and *hagionyms* (religious names). There also are the distinctions between current toponyms and *historical names*, and between (locally official) toponyms, *conventional names* and *exonyms*.

In many geonyms we can spot the *generic* and *specific name parts*, we may differentiate between *simplex names* (single word toponyms) and *composite names*, the latter consisting of more elements. Then there are name attributes like *name gender* (is it le fleuve or la fleuve?) and *name number* (is it plural or singular?). In some languages the form of the toponyms changes, depending on its function in the sentence, leading to different case endings: we go to Kiev: до Києва (do Kieva) vs. we are staying in Kiev: в Києві (v Kievi) vs. We have left Kiev: Київ (Kiev). Apart of such changes in case there may also occur the joining of affixes or suffixes, like *postpositional generics* (like the ending on øya, that is –island, e.g. Hinnøya, Flakstadøya) or the names may occur in definite or indefinite form: elv is a river, but elven is the river in Swedish (-en here is the *definite article*).

When dealing with toponyms in other countries then our own (for instance when producing atlases or when reading news bulletins) we are confronted with the existence of *exonyms* versus *endonyms*. We might tackle this issue by referring to the *Local names policy* (the use of locally official names) already adhered to in the 1880s. The next step would be to ascertain whether the names we want to use are in the same alphabet as ours or in different writing systems. The *Roman alphabet rule* states that, when confronted with names from other countries using the Roman alphabet, we have to copy their spelling, including all the *diacritical signs* they may contain. For countries not using the Roman alphabet, we have to turn to conversion systems, such as *transcription* or *transliteration*, decided by either one or the combination of the *source country* and the *receiver country*. When such systems are used to convert names in non-Roman alphabets or writing systems to ours, we call them *Romanization systems*.

Onomastics and *etymology* study the original meaning (the *semantics*) of toponyms, which may be *descriptive names, commemorative names, commercial names, statutory names* (names given according to the statutes, mostly of administrative entities.

Toponyms in specific areas belong to specific languages, and these languages, each with its specific *gramm*ar and *syntax*, decide how these names are written, spelled, abbreviated or divided. Different languages have different alphabetisation rules (see figure 2-5), and different for capitalisation and hyphenation as well. Therefore we need to know what languages are either *official, dead, disappeared, literal, national or minority languages* or *dialects* (local varieties of a language) in that area.


Værløse 1 F5
Värmland 10 F2 G2
Värnamo 10 G4
Västerdal 8 E4
Västerdalälven 10 F2
Västergötland 10 F3
Västerhaninge 11 I3
Västervik 10 H4
Västerås 10 H3
Västmanland 10 G3
Vättern 10 G3
Växjö 10 G4
Võru 11 M4
Våg 12B B2
Vågø 12B A1
Vågåmo 10 D2

X

Xanten 21D
Xiamen (Amoy) 38 E3
Xian 38 D2
Xi Jiang 38 D3 E3
Xining 30 L6
Xinjiang-waiwur (Sinkiang) 36 D1
Xizang 36 D2
Xuzhou 38 E2

Y

Yam Kinneret (Tiberiassøen) 35A B2
Yamoussoukro 40 C5
Yampi Sound 52A C3
Yanan (Yenan) 38 D2
Yangtsekiang (Chang Jiang) 38 C2 E2
Yankee Stadium 47C C2
Yantai 38 F2
Yaoundé 40 E5
Yap 32 O9
Yaqui 46 C4
Yarkand (by) 30 I6
Yarkand (Yarkant He) (flod) 36 C2 D1
Yarkant He (Yarkand) (flod) 36 C2 D1
Yarkon 35A B2
Yarmuk 35C
Yasawa 52C A2
Yding Skovhøj 1 C5
Ydre Mongoliet 33C
Yellowknife 44 H3
Yellowstone 46 C2
Yellowstone Nationalpark 46 C2
Yemen 32 F8
Yenan (Yanan) 38 D2
Yezd 36 A2
Yinchuan 30 L6
Yining 36 D1
Yizreel 35B
Yogyakarta 38 D6 E6
Yokohama 38 G2
York 22 F4
Yorkshire 23C
Yosemite Nationalpark 46 B3
Youngstown 47A
Ypres 24 A4
Üsküdar 27 I2
Ystad 10 F5
Yucatán 48 A2 A1
Yucatánstrædet 48 A1
Yukon (flod) 44 D3
Yukon (provins) 44 F3
Yuma 46 B3
Yumen 38 C1 C2

Z

Zaanstad 24 C2
Zadar 27 G2
Zagreb 27 G1
Zagrosbjergene 30 F6 G6
Zahedan 36 A2 B2
Zahle 35A B1
Zaire (Congo) (flod) 40 F6 E6
Zaire (stat) 40 E6 F6
Zakopane 20 H4 I4
Zambezi 40 F7 G7
Zambia 40 F7
Zandvoort 24 C2
Zanzibar 40 G6 H6
Zaragoza 26 C2
Zaria 42C
Zaventem 24 C4
Zeebrugge 24 B3
Zeeland 24 B3
Zelten 27 G5
Zepernick 21E
Zerka 35A C2
Zermatt 17C
Zhanjiang 38 E3
Zhengzhou 38 E2
Zielona Góra 20 F3
Zimbabwe (Rhodesia) 40 F7 G7
Zinder 42C
Zlin 20 G4
Zoetermeer 24 C2
Zonguldak 27 J2
Zouerate 40 B3
Zugspitze 20 D5
Zuidersødæmningen (Afsluitdijk) 24 D2 D1
Zuidersøen (Ijsselmeer) 24 D2
Zutphen 24 E2
Zwickau 20 E3
Zwolle 24 E2
Zürich 20 C5

Æ

Æbelø 1 D5
Ægæiske Hav 27 H3 I3
Ækvatorial-Guinea 40 D5 E5
Älmhult 10 G4
Älvdalen (by) 10 G2
Älvdalen (landskab) 10 F2
Ängelholm 10 F4
Ærø 1 D7
Ærøskøbing 1 D7

Ø

Öland 10 H4
Ølstykke 1 F5
Örebro 10 G3
Öregrund 11 I2
Øresund 1 F5
Øsel (Saaremaa) 11 K3
Østafrikanske Højland 54 J23 J24
Østerbro 7A
Österdalen 10 E2
Österdalälven 10 F2
Östergötland 10 G3 H3
Øster-Schelde 24 B3 C3
Östersund 8 E3
Østersøen 1 E8 F7
Østervold 7A
Øster Voldgade 7B
Østerø 12B B1
Øst-Flandern 24 B4
Øst-Flevoland 24 D2
Østfrisiske Øer 20 B2
Øst Ghats 36 C4 D3
Øst-Ghorkanalen 35C
Østgrønland 12A E3
Östhammar 11 I2
Østindiske Oceanryg 55 N29 N32
Østkinesiske Hav 38 F2 F3
Östlandet 10 E2
Østlige Cordillerer 48 C5 D4
Østlige Sierra Madre 44 I7 J7
Østpacifiske Ryg 54 M9 J10
Østprøjsen 15C
Østrig 14 F4
Østrig-Ungarn 15B
Østsibiriske Hav 30 P2 S2
Øst-Tyskland 15D
Øvre Sø (Lake Superior) 46 E2
Öystese 10 B2 C2

Å

Åbenrå 1 C6
Åbenrå Fjord 1 C6
Åbo 11 K2
Åhus 10 G5
Åkersberga 11 I3
Åkirkeby 1 H9 I9
Åland 11 J2
Ålandsøerne 8 F4 G4
Ålborg 1 C2
Ålbæk Bugt 1 D1
Ålesund 8 C3
Åmosen 2 E5
Åmot 10 E3
Åmål 10 F3
Åndalsnes 8 C3 D3
Ångermanälv 8 F3
Årdal 8 C4 D4
Årdalstangen 10 C2 D2
Århus 1 D4
Århus Bugt 1 D4
Årjäng 10 F3
Årnes 10 E2
Års 1 C3
Åseda 10 G4
Åsnen 10 G4

Figure 2-5 Part of the names register of a Swedish school atlas, showing the sequence of names starting with a letter to which diacritical signs have been applied. (With permission of Noordhoff Atlas productions

In areas where minority languages are spoken, we might witness the effects of *topographic bias* in the collection of geographical names, leading to an under-representation of minority names. In areas where the right of speakers of the *minority language groups* are recognised, the maps produced will show both the toponyms and the *marginal information* (legend, scale indicator, imprint, reliability diagram on the map sheet outside the map area) *bilingually* or even *biscriptually*, that is both in the official and the minority language and their writing systems.

Languages are written in different scripts (see figure 2-6), and it is considered relevant to know how these developed. Originally, *pictograms*, that is purely pictorial representations of concepts (*ideograms)* or words (*logograms*) were used, but for many languages these logograms developed into *phonograms* in which the sound value of mono-syllabic words becomes attached to the symbol representing these words. Finally the *syllabic script* develops into an alphabetic script in which symbols represent single *phonemes* instead of syllables. As a result of this development we have *ideographic* and *logographic scripts*, *syllabic scripts*, and *alphabetic scripts* that can either be vocalised or non-vocalised (see figure 2-7).

Figure 2-6 Part of Southeast Asia, showing the different scripts in use. © Menno Bolder

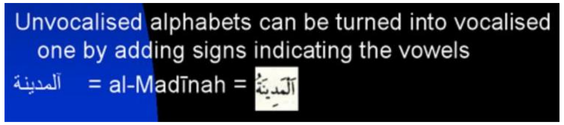


Figure 2-7 Non-vocalised and vocalised names

When producing toponymical databases, from which later gazetteers and/or name servers can be derived, one has to make sure that the necessary attribute information for both the name and the accompanying named object is stored in the database. Apart from ID numbers, these necessary attributes may consist of the feature code, coordinates, and extent of the named object, and the language, gender, number and pronunciation of the name.

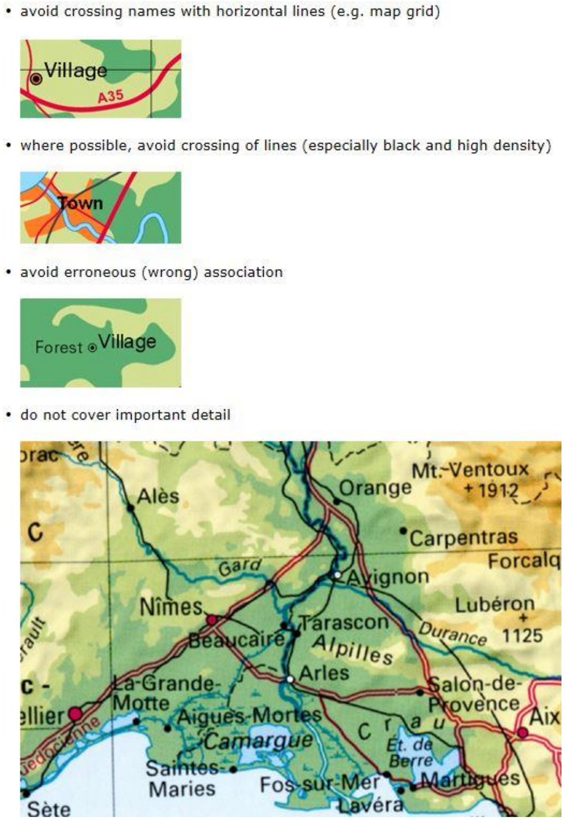


Figure 2-8 Capture of an illustration from the module on names placement in the web course. © Noordhoff Atlas Productions.

When the object of our toponymical databases also is producing maps, we might add information on the map sheet(s) the named object will occur on, and its relative importance, for incorporation on derived, smaller scale maps.

The toponymical databases also serve to help us to correctly name the objects or features on our maps. So the course also deals with applying the names to the maps, in such a way that they are legible, placed unambiguously (so that it will be clear to which mapped object the names refer, see figure 2-8), and that by diversifying the fonts used, each matched to a specific feature category (such as lower case italics used to render river names) it becomes easier to identify a name belonging to a specific feature category on the map.

Kerning (the adjustment of the space between two consecutive letters) is a concept from typography, as are *serifs*.

Finally, in the module on Names as cultural heritage, we deal with concepts such as (toponyms as*) landscape identifiers*, leading to mental or emotional associations, linked to the connotations of names discussed in the module on name functions.

2.4 Processes included in the course

When doing the web course, students were asked to look up definitions, find literature, follow links to other relevant material on the website, and generally explore the wealth of toponymical material available on the UNGEGN-website. Especially the following, downloadable publications should be mentioned:

Glossary of Terms for the Standardization of Geographical Names (UN-New York 2002) / pdf

Manual for the national standardization of geographical names (UN - ECOSOC, New York, 2006 available in the 6 UN languages) / pdf

Technical reference manual for the standardization of geographical names (New York, 2007) / pdf

Resolutions adopted at the ten UN Conferences on the standardization of geographical names (English (pdf) / French (pdf)

Apart from these educational publications, individual working papers handed in by delegates for UNGEGN-sessions or UNCSGN-conferences were referred to as well, and had to be accessed by course participants.

Moreover, participants had the opportunity of doing exercises in looking up data fields, georeferencing, in matching maps, identifying writing systems, in name transcription, the creation of a names data base, producing a geographical names index, solving editorial issues and adapting a map to a new dispensation.

2.5 Final remark:

In order to highlight the educational possibilities of the web course, access to it has been eased, by displaying it more prominently on the UNGEGN website (under Web Based Training), and simplifying its url: as can be seen in figure 2-9. The URL to access it directly remains: https://unstats.un.org/unsd/geoinfo/UNGEGN/docs/_data_ICAcourses/index.html

Figure 2-9 Easy access to the Web based training course

Section 2 Management of a national names programme:

Chapter 3 **Maintaining and publishing the geographical names database in Germany**

Pier-Giorgio Zaccheddu

3.1 Introduction

In Germany, all geographical names information of several (topographic) vector data products based on scale level 1: 200,000 and smaller is maintained in a single database called "*Geographical Names of Germany* (GN-DE)". Through unique identifiers the names entries of the GN-DE are linked unambiguously to all databases and products provided by the Federal Agency for Cartography and Geodesy (BKG). At present, the database contains about 165 000 entries including more than 700 names in the minority languages Sorbian and Frisian. The administration and maintenance tasks of the GN-DE are conducted in Frankfurt am Main at the headquarters of BKG, the German national mapping agency responsible for small to medium scale mapping, whereas the publication of the geographical names data is facilitated through a web (gazetteer-) service by the BKG Central Service and Distribution Centre for Geoinformation in Leipzig [1]

The following information summarizes the tasks and resources needed to set up and maintain the management of a national names programme. The figures are based on rough estimations as the logistics, production time needed or necessary upkeep costs can vary significantly according to the time and resources needed for the adaptation of the database, the services and applications to new technical developments or to the services-oriented architecture in the organization.

3.2 The Geographical names standardization process in Germany

In Germany, a decentralized responsibility between the federal government and the Länder (federal states) exists and this must be considered for the elaboration and assessment of the management of a national geographical names programme. The government in Germany has three distinct levels of public authority: local, regional and national, all of which are generators and holders of public and geospatial information. Surveying and mapping is the task of the 16 Länder. Each of the Länder is responsible for its own topographic and cadastral service, environmental and statistical data collection, and in general for data policies. Traditionally, data collection is largely decentralized and carried out mostly on the regional and local level, which means that the processing and maintenance of data is mostly tailored to local and regional requirements, which leads to a built-in incompatibility. The different Länder have issued laws that regulate the work, which the regional and local authorities are carrying out. The "Surveying and Cadastral Acts" may serve as an example. The surveying and mapping administrations of the Länder are responsible for creating and maintaining the geospatial (reference) data describing real estate and the landscape, including geographical names. Whereas the mapping agencies of the Länder are responsible for providing large and medium scale reference data, the cadastre offices in the Länder support them by performing the tasks of the real estate cadastre.

The responsible regional mapping authorities are collaborating in the Working Committee of the Surveying Authorities of the Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland – AdV [2]). The AdV provides joint documentation regarding the geospatial reference data available as well as a joint pricing and data collection policy. ATKIS® (Authoritative Topographic-Cartographic Information System), the central topographic programme of the German national survey is the major joint project of the 16 Länder surveying authorities and BKG. The product line comprises digital landscape models (DLM), digital terrain models (DGM) and the digital topographic map series 1: 25,000 to 1: 1,000,000. There is an authorization by agreements between the Federal Administration and the Länder on the production of topographic reference data (including maps). Amongst others, all geospatial reference data sets based on scales larger than 1: 200,000 are done by the federal states, while the data sets with equal and smaller scales are compiled by BKG.

The 'Bundesgeoreferenzdatengesetz (BGeoRG)', the Federal Geospatial Reference Data Act, regulates the quality proofed and standardized provision of geodetic reference systems and frameworks as well as of the geospatial reference data products of the federal administration. Currently, geospatial reference data has varying degrees of quality e.g. in terms of content density. Much effort in developing technical methods and personnel resources is needed to aggregate the geospatial reference data in a way that allows for a solid and quality proofed geo-referencing of geospatial thematic data (e.g. statistical analysis or environmental indicators). For improving this unsatisfactory situation, BGeoRG instructs BKG to evolve its competence and capability to a central geospatial reference data and information service provider for the federal administration. Furthermore, the BGeoRG assigns BKG to advise the Federal Government in all questions related to geodesy and geospatial information and it safeguards the relevant German interests at the European and international level, e.g. within the United Nations.

Focusing on the geographical names standardization and maintenance process, the following institutions are involved in Germany:

- communes, counties, Länder (for populated places and administrative units)
- Länder survey administrations (for geographic regions/landscapes)
- federal and state hydrographic administration (for hydrographic features)
- federal and Länder transport administrations (for streets and railways)
- StAGN – Permanent Committee on Geographical Names (responsible for orthographic rules and principles for place names)

3.3 Maintaining and publishing the national geographical names database (GN-DE)

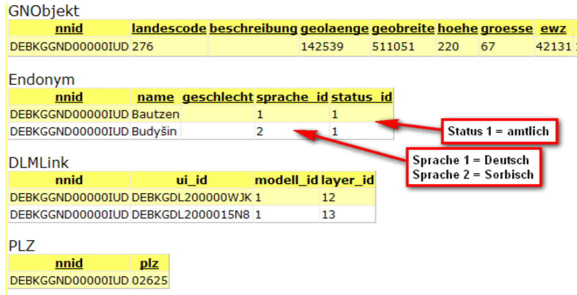
The following information related to finances, administration, legal issues and human resources only refers to the federal level. All tasks for the maintenance of the national geographical names database and its publication through web services and applications are conducted by the BKG, supported by the Permanent Committee on Geographical Names (StAGN).

In 1981 the first edition of the Gazetteer Federal Republic of Germany („West Germany ") was published by the Institute for Applied Geodesy (IfAG; its name was changed to BKG in 1996). It comprised some 21.500 geographical names contained on the 1:500 000 General Topographic Map series. The data was available both in printed and digital form. At that time, it was the first gazetteer in Germany which had been prepared in accordance with the United Nations resolutions.

After the German reunification in 1990 this dataset was supplemented by the geographical names of the five new Länder (federal states) Mecklenburg-West Pomerania, Brandenburg, Saxony-Anhalt, Thuringia, Saxony and the eastern part of Berlin. The database was extended with the records of some 25,000 additional names rendered on the 1:250 000 general topographic map series. Since then, the national geographical names database called "Geographical Names of Germany" (GN-DE) has been developed and maintained by BKG. Today, it includes all relevant object/feature types from the digital landscape model scale 1: 250,000 (DLM250), which have one or more geographical names (for reasons such as multilingualism).

The object/feature themes of the GN-DE geographical names database are: ‚settlement', ‚transportation', ‚vegetation', ‚hydrography', relief', areas and regions' with attributes regarding the administrative hierarchy (administrative code), river system (hydrographic key number in accordance with the German Working Group on water issues (Bund/Länder-Arbeitsgemeinschaft Wasser -LAWA), number of inhabitants, size (area, length), position (geometry data from DLM250 as well as additional attributive geographical coordinates), postcodes, status, language of the geographical name and height data (in meters).

The data model GN-DE was developed and has been evolved continuously by BKG following UNGEGN resolutions. Today, all geographical names entries of the GN-DE are linked unambiguously to all databases and products provided by BKG through unique identifiers UID).


GNObjekt

nnid	landescode	beschreibung	geolaenge	geobreite	hoehe	groesse	ewz
DEBKGGND00000IUD	276		142539	511051	220	67	42131

Endonym

nnid	name	geschlecht	sprache_id	status_id
DEBKGGND00000IUD	Bautzen		1	1
DEBKGGND00000IUD	Budyšin		2	1

DLMLink

nnid	ui_id	modell_id	layer_id
DEBKGGND00000IUD	DEBKGDL200000WJK	1	12
DEBKGGND00000IUD	DEBKGDL2000015N8	1	13

PLZ

nnid	plz
DEBKGGND00000IUD	02625

Figure 3-1 GN-DE database table showing that one GNObject may be associated to one or more geographical names (endonyms) in different languages. The example shows the GNObject associated with 'Bautzen' (in German language) and 'Budyšin' (in Sorbian language). Both names are treated equally in the GN-DE and have equal official status.

At present the GN-DE database contains about 165,000 entries including more than 700 names in the languages of the national Sorbian and Frisian linguistic minorities. About 27,000 entries indicating the gender of hydrographic features are contained as well. The administration and maintenance tasks of maintaining the GN-DE are conducted at BKG in Frankfurt am Main, whereas the web service and applications are facilitated through the 'Central Service and Distribution Centre for Geoinformation' at the BKG department in Leipzig [1].

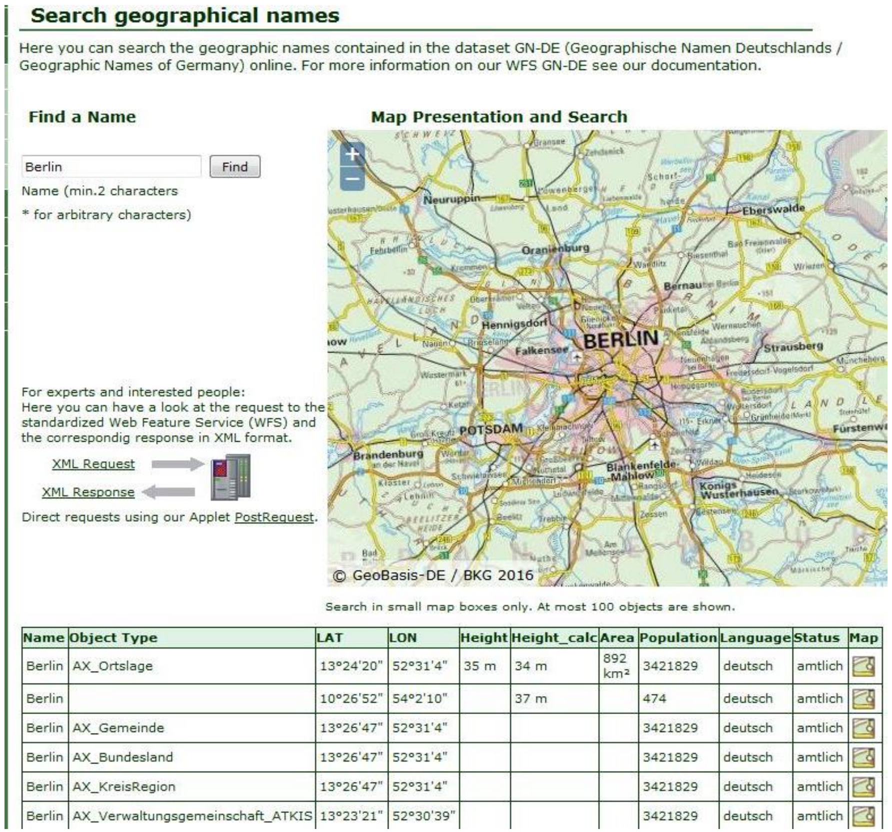


Figure 3-1 Screenshots from the German geographical names application based on a Gazetteer Web Feature Service (WFS)

BKG has been providing geographical names to the general public through the Internet since 2006. The gazetteer web service application 'Geographical names' is based on a Web Feature Service (WFS) called 'WFS GN-DE' and is thus compliant with the corresponding Open Geospatial Consortium (OGC) specification [3].

With the web service application, you can search the geographical names contained in the dataset GN-DE online.

The following financial and human resources can be roughly estimated for maintaining the national geographical names programme by BKG. The figures can vary significantly according to the time and resources needed for the adaptation of the database, the services and applications to new technical developments or to the services oriented architecture in the organization.

BKG personnel costs: about 90,000 Euros / year;

- Office of Permanent Committee on Geographical Names, 1 person, full time, about 40,000 Euros / year
- Coordination of names activity at BKG, 2 persons, 30 %, about 10,000 Euros / year
- Technical staff for administration of databases, 1 person, 50 %, about 20,000 Euros / year
- Technical staff for administration of web (gazetteer) service, 1 person, 50 %, about 20,000 Euros / year

BKG investments: about 20,000 Euros / year

- Database (hard- and software), 2 databases, updates per year for 5,000 Euros / year
- Server (hard- and software), 1 server update per year for 5,000 Euros / year
- IT-specific training, 2 persons, about 10,000 Euros / year
- Soft- and hardware amortization is different for server or database and varies between 3-5 years.

3.4 GN-DE within the German Spatial Data Infrastructure (GDI-DE)

The GN-DE dataset is currently used for the geographical names data provision within the '*German Spatial Data Infrastructure* (GDI-DE)'. GN-DE is the rationale for the INSPIRE data provision for European purposes as it is the only dataset in Germany that includes the mandatory and almost all 'voidable' INSPIRE attributes (like language, status, gender …). A wide range of different applications (e.g. geoportals) has been built up that provides discovery functionality for spatial data sets and services and generally make use of the functionality. In Germany, the Geoportal.de provides insight into GDI-DE. It utilizes several Web Services from GDI-DE. Users can promptly search within the central search engine of GDI-DE, which currently contains around 87,000 decentrally maintained sets of metadata on spatial data and spatial data services from across all levels of public administration in Germany. The Geoportal.de application is completed by a search utility for places and addresses. This search utility has been developed at the BKG Central Service and Distribution Centre for Geoinformation and is widely used in the portal application. It builds upon the '*Georeferenced Address*

Data Federation' data set and parts of the very comprehensive names dataset 'Geographical Names of Germany' (GN-DE) derived as dataset 'GN250', Geographical Names of Germany (1: 250,000) [4]. As mentioned before, GN-DE provides the unique source for all BKG products and naturally for the national (gazetteer) web feature service, too. GN-DE also provides geographical names in the officially recognized Sorbian and Frisian minority languages. The integration of those minority language names in the search utility

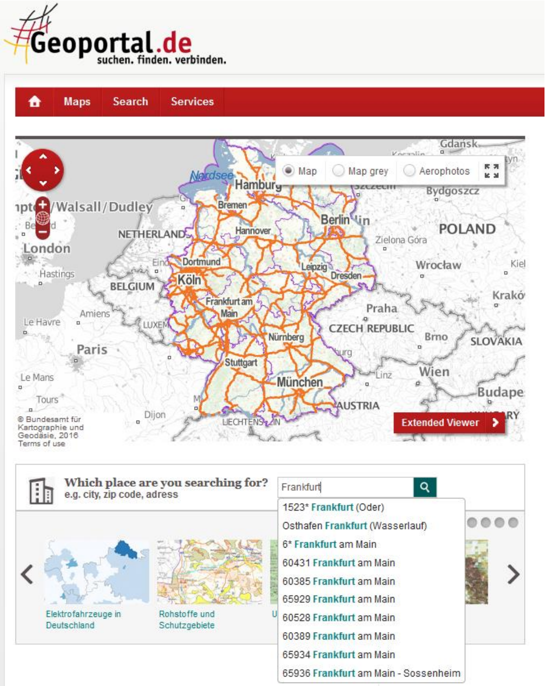


Figure 3-2 The German national geographical names database (GN-DE) published as a web service which is part of the national spatial data infrastructure (GDI-DE) and visualised through the Geoportal application (Geoportal.de)

for places and addresses of the Geoportal.de is envisaged, but not realized as yet (2016).

The 'Georeferenced Address Data Federation' dataset mainly consists of data from the 'Association for the Distribution of House Coordinates' run by the surveying authorities of the Länder. After processing, this dataset provides point coordinates for house addresses and encircling bounding boxes of streets, places and zip codes. The search utility implements two search strategies: a high performance live suggest search, i.e. to deliver real time suggestions for search terms and/or other relevant information based on live user input in an application, on the one hand, and a fault-tolerant search on the other. Words matching the search term are highlighted. While in the results of the live suggest search only the elements of the objects' name as part of the address are shown, the fault-tolerant search provides the complete dataset including geometry. If the match between search term(s) and results is weak, a list of similar places or addresses is offered. The search utility is able to handle unstructured input of addresses.

Figure 3-3 The search utility within the Geoportal.DE – live suggest during user input

The search utility has been implemented as Java-Servlet and offers an HTTP/Key-Value Pair interface. For data indexing the Apache Lucene library is used [5]. For geometry handling and implementation of OGC-standards the GeoTools library is used [6]. Overall, there is a total of 23 million indexed datasets with an aggregate data volume of approximately 3 GB. The index is built once before service start-up and is loaded into the central memory.

Meaningful combinations of map layers available within GDI-DE are placed prominently on the Geoportal.de Welcome page as featured maps. Especially by overlaying spatial information layers from different institutions the value added through the use of interoperable spatial data services is demonstrated. The selection of thematic maps covering Germany will be extended continually.

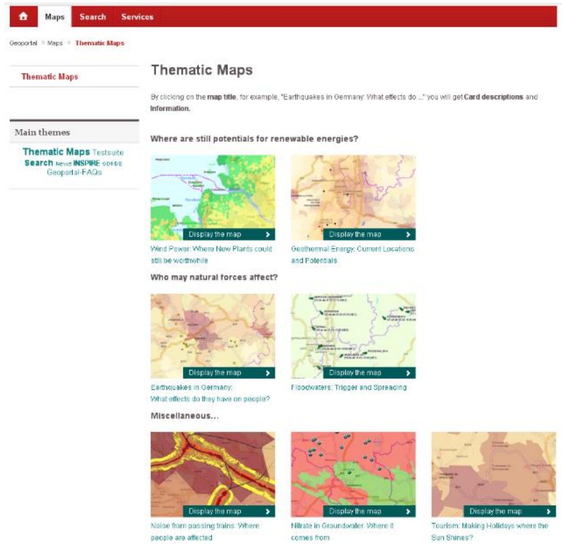


Figure 3-4 Geoportal.DE as a showcase for Spatial Data Infrastructure (SDI) through featured maps

In due course, a map showing the physio-geographic regions of Germany [5] combined with touristic information will be integrated. In this way, one important product created by the German Permanent Committee on Geographical Names (StAGN) will be published and disseminated in the broader GDI-DE context. Considering multilingualism and the cultural heritage in Europe, people often use different spellings and languages when talking about one and the same place. Even within a country more than one (official) spelling is used. These INSPIRE-supported principles also underline the United Nations resolutions on the standardization of geographical names which are reflected in Germany in the work of the Permanent Committee on Geographical Names (StAGN).

3.5 GN-DE brought into conformity to INSPIRE

In 2015, the geographical names database GN-DE and the Implementing Rule for ‘Interoperability of Spatial Data Sets and Services’ (this is, data specification of geographical names, Annex I) of the European geospatial data infrastructure (keyword: INSPIRE) have been brought into conformity [8]. Thus, BKG has been one of the first institutions in Europe providing an INSPIRE-compliant national geographical names data service. According to the INSPIRE schedule EU countries are only obliged to realize this by the year 2017.

The INSPIRE-conform geographical names dataset of BKG is already available via the European data portal (http://www.europeandataportal.eu/de/). The European data portal collects metadata from information of the public sector which is available in European countries. This also includes information regarding the provision of data and the advantages of their further use.

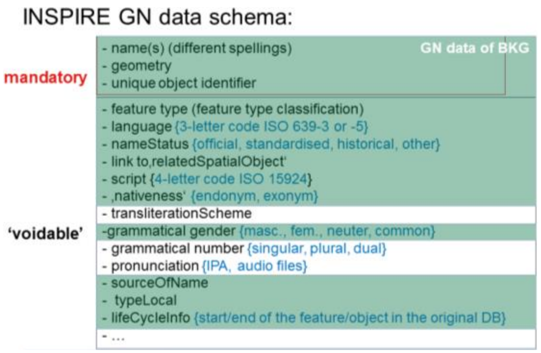


Figure 3-5 INSPIRE Data schema fulfilled by BKG

3.6 GN-DE published as Open Data

Every day, public authorities all over the world gather a multitude of data and information such as spatial data on transport networks, addresses and cadastral information or economic, environmental and statistical data. In the recent past, more and more government authorities admit to the principle of "Open Government" and “Open Data” policy.

The Geodata Access Act (Geodatenzugangsgesetz, GeoZG) passed in Germany by the federal government on 30 July 2008 promotes the further development of the German Spatial Data Infrastructure (GDI-DE). Germany declared the federal government's geodata to be open data in the course of the revision to the Geodata Access Act (GeoZGÄndG) on 16 November 2012. The Ordinance to Determine the Conditions for Use for the Provision of Spatial Data of the Federation (Verordnung zur Festlegung der Nutzungsbestimmungen für die Bereitstellung von Geodaten des Bundes) (GeoNutzV), which allows commercial and non-commercial use of the data free-of-charge, was passed on 19 March 2013. In particular, the GeoNutzV ordinance lays down provisions relating to the provision of spatial data, and to spatial data and metadata services on behalf of spatial data holding agencies [9].

In Germany, the open data declaration refers to geospatial data produced by the Federal Agency for Cartography and Geodesy (BKG). Consequently, view and download services based on international open standards and open source software of digital landscape model (DLM) and digital terrain model (DGM) as well as digital topographical maps (DTK), general maps and maps of municipality areas based on small to medium scale are provided free of charge. Hence, the gazetteer web service of BKG is available free of charge, too [10].

3.7 Further discussion on logistics, production time needed and necessary upkeep costs

The surveying and cadastral administrations of the Länder, the statistical offices and the Federal Statistical Office, private and academic institutions for language and culture etc. are all potentially data-holding authorities for INSPIRE (in accordance with the German law ‘Geodatenzugangsgesetz – GeoZG’) for geographical names data, though so far with highly different names density.

The financial and human resources roughly estimated by BKG as necessary for maintaining the national geographical names programme amounted to about € 110,000 per year for both personnel and investments. The figures vary according to the time and resources needed for the adaptation of the database, the services and applications to new technical developments or to the services oriented architecture in the organization. Considering that most of the German Länder might invest a similar small amount in logistics, the maintenance of geographical names database and the development and up keeping of services and applications, about the overall yearly costs would amount to € 1,8 million.

Considering that the data content for GN-DE is based on a map scale 1: 250,000 a cooperation of the Federal Government with the Länder– (the latter are responsible for data based on maps at scales 1: 25,000 – 1: 10,000) is envisaged in order to provide a unique source for geographical names data in Germany and increase the content of the national web service and integrate names data taken from large scale maps.

The Länder do not and will not add any toponymic attributes like status or language to the geographical names in their databases, as this has never been part of their tasks. The geographical names data content is based on the ATKIS programme and consists of the geographical names as an associated attribute to the object/feature and by that inherits only its feature type and its geometry. Thus, joining the names datasets of BKG and of the Länder and so providing one unique geographical names data source for Germany would be beneficial to everyone – including data users. Furthermore, the INSPIRE data provision could be facilitated by BKG for the federal and the Länder administrations with one national web service.

3.8 References

[1] Central Service and Distribution Centre for Geoinformation at BKG, www.geodatenzentrum.de, last accessed 09/2016

[2] Working Committee of the Surveying Authorities of the Länder of the Federal Republic of Germany http://www.adv-online.de, last accessed 09/2016

[3] Geographische Namen Deutschland – GN-DE, the (gazetteer) WFS service (wfs_gnde), http://www.geodatenzentrum.de/xml/WFS_gnde.xml, last accessed 09/2016

[4] Geographische Namen 1: 250,000 (GN250), documentation of the dataset (in German only), http://www.geodatenzentrum.de/docpdf/gn250.pdf, last accessed 09/2016

[5] Apache Lucene: http://lucene.apache.org, last accessed 09/2016

[6] http://geotools.org/ , last accessed 09/2016

[7] Bundesrepublik Deutschland 1: 1,000,000, Landschaften - Namen und Abgrenzungen, 6st edition (Federal Republic of Germany at 1: 1,000,000 scale, Geographical regions - names and boundaries, 6st edition. Federal Agency for Cartography and Geodesy (BKG)

[8] INSPIRE Implementing Directive (COMMISSION REGULATION - EU - No 1089/2010 of 23 November 2010 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial data sets and services), http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:323:0011:0102:DE:PDF, last accessed 09/2016

[9] Ordinance to Determine the Conditions for Use for the Provision of Spatial Data of the Federation (Verordnung zur Festlegung der Nutzungsbestimmungen für die Bereitstellung von Geodaten des Bundes – GeoNutzV): http://www.geodatenzentrum.de/docpdf/geonutzv_eng.pdf, last accessed 09/2016

[10] Open Data - Freie Daten und Dienste des BKG: http://www.geodatenzentrum.de/geodaten/gdz_rahmen.gdz_div?gdz_spr=eng&gdz_akt_zeile=5&gdz_anz_zeile=0&gdz_user_id=0, last accessed 09/2016

Section 2 Management of a national names programme

Chapter 4 **Compilation and updating of the swisstopo geographic names database “swissNAMES^{3D}”**

Stefan Neudeck and Alfred Gut

4.1 Glossary

swissNAMES^{3D} is the new dataset of geographic names in Switzerland and the Principality of Liechtenstein. It is suitable for a broad variety of uses and visualisations, including updating with names. It replaces the “SwissNames” database that was updated until 2008 and was based on the names used in the set of national maps.

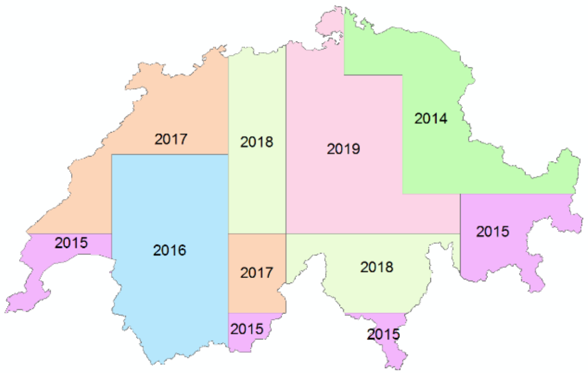


Figure 4-1: Updating cycle for Switzerland’s topographic landscape model (the indicated years refer to the status of the fundamentals)

Figure 4-2 Names displayed as an orientation aid on an aerial image of Interlaken

4.2 Compilation, updating and fundamentals

The data contained in the “swissNAMES^{3D}” database (see figure 4-1 for an example) was compiled within the scope of the production of the topographic landscape model with the aid of the “TopGIS” topographic-geographic information system. The database is to be updated every six years (cf. figure 4-2).

The content of “SwissNames” and the objects resulting from the transfer of the topographic landscape model (TLM) that replaced the former VECTOR25 landscape model into the TLM production database formed the basis for the compilation of the initial dataset. These data formed the original basis for the placement of names on the 1:25,000 national map from which they were taken. “SwissNames” was a point database. All objects that were now modelled as lines or polygons were initially defined as “standard geometry”. The resulting basis facilitates the localisation of the objects to be recorded in the aerial image, which is subsequently used for carrying out the geometric measurement. The high degree of accuracy required for the topographic landscape model calls for the exact geometric referencing of all identified objects and the applicable perimeters for all names, as well as the photogrammetric measurement of their height or allocation of their height based on the digital terrain model that also forms an integral part of the production environment. With respect to objects that cannot be identified in the aerial image, and for the verification of orthography, other principles apply as described below.

Place names in the cadastral survey: The place names defined as polygons in the cadastral survey data model specify the applicable perimeters for place names in the large-scale range. The applicable perimeter refers to the area recorded in the cadastral survey that can be used for an indicated place name. These new geometries and names are to be transferred to the new database. Around 130,000 objects included in the new database are place name polygons. Currently, only around 35 percent of the production perimeters are filled with name data from the cadastral survey. The orthography of the place names in the cadastral survey is also being transferred to other object categories of swissNAMES^{3D} (names of regions and terrains, point objects).

Official index of towns and cities, with postcodes and perimeters from the cadastral survey: Complete existing cadastral survey dataset. Lists towns and cities with their own postcode as polygons with attributed names. This index serves as the reference work for the spelling of the names of towns and cities that have their own postcode. In densely developed areas, the geometries of towns and cities with postcodes are used for the internal delimitation of neighbouring development polygons.

geoSTAT: This is the distribution model for official federal geodata (produced by the Federal Statistical Office). Alongside datasets relating to land use and ground cover, together with data from population and business censuses, municipal boundaries and terrain data, the boundaries of suburbs, districts and residential zones in Switzerland's 17 largest towns and cities are depicted in the GEOSTAT model. The interior boundaries of towns and cities, together with the corresponding names, are used for the definition of development perimeters in swissNAMES^{3D}.

Public transport stations and stops: The complete dataset of the Federal Office for Transport is to be transferred to swissNAMES^{3D}.

Addresses from the register of buildings and residential dwellings: These are used for determining the perimeters and names of developed areas.

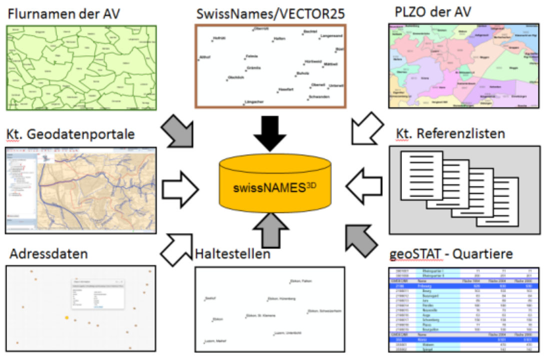


Figure 4-3 Bases for the compilation of swissNAMES^{3D}

Cantonal geodata portals and reference lists: These provide the names of objects to be included in the database (sites, bodies of water, buildings, individual objects, etc. – cf. figure 4-3).

4.3 Example of input into the database

The example below describes the input of areas relating to object type "Place" (object category "TLM_Placename"). The basis for entering the polygon geometry is the corresponding orthophoto. The perimeter encompasses buildings that are occupied permanently or seasonally, as well as infrastructure (e.g. roads and paths). In the case of larger towns and cities, the perimeter also includes objects such as bodies of water, parks and other unoccupied zones up to a certain size. The external boundaries of development polygons are defined on the boundaries of occupied plots of land or may be defined by neighbouring objects that are not part of the development (e.g. bodies of water, forest – cf. figure 4-4).

Figure 4-4 The place names "Grueben", "Muheimere" and "Toffenholz" were transferred to the production database from the previous "SwissNames" database (points) as standard polygons and...

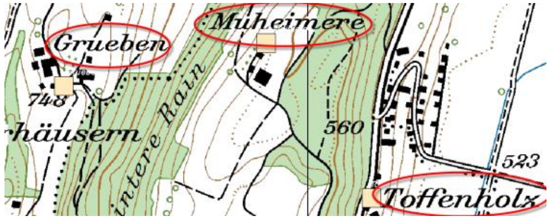


Figure 4-5 ... they correspond to the place names in the 1:25,000 national map.

Figure 4-6 The perimeters are entered on the basis of the aerial image depicting the developed area. The z values are automatically transferred during digitisation from the digital terrain model to the base points of the development contours.

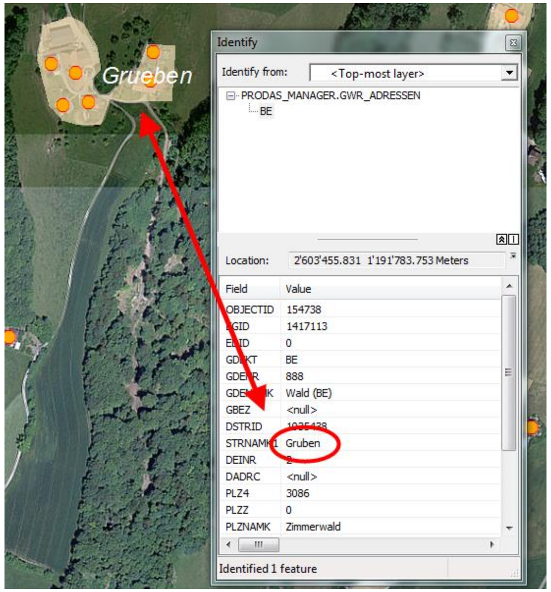


Figure 4-7 The orthography of the place names then has to be verified. This is done by consulting an addresses dataset. The corresponding datasets are the official index of towns and cities (for places with their own postcode) and the official register of buildings and residential dwellings (for smaller places without their own postcode). This means that "Grueben" and "Muheimere" have to be changed to "Gruben" and "Muheimern".

Names of places with an own postcode are binding. However, smaller places do not have an own postcode and therefore do not have a place name with postcode (according to the abbreviated name of the town with a postcode, "Gruben" belongs to "Zimmerwald" [attribute, "PLZNAMEK"]).

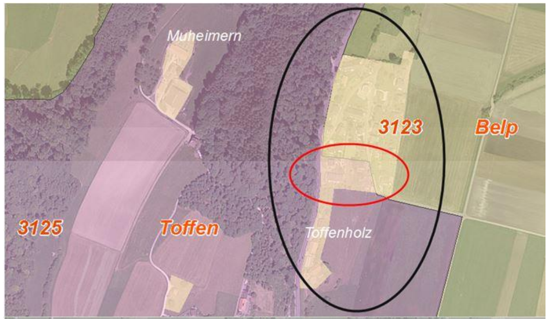


Figure 4-8 Nonetheless, the postcodes dataset still has to be consulted. According to the input guidelines, the boundaries of the town with an own postcode have to be used for the interior delimitation of the development polygons. This now also applies to small places. In the example presented here, "Toffenholz" has to be divided because the northern zone belongs to "Belp" (which has its own postcode) and the southern zone belongs to "Toffen" (which also has its own postcode).

The result is shown in figure 4-9.

Figure 4-9 Example of input

4.4 Geometric characteristics and special features

swissNAMES^{3D} is based on the topographic landscape model. This is Switzerland's new official geodata set and it corresponds approximately to a degree of abstraction of 1:10,000. The data are three-dimensional. The recording of heights is carried out through the automatic adoption of height data during the digitisation of the location base points. Certain object heights are measured photogrammetically. The targeted degree of geometric accuracy is one metre (location and height).
With very few exceptions, all objects in the topographic landscape model that have a specified name are to be transferred to swissNAMES^{3D}. The name is the decisive criterion for the input of an object. While in developed areas there are no neighbouring objects with the same name and same object category, this is not the case when it comes to the network of bodies of water. Neighbouring objects with the same name are merged to form a single object. Figure 4-10 shows a comparison between the development of the topographic landscape model and the swissNAMES^{3D} database. Each colour represents an object with a single name and is thus taken from the topographic landscape model for swissNAMES^{3D}. The topology is not transferred to swissNAMES^{3D}, which means, for example, that for estuaries of the "Fildrich" and "Goldbach" streams into the "Kirel", the latter is not divided. So the green line depicts a single object with the name "Kirel" (and no longer several objects, as would have been expected in a topological dataset). And of course, the same applies with respect to the "Simme" and the "Alpbach".

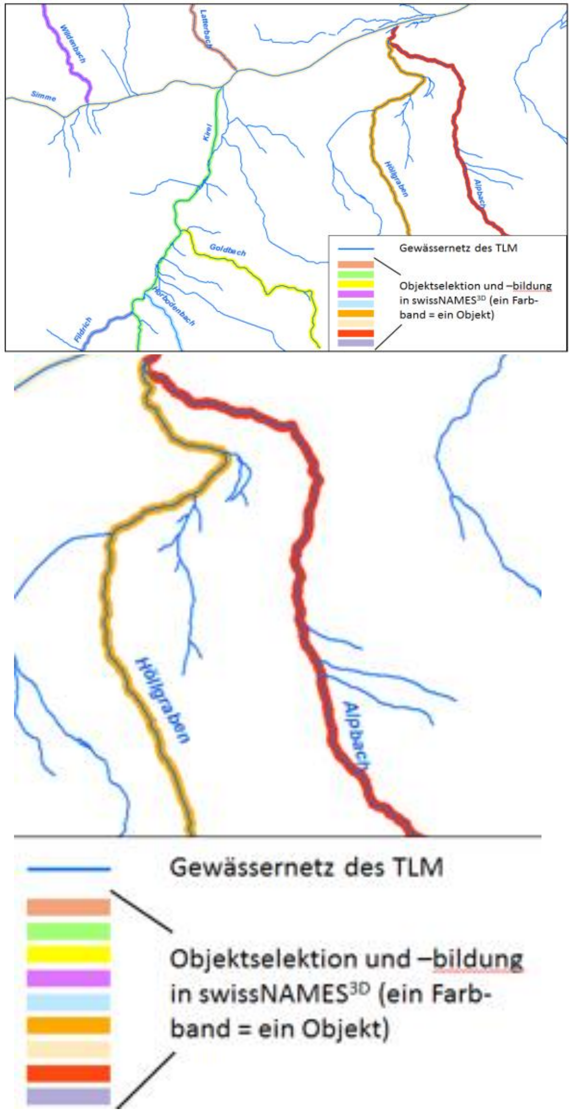


Figure 4-10 Comparison between the development of swissNAMES^{3D} and the topographic landscape model

4.5 Multilingual aspect

Multilingualism is one of Switzerland's main identification characteristics: In the Federal Constitution, German, French and Italian are defined as official languages, together with Romansh. According to a 2013 survey conducted by the Federal Statistical Office, German is the main language for 63.5 percent of the population, French for 22.5 percent, Italian for 8.1 percent and Romansh for 0.5 percent. There are three bilingual cantons (Valais, Bern and Fribourg: French & German), and one trilingual canton (Grisons: German, Italian and Romansh). See figure 4-11.

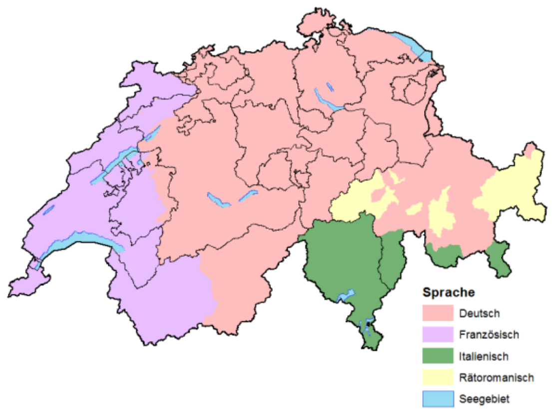


Figure 4-11 Switzerland's language regions

In the current set of national maps, some objects (primarily housing developments, but also the names of some regions) located in the vicinity of the borders of language regions have more than one name (see figure 4-12).

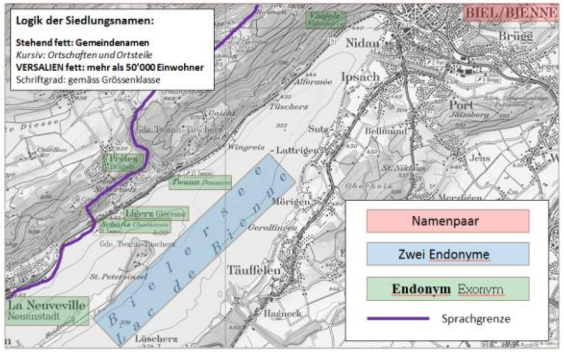


Figure 4-12 Extract from the 1:100,000 national map showing names in more than one language

The following terms that are used in connection with multilingual nomenclature are defined below.

<u>Endonym:</u> The name used in a region within which the designated object is located

<u>Exonym:</u> The name used in another region than that within which the designated object is located

In the context of Switzerland's geographic nomenclature, the term "region" refers to "language region".

<u>Pairs of names</u> are also used exclusively for municipalities and/or their districts, zones, etc., that are officially declared as bilingual. The names in the respective languages are separated by a forward slash (e.g. Biel / Bienne).

swissNAMES^{3D} supports the use of multiple languages by attributing each name with a "language code" (GER, FRA, ITA and ROH, see figure 4-13) based on ISO 639-2. The language code incorporates both the respective standard language and its dialects, i.e. a name in a dialect is attributed the same language code as a name written in the respective standard language.

In swissNAMES^{3D}, the processing of objects for which more than one name exists in the official languages is extended to include major objects of national importance. In other words, it is not only objects in the vicinity of language region borders that may be given more than one name. This may concern several names in the same language as well as names in different official languages. Similarly, in the case of composite names, endonyms and exonyms are explicitly indicated. The data are listed in a separate table as name components. Under "NAME_TECHNICAL", all names of a given object are listed together with their respective language codes and the indication whether they are endonyms or exonyms (cf. figure 4-13).

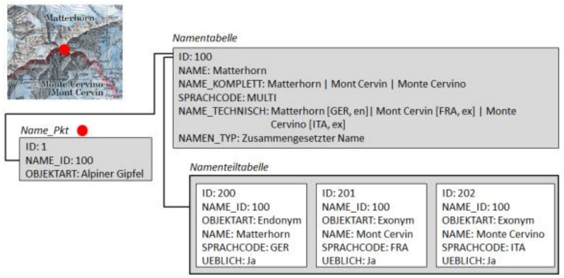


Figure 4-13 Depiction of objects with several names in the production database

4.6 Data model / content and scope of data

The structure of the data model is very simple. For each geometric type (point, line, area) there is one object category which comprises all the named objects in the same geometric category. The names take the form of attributes (Shapefile or CSV format), or are listed in a separate table referenced to the geometric tables (geodatabase). The content of swissNAMES^{3D} is presented in table 4-1.

4.7 Downloading the data

swissNAMES^{3D} will be published annually with effect from 2015 and may be downloaded free of charge from the swisstopo online shop. It may only be used for the purposes that are specified in the licence for obtaining geodata free of charge from swisstopo. With each order, the entire dataset will be delivered in three formats (ESRI Filegeodatabase / ESRI Shapefile, CSV text file) and two coordinate systems (CH1903 and CH1903+).

4.8 Options for use

swissNAMES^{3D} can be used for a broad variety of purposes, especially in combination with other geodata. Because the dataset contains data in the large-scale range as well as for overview purposes, swissNAMES^{3D} is suitable for use at various degrees of abstraction. Examples for the use of the database include:

- As a nomenclature database for Internet solutions (geographic searches)
- As an orientation aid on aerial images and other geodata sets
- For the compilation of catalogues of geographic names (gazetteers)

TLM_NAME_LIN 10'000 Objekte	• Namen wichtiger Verkehrsbauwerke (Brücken, Tunnel) • Seilbahnen • Skilifte • Namen von Sportanlagen (Rodelbahn, Bobbahn, Sprungschanze) • Gewässernamen
TLM_NAME_PKT 225'000 Objekte	• Gipfel • Pässe • Points of Interest (Bildstock, Brunnen, Denkmal, Aussichtspunkt u. a.) • Einzelobjekte • Flur- und Lokalnamen • Haltestellen des öffentlichen Verkehrs • Gebäudenamen • Namen von Infrastrukturanlagen (Ein- und Ausfahrten, Zollstellen)
TLM_NAME_PLY 76'000 Objekte	• Siedlungsnamen (Orte und Quartiere) • Gewässernamen (Seen und Seeteile) • Naturräumliche Gliederung (Grossregionen, Haupttäler, Landschaften) • Geländenamen (Täler, Gräben, Gletscher, Grate, Massive) • Nutzungsareale (Abwasserreinigung, Deponien, Friedhöfe, Klöster, Kraftwerke, Schulen, Spitäler, Messeareale, Pärke, Zoos) • Verkehrsareale (Flugplätze, Heliports, Parkplätze, Rastplätze) • Freizeitareale (Camping- und Golfplätze, Sportanlagen) • Infrastrukturanlagen (Staumauern und –dämme, Wehre)

Table 4-1: swissNAMES^{3D} contains more than 300,000 objects, of which 180,000 have place names or local names.

- As a basis product for publications (e.g. maps, plans), including for all names in the set of national maps
- As an orientation aid in 3D visualisations

4.9 swissNAMES3D in the federal geodata infrastructure

Search tool

swissNAMES^{3D} is used within the federal geodata infrastructure. Together with addresses, the new geographic nomenclature database functions as the main source of data for the SwissSearch geo-coding service, which can be used for carrying out searches based on the geographic names in the federal geoportal (see figure 414; http://map.geo.admin.ch). The “Geographic Names” section of the federal geoportal is also available for use. With this function, users can display names as an orientation aid, for example on an aerial image (see figure 4-1).

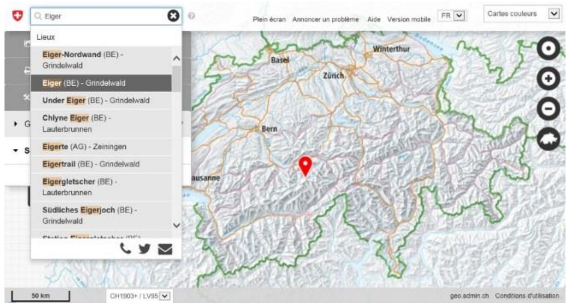


Figure 4-14 swissNAMES^{3D} as the basis for a search tool in http://map.geo.admin.ch

Raster-based data layer

swissNAMES^{3D} is integrated into the federal geodata infrastructure as a data layer. The dataset can be displayed as an orientation aid for other data layers, especially aerial images. The federal geoportal enables the visualisation of data layers at 14 zoom levels or scales ranging from 1:500 to approximately 1:2,000,000. Visualisation of texts (annotations) is raster-based. For each zoom level, the swissNAMES^{3D} annotations to be displayed were initially defined in ArcMap and graphically structured, then exported as a raster image. For the larger scales / higher zoom levels, the raster image has to be tiled due to the higher data density.

swissNAMES^{3D} as a data layer in the federal geodata infrastructure - Data layer visualisation of object geometry

For each zoom level, precisely those object geometries are visualised that are foreseen for annotation. A data layer is required for each geometry type, in which the content and annotations vary according to the zoom level. Basic icons are used for graphic presentation.

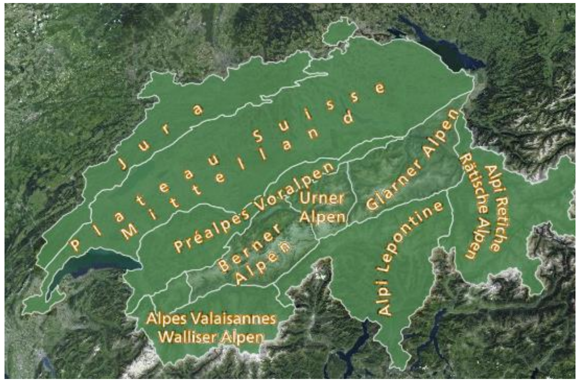


Figure 4-15 Presentation of swissNAMES at the lowest federal geodata infrastructure zoom level

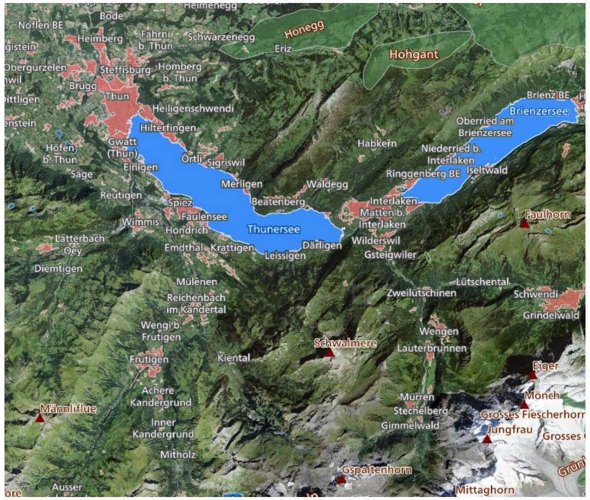


Figure 4-16 Presentation of swissNAMES at the scale of approx. 1:200,000

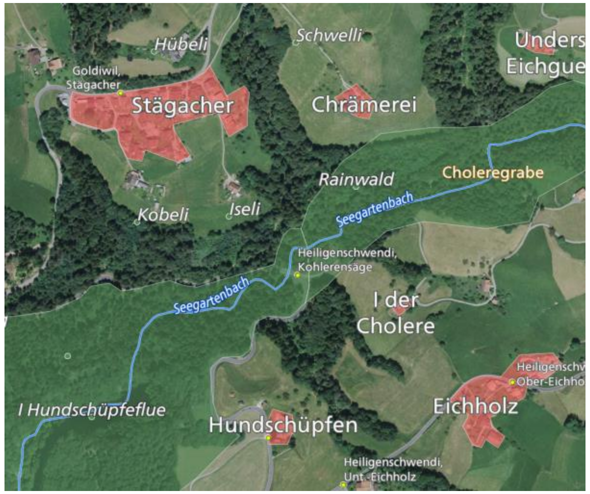


Figure 4-17 Presentation of swissNAMES in the large scale range

Data layer feature information

Tool tips provide object data (feature information) when a swissNAMES^{3D} object is activated in the raster image. The selected object is highlighted and the most important attributes stored in the vector layers are displayed.

Figure 4-18 swissNAMES^{3D} object information in the federal geodata infrastructure

4.10 Development of the geographic names database: required time and personnel

The names database was developed over a period of six years (2009 to 2014, cf. “Development, updating and fundamentals”). An annual total of around 6,500 hours was required, which corresponds to approximately 4 full-time equivalent jobs.

4.11 Links to other sources

http://map.geo.admin.ch: federal geodata portal map viewer

http://www.swisstopo.admin.ch/internet/swisstopo/en/home/products/landscape/swissTLM3D.html - swissTLM^{3D} - the topographic landscape model

http://www.swisstopo.admin.ch/internet/swisstopo/de/home/products/landscape/swissNAMES3D.html: swissNAMES^{3D}

http://www.bfs.admin.ch/bfs/portal/de/index/dienstleistungen/geostat.html: GEOSTAT

http://www.bfs.admin.ch/bfs/portal/de/index/themen/00/05/blank/01.html: Federal register of buildings and dwellings

http://www.cadastre.ch/internet/kataster/de/home/services/service/plz.html: Official index of towns and cities

http://www.cadastre.ch/internet/kataster/de/home/av/names.html: Geographic names

http://www.geo.admin.ch/internet/geoportal/de/home/services/geoservices/display_services.html: Description of the presentation services in the federal geodata infrastructure

Section 3 Regional activities and different applications

Chapter 5 **Names in nautical charts**

Sylvia Spohn

5.1 Introduction

This chapter is directed to cartographers that have to include names of maritime features (seas, gulfs, straits, inlets, shoals, banks, etc) into nautical or similar charts. It is based on the Regulations of the International Hydrographic Organisation (IHO) for its International (INT) Charts and Chart Specifications, edition 4.7.0 July 2017 (https://www.iho.int/iho_pubs/standard/S-4/S-4%20Ed%204.7.0%20July%202017%20EN.pdf) which are used in the German and a lot of other Hydrographic Offices. Before 1982 most national nautical administrations followed their own national rules for the production of nautical charts.

History

The IHO was formed as the result of a desire for standardization of nautical charts and associated publications and consequently for greater safety of mariners. It was felt that this standardization could be achieved in such a way that language differences would be minimized and that a chart produced by one country would be perfectly comprehensible to a navigator from another country.

Although measures have been taken since the formation of the International Hydrographic Bureau (IHB) in 1921 to develop standards to follow when producing charts and publications, it was not until 1967 that the concept of an INTernational chart was proposed. It was felt that, instead of several different Hydrographic Offices each producing different charts of the same ocean area, often with different data, scales and limits, it would be more economic and safer if one Hydrographic Office ('producer nation') would compile and produce an original chart to internationally agreed specifications, and that other Hydrographic Offices ('producer nations') would be able to print the chart, using the basic reproductive material provided by the original producer nation but substituting their own language, if they so wished. After the standardization of the format and symbols in 1967 the International Hydrographic Conference established a Commission which developed the "Chart Specifications of the IHO" which were adopted in 1982. This commission now works as the Nautical Cartography Working Group under the IHO Hydrographic Services and Standards Committee (HSSC) and is still developing the standards further.

5.2 Place names on charts

The consistent use of standardized place names is an essential element of effective communication worldwide and is therefore a requirement for nautical charts.

Terms and definitions

The following terms and definitions are consistent with the UNGEGN 'Glossary of Terms' (2002 edition as updated by addendum 2007 (https://unstats.un.org/unsd/geoinfo/UNGEGN/docs/glossary.pdf) or with the Hydrographic Dictionary (S-32) (http://hd.iho.int/en/index.php/Main_Page):

Toponymy: The place names or the study of place names of a country or district (S-32) or;
The study of toponyms in general and of geographical names in particular, and;
The totality of toponyms in a given region (UNGEGN).
Official language: A language that has legal status in a particular legally constituted political entity such as a State or part of a State, and that serves as a language of administration, for example Spanish in Chile.

National language: A language in widespread and current use throughout a specific country or in parts of its territory, and often representative of the identity of its speakers. It may or may not have the status of an official language.

Minority (Regional) language: In a specific region, a language that is different from the official language of State administration and that is spoken by a national minority. It may or may not have official status. Examples: Welsh in UK; Breton in France; Swedish in Finland (Regional languages may give rise to original forms of toponyms in the areas where they are spoken).

5.3 General language rules

English is the working language for navigational purposes and for communication at sea, according to the International Convention for the Safety of Life at Sea (SOLAS, Chapter V Regulation 14). The International Hydrographic Conference decided in 1997 that:
"each hydrographic office which does not issue charts in the English language should give all legends on charts affecting its territorial waters in its national language and in English."
Consequently, English language versions of all notes should be included on all non-English charts.
Navigationally significant legends should also be given in English, or if more convenient, listed in a glossary (key)

on the chart, see figure 5-1. English language versions of notes and legends are particularly appropriate on charts used for international shipping.

DÄNISCH / *DANISH*		
Bg.	Bjerg	Berg
Ga.	Gamle	Alt
Gd., gd.	G(g)ård	Hof, Gehöft
Gr., gr.	Grund, grund	Untiefe
hm(e).	holm(e)	Insel(n)
Hn.	Havn	Hafen
L.	Lille	Klein
N-	Nord(re), Nørre	Nord, nördlich
Ø- (E)	Øst(er), Østlig	Ost, östlich
S-	Syd, Sønder	Süd, südlich
sk.	skær	Schäre, Klippe
St.	Stor	Groß
V- (W)	Vest(er), Vestlig	West, westlich
Vg.	Vig	Bucht, Wiek

Figure 5-1 Example of a glossary from German chart symbology Karte1/INT 1 (2015. Such glossaries are often found in charts if no national symbology booklet has been issued. In this example, Danish generics are translated into German.

States having two or more official languages may adopt double or multiple legends (including toponyms) on charts of their own territory.

Merkkien selitykset ja kansalliset erityismerkit löytyvät julkaisusta KARTTA 1.	Teckenförklaringar och nationella symboler redovisas i den finländska publikationen KORT 1.	For explanations of international symbols and national symbols see finnish CHART 1.
KULKUSYVYYDET - Väyläviivaan merkityn suurimman kulkusyvyyden voi ylittää väylällä, jolla tälle kartalle on merkitty väyläalue ja sille haraussyvyys, huomioiden Liikenneviraston ohje 4955/1021/2011.	LEDDJUPGÅENDE - Det största leddjupgåendet markerat på farledslinjen kan överskridas i en farled, vars farledsområde och ramade djup angivits i detta sjökort, beaktande Trafikverkets anvisning 4955/1021/2011.	AUTHORISED DRAUGHT - The charted maximum draught may be exceeded in a channel, for which the fairway area and its swept depth have been indicated on this chart, observing FTA Guideline 4955/1021/2011.

Figure 5-2 Legend example from Finnish chart 57 (2011) showing a note in the two official languages Finnish and Swedish. English is added for international shipping.

Language on adopted charts. Printer nations may translate any elements of adopted charts into their own national language, in whole or in part, or add to those elements such translations, explanations, annotations, etc. as they may deem appropriate. However, toponyms and international abbreviations should not be translated. Printer nations may add alternative name variants to those incorporated by the producer nation. Printer nations may add partial glossaries to adopted charts in order to permit their user to understand the generic terms appearing in legends and composite names.

5.4 Writing Systems

If the national language in which a toponym is normally expressed does not use the same alphabet or writing system as that of the producer nation, the toponym may be rendered into the printing nation's language by means of a transliteration or transcription system. Various different transliteration systems exist, including UNGEGN, ISO 9:1995 and regional systems such as Russia's ГОСТ (transliterated GOST) and BGN/PCGN (US Board on Geographic Names/UK Permanent Committee on Geographical Names). Hydrographic offices may be subject to national government requirements regarding the system to be used for charts.

The rules presented here are intended to standardize and simplify the Roman alphabet used on international charts as much as possible, for the benefit of printer nations and users. In the case of nations which do not use the Roman alphabet, alternative forms of toponyms (transliterated using the Roman alphabet) and legends (translated into English) should be shown, in addition to showing them in the national writing system.

The Roman alphabet used on international charts comprises the 26 letters of the basic Roman alphabet, augmented if necessary by the special letters, or the letters modified by diacritical marks, particular to certain languages. Accents and diacritical marks must normally be retained. A letter of the basic Roman alphabet with diacritical mark should always be used in preference to a special (non-Roman) letter. The standard alphabetical order, for example as used in glossaries should be derived from the usual order of the basic 26-letter Roman alphabet. The order of words in an alphabetical list should not be affected by the presence of accents or diacritical marks. Special (non-Roman) letters may be placed after the basic alphabet, or incorporated into appropriate places in the basic order, according to usual national practice.

5.5 Numbers

All numbers should normally be in Arabic numerals, except:
• those forming an integral part of composite names, which must be spelled in full, for example:
Les Sept Îles, *Three Kings*.
• Roman numerals which are part of composite names in the convention of the national language, for example:
King George V Land.
• Roman numerals may be used for the numbering of paragraphs, sub-paragraphs, and columns (for example in the title block and related notes, in tables) and outside the chart border. When a number consists of only four digits it may be shown without using a space to isolate a single digit, for example either 1 500 or 1500 are acceptable. A year date must be written without a space, for example, 2014.

5.6 Abbreviations

For those terms for which an international abbreviation exists, full words and differing national abbreviations should not normally be used, except in toponyms. However, full words may be used in the chart title, tables and notes. For example, the chart title „Norderpiep und Süderpiep "is prefered to „Norder- und Süderpiep ".

D		
DG	Degaussing	N25, Q54
DGPS	Differential Global Positioning System	S51
Dia	Diaphone	R 11
Dir	Direction light	P 30, 31
dm	Decimetre(s)	B 42
Dn, Dns	Dolphin(s)	F 20
DW	Deep Water route	M 27, N 12.4
dwt	Dead Weight Tonnage	
DZ	Danger Zone	Q 50

Figure 5-3 Chapter D of the INT List of abbreviations with reference to INT 1 symbols at German chart symbology Karte 1/INT 1(2015)

Abbreviations should not be used within composite names except where necessary, for example because of lack of space. Where such abbreviations are necessary, national abbreviations may be used irrespective of the existence of a corresponding international abbreviation.

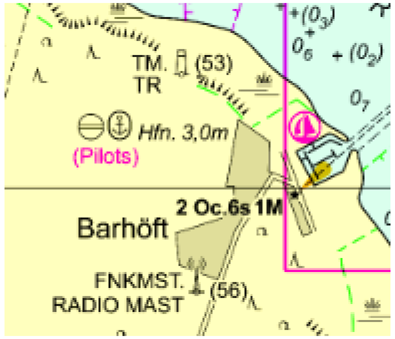


Figure 5-4 - Example for German and English abbreviations in German chart 1622 (2012): TM for Turm, TR for Tower, Hfn. for Hafen (Harbour) FNKMST. for Funkmast (Radio Mast). 2Oc.6s1M is the light description which means: 2 lights in line occulting all 6 seconds with the nominal range (maximum distance a light can be seen in clear weather) of 1 nautical mile

Abbreviations and the use of full stops. International abbreviations should not be terminated by full stops, except where forming the generic part of a toponym or for separating seabed qualities and light descriptions. Abbreviations of national generic terms in toponyms should be terminated by full stops in order to indicate clearly to chart users, that the abbreviation is not a full word. Full stops should also be used for other national abbreviations (see figure 5-3).

5.7 Toponym selection rules

A toponym should serve an identifying or a reference function for the chart user. It may also be required to identify features mentioned in other nautical publications. Chart producers must ensure agreement on toponyms across charts of different scales and between charts and other nautical publications. To ensure consistency of toponyms appearing on official charts, the following general rules apply:
a) On charts of its own territory, a national hydrographic office must show land and sea names as prescribed by the most authoritative source, in its own official writing system.
b) Where names are in Roman script, other hydrographic offices should show names (including generic parts of names) in agreement with the charts and publications of the national hydrographic office of the country having sovereignty, including all diacritical marks.
c) Where the names shown on the charts of the national hydrographic office of the country having sovereignty are not in Roman script, other hydrographic offices should transliterate or transcribe the names according to the rules of their national government.
d) As an exception to the above general rules, a hydrographic office may apply its conventional national usage to names of:
• countries
• major territorial divisions and boundary features (for example: mountain ranges, deserts extending through more than one sovereign state)
• oceans
• water areas which are not confined within one nation's territorial waters.

If a nation is charting an area in which the original toponyms are not in that nation's own language, it may add an exonym to an original name form. The alternative form should be shown in a font and/or text size which is different from and subordinate to that of the original toponym.

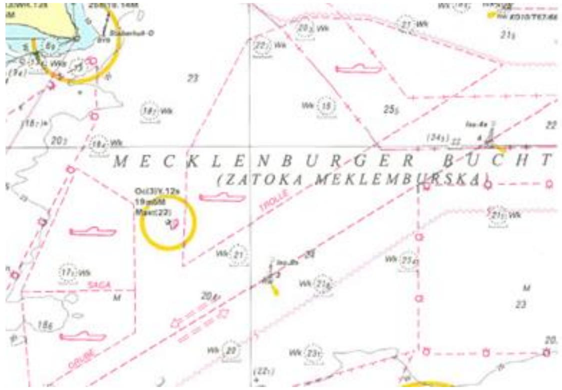


Figure 5-5 Example from Polish Chart 311 (2007), where the German sea area name MECKLENBURGER BUCHT is shown with the Polish exonym in brackets and smaller size (ZATOKA MEKLEMBURSKA)

5.8 Descriptive terms and toponyms

On charts of foreign coasts, toponyms should not be translated. Descriptive terms should be translated, if they are not the generic part of a toponym. However, it is not always easy to distinguish which legends are toponyms and which are descriptive terms. Their usage may provide a guide:

• If the requirement is to know the function of a feature for navigational purposes, for example *Puerto pesquero*, it is a descriptive term and should be translated.
• If it is required for reference, for example Vorhafen, it requires a distinctive designation and should be treated as a toponym and not translated. However, a translation of the generic term may be included in a glossary on the chart, in the index of abbreviations in the national version of INT1 or in another publication.
Descriptive terms should be reduced to a minimum by the use of symbols wherever possible, for example use of symbol F10 (from INT1) instead of descriptive term 'Fishing Harbour'.

5.9 Toponym references and authorities

International references.
The most important international reference for limits (but not names) of sea areas is: **IHO S-23 'Limits of Oceans and Seas'** (https://www.iho.int/iho_pubs/standard/S-23/S-23_Ed3_1953_EN.pdf). This publication names and defines the exact limits of the world's oceans, seas and major gulfs and straits, for the convenience of hydrographic offices when compiling charts and nautical publications. Generic terms are in English only. How those areas are named on a nation's cartographic products will depend on the national policy of each country. During the last 30 years several attempts were made to agree on an updated version of this IHO special publication but it was impossible to achieve consensus, especially in naming the sea area between the Korean peninsula and the Japanese archipelago. The Regional Hydrographic Commisions can decide for their area of responsibility to adopt parts of the 2002 draft (https://www.iho.int/mtg_docs/com_wg/S-23WG/S-23WG_Misc/Draft_2002/Draft_2002.htm). For the Baltic Sea, the Hydrographic Commision agreed to use Chapter 2. At the Baltic Sea International Chart Coordinating Working Group (BSICCWG), the work is still continuing to define sub-areas and their limits more precisely.

The most important international references for names of undersea features are:

- **IHO B-6 'Standardization of Undersea Feature Names'** (http://www.gebco.net/data_and_products/undersea_feature_names/documents/b6_ef_e4.1.0_2013_rev2017.pdf **).** This IHO-IOC publication has been published through collaboration between the Joint IHO-IOC Guiding Committee for GEBCO and the United Nations Group of Experts on Geographical Names. It is intended to ensure maximum international standardization of the names of undersea features, and includes in particular Guidelines for naming features, a name proposal form and a list of terms and definitions.
- **IHO B-8 'Gazetteer of Geographical Names of Undersea Features'.** (https://www.iho.int/iho_pubs/misc/gebco_system_upgrades.htm) This is a Gazetteer of Geographical Names of Undersea Features shown on the GEBCO sheets, on the small-scale international (INT) chart series, on the sheets of the Regional International Bathymetric Chart projects or of significance to charts at small scale. Proposals for new names should be checked first against any published gazetteers and then submitted for clearance, either to the appropriate national authority or, where no such authority exists, to the IHB or IOC for consideration by the GEBCO Sub-Committee on Undersea Feature Names (SCUFN), which may advise on any potentially confusing duplication of names.
- **IHO B-9 'GEBCO Digital Atlas'.** The GEBCO Digital Atlas (GDA) contains, amongst other data sets, the GEBCO gazetteer of undersea feature names. The IHO website should be consulted for the date of the most recent edition of the publications detailed above. See http://www.iho.int/iho_pubs/IHO_Download.htm and 'Data and products' at www.gebco.net.

Names of routeing measures (as specified in IMO publication Ships' Routeing) may not conform to official names. These reference names should be shown in capitals, e.g. *TSS SOUTH OF GEDSER,* on charts, specifically linked to the feature they are referencing, to assist the chart user, see figure 5-6.

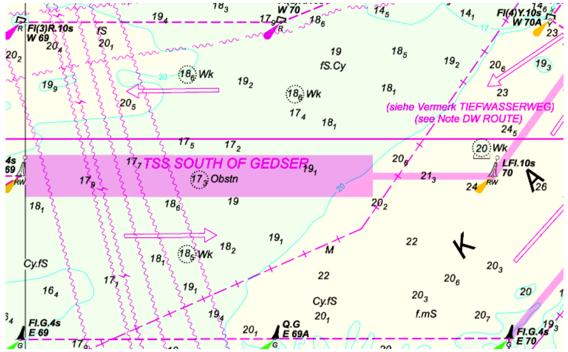


Figure 5-6 The name for the routeing measure „Traffic Separation Scheme TSS South of Gedser" in the German chart 163 (2016) according to IMO publication Ships' Routeing

National authorities. Most nations have set up a permanent or semi-permanent bureau responsible for determining toponomy for national mapping and charting. It is particularly important that states ensure that toponyms on charts of their own territory conform to forms authorized by these national names bureaus.

Non-national waters. The national bureaus will advise on the existence of official toponymy in other states and may be able to specify maps or other official publications or sources of toponyms for topographic and major sea features. It is possible in such cases, or if no national geographic names body exists, that there will be no official forms of toponyms of undersea and minor sea features. If so, charts of the area will probably have been produced by other hydrographic offices (which may have made hydrographic surveys in the past) and the names incorporated there should be used or adapted until official forms of toponymy are assigned to such features.

5.10 Topographic features

Agreed international forms must be used for the names of topographic features of continental or international extent and the names of oceans, seas, arms of the sea and major gulfs. If such forms do not exist, the form in common maritime use by the producer nation should be used. Alternative names may be added if the printer nation considers it useful to do so.

Country names
Names of sovereign states should be given in the original (endonym) form. The variant name (exonym) used by the printer nation may be added, underneath, unbracketed and in a text style and/or size which is subordinate to the original form.

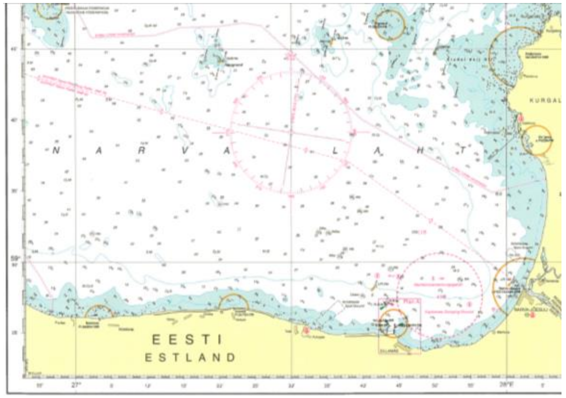


Figure 5-7 - Example in German chart 152 (2014): EESTI is the endonym and ESTLAND the exonym (in smaller size). The endonym for the sea area NARVA LAHT is shown.

Features marking or containing boundaries
If adjacent states do not agree on the names of features (for example: headlands, rivers) which mark an international boundary, or which contain an international boundary (for example: bay, strait), both forms should be given, in the same text style and size.

5.11 Chart lettering principles

In the context of this section, 'names' is used to include toponyms, descriptive terms and abbreviations.
The selection of names to be shown on a chart is an integral part of its design. Position, font, size, style, spacing and placement of names are also important to ensure optimum clarity. A chart must carry all significant names according to its scale. Names should be included on the chart only if they have some significance for the mariner. Names may be used to emphasise features, but irrelevant names may interfere with, or distract from, the reading of essential information. The insertion of names on charts should follow a number of well-established cartographic 'rules', most of which are consistent with other chart or map producing organizations. These rules are listed below, adapted to the specific case of nautical charts; if followed, they will significantly improve the clarity of the chart for the benefit of the user.

Placement and arrangement.
Each name must clearly refer to the relevant object; ambiguity must be avoided. Therefor a name must not be separated from the object to which it relates and must not be placed between two objects in such a way that it is unclear to which it refers. Letters must be evenly spaced. Spacing, if unduly wide, could make a name difficult to read as a unit because of other detail; against a blank background, there is almost no limit to the spacing, but as a general guide it should not be more than 5 times the height of the letters. Inter-letter spaces are measured between the adjacent edges of letters, not between their centres. Where names or descriptions consist of more than one word in line, the spaces between words must be consistent and should be approximately three times the space between consecutive letters of a word. Care must be taken when placing names on water areas to ensure that channels, estuaries, harbour entrances, etc, are not obstructed by names. In most cases, it is better to place names of towns, headlands, etc, on the land rather than on the water. Where possible, place the names of islands, lakes, large shoal areas, etc, on the features concerned rather than adjacent to them. Names and legends should be placed so that they do not break meridians and parallels unnecessarily or clash with detail in other colours. Names should not break the coastline unless unavoidable.
Examples of many of the rules governing the placement and arrangement of names are to be found on INT3 'Use

of Symbols and Abbreviations' published by the United Kingdom Hydrographic Office on behalf of the IHO (https://www.iho.int/iho_pubs/standard/S-4/index_INT3.htm) as a working tool for nautical cartographers.

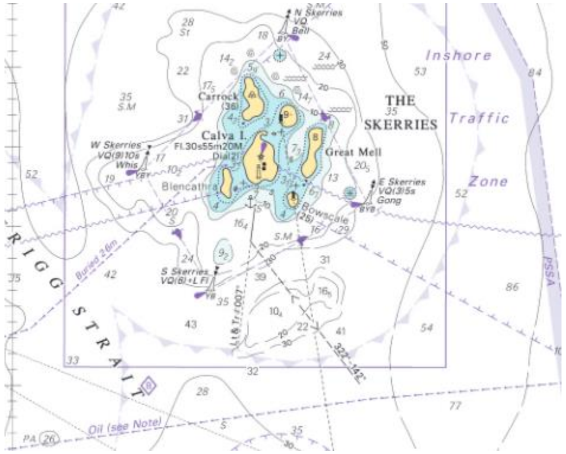
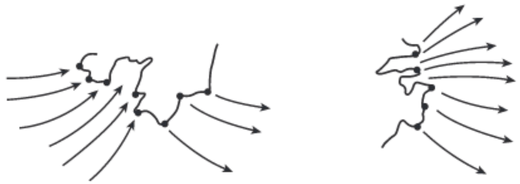


Figure 5-8 - Example for name placement in INT 3(2011)

Straight or curved names

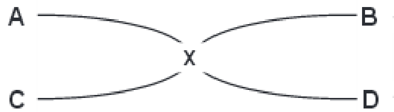
For maximum clarity, horizontal, straight names are preferred. Curved names tending to the horizontal are more readable than oblique straight names, except along linear features. Curved coastal names should meet the coast nearly at right-angles. Where curved names are used, the curve must become more nearly horizontal as it moves away from the point named. Where the trend of the coast changes, adjacent curved names should be either parallel with each other, or gradually diverge.



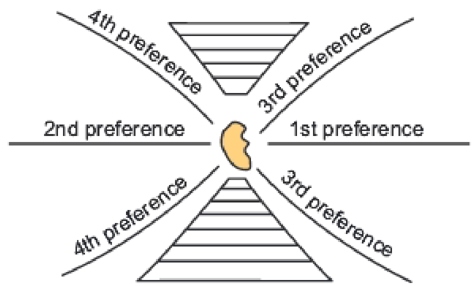
With the increasing extension of digital name placing curved names disappear from nautical paper charts.

Figure 5-9 - Example for curved name placement with curved light depiction in INT 3(2011)

It is useful for a national Hydrographic Office to provide preference guidance to its cartographers, while recognizing that other factors may dictate positioning. Here are some examples:
In the absence of other considerations curves B and C above are preferred in the western part of the chart, A and D in the eastern part (for ease of reading from the centre of the lower border).



The diagram below covers straight and curved names:



Point names and area names where the name cannot be fitted within the area

Ambiguity must be avoided by placing names close to the symbols or areas they describe, without obscuring them. Land names should, if possible, be placed on the landward side of symbols so that lines drawn from seaward will reach the symbols first. Names of towns, islands and headlands should, if possible, be placed on the land. If they have to be placed on the sea, headland names must not obscure depths or other important detail close in to the headlands. At sea, channel buoy and beacon descriptions should, if possible, be clear of the fairway, and must leave space for flares where buoys or beacons are lit. A name or legend describing a point symbol should, if possible, start immediately right of, or end immediately left of, the symbol. If a name cannot be placed immediately right or left of the symbol, the name may be placed above the symbol or, providing no detail is obscured, below it. Names and legends should be placed clear of radio circles, where that can be done without ambiguity.

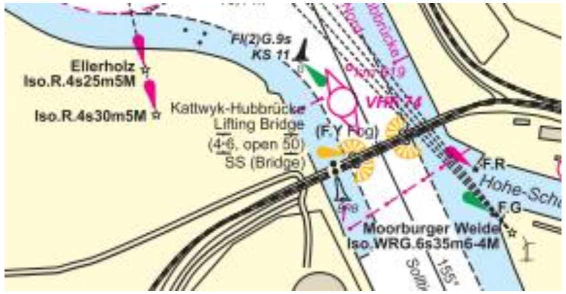
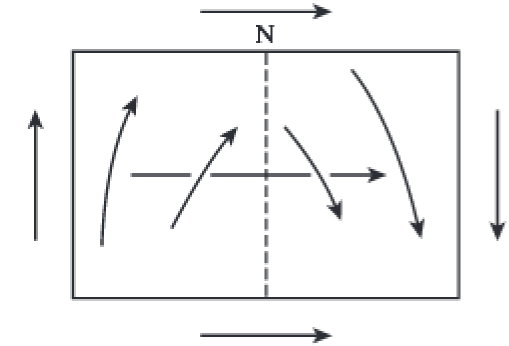


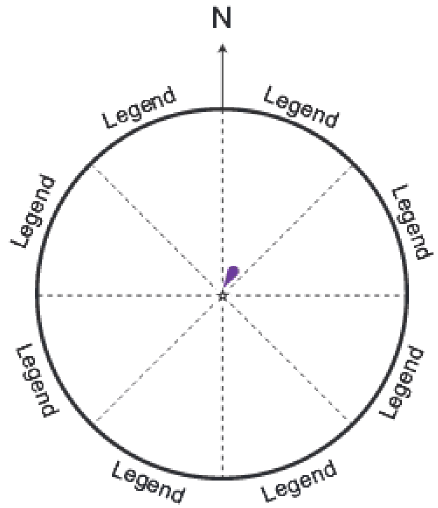
Figure 5-10 - Example for name and light description placement in German chart 48 (2017). The light description "F.R" is placed right of the symbol. The name "KS 11" is placed left of the symbol. The light description and the light name "Moorburger Weide" were moved down to prevent obscuring of light detail.

Consideration should be given to placing long names (consisting of more than two words) in two or more lines if they refer to a small feature. Names in two or more lines, such as the description of a light float, should be aligned vertically, justified on the side nearer the symbol or feature. The alignment of names of island groups, etc, occupying more than one line may, however, reflect the trend of the group rather than be left- or right- justified.

Names of Linear Features (for example: rivers; canals; pipelines) and along area limits.
Names are generally better placed above linear features rather than under. For rivers, name a section where it bends least and is as horizontal as possible. Very exceptionally, a double curve may be acceptable for winding rivers. Names and legends associated with lines representing area limits must be parallel to the line and within the area which the limit defines rather than outside it. Names and legends associated with area limits which run north-south should be oriented so as to be most easily read from the bottom centre of the chart (that is: if a name must be vertical), it should read from south to north in the western half of the chart, north to south in the eastern half. If curved, some of it may be partly inverted, but only if unavoidable. This may be illustrated thus:



Legends on arcs of light-sectors should be outside the arcs, rightway reading from the bottom of the chart, in accordance with the following diagram.



Where the arc is too narrow to contain the text along the arc, text should be rotated at right angles to the arc.

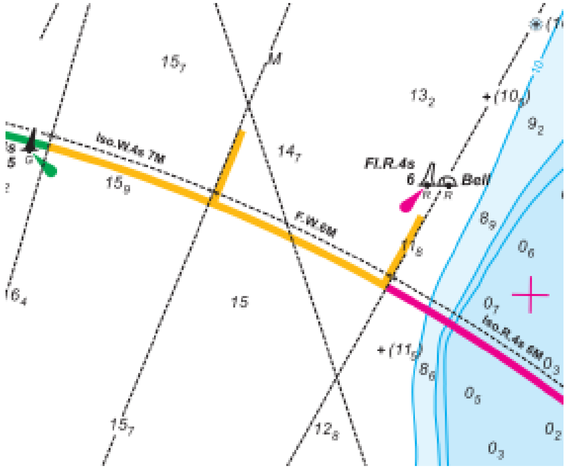


Figure 5-11 - Example for light description placement at light sectors in German chart 33 (2016)

Area names which can be fitted within the areas
Area names must reflect the extent of the area but must not be so fragmented or open-spaced that the letters or words are difficult to associate. The reading of important fine detail within the area must not be obscured. There should always be a close relationship between text size, weight and spacing. Names should cross the centre of the area. If names are divided into more than one line, each line should consist of whole words. Lines should appear to be linked together; if there are three or more lines, the spaces between them must be equal.

Where an area does not have charted limits:
• The name should be spread along the longer axis of the area to show its extent.

• The name of a mountain range or ocean ridge should follow the trend of the crestline, although offset to allow portrayal.
The name of a bank should indicate the overall mass of the bank, not necessarily be confined by the shoalest contour. It may be preferable to offset the name of a bank to display a clear sounding pattern. Exceptionally, names may be repeated over a large area with uncharted limits in order to suggest the extent.
Where limits are charted:
• The name should extend for more than half the length of the area.
• A space should be left at each end of a name or legend of at least 1.5 times the inter-letter space; longer spaces at the ends are acceptable to avoid unduly wide spacing.
The name of a strait takes the coastlines as its limits.
• The name of a channel takes either coastlines or the banks which confine it as its limits.

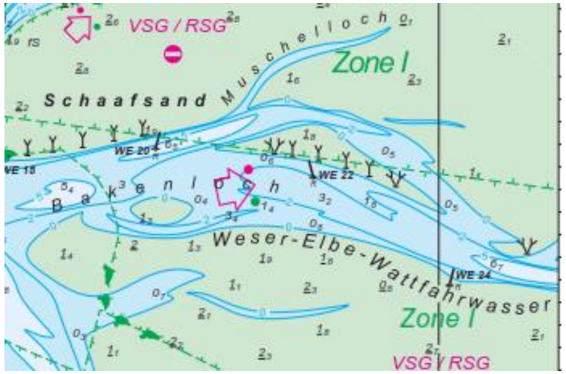


Figure 5-12 - Example for a bank name Schaafsand in German chart 2 (2017). There is a national rule to offset the Wadden fairway names as Muschelloch or Elbe-Weser-Wattfahrwasser to read soundings and buoyage better

5.12 Text styles and fonts

The choice of text styles (for example: upright; sloping; light; bold) and fonts (for example: Arial, Times) is an integral aspect of nautical chart standardization and the creation of a homogeneous set of international charts. A certain degree of standardization is required to achieve a reasonable level of compatibility between the charts of different nations, so that users moving from one to another will feel no practical difference and so that international or national charts can be adopted by nations with a minimum of change: it should be easy for printers to match fonts sufficiently close.

Text styles: general rules
Except for the title, explanatory and cautionary notes and marginalia, the use of different text styles on charts should satisfy the following general rules:
a) A distinction should be made between geographic names referring to land features (including islets, above-water rocks, fixed marks, etc) and those referring to water features (whether on land or part of the sea).
b) A distinction should be made between those legends, words and abbreviations which relate to general or 'background' geographic information and those which relate to significant navigational information. The latter should be shown prominently while the former should be subordinated.

Distinction between land and water names and legends
Names and legends referring to land features should be in an upright (Roman) style and those relating to water features in a style sloping to the right (italic). For features, such as pontoons or locks, which are difficult to define as either 'land' or 'water' objects, there are special specifications.

Distinction between general geographic and navigationally significant names and legends should be achieved by making use of either different weights and/or sizes within a single font family (all based on the same design), or of different families, for example serif and sans-serif fonts.

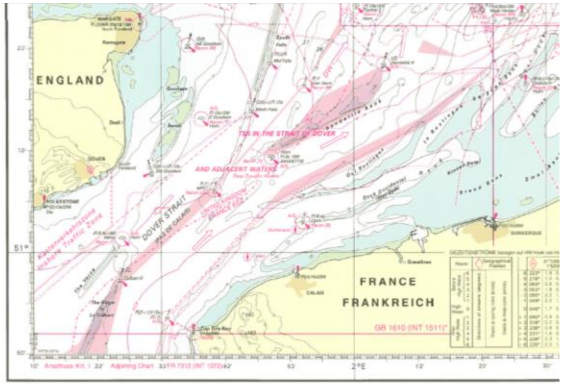


Figure 5-13 Distinction between upright land and sloping water names in German chart 53 (2013)

Prominence of navigational information. Text weights and sizes must be selected according to the relative importance of the various names and legends to be shown, whether general geographic or navigational. However, it is important to ensure that navigational information, such as for example the traffic Separation Schemes in figure 5-7, predominates.
Certain classes of features should, irrespective of their relative importance, be shown consistently in a particular size and weight of lettering. This applies particularly to such features as light descriptions and legends describing landmarks.

Choice of font
The choice of font should be limited to a few, common families. If a single font style is used it should be a sans-

serif family, and if two font styles are used they should comprise one sans-serif and one serif family.
To assist chart adoption by printer nations, fonts should be selected from those which are universely available, for example: Univers (sans-serif), Times (serif). Note: The font that currently (2013) contains the broadest range of Unicode compliant encodings is Arial Unicode MS.

Use of sans-serif fonts.
If a single, sans-serif font is used for navigationally significant features (for example for landmarks, summit names and heights, aids to navigation and underwater hazards), these should be differentiated by thickness (the use of bold or light lettering).

Use of serif fonts.
If a serif font is used (in addition to the sanserif font), its use should be restricted to the more important geographic names which are not navigationally significant features. Minor non-navigational names and legends should be in a light sanserif font.

Title. The choice of font used for the title, explanatory and cautionary notes and marginalia is left to national discretion, subject to being universely available (see figure 5-13).

In German charts, there exist no more serif fonts – a sans serif family is used.

5.13 Chart titles

The titles of charts, in English or national language of the producer, shall preferably be arranged in one block, located in the land area if possible, clear of essential detail. The title block shall include the following items, reading from top to bottom:

- The seal of the producer nation and the IHO seal shall be placed above the title, side by side and of equal height, with the producer nation's seal on the left.
- The words 'INTERNATIONAL', or equivalent, shall be shown above and 'CHART SERIES', or equivalent, below the seals.
- The name of the ocean area on the chart according to S-23, Limits of Oceans and Seas.
- The geographical area reference, if appropriate for example: Asia; South America; Europe, etc.
- The scale and the common mid-latitude
- The unit of measure for depths and for heights
- The projection used and
- a note citing the producer nation and the sources used in the compilation.

INTERNATIONAL CHART SERIES

BUNDESAMT FÜR SEESCHIFFFAHRT UND HYDROGRAPHIE

NORTH SEA, GERMANY

ENTRANCE TO THE RIVER ELBE

SCALE 1 : 50 000 (54°)

EXPLANATORY NOTES
HEIGHTS AND DEPTHS IN METRES
PROJECTION: Mercator
POSITIONS: World Geodetic System 1984 (WGS 84)
HEIGHTS are referred to Mean Sea Level
DEPTHS are referred to Lowest Astronomical Tide (LAT)
VERTICAL CLEARANCES are referred to Highest Astronomical Tide (HAT)
SOURCES: German surveys
SYMBOLS AND ABBREVIATIONS see INT 1

Figure 5-14 English title and explanatory notes
In German chart 44 (New Edition planned 2017)

5.14 Antarctic Chart Names

Topographic names in German Antarcti charts are used according to the StAGN principle: Who discovered the land area first and published a name for it – this name has to be used in this language; so, the generic parts are not translated.
In the case of Germany, sea names on Antarctic charts have been decided in agreement with the local institute in charge of (ant)arctic research (here the Alfred Wegener Institute Helmholtz centre for Polar and Marine Research) to show in English language as the preferred language for nautical use. If a German well established exonym exists, it can be added. The German charts for the Antarctic region are part of the international INT chart system of the International Hydrographic Organization and especially the Electronic Nautical versions are used by the English-speaking sailor and research community.
Furthermore, the Antarctic region is an area offshore from territorial waters and economic zones and south of 60°S the Antarctic Treaty is valid. The principle to use an official name of a corresponding country is not applicable.
Here are some examples for names from different sources and languages used in German Antarctic charts.

Norwegian sources:
- Auståsen
- Ekströmisen
- Halvfarryggen
- Unneruskollen
- Jelbartisen
- Blåskimen
- Fimbulisen
- Trolltunga
- Dronning Maud Land
- Guttom Jakobsenbukta
- Muskegbukta

• Byrdbukta

German sources:
• Neumayereishöcker
• Rüsseleishöcker
• Atka-Eiskuppel
• Sicheleishöcker
• Neuschwabenland

Russian sources:
• kupol Novyj
• zaliv Pribylova
• kupol Moskovskij

US American and South African sources:
• Atka Iceport
• Apollo Island

Finally, the German charts for Antarctica incorporate endonyms in Norwegian, English and Russian, apart from German endonyms and exonyms.

Examples of names used on German Antarctic Charts
[Name - Source - description]
Atka Bank SCAR CG Endonym
Atka Iceport SCAR CGA Endonym
Austasen Bank GEBCO-SCUFN Gazetteer Endonym
Bayer Canyon GEBCO-SCUFN Gazetteer Endonym
Bungenstock Plateau GEBCO-SCUFN Gazetteer Endonym
Byrdbukta SCAR CGA Endonym
Drygalski Canyon GEBCO-SCUFN Gazetteer Endonym
Ekström Basin SCAR CGA Endonym
Fimbul Canyon GEBCO-SCUFN Gazetteer Endonym
Imhof Knoll GEBCO-SCUFN Gazetteer Endonym
Jelbart Basin SCAR CGA Endonym
Mercator Knoll GEBCO-SCUFN Gazetteer Endonym
Muskegbukta SCAR CGA Endonym
Neumayer Canyon GEBCO-SCUFN Gazetteer Endonym
Pribylova Zaliv SCAR CGA Endonym
Sanae Bank GEBCO-SCUFN Gazetteer Endonym
Sanae Canyon GEBCO-SCUFN Gazetteer Endonym
Torge Plateau SCAR CGA Endonym
Weiken Basin GEBCO-SCUFN Gazetteer Endonym
Lazarevmeer StAGN Exonym
König-Håkon-VII-Meer StAGN Exonym

[Name – Source – Description]
Apollo Island CGA Endonym
Atka-Eiskuppel StAGN, CGA Endonym
Austasen StAGN, CGA Endonym
Blåskimen StAGN, CGA Endonym
Dronning Maud Land StAGN, CGA Endonym
Ekströmisen StAGN, CGA Endonym
Fimbulisen StAGN, CGA Endonym
Halvfarryggen StAGN, CGA Endonym
Jelbartisen StAGN, CGA Endonym
Kronprinsess Märtha Kyst StAGN Endonym
kupol Moskovskij StAGN, CGA Endonym
kupol Novyj StAGN, CGA Endonym
Neumayereishöcker StAGN Endonym
Neuschwabenland StAGN Endonym
Rüsseleishöcker StAGN Endonym
Sicheleishöcker StAGN Endonym
Trolltunga StAGN, CGA Endonym
Unneruskollen StAGN, CGA Endonym

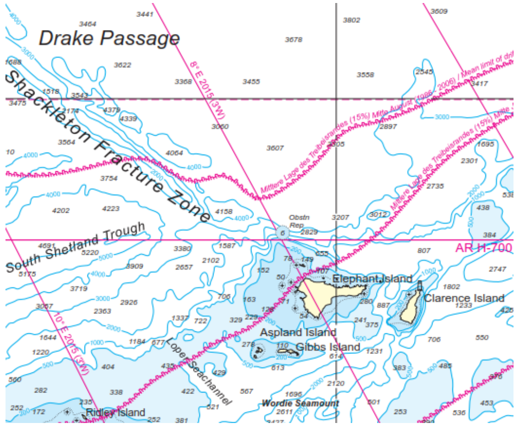


Figure 5-15 Sea and land names in the German chart 1700 (2017)

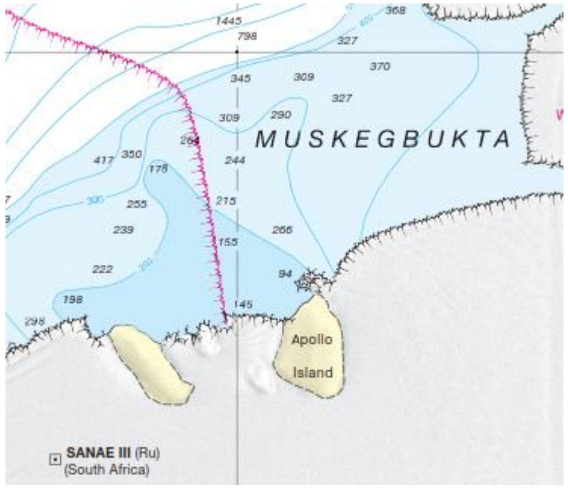


Figure 5-16 Sea and land names in the German chart 1701 (2009)

5.15 List of acronyms used

BSICCWG Baltic Sea International Chart Coordinating Working Group
CGA Composite Gazetteer of Antarctica
GEBCO General Bathymetric Chart of the Oceans
HCA Hydrographic Commission on Antarctica
IHB International Hydrographic Bureau
IHO International Hydrographic Organisation
IOC Intergovernmental Oceanographic Commisiion
SCAR Scientific Committee on Antarctic Research
SCUFN Sub-Committee on Undersea Feature Names (of the IHO)
StAGN Ständiger Ausschuss für geographische Namen: Body that standardizes the spelling of geographical names in German

Section 3 Regional activities and different applications

Chapter 6 **Geographical Names for Tourist Maps**

Roman Stani-Fertl

6.1 Introduction

Tourism is subject to marketing tactics. A region, a town wants to be attractive and interesting to potential visitors. It intends to wake the interest of the people to visit. Those toponyms that indicate the geographical entities play an important part. In addition to traditionally established geographical names, areas create new artificial names, which stand for the characteristics of the respective region, highlight the unique landscape, outstanding objects, or reflect the current spirit of the time. Both traditional names and these unofficial names are aimed to be easily recognized and memorized for the target group. Additionally, the traveller may enter a region of different customs or even a disparate culture. Apart from living conditions many other things might be different, strange, or simply other than at home. This could be such factors as diverse shopping hours, a complex code of behaviour, another language or a different script. Many tourists are not well prepared for this cultural change. Tourist maps may add a valuable contribution to make this change more conscious (see figure 6-1).

6.1 What is a tourist map?

Tourist maps serve for orientation of non-local people when planning a trip to a certain region and also when visiting this region for a more or less short period. Their contents are highly targeted. They are specifically oriented to tourism and are specifically tailored for the obvious requirements of travellers, for planning the journey and when travelling at the holiday destination. They illustrate the destination at its best. It is helpful to differentiate two types of tourist maps.

Figure 6-1 Detail from the pictorial tourist map of the Isle of Skye in Scotland, United Kingdom. The map is using pictorial symbols to show the beauty of the island and stress the picturesque landscape. What is missing in this map are the Scottish Gaelic place names which are also official in this region. Those names are valuable evidence for the cultural background of the history of the Isle of Skye. Places are only shown with their English names. ©Wyvex Media Limited

The first one is a tourist map in a narrower sense. It is designed to provide information like scenic spots, sights and facilities like hotels, public toilets, souvenir shops etc. to its users. This type is tailored for tourists. An example for such a map is a small scale general map shown in figure 6-1.

The second type of tourist maps must be regarded in a broader sense and comprises city maps, street maps and so on, which are not especially targeted for tourists only, but also for other user groups like businessmen, day trippers, and residents. These maps are used predominantly for orientation on the spot.

6.2 Points to observe in tourist maps

An additional aspect, which characterizes tourist maps, is that the presentation of its content is well designed to the expectation of the target group. Symbols chosen for maps will fit with the convention of the potential visitors and in many cases, this will not correspond with the locally used convention. Tourists arriving at their destination sometimes will be confronted with a symbolization different from that found on the material used for planning the itinerary. Especially the writing of geographical names might be very different.

6.3 Spelling of toponyms

Tourist maps are user oriented. They are published in the language affiliated to the target group or at least in a language which is well known by the group. So many geographical names either in leaflets and catalogues or in travel publications do not correspond with the name form at the destination.

a. Conventional names

Many languages have adapted different name forms for famous (historical) sites. These names differ in many cases from the locally used name form. They are called "conventional names" or "exonyms" and the exonym usually is the only name form was known in the language of the country of origin of the tourists. Furthermore, it is mostly unknown to visitors that there exists a different local name for the particular (historical) site. The use of these exonyms is no problem when planning a journey because most of the established cartographic products, atlases or web sites like Google maps or Bing maps use these conventional names. Problems may arise on the spot when the tourist will not find the exonym but only the local endonymic form. On the other hand, on a guided tour with a local tour guide speaking in the language of the traveller, he or she will probably will never realize that there is a different local name form existing.

b. Script

Orientation is getting more complicated when the script the traveller is used to, differs from the script used at the holiday destination. In some regions, the Romanized name form can be found below the locally used one on sign posts and street signs. Other areas do not offer this comfort and orientation is hard. Either the traveller must

Figure 6-2 shows a detail of the tourist map of Macau. In Macau, there are three languages of relevance for tourism using two different scripts: Chinese, Portuguese, and English. The map provides the toponyms in all three languages. Chinese and Portuguese are official languages and English is the language of publication. One can realize that it would sometimes be very difficult to identify a feature unless all the languages were published together. (c) Macau Government Tourist Office.

attempt to need the signs himself, or will feel awkward asking the local people.

With different scripts, a tourist may encounter two additional problems which are described as follows:

The locally used script may be so complicated and strange that it is almost impossible for the potential reader to identify the spelling on the sign post that corresponds with the name on his map, unless he has at least a basic knowledge of the script (e.g. Chinese, Devanagari, ...) (see figure 6-2).

Another problem we are addressing with travel publications in general and tourist maps in particular is when the language of the publication is other than the mother tongue of the user. Nowadays English is seen as a kind of lingua franca and many travel guides and tourist maps are published in English and are sold also outside English-speaking countries. Also, many tourist spots indicate descriptions and information in English beside the local language. The traveller using non-English publications may be confronted with different forms of spelling for one feature (local name and English exonym). Additionally, confusion can exist through the use of different Romanization systems. A user with a non-English speaking background may be faced with the Romanized spelling of a name having a different type face than is used in his own language.

A further complication is found where there is an exonym in common use in just one language and another language is using the endonym. For example: In the Russian town Kaliningrad, there is a tourist site called „Закхаймские ворота “[rus]. The German exonym is “*Sackheimer Tor*” and there is no common English exonym existing. Hence English is using the endonym in the English Romanized form “*Zakkhaymskie vorota*” [eng] or the adapted form of “*Zakkhaymskie gate*” with the generic term translated. If the endonym were to be used in German the Romanized form would read quite differently as “*Sakchaimskije worota*” [ger]. The UN- approved form would be “Zakhajmskie vorota”.

6.4 The function of toponyms in tourist maps

To promote conscious and intercultural travel that creates not a confrontation but an open encounter of the traveller with the local people and its culture, it is necessary to start this encounter during the preparation for a journey. The tourist has to be confronted with these cultural differences while he is at home planning his itinerary.

Tourist maps provide an excellent means to picture these cultural changes. By using the endonymic and exonymic name forms the cultural differences can be made very obvious on such maps. The use of the endonym in its original spelling, its Romanized form and the exonym of publication is a helpful step in the direction of conveying diversity (see figure 6-2 and figure 6-3).

It is necessary to use exonyms in tourist publications to attract the potential customer at his home location. Tourism is oriented to the customer and this requires the publications to be as understandable as possible, in order to express the various cultural differences, as they manifest themselves in language and writing, beside many other aspects. It is the responsibility and duty for editors of travel publications and tourist maps to provide both name forms of a geographical name. The user must find the endonym(s) and the exonym. If the endonym is written in a different script the original local spelling and a Romanized form must be given too.

Figure 6-3 Seoul Metropolitan Government is publishing a tourist map in English which is a good combination of local toponyms and English exonyms. All geographical features (streets, public buildings ...) are labelled with their local Korean spelling and with the Romanized name form (강남구 / Gangnam-gu; 봉은사 / Bongeunsa). Additionally, some generic elements describing the character of a feature are translated into English (e.g. 선정릉 / Seonjeongneung / Royal Tombs; 봉은사 / Bongeunsa Temple). This will certainly assist tourists to orient themselves and to find interesting spots to visit. (c) VisitSeoul

6.5 References:

- Macau Government Tourist Office: *Macau*. http://content.macautourism.gov.mo/uploads/mgto_interactive_corner/mapx1_macau.pdf

- Nafla, Kh., Salman A. (2011): *The English Transliteration of Place Names in Oman.* Pp. 1-27 in Journal of Academic and Applied Studies, Vol. 1(3). September 2011. https://www.academians.edu/1342044org/Tthe-Eenglish-Ttransliteration-of-Pplace-Nnames-in-Ooman

- Wyvex Media: *Skye North*. http://www.scottishtouristmaps.co.uk/2015/SKYENORTHMAP2015.pdf

- Stani-Fertl, R. (2009): *Geographische Namen als redaktionelles Problem – Systematische Untersuchung der Exonyme mit einer Anleitung zu deren Verwendung in der redaktionellen Praxis.* Südwestdeutscher Verlag für Hochschulschriften. Saarbrücken.

- Stani-Fertl, R. (2015): *Implementation of Interlingual Allonymy in Cartographic Products.*
Pp. 325-341 in: Jonas Löfström, Betina Schnabel-Le Corre (eds.), Challenges in Synchronic Toponymy. Narr Francke Attempto, Tübingen.

- VisitSeoul: *Seoul Tourist Map*. file:///D:/Downloads/Seoul%20Tourist%20Map_0420.pdf

- Wobek B. (2001): *Deutsche Exonyme in Reiseangeboten. Beispiel Ostmitteleuropa*. Diploma thesis at the University of Vienna, Wien.

All web sites checked on 4th November 2015.

Section 3 Regional activities and different applications

Chapter 7 **Street name standardisation**

Brahim Atoui

7.1 Introduction

Street name standardisation is just as necessary, for both national and international communication, as place name standardisation.

In place name standardisation, place names are processed by a recognised toponymical authority on the basis of a number of rules and normative criteria that are essentially aimed at guaranteeing the univocity of these place names. Univocity is to have only one standardized spelling form for the name of each geographical feature in each writing system. Street name standardisation answers the same criteria, but also respects some specific rules both regarding the standardization of the generics that form part of the street names and also regarding the standardisation of the material support of these street names, that is the street signs.

But there is a difference: place names often are the result of spontaneous action by the local population (it is the inhabitants themselves that name a place, on the basis of its usefulness or its specific characteristics. This leads to place names that result from the practical usage of these named features, based on free, popular imagination. This actually seems to be a natural process. And this is opposed to the attribution of street names, which often, if not always, is a process steered by the administrative or political authorities of a country.

Before we continue we should first define the term ‘odonym’ or ‘street name’:

“An odonym or street name is a proper name that designates a street.”

In the same way one could define odonymy as the study of street names, or, more general, as the study of urban names, whether it is street names, names of buildings, institutions, constructions like bridges, public parks, or even monuments that have an important commemorative character.

7.2 Short historical development of street naming: between popular spontaneity and political interventions

In the past, it did not use to be the task of the administrative or political authorities in a country to take charge of the street naming process, just as it did not take charge of toponymy either. Place names in the past did not only allow for the location and identification of topographical features, but they also used to inform us about the functions of these features. This resulted in spontaneous names in the local languages, for example Cedar Mountain (the mountain where cedars grow) or Boar River (the river on which banks wild pigs roam).

It is only in the last two centuries, especially so for developing countries, that street naming has been taken up by (local) authorities. Since then it has become the exclusive privilege of these authorities to decide about and select the street names.

This is why official street name standardisation, from its inception, also has been subjected to a number of specific administrative criteria that will be exposed below.

7.3 General principles of street name standardization

The attribution of street names should be significant, taking into account the diverse contents that have a direct or indirect relationship with the named object: the names selected should refer to the history of the country or the city and its memories, to the flora and fauna in its environment. These street names should open up the world nearby or further away by referring to places or countries with which relationships or bonds of friendship exist, with whom cooperation has been established, or with the world of science, of arts or of sports., etc.

The attribution of names can be an act of homage or recognition for the great personalities that have marked the history of the location, of the city or the country as well as of persons with a global impact such as Nobel prize laureates who, by their work or by their acts, have delivered an important contribution to the development and well-being of humanity.

That is why, in order to give concrete form to this approach, criteria for street name selection will be favoured that support the sense of belonging to a town or to the country and its opening to the world.

To this end every country should be able to establish its own specific criteria. Those listed below just serve as an example.

Whenever there is a need for the attribution of street names, there immediately is also a need for its standardization.

Just like toponyms, odonyms (street names) can be constituted from a generic and a specific part. Each of these parts needs standardisation.

Standardisation of the orthography of specific place name parts

Stemming from the fact that the specific parts of street names usually derive from personal names, one should take great care to respect the orthography of these personal names, and copy the exact orthography from the population register. No deviations should be allowed. The name should be written in capital letters, or at least the first letter of the name should be a capital, and the name never should be abbreviated.

Standardisation of the orthography of the generic parts of place names

Here one first should define the various levels of streets and roads, and link them with matching generics. But one also should take account of the fact that some countries use the same definition for all types of streets and roads. Consequently, they would not distinguish between avenues, streets, boulevards, etc.

The generic parts one would encounter in an agglomeration could be the following (this list is not complete, but just serves to give some examples):

- ✓ **Lane:** street, with trees on both sides, flower beds and lawns that allows for the passage of vehicles and walkers into a garden, a park or wood.
- ✓ **Avenue**: Artery of the road network that serves as collector. This artery would provide access to the local streets.
- ✓ **Boulevard**: Major artery of the road network that links several town sectors
- ✓ **Intersection:** Intersection point with obligatory driving direction
- ✓ **Lane**: public country road
- ✓ **Blind alley**: road that comes to a dead end,
- ✓ **Arcade**, small covered street usually prohibited for vehicles.
- ✓ **Foot-bridge**: narrow bridge reserved for pedestrians.
- ✓ **Bridge:** construction by which a road, aqueduct, or a conduit crosses a water course, a depression or another road**.**
- ✓ **Promenade**: Road especially adapted for promenaders
- ✓ **Street**: local public road part of the local network
- ✓ **Alley**: small narrow road
- ✓ **Footpath**: narrow road destined for pedestrians
- ✓ **Rotunda/Roundabout**: circular crossing.
- ✓ **Square**: exposed small or large open public space in a city or village, usually surrounded by public buildings, where several streets or avenues would end, and where frequently commercial, festive or public events are held.
- ✓ **Square park**: small public garden, usually fenced, in the centre of a square

7.4 Selection of names reflecting the importance of the parts of the road network.

The importance of the parts of the road network should be matched by the importance of the persons or concepts these road network elements are named for. According to a logical and standardised hierarchy one should define the order of attribution of names, in function of the criteria linked to the type of road concerned, such as the importance and quality of the name selected (international, national, regional or local renown, and related to a political, scientific, artistic, sports or military background).

In this way, for the grand arteries like boulevards and avenues, names should be selected of personalities that own their renown to an exceptional contribution towards human development.

It is also suitable to go for favoured name themes, that are linked to the specific characteristics of the site – if these can be found – for instance select names of scientists for those streets that run towards the university or to research institutions, select names of famous medical scientists to those streets that run towards hospitals, and select names of animal or floral species for roads that are close to a zoological or botanical garden, etc. This procedure can be called systematic appellation. In order to establish an equilibrium, and to diversify the names, it is also recommended to select names that refer to vegetation species.

The obligation for a direct (tangible) link to exist between the name and the named object, i.e. the road or street, is not required. It is perfectly suitable to name a road *lilac tree street*, even if both in the past and the present no single trace of this vegetable species can be found there! Here tribute would be paid to the beauty of the tree, not to its presence.

Finally, only those persons can have a street named after them, that died a certain number of years ago (the number would be different per country).

7.5 Standardization of the street sign supports

A good street name signage will help towards a better traffic circulation and will increase the possibilities for orientation and locating a destination. A place name sign is an essential location or orientation element, as it helps to guide users in finding their destination. Street name signage facilitates and accelerates access to first aid, distribution of the mail, and rapid intervention in case of disasters.

A proper signage also contributes in providing a town or a municipality with a feeling of identity; and through this common identity in turn it provides the town with a sense of coherence. The signage supports a valorisation of the cultural and historical heritage of the municipality or the town.

The street sign itself can become a true symbol of the city and even a tourist gadget. The Paris name sign for instance can be cited as an example of a material object that constitutes an integral part of the cultural heritage of the city and a true touristic object.

That is why great care should be applied to the choice of sign supports, both from the point of view of the selection of the materials (that should be sustainable), of the script (legibility, sustainability), the colours (contrast), the additional information that may serve as an explanation of the name selection) as well as the place (on the wall close to a street corner or on a freestanding pole) where the sign is applied.

Several choices can be made here, and there is evidence of international experiences here as well.

7.6 Some international examples of street name signage

The example of France

The example of France shows that every municipality may have its own particular type of street signs, which may lead to a sense of identity.
We may refer to the following examples:

Municipality of Paris

- Street signs are applied to walls
- The blue street name signs with white script in a green rectangle with a crescent on top indicating the arrondissement provides three types of information:
- The generic and specific parts of the street name written in white capital letters
- The number of the arrondissement (city quarter) written in white letters
- The generic and the specific are written one beneath the other (see figure 7-1).

Figure 7-1 Example of a Paris street sign

Municipality of Vitry sur Seine:

Vitry is a municipality in the southern suburbs of Paris, about 7 kilometres from Paris centre. This white street name sign (see figure 7-2) hangs on a dark brown post and bears the following information:

- The municipality name written in red letters
- The municipality emblem
- The generic and specific street name parts written in black (big characters written in lower case letters only)
- Information about the person the street was named for
- His or her date of birth and demise
- The specific parts have been placed beneath the generic parts

Figure 7-2 Example of a street name sign from Vitry sur Seine

Municipality of Kremlin-Bicêtre

The town of Kremlin-Bicêtre is located in the southern suburbs of Paris about 5 kilometres from Paris centre. Here, the rectangular street name signs with a blue background are attached to black poles. The signs have a white rectangle of a distinct shape, and bear the following information (see figure 7-3):

- On top the name of the town, below both the generic and specific street name parts, written in white capital letters, followed by the qualifications of the person the street was named after
- The generic and specific parts of the street name have been placed one beneath the other

Figure 7-3 Street name sign from Kremlin-Bicêtre

Examples from Belgium

In the Belgian capital Brussels, which is officially bilingual, the rectangular street name signs are bilingual French/Dutch (see figure7-4). White characters as well as the logo have been placed on a blue or green ground. In figure 7-5 the complete street name has been translated (Rue des Néfliers (French) = Mispelaarsstraat (Dutch)). In English, this signifies ‘Medlar tree street’. Note that the specific and generic street name parts are on different lines. If the specific part is a personal name (as for instance Mr Boileau), this of course would not be translated. Note that in French the generic (rue or avenue) is not attached to the specific, while in Dutch it is. Figure 7-5 also is proof of the attempt to have as short a place name sign as possible, and to have the specific part stand out, as that would be the part the travellers would be looking for. When the specific part needs no translation, it should not be repeated, so instead of Avenue Boileau/Boileaulaan, the joined solution in figure 7-5 has been selected, instead of the complete translation.

Figure 7-4 Brussels bilingual street name sign

Figure 7-5 Brussels bilingual street name sign, with the specific untranslated

Figure 7-6 Brussels bilingual street name sign

In figure 7-6 a ‘traquet’ or ‘zwartkeeltje’ is a bird species called ‘wheatear’ in English.

The example of Canada

Québec city

Here the street name sign is attached to a building. It has a pronounced rectangular form, with a blue foundation and a white rectangle which contains the name starting with a capital letter, on the same line (see figure 7-7).

Figure 7-7 Street name examples from Québec

The example from the Republic of Korea

Here the blue street name sign with white characters bears:

-an indication of the (range of) numbers of the row of buildings in a block

-the street name in Korean (and Chinese, see figure 7-10) characters with its transcription in Roman letters

-an indication of the (16th) district in which the street is situated.

Figure 7-9 Street name sign from Seoul (ROK)

Figure 7-10 Triscriptual name sign from Seoul, with street names in Korean, Roman and Chinese characters.

The example from the Czech Republic

Figure 7-11 Street name sign from Prague

The street name sign in figure 7-11 is from Prague. Here the rectangular red street name sign with a stylised white border bears the name in the same colour as the border. Below the street name in large characters, the name of the city quarter in which the street is situated is listed in smaller characters, also in capital letters.

Figure 7-12 Street name from Istanbul

Example from Turkey:

Figure 7-12 shows a street name sign from Istanbul. The red part indicates the street name and the numbers of the buildings on this block; the white part contains the name of the neighbourhood or city quarter and the lower, blue part contains the name of the city district.

Example from the United States

UNGEGN-participants would be familiar with the street name signs in New York city, particularly those close to the United Nations Building where the UNGEGN-meetings are being held.

Figures 7-13 and 7-14 show two types of street name signs used in New York City: one on a green background with the image of the statue of Liberty, and one on a blue background, with a white border. Both signs have their lettering in white.

Figure 7-13 street name sign from New York city, with logo (freedom statue) attached.

Figure 7-14 Street name sign from New York City, with the numerals in Arabic or Roman characters

New York also has some examples of street names written on the pavement (see figure 7-15).

Figure 7-15 New York street name sign applied to the pavement

Example from Spain

Figures 7-16 and 7-17 show Spanish examples on a marble support with engraved letters enhanced in black.

Figure 7-16 and 7-17 Spanish place name signs

Why is it necessary to add Roman transcriptions to street names and place name signs?

In a world without boundaries one should try to avoid confusion regarding toponyms in general and street names in particular – to improve both information transfer and communication. That is why it is advocated to write toponyms and street names not only in the local script, but also in the Roman alphabet, following a conversion system approved by the competent authorities of each country or by the United Nations Group of Experts on Geographical Names.

Figures 7-18 and 7-19 show unilingual examples, and signs such as these are to be avoided.

Figure 7-18 Street name sign from China

Figure 7-19 street name sign from Algeria

One should rather favour to have street name signs in the local language with their Roman script version added, as in the following examples (figures 7-20 - 7-22):

Figure 7-20 Street name sign from Kuala Lumpur

The street name sign from Malaysia (see figure 7-20), from the city of Kuala Lumpur, is rectangular and has a blue background with Arab script in yellow characters and Roman script in white letters with, in orange Roman characters, the name of the city and the local postal code.

The street name sign from Algeria in figure 7-21 has the street name in Arabic script on top, followed by its conversion into the Roman alphabet. Beneath, information has been added on the person the street was named after. The city arms including the Atlas lion, have been added to the name sign.

Figure 7-21 Bilingual and biscriptual name sign from Alger, with the name in the Arabic alphabet (official language alphabet) and in the Roman script.

Figure 7-22 Name sign from Den Pasar, Indonesia with the name in Roman alphabet (official alphabet for the Indonesian language) and in the Balinese script.

Section 4 Regional and global initiatives

Chapter 8 **Building the New Zealand Gazetteer**

Wendy Shaw

8.1 Introduction

An important function of any country's national naming authority is to produce a publicly available list of all authoritative geographic names – both official and unofficial. This training module describes the reasons for having an easily accessible Gazetteer and the processes followed by the New Zealand Geographic Board (NZGB) to produce its online Gazetteer in June 2013.

8.2 Statutory Requirements

The NZGB is governed by the New Zealand Geographic Board Act 2008 (NZGBA) which replaced the former 1946 Act. Section 13 of the NZGBA refers:

Gazetteer

13 Gazetteer to be created by Board

(1) The Board must establish and maintain a publicly available record known as the New Zealand Gazetteer of Official Geographic Names.

(2) The Gazetteer must record all official geographic names and the relevant *Gazette* or statutory reference for each official geographic name.

(3) The Gazetteer—

(a) must include a record of—

(i) the type of geographic feature or Crown protected area that is named; and

(ii) the positional reference for the feature or area; and

(b) may include—

(i) any background information relevant to the history and name of the geographic feature or Crown protected area that is named; and

(ii) information on the spatial extent of the geographic feature or Crown protected area that is named.

Access to an authoritative gazetteer of geographic names was a new requirement of the NZBGA, reflecting how important Gazetteers are in managing and disseminating geographic names. In addition, sections 32 and 33 of the NZGBA (compliance and enforcement) require Crown agencies to use official names in all official publications.

8.3 Why have a Gazetteer?

Gazetteers provide structure to the outcome of the geographical naming process, helping names become accepted, used, and to endure over time. A Gazetteer provides a single point of truth – an authoritative source where people can search for geographical names and their associated information. Including the physical extent of a named feature is also important to know exactly what the name applies to.

Geographical names are a fundamental dataset of geospatial information. They are one of the base layers for location reference and identification. Gazetteers must be:

- accessible,
- reliable,
- accurate,
- consistent, and
- standardised.

Government and public access to the data

A purpose of the NZGB's Gazetteer is to encourage the use of official names by government agencies and the private sector. There are a variety of ways to do this, but it needs to be as simple as possible for people to access the data and therefore increase use and awareness of official names across New Zealand.

Multiple access levels

Consideration was given to different ways to provide access to the information with three main areas identified:

- Simple web-based text searches and map views for casual queries by the public and non-technical users.
- Data downloads in spatial and non-spatial formats for data analysts.
- Live data services that allow machine to machine access for developers and advanced technical users.

8.4 The Case for a New Gazetteer

The NZGB Gazetteer data was formerly held in six Microsoft Excel spreadsheets published on the internet as CSV and PDF formatted documents (New Zealand, Offshore, Antarctica, Railway Lines, Crown Protected Areas and Undersea). Although this met the minimum legal requirements of the NZGBA, it was not a robust way to maintain a fundamental dataset of national significance. The spreadsheets were implemented as an interim solution to meet the legislative timeframe.

Spreadsheets are considered to be an inefficient and ineffective way to record geographical name information because they require significant manual manipulation to produce the output files and it is not possible to track the changes made to them.

Other issues with the former Gazetteer spreadsheets included:

- Making compliance with the NZGBA difficult for users especially government agencies.
- Users could not spatially search for official names.

- Difficult to show non-official names e.g. 'recorded' or 'historical' names especially for multiple names associated with one feature.
- Not best-practice by the geospatial community.

Official Name	Name Status	Feature	Statutory Reference	Land District
Abbess Peak	Approved	Peak	1925 (HON) p.1	Canterbury
Abbey Pass	Approved	Pass	1972 (18) p.379	Canterbury
Abbot Peak	Approved	Peak	1925 (HON) p.1	Canterbury
ABC Cave	Approved	Cave	1995 (69) p.1847	Otago
Abel Glacier	Approved	Glacier	1962 (73) p.2021	Westland
Abel Janszoon Glacier	Approved	Glacier	1982 (76) p.2163	Westland
Abel Lake	Approved	Lake	1987 (62) p.1964	Westland
Abel Tasman National Park	Approved	National Park	1943 (HON) p.1	Nelson
Abel Tasman Point	Approved	Point	1949 (22) p.858	Nelson
Abel Tasman Roadstead	Approved	Strait	1990 (191) p.4192	Marlborough

Figure 8-1 Extract from the Gazetteer spreadsheet

In December 2010, a 'Request for Proposal' (RFP) was issued for a service provider to design and build a new Gazetteer solution for the NZGB. An initial business case was developed in April 2011, being based on the successful bidder's proposal. This first business case was rejected on the grounds that the cost was too high. Therefore, a cheaper solution was sought. The project revisited its approach with a view to finding a lower cost solution that offered better value for the NZGB.

8.5 Solution Agreed

The solution was agreed in November 2011, leveraging off:

- the Land Information New Zealand (LINZ) Data Service,
- Datacom (IT service provider to LINZ) web development and infrastructure support, and
- a significant level of internal LINZ expertise.

In summary, the solution for the new Gazetteer database would have enhanced searching capability for users, easier administrative functionality, and connection to the LINZ Data Service, allowing automatic access to Gazetteer data via web services.

The solution would utilise existing infrastructure (no new hardware), leveraging off existing infrastructure support regimes, and taking advantage of open source software and the associated cost benefits.

8.6 Benefits

- Improves compliance with the NZGBA beyond the legal minimum achieved through the spreadsheets.
- Provides greater discoverability and access of the Gazetteer data to New Zealanders, especially government and other agencies required, under the NZGBA, to use official names in official documents.
- Enhances the value of the Gazetteer dataset by incorporating spatial information.
- Aligns with the general across-Government initiatives on dissemination and reuse of Government data.
- Adopts database functionality including better storage and maintenance – geographical names in single location and including background information.
- Enables the public to find out information about names themselves, especially those wishing to make name proposals.
- Assists the NZGB Secretariat to perform its own research and investigation e.g. duplication checks and historical research.
- Significantly reduces the resource requirement to publish updates to the spreadsheets Gazetteer.
- Reduces the resource requirement for annual reporting to the Minister.
- Contributes to a Ministerial priority to support Treaty of Waitangi Settlements.
- Advances good governance, and thus jurisdiction, over the Ross Sea Region of Antarctica.
- Emphasizes New Zealand's jurisdiction and governance over its continental shelf.
- Supports the United Nations Gazetteer Framework Project which aims to improve access to, and integration of, data held in different national gazetteers throughout the world.

8.7 Risks and Issues

The former spreadsheets based provision of the Gazetteer exhibited several notable limitations that impair compliance with the spirit of the NZGBA. Specific issues were:

- Users, notably government agencies, have a statutory requirement to use the official names of features and places in official documents. The former spreadsheets method of providing the Gazetteer made compliance with this requirement difficult and inefficient.
- The NZGB has a statutory requirement to collect information about official names. If data collection had been carried out before the design of the enhanced Gazetteer was completed, the collected data would have had to have been manipulated or translated or re-entered.

- Linear (rivers, ranges) or area (lakes, suburbs, forests) features cannot be effectively described by a point location. Lines and polygons (complex geometry) are better spatial descriptors. In the former spreadsheets, all positions were shown as points.

- Users could not spatially search for official names in the former spreadsheets. Also, with the former spreadsheets solution it was necessary for users to know the name of the feature in order to find it in the Gazetteer. It was not possible to determine 'what is the official name (if any) of the feature based on position'.

- Lack of standard audit control measures made it difficult to confirm the reliability of the authoritative record. Maintenance of the Gazetteer spreadsheets was a manual process that introduced transcription errors, which could potentially undermine confidence in the data.

- The use of spreadsheets to provide official data and information is not recommended best-practice by the geospatial community. It is the intention of all Australia New Zealand Land Information Council (ANZLIC) jurisdictions to move to Web Feature Services (WFS) and ANZLIC Standards, tying in with its Fundamental Spatial Data Framework (FSDS).

- Annual reporting on the NZGB activities for the Minister for Land Information was time consuming with the scattered distribution of data/information (both hard and soft copy)

8.8 Project Scope

The project scope included:

- Building a relevant database, populated with information held in the former Gazetteer spreadsheets and containing point, polygon and line geospatial reference data.

- Administrative back end allowing the NZGB Secretariat to maintain, administer, and report on, the Gazetteer database.

- Web based front end allowing textual and spatial queries of Gazetteer by users.

- Connection to the LINZ Data Service and development of processes supporting delivery of the Gazetteer through the Data Service.

- Development or update of relevant Secretariat business processes to support the new Gazetteer solution.

- Development or update of metadata to support the Gazetteer database.

- Migration of data from the existing Gazetteer spreadsheets to the new database.

8.9 Project Assumptions

The project assumptions were:

- The LINZ Project Management Framework (PMF) and the Information Technology System Development Life Cycle (SDLC) would be followed.

- Relevant NZGB Secretariat and LINZ staff would be available to undertake the project, as required.

- The Gazetteer database would be hosted on an existing server.

- The Gazetteer web application would be hosted on an existing server.

- The implemented system would be supported by LINZ and Datacom.

8.10 The Application Solution

The Gazetteer database was developed using PostgreSQL, a database management system that allows spatial objects (points, lines and polygons) to be stored directly in the database. A connection to the LINZ Data Service allows users to query the Gazetteer data against other datasets including Topographic and Hydrographic.

With a clear intention to design and build a new Gazetteer database with enhanced search capability to enable users to access the data (textual searches, spatial searches using a map interface, bulk data in several formats via the LINZ Data Service), this diagram represents the linkages and relationships with how the system operates:

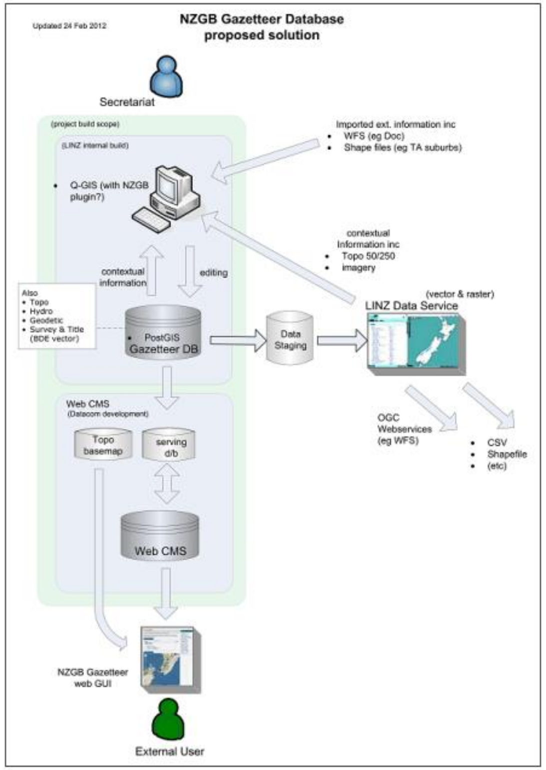


Figure 8-2 NZGB Gazetteer – proposed solution

Importantly, the diagram shows a system that provides separately for a simple enquiry in the WebApp from someone just wanting to get some information about a name, or that they want to know about names in a particular are of the country, and it provides for the GIS professionals to link in real time or obtain downloads from the LINZ data Service.

The GIS application used by the NZGB Secretariat is a package that is simple to use, is supported by LINZ, is open source (low cost), and will have widespread practitioner use.

8.11 Contextual Model

The NZGB Secretariat interacts with the Gazetteer through a purpose-built interface. This interface provides access to the data contained in the Gazetteer database and enables it to be viewed spatially in conjunction with other contextual data. This enables the NZGB Secretariat to create new records of place/features names, search for, query and update existing records, print/copy 'name' records and obtain reports from the database.

External users of the system interact with the Gazetteer through a purpose-built web interface. This interface provides 'read only' access to some of the data contained in the Gazetteer database. This enables the external user to search for and query existing records, and request output (printed/electronic) from the system.

External users are also able to download pre-compiled extracts of data from the LINZ Data Service web service. The LINZ Data Service provides web services to enable machine-to-machine connections to the database (eg. WFS, KML).

This model is a high-level view of that shows users interaction with the Gazetteer application.

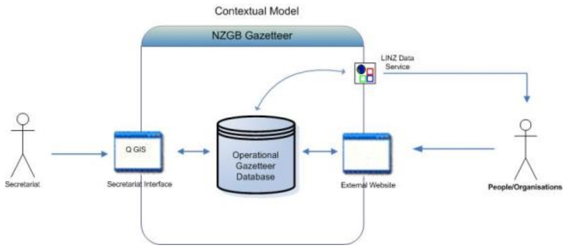


Figure 8-3 Gazetteer application interaction model

8.12 Value for Money

- Open source software – the Gazetteer was built using open source software meaning no or very low cost for application software. The applications used are:
  - PostgreSQL – database
  - Quantum GIS – NZGB Secretariat administration
  - Drupal – web user interface leveraging off LINZ web Customer Management System (CMS)
- LINZ Data Service – the LINZ Data Service offers minimal costs to providing Gazetteer data via web services e.g. WFS. Also, other LINZ Data Service functionality is available to users of the Gazetteer including:
  - Mashing Gazetteer data with other datasets
  - Downloading subsets of Gazetteer data with the LINZ Data Service cropping feature
- 'Internal' Development – the Gazetteer was built using 'internal' LINZ resources:
  - The user web interface was built by Datacom (service provider to LINZ)
  - The Gazetteer database and administration application were built within LINZ
- Shared infrastructure – the Gazetteer database utilises existing infrastructure:
  - Database – on existing Prod/Geo1 server
  - Web interface – on existing CMS server
- Internal support – the Gazetteer is supported by both Datacom and LINZ.

8.13 Data Migration

The Gazetteer data was converted from the set of six spreadsheets formerly maintained with layers in an ESRI database, exported as a set of shapefiles. The data was imported into a postgres database, processed and finally loaded into the Gazetteer administration tables. The migration process created a spreadsheet, error.xls, summarising data conversion issues and inconsistencies identified by the migration.

The basic steps of the migration were:

1) create schema, gazetteer import, for loading and processing the data

2) import spreadsheet of feature types. These were loaded from the spreadsheet 'Feature Type and Classification Summary.xls'. The script converted the list of feature types to unique four-character system codes.

3) Load the migration spreadsheet data. This involved first loading the '.status_mapping.xls' spreadsheet, which defined the mapping from status values in the data spreadsheets to those in the administration database, and then loads each spreadsheet in turn. All spreadsheets were loaded into a table called data. Each row was identified by a source spreadsheet id (e.g. NZON), and a line number in that spreadsheet. No data conversion was done during loading.

4) Load the GIS data into a table called 'gis'. This loaded the data without modification. It kept only the feature id, name, coordinate system, and spatial definition from the shapefiles. Other fields were ignored.

5) Load a spreadsheet 'superceded.xls' which was used to record features that should be merged, (where several names referred to the same feature, even though they may have had different feature ids in the source data).

All the source data was loaded into the postgres database. The migration was then performed by running a series of postgres scripts to process the data and load it into the tables in the Gazetteer schema.

The scripts ran in two stages, firstly a series of scripts to process the data into a normalized set of tables, including error checking, and then installing this data into the Gazetteer tables. The initial stage itself had two passes. In the first pass it processed features, geometries, and names. It then looked for names which should refer to the same feature id, and merged these feature ids. In the second pass it repeated the feature, geometry, name imports, and then also processed events and annotations.

At each stage errors/inconsistencies were reported into the error table. Finally, a python script extracted the error information into a summary spreadsheet for analysis and manual resolution by NZGB Secretariat staff. There was also a log file generated by the script which reported processing errors.

Improved location data

Before running the data migration, the NZGB Secretariat spent nine months cleansing location data for all the official names by adding line and polygon geometries where appropriate, for accurate spatial extent for each officially named feature.

8.14 Database Schema

The schema in figure 8-4 was developed and designed by LINZ. It meets OGC Standards for WFS-G, WMS and WMTS. Metadata is provided with the open layer on the LINZ Data Service. It is simple and uncomplicated, and has been designed to allow for additional fields to be added, as required. Annotations can be simply dropped into the Web Application.

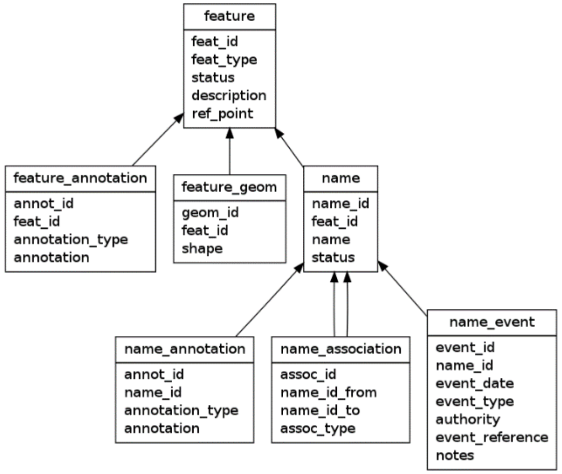


Figure 8-4 Database scheme

8.15 Timing

The estimated times for major phases of the project were:

Business case signed off	0 week
Planning and Design	+3 weeks
Build	+6 weeks
Implementation Phase	+2 weeks
User Acceptance Testing phase	+4 weeks

From when the revised Business Case was accepted in November 2011, it took 1½ years until going live with the new Gazetteer in June 2013.

8.16 Data Cleansing

Errors reported after the automated data migration phase were processed manually; They were investigated and resolved case by case.

In making the new Gazetteer service easy to use and accessible data errors and limitations across the three public access points (WebApp, CSV, LINZ Data Service) were exposed.

A significant number of official names with migration errors were updated just prior to the June 2013 launch. But there were several thousand unofficial names (mostly 'recorded') which required editing (investigation and cleansing).

A concerted effort was made to resolve these data issues in the Gazetteer from March to June 2014, to ensure users have confidence in the authority and accuracy of the Gazetteer, to ensure their on-going use and application.

8.17 Testing

Test phases for the system were undertaken in the latter half of 2012, with:

- User Acceptance Testing (UAT) Planning and Control
- Functional Testing
- System Testing
- Regression Testing
- UAT Execution

Test Plans were compiled largely based on the user requirements and the development changes.

8.18 Challenges and Constraints

Initially Open Street Mapping (OSM) was considered for the base layer, which is currently used for the LINZ Data Service. However, in rural areas, that product does not show sufficient geographical detail.

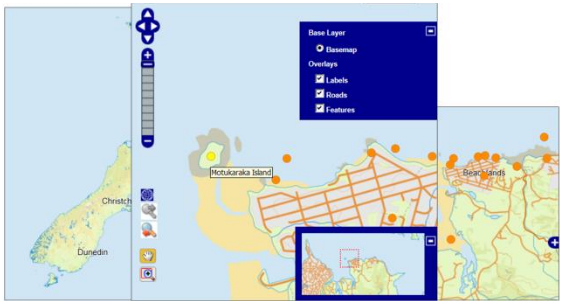


Figure 8-5 Base map. This tile was produced from several topo layers extracted from LINZ's 1:50K topo mapping. It includes 'geographical names', with the density controlled by zoom scale.

Another option was to use an open source product from Landcare NZ, who had developed a similar tile also based on LINZ's 1:50K topo mapping. But this product did not include the 'geographical names' layer, plus the soil layers detracted from interpreting geographic features.

In 2015, LINZ produced its own base mapping tiles, which will be added to the Gazetteer WebApp as options in a future enhancement.

Updates to the base layer tile will be managed through the service arrangement that LINZ has with Koordinates Enterprise.

As to the base mapping for the Ross Dependency of Antarctica and the continental shelf of New Zealand, the WebApp utilises KML links to Google Maps. A problem for the WebApp was in producing either one tile for all three areas, which would cause projection issues as well as size and browser time issues.

Quantum GIS (Q-GIS) will use LINZ's digital mapping for Antarctica, and geo-referenced bathymetric charts from National Institute for Water and Atmospheric Science (NIWA) for continental shelf areas. The administration of the Q-GIS Gazetteer by the Secretariat is separated for the three areas (New Zealand, Antarctica, Continental Shelf), without causing problems.

Requirements not met

- WebApp does not have a base map for Antarctica or the Continental Shelf of New Zealand (Google maps used)
- WebApp does not include advanced searches, e.g. searching for just official names. Rationale: cost and simplicity, noting that external users can download the whole dataset as a CSV file or use the open layers in their GIS applications
- WebApp does not have all geometries

Staff training and capability

- Data quality issues from current data formats
- Staff working on other priorities

Compromises had to be made as the project progressed, in order to retain and value for money solution.

The things that have had a significant effect were the need to keep the costs down and LINZ staff having other core work priorities.

While there are many examples of online Gazetteers and other mapping applications that we were able to draw on, the conditions of our user requirements and the collaborative approach has led to a unique outcome, tailored specifically for New Zealand.

8.19 Ongoing Support

The Gazetteer is supported internally within LINZ, providing support for the postgres Gazetteer database and NZGB Secretariat administration application (Q-GIS), and with Datacom providing infrastructure support within existing maintenance regimes for the web external user application.

8.20 Using Quantum GIS

- Starts with opening the plugin database and searching for a place name (figure 8-6).

Figure 8-6

- Wildcards work with at least three characters.
- The search returns a list of place names (figure 8-7).

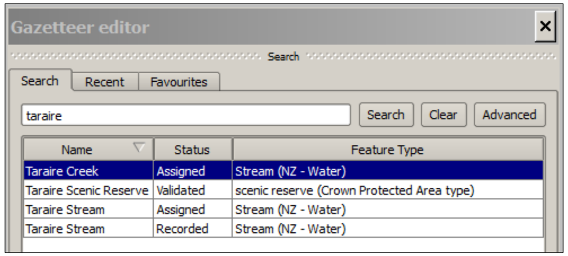


Figure 8-7

- The spatial window (figure 8-8) takes you to the feature/name

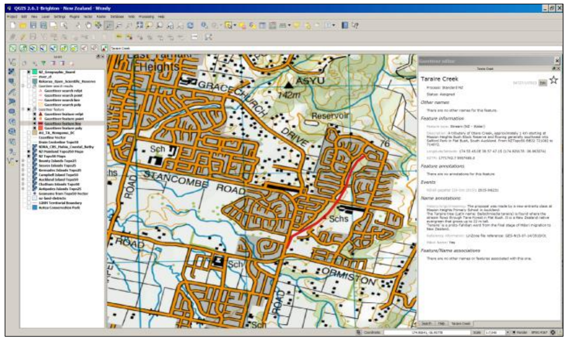


Figure 8-8 Spatial window displaying the named object

- The layer list (figure 8-9) shows the usual GIS layers including any that have been imported or WFS'd or WFTM'd

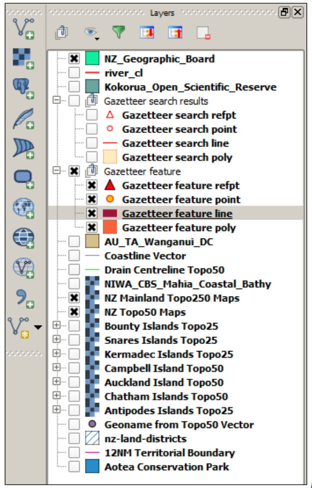
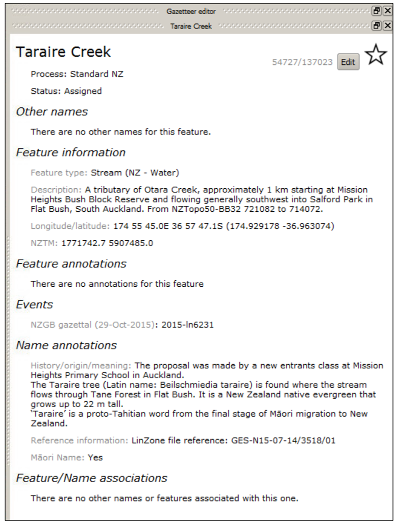


Figure 8-9

- The Name Record (figure 8-10) presents the fields of information that has been captured. Annotations are used to capture the relevant information.


Gazetteer editor
Taraire Creek

Taraire Creek

54727/137023 Edit

Process: Standard NZ

Status: Assigned

Other names

There are no other names for this feature.

Feature information

Feature type: Stream (NZ - Water)

Description: A tributary of Otara Creek, approximately 1 km starting at Mission Heights Bush Block Reserve and flowing generally southwest into Salford Park in Flat Bush, South Auckland. From NZTopo50-BB32 721082 to 714072.

Longitude/latitude: 174 55 45.0E 36 57 47.1S (174.929178 -36.963074)

NZTM: 1771742.7 5907485.0

Feature annotations

There are no annotations for this feature

Events

NZGB gazettal (29-Oct-2015): 2015-ln6231

Name annotations

History/origin/meaning: The proposal was made by a new entrants class at Mission Heights Primary School in Auckland.
The Taraire tree (Latin name: Beilschmiedia taraire) is found where the stream flows through Tane Forest in Flat Bush. It is a New Zealand native evergreen that grows up to 22 m tall.
'Taraire' is a proto-Tahitian word from the final stage of Māori migration to New Zealand.

Reference information: LinZone file reference: GES-N15-07-14/3518/01

Māori Name: Yes

Feature/Name associations

There are no other names or features associated with this one.

Fig.8-10

- Publishing from Q-GIS (figure 8-11) sends the latest information out to the WebApp, creates the CSV file, and updates the data layer ‘NZ Place Names (NZGB)’ in LDS.

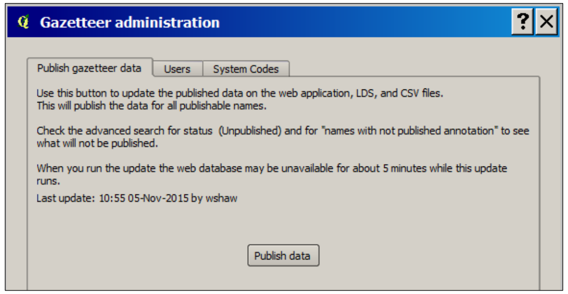


Figure 8-11

8.21 Using the Web Application

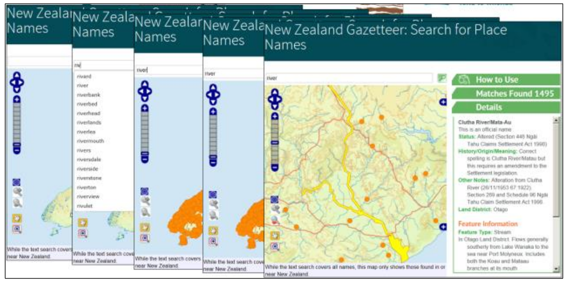


Figure 8-12

- Launched June 2013, minor enhancements May 2015
- A brief pilot for external users was undertaken
- Text search, interaction between the spatial view and the Matches Found / Details views, type of information included in the panel views, and mapping tool icons.
- The standard LINZ border information is in Drupal.

Analytics

- In the first half of 2015 the WebApp had about 4600 pageviews with an average of 4.5 minutes each.

Where do I find the Gazetteer

http://www.linz.govt.nz/regulatory/place-names/find-name/new-zealand-gazetteer-official-geographic-names/new-zealand-gazetteer-search-place-names#zoom=7.6133333333278195&lat=-35.27323&lon=174.12763&layers=BTTT

Or navigate from the LINZ home page (figure 8-13):

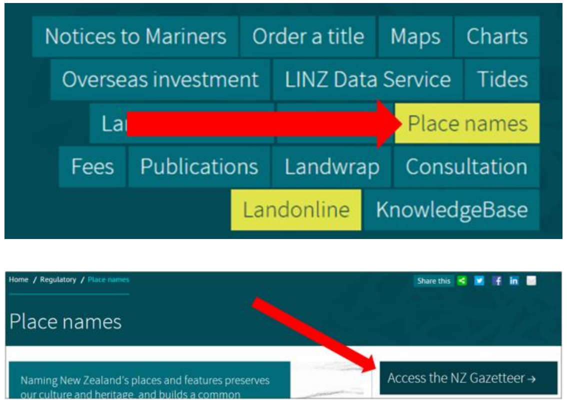
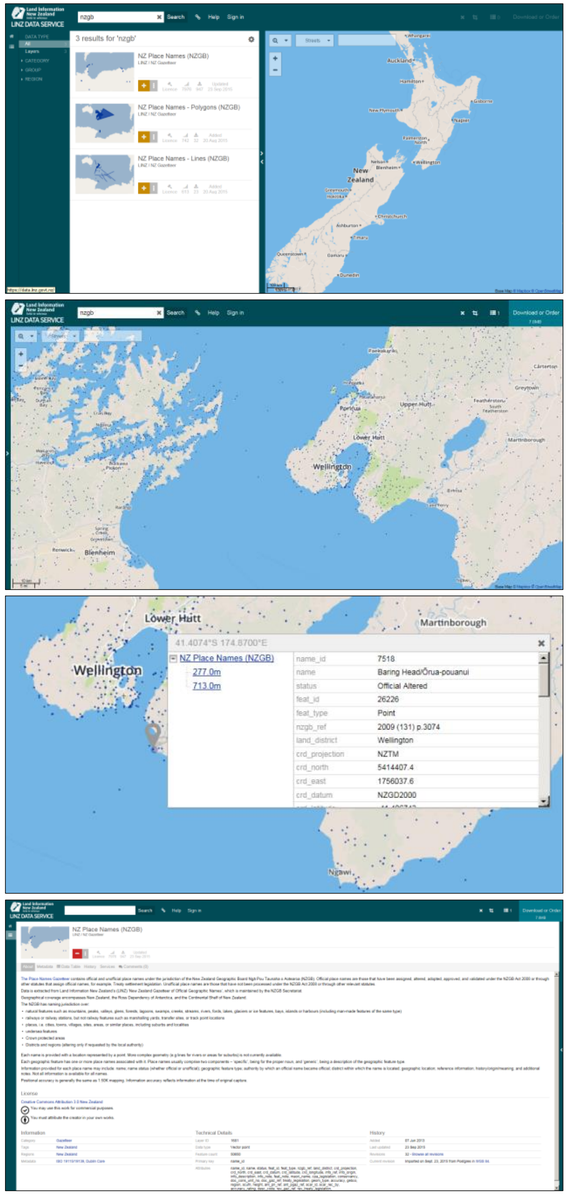


Figure 8-13 LINZ home page

8.22 Using the LINZ Data Service



Open layers of the Gazetteer have been added to the LINZ Data Service, which currently has over 2000 layers of topographic, hydrographic, geodetic, cadastral and title data via the web interface, all freely available, for downloading or linking to real time.

With simple registration procedures, an allocated API (Application Program Interface) allows users to tailor the layers they want into their own GIS applications, i.e. projections, formats. For more information see http://www.linz.govt.nz/data/linz-data-service

Usage

Use of NZ Gazetteer data has increased marginally over the past 6 months. This is reflected in the average monthly downloads of 36 to 30 June 2015, compared to 31 for the previous six months.

	Jan-Jun 2015	% of LDS
Total downloads	218	0.3%
Download size	45 GB	0.3%
Web service request count*	63	0.004%
Web service volume**	813 MB	0.1%
Unique users downloading	177	1.9%

* WFS and Spatial and table queries by API ** WFS only

Top users of Gazetteer data by download and web service count hail from the following sectors:

- Central Government – Ministry of Justice and New Zealand Transport Agency (NZTA)
- GIS – Critchlow & Eagle Technology
- Web – Topomap

8.23 Outcomes for External Users

The focus has been on the delivery for customers and stakeholders, which has concentrated our efforts on the Web Application, however the NZGB's Gazetteer remains the administrative tool that feeds both the Web Application, creates a CSV download and provides an open layer in the LINZ Data Service.

One of the key benefits for the Secretariat in providing the new Web Application, is that people can see for themselves information about names and where they are located on a map.

Other customer advantages:

- Text searches can include macrons
- CSV download of full dataset
- Other names associated with a feature
- Authoritative for official names and comprehensive source of unofficial names
- Positions more accurately on feature and extent is defined
- Extent of NZGB naming jurisdiction is known
- Accurate, knowledgeable, consistent, sustained, reliable, standardized, transparent, enduring, agreed, official

8.24 What's next?

- Enhancements:
  - Options for base maps: NZTopo50, NZTopo250, Parcel, Cartographic render
  - Extend coverage over Antarctica and Continental Shelf
  - Add geometry layers for lines and polygons
  - Improve the user experience in the spatial navigation
  - Provide for more complex searching, eg find all features of type 'hill' that are 'official'
- Continue data cleansing
- Ongoing updates and improvements
- Promote to the public and geospatial community
- LINZ to investigate a Gazetteer Service to combine each of its 'Names' datasets (Electoral, Topographic, Hydrographic, NZGB).

Section 4 Regional and global initiatives

Chapter 9 **UNGEGN World Geographical Names Database**

Helen Kerfoot

9.1 Introduction

The UNGEGN World Geographical Names Database was initiated in 2004 and has continued with the support of resolution IX/6 of the Ninth UN Conference on the Standardization of Geographical Names in 2007.

It is a multilingual, multi-scriptual geo-referenced database containing names of UN member states, capitals, and cities/towns with a population over 100,000. All entries provide endonyms, as well as forms used by the United Nations in Arabic, Chinese, English, French, Russian and Spanish for the countries (UN Member States) and capitals. The data, now uploaded quarterly, is accessible on the UNGEGN website at http://unstats.un.org/unsd/geoinfo/geonames/ through a world map interface and tables (Figure 9-1).
The UNGEGN Secretariat is responsible for maintaining the database and development of the web interface.

9.2 History behind the development of the database

At its twenty-second session in 2004, UNGEGN recommended that the Secretariat take the lead in developing a world database to collect, manage and disseminate authoritative data on country and major city names. In particular this would use the UNGEGN website to make available information that would help respond to toponymic questions received by the Secretariat and would provide a vehicle for countries to have their city names displayed in standardized form within a worldwide framework.

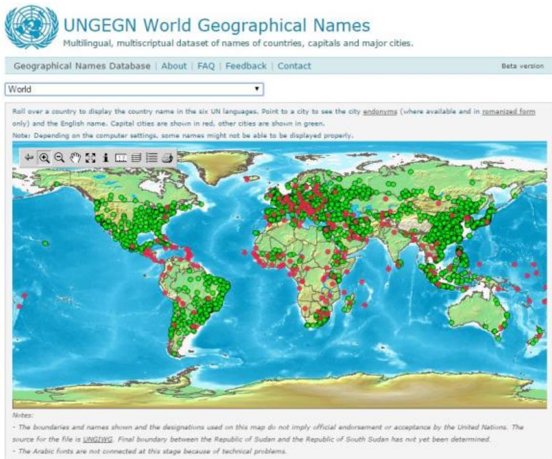


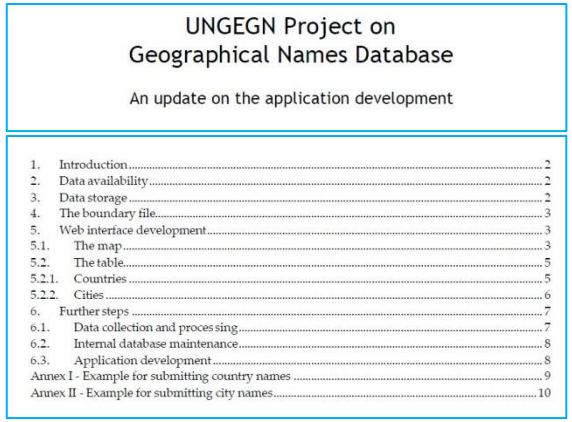
Figure 9-1 World map; starting point for searching the UNGEGN World Geographical Names Database

UNGEGN experts are responsible for supplying (or updating) the city/town data from their countries together with the recognized coordinates of latitude and longitude. In addition, experts are encouraged to supply audio files for the pronunciation of each city name; these are attached to the individual entries and are available to web users.

As a result, the Secretariat with advice from UNGEGN initiated the process of building a multilingual, multi-scriptual geo-referenced database, designed to represent the reality of geographical names in a variety of languages and scripts. The database had to be available to UNGEGN experts and the general public through a web interface. Names for places would be linked to a map, and standardized names, their spelling and pronunciation, would be displayed as tables.

At the time, the database was created in SQL Server 2005 which could store all the information necessary for populating the map and providing data in tabular format (including city and country names, ISO 3-letter country and language codes, variants, coordinates, comments and pronunciation audio files).

Following a special presentation to the Ninth UN Conference on the Standardization of Geographical Names in 2007 (Figure 9-2), the Conference passed resolution IX/6, recommending that the UN Statistics Division, in cooperation with the UN Cartographic Section, the UN Second Administrative Level Boundaries (SALB), UNGEGN and member States "further develop, populate and maintain the geographical names database" of UNGEGN, "initially containing names of countries, capitals and major cities".


UNGEGN Project on Geographical Names Database

An update on the application development

1. Introduction ... 2
2. Data availability ... 2
3. Data storage ... 2
4. The boundary file ... 3
5. Web interface development ... 3
5.1. The map ... 3
5.2. The table ... 5
5.2.1. Countries ... 5
5.2.2. Cities ... 6
6. Further steps ... 7
6.1. Data collection and processing ... 7
6.2. Internal database maintenance ... 8
6.3. Application development ... 8
Annex I - Example for submitting country names ... 9
Annex II - Example for submitting city names ... 10

Figure 9-2 Special presentation available in text and slides at http://unstats.un.org/unsd/geoinfo/UNGEGN/ungegnConf9Add.html

9.3 Geographical names data included

So far, the data includes:

(1) Country names - formal and short forms
    a. In the language(s) and writing system(s) of the UN member state itself (source: UNGEGN Working Group on Country Names)
    b. As used by the UN in Arabic, Chinese, English, French, Russian, and Spanish (source: UN Term database)
(2) Capital cities
    a. In the language(s) and writing system(s) of the UN member state
    b. As used by the UN in Arabic, Chinese, English, French, Russian, and Spanish (source: UN Term database)
(3) Cities/towns with a population over 100,000
    a. Names (endonyms) as supplied by each UN member state in its own language(s) and writing system(s)
    b. Romanized forms of the city/town names (where possible through systems recommended through UN resolutions)

For each country or city name stored, coordinates of latitude and longitude are indicated, the language in which the name is used is supplied, the data source is noted, and audio files for pronunciation can be included if they are provided by the UN member state.

Convenors of the UNGEGN Working Group on Romanization Systems and the Working Group on Country Names help in verifying data before it is loaded into the database. In addition, the Working Group on Toponymic Data Files and Gazetteers has provided advice during the development stages.

To date (June 2015) the following countries have supplied city data:

A. City/town data sets - some with updates (see Figure 9-3):

Argentina, Australia, Austria, Belarus, Belgium, Botswana, Brazil, Bulgaria, Burkina Faso, Cameroon, Canada, Chile, China, Croatia, Cuba, Cyprus, Czech Republic, Denmark, Egypt, Estonia, Ethiopia, Finland, France, Gambia, Germany, Greece, Hungary, Iceland, Indonesia, Iran (Islamic Republic of), Ireland, Israel, Italy, Japan, Kenya, Kyrgyzstan, Latvia, Lithuania, Madagascar, Malaysia, Mali, Mexico, Nepal, Netherlands, New Zealand, Niger, Norway, Philippines, Poland, Republic of Korea, Romania, Russian Federation, Saudi Arabia, Serbia, Slovakia, Slovenia, South Africa, Spain, Sri Lanka, Sweden, Switzerland, Tajikistan, The former Yugoslav Republic of Macedonia, Tunisia, Turkey, Ukraine, United Arab Emirates, United Kingdom, United States of America, Uzbekistan, Viet Nam

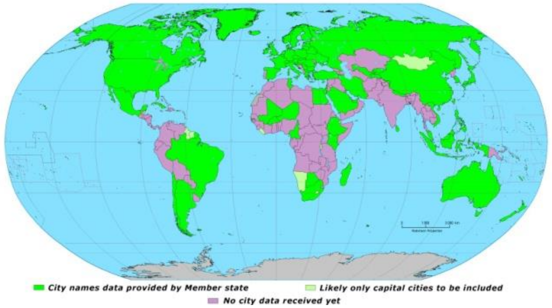


Figure 9-3 UNGEGN World Geographical Names Database – data provided for population centres over 100,000

B. Audio files for pronunciation of city/town names (see Figure 9-4):

Austria, Belgium, Brazil, Bulgaria, Burkina Faso, Canada, Croatia, Cyprus, Czech Republic, Denmark, Egypt, Finland, France, Gambia, Germany, Hungary, Iceland, Ireland, Israel, Latvia, Madagascar, Netherlands, New Zealand, Norway, Philippines, Poland, Republic of Korea, Romania, Saudi Arabia, Serbia, Slovenia, Spain, Sweden, Tunisia, Ukraine

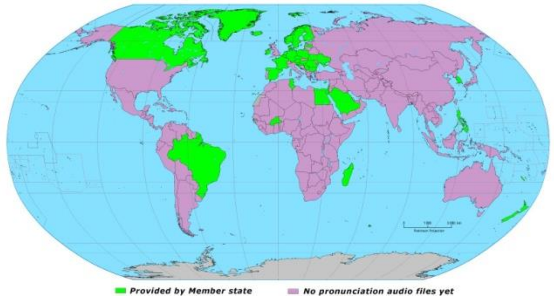


Figure 9-4 UNGEGN World Geographical Names Database – pronunciation audio files provided

The current statistics of available data for UN member states (as of September 30, 2015) are as follows:

- 5947 name records[1]
    - 193 countries
        - 1304 country names (including names in the six official UN languages from UN Termium)
            - 273 endonyms
            - 1031 exonyms
    - 3393 cities

[1] If both short and formal country names and romanized forms of country and city names are counted as separate items, the total count of names would be 8884.

- 4643 city names
    - 3437 endonyms
    - 1206 variants

- Romanization: 50 systems ... 1587 romanized forms of names
- Sound files: 1007 cities from 35 countries
- Languages: 116 (with Chinese- and English-language forms of names having the highest counts)

9.4 Boundaries

According to Sibylle Marxgut in her presentation to the Ninth Conference in 2007, the shapefile used is sourced from the UN Geographical Information Working Group (UNGIWG) 1:1 million dataset reflecting cartographic practices of the United Nations.

9.5 Some technical aspects of the data storage and web interface[2]

A Web Map Server (WMS) was set up to show the names on a world map. The WMS protocol is based on simple query syntax for posting a request for the desired layers and zoom window to the server, which returns a map as a standard picture. It has been developed in compliance with standards and protocols of the Open GIS Consortium.

The web interface was developed in ASP Net 2.0 and Macromedia Flash 8 to provide a map application and table format.

So far, the user interface has been developed in English, although the data is multilingual.

[2] From information from Sibylle Marxgut's presentations to the Ninth Conference in 2007 and to UNGEGN in 2009.

9.6 Web access

With the initial log-in, the user sees the world map of UN member states, displaying the location of capitals (red), and other cities (green) that have been uploaded with data received from the UN member states (as shown in Figure 9-1). A tool bar provides map zooming and dragging capabilities, selection/de-selection of map layers, a distance measuring tool and printing icon.

Also from this view, a small amount of background information can be accessed above the map: a statement about the database, FAQs, Feedback and Contacts.

Queries can be made via the map interface or by using the pull-down menu.

(1) Country names (Figures 9-5, 9-6, 9-7, 9-8).

Figure 9-5 Rolling over a country will show the short form of the name of the country in the six languages of the United Nations

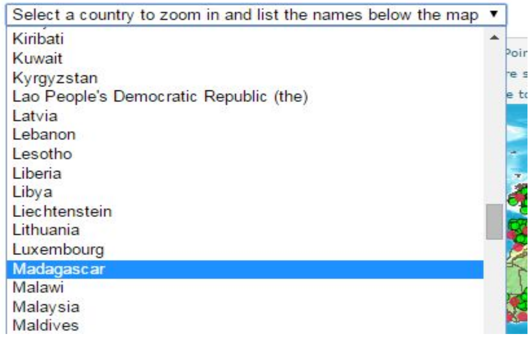


Figure 9-6 Individual countries can be selected from the pull-down menu

When a country is selected from the pull-down menu a map showing the country will be displayed (example in Figure 9-7) and below will appear tabular data (example in Figure 9-8) showing information on the short and formal names for the country:

- In the language(s) of the country, both in the writing system of the language and in the Roman alphabet, and
- In the United Nations languages (Arabic, Chinese, English, French, Russian and Spanish)

Figure 9-7 Example of country map: Greece

Country names

Endonyms

Language	Short name	Formal name
Greek	Ελλάς *Ellás*	Ελληνική Δημοκρατία *Ellinikí Dimokratía*

United Nations languages

Language	Short name	Formal name
Arabic	اليونان	الجمهورية الهيلينية
Chinese	希腊	希腊共和国
English	Greece	the Hellenic Republic
French	Grèce (la)	la République hellénique
Russian	Греция	Греческая Республика
Spanish	Grecia	la República Helénica

Figure 9-8 Tabular data providing short and formal country names: Greece

(2) Names of capital cities

For names of capitals, rolling over the red circles displays the endonym(s) (romanized) and the name as used in English (see Figure 9-9)

Figure 9-9 Colombo, Sri Lanka, shown in Sinhalese, Tamil and English

Tabular data for capital cities (example in Figure 9-10) shows:

- The capital city endonym(s), with romanized form where applicable
- The name as used in the United Nations languages (Arabic, Chinese, English, French, Russian and Spanish)
- The latitude and longitude in degrees and decimal degrees
- The data source
- Audio pronunciation files, if provided by the UN member state; they are attached to the name and can be heard by clicking on the audio icon.

Capital and other major cities

Capital city

Lat: 6.93 | Long: 79.84 කොළඹ - Kōḷamba

Endonym

Language	City name	Source
Sinhalese	කොළඹ *Kōḷamba*	UNGEGN
Tamil	கொழும்பு *Kōḷumpu*	UNGEGN

United Nations Languages

Language	City name	Source
Arabic	كولومبو	UN terms
Chinese	科伦坡	UN terms
English	Colombo	UN terms
French	Colombo	UN terms
Russian	Коломбо	UN terms
Spanish	Colombo	UN terms

Variant name

Language	City name	Source
English	Colombo	UNGEGN

Figure 9-10 Tabular data for the capital city of Sri Lanka

(3) Names of major cities (population over 100,000)

Cities with a population over 100,000 (other than capital cities) are shown with green location symbols on the map and their names can be highlighted in the romanized form(s) of the endonym(s). (See Figure 9-11.)

Figure 9-11 A city in Cyprus with its name shown as endonyms: in Greek (romanized) and in Turkish.

Cities are displayed as geo-referenced endonyms in the tables. Audio files are included, if provided by the UN member state, and are accessed through the audio symbol against the city name in the tables (Figure 9-12).

Other major cities

Lat: 34.68 | Long: 33.03 Λεμεσός - Lemesós

Endonym

Language	City name	Source
Greek	Λεμεσός *Lemesós*	UNGEGN
Turkish	Leymosun	UNGEGN

Figure 9-12 Tabular data for a city in Cyprus, including the icons to click for audio pronunciation files

9.7 Current work

The UNGEGN Secretariat continues to be responsible for the design and development of the database, as well as for entry of data that has been supplied by UN member states through UNGEGN experts, or comes from UN Termium.

The website is now updated four times a year to reflect the information loaded on an ongoing basis into the UNGEGN database.

Since 2012, modifications have been made to improve the representation of a number of map views at the country level (for example, for some island countries).

Although we would like to expand the content coverage of the UNGEGN World Geographical Names Database (for instance to include more variant names or to include major physical feature names), for the time being the UNGEGN Secretariat and UNGEGN experts will continue to focus on fundamental aspects of the database.

The main requirements at present are:

- City data (in an Excel table with coordinates) from a number of UN member states. As well, updates are welcomed.
- Audio files (preferably .wav files) from member states to be attached to the city, capital and country name records to enhance users' understanding of the endonyms.
- Changes in names of UN member states as used by the United Nations and stored in the UN Termium database.
- Updating of boundaries as new UN member states come into existence.
- Expanding FAQs, upgrading the web interface and character representation, to be undertaken as appropriate.

9.8 References and further information

This article is based on Working Paper 56 presented at the 28th UNGEGN Session in 2014.
More information and progress reports on the UNGEGN World Geographical Names Database are available:

(1) Special presentations:

a. 2007: Ninth UN Conference "UNGEGN project on geographical names database", http://unstats.un.org/unsd/geoinfo/UNGEGN/docs/special-presentations-docs/geonamesdatabase.pdf

b. 2009: 25th UNGEGN Session "The UNGEGN Geographical Names Database", http://unstats.un.org/unsd/geoinfo/UNGEGN/docs/25th-gegn-docs/presentations/ungegn%20db%20project_2009-04-24.pdf

(2) Included in Secretariat reports to UNGEGN (http://unstats.un.org/unsd/geoinfo/UNGEGN/sessions.html)

a.	2004	GEGN/22/9	last page
b.	2006	GEGN/23/11	section 3
c.	2009	GEGN/25/11	section on Database
d.	2011	GEGN/26/11	section on Database

(3) Included in UNGEGN Bulletins (http://unstats.un.org/unsd/geoinfo/UNGEGN/bulletin.html)

a.	#35	August 2008	p. 7-8
b.	#36	March 2008	p. 6
c.	#37	Sept. 2009	p. 6-7
d.	#38	March 2010	p. 8
e.	#39	Fall 2010	p. 6

Section 4 Regional and global initiatives

Chapter 10 – **Working with geonames.org**

Marc Wick

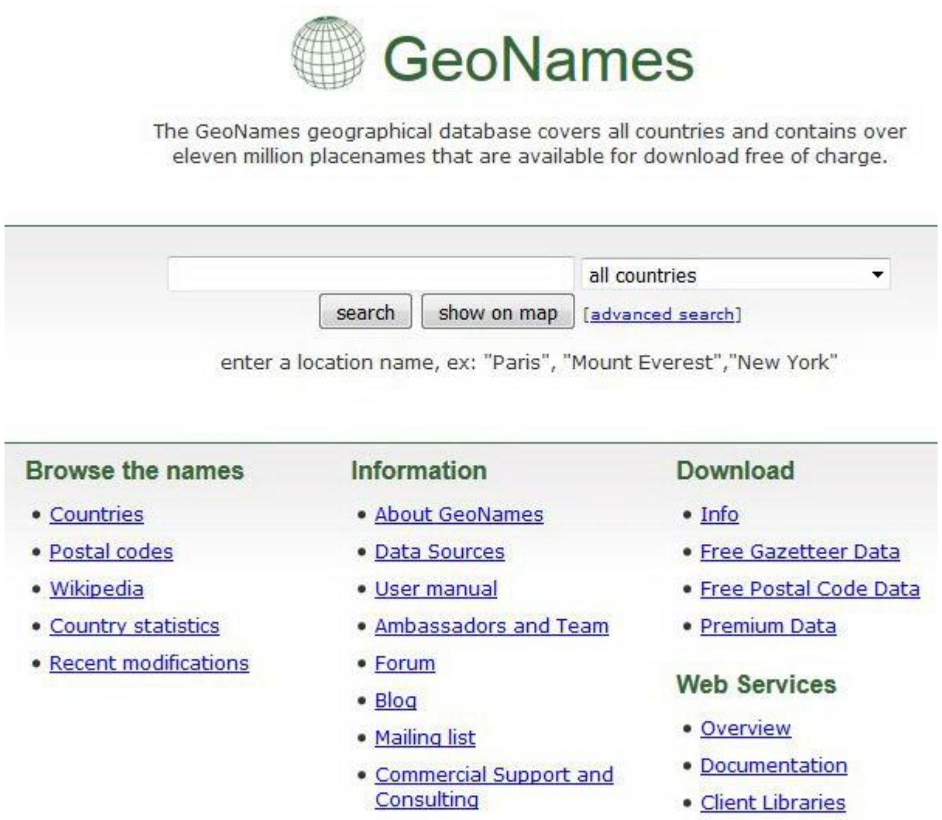


Figure 10-1 - GeoNames application

10.1 Introduction

This chapter will give an overview of the geonames.org gazetteer and how to work with it. We will introduce the data model and the classification system. We will describe the extract files, which can be downloaded to have a full copy of the gazetteer locally and also the web services, which can be used to build an application with the geonames.org ([1]) data set without having to download the data. We will list the most important sources and how data can be contributed to the gazetteer either in making edits directly with the wiki interface or with providing data sets to be imported into GeoNames.

10.2 What is GeoNames?

GeoNames is an open global gazetteer, a database with records of geographical features. At the time of writing (Summer 2016) it contains over 11 million features with additionally 11 million alternate names. The gazetteer data is downloadable in tab separated csv files in utf8 encoding under a liberal Creative Commons Attribution (cc-by) license. The only condition is to somehow give credit to GeoNames and in turn to the sources. It is left to the user how the attribution is implemented, it can be a link on a website, a phrase in the documentation or another form. GeoNames is aggregating data from many sources, the most important of which is the United States National Geospatial-Intelligence Agency. Other important sources are the national mapping agencies or the national statistical offices of all countries whenever they publish data compatible with the cc-by license. A wiki interface on the www.geonames.org website allows users to browse the data and to quickly add features, attribute values or correct errors. GeoNames is having over 160 000 users many of which contribute with a total of several hundreds of edits every day. Nevertheless, the main data volume are batch imports from the national mapping agencies. A quality assurance process in a monthly release cycle ensures the data quality and integrity, in particular with the wiki interface where everybody can contribute. Last but not least GeoNames offers an api with nearly 40 web services. The web services give direct programmatic access to the gazetteer data set. On the other hand, many services used internally to enhance the data when adding to the database are also made publicly available as a web service. Examples are various digital elevation services or time zone reverse lookup.

10.3 Data model

The data model of GeoNames is quite simple and consists of two main tables. A ***GeoName*** table with the feature attributes and a second table ***AlternateName*** with translations of the feature name. Each feature has a main toponym name, which is either in English or an internationally understood name. Names in other languages are found in the alternatename table. The language is identified with the ISO language code [3].

GeoName – Feature Attributes

The most important table is the GeoName table. It contains the basic attributes of a geographic feature. The geonameId as the primary key is a sequence number and never changes. When inserting a new feature, the sequence is incremented and the new feature gets the next integer number. Each feature is represented by a lat/lng coordinate pair. The location of the lat/lng point is indicating the location on the globe, it may be the geometric centre, the location of the administration or some other position. The type of feature is described with a two-level classification system. The featureClass is a first rough classification and puts the features into one of nine classes. The nature of the feature is described in more detail with the assignment of one of the 660 feature codes on the second classification level.

GeoName	
geonameId	integer id of record in GeoNames database
name	name of geographical point (utf8) varchar(200), English or internationally understood name.
asciiname	name of geographical point in plain ascii characters, varchar (200)
alternatenames	alternatenames, comma separated, ascii names automatically transliterated, convenience attribute from alternatename table, varchar (10000)
latitude	latitude in decimal degrees (wgs84)
longitude	longitude in decimal degrees (wgs84)
feature class	char (1)
feature code	varchar (10)
country code	ISO-3166 2-letter country code, 2 characters [3]
cc2	alternate country codes, comma separated, ISO-3166 2-letter country code, 200 characters
admin1 code	fipscode, see exceptions below, see file admin1Codes.txt for display names of this code; varchar(20)
admin2 code	code for the second administrative division, see file admin2Codes.txt; varchar(80)
admin3 code	code for third level administrative division, varchar (20)
admin4 code	code for fourth level administrative division, varchar (20)
population	bigint (8 byte int)
elevation	in meters, integer
dem	digital elevation model, srtm3 or gtopo30, average elevation of 3''x3'' (ca 90mx90m) or 30''x30'' (ca 900mx900m) area in meters, integer. srtm processed by cgiar/ciat.
Time zone	the iana [2] time zone id (see file timeZone.txt) varchar(40)
modification date	date of last modification in yyyy-MM-dd format

Table 10-1 – GeoName table

AlternateName – name in other languages

A feature may have more than just a single name. It may have name variants in other languages, it may have had other names in the past, it may have short and long names and it may even be known with colloquial names. These names are modelled in the second table, the *AlternateName* table.

In order to identify the language an alternate name stands for the ISO 639 language code is used [3]. Furthermore, a couple of pseudo language codes describe other names like: '*post*' for postal codes, *'iata'*, *'icao'* and *'faac'* for the respective airport codes. *'link'* stands for an url pointing to a website. The most often used links are to the corresponding Wikipedia article. Over 500000 alternatenames are links to Wikipedia pages.

Four flags help to further describe an alternate name: short, preferred, historic or colloquial. *'Big Apple'* is a

AlternateName	
alternateNameId	the id of this alternate name, int
geonameId	geonameId referring to geonameid in table 'GeoName', int
isolanguage	iso 639 language code 2- or 3-characters; 4-characters 'post' for postal codes and 'iata', 'icao' and 'faac' for airport codes, fr_1793 for French Revolution names, 'abbr' for abbreviation, 'link' for a website, varchar(7)
alternate name	alternate name or name variant, varchar(400)
isPreferredName	'1', if this alternate name is an official/preferred name
isShortName	'1', if this is a short name like 'California' for 'State of California'
isColloquial	'1', if this alternate name is a colloquial or slang term
isHistoric	'1', if this alternate name is historic and was used in the past

Table 10-2 – AlternateName table

colloquial name for *New York City*, whereas *'Karl-Marx-Stadt'* is a historic name for *Chemnitz*.

Hierarchy - Administrative Hierarchy

The administrative division a feature belongs to is modelled with the attributes countryCode and adminCode1 to adminCode5. The countryCode attribute is the two character ISO country code and contains the countryCode of the country the feature belongs to. The adminCode1 point to the first order administrative division of the same country. The combination of countryCode, adminCode1 and adminCode2 gives you the second order administrative division. The administrative divisions are like all features part of the GeoName table, differentiated by their featureCode.

The hierarchical structure modelled with the adminCodes is treelike, every feature can have only one direct parent. For administrative hierarchies, this is normally sufficient.

For super-national features, a second countryCode attribute contains a comma separated list of all countries with a relation to the feature. The same for border features (mountains, lakes), which belong to more than one country.

Example for administrative hierarchy: Rome (geoNameId 3169070), the capital of Italy, has countryCode *'IT'* pointing to Italy (geoNameId 3175395), the adminCode1 is *'07'*, which points to the region *'Lazio'*. The second order administrative division is the province Rome with adminCode2 'RM'. On the third level, we have adminCode3 '058091' for the Commune (municipality) of Rome. The adminCode3 ' 058091' is the code used by the *Italian National Institute of Statistics [4]* for the third order administrative divisions.

Non-Administrative Hierarchy

The admin hierarchy cannot model all hierarchy types. An additional table contains relations between features outside of the administrative hierarchy. These relations are for instance relations between spot features and populated places, between neighbourhoods and cities or also regions consisting of administrative divisions. Example for tourism region: Spain has defined a couple of tourism regions which are defined as a group of municipalities. The Costa Brava (geoNameId 3127668) is made up of 26 municipalities.

10.4 Feature Classes and Feature Codes

FeatureClass	
A	Administrative features: country, state, region, ...
H	Hydrographic features: stream, lake, ...
L	Area features: parks, area, ...
P	Populated places features: city, village, ...
R	Road/Railroad features: road, railroad
S	Spot features: spot, building, farm
T	Hypsographic features: mountain, hill, rock, ...
U	Undersea features
V	Vegetation features: forest, heath, ...

Table 10-3 – Feature Classes

The *featureClass* is a rough categorization further enhanced by the *featureCode* which describes the feature in more detail. Each feature may belong to one of 660 feature codes. We distinguish populated places by size and function, whether a populated place serves as a seat of an administrative division or even as a country capital. Similarly, we separate protuberances into hill, mountain, peak, range, rock, pass or another of 99 feature codes.

FeatureClass			
		Number of Feature-Codes	**Number of Features**
A	Administrative features	24	357,767
H	Hydrographic features	134	2,134,794
L	Area features	49	379,109
P	Populated places features	18	4,349,577
R	Road/Railroad features	21	40,356
S	Spot features	244	2,276,788
T	Hypsographic features	99	1,516,560
U	Undersea features	63	14,476
V	Vegetation features	17	39,478

Table 10-4 – Feature Classes and its Feature Codes

Each feature is unique, there is only one entry in the GeoName table for a feature. Important to understand in this respect is that administrative divisions and populated places are considered two different concepts, each requiring its own feature in the gazetteer. Cities therefore often have two entries, one for the populated place and the other to represent the administrative division (municipality, commune, etc).

Most often used Feature Codes			
Number of Features	**Description**	**Feature Code**	**Feature Class**
3,982,992	populate place	PPL	P
856,325	stream	STM	H
382,137	mountain	MT	T
359,720	hill	HLL	T
320,048	farm	FRM	S
276,737	school	SCH	S
262,085	lake	LK	H
245,001	church	CH	S
241,238	hotel	HTL	S
194,191	intermittent stream	STMI	H

Table 10-5 – Most often used Feature Codes

10.5 Download – Extract Files

The GeoNames data is exported daily into a download directory where it can be downloaded for free. A username/password is not required for the data download.

The data model with the two main tables GeoName and AlternateName is reflected in the extract files. The main information is included in the two files allCountries.txt and alternatenames.txt, with the former being the export of the GeoName table and the later the export of the AlternateName table. The hierarchy information is found in the hierarchy.txt file.

The other files in the download directory are reference and convenience files.

10.6 Reference Files

A couple of files in the GeoNames download directory are reference data. These are lookup files for codes used by GeoNames. GeoNames is using the ISO 639 languages codes in the alternatename table to identify the language of a name variant. The timezone of a feature is identified by the iana timezoneId (see also the section 'TimeZone') [2]. The feature codes used by GeoNames are described in a handful of languages. Attributes specific to countries and therefore not part of the data-model for geoname

Reference Files	
iso-languagecodes.txt	iso 639 language codes, as used for alternate names in file alternateNames.zip
timeZones.txt	countryCode, timezoneId, gmt offset on 1st of January, dst offset to gmt on 1st of July (of the current year), rawOffset without DST
featureCodes_<lang>.txt	name and description for feature classes and feature codes in a couple of languages (bg, en, nb, nn, no, ru, sv)
countryInfo.txt	country information: iso codes, fips codes, languages, capital, ...

Table 10-6 – Reference Files

features are found in a separate countryInfo file. It contains various country codes, the languages spoken in the country, postal code format, internet top level domain and neighbouring countries.

Convenience Files

The two main files allCountries.txt and alternatenames.txt are quite big. The allCountries file is uncompressed about 1.3 GB large. For users who are only interested in a subset of the data a couple of convenience files are available.

Convenience Files	
XX.zip	features for country with iso code XX, see 'geoname' table for columns
Cities1000.zip	all cities with a population > 1000 or seats of adm div (ca 150.000), see 'geoname' table for columns
Cities5000.zip	all cities with a population > 5000 or PPLA (ca 50.000), see 'geoname' table for columns
Cities15000.zip	all cities with a population > 15000 or capitals (ca 25.000), see 'geoname' table for columns
admin1CodesASCII.txt	ascii names of admin divisions.
admin2Codes.txt	names for administrative subdivision 'admin2 code' (UTF8), Format: concatenated codes <tab>name <tab> asciiname <tab> geonameId

Table 10-7 – Convenience Files

These files are subsets of the allCountries file with the same file layout and attributes, but with fewer rows. Users interested only in a single country can download the features for this country in a file named with the iso two letter country code. Users only interested in major cities can download one of the three citiesXXX files.

The alternateName table has a couple of flags, which need to be evaluated in order to determine the best name variant for display in a particular language (see also the section ‘Using GeoNames - Tips and Tricks - Which name to use’). For the first and second level administrative divisions, two convenience files (admin1CodesASCII.txt and admin2Codes.txt) contain the best name for display in English.

Modification Files

The download directory also includes a couple of convenience files with the modifications of the previous day. These may be used to update the local copy of GeoNames on a daily basis. The downside of this approach is that no file must be skipped.

We consider it preferable to download the full dataset periodically instead of the daily modification files.

Each feature has at least the main name in the geoname table, often additional name variants are found in the alternatename table. There might also be various name variants in a particular language and we need to decide which name to use.

Modification Files (With <date> = <yyyy-MM-dd>)	
modifications-<date>.txt	all records modified on the previous day
deletes-<date>.txt	all records deleted on the previous day, format: geonameId <tab> name <tab> comment. Duplicates usually have a delete comment of the form ‘duplicate of <geoNameId>’ with the geoNameId of the remaining feature.
alternateNamesModifications-<date>.txt	all alternate names modified on the previous day
alternateNamesDeletes-<date>.txt	all alternate names deleted on the previous day, format: alternateNameId <tab> geonameId <tab> name <tab> comment.

Table 10-8 – Modification Files

10.7 Using GeoNames - Tips and Tricks – Which name to use?

The recommended approach is to check the alternate name table for name variants in the desired language, ignoring the historic and colloquial names. From these name variants, we then use the one marked as 'short' for display. If no name is available with the desired iso language code we see whether we have an alternate name without language code. Last but not least we use the name from the geoname table if we do not find an appropriate entry in the alternatename table.

Duplicates

Place names are not unique, they even tend to cluster in an area. People often complain about 'duplicates' in GeoNames because they erroneously assume that a feature name may only exist once per country and first admin division. Unfortunately, this is not the case and village names are often repeated in an area.

10.8 Webservice vs Data Download

Besides the freely available daily database extracts GeoNames also offers a wide range of web services in a freemium model. A number (currently 20000 credits) of requests per day is free, for higher usage or a service level agreement a premium offering exists.

The GeoNames web services have the advantage that they offer a ready-made and direct access to services based on the gazetteer data. Many additional services, internally used to enhance and verify the gazetteer data, are also available. Among these: elevation, time zone, Wikipedia, weather and street reverse geocoding services.

Provide Data – Contribute to GeoNames

Contributions to GeoNames can be made directly with the browser based wiki interface. Only a user account is needed. Experienced users may be upgraded to higher user levels to be able to also update critical attributes and features.

For contributing entire data sets it is preferred to import the data with a bulk import instead of the wiki interface. It is sufficient to make the data available in flat csv files, similar to the export files.

TimeZone

One of the attributes of a feature is the timeZoneId. It refers to the iana timezone project. Example Europe/Paris for the time in continental France.

The iana timezone project maintains all time changes since 1970 and is included in most programming environments. The iana timeZoneId is therefore sufficient to determine all time changes (daylight saving time) since 1970.

Countries used in GeoNames

GeoNames relies on the ISO [3] country codes. Each ISO country code has a corresponding entity in the GeoNames database.

The only exception so far is Kosovo, which does not yet have an ISO country code assigned. GeoNames is temporarily using XK, till an ISO country code will be officially assigned.

Admin codes used by GeoNames

On the first administrative level GeoNames is mainly using the US coding standard (formerly FIPS). On lower levels the national code, often provided by the national statistical office, is used.

10.9 Data Sources

GeoNames is accumulating data from a large variety of different data sources together with user contributions via the wiki style browser based edit interface. The largest single data source is the United States NGA GEOnet Names Server (GNS) from the National Geospatial-Intelligence Agency. [5] It contains data for all countries except the US. GeoNames is importing the NGA data set on a monthly basis.

As more and more national mapping agencies and national statistical offices become aware of the importance of free access to geographical data the number of national agencies offering data for free under an open license is growing year by year. Thus, continuously increasing the number of data sources GeoNames is integrating.

In developing countries where national mapping agencies do not have the same resources as in industrialized countries relief organizations working with GeoNames data often contribute their own data.

When periodically importing a data set from a data provider, GeoNames checks each modification with the modification history of the same attribute to make sure corrections done by users via the wiki interface are not directly overwritten and set back with the following data import.

10.10 References

[1] www.geonames.org, last accessed 09/2016

[2] Internet Assigned Numbers Authority (IANA), http://www.iana.org/, last accessed 09/2016

[3] International Organization for Standardization (ISO), http://www.iso.org/ last accessed 09/2016

[4] Italian National Institute of Statistics, http://www.istat.it/ last accessed 09/2016

[5] NGA GEOnet Names Server (GNS), http://geonames.nga.mil/gns/html/index.html last accessed 09/2016

Section 5 Progress of toponymic databases

Chapter 11 **Standardization of Statistical Area Names**

Centro de Referência em Nomes Geográficos, Instituto Brasileiro de Geografia e Estatística

11.1 Introduction

A country with Brazil's dimensions and regional differences demands careful planning of census operations, which are initiated with:

- Characterization of the national territory, in its fundamental aspects, for adequate assessment of its volume of data and costs for the collection operation;
- Correct representation of present boundaries (federation units, cities, subdistricts, and other territorial structures) to ensure adequate ascertainment and disclosure of results;
- Delimitation and description of small territorial collection units (enumeration areas), so as to ensure perfect recognition of work areas.

One of the biggest concerns during planning for the census is avoiding omission problems and/or coverage duplicity of census operation, which is attained by improving the maps used for operation and by a registry system; together they constitute what is conventionally called the Territorial Base.

Therefore, to attain the Territorial Base Planning goals, the territory must be split up in smaller spatial units, called territorial units or enumeration areas. The delimitation process for enumeration areas aims to support census infrastructure planning, ensure coverage of all the national territory in the period established for census collection, allow comparability of information between censuses, as well as promote ascertainment and disclosure of results for political-administrative divisions (DPA) and other national territorial structures.

In this context, it is important to highlight the definitions of Disclosure Areas and Ascertainment Areas.

Disclosure Areas are areas legally instituted at states and cities, related to DPA, in force at the trigger date of census surveying, which meet the disclosure requirement of census data. DPA portrays state borders and limits for cities, districts and subdistricts. According to the 1988 Constitution, DPA management is the responsibility of states and cities; intracity territorial structure management is under the competence of cities as well.

Beyond legal obligation, the Brazilian territorial structure constitutes one marker for the processes of occupation and regional distinction of the territory, since the territory is, largely, the direct or indirect result of State action. Beyond that, being it a dynamic structure, it goes through constant changes, demanding an effort for updating of the limits of territorial units which compose the Brazilian federation.

Disclosure areas portrayed at IBGE surveys are:

•Federation Units - UFs (Federal District and states);
•Cities;
•Districts;
•Subdistricts, Administrative Regions (RAs) and zones;
•Urban areas;
•Isolated Urban Areas (AUI's) and
•Rural areas.

On the other hand, ascertainment areas are geographical areas defined on maps and logged on the Territorial Base to serve as spatial units to ascertain statistic data in different territory environments, aiming to meet the growing demand for disclosure on these subjects. Disclosure of ascertainment areas depends, however, on partnerships established between IBGE and the agencies in charge of ascertainment areas (when they exist) such as, for example, Fundação Nacional do Índio (FUNAI), in charge of Indian Lands (TIs).

In this classification, territorial sections legally defined for cultural and environmental conservation reasons as well as those sections more directly involved with urban and rural politics stand out. IBGE currently considers the following ascertainment areas:

- Urban Cities or Villages;
- Non Urban Cities or Villages ;
- Neighborhoods;
- Subnormal Clusters;
- Rural Clusters;
- Settlement Projects (PA), Agro-villas;
- Conservation Units (UCs) ;
- Indian Lands (TIs) ;
- Indian Villages ;
- Quilombola Lands (TQs) ; and
- Quilombola Communities.

(Quilombola Lands are agricultural communities established by runaway slaves and Indians in the Interior in the 18th and 19th century, whose land rights have been officially recognized)

That said, it's necessary to emphasize the importance of correct and precise coding of enumeration areas to recover disclosure and ascertainment areas and consequently integrate statistic and geo- spatial info, summarized in figure 11-1:

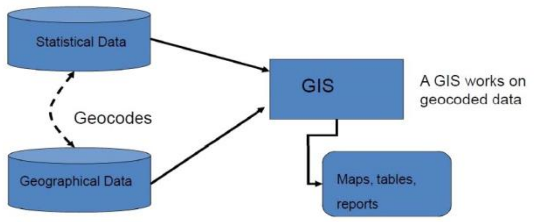


Figure 11-1 Integration of statistical and geographical data. Source: Laaribi, A., 2014

11.2 Criteria for the definition of statistical areas

Enumeration areas are delimited taking into consideration the divisions of disclosure and ascertainment areas, in this order. This prerogative is justified to ensure viability of the survey, ascertainment and dissemination of statistic info on the foreordained territorial structure. They are the smallest areal units in which the National Territory is fragmented for the purposes of IBGE's statistic collection. At the last demographic census, in 2010, Brazil was divided in 314 thousand enumeration areas.

Enumeration areas are, therefore, continuous areas (with the exception of island territories, which do need to obey quantitative home criteria and can either be isolated in a enumeration area or be connected to a enumeration area on land) which must be classified according to their situation as urban or rural (see figure 11-2). This way, it stands out that according to spatial logic, urban enumeration areas tend to have smaller territorial extensions, being more populous, while rural sectors have bigger extension but sparser population. It should be highlighted that according to information from IBGE's Coordenação de Estruturas *Territoriais* (CETE), responsible for the Territorial Base, in January 2016, Brazil counted 344,220 enumeration areas, 252,873 of them classified as urban and 91,347 of them classified as rural.

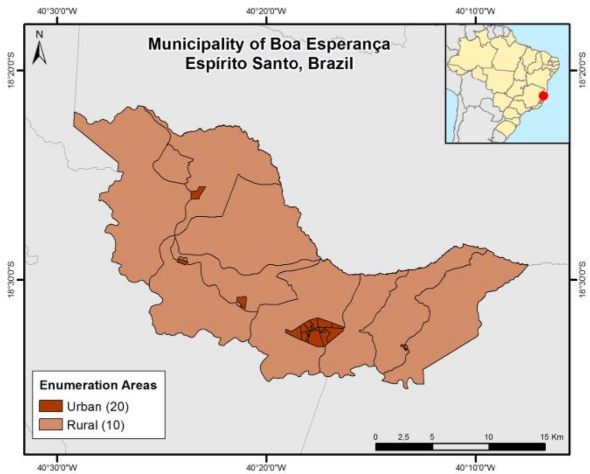


Figure 11-2 Urban and Rural Enumeration areas Boa Esperança City, Espírito Santo, Brazil Source: IBGE, 2015

An enumeration area must be even more adequately classified according to geopolitical, administrative, morphologic, socioeconomic and cultural characteristics, so that its delimitation fulfills the goal of dividing Brazilian territory, according to the assessment of data volumes and collection operation costs, also representing adequately the current territorial structures to ensure ascertainment and disclosure of results, which will offer fundamental elements for geospatial analyses.
Delimitation of geographic space of enumeration areas is defined by a sequence of topographic or artificial landforms, preferably stable ones that can be easily identified on the field, so as to avoid that collection agents invade the enumeration area of other collection agents or omit collections in the enumeration area under their responsibility.

In short, an enumeration area is the territorial unit established for purposes of registry control, formed by a continuous area, located in a single urban or rural framework, with dimension and number of homes that allow surveying by one census agent. That way, all census agents will proceed with data collection having coverage of their designated enumeration area as a goal.

11.3 Presentation of the "Geocode"

In Brazil, enumeration areas have no names. They are identified by numbers, which in turn connect them to the names of regions, states, cities and districts. In this chapter, we will explain how these numbers are composed.

Each enumeration area has a number with allows its identification in relation to other enumeration areas in a unique way: geocode. The geocode aims to allow reference of lots of information by collection territorial unit, being used as a recovery key of several registries in the Territorial Base. This number is formed by 15 digits.

Beyond this number referring unequivocally to the sector, its composition allows for recognition of the region, state, city, district and subdistrict (or RA or zone) where the enumeration area is located. The example below shows the geocode composition for a sector:

Figure 11-3 Example of the Geocode

In figure 11-3, starting from the enumeration area geocode (411520005010010) and from tables matching the geocodes to state-, city-, district- and subdistrict names, it is possible to know that enumeration area 0010 is located in subdistrict 01 (Zone 1), in district 05 (Maringá), in city 15200 (Maringá), which is located in state 41 (Paraná), in Region 4 (South Region), Brazil.

That way, it is shown that territorial units that compose Brazilian DPA will always be associated to a unique code that allows integration of the territorial base to the most diverse statistics date base. It should be highlighted that when a DPA alteration occurs, a recoding of territorial units involved is performed so as to preserve the previous geocode making it possible to reconstruct territorial organization of the past. For example, districts that were split off could not have their geocodes reused at their origin cities; that way it is possible to rescue previous information, that is, when time passes, it is possible to recover the district's origin, beyond that all levels inserted in the district, among them enumeration areas, must be recoded.

Geocode determination to different Brazilian territorial units is managed by *Coordenação de Estruturas Territoriais* (CETE), part of IBGE, and follows specific criteria for each DPA hierarchic level.

Figure 11-4 Geocodes of the Brazilian Regions (Source: IBGE)

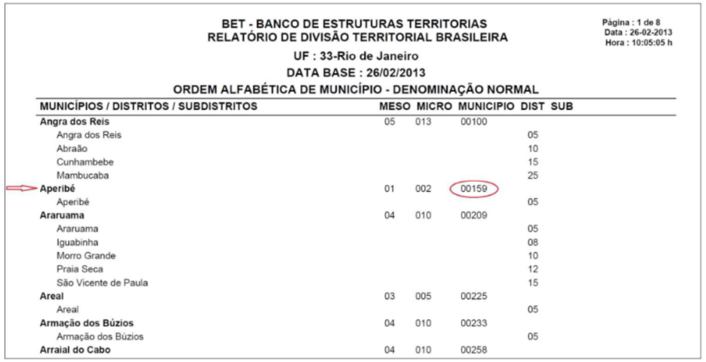
In case of regions, also called great regions; their identification is done by the 1st digit. Numbering order started by the northernmost region, following clockwise from there, applying the following numbers: 1 - North Region; 2 - Northeast Region; 3 - Southeast Region; 4 - South Region and 5 - Center-West region.

Figure 11-5 Geocodes of the Brazilian Federation Units (Source: IBGE)

The following territorial level refers to Federation Units (UFs). In total, there are 27 UFs whose geocodification also occurs clockwise, starting at 1, and joins with the Region Geocode in which the UF is located as presented in figure 11-5.

Figure 11-6 Example of UF Geocodes located at Brazil's North Region (Source: Banco de Estruturas Territoriais– BET, IBGE.)

The next five geocode digits refer to city, the fifth one being called a verifier digit. The rules for geocode attribution to a new city are as follows:



BET - BANCO DE ESTRUTURAS TERRITORIAS
RELATÓRIO DE DIVISÃO TERRITORIAL BRASILEIRA

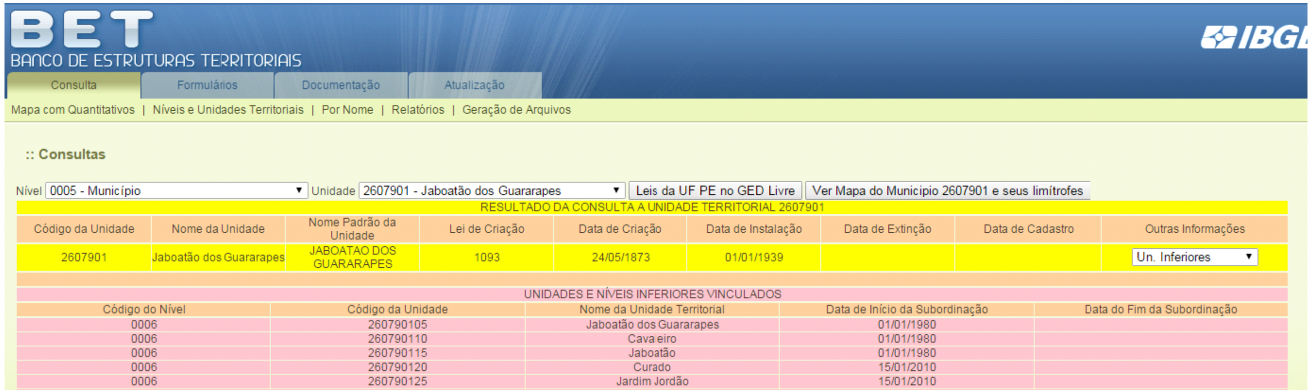
Página : 1 de 8
Data : 26-02-2013
Hora : 10:05:05 h

UF : 33-Rio de Janeiro
DATA BASE : 26/02/2013
ORDEM ALFABÉTICA DE MUNICÍPIO - DENOMINAÇÃO NORMAL

MUNICÍPIOS / DISTRITOS / SUBDISTRITOS	MESO	MICRO	MUNICIPIO	DIST	SUB
Angra dos Reis	05	013	00100		
Angra dos Reis				05	
Abraão				10	
Cunhambebe				15	
Mambucaba				25	
Aperibé	01	002	00159		
Aperibé				05	
Araruama	04	010	00209		
Araruama				05	
Iguabinha				08	
Morro Grande				10	
Praia Seca				12	
São Vicente de Paula				15	
Areal	03	005	00225		
Areal				05	
Armação dos Búzios	04	010	00233		
Armação dos Búzios				05	
Arraial do Cabo	04	010	00258		

Figure 11-7 DTB report with example of Rio de Janeiro cities and districts (Source: IBGE (2010 Census))

•Add the new city in the interval in which its name fits in alphabetic order on Brazilian Territorial Division report for the state to which the city in question belongs. In figure 11-7, one may find the example of the city called Aperibé in Rio de Janeiro state which should be between the cities of Angra dos Reis and Araruama.



BET
BANCO DE ESTRUTURAS TERRITORIAIS
IBGE

Consulta | Formulários | Documentação | Atualização

Mapa com Quantitativos | Níveis e Unidades Territoriais | Por Nome | Relatórios | Geração de Arquivos

:: Consultas

Nível 0005 - Município | Unidade 2607901 - Jaboatão dos Guararapes | Leis da UF PE no GED Livre | Ver Mapa do Municipio 2607901 e seus limítrofes

RESULTADO DA CONSULTA A UNIDADE TERRITORIAL 2607901

Código da Unidade	Nome da Unidade	Nome Padrão da Unidade	Lei de Criação	Data de Criação	Data de Instalação	Data de Extinção	Data de Cadastro	Outras Informações
2607901	Jaboatão dos Guararapes	JABOATAO DOS GUARARAPES	1093	24/05/1873	01/01/1939			Un. Inferiores

UNIDADES E NÍVEIS INFERIORES VINCULADOS

Código do Nível	Código da Unidade	Nome da Unidade Territorial	Data de Início da Subordinação	Data do Fim da Subordinação
0006	260790105	Jaboatão dos Guararapes	01/01/1980	
0006	260790110	Cava eiro	01/01/1980	
0006	260790115	Jaboatão	01/01/1980	
0006	260790120	Curado	15/01/2010	
0006	260790125	Jardim Jordão	15/01/2010	

Figure 11-8 Example of district Geocode at the city of Jaboatão dos Guararapes, Recife, Brazil. (Source: Banco de Estruturas Territoriais – BET, IBGE.)

BET - BANCO DE ESTRUTURAS TERRITORIAIS
RELATÓRIO DE SUBDISTRITOS
UF : 43-Rio Grande do Sul
DATA BASE : 09/03/2016
ORDEM ALFABÉTICA DE MUNICÍPIO - DENOMINAÇÃO NORMAL

Página : 1 de 6
Data : 09-03-2016
Hora : 11:16:56 h

MUNICÍPIO	DISTRITO	CÓDIGO	NOME	LEI_CRIA	DATA_CRIA	DATA_INST
4300406-Alegrete	05-Alegrete	06	Zona Central	2885	29/03/1999	29/03/1999
		07	Zona da Coxilha ou Cidade Alta	2885	29/03/1999	29/03/1999
		08	Zona Norte	2885	29/03/1999	29/03/1999
		09	Zona Sul	2885	29/03/1999	29/03/1999
		10	Zona Leste	2885	29/03/1999	29/03/1999
		11	Zona Oeste	2885	20/03/1999	20/03/1999
		12	Alegrete	4850	24/10/2011	
		13	Catimbaú	4850	24/10/2011	
		14	Durasnal	4850	24/10/2011	
		15	Guassu-Boi	4850	24/10/2011	
		16	Inhanduí	4850	24/10/2011	
		17	Itaporó	4850	24/10/2011	
		18	Rincão de São Miguel	4850	24/10/2011	
		19	Vasco Alves	4850	24/10/2011	

TOTAL P/MUNIC : 14

Figure 11-9 Table with example of subsdistrict Geocoding located at the city of Alegrete, Rio Grande do Sul, Brazil. (Source: Banco de Estruturas Territoriais – BET, IBGE.)

•Choose Aperibé's geocode, preferably, in the interval of city geocodes of Angra dos Reis and Araruama (if not possible, geocode should be last in the list for UF's cities). In this case, Angra dos Reis geocode is 0010 and Araruama is 0020, with geocode 0015 being chosen (0010<0015<0020), leaving, if possible, an interval to register new cities.

•The last city geocode digit, DV (9, in Aperibé's case), is calculated by computer algorithms. Geocode for the city of Aperibé, therefore, will be as follows: 33 0015 9.

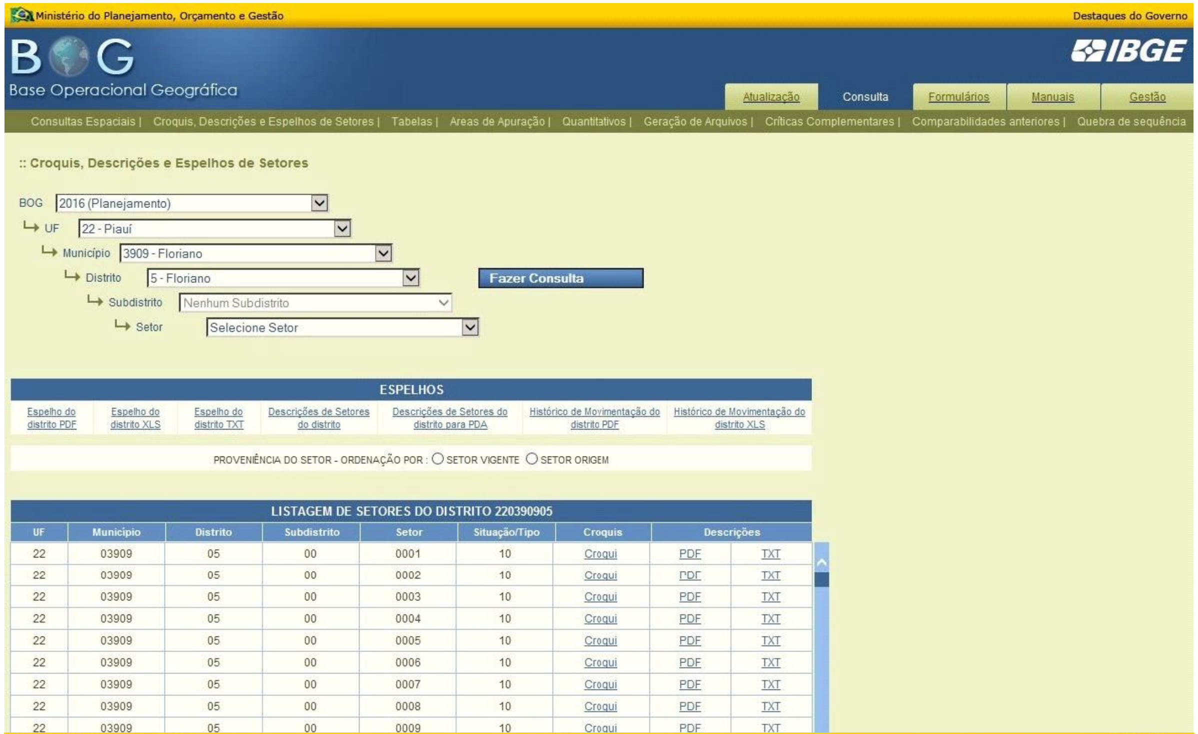
After the geocode associated to the city there are two digits that form the district geocode, beginning with 05 for the base district. It should be highlighted that all of Brazil's cities have, at least, the base district and that it has the same name as the city. In case the city has other districts beyond the base district, geocodes for these districts should keep, if at all possible, alphabetic order of its respective names.

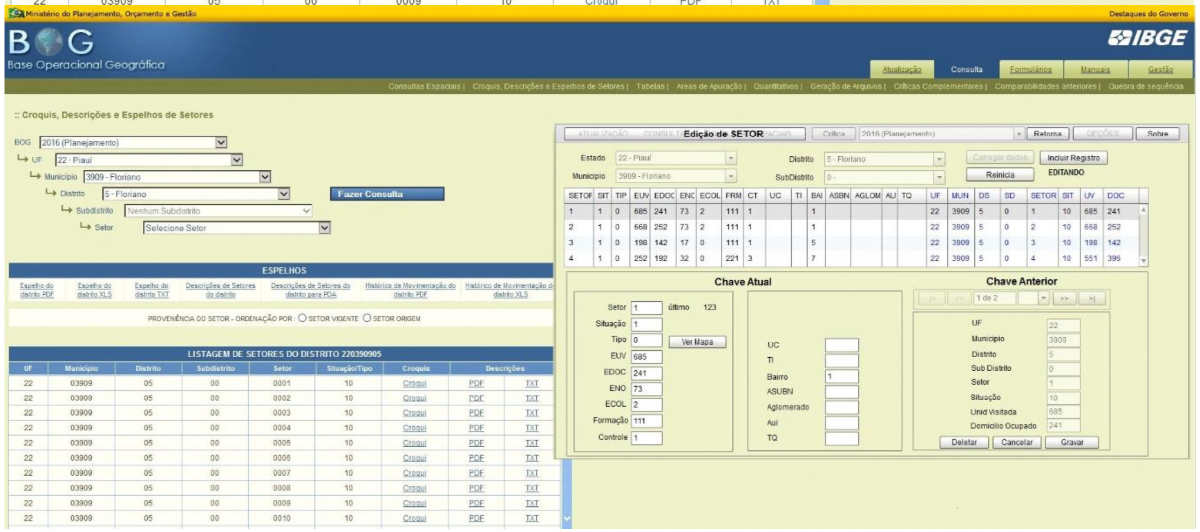
Next Brazilian DPA level refers to subdistricts, RAs or zones to which two digits are attributed, starting preferably at 06. When a city does not have territorial units of this nature, the geocode must be 0. For example, in figure 11-9 it is possible to see the existent subdistricts on the city of Alegrete, state of Rio Grande do Sul, in detail.

Completing the geocode, the four digits referring to the enumeration area are attributed using crescent sequential criteria, inside the smallest existing area (district and subdistrict, RA or zone). According to this rule, numbering activity of census factors follows the following steps:

- In a district, subdistrict, RA or zone, numbering should be unique, sequential and crescent, starting at 1;
- Numbering should start at the urban framework and continue to rural framework in the same district, subdistrict, RA or zone, starting numbering from the area considered as the center of city or village; and
- Numbering should start at 1, following spirally clockwise and towards the rim; after all enumeration areas in a district, subdistrict, RA or zone are numbered. In case the city has several districts or subdistricts, they should number them as well, starting again at number 1, and repeating every numbering operation, until all sectors of all districts, subdistricts, RAs or city zones are numbered.

While creating new sectors, preferably, the numbering sequence should be kept, differentiating urban and





rural frameworks.

All census geocodes find themselves documented in what is called an Enumeration Area Spreadsheet, available at the Geography Operational Base (BOG). Such documents allow for the identification of existing homes in the sector, its situation - urban or rural, its kind and location in relation to disclosure and ascertainment areas. It should be highlighted that there is a sector spreadsheet for each Brazilian city. Figures 11-10 and 11-11 present the city of Floriano, in the state of Piauí, as an example, using information contained in the BOG.

Figure 11-10 List of Enumeration Areas at the city of Floriano, state of Piauí, through BOG query.

Figure 11-11 Spreadsheet of one of the Enumeration Areas of the city of Floriano, state of Piauí, through BOG query.

11.4 Some considerations about Territorial Structure Database (BET)

In the context of registries, files and databases that compose the Territorial Base is included the BET - Territorial Structure Database, registration reference for DPA, for IBGE's and several other public and private bodies' projects and activities.

Its conception with territorial levels, level hierarchy, dates and information about legislation on creation, installing, extinction, alteration of toponymy and other attributes makes the recovery of DPA's historic evolution viable, as well as recovery of territorial snapshots for dates of interest.

Beyond DPA's territorial structures, Territorial Structure Database contemplates, still, mesoregion and geographic microregion levels, including neighborhoods and like, rural clusters, subnormal clusters, special areas and respective hierarchies and attributes.

On the scope of Census, BET is the registry responsible for territorial information (codes, names and territorial sub ordinance) of disclosure and ascertainment units registered at Territorial Base which are disseminated at IBGE's publications and products.

During the 2010 Census, 5,565 cities, 10,283 districts, 662 subdistricts, 10,665 rural clusters, 6,329 subnormal clusters, 14,402 neighborhoods, among other territorial levels, were registered in the Territorial Structure Database.

BET was updated during the preparation for the Territorial Base 2010 with territorial changes after the 2000/2007 Census, aiming not only to make territorial information related to IBGE's ongoing activities available, but mainly making the territorial structure of the 2010 Census available for dissemination.

A vision of BET's levels and territorial hierarchies is presented as follows: Levels:

BET has available, at the moment, 20 territorial levels, being that 14 of them refer to current territorial structures and 6 are called historic levels.

The current levels are:

1- Brazil (level 0000);

2- Levels referring to regional division:

- Geographic Region (level 0001)
- Geographic Mesoregion (level 0008) and Microregion (level 0009)

3- Levels referring to territorial division, registered according to legislation:

- Federative Unit (level 0002)
- City (level 0005)
- District (level 0006)
- Subdistrict (level 0007);

4- Levels referring to Census ascertainment areas:

- Neighborhoods (level 0102), Indian Lands (0105) and Environmental Conservation Units (0106)® territorial structures legally established
- Rural Clusters (level 0103), Isolated Urban Area (level 0107) and Subnormal Clusters (level 0101) ® registered according to concepts adopted during construction of Territorial Base

Código do Nível	Nome do Nível	Código da Unidade	Nome da Unidade	Nome Padrão da Unidade	Data de Criação	Data de Instalação	Data de Extinção
0104	Área Especial	1546	PAEMU DO RIO DE JANEIRO	PAEMU DO RIO DE JANEIRO	31/03/1988	31/03/1988	31/07/2009
0012	Unidade Federativa75	51	Rio de Janeiro	RIO DE JANEIRO			01/01/1975
0006	Distrito	330455705	Rio de Janeiro	RIO DE JANEIRO		01/01/1980	
0005	Município	3304557	Rio de Janeiro	RIO DE JANEIRO	01/03/1565	15/03/1975	
0002	Unidade Federativa	33	Rio de Janeiro	RIO DE JANEIRO		01/01/1975	
0005	Município	3099901	Rio de Janeiro	RIO DE JANEIRO	01/03/1565	01/01/1939	20/04/1960
0005	Município	3499901	Rio de Janeiro	RIO DE JANEIRO	01/03/1565	21/04/1960	14/03/1975
0003	Mesorregião Homogênea	54	Grande Rio de Janeiro	GRANDE RIO DE JANEIRO		01/01/1980	31/05/1990
0103	Aglomerado Rural	241335900022	Vila Rio de Janeiro	VILA RIO DE JANEIRO	28/09/2006	28/09/2006	
0103	Aglomerado Rural	241335900010	RIO DE JANEIRO	RIO DE JANEIRO	31/03/1988	31/03/1988	24/08/1999
0008	Mesoregião Geográfica	3306	Metropolitana do Rio de Janeiro	METROPOLITANA DO RIO DE JANEIRO	01/06/1990	01/06/1990	
0015	Município75	53001	Rio de Janeiro	RIO DE JANEIRO			01/01/1975
0009	Microregião Geográfica	33018	Rio de Janeiro	RIO DE JANEIRO	01/06/1990	01/06/1990	
0014	Microrregião Fisiográfica	53000	Rio de Janeiro	RIO DE JANEIRO			01/01/1975
0101	Aglomerado Subnormal	35503080891	Rio de Janeiro	RIO DE JANEIRO	08/12/2009	08/12/2009	
0016	Região Metropolitana	3301	Região Metropolitana do Rio de Janeiro	REGIAO METROPOLITANA DO RIO DE JANEIRO	01/07/1974	01/07/1974	

Figure 11-12 BET Name Query - Example of a search for "Rio de Janeiro".

Relation between Territorial Structures:

To recover subordination of units such as, for example, the relation of districts in a certain city, there is a pre-established level hierarchy. Each level has a unit code composition which is standard for DTB levels and the same as Operational Bases for ascertainment areas.
The diagram in figure 11-13 (next page) presents the structure and composition for current levels.
BET allows that queries to territorial units be realized through territorial level or name. Hereinafter is an example of query by name which evidences how geocode helps to differentiate between units with the same name but different level.

11.5 Conclusion

The demand for stronger integration between statistic and geospatial data is growing, therefore it is indispensable for methods that allow such integration to improve. When it comes to IBGE, it is possible to notice the existence of a geocode system already well-structured and widespread in the institution for such goals.
These geocodes make identifying territorial units whose names are not yet standardized easier. This way, possible divergences in geographic names used by statistics department and geography department are voided allowing unequivocal reference to spatial units.
The use of geocodes meets the principle of univocity, however, it does not meet the need for standardizing geographic names since this activity requires analysis of linguistic, cultural, historic and ethnic aspects, among others.

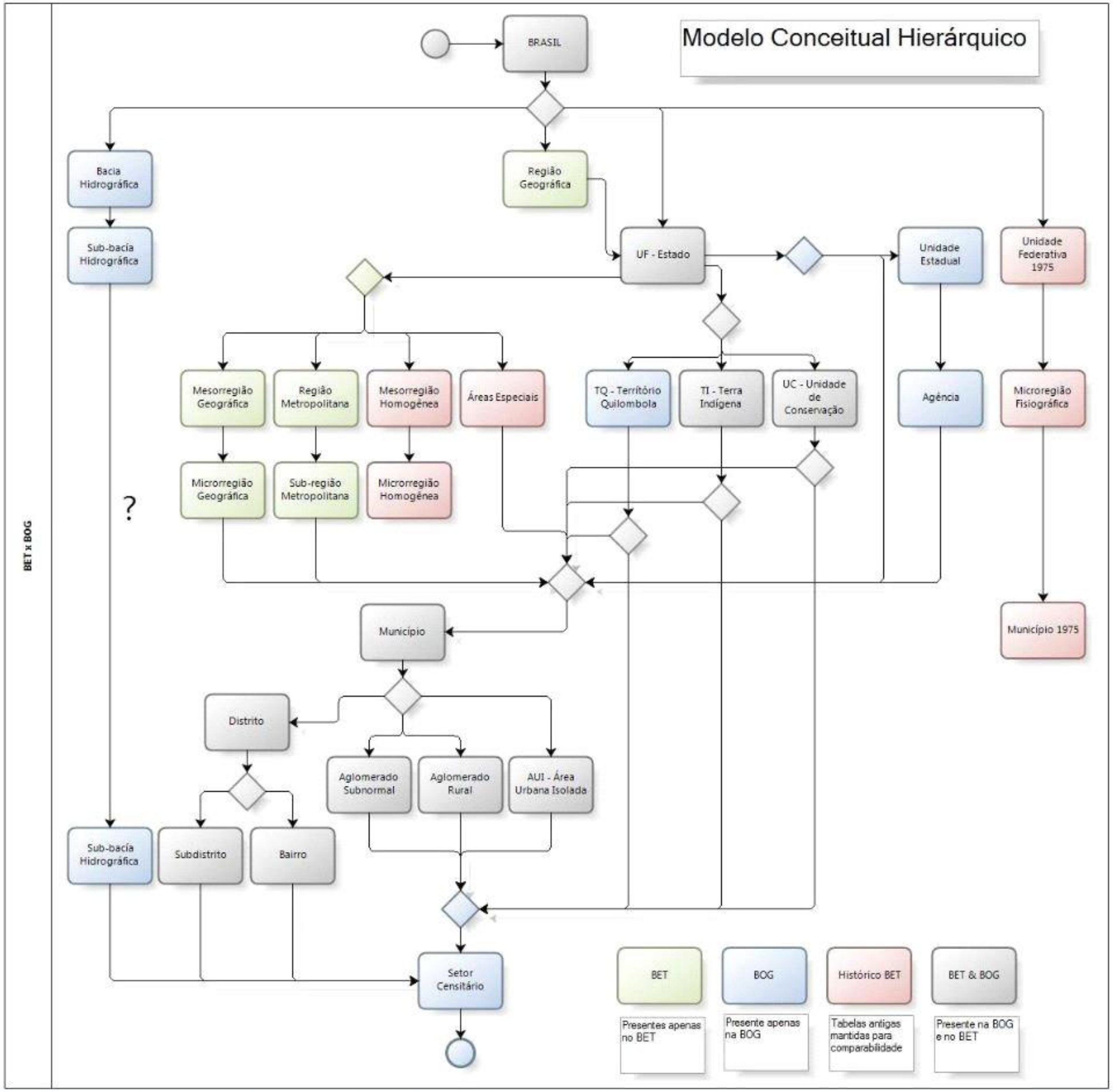
11.6 References

GUIA do censo. Rio de Janeiro: IBGE. Available at: <http://censo2010.ibge.gov.br/materiais/guia-do-censo/operacao-censitaria.html> Access in April, 2016.

LAARIBI, A. Linking People and Socio-Economic Information to a Location: Integrating Statistical and Geospatial Information, 2014. Available at: <https://unstats.un.org/unsd/geoinfo/UNGEGN/docs/28th-gegn-docs/28th%20Session-UNGEGN.pdf> Access in April, 2016.

MANUAL da Base Territorial. Rio de Janeiro: IBGE, 2013.

Figure 11-13 Territorial Structure Database Structure



Section 5 Progress of toponymic databases

Chapter 12 **Volunteered Geographic Information (VGI)**

Pier-Giorgio Zaccheddu

As methods of accessing geospatial information change and public interaction with such data becomes more dynamic, national mapping agencies and place-name standardization organizations are looking to take advantage of the willingness of the local population to provide information on location and associated attributes.

"The reduction in barriers to entry, the growth of Web and mobile mapping and the enthusiasm for crowd-sourced geospatial data have massively increased the role of the private sector and the volunteer community over the last decade. The fact that Google and Bing Maps are global brands highlights that it is the private sector that has made digital mapping accessible to the masses. Alongside the massive explosion in the use of geospatial information, 'Volunteered Geographic Information (VGI)' groups such as 'OpenStreetMap (OSM)' have gone some way towards popularising, albeit within a relatively niche community, the collection of geospatial data" [1].

For the utilization of crowd-sourced geospatial data, including geographical names, there is no single established approach [2]. The following information in this section is limited to the use of VGI for the publication and dissemination of geographical names data and more specifically, the use of one VGI community, the OpenStreetMap (OSM) community.

OSM data is subject to copyright and database rights, and is available under a license called Open Database License (ODbL). In a nutshell, ODbL is "share-alike license" which says that if one publishes works based on the OSM data, it must be attributed to OSM, and if the OSM data is enhanced with one's own data, the enhancements must be shared with the community. Organizations have the choice of doing this by making available a derived database, or at least by publishing the method (code) used to make it, or both. The two key concepts for legal purposes are 'produced works' (published works that use OSM in their production) and 'derivative databases' (data sets which include data which is derived from, based or dependent on OSM's data in any way). How the license applies to the use case depends largely on the purpose foreseen for the data. In all cases, the usage of OSM data is free of charge, and organizations are free to charge others for any enhancement work and/or set own licensing conditions, but any derivative database from OSM must be available, for free, to anyone who requests it. [3].

Often OSM is used to publish a map which shows data collected by others in a geographical context. It makes no difference whether this is on a web map, in a printed publication, as a graphic in a TV programme, or in any other media. Non-map examples include creating a spreadsheet or database that shows a public transport route along with information from timetables to illustrate arrival/departure times from certain places, or a mobile app to help users locate points of interest in the real world.

Related to the 12 January 2010 earthquake in Haiti, the OSM community created a map coverage of the worst hit cities of Port-au-Prince and Carrefour, and the neighbouring countryside thanks to the "Haiti earthquake response mapping project" [4]. Using the free data, OSM developers and others around the Web have created a set of earthquake map resources in addition to OSM's basic online map:

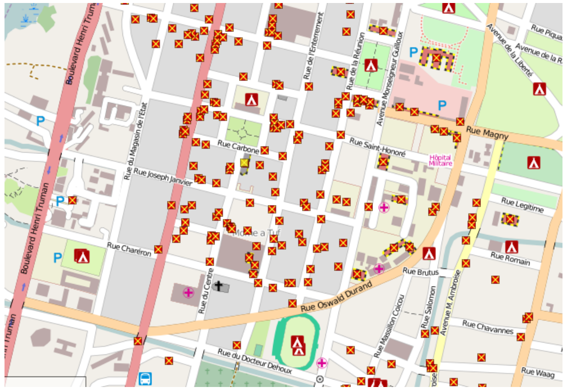


Figure 12-1 Example of an area zoomed on Port-au-Prince rendered on openstreetmap.nl The customized rendering set up by User:Ldp which shows damaged buildings and refugee camps mapped within OpenStreetMap using special GeoEye/DigitalGlobe imagery. See WikiProject Haiti#2010 Earthquake Response, [*http://wiki.openstreetmap.org/wiki/File:Haiti_earthquake_damage_map.png*](http://wiki.openstreetmap.org/wiki/File:Haiti_earthquake_damage_map.png)

The ODbL might be an obstacle or at least a reason to hesitate for an organization before using OSM for the publication and dissemination of its geographical names data or other geospatial data.

Generally, VGI and crowd-sourced data is likely to include valuable additional information in those

countries where well‑established geospatial sources are already available. The speed at which data can be gathered and compiled with VGI in times of crisis can be of importance in countries where no or few geospatial sources exist, although the accuracy and acceptability of the geographical names (for example) may be questionable in the long term. This VGI has at least the potential to provide a user's view of their geography. The supporting community states that more tailored public services and even more effectively targeted interventions by policy and decision‑makers could be possible. Very likely, VGI and crowd-sourced content will decrease cost, improve accuracy and increase availability of rich geospatial information. It is envisaged that progress will be made in future on bridging the gap between authoritative data and crowd-sourced data, moving towards true collaboration.

References

[1] UN-GGIM publication, 'Future trends in geospatial information management: the five to ten-year vision', chapter '4.1 Making mapping accessible to the masses', http://ggim.un.org/docs/Future-trends.pdf, last accessed 09/2016

[2] Online discussion forum of the Working Group on Toponymic Data Files and Gazetteers of UNGEGN, forum 1: 'Volunteered geographic information (VGI) / crowd-sourcing', https://wiki.gdi-de.org/display/wgtdfg/Forum+1+-+Volunteered+geographic+information+and+crowd-sourcing, last accessed 09/2016

[3] Introduction to OSM, http://wiki.openstreetmap.org/wiki/License/Use_Cases#Introduction, last accessed 09/2016

[4] Haiti earthquake response mapping project, http://wiki.openstreetmap.org/wiki/WikiProject_Haiti, last accessed 09/2016

Section 5 Progress of toponymic databases

Chapter 12 **Volunteered Geographic Information (VGI)**

Pier-Giorgio Zaccheddu

As methods of accessing geospatial information change and public interaction with such data becomes more dynamic, national mapping agencies and place-name standardization organizations are looking to take advantage of the willingness of the local population to provide information on location and associated attributes.

"The reduction in barriers to entry, the growth of Web and mobile mapping and the enthusiasm for crowd-sourced geospatial data have massively increased the role of the private sector and the volunteer community over the last decade. The fact that Google and Bing Maps are global brands highlights that it is the private sector that has made digital mapping accessible to the masses. Alongside the massive explosion in the use of geospatial information, 'Volunteered Geographic Information (VGI)' groups such as 'OpenStreetMap (OSM)' have gone some way towards popularising, albeit within a relatively niche community, the collection of geospatial data" [1].

For the utilization of crowd-sourced geospatial data, including geographical names, there is no single established approach [2]. The following information in this section is limited to the use of VGI for the publication and dissemination of geographical names data and more specifically, the use of one VGI community, the OpenStreetMap (OSM) community.

OSM data is subject to copyright and database rights, and is available under a license called Open Database License (ODbL). In a nutshell, ODbL is "share-alike license" which says that if one publishes works based on the OSM data, it must be attributed to OSM, and if the OSM data is enhanced with one's own data, the enhancements must be shared with the community. Organizations have the choice of doing this by making available a derived database, or at least by publishing the method (code) used to make it, or both. The two key concepts for legal purposes are 'produced works' (published works that use OSM in their production) and 'derivative databases' (data sets which include data which is derived from, based or dependent on OSM's data in any way). How the license applies to the use case depends largely on the purpose foreseen for the data. In all cases, the usage of OSM data is free of charge, and organizations are free to charge others for any enhancement work and/or set own licensing conditions, but any derivative database from OSM must be available, for free, to anyone who requests it. [3].

Often OSM is used to publish a map which shows data collected by others in a geographical context. It makes no difference whether this is on a web map, in a printed publication, as a graphic in a TV programme, or in any other media. Non-map examples include creating a spreadsheet or database that shows a public transport route along with information from timetables to illustrate arrival/departure times from certain places, or a mobile app to help users locate points of interest in the real world.

Related to the 12 January 2010 earthquake in Haiti, the OSM community created a map coverage of the worst hit cities of Port-au-Prince and Carrefour, and the neighbouring countryside thanks to the "Haiti earthquake response mapping project" [4]. Using the free data, OSM developers and others around the Web have created a set of earthquake map resources in addition to OSM's basic online map:

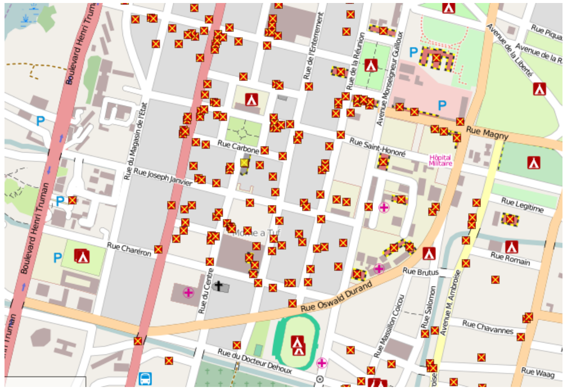


Figure 12-1 Example of an area zoomed on Port-au-Prince rendered on openstreetmap.nl The customized rendering set up by User:Ldp which shows damaged buildings and refugee camps mapped within OpenStreetMap using special GeoEye/DigitalGlobe imagery. See WikiProject Haiti#2010 Earthquake Response, http://wiki.openstreetmap.org/wiki/File:Haiti_earthquake_damage_map.png

The ODbL might be an obstacle or at least a reason to hesitate for an organization before using OSM for the publication and dissemination of its geographical names data or other geospatial data.

Generally, VGI and crowd-sourced data is likely to include valuable additional information in those

countries where well‑established geospatial sources are already available. The speed at which data can be gathered and compiled with VGI in times of crisis can be of importance in countries where no or few geospatial sources exist, although the accuracy and acceptability of the geographical names (for example) may be questionable in the long term. This VGI has at least the potential to provide a user's view of their geography. The supporting community states that more tailored public services and even more effectively targeted interventions by policy and decision‑makers could be possible. Very likely, VGI and crowd-sourced content will decrease cost, improve accuracy and increase availability of rich geospatial information. It is envisaged that progress will be made in future on bridging the gap between authoritative data and crowd-sourced data, moving towards true collaboration.

References

[1] UN-GGIM publication, 'Future trends in geospatial information management: the five to ten-year vision', chapter '4.1 Making mapping accessible to the masses', http://ggim.un.org/docs/Future-trends.pdf, last accessed 09/2016

[2] Online discussion forum of the Working Group on Toponymic Data Files and Gazetteers of UNGEGN, forum 1: 'Volunteered geographic information (VGI) / crowd-sourcing', https://wiki.gdi-de.org/display/wgtdfg/Forum+1+-+Volunteered+geographic+information+and+crowd-sourcing, last accessed 09/2016

[3] Introduction to OSM, http://wiki.openstreetmap.org/wiki/License/Use_Cases#Introduction, last accessed 09/2016

[4] Haiti earthquake response mapping project, http://wiki.openstreetmap.org/wiki/WikiProject_Haiti, last accessed 09/2016

Section 5 Progress of toponymic databases

Chapter 12 **Volunteered Geographic Information (VGI)**

Pier-Giorgio Zaccheddu

As methods of accessing geospatial information change and public interaction with such data becomes more dynamic, national mapping agencies and place-name standardization organizations are looking to take advantage of the willingness of the local population to provide information on location and associated attributes.

"The reduction in barriers to entry, the growth of Web and mobile mapping and the enthusiasm for crowd-sourced geospatial data have massively increased the role of the private sector and the volunteer community over the last decade. The fact that Google and Bing Maps are global brands highlights that it is the private sector that has made digital mapping accessible to the masses. Alongside the massive explosion in the use of geospatial information, 'Volunteered Geographic Information (VGI)' groups such as 'OpenStreetMap (OSM)' have gone some way towards popularising, albeit within a relatively niche community, the collection of geospatial data" [1].

For the utilization of crowd-sourced geospatial data, including geographical names, there is no single established approach [2]. The following information in this section is limited to the use of VGI for the publication and dissemination of geographical names data and more specifically, the use of one VGI community, the OpenStreetMap (OSM) community.

OSM data is subject to copyright and database rights, and is available under a license called Open Database License (ODbL). In a nutshell, ODbL is "share-alike license" which says that if one publishes works based on the OSM data, it must be attributed to OSM, and if the OSM data is enhanced with one's own data, the enhancements must be shared with the community. Organizations have the choice of doing this by making available a derived database, or at least by publishing the method (code) used to make it, or both. The two key concepts for legal purposes are 'produced works' (published works that use OSM in their production) and 'derivative databases' (data sets which include data which is derived from, based or dependent on OSM's data in any way). How the license applies to the use case depends largely on the purpose foreseen for the data. In all cases, the usage of OSM data is free of charge, and organizations are free to charge others for any enhancement work and/or set own licensing conditions, but any derivative database from OSM must be available, for free, to anyone who requests it. [3].

Often OSM is used to publish a map which shows data collected by others in a geographical context. It makes no difference whether this is on a web map, in a printed publication, as a graphic in a TV programme, or in any other media. Non-map examples include creating a spreadsheet or database that shows a public transport route along with information from timetables to illustrate arrival/departure times from certain places, or a mobile app to help users locate points of interest in the real world.

Related to the 12 January 2010 earthquake in Haiti, the OSM community created a map coverage of the worst hit cities of Port-au-Prince and Carrefour, and the neighbouring countryside thanks to the "Haiti earthquake response mapping project" [4]. Using the free data, OSM developers and others around the Web have created a set of earthquake map resources in addition to OSM's basic online map:

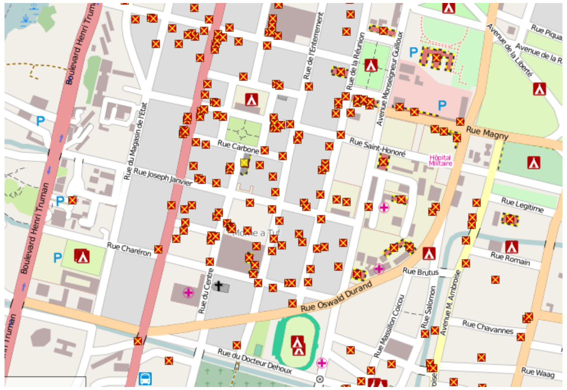


Figure 12-1 Example of an area zoomed on Port-au-Prince rendered on openstreetmap.nl The customized rendering set up by User:Ldp which shows damaged buildings and refugee camps mapped within OpenStreetMap using special GeoEye/DigitalGlobe imagery. See WikiProject Haiti#2010 Earthquake Response, http://wiki.openstreetmap.org/wiki/File:Haiti_earthquake_damage_map.png

The ODbL might be an obstacle or at least a reason to hesitate for an organization before using OSM for the publication and dissemination of its geographical names data or other geospatial data.

Generally, VGI and crowd-sourced data is likely to include valuable additional information in those

countries where well‐established geospatial sources are already available. The speed at which data can be gathered and compiled with VGI in times of crisis can be of importance in countries where no or few geospatial sources exist, although the accuracy and acceptability of the geographical names (for example) may be questionable in the long term. This VGI has at least the potential to provide a user's view of their geography. The supporting community states that more tailored public services and even more effectively targeted interventions by policy and decision‐makers could be possible. Very likely, VGI and crowd-sourced content will decrease cost, improve accuracy and increase availability of rich geospatial information. It is envisaged that progress will be made in future on bridging the gap between authoritative data and crowd-sourced data, moving towards true collaboration.

References

[1] UN-GGIM publication, 'Future trends in geospatial information management: the five to ten-year vision', chapter '4.1 Making mapping accessible to the masses', http://ggim.un.org/docs/Future-trends.pdf, last accessed 09/2016

[2] Online discussion forum of the Working Group on Toponymic Data Files and Gazetteers of UNGEGN, forum 1: 'Volunteered geographic information (VGI) / crowd-sourcing', https://wiki.gdi-de.org/display/wgtdfg/Forum+1+-+Volunteered+geographic+information+and+crowd-sourcing, last accessed 09/2016

[3] Introduction to OSM, http://wiki.openstreetmap.org/wiki/License/Use_Cases#Introduction, last accessed 09/2016

[4] Haiti earthquake response mapping project, http://wiki.openstreetmap.org/wiki/WikiProject_Haiti, last accessed 09/2016

Section 6 Technical issues: databases and pronunciation

Chapter 13 **Technical issues: database management**

Pier-Giorgio Zaccheddu

13.1 Introduction

After geographical names information has been gathered in the field, this information needs to be stored and prepared for dissemination. Several methods exist for storing the data, from the paper maps being an old-fashioned but proved medium, to the advanced digital techniques using databases.

For storage of names information in a digital format, one has several options, on the one hand text files or spreadsheets and on the other hand databases.

Some decisions have to be made on the used soft- and hardware side before starting the storage of the geographical names data in databases. It has to be considered that for proprietary database software solutions high license costs might need paying. In contrast, open source software products are free, but some maintenance and updating skills are required.

This section touches on the implications of the geographical names database management used for the output options, e.g. printing a gazetteer from the database or importing the database content into other systems, like Geographic Information Systems (GIS). The publication of geographical names through the internet is not part of the database management and will therefore not be explained in this chapter, but will be tackled in the following section 8.

13.2 Preliminary discussion on structure and content of toponymic databases and gazetteers

What does one have to consider (or to know) before one starts the creation of a database? The following information is a selection of possible issues to deliberate. This information is not intended to be exhaustive. It is envisaged to raise awareness for technical and organizational implications that may result from the decisions made.

For storage of names information in a digital format, one has several options:

- text files (such as Microsoft Word) or spreadsheets (such as Microsoft Excel)
- databases

Text files are easy to handle, but the methods have very limited capabilities in digital processing. Databases might be more complicated for non-skilled people, but the data can be connected with information stored in other databases (such as Geographic Information System repositories), and therefore it can be processed in many ways. That’s one of the major reasons to store data in databases.

The advantage of using a database management system (DBMS) is that it can be used to impose a logical, structured organization on the data. A DBMS delivers economy of scale for processing large amounts of data because it is optimized for such operations.

Apart from the technical software support which will be tackled in the following section III, the layout and the structure of the database tables and – which is much more important – the purpose of the database, has to be considered. Examples of purposes are:

- Names database to be used in a Geographic Information System (GIS) for map production
- Names database to be used for the publication of geographical names as gazetteer through the Web
- Names database to be used for Web applications support

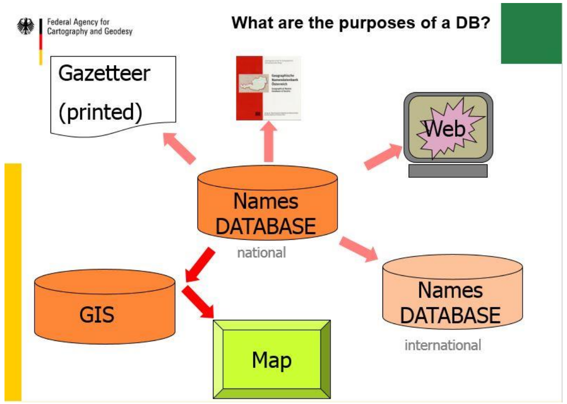


Figure 13-1 Purposes for geographical names databases

The database model design depends on the envisaged complexity of the geographical names content to be maintained, i.e. what object/features are to be captured, which attributes are associated to the object/feature and to the name? Therefore, the objects/features have to be selected and defined. The following object/feature types might be included: settlements, transport, vegetation, water bodies, relief areas, etc. This is related to the decision about the respective attributes associated with the objects/features. Examples are:

administrative division (statistical key-number), stream system (hydrological area code number), population size, status of the geographical name, language of the geographical name, height (height in metres above sea level), etc.

Within UNGEGN's 'Technical reference manual for the standardization of geographical names' a toponymic data transfer standard and format is mentioned in its second section, providing tables of Roman characters used in different languages, and a draft toponymic data exchange standard. The manual was published in 2006 [1]. The draft toponymic data exchange standard has been updated and adapted within several toponymic training courses from 2006 until today. A summary of the discussions can be displayed as follows:

Field Name	Data Type
RecordID	Index
UID	Number (long integer)
Name	Text (50 char.)
Latitude	Number (double)
Longitude	Number (double)
FeatureCode	Text (50 char.)
AdminUnit	Text (50 char.)
Language	Text (50 char.)
Description	Text (255 char.)
VariantName	Text (50 char.)
MapSheet	Number (long integer)
Source	Text (255 char.)
Status	Text (50 char.)
Pronunciation	OLE-Object
Location information	OLE-Object

Figure 13-2 UNGEGN Data model content (status: 2016)

The key for each database model is the inclusion of unique identifiers (UID) to uniquely identify each named feature.

Associated with the envisaged content of the geographical names database a feature classification has to be determined. Generally, a database should allow for different feature classifications to be used. Usually, a national feature classification is available and can be used for describing and classifying the objects/features in the database. A national feature classification recognizing the situation in the respective country and with feature code names in one (all) official language(s) as well as translations into other language(s) (e.g. English) would be helpful.

Sometimes regional views on and publication of geographical names data are needed, e.g. for regional projects or infrastructures. One good example is the European project "EuroGeoNames (EGN)" (funded by the European Commission from 2006 – 2009). Within EGN the "EGN feature classification" was developed because all other classifications available were deemed to be limited. However, the EGN classification was a compromise as well (the "highest common denominator") and it was satisfactory for the purposes for which it had been required, essentially, query filtering of the web service publishing the data of the different European databases. The EGN classification consisted of 8 classes and 27 sub-classes:

Code		*Feature Type*	*Short Definition*	*Feature Type Examples*
1		**COUNTRIES, ADMINISTRATIVE UNITS AND OTHER AREAS**	Countries, territorial units of a country for administrative purposes and other man-made areas	
	1.1	**Country**	Country of Europe	
	1.2	**Administrative units**	Territorial units of every country for statistics and administrative purposes *Including:*	
			Nomenclature of Territorial Units for Statistics in EU (NUTS 1, 2 and 3) Local Administrative Units (LAU 1 and 2)	länder (Germany) autonomous region (Spain) province municipality
			Other administrative units	
	1.3	**Other non-administrative units**	Other type of man-made areas like economic, cultural, linguistic or tourist areas	
2		**POPULATED PLACES**	Buildings for housing of any category like cities, towns, villages, etc.	
	2.1	**Administrative capitals**	Populated places with capital status *Including:*	
			Administrative capitals of NUTS 1, 2 and 3	capital of country, autonomous region (Spain), province
			Administrative capitals of LAU 1 and 2	capital of municipality
			Other administrative capitals	

Figure 13-3 EGN Feature Classification - extract

Another important item to be touched on in relation to the storage of geographical names data in a database is the definition of a metadata profile/scheme for explaining the content of the database to users, customers, etc. The metadata profile should be created according to common international standards. Many different metadata schemes are being developed as standards across disciplines, such as library science, education, archiving, e-commerce, and arts. Concerning geographical names data as part of geospatial data the ISO 19115:2003 Geographic information -- Metadata standard is one example of how to describe geographical information and associated services, including contents, spatial-temporal purchases, data quality, access and rights to use. It is maintained by the ISO/TC 211 committee.

13.3 Database design and management

A database management system (DBMS) is a system software for creating and managing databases. The DBMS provides users and programmers with a systematic way to create, retrieve, update and manage data. A DBMS makes it possible for end users to create, read, update and delete data in a database [2]. The DBMS essentially serves as an interface between the database and end users or application programs, ensuring that data is consistently organized and remains easily accessible. The DBMS manages three important things: the data, the database engine/tool that allows data to be accessed, stored and modified -- and the database schema, which defines the database's logical structure. These three foundational elements help to provide concurrency, security, data integrity and uniform administration procedures.

The DBMS is perhaps most useful for providing a centralized view of data that can be accessed by multiple users, from multiple locations, in a controlled manner. A DBMS can limit what data the end user sees, as well as how that end user can view the data, providing many views of a single database schema.

The DBMS can offer both logical and physical data-independence. That means it can protect users and applications from needing to know where data is stored or having to be concerned about changes to the physical structure of data (storage and hardware). As long as programs use the so-called application programming interface (API) for the database that is provided by the DBMS, developers won't have to modify programs just because changes have been made to the database. For relational DBMSs (RDBMSs), this API is SQL, a standard programming language for defining, protecting and accessing data in a RDBMS [2].

Using a DBMS to store and manage data comes with advantages. One of the biggest advantages of using a DBMS is that it lets end users and application programmers access and use the same data while managing data integrity. The latter is the maintenance of, and the assurance of the accuracy and consistency of data over its entire life-cycle. Data is better protected and maintained when it can be shared using a DBMS instead of creating new iterations of the same data stored in new files for every new application. The DBMS provides a central store of data that can be accessed by multiple users in a controlled manner.

If steps for designing a (geographical names) database were defined, the following ones should be included:

1) Determine the purpose of your database.
2) Determine the tables you need in the database.
3) Determine the fields you need in the tables.
4) Identify fields with unique values.
5) Determine the relationships between tables.
6) Refine your design.
7) Add data and create other database objects.

After the purpose of the geographical names database has been determined (step 1), the structure of the database becomes crucial. This means that the database schema or model, the tables and fields have to be selected and their relationships to be set (steps 2 – 5).

A geographical names database schema or model should allow flexibility. Anyway, the database schema should explain whether the geographical names will be used as an attribute of a spatial object/feature ("geospatial-based"), or whether the geographical names entries are the main database entries ("attribute-based"). In the geospatial community the first option is favoured, whereas the geographical names and language community very often model the databases while focusing on the names entries.

Concerning the latter option, the core table will hold one row for each geographical name. The definition of columns (fields) may vary considerably from country to country, but there are general rules that apply to most databases of geographical names. Typical fields associated with a geographical name are feature type, coordinates, variant names, textual description, source of the name information and status of the name. The following simple setup for rows and columns in the database explains the general rules:

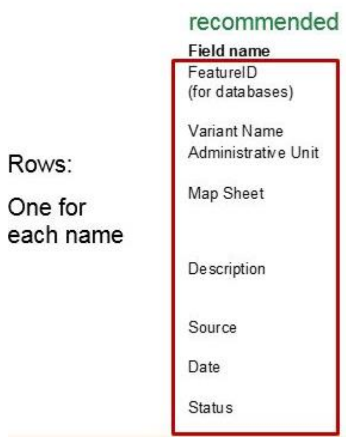


Figure 13-4 General rules for a simple database - rows

Columns:
attribute information to the name

Data type	Description
Integer	A unique identifier assigned to the name. This ID will be used to link the name with other database tables.
Text	Other names assigned to the feature, if any
Text	Name or Code of the Administrative Unit where the name is situated in.
Text	Reference to a map sheet in a topographic map series. The data type may be Integer if the sheet name contains only numbers no letters.
Text	Comments, e.g. on the history of the name, and verbal statements on the extension of the feature.
Text	Source of the name. e.g. captured in the field by interview
Date	Date of the entry to the Database. Other option: date of approval by the Board.
Text	Comment, e.g. name is approved or not approved by the Board.

Figure 13-5 - General rules for a simple database - columns

The possibility of linking more names with the same named place gives the opportunity to integrate minority languages and exonyms, which are an important contribution to multilingualism.

Within the European initiative “Infrastructure for Spatial Information in Europe – INSPIRE” the INSPIRE data specification on geographical names was prepared following the participative principle of a consensus building process, also involving UNGEGN experts. In INSPIRE the concept that the same place can be referred to by several names was reflected. In order to reflect this approach the central element of the INSPIRE geographical names data model is the spatial object “named place” that can carry one or more names. The specifications of geographical names can be used for modelling names in any other INSPIRE theme. Each named place has a unique INSPIRE identifier. It is further characterised by the eventual name(s), geometrical representation and if available, type, local type, indicative scale of usage, and the possibly related spatial objects. The latter helps to preserve consistency between data at different levels of detail. In addition, life-cycle information – i.e. when the named place has been inserted / changed, or eventually superseded / retired in the spatial data set – should be given if available [3].

Within any database design and management it shall be considered as well that geographical names data are used at all levels of resolution. The spatial resolution of a geographical names data set is typically described by the scale of the map where it has been captured from, or for which it has been captured.

The core of the INSPIRE geographical names application schema is described in figure 6 showing its non-voidable elements.

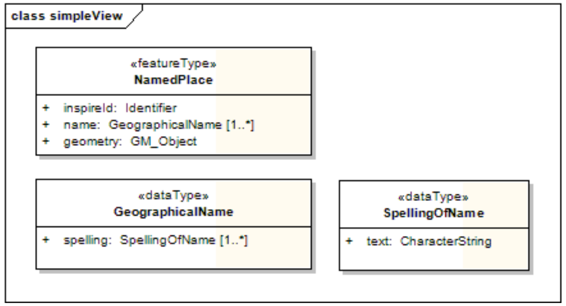


Figure 13-6 - core of the INSPIRE geographical names application schema

The only feature type of the schema is the feature type NamedPlace, representing any real world entity referred to by one or several proper nouns. Each NamedPlace is associated with one or several geographical names, i.e. proper nouns applied to the spatial object, modelled with the data type GeographicalName. The different geographical names of one given spatial object may be for example the names in different languages or in different forms (e.g. complete and short forms of country and administrative unit names). Each GeographicalName may have one or several spellings, i.e. proper ways of writing it, in one or several scripts like the Latin/Roman, Greek and Cyrillic scripts, modelled with the data type SpellingOfName. The following example explains how geographical names are modelled:

- The city of Athens may be modelled in the schema as one NamedPlace.
- The endonym “Athína” (Greek language) and exonym “Athens” (English language) are two different GeographicalName of this unique NamedPlace.
- “Αθηνα" (Greek script) and its standard romanisation "Athína" (Latin script) are two different SpellingOfName of the same GeographicalName “Athína”.

The following information shows one example for a characteristic which might be encountered when designing and managing the database. Different computer systems may cause and support different characteristics. For example, the operating system (OS) of the computer may cause displaying errors when geographical names data are exchanged and the settings are not known and not considered properly. An OS is a software program that manages the hardware and software resources of a computer. The OS performs basic tasks, such as controlling and allocating memory, prioritizing the processing of instructions, controlling input and output devices, facilitating networking, and managing files. The dominant desktop operating system is Microsoft Windows with a market share of around 85%. OS X by Apple Inc. is in second place (9%), and Linux is in third position (1.5%) [4].

The OS uses code pages, i.e. a table of values that describe the character set used for encoding a particular set of glyphs – i.e. graphic symbols that provide the appearance or form for a character –, usually combined with a number of control characters. The Unicode Standard is an effort to include all characters from previous code pages into a single character enumeration that can be used with a number of encoding schemes [5]. If the produced geographical names database is set with the Unicode character set and the receiving operating system supports the Unicode character set, no displaying errors should appear.

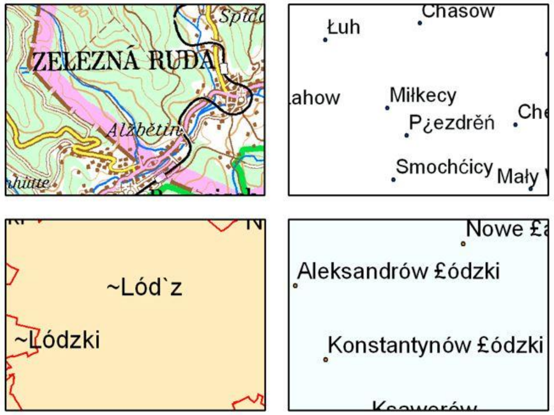


Figure 13-7 - Possible displaying errors of character sets in GIS

Today such displaying errors do appear rarely as the common data ETL (Extract Transform and Load) routines used by GIS or other applications have integrated the Unicode standard and are updated regularly.

13.4 Database open source options vs commercial options: criteria for selection

There is a huge choice of database management systems (DBMS), which includes commercial and open source database products or tools.

This section provides possible criteria for the selection of database software tools. The criteria are the result of an overall comparison of open source and commercial options. It is exemplary only and does not rely on a comprehensive scientific research. Therefore, the following criteria can be reflected and used for the main questions to be answered before extracting geographical names from the database and displaying them in a GIS, on a satellite image or in other applications.

Examples for commercial options are, amongst others, Microsoft Access and Oracle and for open source options MySQL. Microsoft Access is a pseudo-relational database management system from Microsoft that combines the relational Microsoft Jet Database Engine with a graphical user interface and software development tools. Microsoft Access stores data in its own format based on the Access Jet Database Engine. It can also import or link directly to data stored in other applications and databases [6]. MySQL "The world's most popular open source database is a relational database management system (RDBMS) that runs as a server providing multi-user access to a number of databases. The MySQL development project has made its source code available under the terms of the GNU General Public License, as well as under a variety of proprietary agreements. MySQL is a popular choice for using databases in web applications. PostgreSQL, often simply Postgres, is an open source object-relational database management system (ORDBMS). PostgreSQL is not controlled by any single company — a global community of developers and companies develops the system. PostGIS – as an extension for PostgreSQL – adds support for geographic objects to the PostgreSQL object-relational database. In effect, PostGIS "spatially enables" the PostgreSQL server, allowing it to be used as a backend spatial database for geographic information systems (GIS) [7].

Databases are meant to be interoperable, but they all have their own operational procedures and processes for storing data etc. It should be a last ditch choice to change the database, although there are sometimes good financial reasons for changing, such as licensing issues, acquisition or mergers.

A possible list of criteria for selection for both – commercial and open source tools – may be the following:

Requirement	Commercial	Open Source
Data abstraction and independence using the same data	Yes	Yes
The ability to swiftly recover from crashes and errors	Yes	Yes
Data security	Yes	Yes
A locking mechanism for concurrent access	Yes	Yes
An efficient handler to balance the needs of multiple applications	Yes	Yes

Robust data integrity capabilities	Yes	Yes
Logging and auditing of activity	Yes	Yes
Simple access using a standard application programming interface (API)	Yes	Yes
License costs	Yes, eventually very high	No
Technical support	Yes, but additional effort and usually not included	Yes, support by IT companies can be contracted

Figure 13-8 - Possible list of selection criteria

In a nutshell, the most relevant technical and organizational requirements seem to be fulfilled by commercial and open source database tools. Many organizations are turning to open source databases to reduce database management (licensing) costs and avoid supplier lock-in. The maturity of open source databases is at its highest level ever, with more choices, better support and comprehensive ecosystems. Open source DBMS products continue to improve in terms of functionality and scalability, and DBMS tool suppliers are beginning to provide support for these offers. Software companies do provide support to DBMS implementations.

Thus, the big advantage of open source software tools is the licensing cost savings, which might be significantly higher using commercial software tools. It is up to the organization to decide whether the licensing cost savings using open source software tools can be invested in capacity building of the employees or for subcontracts with private companies for providing substantial IT support. However, this support might be needed for commercial software tools as well.

If the geographical names database – be it proprietary or open source – will be imported into a Geographic Information Systems (GIS), the same decisions as for the database have to be made. The main purpose for doing this import is the map production, i.e. to link and overlay the geographical names data with other spatial (topographic) data.

The following information is the result of the overall comparison of most popular open source and commercial options. It is exemplary only and does not rely on a comprehensive scientific research. The most popular commercial GIS software product is Esri® ArcGIS®. It facilitates collaboration and lets one author data, maps, globes, and models on the desktop and serve them for use on a desktop, in a browser, or in the field, depending on the needs of one's organization [8].

ArcGIS support and educational services consist of technical maintenance programs, software releases and updates, technical support, online support services, publications, training, and consulting services. As an open source alternative Quantum GIS (QGIS) can be mentioned. QGIS provides data viewing, editing, and analysis capabilities. QGIS is a user friendly open source GIS licensed under the GNU General Public License. QGIS runs on different operating systems like Linux, Unix, Mac OSX, and Windows and supports numerous vector, raster, and database formats and functionalities [9].

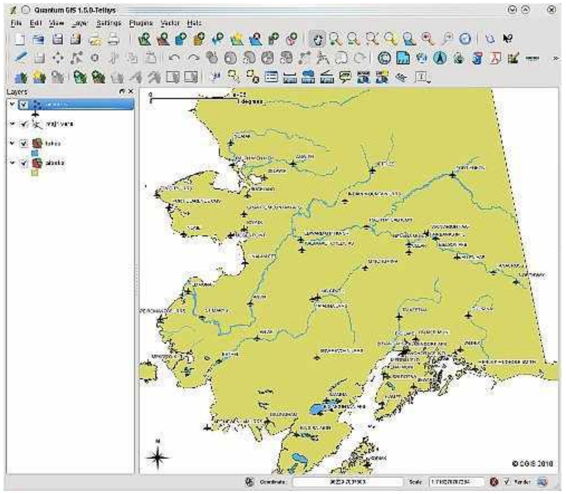


Figure 13-9 - Screenshot from Quantum GIS (QGIS)

Both, proprietary and open source GIS software products do fulfil the requirements for a high-quality spatial data analysis and map production. Again, the big advantage of open source software tools are the licensing cost savings.

13.5 References

[1] UNGEGN Technical reference manual for the standardization of geographical names, http://unstats.un.org/unsd/geoinfo/UNGEGN/publications.html, last accessed 09/2016

[2] What is a database management system (DBMS)? - Definition from WhatIs.com, www.whatis.com, last accessed 09/2016

[3] INSPIRE Data Specification for Geographical Names, http://inspire.ec.europa.eu/index.cfm/pageid/2, last accessed 09/2016

[4] Wikipedia (2016) Operating system, https://en.wikipedia.org/wiki/Operating_system, last accessed 09/2016

[5] Unicode Consortium, codepages, http://www.unicode.org/ , last accessed 09/2016

[6] Microsoft Access, Website: http://office.microsoft.com/en-us/access/ , last accessed 09/2016

[7] MySQL, Website: http://www.mysql.com/ and http://www.postgresql.org/ and http://postgis.refractions.net/ , last accessed 09/2016

[8] Esri® ArcGIS, Website: http://www.esri.com/software/arcgis/index.html, last accessed 09/2016

[9] Quantum GIS, Website http://www.qgis.org/en.html, last accessed 09/2016

Chapter 14 **Building a database for names management and index production for atlases**

Tjeerd Tichelaar

14.1 Introduction

The capstone of the commercial production of an atlas has for long been the compilation of an exhaustive index of geographical names. Falling outside the core discipline of the cartographer and typically subject to the toughest time pressure – ever occurring delays in the preceding cartographic process are rarely allowed to push backward deadlines imposed by the arrangements with the printing press – the production of the index of names was in pre-digital times known to be the most stressful, if not right-out dreaded part of the entire atlas production workflow.
Being both complicated and extremely labour-intensive, it was not uncommon for parts of the index production to be automated even before computers were called in to support the cartographic process. Whereas initial applications may have been limited to storage and sorting of names, it was the development of relational databases and SQL that would enable nearly complete automation of the editorial steps required for the production of a print-ready index of names.

14.2 The atlas and its index

The geographical contents of the interrelated maps constituting an atlas, labelled by geographical names, are unlocked in the average atlas by a number of search facilities. A list of contents is generally provided, along with a graphical index key to the maps. The most complete and detailed access portal to the objects presented in the maps is, however, the index of names. The names index of an atlas differs significantly from indexes provided in non-cartographic reference books and text books in general. Rather than to retrieve occurrences of text strings throughout the work, the entries in the atlas index are meant to localize named objects. The geographical name is not a concept one might be interested in for its own sake, but serves as a bridge to the object of our concern. We search the name to find the object it refers to. Once we localize it in geographical space, we have access to all the information pertaining to it in all the maps in which it occurs, which in the case of a school atlas or a national atlas may be a multitude of geographical and thematic maps at varying scales.

The consequence of the function of geographical names in the atlas, as explained above, is that we want just one index reference for each combination of object and name. As many names will occur in many different maps throughout the atlas, we must select the occurrence we believe to be most appropriate for localizing the object described by the name. This selection should be made in a consistent, systematic way, so that it is predictable to the user and automatable to the producer.

To effectively isolate the searched object, it may not always suffice to quote its name in the index: homonyms are quite common, and these must be made discernible. This can be done by adding object-specific information, which may be geographic (a country or region name) or generic (an object category or class).

The choices to be made to populate and structure the index of geographical names in an atlas and the requirements concerning data structure and definitions of the database employed can be illustrated by the example of the system employed for the Dutch high school atlas *De Grote Bosatlas* and its domestic and international spin-offs. This atlas has been published starting from the 19th century to support the geography curriculum at schools nationwide, with separate editions for primary schools and different levels of secondary education as well as foreign co-editions in Belgium, France, Denmark, Sweden, Switzerland and Canada. The cartographic bases of its constituting maps were reused wherever product specifications allowed. Production of maps and indexes was gradually digitized in the 1990s, and a strictly relational database for names management and index production was developed in 2004.

14.3 Index production: principles and rules

Traditionally, the Dutch school atlas presented two types of maps: oro-hydrographic overview maps providing a synoptic topographic reference, and thematic maps covering the themes dealt with in the contemporary geography curriculum. Due to the evolution of the demands of geography teaching over the last half century, the proportion of thematic maps was continuously increased. For each new edition, the thematic map content might be replaced, while the overview maps (the time-resistant 'backbone of the atlas'), would typically be updated and incidentally restyled. The geographical names in the overview maps were the first, and remained the most important to be listed in the index of names.
The compilation of the index of names for the Dutch school atlas has been given ample attention ever since it was first produced (1924). In its structure and composition, it follows long established conventions that were strictly maintained over the successive editions of the atlas. In the same way as a dictionary or a telephone directory, changes in composition or design can be expected to be unfavourably received by the customary

users: in this case the teachers in the schools. The rules to be followed are therefore meticulously defined.

These rules can be summarized as follows:

1. Every name-object combination (a name referring to a unique object) is represented in the index with one reference only. When a name-object combination occurs in the atlas more than once, the map referred to in the index is defined by a set of selection rules.
2. All name-object combinations occurring in the atlas are included in the index in a similar way and unabbreviated, regardless of whether in the maps referred to they are parenthesized (secondary names) or abbreviated.
3. The names are listed alphabetically, following the currently customary sequence of the alphabet of the language in which the atlas is published.
4. When a name consists of a specific element preceded by a generic element, the generic element, including prepositions where they occur, is inverted in the index. In other words: names are listed by their specific elements, irrespective of their position in the name. *Example*: Negro, Rio (a river name in which Rio is a generic element) vs. Rio de Janeiro (a city name, in spite of its etymological meaning, Rio is not a functional generic element here).
5. When a name consists of a specific element preceded by an article or a preposition, the article or preposition, is in the index inverted as well. (Example: Hague,The)
6. To differentiate between homonyms, regional and/or generic specifications are added to the names of homonymous objects. *Example*: Victoria (city, Canada) - Victoria (city, Seychelles) - Victoria (state).

Cross-references may be included in the index, for instance to accommodate for recent name changes or frequently occurring alternative writings. These refer to the map name using an =-sign; they are inserted into the alphabetical order of the regular names. *Example*: McKinley, Mount = Denali.

Index production: the selection of entries
The principles by which index references are selected in the Dutch school atlas can be summarized as follows:

1. Priority is given to the overview maps above thematic maps. As an exception, for country names priority is given to political continent maps.
2. Priority is given to the map showing the entire object (for instance in the case of rivers or mountain ranges).
3. Priority is given to the maps in which the named object is included in the map's subject area. The subject area is generally described by the map's title: New York is in the subject area of the maps titled 'United States' and 'North America', but outside the subject area of the map titled 'Canada', even if because of this map's incidental layout, the object is also displayed there. Where the map's title doesn't suffice, the subject area is explicitly defined internally by the editorial staff.

Priority is given to the map with the largest map scale, except in the case of country names (see rule 1).

14.4 A database for names management and index production

A well-designed relational database can meet both the demand of efficient index production and a more comprehensive objective to provide the ability to store and query all possible information on geographical names (names management). The rationale behind the latter is the observation that knowledge of or even affinity with toponymy cannot be expected to be widespread within the editorial staff responsible for the maintenance of the atlas, while individual editors might still be expected to answer questions about decisions on the rendering of geographical names taken in the past. In the Dutch case, the pseudo-institutional status the Bosatlas historically acquired in the Netherlands – by many it is considered the most trusted work of reference about foreign geography, including the confusing subject of name orthography – gives rise to a continuous stream of questions forwarded to the publisher on the contents of its maps, and the bulk of these involve geographical names. Moreover, when for whatever reason a new name has to be added in any of the existing maps – or a new map is to be made – the editors must have access to information about the principles (including transliteration keys) and sources the existing names were derived by.
From a practical point of view, the database should ideally meet the following conditions:

1. It should be object-oriented.
2. The data structure should be relational, and normalized to the greatest possible extent: in other words, every piece of information should occur only once in the database. Violations of this condition expose the database to contamination: errors and inconsistencies will be imminent.

3. It should be portable, in the sense that only commonly available market software should be used.
4. The database should be able to communicate with other cartographic databases that may be employed to produce both paper and digital maps. Meanwhile, the names and indexing database should remain independent.
5. Its maintenance should require no special technical skills.
6. Data safety should be sufficient.
7. The character coding should be Unicode.
8. The result of the index production process should be an output file (txt) holding all the information required for automatic page layout with commonly available DTP software.

The index production application should incorporate the principles and selection procedures we choose to follow when preparing our index. It can be built into the database by adding a script and all the required queries to the necessary tables.

14.5 The data structure: core tables

The core of the database should consist of two separate tables, one holding the names and another the objects. A name can, in this definition, refer to only one object: in other words, homonyms are considered different names. An object, on the other hand, can be referred to by different names. The table NAME consists of a key field (name identifier) and a number of attributes directly dependent of the name. The table OBJECT consists of an object identifier and a number of attributes directly dependent of the object.

To allow the index production application to collect the information it requires, the database must include tables containing relevant data about the atlas, the maps it includes and the position of these maps within the atlas, and the names and their location (grid reference) occurring in the maps:

14.6 What is a name?

NAME	
Field	*Data Type*
Name identifier*	Number
Object identifier*	Number
Language identifier	Number
Script identifier	Number
Orthography identifier	Number
Name (main specific element) *	Text
Non-specific name element identifier*	Number
Grammatical fields (gender, number etc.)	Number
Is the name an official name?	Yes/No
Is the name a primary name?	Yes/No
Does the name contain a generic element?	Yes/No
Is this a long /complete version of the name?	Yes/No
Is this a short /abbreviated version of the name?	Yes/No
Is this a redundant or historical name?	Yes/No
Source identifier	Number
Adjective (masculine singular) form of the name	Text
Pronunciation (IPA notation)	Text
Syllable of primary emphasis	Number

Table 14-1 the NAME table

The table NAME could contain the following fields (fields marked * are mandatory, fields in bold type are required for the index production process).The table contains one row for every geographical name we recognize. Its key attribute is a name identifier. Note, that the attributes collected for every name presuppose a subjective definition of 'name'. A name in the context of the database is a *written* code referring to one single geographical object. This may be self-evident in the light of the purpose of the database, which is to help produce a printed atlas, but it is neither the only possible nor the original definition of a name. As writing is a relatively new addition to the systems of oral communication we refer to as languages, the original description of a name would be 'a sequence of sounds conventionally used to refer to a single person or object' or, in database terms, '... a single instance of a certain entity-type'.

The decision to define 'name' as a written instead of a sonic code has far-reaching consequences, some of which may be considered undesirable. It implies, for instance, that different writings/alphabet versions or transliterations of a single name are different names in database-terms, even when they refer to the same object and their pronunciation is per definition the same. This decision thus leads to imperfect normalization, as some attribute values (for instance Pronunciation) are shared between different written versions of the name: it will be a challenging effort to address this imperfection, by grouping together those written forms of a name that are essentially (pronunciation-wise) the same. On the other hand, the French name Paris for the capital of France (IPA: paˈʁi) and the English name for the same object (IPA: ˈpærɨs), which happen to be similar in writing but are certainly not so in pronunciation, are to the database one and the same name. The addition of the attribute holding the name’s pronunciation in IPA notation, makes it an obvious step to consider ‘Paris’ in English and ‘Paris’ in French two different instances of the entity type Name, in which case the language of the first is English, not French. In other words, we would consider Paris in English a sonic exonym to the homonymous French endonym in its written form. This is a matter of case-to-case definition: by defining Paris in its English pronunciation a name in its own right, we recognize it as being by convention part of the English language, and not merely a mispronounced French name. Just as is the case with written exonyms, the number of recognized sonic exonyms should preferably be confined to some well-known names in frequent use in the host language.

The question of whether an exonym in a certain language should be defined as a Name in its own right not just involves names that in their written form are homonymous to the endonym. Sometimes non-homonymous exonyms are in their written form borrowed from another than the locally official language. A representative example of this kind of situation is the French name for the Italian city of Firenze (endonym). The French exonym Florence is borrowed by several other languages, among them English and Dutch. The pronunciation of the English name (IPA: /ˈflɒrəns/) clearly differs from the French one, so English Florence and French Florence might each be considered separate exonyms. The Dutch pronunciation however sticks to the Original French one, so in this case it is debatable whether Dutch Florence is to be considered a Dutch exonym and thus a separate Name in the database, or the Dutch language merely borrows the French name. Whether or not a graphically homonymous and synonymous name should be considered a separate name, might be decided after the rules of the normalized database. These define that it can only be the same name if the values of all attributes are identical. A typically discriminating attribute would in this example be Adjective. In the case of Florence, its French adjective is ‘florentin’. The English adjective is ‘Florentine’. Dutch uses an adjective form ‘Florentijns’. The value of this attribute thence requires Dutch Florence to be included in this database as a name separate from the original French exonym.

Taking this reasoning a step further, local or any other viable pronunciation variants deviating from the standard pronunciation (if existent) of the name in the official language might likewise be included in the database as separate name. The name(s) of the American city of New Orleans, in ‘standard’ American English pronounced /nuː ɔrˈliːnz/ or /nuː ˈɔrliənz/ but locally /nuː ˈɔrlənz/ or more commonly /ˈnɔːlənz/ (IPA) sets a fine example, those of the Brazilian city of Rio de Janeiro (English pronunciation /ˈriːoʊ deɪ ʒəˈnɛəroʊ/ or /ˈriːoʊ deɪ dʒəˈnɛəroʊ/, in modern European Portuguese /ˈʁi.u dʒi ʒaˈnejɾu/, but in the local Carioca dialect of Brazilian Portuguese /xiudʒiʒaˈneɪ̯ɾu/) another.

Generic elements

A field defining whether or not a name contains a functional generic element will be required by the index production application if for a certain atlas we choose to add a generic specification to all names in index entries where this might be useful. If such a specification is textual instead of graphical (an icon), it will be redundant and hence undesirable to do this where the name itself already shows which generic category the object belongs to: for instance ‘Vänern *(lake)*’ makes sense, but in ‘Michigan, Lake (lake)’ the addition is redundant, as well as ‘Bodensee (lake)’ – the latter presuming that the meaning of See in the German language is familiar to the public, or is explained in a list of geographical terms.

Grammatical fields

Some grammatical details may be useful to store in the table NAME as well. Although they will rarely be relevant to the rendering of the name in the map or in an index, they do become relevant when names are used in written text or oral discourse: for instance, when they carry a declinable adjective or an article, or when it needs to be inflected when occurring in a particular grammatical case. For instance, the possibility to store grammatical number (singular, plural) and grammatical gender (masculine, neutral, feminine) can be easily accounted for.

In the case of grammatical number, geographical names incidentally behave in a way not matching the elements they were etymologically constructed from, even when these elements were left unchanged, are clearly recognizable yet, and call for the use of a proper article in speech: ‘United States’ and ‘Netherlands’, for instance, are plural names that in English usage in speech, but not in the map, go with the article ‘The’ – their French equivalents carry the plural article Les (États-Unis, Pays-Bas). In English speech, however, they are treated as singular words whenever they carry verbs: ‘the United States is a big country in America, the Netherlands is a small country in Europe’. Such knowledge could be stored in the database.

Much relevant grammatical information is language-specific, for instance lexical roots of names in languages that practice inflexion depending on grammatical case (Latin, Lithuanian), or definite-indefinite distinctions in languages like Turkish, Romanian and Albanian. An inventory of such issues per (relevant) language would be required to decide which additional attributes might be desirable to complement the database.

Adjectives

Adjective forms of geographical names are in many languages different from the nouns. Quite often they are not formed in a systematic way, sometimes they are even derived from predecessors of the current name or a different language version (e.g. Latin in Romance languages). They are often historically defined and not seldom influenced by non-official regional or local language. Homonymous names may thus carry different adjectives: the adjective going with the French city name of Beaumont in the department Manche, for instance, is 'beaumontais', while the city of Beaumont in the department Vienne carries the adjective 'beaumontois'. The adjective of Beaumont-en-Véron is 'bellimontois', while what belongs to Beaumont-sur-Sarthe is known as 'belmontais'. Knowledge of the adjective forms of a geographical name may be relevant if the name is to be used in prosaic texts, and adjectives may also be used to make demonyms.

Pronunciation

IPA notation (or, if the database format can accommodate this, an audio file) and accents are useful to know when names need to be pronounced. If more than one pronunciation is encountered, which in real life occurs more often than not, a decision has to be made whether this should be allowed for in the database. If pronunciation is not officially standardized, and it is possible to store more than one IPA string in this field, a case per case decision must be made whether or not we are still talking of the same name.

Syllabic stress (accent) may or may not follow systematic rules in a language. In Finnish, the stress (primary stress in the case of compound names) always falls on the first syllable of the word. In Italian, it is commonly the penultimate syllable which receives the stress, but under the influence of regional languages there are many exceptions to this rule. In some languages (for instance Spanish), diacritics are mandatory to mark syllabic stress not following the general rule, in others (Italian) they are optional and most commonly not used. In the database, a numerical field may be reserved to store information about the syllable receiving the (primary) stress.

Additional name attributes for names selection

Some other attributes included in the exemplary table NAME are not required for index production either, but may be useful for the selection of the right name versions as the map is compiled or revised: for instance, language, script, orthography, and the 'flags' (on/off-switches, stored as Boolean) official name, primary name etc.

Specific and non-specific name elements

If we want to be able to invert generic elements and/or articles and prepositions before alphabetizing the index output for certain atlases, these should be stored in a separate table in those cases where they precede the specific element of a name. In Dutch or French school atlases, for instance, the river name Río de la Plata is rendered as Plata, Río de la – while the town name Rio de Janeiro is simply rendered as Rio de Janeiro, because in this name the element Rio is not a functioning generic element even when it was derived from one. In Dutch school atlas indexes, articles are inverted as well, but in some Dutch-language Belgian primary school atlases they are not. The table holding the non-specific name elements should contain the following fields:

NON-SPECIFIC NAME ELEMENT	
Field	*Data Type*
Name element identifier*	Number
Name element (description)*	Text
Element type identifier	Number

Table 14-2 Non-specific name elements

The element type identifier should connect to a small table listing the options, for instance 'Generic', 'Article' and 'Article connected': the latter needs to be discernible by the indexing application because in some cases the element should not be separated from the rest of the name by a blank if it is not inverted in print (examples: elements ending in a hyphen or an apostrophe like the article al- in transliterated Arabic names or the contracted article l' in French; suffixes in Irish or transliterated Hebrew directly connected to the specific name like An t or Ha).

Language

According to the principles of the database, every Name belongs to a Language. The definition of Language in the sense of the database is not necessarily merely linguistic-taxonomical: a Language here may be a set of names belonging to a certain language or dialect in a historical context. The instances of the entity-type Language must suffice to store any geographical name we get hold of, allowing also for a lack of specific knowledge where it occurs. So, the table may include both very specific dialects like 'North Straits Salish (Lummi)', more general categorical instances like 'Occitan (local dialect)', 'Pre-

Columbian Native American (unspecified)' or 'British Indian (unspecified)', and even 'unknown'.

LANGUAGE	
Field	*Data Type*
Language identifier*	Number
Sorting system*	Number
Language (description)	Text
ISO-code fields	Text
Year of origin	Number
Year of end	Number

Table 14-3 Language table

The table LANGUAGE allows us to store information we might require when selecting the names that should appear in a map: the more we know, the better. It also plays a role in index production, as next to the table NAME, the table ATLAS also refers to it in connection with the addition of geographical or generic specifications. Moreover, the language of the atlas must be known in order to select the alphabetization rules applied in the output phase (sorting system), as these are language-specific. In Swedish indexes for instance, Å, Ä and Ö are inserted into the alphabet as separate letters after Z, while Ü is considered the same as Y and W the same as V; in the Danish alphabet, the sequence is Z-Æ-Ö-Å.

ISO codes of the languages and dialects may be included to facilitate communication with external parties. Data about the historical range of earlier language phases (the approximate years of origin and end according to the definition we apply) may be useful in the editorial phase of the compilation of a historical maps series, to avoid anachronistic names.

Name vs. Language

The starting point that any name can be assigned to one single language, as commanded by the data model of the database, poses theoretical problems that must be addressed in an unambiguous way. Debatable as it may be, for the practical use the database will be maintained for, the relationship should for each single name be explicitly defined by the Language and Orthography attributes assigned. Some examples are listed below to demonstrate both problems and solutions.

The language of the names of the American cities of Los Angeles, California (originally Spanish) and Terre Haute, Indiana (originally French) is English. They are the names by which the speakers of the official language of the United States of America, English, refer to them. In spite of their clearly recognizable Spanish, resp. French origin, in spite also of the fact that many inhabitants of the city of Los Angeles are till this day Hispanic, the English rules of pronunciation apply to them (IPA: lɒs ˈændʒəˌləs). The original Spanish name now functions as an exonym, insofar as Spanish is not considered a local language in the town: in the database, it will be a separate name (IPA: los ˈaŋxeles). In writing it differs by the acute accent on the A. Another difference is, that the element Los in

Name identifier	Object identifier	Language	Script	Orthography	Name (main specific element)	Non-specific name element
14282	79150	English (Modern)	Roman		Los Angeles	
14282	79150	Spanish (Castilian)	Roman		Ángeles	Los

Figure 14-1

the Spanish name is an article that might be inverted in the index, while in the English name it no longer functions as such (see figure 14-1): Likewise, the name by which the Romans referred to the settlement growing around their army camp where at present the Dutch city of Nijmegen stands, Noviomagus, may in the database be defined as Classical Latin because it was quoted by this name by sources written in Classical Latin language (figure 14-2):

Name identifier	Object identifier	Language	Script	Orthography	Name (main specific element)
206038	232601	Latin (Classical)	Roman		Noviomagus

Figure 14-2

Or it might, as historical linguists might advise, be defined as a Celtic name rendered in a Roman Latinised way (figure 14-3). The latter can be specified in the field 'Transliteration':

Name identifier	Object identifier	Language	Script	Orthography	Name (main specific element)
206038	132601	Celtic (Gallic-Transalpine Gaulish)	Roman	Roman Latinised	Noviomagus

Figure 14-3 language definition

Script

Unicode fonts are available to store and visualize names in all known writing systems. The description of the script should be stored in a separate table. Again, the instances maintained should accommodate for any name we encounter, also when specific knowledge as yet falls short. The table may consist of just a key field and a textual description. The purpose of including different script versions of names in the database, even when the atlases we produce won't show them in print, is that it allows us to store the original writings of names that we transliterate: we might need these in case official of UNGEGN-promoted transliteration keys are replaced and we need to re-transliterate.

Orthography

The definition of Orthography in the context of the database may include both transliteration, transcription, orthographic standards and optional or unofficially adapted variants like accentuated, vocalized or simplified spellings. Many instances may be unspecified and possibly unofficial/non-standardized transliterations and transcriptions matching the pronunciation-to-writing conventions of a certain language, but ISO-norms and transliterations recommended by the UNGEGN Working Group on Romanization should be included as well. An instance 'unknown' may be used temporarily until knowledge is gained. To avoid confusion, the target script may be included as an attribute.

ORTHOGRAPHY	
Field	*Data Type*
Orthography identifier*	Number
Orthography (description)*	Text
Target script identifier	Number

Table 14-4 Orthography table

Objects

The separation of attributes dependent of objects from those dependent of names is the most important step of normalization to be applied to the database. The indexing application will require, object attributes to add generic or geographical specifications to index entries in the case of homonyms. Regional identifiers specifying the country, state/province and possibly lower administrative units in which the object is located should connect to separate tables holding the descriptions and other relevant data about these units. Other attributes might be useful for specific index production purposes or other applications, for instance an electronic atlas (coordinates), cartographic presentation (capital underlining, representation of highest elevations per country or state), or optional additions to index output.

OBJECT	
Field Name	*Data Type*
Object identifier*	Number
Display name identifier	Number
Object type identifier*	Number
Regional identifiers*	Number
Latitude and longitude fields	Number
Years of object origin and end	Number
Elevation	Number
Flags required for printing the names in the map	Yes/No

Table 14-5 Object table

For the ease of database maintenance, one of the names under which the object is known may be selected from the Name table, so that we don't need additional queries to know what objects we are handling with.

A separate table listing the recognized object types should allow us to assign each object instance to an object class or category, like for instance 'populated place', 'independent first order area', different types of second, third and fourth order administrative areas, mountains, lakes, rivers, etc. When building the indexing application, the level of detail should suffice to keep homonymous objects apart in the indexes.

It should be noted that there is no universal way to cut up geographical reality into object categories. If the system must be able to produce atlas indexes for atlases in different languages and for different markets, the list of object types should therefore accommodate for distinctions made in each of these. *Example*: for specifying additions to index entries in some French language atlases, a distinction may be required between streams discharging into another stream or a lake (French: *rivière*) and streams discharging into an ocean or sea (French: *fleuve*), a distinction English or Dutch practice does not make. Likewise, distinctions between certain kinds of water bodies (straits, sounds, channels, bays, gulfs, creeks, coves) or mountain features (mountains, hills, rocks, peaks, mountain ranges, massifs, ridges) are language and society-dependent.

Objects vs. Names

One geographical object can be referred to by many different names. Different languages, scripts, transcriptions and transliterations, long and short forms, official and colloquial, old and new names may all be applied to one single object. In some cases, objects are subjectively defined by names: this holds, for instance, for the complete object class of streams as well as those of open water bodies.

14.7 Localizing the names

A table LOCATION is required to bring together the name, object and all relevant data connected to them with the name's grid reference in the map and other map-specific data:

LOCATION	
Field	*Data Type*
Location identifier*	Number
Name identifier*	Number
Map identifier*	Number
Does the named object occur in the map's subject area? *	On/Off
Does the name figure as a primary name?	On/Off
Grid reference fields*	Number and Text

Table 14-6 LOCATION table

This is the longest but also narrowest table of the database: it contains one row for every single name occurrence in every single map in every single atlas we produce, but apart from the usually one-letter grid references it includes no text strings. Essential to the names selection procedure of the indexing system is the field in which we define whether the name is part of the subject area (the main area of focus of the map) or not. In a map titled Indonesia, for instance, the names of all objects within the national territory of Indonesia are within the subject area, but the Malaysian capital Kuala Lumpur is not. In a map titled South-East Asia, the names of all objects located in South-East Asian countries are in the subject area. To qualify for index reference, maps in which the named object's position is within the subject area should receive priority above maps in which it is not.

Grid reference data may be stored in different fields containing the letters and numbers of either the grid section in which we find the first letter of the name, the grid sections where we find the first and the last letters of the name, or the grid sections that hold both ends of the named object. In the last case, coordinate information from the OBJECT table may also be used. Usually the location of the name is preferred because it can easier be found in the map visually, but more importantly because it is not always obvious where to logically locate beginning and end of areal objects.

The question whether a name is primary or secondary in a map is relevant in the cartographic process only, to decide whether a name should be parenthesized or printed in smaller type. If we choose to apply this field in a more sophisticated way, we could make it numerical and distinguish whether a name is primary, an exonym, a secondary, tertiary or quaternary name, whether it is preferably translated or not, etc.

14.8 The Atlas table

ATLAS	
Field	*Data Type*
Atlas identifier*	Number
Language identifier*	Number
Composition identifier*	Number
Atlas name	Text
Meta-information fields	Text

Table 14-7 ATLAS table

Of the atlas-dependent information that needs to be stored, the index production application requires the language of publication to select the proper text strings to be added to names in case of homonyms, as well as index composition details. The latter entail specifications about the types of name elements that will be inverted in the index, and whether or not both inverted and non-inverted index references will be included. For maintenance purposes, the name or title of the atlas and other meta-data (publisher, year of publication, edition number etc.) may be included.

14.9 Maps and Products

When a map is included in an atlas, we consider it a Product: within the atlas, it receives a title and possibly a subtitle, one or two-page numbers and possibly a map letter or number. Once we make clear for which atlas an index should be produced, the very first action of the index production application will be to consult the table PRODUCT and collect all the maps composing the atlas. The same table is also required to add page numbers and letters to the index entries in the end.

The MAP and PRODUCT tables should contain the following fields:

MAP	
Field	*Data Type*
Map identifier*	Number
Map description*	Number
Map type identifier*	Number
Map scale*	Number

Table 14-8 MAP table

PRODUCT	
Field	*Data Type*
Product identifier*	Number
Map identifier*	Number
Atlas identifier*	Number
Product title*	Text
Product subtitle	Text
Page reference fields*	Number
Map letter or number	Text or Number

Table 14-9 PRODUCT table

These tables are required to extract the name locations of all the names included in the atlas for which the index is being produced. The map type identifier refers to one of a set of distinguished map types, the descriptions of which are stored in a separate table MAP TYPE. It is required if in the selection procedure of our indexing system priority is given to occurrences in certain types of maps (in the Dutch case: the oro-hydrographic survey maps) above occurrences in other maps (for instance thematic maps).

The map's scale can be most simply stored as the denominator of the fraction of 1, at prevailing atlas map scales divided by 1000 (for instance: 6000 at a scale of 1/6,000,000). This value too is essential in the selection process of index production, where for the names of most categories of objects priority is given to the largest scale if the name occurs within the subject area of several maps of the same type.

14.10 Specifier Tables

SPECIFIER TABLES	
Field	*Data Type*
Regional or thematic addition identifier*	Number
Country, administrative unit or object type identifier *	Number
Language identifier*	Number
Text string to be added*	Text

Table 14-10 Specifier table

In order to allow the index production application to select the proper textual additions to homonyms, the database should contain one table for generic (object type) specifications, and several for geographical (regional) specifications.

14.11 Additional tables not used for index production application

Apart from the tables required by the indexing process, the database can be augmented with a number of tables to store information used in the editorial stage of cartography, or for referential purposes only. Dated population figures can be kept in a table that may include several rows per object, for different years and different object definitions (cities proper vs. metropolitan areas): these are required when population size determines the map symbol or type font or size cities and towns are represented with in the map. Object or other identifiers connect these tables to the core tables of the database.

The biggest advantage of maintaining a maximally normalized relational database for names management in an atlas producing environment is, that every piece of information having any relevance to the indexing or editorial process is stored and can be mutated at one place only within the structure of interconnected database tables, which reduces the chance of error because of unintentionally introduced inconsistencies to a minimum.

14.12 Reference:

Rennau, Günter – Register von Karten und Atlanten. Gotha: Haack 1976

Section 6 Technical issues – Data bases and pronunciation

Chapter 15 **UNGEGN involvement in dealing with pronunciation aspects of geographical names**

Tjeerd Tichelaar

15.1 Pronunciation and names standardization

A little over half a century ago, the United Nations Group of Experts on Geographical Names (UNGEGN) was instituted in order to promote the standardization of geographical names, both on national levels and – making it a self-evident UN concern – in international communication. Although this was not explicitly expressed, the standardization effort was practically directed at the *written* form of names, as it was a problem of written communication the expert group was meant to address.

Because names are linguistically non-generic nouns and thus, as all words, essentially sequences of sounds rather than graphic characters, the standardization pursued by UNGEGN cannot intend to 'freeze' them for eternity: both speech and writing conventions are dynamic, they evolve and change over time. Shakespeare's 16th century English we now find hard to understand, and likewise the names in the maps of his contemporaries Ortelius and Blaeu are no longer the names we put in our maps today. The standardization endeavoured by UNGEGN, in other words, is not a one-time job but a process that cannot be expected ever to end. As long as our nations unite, the relevance of the expert group's mandate will not decrease.

Although it is the graphical expression of geographical names which is for practical reasons the central concern of UNGEGN, pronunciation aspects both underlying and interacting with the written forms may not be ignored when dealing with names. Hence the creation of the Working Group on Pronunciation in response to resolution no. 11 of the Eighth United Nations Conference on the Standardization of Geographical Names (Berlin 2002). As standardization of pronunciation is rarely attempted by national governments worldwide and, admittedly, way over the head of any international expert group whether or not it operates under the umbrella of UN – its very desirability being debatable, in this respect – the proper allocation of the subject of pronunciation within the focal range of UNGEGN deserves some discussion yet. The following is meant to support the debate.

15.2 Written vs. spoken language

Of the current 192 UN member states, the nationwide official languages of 190 exclusively use so- called *phonographic* writing systems. The written form of these languages essentially represents their pronunciation, as opposed to languages using *logographic* writing systems, in which the writing directly refers to the meaning of what is written. The latter are nowadays still in use for Chinese and (partly) Japanese.

Although phonographically written words and names are thus supposed to reflect their pronunciation, the relationship between writing and pronunciation is language-specific and in many languages far from straightforward. Similar letters and combinations of letters or letters combined with diacritical marks are pronounced differently in different languages. Even within a single language they may be pronounced in more than one way, depending on their position within a word or the sounds surrounding them, or, especially in the case of names, by historical influences the writing itself does not disclose. The letter combination *ough* in English, for instance, may be pronounced in multiple ways: cf. r*ough*, thr*ough*, th*ough*, th*ough*t, or, to list some English geographical names, Br*ough* (IPA notation: ʌf), *Ough*tibridge (uː) and Scarbor*ough* (ə) in Yorkshire (but Scarbor*ough* in Ontario: oʊ), Sl*ough* in Berkshire and L*ough*ton in Essex (aʊ), Br*ough*ton in Buckinghamshire (ɔː), St*ough*ton in Massachusetts (oʊ), L*ough*or in Wales (ʌx), Cl*ough* (ɒk), Cloughey (ɔːx) and Kill*ough* (ɒx) in Northern Ireland, Y*ough*al in Ireland (ɒh), G*ough* Island in the South Atlantic Ocean (ɒf). Sometimes this 4-letter combination is even pronounced differently within one name, as in L*ough*bor*ough* in Leicestershire (ʌ and ə). Likewise, the letter combination *augh* is pronounced differently in the names of three (written) homonymous places Cl*augh*ton, all located within miles from each other: Cl*augh*ton in the City of Lancaster (æf) and Cl*augh*ton in the Borough of Wyre (aɪ), both in Lancashire, and Cl*augh*ton in neighboring Merseyside (ɔː).

Figure 15-1 The British Isles: many ways to say ough (lines connect similar pronunciation)

The twelve different pronunciations of *ough* in English geographical names are obviously an extreme example. Although many languages do maintain a more systematic correspondence between writing and pronunciation than English, it is quite common for languages to either use the same character for several different sounds, or apply letters, diacritical marks and combinations thereof in a language-specific way to accommodate at least the meaningful sounds the language discerns. The reason for this is, that the writing systems applied for languages were more often adapted than specially created for the language employing them. Writing systems typically spread in the same way most technological innovations do: borrowed at first from foreign creators, then gradually adapted to the specific requirements of the borrowers – in this case the borrowing languages.

In the case of the writing systems called alphabets, officially applied now by 158 UN member states, a complicating factor is that the ancient Phoenician script all these systems ultimately trace back to, was a so-called *abjad* rather than an alphabet itself: a script representing consonants only. This must have sufficed for the purposes this script was originally devised for, which may have involved the administrative identification of a limited number of generally known objects and geographical names. The widely travelling Phoenician merchants undoubtedly needed to write down names that were foreign to them, and thus lacked the meaning allowing them to be written down in the logographic script of the time. Nevertheless, the letters they devised represented the consonants of their own Canaanite language, to which foreign sounds were equated in accordance with what the Phoenicians believed to hear.

At this point, it is instructive to realize that of the numerous sound distinctions human beings are physically able to make, communities sharing a language typically use a limited number only to communicate. The sounds they set apart by such (to them) meaningful distinctions are called *phonemes*.

Phonemes are defined by inherited consensus within the community of speakers of the language. Every language thus possesses its own specific set of phonemes. The members of a language community develop sensitivity towards their own phonemic sound distinctions (the sound distinctions meaningful to them), and are simultaneously trained to ignore any other distinctions that might be heard. People speaking different languages don’t just fail to understand each other’s words: they neither recognize each other’s phonemes, to a level that they may believe they don’t hear the difference between all of each other’s sounds. This mechanism is nicely demonstrated through the word by which ancient Greeks generalized all non-Greek speakers: these people, according to their judgement, did not really speak a language but produced ‘bar-bar-bar’-sounds instead (i.e.: sounds that to Greek ears all sounded the same). This habit reduced them to ‘barbarians’, a brand of people occupying a lower step of civilization. Similar references were made by foreigners in later times to indigenous people of northern (‘berbers’) and southern (‘hottentots’) Africa.

Ethically speaking, most of us will currently agree that such appellations expose an intolerable degree of ignorance and indifference on the side of the name-givers, but actually it is an important quality to be insensitive to the sounds of others in order to be able to understand one’s own.

15.3 Roman alphabets

The Roman alphabets applied internally by 75% of the UN member states today and for transliteration purposes by many others were all derived from the alphabet the Romans once optimized to represent the phonemes of their own Latin language. The Roman script was an adaptation of an Etruscan predecessor, which had been adapted again from the alphabet of the Greek city state of Chalcis, an early adaptor of the phonograms of its Phoenician trade partners in the first half of the first millennium BCE. By adding vowel signs, the Chalcidians (and other Greeks) had turned the Phoenician abjad into a proper alphabet.

The fact that the Greek phonetic system differed to a considerable degree from the Phoenician, the Etruscan from the Greek and the Latin from the Etruscan meant that at every adoption some letters were considered superfluous or redundant, while for some phonemic sounds the source script lacked the letters required. This problem was incidentally resolved by inserting characters taken from a different script (for instance the runic Þ in Gothic, Old English, Old Norse and Icelandic), but more often by making digraphs and ligatures (Ll in Welsh, IJ in Dutch, Æ in Danish, Œ in French, and ß in German – the latter initially used as a typographic ligature of 'long s' ſ and 'short s' z or 'round s' s) and/or combining letters with diacritical marks (Ð in Old Irish and again old Germanic alphabets and Icelandic, Ł in Polish, letters with acute, grave and circumflex accents etc.). In some languages, like German, French and also English (where this is the exclusive resolve), combinations of two to four letters were customarily applied to represent phonemes non-existent in the Roman Latin source of their alphabet. As a result of such local solutions to similar problems invented all over the world over a period of roughly a century and a half, the world now knows dozens of different Roman alphabets, between which the sound values of many letters and diacritic signs differ in a sometimes-unpredictable way.

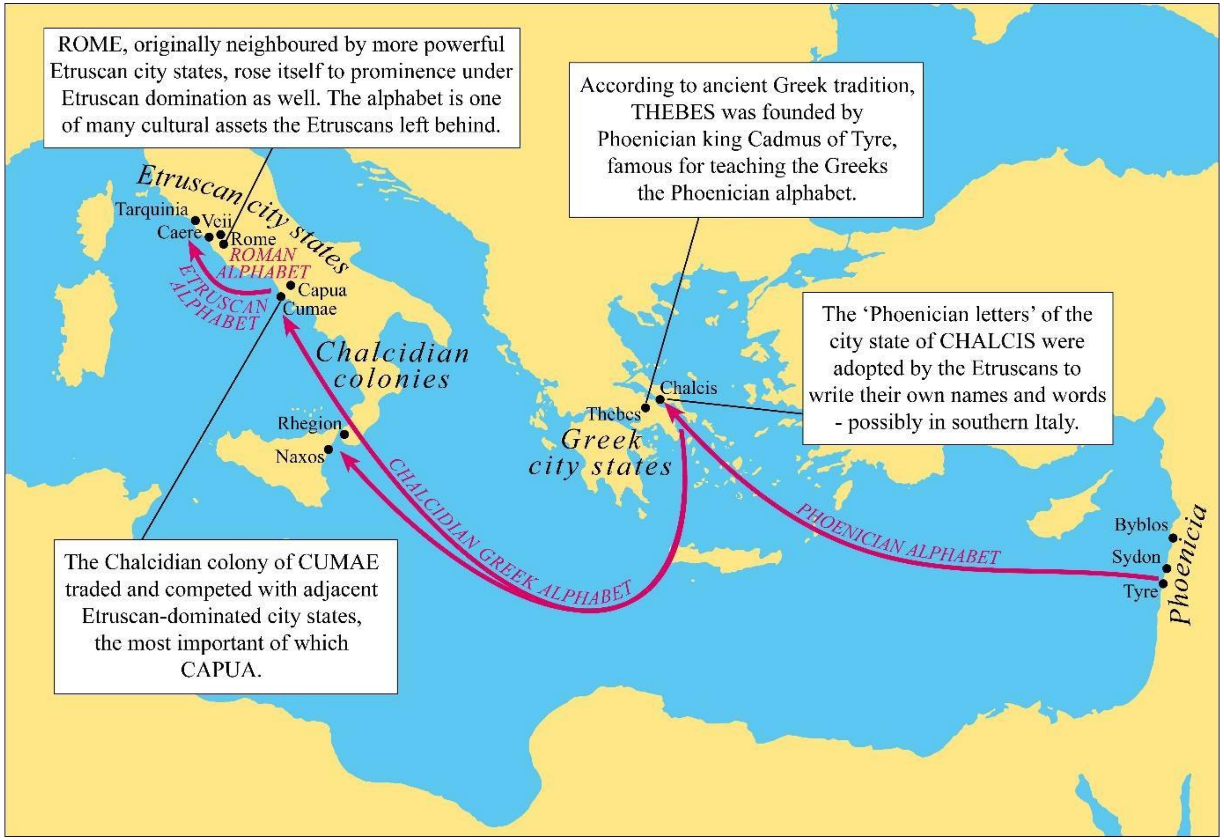


Figure 15-2 Spread of the innovation of phonetic script: the origin of the Roman alphabet.

15.4 Synchronization of writing and pronunciation: spelling reforms

To maximize the accessibility of written language, national regulatory bodies regularly enact spelling reforms to ensure that the correspondence between speech and writing remains as uncomplicated as possible. Because of political and other reasons – such authorities do not always exist, and their mandates are not everywhere as far-reaching – languages quite spectacularly differ in the degree to which this policy is effective. Orthographies accurately reflecting current pronunciation are labelled *phonetic* – for instance in Finnish, as well as in Indonesian and other languages relatively recently put to (Roman) script – as opposed to *archaic* spelling where pronunciation and writing grew wide apart. To the detriment of those concerned with geographical names, in quite a few cases (for instance in the UK and in the Netherlands), names tend to be completely or partially exempted from general pronunciation-based spelling reforms. Even when new spelling rules are meant to include toponyms as well, the mandate of language authorities does not always extend to legally registered names of administrative entities. In other cases, however, geographical names neatly follow suit: this happened, for instance, in the Swedish reform of 1906, the Indonesian reforms of 1947 and 1972, the Romanian replacement of î with â in 1993, and the simplification of the Bulgarian Cyrillic alphabet in 1945.

15.5 Treatment of pronunciation issues by UNGEGN

Although the standardization objectives of UNGEGN cannot be realistically extended to standardization of names pronunciation, the issue of pronunciation is both essential and important enough from a user's point of view to justify incorporation in the expert group's supportive activities. The involvement of the working groups in the subject of pronunciation could (and partly already does) focus on the following aspects:

15.5.1 Toponymic Data Files and Gazetteers

Although the entity Name would be logically defined as a string of phonemes (a sound sequence) that may be expressed by one or more strings of graphic characters, the fact that written name forms are the main focus of names standardization as promoted by UNGEGN makes it an obvious choice to include pronunciation among the attributes of the written name in a toponymic database. In this structure, it should be noted that there may be a '1 to n' relationship between written name and pronunciation. In many cases, it might be worthwhile to facilitate the inclusion of both a (preferred) standard pronunciation and a local pronunciation: for instance, New Orleans (/nuː ɔːrˈliːnz/, locally /nuː ˈɔːrlɪnz/), Toulouse (/tu.luz/, locally /tuˈluzə/), Toronto (/təˈrɒntoʊ/, locally also /ˈtrɒnoʊ/).

It should be reminded that it may occasionally be debatable and/or politically sensitive to make an implicit statement about what should be considered standard and preferred. In any case, pronunciation is never merely dependent upon writing, but always at least upon a combination of writing and language: both in English and in Dutch, the writing of Amsterdam is the same but its pronunciation differs (/ˈæmstərdæm/ in English, /ˈɑmstərˈdɑm/ in Dutch), while Berlin is pronounced /ˈbərlɪn/ in English but /bɛɐ̯ˈliːn/ in German. As local pronunciation may actually belong to another language, confusion in this respect is imminent: Barcelona is pronounced /barθeˈlona/ in Spanish, but locally and in Catalan /bərsəˈlonə/.

Likewise, in spite of its very obvious Spanish origin, the name of America's second largest city Los Angeles is pronounced /lɑˈsændʒələs/ by English-speakers but /losˈaŋxeles/ by its 1.5 million Spanish speaking inhabitants; similar situations occur all over the State of California and other parts of the southwestern United States, abounding as they are both in geographical names of Spanish origin and in Spanish-speaking inhabitants. As neither the United States federally nor the State of California constitutionally defined an official language, the location of the objects these names refer to does not automatically imply the language they belong to. Moreover, in cases where the connection between writing and pronunciation of names has essentially been lost, like in the English of the UK, the need to apply another writing to the dialectal form of a name will not be perceived, making it impossible to assign a language to a written name: the language of Newcastle (upon Tyne) is 'English English' (Received Pronunciation) when pronounced /ˌnjuːkɑːsəl/, but Northern (Geordie) dialect when pronounced /njuːˌkæsəl/. On the other hand, the dialectal form of the name of the southern Dutch city of Maastricht (Dutch pronunciation: /maːˈstrɪxt/) is written Mestreech, to approximate (according, although debatably so, to Dutch writing standards) the local pronunciation /məˈstʁeːç/.

In a digital database, pronunciation may be included as a character string (IPA or any language-specific notation) and/or as a recorded or automatically generated sound file. The advantage of IPA notation is that, unlike sound, it does not depend on the trained ear: as argued above, the sounds one recognizes are to a certain degree dependent on the phonemes one is familiar with. Disadvantages are the unfamiliarity of the general public with the sound values corresponding with the IPA characters, and the fact that there is no assured unity or consensus (yet) concerning the proper way to represent every single pronounced name in IPA.

Automatically generated sound files depend on the character sequences used as input to the sound generation application, so they share this latter disadvantage with IPA. An additional disadvantage of recorded sound files is the influence of the voice and the native language of the owner of the recorded voice. Language-specific complications may occur: in Chukchi, an indigenous language of Eastern Siberia written in a Cyrillic script, the proper pronunciation for the character combination кр is reportedly /rk/ when pronounced by men, but /ʦʦ/ when the speaker is a woman[ii].

Pronunciation details provide in either way a useful addition to serve the needs of those needing to communicate the names in speech (school teachers, news readers and reporters etc.). When including pronunciation details in a toponymic database considered to have official status, they should however be accompanied by a disclaimer explaining the exact status of the pronunciation details offered – unless these are official as well.

15.5.2 Toponymic Terminology

The Glossary of Terms for the Standardization of Geographical Names published in 2002 (ST/ESA/STAT/SER.M/85) does include terms required to discuss the relationship between writing and pronunciation.

15.5.3 Romanization Systems

Although, as demonstrated above, even in originally Roman-written names the relationship between writing and pronunciation is never self-evident, it was the observed disadvantage of transliteration as opposed to direct transcription of sounds to graphics – the obvious difficulty to interpret sound values of transliteration alphabet letters to those unfamiliar with the language involved – that made the 8th UNCSGN issue its resolution recommending the institution of the Working Group on Pronunciation in 2002. In the Romanization tables listed in the Technical reference manual for the standardization of geographical names (ST/ESA/STAT/SER.M/87), details of pronunciation were thus far nevertheless explicitly left out because of the difficulty to find complete and authoritative (official) sources for all the languages presented. When such sources become available, the sound values of single characters might be explained using IPA in print or pdf, or hyperlinks to exemplary sound files offered on the Romanization website. As the tables are exclusively meant for transliteration from non-Roman to Roman script and not for direct transcription of names from unwritten languages, however, such additional functionality in a way overshoots the target of the tables. Also, as was made clear above, in many a case there is no simple or unambiguous correspondence between a single written character and the pronunciation of a phoneme.

15.5.4 Training Courses

Instructions on the recording of pronunciation details when collecting geographical names in the field are essential to the toponymic component of the base mapping process. Language and situation-specific directives should be drafted up and practiced using the expertise of local language specialists. Properly recorded pronunciation details are required during the office treatment of the names, when recommendations on the correct or most appropriate writing are being prepared for official approval.

15.5.5 Exonyms

If pronunciation is to be included in the discussion about the discrimination between exonyms and endonyms, the definition of an exonym should be broadened from the written form of the name to its pronounced original. This would not be without consequences for the prevailing lists of exonyms: in spite of its coincidentally identical writing, English /ˈæmstərdæm/ (Amsterdam) would then for instance be considered an exonym of Dutch /ˈɑmstərˈdɑm/, and English /ˈbərlɪn/ (Berlin) an exonym of German /bɛɐ̯ˈliːn/. Although theoretically defendable, this would not serve the interest of (written) names standardization, and might thus not be advisable.

15.5.6 Geographical Names as Cultural Heritage

For the cultural heritage content of a name, its pronunciation might be considered to weigh as heavily as its written form. As there is a tendency for the pronunciation of names to change faster than their writing, especially where the latter does not follow spelling reforms but keeps reflecting a pronunciation that no longer exists, written names sometimes expose a part of an object's heritage that its pronounced equivalent no longer reveals. On the other hand, there are many historical examples of name writings that were at one point of time adapted to a reinterpretation of the name's meaning following language change (a process called folk etymology). The mutual influence of writing and pronunciation is a central theme when it comes to the study of a name's etymology in order to reconstruct its cultural background.

15.5.7 Task Team for Africa

In order to support the collection and standardization of names belonging to thus far unwritten languages of Africa – as well as other multilingual parts of the world that fell under foreign domination until recent times – it is important to map the phonemes of the languages concerned and agree on a way to systematically transcribe them. If no written name is available yet, pronunciation is obviously all we can fall back on.

Country	Year of publication/ latest update	General information on pronunciation	Standard pronunciation	Standard accentuation	Exceptions	Minority language/ dialect pronunciation
Austria	2012	x	x	-	-	x
Belgium	2009	-	-	-	-	-
Chile	2007	-	-	-	-	-
Czech Republic	2007	x	x	-	-	-
Denmark	2012	x	x	-	-	-
Estonia	2012	x	x	x	-	x
Germany	2010	-	-	x	x	-
Netherlands	2012	x	x	x	-	x
Ukraine	2012	x	x	(none)	-	-
Republic of Korea	2012	x	x	-	-	(none)
Hungary	2012	x	x	-	-	-
Poland	2010	x	x	x	-	x
United Kingdom	2009	x	(none)	(none)	(not listable)	(references)
Norway	2014	x	x	x	-	x
Finland	2014	x	x	x	-	x

Table 15-1 Survey of toponymical guidelines

The experience of the Working Group on Training Courses may prove useful here.

15.5.8 Toponymic Guidelines

Especially for languages where the correspondence between writing and pronunciation essentially remained systematic, toponymic guidelines can provide valuable assistance to those required to pronounce a name. As in many situations there are at least some archaic remains among the written names, or names of which the pronunciation does not follow the contemporary rules for reasons unknown, listings of the general rules for all languages in official use within the territory covered by the guidelines should be augmented by a list of the most commonly encountered exceptions, i.e. irregularly pronounced names of the most prominent geographical objects. This will only be practicable for languages where a fixed set of rules is generally followed: as demonstrated before, English is not one of them.

A quick survey of Toponymic Guidelines recently published online discloses the following:
Whereas some of the Guidelines provide quite extensive information on standard pronunciation details of the major official language, regional minority languages are typically treated more summarily, and pronunciation details are mostly omitted. Notable exceptions to general pronunciation rules were only incidentally listed; it would be recommendable to mention at least whether they occur, which in some cases might not be the case (see table 15-1).

Amongst the pronunciation details accentuation deserves special mention. Although in some languages it does not play a very prominent role, in many it is phonemic. Where standard accentuation rules apply, it is helpful when these are explained. They might be reflected in writing, either mandatory or optional, and either universally or just in case of exceptions to the general rule (e.g. Spanish): this too is helpful to know.

15.5.9 Pronunciation

The Working Group on Pronunciation, originally primarily instituted to produce pronunciation guides for sets of geographical names, facilitates the sharing of insights and experiences regarding the handling of all thinkable pronunciation aspects of geographical names between the UN member states. As language situations and the relationship between pronunciation and writing conventions vary widely between nations, the complexity and the exact nature of the problems faced as well as the direction into which solutions are to be sought and found differ accordingly. Parallels nevertheless occur worldwide, and where they do the wheel doesn't need to be invented twice: the opportunity to learn from each other should never be left unused. Apart from being an obvious platform to share expertise in the recording, storing and disseminating of pronunciation details on a national scale, the Working Group is also mandated to provide support for the collection of audio files for the international UNGEGN Geographical Names Database.

15.6 Future direction and short-term goals for the UNGEGN Working Group on pronunciation – a proposal

In many cases the subject of pronunciation may be considered slippery or even politically sensitive, but in the context of standardization of the writing of names it is, as recognized by the UN Conference on the Standardization of Geographical Names – cf. resolution 7 of the 3rd (1977) and resolution 11 of the 8th UNCSGN (2002), certainly not irrelevant. By gathering accounts of practical experience and bringing together relevant expertise from the greatest possible variety of national situations concerning written vs. pronounced geographical names, as has been done before on a limited scale, the UNGEGN Working Group on Pronunciation can live up to a wide interpretation of its mandate to offer guidance in the treatment of pronunciation aspects of names. As the technical aspects of the storage and dissemination of digital names information, to which for instance audio files representing pronunciation may be reckoned too, are covered by the WG on Toponymic Data Files and Gazetteers, a logical focus of the Working Group on Pronunciation might be the theoretical underpinning of all pronunciation-related activities as well as their coordination on the input and output side, and the production of guidelines and directives to serve the needs of geographical names authorities worldwide. At short notice the Working Group should yield a document stimulating national names authorities to accept the challenge to capture, store and disseminate any kind of names pronunciation detail, by providing widely applicable best practice advice. This Best Practice document should ideally be presentable at the 11th United Nations Conference on the Standardization of Geographical Names scheduled for 2017.

15.7 Reference

Katzner, K. (1975) The languages of the world. London: 1975

Chapter 16 **Web services and applications - open source options vs commercial options: criteria for selection**

Pier-Giorgio Zaccheddu

16.1 Introduction

Access to consistent and reliable multilingual geographical names is essential for a number of uses including postal services, emergency services, navigation, tourism, property purchases, the mass media and applications such as Google Earth/Maps. In all of these areas, geographical names provide one of the most important keys for referencing and accessing a variety of related information. However, today, a patchwork of heterogeneous international, national and regional geographical names related web services and applications exist. Not all of them are compatible with the requirements and standards for integration in international, national and regional spatial data infrastructures.

From a spatial data infrastructure or database point of view, the entity 'name' might be one of many attributes related to the geographical object/feature. UNGEGN has continuously adapted its geographical names standardization programme as an essential part of spatial data infrastructures. It leads to a significant improvement in this portion of the geospatial data management framework promoted by the United Nations Global Geospatial Information Management (UN-GGIM).

This section deals with the display and publication of the content of geographical names databases in different output options, through web services or integrated in web applications.

The following sections will provide information about the prerequisites and tools for creating web services and applications for the publication of geographical names databases within regional, national and international spatial data infrastructures. The information is far from being exhaustive, but it might provide an overview of issues and considerations for geographical names experts.

16.2 The design of web services and applications

First of all, a web service and a web application need a database to be based on. The database needs a certain structure in order to be understood by the web service or application. How a geographical names database should be designed is explained in chapter 'Technical issues: database management'.

Web services are the vehicle for accessing geographical names database content through an application, integrated in a Geographical Information System (GIS) or as an essential part of a spatial data infrastructure.

The German geoportal (Geoportal.de) reveals the contents of the national spatial data infrastructure (GDI-DE). It is the central point of access to the data and services of the GDI-DE. Users can search within the central search engine, which contains around 140,000 decentrally maintained sets of metadata on geospatial data sets and services from across all levels of public administration in Germany.

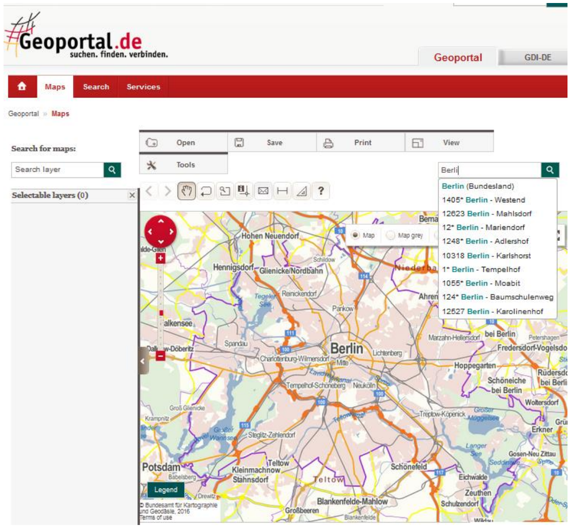


Figure 16-1 The German national geographical names database (GN-DE) published as a web service as part of the national spatial data infrastrucuture (GDI-DE) and visualized through the Geoportal application (Geoportal.de)

Different books and different organizations give different definitions of Web Services. Some of them are listed here [1]:

"A web service is any piece of software that makes itself available over the internet and uses a standardized Extensible Markup Language (XML) messaging system. XML is used to encode all communications to a web service. For example, a client invokes a web service by sending an XML message, then waits for a corresponding

XML response. As all communication is in XML, web services are not tied to any one operating system or programming language."

"Web services are self-contained, modular, distributed, dynamic applications that can be described, published, located, or invoked over the network to create products, processes, and supply chains. These applications can be local, distributed, or web-based. Web services are built on top of open standards such as TCP/IP, HTTP, Java, HTML, and XML."

"A web service is a collection of open protocols and standards used for exchanging data between applications or systems. Software applications written in various programming languages and running on various platforms can use web services to exchange data over computer networks like the Internet in a manner similar to inter-process communication on a single computer. This interoperability (e.g., between Java and Python, or Windows and Linux applications) is due to the use of open standards."

To summarize, a complete web service is, therefore, any service that:

- Is available over the Internet or private (intranet) networks
- Uses a standardized XML messaging system
- Is not tied to any one operating system or programming language
- Is self-describing via a common XML grammar/syntax
- Is discoverable via a simple find mechanism

In a nutshell, web services allow various applications to talk to each other and share data and services among themselves. Thus, different applications can use the same web service if the required standards and protocols are used.

In computing, a web application or "web app" is a client-server software application in which the client (or user interface) runs in a web browser (e.g. Microsoft Internet Explorer or Mozilla Firefox) [2]. Unlike traditional desktop applications, which are launched by the operating system, web apps can only be accessed through a web browser.

The difference between a website and a web application is simple. A website is defined by its (static) content while a web application is defined by its interaction with the user. E.g. a website can plausibly consist of a static content repository that's accessible for all visitors, while a web application depends on interaction and requires user input and data processing. For example, a daily news site would be a website, but a spreadsheet or a collaborative calendar would be a web application.

Web browsers support many programming languages. Web apps have to be programmed in a language the browser can understand. There is no "best language for developing web applications". The following is a list (not complete) of languages that are commonly used to programme web applications [3]:

- Python: high-level programming language focusing code readability.
- Java: the server-side language for large-scale websites with a high volume of traffic.
- JavaScript: used to express animation, interactivity and other dynamic effects.
- Ruby: similar to Python, emphasizing short and simple code.
- PHP: a general-purpose server-side scripting language.
- CSS: allows a programmer to define the layout of multiple webpages at once.
- HTML: the standardized markup language that structures and formats content on the web.

HTML and CSS are not really programming languages. HTML is simply the format consumed by web browsers to parse and render webpages. CSS is used to write rules that specify the appearance of the page (fonts, colours, background, etc.).

A special web service, addressing a specific profile or use case for publishing geographical names data, is the so-called "gazetteer service". In this context, geographical names databases and gazetteers should be distinguished from each other, as gazetteers are only one of many possible outcomes derived/produced from a names database. A gazetteer can be defined as a list, a report or a repository of location information that is used to search for specific locations.

The term 'gazetteer' in a spatial data infrastructure (SDI) context is considered as *"any geospatial dataset which contains 'spatial identifiers'"*. These can be geographical names, postal codes or other indexes for indirect spatial referencing. The intended use of 'gazetteers' in the European INSPIRE initiative (using 'geographic identifiers') followed the International Organization for Standardization (ISO) 19112 standard. The schema from ISO 19112 was slightly amended to correct errors in that schema and to allow for a better integration in INSPIRE as an SDI. 'Gazetteers' here were simply intended as a channel to publish spatial data from the INSPIRE themes that allows others to use them in indirect spatial referencing. It is obvious, that this technical SDI view on 'gazetteers' is different from the UNGEGN view on 'gazetteers': *"List of toponyms arranged in alphabetic or sequential order, with an indication of their location and preferably including variant names, type of (topographic) feature and other defining or descriptive information."* [4]

Within UNGEGN recommendations and policies, and indeed within the wider research literature, commonly accepted definitions for the terms 'official' and 'unofficial' do not seem to be available as they relate to gazetteer data. Rather, there seems to be a proliferation of terminology used to define both types of data which are incorporated into gazetteers, and the gazetteers themselves – ranging from official and authorised to unofficial and informal. The need for the officially sanctioned gazetteers to be of a high quality in terms of accuracy and completeness of available data is increasing rapidly [5]. In a nutshell, a gazetteer enables the user to search and find a location of interest within an application. A gazetteer service uses location data (usually in object/feature based format) with information related to location like a location name, coordinates and/or postcodes [6].

There is growing interest in the development of a common object/feature-based model for access to named objects/features, often referred to as a 'gazetteer' as well. Two major activities form the basis of this standard, an Open Geospatial Consortium (OGC) Best Practice for Gazetteer Services – Application Profile of the Web Feature Service Gazetteer (WFS-G) Implementation Standard, which is an OGC Discussion paper on gazetteers, and an ISO draft standard for geographic identifiers [7]. Since this resource is so important, there is growing global interest in sharing and updating geographical names across a standard web-based service interface that is not controlled by any one organization or group. OGC has met this need for open accessibility to geographical feature data via its creation of the Web Feature Service (WFS) standard. The WFS standard defines an interface for specifying requests for retrieving and updating geographical features across the Web using platform-independent calls. And of interest to the geographical names community, members of the OGC have been actively developing a special profile of the WFS which is being designed specifically to support geographical names data through a WFS-G profile. However, the WFS-G profile is still a Best Practice document and has not been adopted as an OGC standard for different reasons, amongst others, that schema elements, attributes, etc. should be defined and described in more detail. Furthermore, the suggested closed code list for type/status of name ("official", "variant") needs to be reconsidered in order to make the schema applicable beyond the USA and Canada. A compromise (INSPIRE) for European purposes was "official", "standardized" and "other" for current names, and "historical" for names not in use anymore. Only a few countries in Europe provide their geographical names with status 'official' while the names are 'standardized'. Also, the value "variant" is out of the scope of name's status (e.g. "official").

16.3 Open source options vs commercial options: criteria for selection

Generally, one can find commercially oriented and open source oriented web feature (gazetteer) services software and implementations. It is deemed impossible to define selection criteria for any decision-making regarding their selection. From the technical point of view there might be advantages and disadvantages with every software programme or tools considered for the services architecture, but these are very specific and depend on the envisaged implementation. Very often the existing IT infrastructure and services architecture in an organization determine the usage of web services software/tools. For example, if an organization uses ESRI Arc GIS for the maintenance of the spatial data – including geographical names – the publication of web services through the tool ESRI Arc GIS Server is obvious. Nevertheless, technical requirements can make the usage of open source software necessary as well.

Organizations providing a commercial web services' offering and charging for its usage, most probably do offer a free-of-charge (light) version of the services or applications as well. This is general practice. One example for an open gazetteer service is provided via Ordnance Survey (OS). OS OpenData provides access to a 1:50,000 scale gazetteer product that can be used to generate gazetteer services [8]. Another example is GEOnet Names Server (GNS), provided by the United States National Geospatial-Intelligence Agency (NGA). It is an official repository of standard spellings of global geographical names. All geographical features within the database contain information about location, administrative division and quality and are ISO compliant. The database can be used for a variety of purposes, including establishing official spellings of foreign place names, cartography, GIS and finding places [9].

Figure 16-2 Search for geographical names through the GEOnet Names Server (GNS)

GNS provides a number of OGC services including a WFS-G. This can be implemented into any WFS enabled client and is therefore extremely useful for geoportals looking to provide gazetteer services and search tools. Furthermore, the GNS repository of global geographical names, coordinates, and extents can be used to create gazetteer services. All data and services are open and available to download free of charge from the website.

In summary, geographical names are used within different web services and applications:

- As search criteria (location), e.g. in a geoportal, for rescue services, geocoding [10], geoparsing [11] and navigation.
- As geographical identifiers, e.g. in gazetteer services.
- For visualization, e.g. as an information layer in viewing services.
- In standardization, translation, and compilation of maps, reports, documents and articles. For instance, reliable information on the correct spelling and the status of names is required by press agencies and map producers.
- For the processing of spatial data sets, e.g. for integration of historical data.
- In human and social sciences, e.g. in linguistic research, onomastic science, archaeology and etymology.

Figure 16-3 News portal of the European Commission using geographical names in different languages to search for information

For all these reasons and in order to allow for these various use cases the publication of geographical names databases as web services or web applications is broached.

From a spatial data infrastructure or database point of view, the entity 'name' is only one of many attributes related to the geographical object/feature. One and the same geographical object/feature may be described by many different names, each one of which may again be pronounced, transcribed, transliterated or otherwise rendered graphically in different ways, which may be considered 'official' at different levels at different times. UNGEGN has continuously adapted its geographical names standardization programme as an essential part of spatial data infrastructures, leading to a significant improvement in this portion of the geospatial data management framework promoted by the United Nations Global Geospatial Information Management (UN-GGIM) as geographical names are definitely very pertinent to it.

Within UNGEGN a trend of activities in the countries/divisions focused on the establishment of multi-functional or multi-usable geographical names databases, services and applications to provide GN data for different purposes has been identified – e.g. providing geographical names as an essential dataset to the national or regional SDI or for the support of specific services and applications. Web services technologies for the geographical names database provision, visualization and dissemination are increasingly used and thus support the vision of a (national or regional) Spatial Data Infrastructure (SDI). This very positive trend of supporting multi-purposes of geographical names use is recognized by UNGEGN Resolution VIII/6 'Integration of Geographical Names data into National and Regional Spatial Data Infrastructures' of the Eighth United Nations Conference on the Standardization of Geographical Names (Berlin, 2002). Res VIII/6 […] *"recommends that standardized geographical names data should be better considered in the establishment of national and regional spatial data infrastructures (SDIs) and included in their design, development and implementation."* [12]

16.4 References

[1] tutorialspoint, http://www.tutorialspoint.com/webservices/what_are_web_services.htm, last accessed 09/2016

[2] Wikipedia, https://en.wikipedia.org/wiki/Web_application, last accessed 09/2016

[3] 10 Best Programming Languages of 2015, Devsaran, https://www.devsaran.com/blog/10-best-programming-languages-2015-you-should-know, last accessed 09/2016

[4] Glossary of Terms for the Standardization of Geographical Names, UNGEGN, 2002, http://unstats.un.org/unsd/geoinfo/UNGEGN/docs/pdf/Glossary_of_terms_revised.pdf

[5] "The Four Faces of Toponymic Gazetteers", Submitted by Australia, Tenth United Nations Conference on the Standardization of Geographical Names, 2012, E_CONF.101_57_The Four Faces of Toponymic Gazetteers

[6] Explanations to gazetteer services, http://learningzone.rspsoc.org.uk/index.php/Learning-Materials/Introduction-to-OGC-Standards/5.3-Example-of-gazetteer-services, last accessed 09/2016

[7] Best Practices Document "Gazetteer Service - Application Profile of the Web Feature Service Best Practice", Copyright © 2012 Open Geospatial Consortium; To obtain additional rights of use, visit http://www.opengeospatial.org/legal

[8] OS OpenData, http://www.ordnancesurvey.co.uk/business-and-government/products/opendata-products.html, last accessed 09/2016

[9] GEOnet Names Server (GNS), http://geonames.nga.mil/gns/html/, last accessed 09/2016

[10] Wikipedia, "Geocoding is the computational process of transforming a (postal) address description to a location on the Earth's surface", https://en.wikipedia.org/wiki/Geocoding, last accessed 09/2016

[11] Wikipedia, "Geoparsing is the process of converting free-text descriptions of places (such as "twenty miles northeast of Jalalabad") into unambiguous geographic identifiers, such as geographic coordinates expressed as latitude-longitude.", https://en.wikipedia.org/wiki/Geoparsing, last accessed 09/2016

[12] UNCSGN Resolutions 1967 – 2012, http://unstats.un.org/unsd/geoinfo/UNGEGN/docs/RES_UN_E%20updated_1-10%20CONF.pdf, last accessed 09/2016

#Section 7 Technical issues: internet, web services infrastructure and applications

Chapter 17 **Visualizing geographical names databases**

Pier-Giorgio Zaccheddu

17.1 Introduction

After the geographical names information has been gathered in the field, this information must be stored in databases and prepared for publication and dissemination. If maps are the method used for publication and dissemination, very likely the content of the databases will be imported into other systems, like Geographical Information Systems (GIS). With GIS geographical names data can be integrated with other geospatial data relevant for the target map.

Concerning the publication and dissemination of geographical names data through the internet we have learned in chapter 'Web services and applications - open source options vs commercial options: criteria for selection' that several technologies can be used: using a simple website or through web services and applications.

In this chapter, the focus will be put on the publication and dissemination of geographical names data using the Google Earth and Google Maps application. Alternatives might be Bing Maps or Yahoo Maps [1].

It is worth noting that other applications can be used free of charge without any commercial interest. Most often 'OpenStreetMap (OSM)' is used to publish a map which shows data collected by others in a geographical context [2].

17.2 The Google Earth/Maps applications

Google Earth is a virtual globe, map and geographical information programme provided by Google in 2004. It maps the Earth by the superimposition of images obtained from satellite imagery, aerial photography and geographical information system (GIS) onto a 3D globe. It was originally available with three different licences, but has since been reduced to just two: Google Earth (a free version with limited function) and Google Earth Pro, which is now free and is intended for commercial use [3].

Google Maps is a Web-based service that provides detailed information about geographical regions and sites around the world. In addition to conventional road maps, Google Maps offers aerial and satellite views of many places. In some cities, Google Maps offers street views comprising photographs taken from vehicles. Google Maps offers several services as part of the larger Web application (e.g. a route planner). The Google Maps application programme interface (API) makes it possible for Web site administrators to embed Google Maps into a proprietary site such as a real estate guide or community service page. The API defines the correct way for a developer to write a programme that requests services from an operating system (OS) or from other applications.

Google Earth and Google Maps (API) also allow users to create and import geospatial data – including geographical names – from a variety of different sources.

In several toponymic training courses organized by UNGEGN the geographical names information which had been gathered in the field, was first of all stored in a database before it was further processed. Usually, the data was imported into a database and then into a GIS to demonstrate how to display names together with geospatial data, e.g. on a satellite image or by using Google Earth or other applications.

The following figure 17-1 shows geographical names gathered within the fieldwork in Madagascar (2013) and displayed in a GIS, here using the proprietary product ESRI ArcGIS and OSM data as background map layer:

Figure 17-1 "OSM Map data", Displaying fieldwork results of the Madagascar UNGEGN toponymy course with ESRI ArcGIS integrating the geographical names data with free OSM data for Madagascar (2013)

If the geographical names data attribute table within the GIS will be displayed further using the Google Earth or Google Maps application the table has to be exported in a file that can be understood by the Google applications. At least, a so-called CSV-file has to be created. A CSV is a comma separated values file, which allows data to be stored in a table-structured format. CSV files can be used with any spreadsheet programme, such as Microsoft Excel, Open Office Calc, or Google Spreadsheets.

The general default file type for spatial data in Google Earth and Google Maps is KML (Keyhole Markup Language) or KMZ (a compressed or "zipped" KML file). KML files are text-based and employ coding tags like those used for XML or HTML programming. Google Earth processes KML files in a way that is similar to how web browsers process HTML and XML files. KML syntax tells applications like Google Earth and Google Maps how to display geographical features such as points, lines, images, polygons, and models. Each location included in a KML file must have longitude and latitude coordinates assigned to it. The following figure 17-2 shows an example from the fieldwork data of Madagascar (2013) displayed with Google Earth.

Other data can make the view more specific, such as tilt, heading, and/or altitude. Some KML files, like placemarks (i.e. one can mark one's own places on the map in Google Earth), ground overlays, paths, and polygons can be created directly in Google Earth and Google Maps. More complex KML elements must be created "by hand" using a text editor and require knowledge of KML tagging.

All GIS software – proprietary or open source – allows for the export of a so-called shapefile from the GIS and conversion of it directly to a KML file. Otherwise open source tools (like Shape2Earth as shown in figure 17-3) exist to convert the shapefile into a KML file. When a CSV or shapefile is saved with a .kml or .kmz extension,

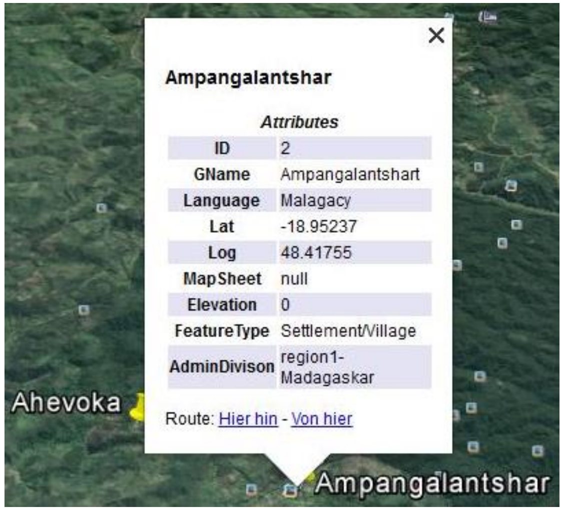


Figure 17-2 "Image © CNES/Astrium", "© Google", "Image Landsat", "Image © 2016 Digital Globe", displaying fieldwork results of the UNGEGN Madagascar toponymy course (2013) with the Google Earth application

applications like Google Earth/Maps, Bing Maps, Yahoo Maps and all the others know how to display it.

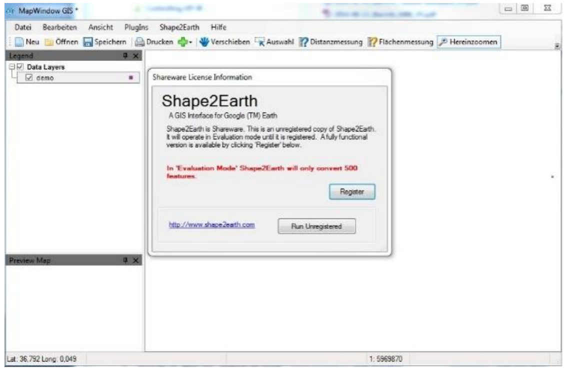


Figure 17-3 Open Source software "Shape2Earth", http://shape2earth.com/

KML files can be used for a wide variety of applications. One may placemark one's home, school or favourite destination; plan or document a trip; map natural resources and scientific datasets – and – visualize geographical names data retrieved from the fieldwork through the internet [4].

The disadvantage of using the Google Earth application is that users are required to download and install the application software on their own computers. If the application has been installed the geographical names data can be displayed using the ortho-imagery of Google Earth as a background layer (see figure 17-2). This data view can either be demonstrated to people as an internal view or shared with other communities.

The same applies for the Google Maps application, but the advantage with this application is that no download and installation of software is needed on the computer. The Google Maps application just requires an existing Google-eMail account. "Maps-user" have to register in order to be able to import their own geographical names data using Google Maps (comprising maps and ortho-imagery) as the background layer.

The following figure 17-4 shows an example from the fieldwork data of Madagascar (2013) and displayed with Google Maps:

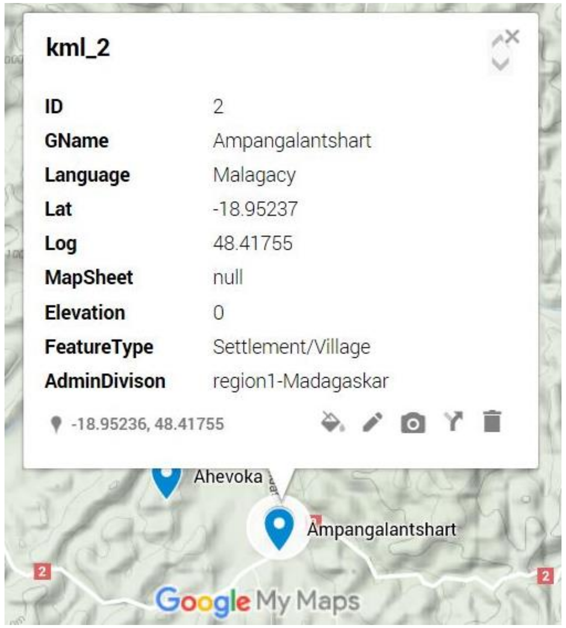


Figure 17-3 "Map data ©2015 Google", Fieldwork results of the UNGEGN Madagascar toponymy course (2013) using the relief map of the Google Maps application, <iframesrc="https://www.google.com/maps/d/embed?mid=14QOZeQaogJHVZLb3GEDGoqSezX4" width="640" height="480"></iframe>

The following figure 17-5 shows another example from the fieldwork data of Indonesia (2012) displayed with Google Maps.

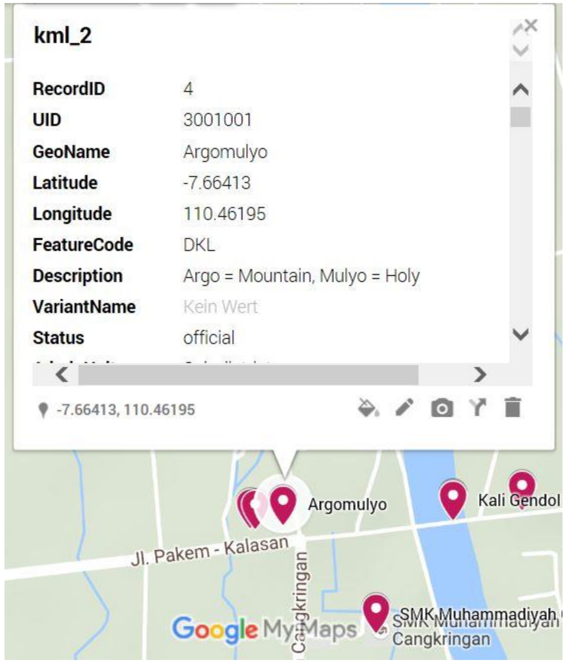


Figure 17-4 "Map data ©2015 Google", Fieldwork results of the 4th UNGEGN toponymy course in Indonesia (2012) using the background map of the Google Maps application, <iframe src="https://www.google.com/maps/d/embed?mid=14QOZeQaogJHVZLb3GEDGoqSezX4" width="640" height="480"></iframe>

The fieldwork results of Indonesia had been initially stored in a Microsoft Access database (figure 17-6), exported as CSV – including all toponymic attributes – and then imported into Google Maps (figure 17-5). The MS Access database is advantageous for maintaining and managing the geographical names information. For its publication and visualization with Google Maps a simple CSV or KML would be sufficient.

GN_object

RecordID	UID	Name	Latitude	Longitude
3	3000999	Yogyakarta	-7,78	110,36
4	3001001	Argomulyo	-7,66413	110,46195
5	3002229	Cangkringan	-7,66419	110,46143
6	3003987	Kali Gendol	-7,66405	110,46465
7	4559983	Jembatan Gendol	-7,66406	110,46465
8	7689912	Mesjid Nur Rohman	-7,6639	110,46574
9	4599723	SMK Muhammadiyah Cangkrigan	-7,66584	110,46342
10	6910036	Majasem	-7,78619	110,48973

Figure 17-5 – Fieldwork results of the 4th UNGEGN toponymy course in Indonesia (2012) stored in a MS Access DB

The data views created with the Google Maps application can be accessed by other people or communities only by invitation of the "Maps-user" (author) who created the views. However, all views created with Google Earth and Maps can generally be shared (published) as "open views" to everybody in accordance with the Google license conditions.

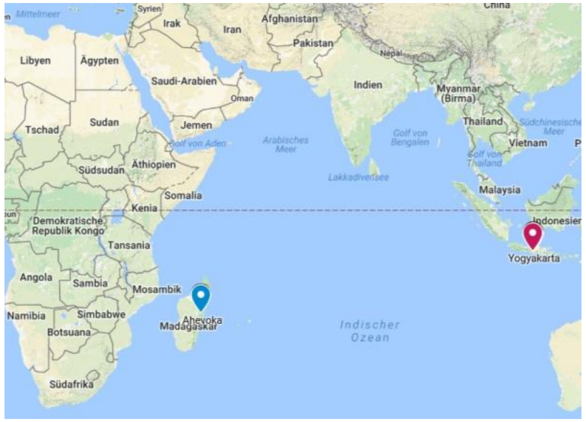


Figure 17-7 "Map data ©2015 Google", "Personal" data views of the author showing the fieldwork results from Indonesia (2012) and Madagascar (2013). This view is restricted and can be shared with other people only on invitation by the author. Although it is rendered with German language toponymy it will be displayed in other languages according to the language indicated by the *operating system of the computer. <iframe src="https://www.google.com/maps/d/embed?mid=14QOZeQaogJHVZLb3GEDGoqSezX4" width="640" height="480"></iframe>*

In a nutshell, with the Google Maps application fieldwork results can be quickly and easily displayed without the need to use any database software or GIS tool. This is interesting for those organizations that would like to show fieldwork results to other people or communities without using a database software and GIS tools. However, it is worth noting that the Google Earth and Maps applications cannot provide a substitute for the solid management of a national names programme.

17.3 References

[1] Top Free Mapping Apps, http://www.appappeal.com/apps/mapping, last accessed 09/2016

[2] Publish something based on OSM and own data, http://wiki.openstreetmap.org/wiki/License/Use_Cases#Case_3:_I_want_to_publish_something_based_on_OSM_and_my_own_data, last accessed 09/2016

[3] Google Earth, https://www.google.com/earth/, last accessed 09/2016

[4] Generate a KML file (see http://code.google.com/apis/kml/articles/csvtokml.html for a possible solution). You can then open that up in Google Maps by storing it online and linking to it from Google Maps as described at http://code.google.com/apis/kml/documentation/whatiskml.html, last accessed 09/2016 for both URLs

Chapter 19 **A Data Repository for Named Places and their Standardised Names Integrated with the Production of National Map Series**

Teemu Leskinen (National Land Survey of Finland)

19.1 Introduction

Finland is a multilingual country. Finnish and Swedish are the official state languages, and three indigenous Saami languages – Inari Saami, North Saami and Skolt Saami – have an official status in the four northernmost municipalities of Finland. In Finland, a specific geographic feature may have several parallel names in one or different languages. In bilingual Finnish–Swedish and Swedish–Finnish areas, a feature typically has two names, and in Lapland's Finnish–Saami language areas, a feature may have up to five names.

The National Land Survey of Finland (NLS) produces, in print, two national map series based on the NLS Topographic Database (TDB): the Basic Map 1:25 000 and the Topographic Map 1:50 000. In addition, seven national small scale map datasets – all incrementally generalised from the TDB data – are being produced at scales 1:100 000, 1:250 000, 1:500 000, 1:1 million, 1:2 million, 1:4.5 million and 1:8 million. In national maps, all standardised names of the feature are presented, if any. Names in bilingual Finnish–Swedish and Swedish–Finnish areas occur in two languages, the name in the municipal majority language placed first. In Lapland, all existing Saami names of a given feature are presented, in a standard language order, together with the (possible) Finnish and Swedish name (*Figure 19-5*). The amounts of geographic names presented in national map series are:

- Basic Map 1:25 000 ca. 810 000 names;
- Topographic Map 1:50 000 ca. 360 000 names;
- Topographic Map 1:100 000 ca. 170 000 names;
- Topographic Map 1:250 000 ca. 45 000 names;
- Topographic Map 1:500 000 ca. 13 500 names;
- General Map 1:1 million ca. 2 500 names;
- General Map 1:2 million ca. 600 names;
- General Map 1:4.5 million ca. 200 names;
- General Map 1:8 million ca. 25 names.

The multilingual and multi-names circumstances in Finland and NLS national map production environment have guided the development of the NLS Geographic Names Register (GNR), initially established in the late 1990s. The GNR was entirely renewed in an NLS development project during 2010–2014. The reform covered the data model and database, the names data management and integration with the production of national maps, the production application as well as the production lines and organisation of work. The GNR named feature classification was also renewed, as were the GNR information services and data products.

19.2 Geographic Names Register

19.2.1 General

The Geographic Names Register of the National Land Survey is the authoritative geographic names data repository of Finland and one of the elements included in the national and international spatial data infrastructure. The GNR comprises the primary Place Name Register (PNR), containing information on over 800 000 named places and their standardised names, and the Map Name Register (MNR), an operational dataset holding the cartographic attributes (placement, typography) related to the PNR names selected to be presented in the NLS cartographic product in question.

19.2.2 Data model

The GNR data model is feature-oriented and includes the objects Place, Place name and Map name. All objects are interconnected and provided with external persistent unique identifiers. A named feature (Place) has at least one name (Place name) – in Finland, commonly several names in different languages – that may have zero, one or multiple occurrences (Map names) in different cartographic products (see figures 19-1 and 19-2).

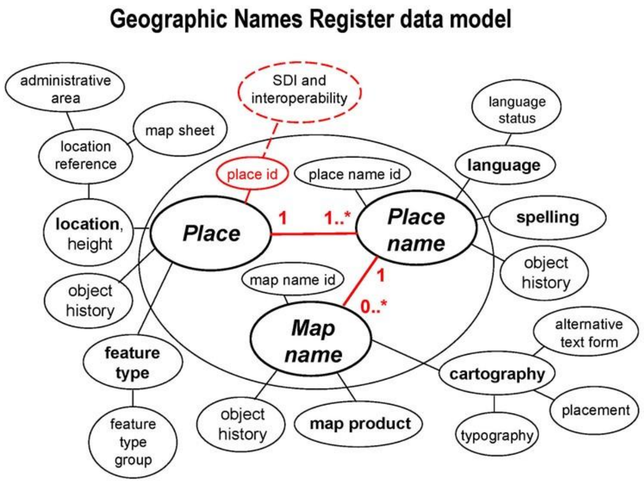


Figure 19-1 Geographic Names Register data model.

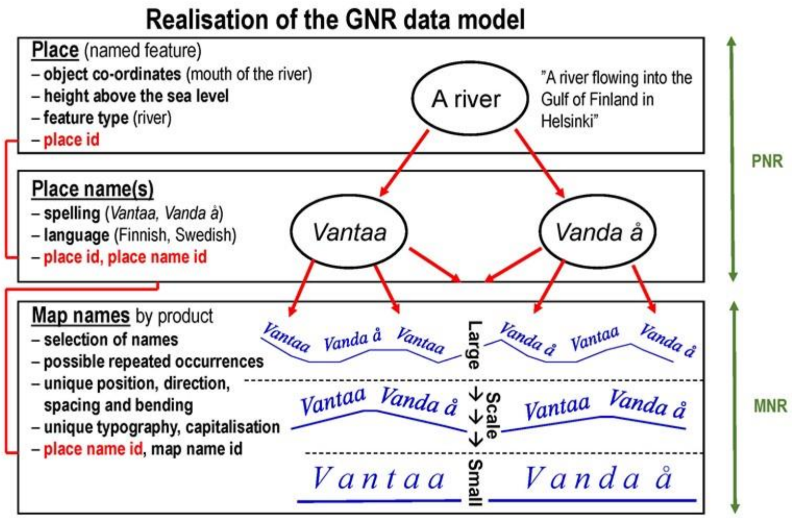


Figure 19-2 Realisation of the GNR data model.

19.2.3 Place Name Register

The objects Place and Place name build up the Place Name Register. The PNR is not based on any map scale and includes no cartographic information.

The object Place carries the following information:

- Place id, a unique identifier as the link to the related Place name(s) and as a possible link to related spatial objects maintained in external datasets;
- The feature type, and the feature type group via code tables;
- The object co-ordinates, centre point (mouth of a watercourse), in ETRS-TM35FIN co-ordinate system;
- The height above sea level, derived automatically from a digital elevation model;
- The municipality, and other administrative areas via code tables, in which the object is situated, derived automatically;
- The map sheet in different map series in which the object is situated, derived automatically;
- Timestamps for the creation, latest modification and possible deletion of the object (see also *Section 19.2.5, Object version management*).

The PNR feature classification was fully renewed in the GNR reform (*Table 19-1*). While the reclassification of the Places is in progress, both the old and the new classification are being maintained as attributes.

The object Place name carries the following information:

- Place name id, a unique identifier as the link to the Map names and as a possible link to related linguistic data elements maintained in external datasets;
- Place id as the link to the Place;
- The spelling of the name, unabbreviated, correct capitalisation;
- The language of the name (Finnish, Swedish, Inari Saami, North Saami or Skolt Saami);
- The official status of the language of the name in the municipality in which the Place is situated (monolingual Finnish, monolingual Swedish, bilingual Finnish–Swedish, bilingual Swedish–Finnish, multilingual Finnish–Saami municipality), derived automatically;
- The source of the name;
- The status of the name, e.g. official, standardised, other;
- Timestamps for the creation, latest modification and possible deletion of the object (see also *Section 19.2.5, Object version management*).

TERRAIN FEATURES
Landforms
Group of islands or islets
Island or islet
Boulder in water or boulders in water
Cape
Isthmus
Shore or waterside
Group of elevations
Elevation
Depression
Shallows
Deep
Terrain areas
Forested area
Marsh or paludified area
Field
Meadow or pasture
Blockfield or rocky area
Detailed terrain features
Boulder
Tree
Other terrain features
Former dwelling place
Other terrain feature

HYDROGRAPHIC FEATURES
Water areas
Baltic sea and its main parts
Part of sea
Group of lakes or ponds
Lake or pond
Part of lake
Reservoir
Part of reservoir
Canal
Watercourses
Watercourse
Part of watercourse
Rapids
Detailed hydrographic features
Spring
Water pit
Other hydrographic features
Other hydrographic feature

PROTECTED SITES
Nature reserves
(3 feature types)
Other protected sites
(6 feature types)

POPULATED PLACES
Population centres and villages
City or town or population centre
Village or neighbourhood
Parts of population centres
Part of city or town or population centre
Park
Market place or square
Houses
House
Other populated places
Other populated place

ADMINISTRATIVE UNITS AND PUBLIC SERVICES
Administrative units
(5 feature types)
Administrative buildings and offices
(5 feature types)
Religious buildings
(3 feature types)
Burial services
(2 feature types)
Social and health services
(4 feature types)
Safety and defence units
(14 feature types)
Educational and research units
(6 feature types)
Waste management services
(3 feature types)
Other administrative units and public services
(1 feature type)

TRANSPORT SITES
(17 feature types)

PRODUCTION AND INDUSTRIAL UNITS
(13 feature types)

CULTURAL AND RECREATIONAL SERVICES
Cultural services
(5 feature types)
Sports and exercise sites
(10 feature types)
Recreational services
(11 feature types)

Table 19-1 Renewed Place Name Register feature classification.

19.2.4 Map Name Register

The Map Name Register object Map name carries the following information:

- Map name id, a unique identifier;
- Place name id as the link to the Place name (and Place);
- Map product, one of the nine NLS map series;
- Typography code, interpreted via code tables and carrying the following information:
  - Text font;
  - Text size, graphic size in mm/100;
  - Text slant, angle in degrees;
  - Text colour;
- Text position, co-ordinates of the lower left corner of the text, in ETRS-TM35FIN co-ordinate system;
- Text direction, expressed as relative co-ordinates (dx, dy) measured from the text position;
- Text bending, up to 32 pairs of relative co-ordinates conducting a curved text;
- Spacing indicator, whether the text direction or text bending parameters also indicate the length of the text;
- Capitals indicator, whether the spelling of the PNR name is turned into upper case during the plotting;
- Parallel name preference indicator, for cases when feature's two parallel names in one language are presented as map names. If the indicator is activated, a space followed by the Finnish, Swedish, Inari Saami, North Saami or Skolt Saami word for the English "or", according to the language of the respective Place name, is appended to the text during the plotting;
- Alternative text form, for cases when the spelling of the name in the map differs from the spelling of the name in the PNR, e.g. for texts split into two lines;
- Timestamps for the creation, latest modification and possible deletion of the object (see also *Section 19.2.5, Object version management*).

19.2.5 Object Version Management

In the GNR reform, a complete version management for Places, Place names and Map names was implemented, based on timestamps for the creation, modification and deletion of the versions and separate database tables for the modified object versions (*Figure 19-3*). Before any modification of a Place, Place name or Map name, the unaltered object version is inserted into the respective version table, i.e. Place version, Place name version or Map name version, and the version is identified by an incremental version number. The object version management enables temporal cross-sections of the GNR data and complete change-only updates for the GNR data dissemination.

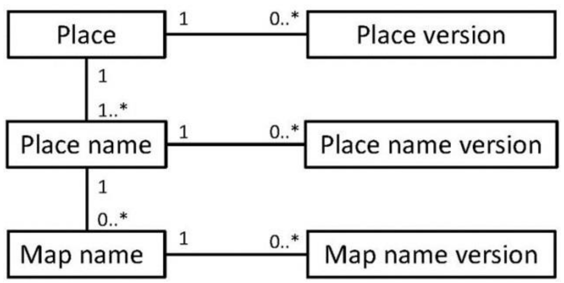


Figure 19-3 GNR object version management.

19.2.6 Database and Transaction Management

The Geographic Names Register is an Oracle Spatial database enabling efficient spatial operations. In addition to the dynamic database tables Place, Place name and Map name, and the respective object version tables, the database includes some 50 code tables for e.g. managing the hierarchical feature type, administrative area and map sheet index structures and interpreting and explaining the codes used in the database for the user interface and data dissemination.

The management of absolutely necessary transactions of long duration, i.e. isolation of data changes during the production and the detection and automatic and semi-automatic resolution of possible object conflicts during the merge, is based on Oracle Workspace technology. The GNR hierarchical workspace structure may have one to three temporary workspace levels each of which consisting of one or several workspaces for different kinds of national, regional, local and object-wise production tasks (*Figure 19- 4*).

19.3 Data Maintenance

19.3.1 Production Application

The Geographic Names Register production functionality is integrated into the NLS Topographic Database production application based on Smallworld technology. Thanks to the integration, all the versatile TDB production functionality developed in the NLS for decades is available for the GNR production. Recently added GNR production functionality covers the GNR workspace management, the Place Name Register data maintenance and the Map Name Register data maintenance. The GNR production application user interface consists of specialised tools for the maintenance of the GNR – forms, lists and the map interface for displaying and altering the data. In total there are over 200 daily NLS users using the application and performing GNR transactions.

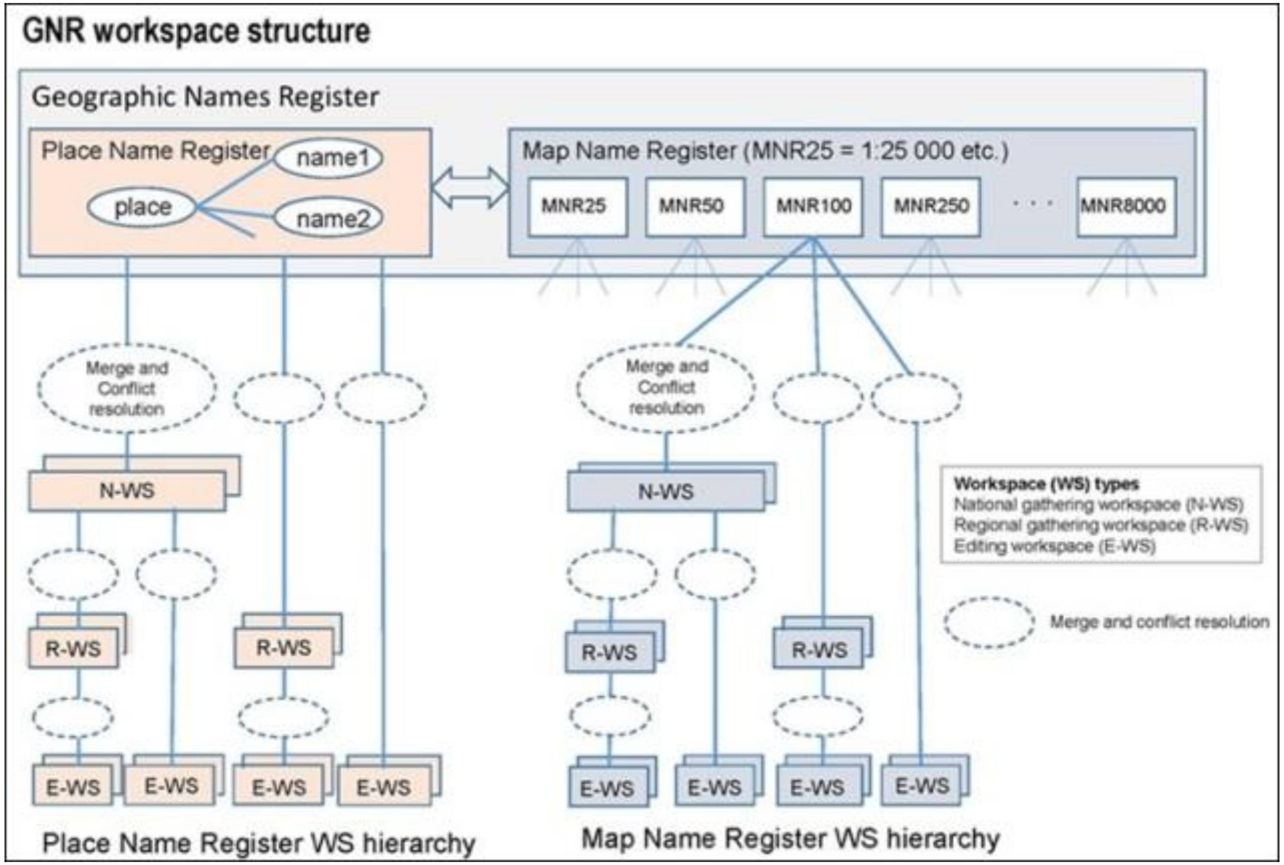


Figure 19-4 GNR workspace structure

Workspace Management

The GNR workspace management user interface is common to PNR and MNR production and includes e.g. the functionality for browsing the workspace structure (*Section 19.2.6*), creating, refreshing and merging the workspaces as well as the detection and automatic and semi-automatic resolution of possible object conflicts during the merge.

19.3.2 Place Name Register Maintenance

The Place Name Register production is maintenance of Places and their attributes described in *Section 19.2.3*. Place names are maintained as attributes of Places, with attributes of their own.

To be able to edit Places and Place names, the user creates a new or opens an existing PNR workspace (*Sections 19.2.6 and 19.3.2*). The existing Places to be edited are fetched from the PNR by using the Place Name Register search form. The form allows the user to combine different search terms freely. The search terms for Places and Place names include e.g. the Place id, the location (a polygon, map sheet or administrative area) and the height of the feature, the feature type, the Place name id, the spelling of the name, the language of the name and the status of the name. A time period for the latest modification of Places and related Place names can also be included as a search term.

A PNR search is search for Places and returns a sortable list of Place names with the essential information on both Places and Place names arranged as columns. All parallel Place names are included in the list even if the search terms would match only some of them. For example, a query for the Finnish name and spelling “Inari” returns all parallel names of the municipality i.e. Inari, Enare, Anàr, Aanaar and Aanar as separate rows, with their parallel names as columns by language in turn.

The map interface of the PNR production application includes the background maps, the portrayal of PNR data on the screen, and the geometry tool for maintaining Places’ locations. As to the background maps, both the TDB vector map data and a complete set of NLS raster maps in different scales are available. In the portrayal of

PNR data, the locations of selected Places are displayed as red symbols and the Place names are automatically placed around Places according to the language of the name: Finnish name(s) appear in upper right, Swedish name(s) in upper left, North Saami name(s) in lower right, Inari Saami name(s) in lower left position and Skolt Saami name(s) under the symbol of the Place (*Figure 19-5*).

The Places and Place names are managed by using the respective object forms and the geometry tool. An existing Place to be edited is selected from the search list or the screen. A Place may be added, modified or deleted, and a Place name may be added to a Place, modified, or withdrawn from a Place by using the Place name form opening from the Place form.

In addition to the object-wise tools, there are three inter-object tools for managing the PNR Places. Two Places with their Place names may be merged as one Place, a Place and its Place names may be split into two Places, and a Place name may be moved from one Place to another. These special operations are necessary in preserving the consistency within the PNR and between the PNR and MNR. There are also special group tools for setting attribute values for several Places at one time.

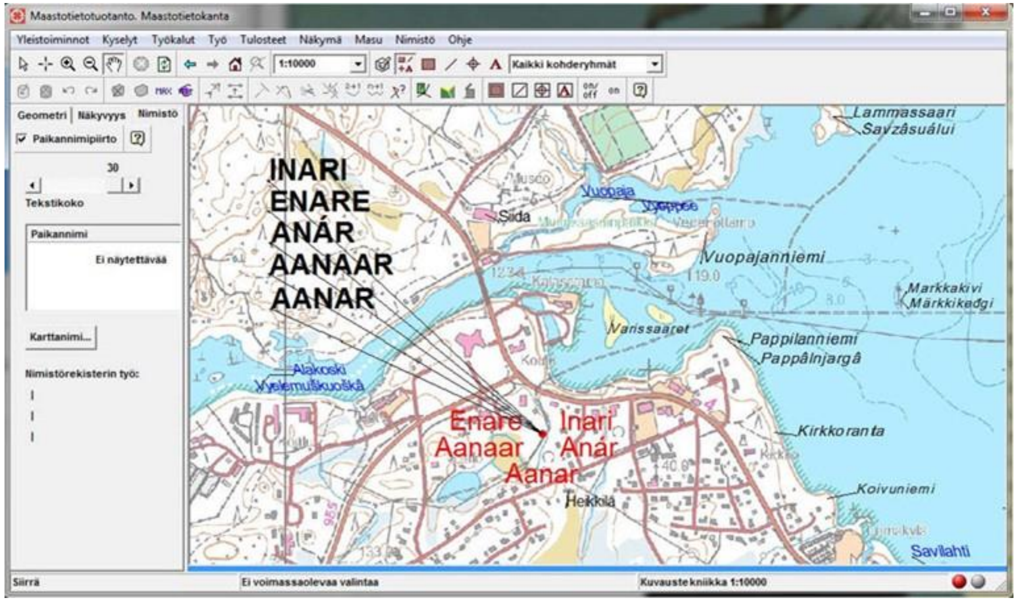


Figure 19-5 GNR production application map interface. A Place and related Place names are displayed in red, and Map names in 1:25 000 (Basic Map) in black and blue.

The version management of the database objects Place and Place name (*Section 19.2.5*) enables queries for the changes made to the PNR during a given period of time. In addition to the time period, the user may combine different criteria for the object changes. The search results may include e.g. added or deleted Place names, changes in the spelling, language or status of the Place names, or they may include changes in the location or feature type of the Places. The query can also be restricted by spatial terms for the location of the Places, such as a polygon or map sheet. The query returns a list of current (or deleted) Place names and Places, similar to the search list in normal PNR production. By selecting a row in the list, another list opens containing all combinations of the respective Place and Place name versions in chronological order to be analysed.

19.3.3 Map Name Register Maintenance

The Map Name Register production is maintenance of Map names and their spatial and other attributes described in *Section 19.2.4*. Map names are always related to a certain cartographic product, e.g. the Basic Map 1:25 000 or the General Map 1:1 million. The map product to be managed is chosen from a pull-down menu opening from the GNR production application main window.

To be able to edit Map names, the user creates a new or opens an existing MNR workspace (*Sections 19.2.6 and 19.3.2*). The existing Map names are fetched for editing by using the Map Name Register search form allowing the user to define the map sheet or map sheets as the search term. As the result of the search, the Map names appear in a list and may be selected to be plotted on the screen. The background map data may be the editable vector data, or a raster map of the cartographic product in question. As to the raster background maps, special versions with no cartographic names in them are available in order to avoid disturbing textual information on the screen.

A Map name may be added, modified or deleted. Names to be added as Map names are always selected from the Place names stored in the Place Name Register, i.e. a Place name must be accepted to the PNR before it can be presented as a Map name in any cartographic product. The Place names to be added as Map names are fetched from the PNR by using the normal PNR search tools described in *Section 19.3.3*.

Map names are managed by using the respective object form, the geometry tool and the map interface. A Map name to be edited can be selected from the search list or the screen. In addition to the editable cartographic attributes, the Map name object form includes the necessary read-only information of the Place and Place name in question.

The GNR production application geometry tool allows the user to control the placement of the Map name in different ways (*Figure 19.6*). A Map name may be placed horizontally by using one geometry point for the lower left corner of the text, directed according to two points, or bended according to three or more points. The text may also be stretched to reach the last point of the given geometry. By using the object form, the name can be split into two lines, or the spelling of the name turned into capitals during the plotting. In the user interface, only a pre-defined set of all existing typography codes (*Section 19.2.4*) is available for a given Map name: the codes allowed for each combination of a cartographic product and PNR feature type are introduced in the GNR code tables.

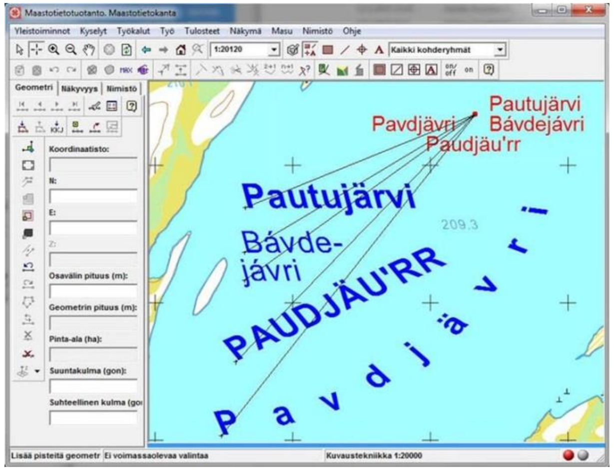


Figure 19-6 Managing the placement and other attributes of Map names (fictional).

To ensure the consistency between the PNR and MNR, and between the MNR Map name layers of different cartographic products, the user interface offers 13 tools for making sure, for example, that recent changes in the PNR (additions, deletions, certain modifications of Places and Place names) are taken into consideration in map compilation, that all parallel Place names of a PNR Place are presented as map names according to NLS cartographic rules and that Place names of all PNR Places of a given feature type are presented as Map names.

19.4 Production Lines

19.4.1 General

The GNR reform covered the PNR and MNR production lines and organisation of work within the NLS, and the procedures for the linguistic treatment of the GNR names data carried out by the Institute for the Languages of Finland and University of Oulu.

19.4.2 Place Name Register Production

The PNR production is carried out by an NLS team devoted to the task. At first, the team concentrates on the reclassification of the PNR Places according to the renewed feature classification (*Section 19.2.3 and Table 19-1*) and complementing the PNR data contents by using different names data sources, including the Names Archive of the Institute for the Languages of Finland and official city and town plans sanctioned and maintained by municipalities.

The quality management of the PNR data and data production covers both linguistic and technical aspects. An overall process for the resolution of reclamations, feedback and change requests concerning the PNR data and the names in national maps has been implemented, with appropriate feedback channels and defined duties, roles and designated persons in charge of the decisions.

19.4.3 Map Name Register Production

The MNR, containing all Map names presented in different NLS cartographic products, is maintained as a part of normal production and compilation of Basic Maps 1:25 000, Topographic Maps 1:50 000 and small-scale maps 1:100 000–1:8 million. The map name compilation may include requests to the PNR production team for the necessary PNR operations, such as the addition of a new Place and/or Place name to be included as a Map name.

19.4.4 Linguistic Treatment

The linguistic treatment of the GNR names is carried out by the Institute for the Languages of Finland (Finnish and Swedish names) and University of Oulu experts (Saami names). These experts can access and alter the PNR data directly online by using the NLS GNR production application.

The names treatment means e.g. the checking of all additions and changes made to the PNR Place names and possible corrections of the spelling of the names. The locations and feature types of PNR Places are checked on request. The most important information sources for the treatment are the Names Archive of the Institute, recent names field collections, and all the treatment material accumulated during the continuous and successful co-operation between the NLS and the Institute since the 1950s.

19.5 Services and Products

19.5.1 WFS Interfaces

The Geographic Names Register datasets are available through NLS Web Feature Service (WFS) interfaces (National Land Survey of Finland 2015). At present, the national GNR WFS products include two XML schemas for the Place Name Register – with equal data contents but different data structure, serving a little different use cases – and one XML schema for the Map Name Register (*Figure 19-7*).

The feature-oriented PNR XML profile (Places) has the Place as the GML Feature Member, with the Place name(s) as attributes, while the name-oriented PNR XML profile (Place names) has the Place name as the GML Feature Member, with the Place data and possible parallel Place names as attributes. The MNR XML profile (Map names) has the Map name as the GML Feature Member, with corresponding Place and Place name information as attributes. The XML schemas also include the explanations for the codes used in the service in Finnish, Swedish and English.

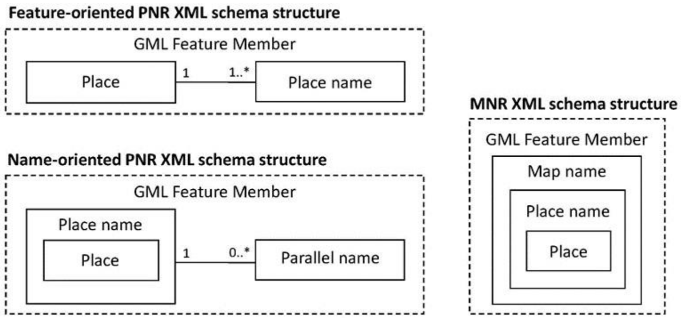


Figure 19-7 GNR Web Feature Service (WFS) XML schema structures.

A Place Name Register WFS query can be filtered e.g. by:

- Place identifier or Place name identifier;
- The location of the Place, a polygon, buffer area, administrative area or map sheet;
- The height of the Place;
- The feature type or feature type group of the Place;
- The spelling of the Place name;
- The language of the Place name;
- The municipal status of the language of the Place name;
- The 'Relevance at Scale' indicator depicting the size or importance of the Place by means of the smallest scale in which the Place name(s) of the Place have been selected as Map name(s) in national map series;
- Places' and Place names' database lifespan information (timestamps for the creation, latest modification and deletion of the objects).

A Map Name Register WFS query can be filtered e.g. by:

- Map product (e.g. Basic Map 1:25 000 or General Map 1:1 million);
- The location of the Map name (e.g. a bounding box or a polygon).

In addition to the national XML schemas for PNR and MNR WFS products, the geographic names XML schema profile sanctioned by the EU INSPIRE directive (Infrastructure for Spatial Information in the European Community) has been implemented for international use. For object-wise interoperability purposes, the GNR data will soon be available as linked data (Uniform Resource Identifiers (URI) for both PNR Places and Place names), according to the guidelines provided by the Finnish Public Administration Recommendations.

19.5.2 GML Products

In addition to the WFS, the GNR datasets are available as downloadable GML files (National Land Survey of Finland 2015). The XML schemas of GNR GML data products follow the corresponding GNR WFS XML schemas. In total, there are 11 GML data products (downloadable files) with nationwide data coverage – two for the PNR (Places, Place names) and nine for the MNR (Map names 1:25 000 to Map names 1:8 million).

19.5.3 National Land Survey Data Policy

The National Land Survey of Finland has made its topographic datasets available to the public and to companies to be used freely and free of charge. The open data products are available without compensation and with extensive and permanent rights of use. The datasets can be used for private and commercial purposes, be published, further licensed and attached to other products and services. The NLS open data products and services include e.g. the Topographic Database, GNR names data, elevation models, orthophotos, laser scanning data, small scale map datasets and raster maps in scales 1:5000–1:8 million.

19.6 Fields of Application

Examples of fields of application of the GNR services and data products are:

- National and international standardisation of geographic names; clear and consistent use of nationally standardised and accurate geographic names in any type of communication;
- Search; finding named places and geographic names by using their attributes (e.g. spelling); map browsing applications; gazetteer services;
- Geocoding; automatic positioning and navigation;
- Interoperability within the national and international spatial data infrastructure, based on unique and persistent Place and Place name identifiers; linked data; geoparsing; ontology;
- Visualisation; map production; geographic names as an information layer in viewing services;
- Research, e.g. onomastics, historical research, natural science, archaeology, genealogy;
- Cultural heritage promotion; safeguarding of the cultural heritage related to inherited geographic names and respective places.

19.7 References

National Land Survey of Finland (2015) Data and services, Digital products, Product descriptions, Geographic names http://www.maanmittauslaitos.fi/en/digituotteet/geographic-names. Accessed 10 April 2015

Section 8 Websites

Chapter 20 **Distribution of standardized names: accessibility on the internet**

Ferjan Ormeling and Pier-Giorgio Zaccheddu

20.1 Introduction

When names have been collected and standardised the work is only half done. As long as the standardised names have not found their way to the public, all the work has been for nothing. The public needs to be informed of the standardised name forms, so that they are used. On the UNGEGN website (http://unstats.un.org/unsd/geoinfo/UNGEGN/geonames.html) there are links to all national websites from which the national geographical names databases can be accessed. We will discuss a number of them in order to get an overview of the way they operate and of their functionality.

20.2 Australia

Geoscience Australia, a department of the Australian Government, has a place name search option (http://www.ga.gov.au/map/names/) where one can type in geographical names which then are retrieved from the central database. A number of matching entries will then come up, with indication of their record ID, feature code, coordinates, status (are they official or not) and the state they are located in. On that basis, one can go to the entry looked for, and when clicking this, a detailed map with the looked-for entry comes up, with a locator map, feature ID and an indication of the topographical map sheets it can be found on.

20.3 Canada

Canada has 3 options, as the geographical names database can be accessed through three different sites, the national atlas (English: http://atlas.gc.ca/site/english/index.htm , the Canadian Geographical Names Database (http://www.nrcan.gc.ca/earth-sciences/geography/place-names/10786), and the Canadian Geographical Names Service (www4.rncan.gc.ca/search-place-names/unique/GBEIN). In all three cases, you will end up at Natural Resources Canada, Earth Sciences, where the option Geography leads to the Canadian Geographical Names button. When we go to the Geographical Names Search Service, we can type in names, coordinates, define rectangular areas or give unique name ID's; this last facility also provides the possibility to type in aboriginal characters. The results can be displayed according to different options, ordered according to feature type or region, and when the looked-for entry has been selected additional information is given, such as a feature identifier; also, the dates when the name changed status (and the institution responsible), for instance when it became official.

Figure 20-1 Result of a query for Kerfoot in Canada

The Canadian National Atlas site provides more ways of access to the names database, through its Toporama mapping tool, as a location may be found here by typing in a place name, a map number (national topographic system number), a postal code (FSA), a street address, street name, or map coordinates. When a name has been typed in, a list of entries will come up, providing access to search-matching place name results (with different locations and feature types).

20.4 Estonia

The Estonian place name database (KNAB), developed by the Institute of Estonian Languages EKI can be accessed through http://www.eki.ee/knab/knab.htm . The database also contains street names, names of institutions, companies and organizations, farms, administrative units and natural features next to the names of populated places. There are detailed photo and map options for showing the named feature in situ. As Estonian names in the past had their German and Russian variants, much attention is paid to languages of the place names. The database records inform on principal name forms and variant names, the status of the names and the respective decisions by local authorities (parish, county or town council, parliament), feature designation codes, present administrative

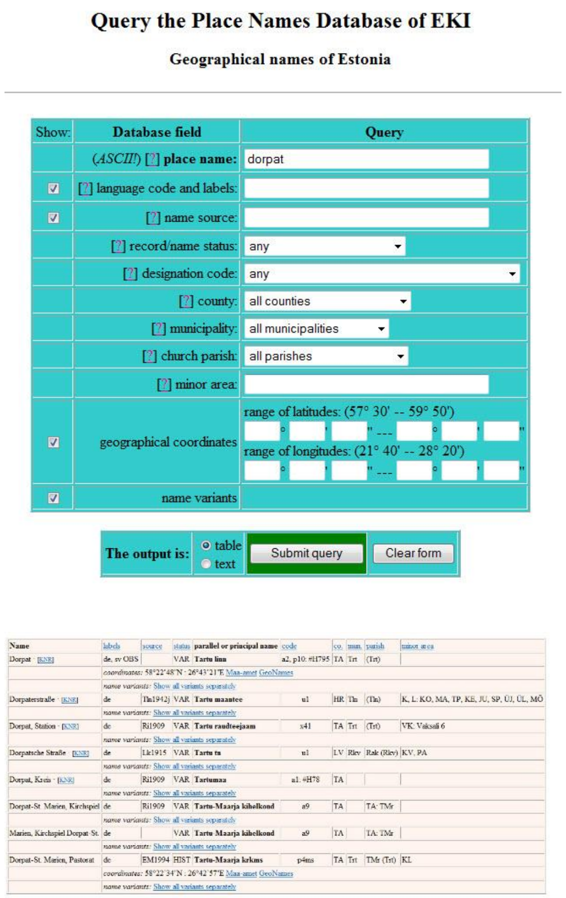


Figure 20-2 Query and entries for Dorpat in the Estonian place name website.

division, and geographical coordinates. The database also contains information not included in the website, such as data on the temporal extension of the names, short textual descriptions of the object, use of locative cases, names of superior features, comparisons to other names and historical background information). The website, which can also be accessed in English and Russian, also provides information for named features outside Estonia (with more place names from adjoining countries than from countries further away); according to the description they will specially collect names for linguistic minorities, such as the Basque, Chechen and Udmurt.

20.5 Finland

The Finnish national names website, found at (http://kansalaisen.karttapaikka.fi/kartanhaku/osoitehaku.html?lang=en-GB), has the option to query by address, by place name, by map sheet number and by

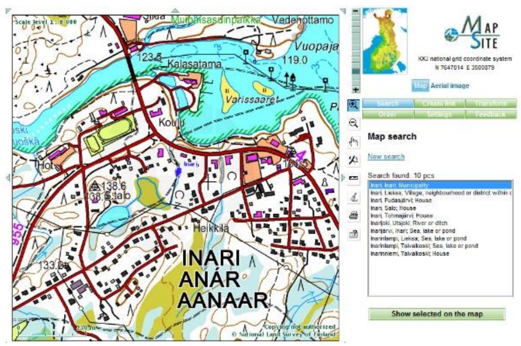


Figure 20-3 Answer for a query of Inari (with its Sami variant names) on the Finnish Place name server

coordinates. When queried by place name, a list of matching answers will come up, with their feature codes and province, giving access to a large scale cartographic representation. This map is very clear and has reproduction-like quality. A layer with cadastral information can be added to the map. Instead of maps, satellite images can also be opted for.

20.6 France

In France, the Géoportail (http://www.geoportail.fr/) provides geographic and cartographic information delivered by the Institut Géographique National (IGN). The names part of this website is displayed against a map or satellite imagery background which can be customised, rendering at will the transport network, land use, heights, hydrography, public services and administrative boundaries. It can also be customised temporally, as the *Carte de Cassini* (1750s) or the *Carte de l'Etat Major 1:40 000* (1860s) can be chosen as map background, thus allowing a check whether the place name spellings have changed over time. The various homonyms are differentiated between by adding the departments they are situated in; highlighting an entry will lead to a large scale photographic or cartographic representation of the named object.

20.7 Germany

As for Germany, a new Gazetteer- web service was developed by the Federal Agency for Cartography and Geodesy (Bundesanstalt für Kartographie und Geodäsie or BKG) in co-operation with the German Permanent Committee on Geographical Names (StAGN) in autumn 2006. This (Gazetteer-) web service is based on a Web Feature Service (WFS) and thus compliant to the respective Open Geospatial Consortium (OGC) specification. The homepage (http://www.geodatenzentrum.de/geodaten/gdz_rahmen.gdz_div?gdz_spr=eng&gdz_user_id=0) accesses both Geographical names and a Historical place names option. The first one provides the content of the single database called Geographical Names of Germany (GN-DE) comprising also geographical names in the Sorbian and Frisian languages. Here, names can be entered (with wildcards), and those that answer the query will be listed in a table, also informing about the coordinates, feature type, coordinates, size, height, number of inhabitants, language and status of the name, with a map reference which, when clicked on, will lead to a large scale portrayal of the named object.

The historic place names section (http://www.geodatenzentrum.de/geodaten/gdz_rahmen.gdz_div?gdz_spr=deu&gdz_akt_zeile=3&gdz_anz_zeile

=5&gdz_user_id=0) informs about the German names of areas formerly (between 1900 and 1945) under German jurisdiction, and allows German civil servants to search for the pre-1945 birthplaces of German citizens.

20.8 Iran

The national geographical names data base of Iran (http://gndb.ncc.org.ir/Pages/EnglishSearch.aspx?d=bBuyOT3PST4LFMPpuL3K10yEAoc7ce4B&r=yrqgaicm) allows for the input of a name in Roman characters as well, with the possibility to select feature types and geographical divisions like provinces, counties and districts. A list of matching names will come up, in Arabic and Roman script, with feature types. Clicking a name will highlight its location on the accompanying map, and can also provide additional information, like coordinates, both transcribed and transliterated Roman name

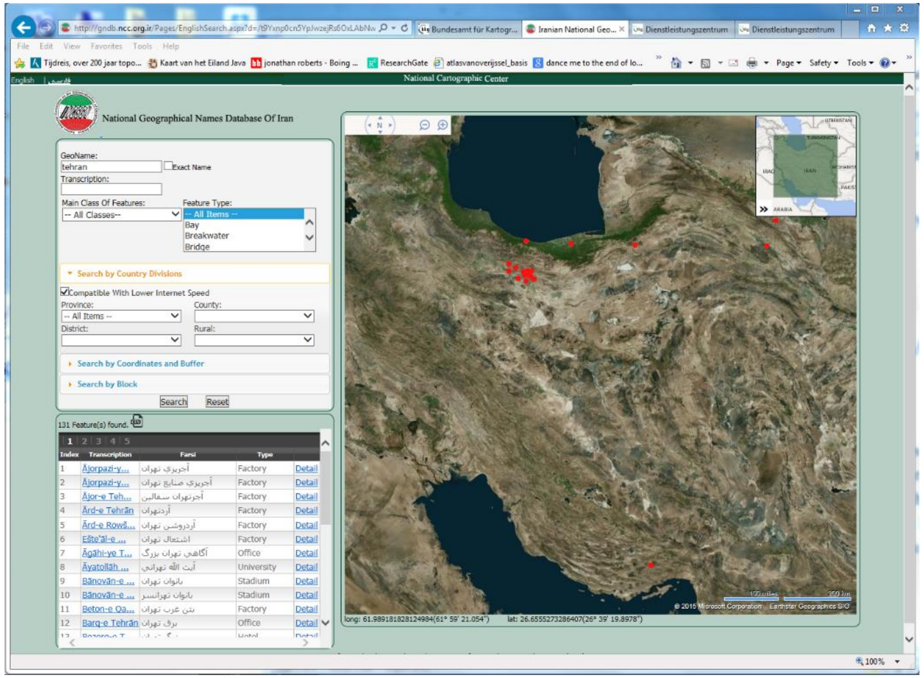


Figure 20-4 Part of all the names containing Tehran, on the Iranian name server.

versions, height information, etc. It is also possible to search names by coordinates with a buffer and by block. In the latter case, all the names on a particular map sheet will come up. The alphabetic sequence of the names delivered is that of the Arabic alphabet.

20.9 Ireland

Fiontar, the Irish Language School of the Faculty of Humanities and Social Sciences of Dublin City University, has created the *Placenames Database of Ireland* (Logainm) in collaboration with the Placenames Branch (Ministry of Arts, Heritage and the Gaeltacht). This is a database for archival records and place names. Its website (http://www.logainm.ie/en/) is aimed at journalists and translators, students and the public at large. Launched in 2008 it was later extended with sound files, mapping tools, Irish-English and reverse translation options, scanned images of archival records, and tools to explain Irish generics in place names. In figure 20-5 an example is given of the distribution of one of the Irish generics, drom (English: Ridge) over the island. There is ongoing cooperation between Logainm.ie and the Northern Ireland Place-Name Project (http://www.placenamesni.org/index.php) to create

Figure 20-5 Distribution map for a generic, from the Irish place name server.

links between the two databases for shared place names. Place names can be entered into the search box, either in English or in Irish, and results will pop up as clickable pins on the map and also as a list below the map. By clicking a pin on the map or a name on the list underneath, more results will appear.

A special aspect of the Irish name server is the clickable map on the home page. Here one can click on any of the counties in the island, and then a list pops up of all the features named there, with their numbers. In Country Kerry for instance there would be 66 named bridges and 45 promontories, and by clicking a category, all the items in it will be displayed by clickable red points on the map, as in figure 20-6.

Figure 20-6 Distribution of all promontories in County Kerry, Ireland, with one of them clicked.

20.10 New Zealand

The website for accessing New Zealand place names has the url http://www.linz.govt.nz/regulatory/place-names/find-place-name . The production of the

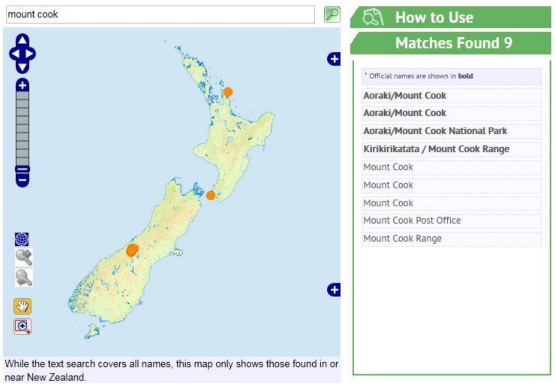


Figure 20-7 Dual names from the New Zealand gazetteer

Gazetteer, on which this website is based is described in another chapter in this manual, by Wendy Shaw of the New Zealand Geographic Board. This body, housed at Land Information New Zealand (LINZ) collects, adopts, approves (or assigns, alters and discontinues) and validates names. When adopted the names are listed in the New Zealand Gazetteer, with information on their status, feature class, coordinates, a short description of the feature, its extent, and – and this is not often found in these name servers – something on the history of the named object or origin and meaning of the name. File and archive references will be added if possible.

Special name categories are dual names, alternative names and recorded names. In dual names, the community has expressed its recognition of the special historical and cultural significance of both original Māori and non-Māori names, as for instance Aoraki / Mount Cook. They would be inseparable on official documents. For names from the alternative name category this special historical and cultural significance for the community is also valid, but they may be separated, in the sense that only one of them may be selected.

Names from the 'Recorded names' category have not officially been approved as yet. However, they have been cited in at least two publicly available authoritative sources. It may simply be the case that NZGB has not had the time as yet to validate them, or that they are beyond the jurisdiction of NZGB, as is the case for names for homesteads, roads, streets, tracks and lighthouses.

20.11 Norway

On the website http://norgeskart.no/ssr/#5/378604/7226208 Norwegian geographical names can be found; The search can be done with the names, with coordinates, bounding boxes or street addresses. The map will show the points that match the queried name, and by clicking them additional information will pop up: language and status of the name (and date of change of status), feature class of the named object, and ID. The maps on this site can also display all names, categorised according to their language (Norwegian and several Sami languages), to their name type (administrative, cadastral, physical, hydrographical, coastal, vegetation and other names), and to their status (authorized and non-authorized names.

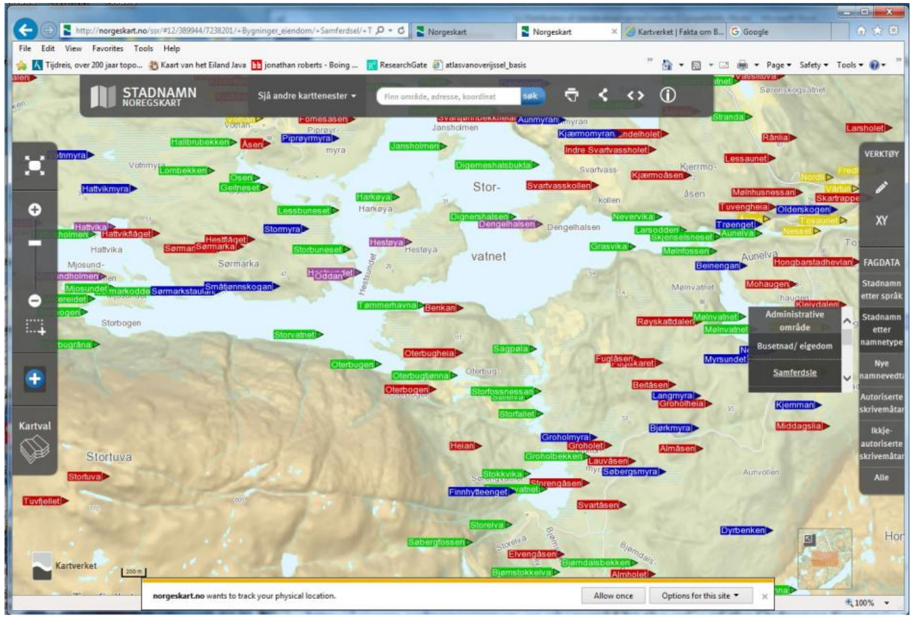


Figure 20-8 All place names, colour-coded according to feature-type, on the Norwegian names server

The names can also be projected on older map series, so that one may see as well whether any changes have been applied. Additional names data provided are the municipality the named object belongs to, and its UTM zone.

20.12 Sweden

http://kso.lantmateriet.se/# shows the location of place names on the map, as well as their language, province, municipality, coordinates and name type. The site provides extra functionality relevant for names research, as it also can show all the names that begin or end with a specific letter string or have them in the middle. Figure 20-9 shows a map of Blekinge province with all the geographical names there ending in –borg.

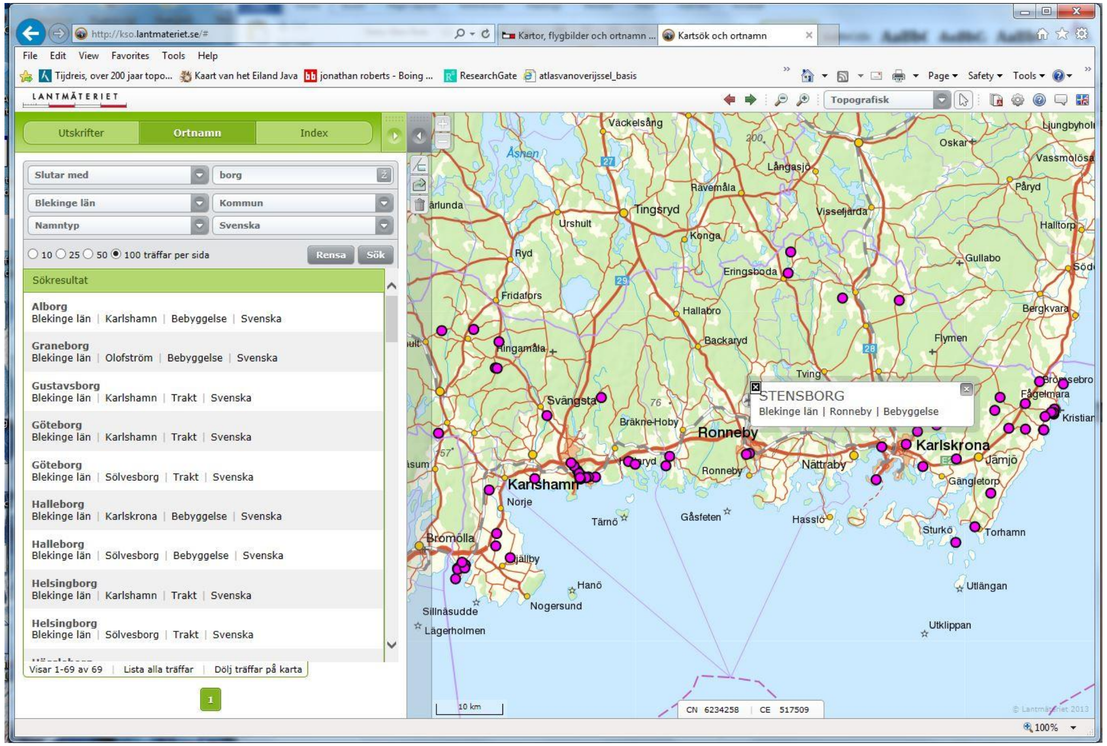


Figure 20-9 Names search result on the Swedish names server

20.13 United Kingdom

Ordnance Survey (OS) geo-referenced street dataset ("National Street Gazetteer" (NSG)) can be accessed at http://www.thensg.org.uk/iansg/welcome.htm, while the place name search can be effectuated at http://www.getamap.ordnancesurveyleisure.co.uk/; they each provide different numbers of matches to the queries, most being given by the explore site. As with the other sites shown, the place name search functionality is part of a service allowing the customers to find the proper map they want to buy. The added functionality here is that also map-related services are provided, such as walking tours. Find a route, create a route, blogs about the routes, are all related services here.

20.14 United States

The USGS Geographic Names Information System (GNIS) of the US Board on Geographic Names allows one to query or download domestic or foreign names. For querying the domestic name server, a feature name can be entered (restricting the search by entering features, IDs, states or counties can speed up the search process) and as a result a table of all the names matching the

Feature Query Results

Click the feature name for details and to access map services
Click any column name to sort the list ascending ▲ or descending ▼

Feature Name ▼	ID	Class	County	State	Latitude	Longitude	Ele(ft)	Map	BGN Date	Entry Date
West Batavia Cemetery	420782	Cemetery	Kane	IL	415025N	0881843W	702	Aurora North	-	15-JAN-1980
Tri - City Ambulance - Batavia 2	2673792	Building	Kane	IL	415046N	0881957W	725	Aurora North	-	26-OCT-2011
Tri - City Ambulance - Batavia 1	2673856	Building	Kane	IL	415104N	0881734W	718	Aurora North	-	26-OCT-2011
Township of Batavia	428640	Civil	Kane	IL	415005N	0881858W	702	Aurora North	-	01-SEP-1995
East Cemetery	407599	Cemetery	Kane	IL	415141N	0881821W	722	Aurora North	-	30-SEP-1999
Congregational Church of Batavia	1998721	Church	Kane	IL	415056N	0881844W	712	Aurora North	-	30-SEP-1999
City of Batavia	2394077	Civil	Kane	IL	415056N	0881830W	666	Aurora North	-	27-FEB-2008
Batavia Wastewater Treatment Facility	2352366	Locale	Kane	IL	415043N	0881831W	659	Aurora North	-	18-SEP-2007
Batavia Senior High School	1816208	School	Kane	IL	415053N	0881948W	732	Aurora North	-	25-MAR-1999
Batavia Post Office	1816207	Post Office	Kane	IL	415059N	0881830W	669	Aurora North	-	25-MAR-1999
Batavia Middle School	1816206	School	Kane	IL	415057N	0881840W	709	Aurora North	-	25-MAR-1999
Batavia Landmark United Methodist Church	1998718	Church	Kane	IL	415101N	0881841W	709	Aurora North	-	30-SEP-1999
Batavia Junction	1771669	Populated Place	DuPage	IL	414743N	0881426W	735	Naperville	-	07-JAN-1998
Batavia Fire Department Station 2	2681947	Building	Kane	IL	415046N	0881957W	725	Aurora North	-	30-JUL-2011
Batavia Fire Department Station 1	2682077	Building	Kane	IL	415104N	0881734W	718	Aurora North	-	30-JUL-2011

row(s) 1 - 15 of 19

Figure 20-10 Result for the query for the toponym Batavia in Illinois state, on the US domestic names server.

query shows up, with info on ID, feature class, county, state, coordinates, elevation (in m or feet), and the map sheet on which the named object can be found. Clicking on the name will lead to additional coordinate information and mention of variant names.

20.15 Overview

Most name servers are related to the national mapping agency. A number of them have as a prime function to find the necessary maps, but some have more extended functionality for scientific purposes (giving more information about the name (etymology, history) or their variants. Several countries have the option to go for historic map series as map background, so that the past orthographies of the queried names can be studied. The Irish site has special attention for generics and their distribution over the island, the Canadian site has special scripts for allowing Inuktitut names to be rendered, as does the Estonian site which also allows for input in Cyrillic characters. Nearly all of them are linked to a map option where the named object can be seen on the map or on a satellite image. These maps or images can be customised, adding information layers at will. The information provided for the names sometimes extends to providing their ID numbers, and dates when they were standardised or entered into the names database.

Many of the sites also allow for downloading the names, in total or per administrative area or bounding box/coordinates.

Depending on the language characteristics additional information is needed, for example in Estonia, the name Talinn would be different depending on its use in a sentence (I am coming from Talinn, going to Talinn or living in Talinn) would have different case ending and thus the locative case is also mentioned.

Some countries have extended the names information to areas beyond their current borders, in order to allow for administrative purposes. Many of the countries have also the option to use English next to their native language, to allow for use by the international community – this is especially admirable when it involves the use of another script – as for instance on the Iranian website where both Arabian and Roman scripts are used.

Finally, some websites also mention variant names – this may not seem to help the standardisation process at first sight, but it actually helps to identify the standardised name version, as by entering the variant name, the official spelling will be found.

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

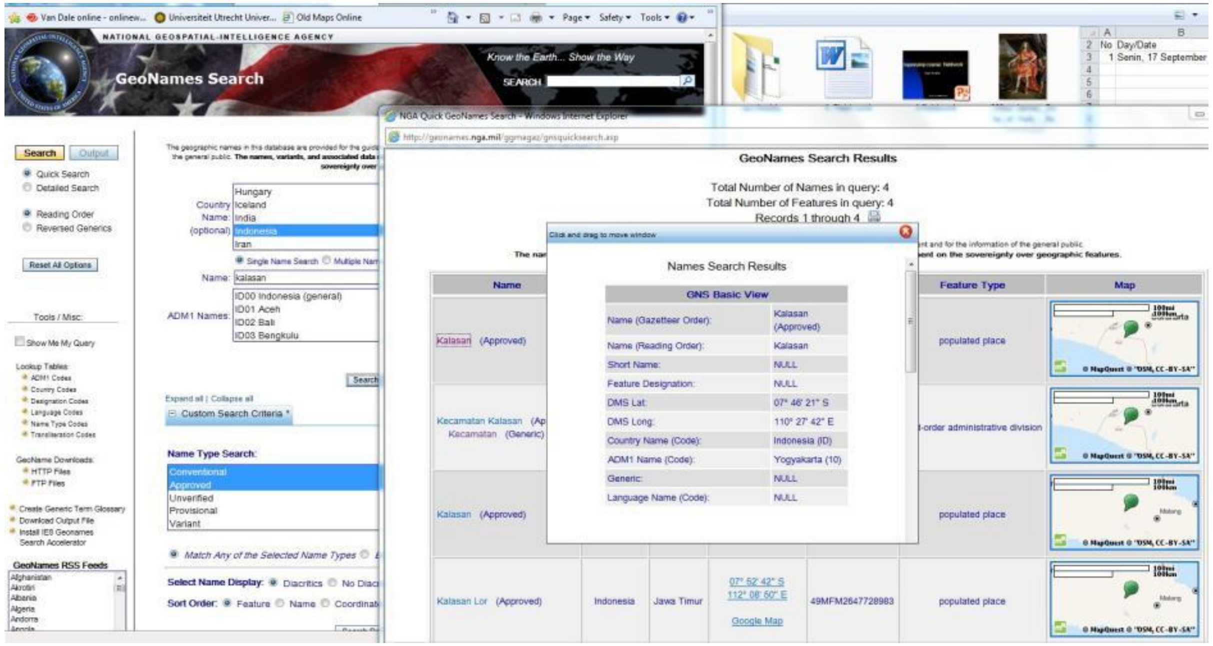


Figure 21-1 Example from GNS: looking for Kalasan, Java

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.

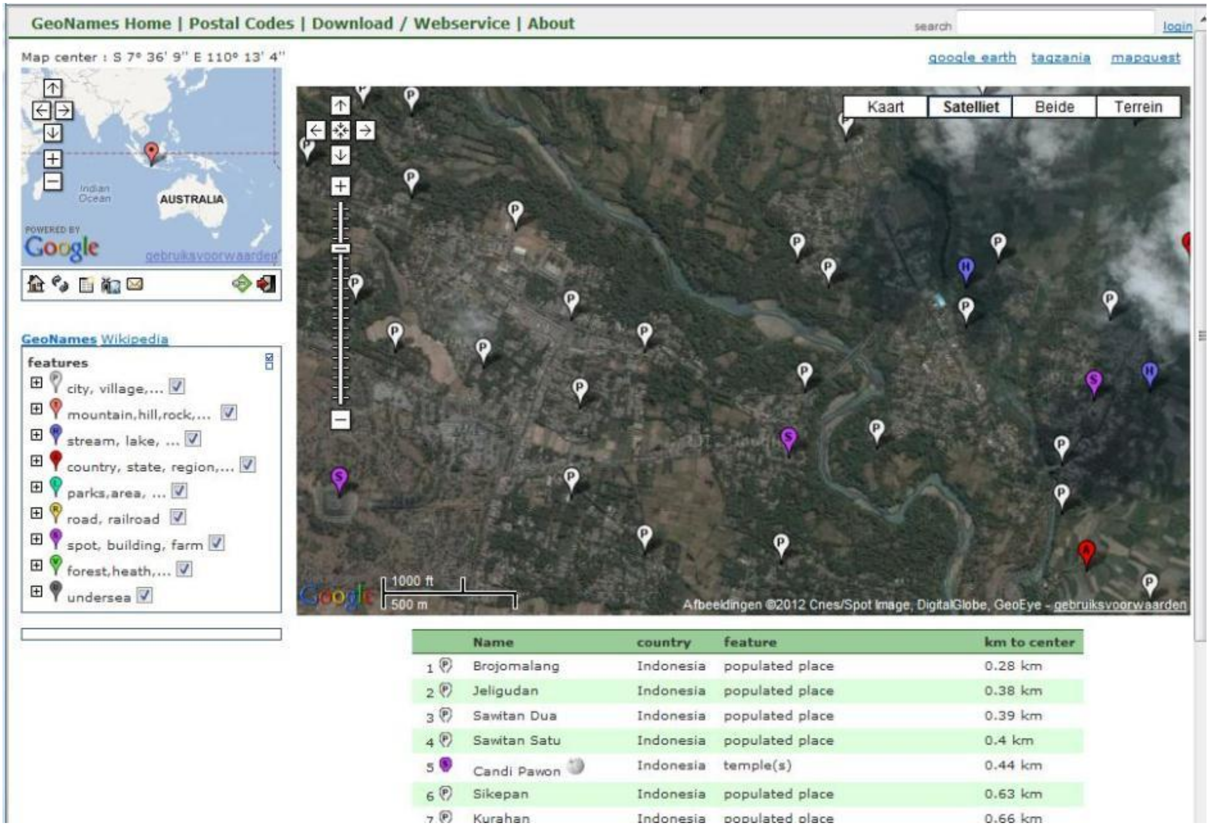


Figure 21-3 -Image from GeoNames server, with satellite image showing the named objects represented in the database, in a search for Candi Pawon, on a line with Borobudur and Mendut (purple S at right) on Java island. On the current GeoNames version (2015) the colour-coded droplets now have changed to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

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	cemeteries 64,535	dunes and beaches 8,854
parks 20,408	historical sites 66,228	mountains and mesas 367,723
political areas 32,623	hydrographic structures 123,991	plains, gaps and valleys 59,932
countries 165	canals 21,482	plateaus and flats 5,887
populated places 2,000,821	dam sites and reservoirs 95,804	seafloor items and reefs 10,382
reserves and tribal areas 8,887	mine sites 24,070	volcanic features 2,262
hydrographic features 636,564	monuments 8,560	regions 120,132
bays, fjords and gulfs 36,974	recreational facilities 7,526	biogeographic regions 44,782
channels 13,874	storage structures 11,969	deserts 588
ice masses 3,569	towers and telecom. features 26,663	forests 19,476
lakes 94,758	transportation features 77,933	grasslands 4,419
seas 273	wells 71,680	wetlands 17,887
streams and rivers 480,921	physiographic features 575,964	Islands 70,094
land parcels 12,424	arroyos 45,359	
manmade features 858,145	bars (physiographic) 12,167	
agricultural sites 174,912	basins 8,957	
buildings, residential sites 243,448	capes and cliffs 25,264	

Table 21- 1. Feature classes in ADL and their numbers

Presently, the site seems to be under reconstruction.

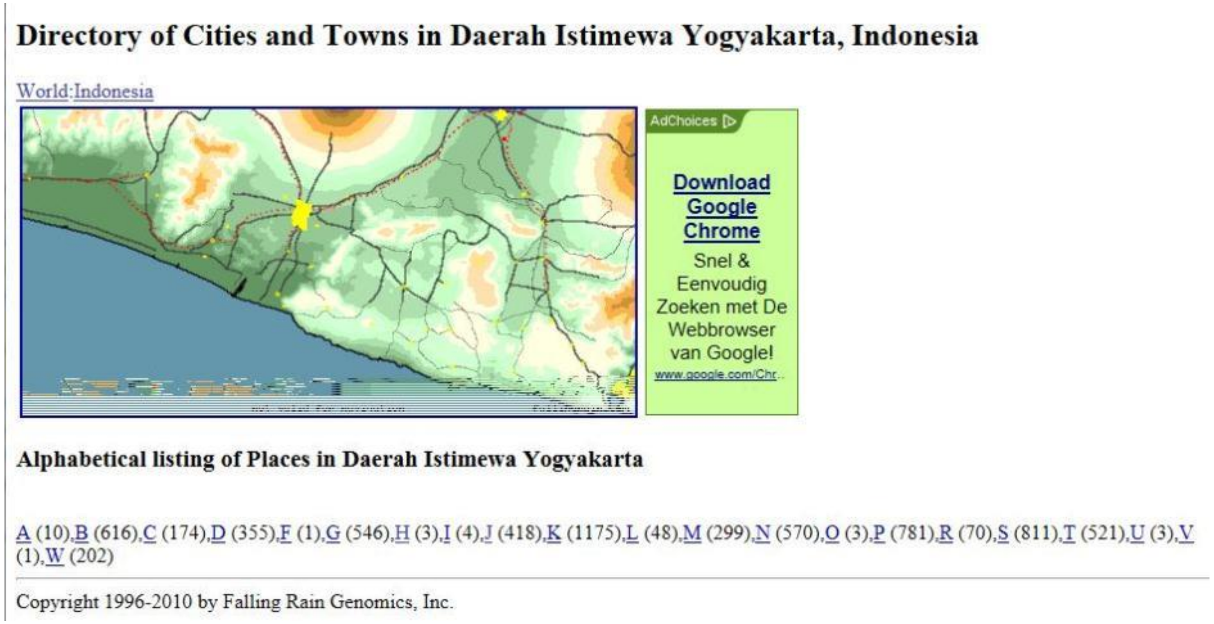


Figure 21-6 Global gazetteer screen dump

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

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

temple name	Getty/TGN	GeoNames	fuzzyG	Alexandria 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

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.

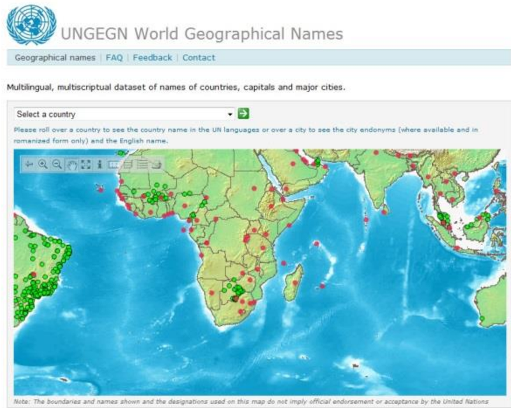


Figure 21-7 Screen dump from UNGEGN World Geographical Names Database

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.

Section 9 Cultural aspects

Chapter 22 **Geographical Names as Cultural Heritage**

Botolv Helleland

22.1 Introduction

What is meant by "cultural heritage", and what makes a geographical name an object of cultural heritage? According to Wikipedia "cultural heritage" is the legacy of physical artifacts and intangible attributes of a group or society that are inherited from past generations, maintained in the present and bestowed for the benefit of future generations. Cultural heritage includes tangible culture (such as buildings, monuments, landscapes, books, works of art, and artifacts), intangible culture (such as folklore, traditions, language, and knowledge), and natural heritage (including culturally significant landscapes, and biodiversity).[1])

The subgroup "intangible culture" also includes geographical names, as names are part of a nation's language(s). However, the intangible aspects of the cultural heritage have attained much less attention than the tangible objects. Although oral traditions, local languages, and names have been looked upon as worthy of protection and preservation, they have not had the same legal status as for example buildings and monuments.

Geographical names (place names) are in fact an important part of the intangible cultural heritage. First of all, it should be underlined that most geographical names were coined at the very spot where they are used – or were used, if they have gone into oblivion. Thus, they are memories of the place as well of the people who gave the names.

Secondly, geographical names normally contain linguistic elements which are semantically related to the scene of naming. That means that they tell something about the place and the circumstances at the time of the coining of the name.

Thirdly, the linguistic material contained in the geographical names is a valuable source for the study of language history.

Every name may be considered as a window to the past; that is to the period of time when the name came into being. Many geographical names have been borrowed and reused in other places where they did not belong originally. Such names are, however, also historical documents as they convey information on the background of the name givers, for instance their original homeland and the emotional ties to where they came from.

Every area, every greater or smaller part of the Earth's surface, has a number of geographical names that together make up an important part of the area's history. It is a matter that touches upon almost every kind of human activity as well as natural circumstances.

22.2 United Nations and geographical names as cultural heritage

Geographical names as cultural heritage have for many years been on the agenda of UNGEGN sessions, and the United Nations Conferences on Geographical Names have adopted several resolutions relating mainly or partly to this topic. The resolutions can be accessed at the following website: https://unstats.un.org/unsd/geoinfo/UNGEGN/docs/RES_UN_E%20updated_1-10%20CONF.pdf . At this point we give the title of the resolutions in question and number of conference as well as the number of resolution in sequence:

V/6 Promotion of national and international geographical names standardization programmes

VIII/9 Geographical names as cultural heritage

IX/4 Geographical names as intangible cultural heritage

IX/10 Support for training and publications

X/3 Criteria for establishing and evaluating the nature of geographical names as cultural heritage

X/4 Discouraging the commercialization of geographical names

Convention for the Safeguarding of the Intangible Cultural Heritage

Geographical names as a part of the world's cultural heritage became more of a focus when UNESCO initiated the *Convention for the Safeguarding of the Intangible*

Cultural Heritage. The convention was adopted by the UNESCO General Conference on 17 October 2003[2]) and entered into force 20 April 2006 for the thirty States that had ratified it on or before 20 January 2006[3]). As at 14 April 2015 the Convention had been ratified by 163 states.

Unlike other UNESCO conventions, this convention begins with stating its purposes which are:

(a) to safeguard the intangible cultural heritage;

(b) to ensure respect for the intangible cultural heritage of the communities, groups and individuals concerned;

(c) to raise awareness at the local, national and international levels of the importance of the intangible cultural heritage, and of ensuring mutual appreciation thereof;

(d) to provide for international cooperation and assistance.

Intangible cultural heritage refers to "traditions or living expressions inherited from our ancestors and passed on to our descendants, such as oral traditions, performing arts, social practices, rituals, festive events, knowledge and practices concerning nature and the universe or the knowledge and skills to produce traditional crafts" [4])

Although geographical names are not mentioned explicitly in the Convention they are included in items related to oral traditions. UNGEGN has referred to the Convention in various contexts and also passed a particular resolution (IX/4) addressing the importance of preserving geographical names as a part of the intangible cultural heritage of a nation. Not the less, this was also UNGEGN's view even before the Convention was signed.

Proceedings of the International Symposium on Toponymy 2014

On 7–9 November 2014 an international symposium entitled *Geographical Names as Cultural Heritage* was held in Seoul, South Korea. Eleven Papers were presented by experts from nine countries. The proceedings were published in 2015[5]) On the back cover of the volume the following introduction to the topic is given:

> "Discussing elements of cultural heritage in toponymic research may not be a new trend. Each place name has its own *raison d'être* in that it was created by name giver's perception of identity of the place. All these identities have been formulated on the basis of cultures of each society; thus we would say every place name has elements of cultural heritage, in whatever form or context.
>
> What brought the cultural heritage issue to the fore was the realization of the United Nations Group of Experts on Geographical Names (UNGEGN) that geographical names with cultural heritage values should be protected from disappearing. Initiated by the focus on respecting diverse identities contained in geographical names, especially of indigenous, minority and regional language groups, interests have been extended to finding aspects of cultural heritage in every group of geographical names and dealing systematically with them in toponymic research and standardization. "

The former UNGEGN Chair Helen Kerfoot traces the Group of Experts' commitment and activity on geographical names as cultural heritage from the First Conference in 1967 up to the Tenth Conference in 2012, and in the work of UNGEGN, whereas her successor Bill Watt examines some Australian examples of cultural aspects of naming practices in the context of the UNGEGN Working Group on Geographical Names as Cultural Heritage. These two articles together with the other contributions in the volume throw light on a variety of aspects related to geographical names as cultural heritage. Furthermore, the topic is treated in a number of other works, journals, and articles worldwide, for instance, the collection *Geographical Names as Part of the Cultural Heritage* (2009)[6]) and *ONOMA* (the journal of the International Council of Onomastic Sciences), if not specifically, then indirectly. In fact, most publications on geographical names touch upon aspects relating to cultural heritage. Internet searches through national and international library bases would give access to an amount of relevant literature.

22.3 Geographical names: the memory of places

Geographical names, or place names, may be said to represent the oldest living part of human cultural heritage, in the sense that they have been handed down orally from generation to generation for hundreds and even thousands of years. In the continents overseas, where Europeans introduced their languages and their names, many aboriginal names disappeared because they were discarded by the new settlers. Luckily many such names have survived as a testimony of earlier civilizations. In many instances, geographical names are the only surviving elements of extinct languages.[7]) When one takes into account that many geographical names have been coined as descriptions of the area or place where they belong, it becomes evident that we are dealing with material of immense historic value. But also more recent names of dwelling places, streets and fields, as well as micro toponyms, form a part of the collective memory and heritage expressed in geographical names. All the geographical names of the Earth put together represent an immeasurable body of human experience and description of natural circumstances.

Although many people do not seem to be particularly aware of the historical richness of the place names in their surroundings they may still feel that the name stock contributes to their link to the land. People who have become acquainted with the etymological contents of names will of course appreciate their historical value to a greater degree. Through the place names of a district small or large, in particular its settlement names, it is possible to read many details of the area's history.

Many place names are also identified with past events and are pegs upon which stories both written and oral can be hung. One can also see geographical names as a reflection of the interplay between humans and nature through different periods of time. Besides, if a person has some meta-linguistic and historic awareness, he may listen to place names as voices of the past, which in its turn may strengthen his feeling of home. Thus, place names can function as a textual representation, in some cases in obsolete words, of the historic landscape.

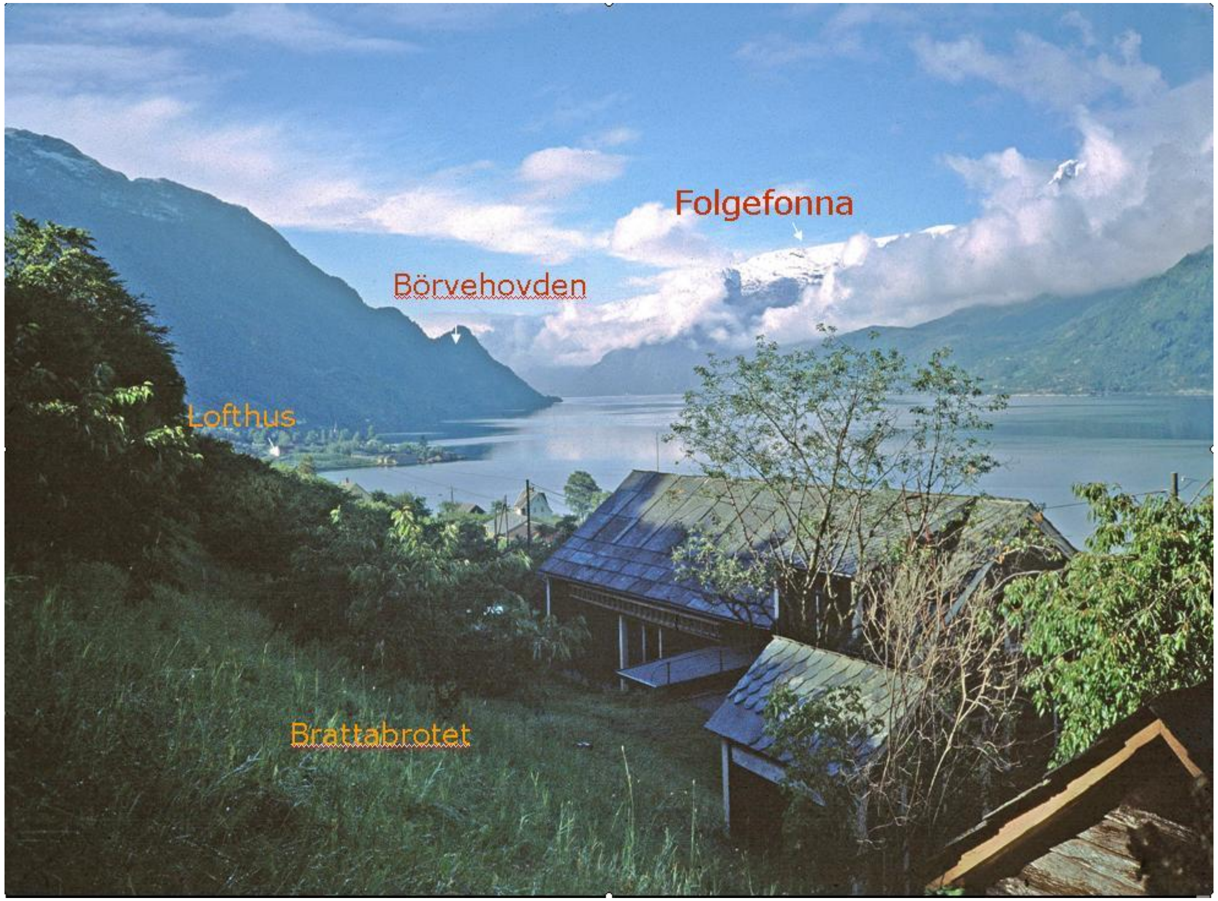


Figure 22-1. The farm Helland in Lofthus. In the background, the hill of Børvehovden and the Folgefonna glacier. In the foreground, the field of Brattabrotet. Photo: B. Helleland

A place name not only points out a place, it also mediates a cluster of qualities and meanings attached to that place, partly valid for a single individual, partly shared by a given social group. Everybody over a certain age who has spent sufficient time in the village of Lofthus in western Norway will identify the hill of *Børvehovden* (see picture) when the name is mentioned. Another example from this setting is the field name *Brattabrotet* "the steep slope" on the small farm of

Helland. This name is known only by the family living on the farm and is associated with the difficulty of mowing and harvesting this field due to its steepness.

A way of elucidating the historical contents of place names is shown in figure 22-2.

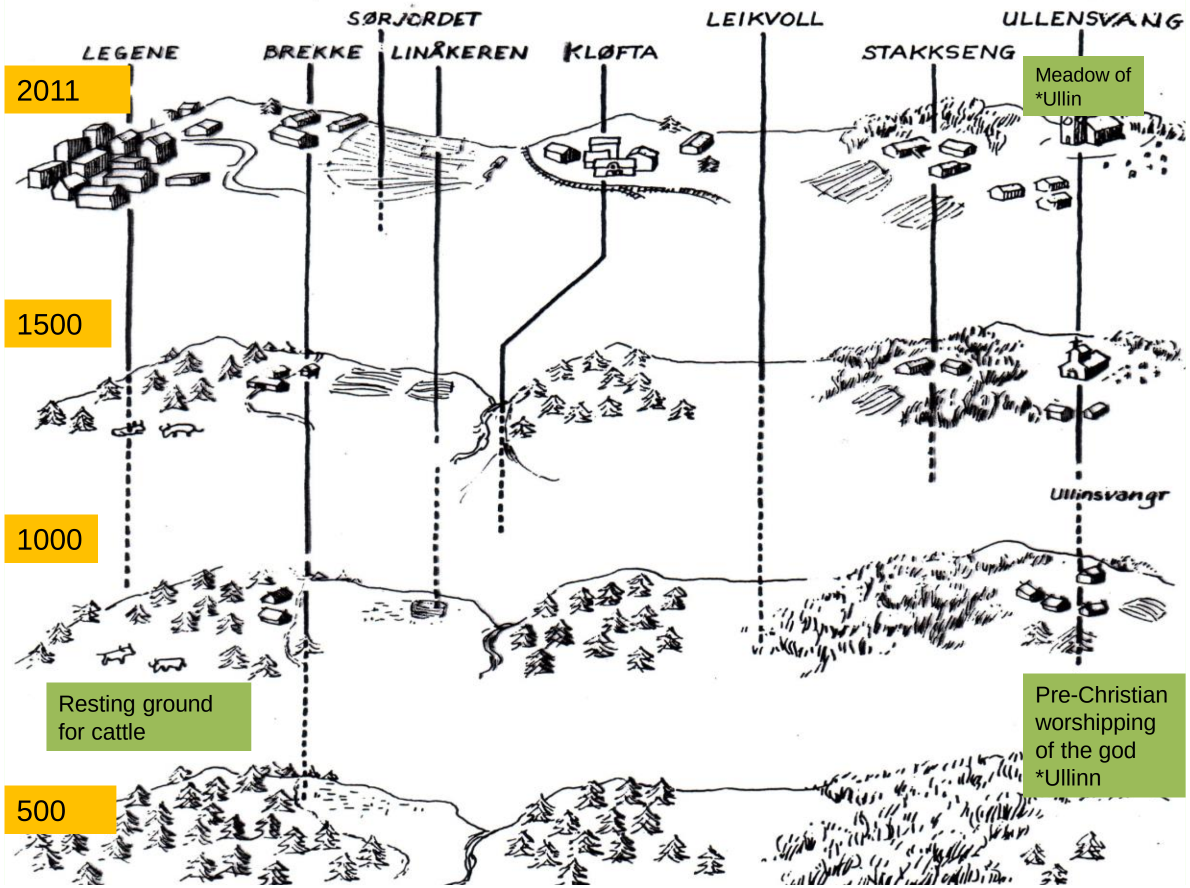


Figure 22-2. A named historical landscape as imagined at various periods of time.

At the top of the drawing in figure 22-2 samples of existing place names are shown, each dating back a certain distance into the past. The suggested period of name coinage is based on written sources and typological characteristics of the names and of the named places. The vertical lines, moving upwards, follow the passage of time. New names may be coined at any time, as old names sometimes fall out of use and become obsolete. The places to which the names refer are also constantly changing, sometimes unnoticeably, sometimes noticeably, but they may still be considered to be the same place.

If we look at the names mentioned in figure 22-2, it is clear that their referents are something else today than they were in the past. *Legene* on the left is, for instance, used for a settlement which is only a couple of decades old. The etymological meaning of this name is “the place where the cattle rest”, and that is what the name most probably referred to when it was coined. The name is not testified in writing before the 19th century, but it is no doubt much older. *Brekke*, reflecting Old Norse *brekka* f. “slope, steep road” is now used in reference to a farm. It is recorded in the 14th century, but the importance of the name and its onomastic typology indicate that it could be several hundred years older. The last example I will mention is *Ullensvang* on the far right of the drawing, the name of a parish and the farm on which the parish church was built (see also the picture and map above). It is recorded many times in the 13th and 14th centuries. As the name is apparently composed of the pagan god’s name *Ullin* and *vang* “meadow where people assemble”, there is reason to believe that it was coined in the centuries before the introduction of Christianity in Norway around the year 1000. Today – as probably a millennium ago – people use the name without thinking of its heathen background.

A representative expression of the viewpoints of communities of name scholars in this field is to be found in a report that came out of a symposium on place name care and planning held in Finland in 1977:

> Place names are an important part of our geographical and cultural environment. They identify localities of different kinds and represent irreplaceable cultural values of vital significance to people's sense of belonging and well-being.[8])

Place names function as a social consensus or a sort of an agreement reached over many generations. This quality, I feel, is another contribution to the well-being and ease of people who are familiar with the area and its names, although the name users may not be very conscious of it. Orally the place names of a local area are mostly used in a dialectal form, at least in Norway, which also reflects a social consensus. When local people see place names written on road signs and on maps they often react negatively because they feel that the standardized written forms of the names break with their idea of the correct name. This conflict, which sometimes makes the standardization of place names problematic, is an issue which needs to be discussed in relation to the preservation of place names as an intangible part of cultural heritage. This lies, however, outside the scope of this paper.

In 1999, Finland chose "Place Names – The Memory of Places" as theme for its European Heritage Days. The aim of Heritage Days is to awaken people to see and evaluate the environment built around them, and to help them appreciate its beauty and variety. Place names constitute an indispensable part of the environment, and the aim of Finland's 1999 Heritage Days was to make people aware of the use of such names in the planning process. The Heritage Days activity was used as a means of educating the public about the importance of standardizing geographical names without losing their historical value. During that year a number of events took place, with many people involved, and the organizers had good reason to call the arrangements a success. As part of these activities, a video was prepared by the Research Institute for the Languages of Finland, which was shown on Finnish television.[9])

22.4 The function of geographical names

A geographical name normally exists in relation to a geographical entity, real or imagined, and if this entity changes character, the reference of the name may also change. This reference address function of geographical names means that they constitute an essential linguistic tool that is used by most people many times a day; and for this tool to function as well as possible in an increasingly complicated society, it is incumbent upon the authorities to regulate naming in certain ways and in certain contexts. For example, it must be determined what name is to be chosen, in which standardized form, and to what entity it applies. This reference address function of geographical names is primary, but like other parts of language geographical names function on several levels, including:

• The cognitive level (identification of an entity), for instance, the North Sea, referring to a sea in Europe.

• The emotive level, for instance, names relating to places that evoke good or bad feelings (example: Paradise and Hell).

• The ideological level, for instance, Bethel as the name of a religious assembly house and Karl-Marx-Stadt for the former (and present-day) Chemnitz.

• The community-constituting level, for example, Saami-speaking people identifying with Deatnu and Norwegian speakers with Tana, both names referring to the same river in northern Norway.

• All persons are surrounded by named features with which they are familiar, and which they know as their own. This also applies to the names of the features. At the same time, there are other names that they share in common with larger groups in what one may call a structural hierarchy.

This can be demonstrated by a simple three-level hierarchy comprising:

• The geographical names that are common to the people in a micro community (family, farm, village).

• The geographical names that are common to two or more micro communities (rural district or town).

• The geographical names that are common to many or most of the communities of a society.

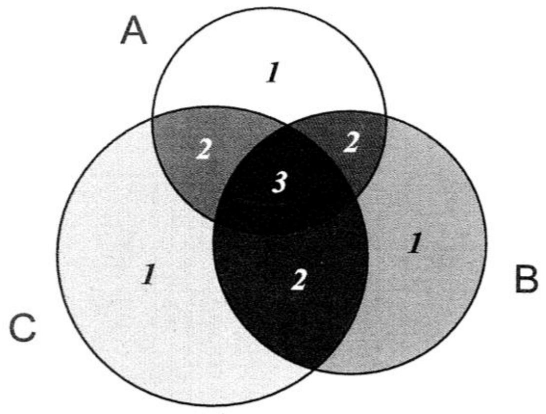


Figure 22-3 Conceptual representation of geographical names knowledge. Note: ***A*****,** ***B*****,** ***C*** *refer to collective geographical names knowledge of individuals in a micro community.*

Figure 22-3 illustrates various levels of names knowledge shared by different communities in this particular model. Each circle (A, B, C) represents the collective geographical names knowledge of individuals in a particular micro-community. People who belong to the same language community will on the whole be familiar with the same names and feel attached to them—even more so if they belong to the same local community. In multilingual societies, people speaking different languages will apply different names. These names not only form a body of reference addresses, but constitute a part of linguistic and local identity. Also, through their close connection with the dialect, geographical names represent an inherent human value; and for the individual, certain geographical names may be connected with life experiences.

Examples of geographical names used in various groups of name users:

1 Names that are common only to the people of a particular micro-community, for example, *Hesthaug* 'horse hill' referring to a small hill; 'Leikvoll 'playing ground' referring to a meadow where locals use(d) to gather.

2 Names that that are known and used by the people of more than one micro-community, but that are not widely known in the area, for example, *Opedal*, for a group of farms; *Eidfjord*, for a local municipality.

3 Names that that are known and used by many or most people of a society or nation, for example *Lofoten*, a well-known group of islands; *Oslo*, a big city and capital.

22.5 Giving geographical names

The preservation of geographical names encompasses first and foremost the inherited and spontaneously given body of names, but also names having resulted from planned naming and become an integrated part of the society's cultural heritage. As geographical names coined in the past represent a part of a society's cultural heritage, the act of giving new geographical names adds to the future cultural heritage. This perspective should be considered by the authorities as well as by private bodies that play a part in name giving processes.[10])

The naming of streets, roads, squares, buildings, etc. is an important administrative activity for local authorities all over the world. The giving of names takes place either through the acceptance of existing names in the area or through the creation of new names. In recent times, the use of various categories of names has been a common occurrence in street naming, for example, names of birds in one area, names of animals in a second, and names of artists in a third.

Traditionally the naming of geographical entities has mainly followed one of two paths: "spontaneous or popular naming" and "baptism". The majority of geographical names that are used today (in Europe at least) are the product of spontaneous naming, the result of a process of lexicalization [the process of making a word to express a concept] with its starting point in a description of the object being named. Normally, a geographical name evolves from a descriptive expression (for example, the hill or the green hill) into a proper name (for example, *The Hill* or *Green Hi*ll). In many cases, the naming will be governed by analogy, that is to say, new names will be produced on the pattern of other names in the area or of other names with which the name-givers are familiar. This is one of the reasons why certain types of names have a tendency to occur with greater frequency than others in particular areas[11]) or in particular periods.

Another way of giving geographical names is to transfer existing names to new entities in order to compare them with or to commemorate the original place, through the process known as baptism. This has in particular been the case in earlier European colonies. For instance, Harlem in New York is named after the Dutch town of Haarlem. Also, people are commemorated in many geographical names, for instance, George Washington in Washington.

Place names such as *Washington* could be considered to be among those that have resulted from a more intentional baptism through which towns, streets, roads, squares and other entities are given prescribed names through special resolutions. It is true that some of the names that are given formalized status in this way have earlier been used informally for all or part of the same named entity. Already in late medieval times, fortresses and towns, like the Scandinavian *Akershus* and *Landskrona*, were named in accordance with already existing naming patterns.

22.6 Geographical Names Studies and Cultural Values

A detailed picture

One can also regard geographical names as a reflection of the interplay between humans and nature through different periods of time. The many regional names studies that have been carried out in the world show that geographical names provide a multifaceted and detailed picture of human existence in relation to our surroundings. In some cases, the written geographical names may tell us in a general way about the time spans of settlement. For instance, Dirk Blok, in referring to Dutch-speaking parts of Europe, has indicated that: names ending in *-burg* or *-ingen* were recorded in the period from the seventh to the ninth century; names ending in *-donk* were recorded after the tenth century; names ending in *-kerke* were recorded in the period from the twelfth to the thirteenth century; and names ending in *-dam* were recorded after the thirteenth century.[12])

Linguistic heritage

In addition to being repositories of natural and cultural history, names represent an important part of our linguistic heritage. This understanding of geographical names as a historical and linguistic source has long been focused on within the community of names scholars and has in the course of time opened up the way for a broader commitment to the protection and preservation of names, as attested in the legislation of several countries. In this perspective, we also note the increasing concern with geographical names in multilingual areas,[13]) and recall several resolutions adopted by the United Nations Conferences on the Standardization of Geographical Names that have supported the collection and use of minority group and indigenous geographical names, for example, Conference resolutions V/22 (1987) and VIII/1 (2002).

Legislation

When place-names are used in a private context they are often used with varying spellings, and this is by no means forbidden. However, place-names belong to the whole society and that entails that they should be standardized in accordance with officially adopted regulations. These names are linguistic terms handed down from generation to generation mostly without identified name givers. Consequently, they should be regarded as common cultural property. This is also in line with United Nations resolutions.[14]) In many cases, however, individuals like to see their own private spellings of place-names on maps and road signs instead of the official spellings, and bitter conflicts sometimes emerge from such disagreements. This applies, not least, to Norway.

A number of countries have passed legislation applying to the protection of geographical names. In 1990 the Norwegian parliament passed an act on the spelling of place-names to safeguard the place-names as a part of the cultural heritage. The main regulation of the act implies that the geographical names should be standardized on the basis of the inherited pronunciation at the same time being in compliance with the spelling principles of Norwegian. However, numerous farm owners protested against the spelling regulations and in 2009 a small group of MPs forwarded a proposal to give the owners of smallholdings the right to decide on the spelling of the name of their smallholdings (which most often are identical with the names of the main farms), whereas the name of the main farm should be kept in its standardized spelling. In 2015 an amendment to the place-name law was adopted by the Parliament, thus legally permitting a divide in the spelling of place name heritage. For example, a family with the surname *Krogvold*, a Danish-inspired distorted form of the standardized spelling *Kråkvål*. *Krogvold* looks like a compound of Danish *krog* ‘hook’ and *vold* ‘meadow’, whereas the original compound *Kråkvål* consists of *kråke* ‘crow’ and *vål* ‘burnt clearing’. By using *Krogvold* instead of *Kråkvål* the message of the name gets lost.

A new Law (N.71(I)/2013), was recently voted on by the Parliament of the Republic of Cyprus, and an amendment was added to the basic law, by creating a new article, which criminalizes the alteration of geographical names and toponyms and the illegal issue, import, circulation, supply, distribution and trade of maps, books or other documents, in conventional or digital form, which contain geographical names and toponyms of the Republic, spelled differently from those specified in accordance with the procedures laid down in the Law, or those contained in the official Toponymic Gazetteer. This was an additional measure taken by the Republic of Cyprus for safeguarding its invaluable cultural heritage.

In Sweden, a new paragraph on good place-names practice was added to the Historic Environment Act in 2000. In the Act, the emphasis is placed on linguistic correctness, but even stronger on the importance of preserving geographical names as a part of the nation's cultural heritage. The paragraph also recommends the use of minority names on maps and signs along with the Swedish names.[15])

22.7 Collection of geographical names

Collecting geographical names is discussed in a separate chapter of the handbook, but it should be mentioned in this chapter in relation to the topic 'cultural heritage'.

It is self-evident that all countries and regions should record and preserve their geographical names, both for their practical use and for the preservation of names as cultural and linguistic heritage. Since around 1900, and even more in the second half of the 20th century an important aim has been to collect and research geographical names for mapping purposes, and also in general either by categories for reference purposes or by in-depth studies in selected areas. In this work, emphasis has been placed on registering the pronunciation that has been handed down by oral tradition, since it best reflects the linguistic basis of names.

At the same time, although the scientific archives have built up systematic collections of older written forms, it is the collection in the field of oral material that has been, and still is, the most important means of preserving, as far as possible, the wealth of names still existing in older tradition. In recent times, the preservation of geographical names from both oral and written sources has been acknowledged politically in many countries. For the general public, it is usually not the historical value of names that is of most concern but rather the unilateral changes made by administrators, without adequate local consultation to names currently in use. Many people feel this to be a form of encroachment on their well-being. That names uphold the social and cultural values of the individual is an important argument for promoting the systematic collection of names.

22.8 Exonyms

One special type of geographical names seen as part of the cultural heritage is that referred to as exonyms. The United Nations Group of Experts on Geographical Names *Glossary of Terms for the Standardization of Geographical Names* defines an exonym as a name used in a specific language for a geographical feature situated outside the area where that language is widely spoken, and differing in its form from the name used in the official language or languages of the area where the geographical feature is situated (*Glossary*, 081). Better-known examples include: *Vienne* (French) and *Vienna* (English) for *Wien* in Austria; *Ginebra* (Spanish) for *Genève* in Switzerland; and *Vilna* (Finnish) for *Vilnius* in Lithuania. Less generally known examples may form part of the vocabulary of smaller language groups (for instance, in Faroese: *Høvdastaður* for *Cape Town*; *Skiloy* for *Sicilia*; *Kili* for *Chile*; and *Simbabvi* for *Zimbabwe*).

All language communities have such names, and in the individual language they are part of the linguistic heritage of those communities. This has been recognized by the United Nations Conferences on the Standardization of Geographical Names (for example, in Conference resolution II/28, 1972). At the same time, however, several United Nations Conference resolutions have recommended that the use of exonyms be reduced (for example, II/29 and IV/20), particularly in cartography, in order to facilitate international communication through the use of nationally standardized geographical names.

In a United Nations context, the question of exonyms is ambiguous. On the one hand, the United Nations Conference on Geographical Names has passed several resolutions pertaining to the use of endonyms; on the other hand, the experts have maintained the value of exonyms as part of a nation's cultural heritage. Even if UN recommends the use of endonyms, it is no contradiction to consider exonyms as part of the cultural heritage. So, one is supposed to collect them, but not use them in maps for general reference (?).

22.9 Concluding remarks

In discussing geographical names as part of the cultural heritage, one must take into account the fact that geographical names in a society are part of the language. Geographical names are functional to the extent that they have a place in a society's communication system as reference tools; when they no longer fulfill a need, they fall out of use. As is apparent from this presentation, geographical names preservation is an important issue of concern to both laymen and experts.

Traditionally, problems in connection with the spelling and writing of geographical names have been the central focus, and indeed they still are; but in recent decades, the issue of protecting geographical names as part of the linguistic cultural heritage has come more to the forefront. Increasing globalization offers the advantage of allowing a wider understanding of local names and heritage. However, in some cases today, local place names appear at risk from the imposition, through both local and global commercialization, of names introduced from outside cultures.

22.10 References

[1]) https://en.wikipedia.org/wiki/Cultural_heritage (5. January 2016)
[2])http://www.wipo.int/edocs/mdocs/tk/en/wipo_iptk_ge_2_15/wipo_iptk_ge_2_15_presentation_mounir_bouchenaki.pdf
[3]) http://www.unesco.org/culture/ich/en/states-parties-00024
[4])https://en.wikipedia.org/wiki/Convention_for_the_Safeguarding_of_the_Intangible_Cultural_Heritage
[5]) Geographical Names as Cultural Heritage. Proceedings of the International Symposium on Toponymy, Seoul, 7–9 November 2014. Ed. by Sungjae Choo. Kyung Hee University Press. Seoul 2015.
[6]) *Geographical Names as Part of the Cultural Heritage* (2009) ed. by P. Jordan, H. Bergmann, C. Cheetham, & I. Hausner. Vienna: Institut für Geographie und Regionalforschung der Universität Wien).
[7]) Naftali Kadmon, Toponymy. The Lore, Laws and Language of Geographical Names (New York, Vantage Press, 2001), p. 47
[8]) "Ortnamn är en viktig del av vår geografiska och kulturella miljö. De identifierar lokaliteter av olika slag och representerar oersättliga kulturella värden med avgörande betydelse för människors hemkänsla och trivsel." (Zilliacus 1978, 211)
[9]) "Place Names – The Memory of Places. The theme of the European Heritage Days in Finland 1999". UNGEGN, 20th Session, New York, 17 – 28 January 2000, Working Paper No. 22, submitted by Sirkka Paikkala, Finland. See also Sirkka Paikkala: "Place names in Finland as cultural heritage" in *Onoma* 35, 2000, pp. 145-164.
[10]) For examples of this type in a South African city, see Colin Clive Smith, "Naming a city: the street and place names of the Roodepoort area 1854-1999", *Nomina* Africana*,* vol. 16 (2002), pp. 56-64.
[11]) See, for instance Wilhelm F.H. Nicolaisen, "Lexical and onomastic fields", in *Proceedings of the Thirteenth International Congress of Onomastic Sciences, Cracow, August 21-25, 1978*, by Kazimierz Rymut, ed. (Wroclaw, Poland, 1982) vol. II, pp. 209-216.
[12]) For an English-language reference, see Dirk P. Blok, "Names in *-iacum*", in *Topothesia: A Collection of Essays in Honour of T.S. Ó Máille*, Breandán S. Mac Aodha, ed. (Galway, Ireland, 1982), pp.111-114.
[13]) See F.J. Ormeling, *Minority Toponyms on Maps. The Rendering of Linguistic Minority Toponyms on Topographic Maps of Western Europe* (Utrecht, Netherlands, Department of Geography, University of Utrecht, 1983).
[14]) http://unstats.un.org/unsd/geoinfo/UNGEGN/docs/RES_UN_E%20updated_1-10%20CONF.pdf, Chapter 3. [15 August 2014.]
[15]) See Annette C. Torensjö in Geographical Names as Cultural Heritage. Proceedings of the International Symposium on Toponymy, Seoul, 7–9 November 2014. Ed. by Sungjae Choo. Kyung Hee University Press. Seoul 2015, p. 70 ff.

Section 9 Cultural aspects

Chapter 23 **Toponyms as carriers of cultural heritage: the history of place names**

Tjeerd Tichelaar

23.1 Localized history contained in place names

A special and extremely valuable aspect of geographical names is the localized history they contain. Unless a toponym is deliberately replaced, it generally changes in appearance through time because of evolutionary and revolutionary changes of language, writing and pronunciation, but essentially it stays the same. The meaning a name carries when it is first applied may seem lost at first sight, but somewhere inside its iconic image and sound, in a fossilized form maybe and only extractable by the trained eye, it remains there to tell its own unique story to anyone who can see through the scars of time. Al-Iskandariyyah (Alexandria) in Egypt, Kandahar in Afghanistan, İskenderun in Turkey are surviving witnesses of local achievements of Alexander the Great, the names of Silifke and Antakya (Antioch) and several Ladiks in Turkey, as well as al-Lādhiqīyah (Latakia) in Syria, commemorate members of the Macedonian dynasties founded by his generals that held sway right there. Oxford and Cambridge tell a story of river crossings in medieval times, important enough to make settlements grow into towns. Town names ending in –chester or –caster in England echo the presence of Roman legions initially responsible for the growth of these towns.

Older toponyms unveil images of local circumstances that no longer exist: natural environments that disappeared, plants and animals no longer around, economic, social and political systems for long replaced; they allow us to trace migrations and to reconstruct boundaries of which no material traces remain. Exactly because they are tied to fixed spots in space, the historical information geographical names contain is a priceless supplement to documented history and archaeological evidence.

23.2 Ancient toponyms: a special class of cultural heritage

Although geographical names are subject to deliberate, often politically motivated change and replacement and are newly devised whenever a new geographical object needs to be named, many of them subsist many centuries. They prove resistant to events like wars and natural disasters all but obliterating the objects they originally describe. Like ghosts of the deceased, they sometimes keep hanging around the locations where they were once brought to life, and not uncommonly find a new body in which to dwell. Some toponyms constitute a category of cultural heritage of nearly unmatched ancientness.

Geographical names sometimes embody a population's historical roots, or even a location's history preceding the presence of the ancestors of those living there now. They are valued by historians, historical geographers and linguists for the light they may shed on a past otherwise dark, for the window they may provide to look back beyond the beginnings of history.

Apart from actual, verified historical details, legends and popular anecdotes often at some time in the past got attached to them. Even if proven untrue, these played or still play a role in the connection people perceive with locations in geographical space.

23.3 Time-resistant ancient names

Most commonly, ancient names over the centuries change in appearance following language change when control of the land and everything on it, including the named objects, passes from one ethnolinguistic group ('people') to another. Names also tend to change beyond recognition as a consequence of the evolution of phonetic, grammatical and syntactic details of languages and their writing conventions. In some cases, however, names seem to have survived all what happened to their environment practically unchanged. The map below shows some of the many hundreds of place names in Italy dating back more than 2000 years: the coloured names printed in roman type are currently still (or again) written in exactly the same way as in ancient times.

What strikes us when looking at the map in figure 23-1 is the impressive diversity of cultures that left their traces in names still in use today, most of them now forgotten and, to the eyes of the unknowing, for long disappeared. Also, the distribution of these hidden heirlooms still gives us a clue of the geographical range of the cultures leaving us their names: Etruscans in the central peninsula, Illyrians and Venetes in the east, Celts in the north, Greeks and Phoenicians in the south, Romans and other Italic peoples nationwide: they all contributed to what we call Italy and Italian today. What also may surprise us, is that exactly among the surviving ancient names in Italy originally Latin names seem to be a minority.

Staying in Italy, the persistence of geographical names may be demonstrated by some examples of toponyms surviving the demise of the objects for which they were designed. The Etruscan city of Caisra, one of the biggest and most important places in Italy by the 6th century BCE, survived as a small provincial town into Roman times under the Latinised name Caere, but started to become abandoned as it fell victim to outbreaks of

Figure 23-1 Some place-names in Italy survived the passing of 20 centuries and multiple substitutions of language largely unchanged.

malaria and Saracen raids after the 6th century CE. By then its name had been transferred to the local bishopric, for which a new see was built 9 kilometres to the east. The new settlement was named Caere Nova ('New Caere'), which now resounds in the name of the village of Ceri. By the 13th century the old city had become a ghost town known as Caere Vetus ('Old Caere'), a name it retained when it became resettled in the 17th century: in modern Italian, this became Cerveteri.

Another notable example is the story of the city of Capua. This ancient place, its name also revealing an Etruscan past, was in the 3rd century BCE the second-largest population centre in Roman Italy. In 851 CE, the city was burned to the ground by Saracen mercenaries sent by the Lombard usurper of the principality of Benevento to which it then belonged, after which a new city was built at the remains of the old Roman town of Casilinum, five kilometres down the Via Appia. The name Capua was consequently transferred to the new site, the location's former name going into oblivion. After a couple of centuries, the ruins of the old town came back to life, when some farmers settled around the old church of Santa Maria Maggiore, the only structure that had survived the Saracen attack. To honour the ancient site, in 1861 the town was renamed Santa Maria Maggiore Capua Vetere ('Old Capua').

Some similarities between the ancient and modern names in Italy are the result of recent renaming as a purposeful policy to boost national pride by reminiscing the glorious Roman past. In Sicily, the Greek town of Akragas had received the Latinised name Agrigentum when it was seized by the Romans in 210 BCE. At the time Arabic was the dominant language in Sicily, roughly from the 9th to the 11th century CE, the place was referred to as Kerkent, a name that was Sicilianized to Girgenti after the Arabic dominance passed. Girgenti was replaced with Agrigento, the modern Italian version of the Roman name, in 1927. Other examples of ancient

Latin names that were recently revived are Tuscania (Toscanella until 1911), Tarquinia (Corneto until 1922) and Aprilia (a new town laid out in 1936).

23.4 The transfer of names from language to language

At the moment a name is given to an object, the language of the name-giver provides both the elements needed and the structure to join them together. The elements consist of semantic and morphologic units – units of meaning and form - called *words* and *morphemes*. The former are the smallest units that may occur independently, the latter even smaller particles, like suffixes and affixes forming part of or joined to them. The structure is provided in the form of a set of rules called *grammar*, that defines the way the language can be used to convey (*communicate*) meaning. An important constituent of grammar is the *syntax*, determining the way words should be linked together into larger semantic conglomerates.
Most names start their existence as such a semantic conglomerate.

The linguistic abracadabra above may easily be clarified by picking the first name that comes to mind. For instance: Stratford-upon-Avon. This English town that became world famous as the birthplace of William Shakespeare clearly consists of three elements, which are, obviously in accordance with some syntactic rule specifically applying to English names, separated by hyphens. Two of the elements start with a capital, the one in the middle does not: again a syntactic rule. As a capital initial letter is commonly used in (Roman) written language to denote that a word is either the beginning of a sentence or a name, we get the idea that both 'Stratford' and 'Avon' are names in themselves, and 'upon' is not. We need to know that 'upon' is a preposition, meant to establish a situational link between 'Stratford' and 'Avon'. Both of the remaining elements of this name also enclose a meaning for themselves, which at the time of the name-giving must have been considered important: this meaning had to ensure that upon mentioning it would make clear which geographical object was meant, without anyone needing to point at it.

'Stratford' appears to be an Anglo-Saxon (Old English) name, consisting again of two semantic units, namely *strat* and *ford*. 'Strat' is adopted from the Latin word *strata*, meaning 'paved road'. It was a Latin (Roman) name for something the Romans made and left behind for the Anglo-Saxons, who did not know it themselves and, as they seem to have done with other Roman things, ignored it except when they could make use of it. The paved road referred to was in this case the Roman road from Alcester (ancient Alauna) to Tiddington, both of them Celtic settlements fortified by the Romans. 'Ford' is an Anglo-Saxon word that still exists in modern English, meaning 'part of a river shallow enough for people to cross it'. So 'Stratford' was obviously the place where one would cross the river when following the Roman road. If one would mention this, anyone would know which site was meant without someone having to go there and point at it.
The addition 'upon Avon' became necessary when the place became important enough to be mentioned to people who might also know other places where Roman roads crossed rivers, or other towns named 'Stratford'. To these people the mentioning of 'Stratford' alone might not provide enough information. The name 'Avon' itself is Celtic; it simply means 'river'; it is still the name of a couple of English rivers, including the one flowing through Stratford-upon-Avon.

It is thus clear that the name of this town really started as a semantic conglomerate, even though today its meaning to most people is simply 'Shakespeare's birthplace'. The Roman road became forgotten, its asphalt successor being not special enough to be mentioned, and the ford lost its importance once bridges were built. The addition 'upon Avon' remained worth mentioning because of the existence of another Stratford, located in the Greater London conurbation.

23.5 The relevance of dead and disappeared languages

Because toponyms generally (although not always) outlive their creators, locally vanished and even 'dead' languages are not per definition deprived of their importance from a toponymic point of view. Dead languages often leave their traces both in the vocabulary of their living successors and, much more so, in geographical names. This is a well-known fact to historical linguists, who make indeed grateful use of toponyms in their efforts to reconstruct so-called proto-languages (disappeared common ancestors of modern languages belonging to the same family), as well as trace *substrates*, residues of local predecessor languages in unrelated immigrant successor languages. Especially hydronyms (water names) have a reputation of being very ancient, often ancient enough even to antedate the presumed 4th and 3rd Millennium BCE Indo-European immigration into Europe. These substrates are held responsible for a major part of the diversification between the branches of the Indo-European language family. The vocabulary of the Germanic languages, for instance, is thought to contain a large number of pre-Indo-European words, maybe inherited from the thriving 4th Millennium BCE society that built the numerous tumuli and megalithic graves in north-western Europe. Also, the Greek geographical generic term meaning 'sea', *thalassos*, is supposedly of pre-Hellenic and pre-Indo-European (by the ancient Greeks themselves labelled 'Pelasgian') descent – suggesting that this famous seafaring people was not yet so familiar with the sea at the time it reached its present homeland.

A quick survey of the geographical names in a well-known country like the United Kingdom will further illustrate the arguments expounded above.

The nationwide official language of the United Kingdom is English. Besides English, the dwindling Celtic languages Welsh and (Scottish) Gaelic also have official status on a sub-national level.
English is a Germanic language, which developed from the closely related languages of Anglian and Saxon immigrants in the 5th century CE. In the part of the kingdom currently called England, Anglo-Saxon and neighbouring continental Germanic invaders, earlier than their Germanic language(s), presumably superseded a mixed Roman and Brythonic Celtic aristocracy ruling a partly Romanised, but largely still Celtic (Brythonic) speaking population. The part of the Brythonic population most strongly opposing assimilation with the Anglo-Saxon language and culture fled the Germanic invaders to take refuge in present-day Wales, the border area of England and Scotland (Cumbria and Strathclyde), the southwestern peninsula of England, and the peninsula of Armorica (thereupon renamed 'Britannia minor', Little Brittany) in continental Gaul (currently France). In Scotland, at the same time, a Pictish population speaking an as yet unknown language that had taken refuge there for the Roman invaders of the island, four centuries earlier, were gradually superseded by so-called Goidelic Celts (Gaels, *Scoti*) invading their homeland by sea from Northern Ireland. The Brythonic and Gaelic newcomers in Scotland were, although both Celtic, distinctive enough not to understand each other's language.

Starting from the 8th century, new Germanic immigrants invaded the country: Norwegian and Danish Vikings took possession of and effectively colonized large parts of both Scotland and England, to be eventually (in the 11th century) expelled again by the Anglo-Saxons. Even before the last Norwegians were ousted, however, Anglo-Saxon dominance itself came to an end by an invasion of yet another Viking aristocracy: this time already Romanised (French-speaking) Normans, from what had recently come to be known as Normandy successfully claiming the English throne.

At present the English language is, apart from being the only nationwide official language, the mother tongue of more than 99% of the native inhabitants of the United Kingdom. But before Anglo-Saxon or English prevailed, Pictish, Brythonic Celtic, and Latin were for centuries the languages of both aristocracy and (part of) the common people, as were Gaelic Celtic, Norwegian, Danish and French (the latter mostly of aristocracy) after the introduction of Anglo-Saxon. The imprint of some of these languages on the geographical names of the British Isles is at least as large as Anglo-Saxon/English: the large majority of names in Scotland is of Gaelic origin, except in Strathclyde, where many names are either Brythonic or Anglo-Saxon, and in the Northern and Western isles (Shetland, Orkney and the Hebrides), where almost all names are of Norwegian descent. The islands were Norwegian from the 8th until the 15th century, which was long enough for a new variety of the Old Norse language to develop there (Norn, spoken in Orkney until the 18th century).
The northern and eastern parts of England show a mixture of Danish - for instance names on *-by* (= 'farmstead, village') - and Anglo-Saxon, while the southeast is predominantly Anglo-Saxon. In the southwest the Brythonic element is dominant. All through England a Brythonic substrate is eminent, although least so in settlement names, as are remains of Latin like the formerly generic elements *caster* or *chester* (Lancaster, Manchester - from *castra* = 'fortress') and -port (from *portus* = 'harbour' or *porta* = 'gate'). Wales is almost completely Brythonic: the Anglicised forms of Brythonic (Welsh) names were with the recent emancipation of the Welsh language returned to their original state, and English names reverted to their Welsh counterparts. In Cornwall in Southwest-England, the Cornish (Brythonic) language, actually extinct (a 'dead language') since 1777, is presently being revived and granted official status next to English on a local level: some Cornish place-names are correspondingly being restored.

The English language itself lost much of its original Anglo-Saxon character because of all subsequent invasions, causing the grammatical structure to be simplified and the vocabulary augmented with a large amount of Scandinavian and French words. Pronunciation was heavily influenced as well.
Geographical names were in writing often adapted to the language passing by. An appealing example is York, going back on a Brythonic personal name Eburos (meaning 'yew man'), maybe the owner of an estate with yew trees where the Romans built a fortress they consequently called Eburacum. The Anglo-Saxons, ignorant of this meaning, transformed the name through etymological misinterpretation ('folk etymology') into Eoforwic, meaning 'wild boar settlement'. The Vikings taking over the place from the Anglo-Saxons contracted the first part of the name, without bothering for a possible meaning that they didn't understand anyway, into 'Hjor', while they thought to understand the second part as the similar sounding Norse generic 'vík', meaning 'bay': not very appropriate for the inland town, but the combination Hjor and vík occurred in their Scandinavian homeland as well. They were the last to bother at all: the Anglo-Saxons ousting the Danes, just before they themselves had to accept francophone Norman rule, left the name as it was remodelled by the Vikings: Hjorvík. The Anglo-Saxon tongue would ultimately erode this into what it is now: York.

The process of subsequent transformations of names illustrated by the case of York shows the significance of 'dead' as much as 'living' languages to the development of geographical names. The 'erosion' ultimately yielding the present form of the name does not follow a random path, but is dependent on the phonological characteristics of the 'new' language (the Anglo-Saxon dialect of Yorkshire) as compared to those of the 'old' language (medieval Norse or Danish). The regional settlement history, as it also culminates in the local dialect, is decisive. Latin *castra* thus used to evolve into *caster* in the areas of Northern England staying for long out of the grip of the Anglo-Saxons, but tended to become *chester* or *cester* in the more thoroughly anglicised parts of the country. It is thus the phonology of the dialect, not the official language, which determines the ultimate form of the name.

23.6 Vanished identities preserved in names

Geographical names more often than not antedate the national entities we discern today. In Europe, for instance, most of the state-based nations Europeans currently derive their identity from (British, French, Germans, Italians, Poles, Dutch, Danes and so forth), crystallized between the 17th and the early 20th centuries, with processes of nation-building peaking in the latter half of the 19th century. Dominant state languages became national, official and standardized, while vernacular 'between the lines' became suppressed. In the geographical names populating the redrawn map, the actual diversity as of old nevertheless persevered. The wealth of regional culture and history fossilized in the names, so to say. A nice example can be observed in France (see figure 23-2).

At the time of Caesar's campaign in Gaul just over 2000 years ago, Gaulish society was divided over numerous Celtic, Celto-Ligurian and Aquitanian tribes and alliances, some of which were more powerful than others but none had succeeded in lastingly unifying a significant part of Gaul. Under Roman rule too their dividedness remained a characteristic feature of the Gauls. The tribes of old became *civitates*, more or less autonomous communities bound together by common laws and each living within their own territories delimited by the Roman provincial authorities and governed from their own capital. The names of the capitals became typically connected to the names of the *civitates*, which were the same as the names of the former tribes, especially when they were common Gaulish names of which many homonyms occurs, like Noviomagus ('new market'), Noviodunum ('new borough') or Mediolanum ('centre of the plain/territory'): in this case they functioned as epithets to discriminate between homonyms, for instance Mediolanum Eburovicum ('Mediolanum of the Eburovices', now Évreux; Mediolanum Santonum ('Mediolanum of the Santones', now Saintes). By the 4th century CE the tribal epithet (Eburovicum, Santonum) had in many cases replaced the original name as the common appellative by which the place was known. The result of this was that both the name of the *civitas* – as the originally 'ethnic' administrative unit typically became a diocese after the Christianization, its name usually lived on after the Franks finally brought unity to Gaul in early medieval times – and its capital became derived from the name of the tribe, but three or four centuries intervened so the derivations followed different roads. The name of the Atrebates thus lived on in the regional name Artois and the city name Arras; the Baiocassi in Bessin and Bayeux; the Cadurci in Quercy and Cahors; and so forth. The map shows the tribal names in their Roman Latinised form (red italic), the derived regional names (black italic) and the city names; note that the word *city* (French *cité*) itself is derived from Latin *civitas*.

Below the map in figure 23-2, the derivations are listed in a table. In this table, the current French demonyms (word indicating the inhabitants of a city or region; mostly they are geographical adjectives as well) are included in the last column. In some cases, these retain the original form of the Latin tribe names: inhabitants of Bayeux are still called Baiocasses, those of Corseul Curiosolites, those of Langres Lingones.

23.7 Of people, activities and circumstances gone by

A nation's collection of geographical names is like a geographical information system storing all that passed by in the different parts of its national territory. A wealth of geo-referenced bits of heritage that collectively recollect how nation and land became what they are today.

The oldest geographical names currently still in use echo the sounds of languages sometimes long forgotten, thus proving that the people speaking these languages were actually present at the named locality at some point of time. The city of Aswan in Egypt was represented by hieroglyphs with the phonetic value swnt.t (Swenett), logographically thought to designate 'trade' or 'market' in Early Egyptian of the Old Kingdom of the early 3rd millennium BCE, where it functioned as a border settlement trading with Nubia. The hieroglyphs representing the name of present-day Asyut read z3jw.tj, its meaning resting unveiled. Many European river names, as well as names of mountain ranges and some settlements, are on the grounds of linguistic comparison (they don't match Indo-European phonological schemes) assumed to be at least of similar age, although no written documentation was handed down. Vestiges of unknown ancestors are they, of cultures we have no memory of. Cultures that did leave traces of textual evidence and material artefacts impressive enough for us to nickname them 'cradle of civilization', also left us toponyms the sounds of which can still be heard today, even if the languages they belonged to went out of use thousands of years ago. Turkish Adana was Adaniyya, Tarsus Tarsa in the language of the Hittites; Iranian Shush was Ŝuŝan in Elamite, Ilam was the Elam of the Sumerians; Syrian Dimashq (Damascus) was Dimašqa in the Akkadian tongue, Turkish Nusaybin Naşibīna. Even when original meanings cannot be retrieved, identifying

the language alone suffices to tell us something that stirs the imagination.
The information contained in names originating in better known languages presents us with a complete history of a region's settlement and cultivation, economic development and political fortunes, wars and migrations and natural environment.

Figure 23-2: Names of French towns and regions derived from former Gaulish (Celtic, Celto-Ligurian and Aquitanian/Iberian) tribes.

Table 23-1 Comparison of the names of Gaulish tribes, derived regional names and city names and region/city demonyms

Gaulish tribe	**Derived regional name**	**Derived city name**	**Demonym of modern region/city name**
Atrebates	Artois < Atrebatensis	Arras < Athrebate < (Nemetacum) Atrebatum	Artésiens (region); Arrageois (city)
Viromandui	Vermandois < (pagus) Viromandensis	Vermand	Vermandois
Ambiani		Amiens	Amiénois
Caleti	Caux		Cauchois
Bellovaci	Beauvaisis < Bellovaci	Beauvais < Bellovacis	Beauvaisien
Suessiones	Soissonnais	Soissons < (Augusta) Suessionum	Soissonnais
Remi		Reims < Remos < (Durocortorum) Rementium	Rémois
Mediomatrici		Metz < Divodurum Mediomatricorum	Messin
Catalauni		Châlons < (Duro) Catalaunum	Châlonnais
Meldi		Meaux < Meldis < Iantinum civitas Meldorum	Meldois
Silvanectes		Senlis < Silvanectis < Augustomagus Silvanectum	Senlisiens
Veliocassi	Vexin < (pagus) Veliocassinus		
Lexovii	Lieuvin	Lisieux < Noviomagus Lexoviorum	Lexoviens
Eburovici		Évreux < (Mediolanum) Eburovicorum	Ébroïciens
Durocassi	Drouais < Dorcassinus	Dreux < Drocas < Durocas < Durocassis	Drouais
Viducassi		Vieux < civitas Viducassensis	Viducasses
Baiocassi	Bessin < Pagus Baiocensis	Bayeux < civitas Baiocassium	Bessinois (region); Bajocasses (city)
Parisii	Parisis	Paris < Lutetia apud Parisios	Parisiens
Carnutes		Chartres	Chartrains
Tricassi		Troyes < civitas Tricassium	Troyens
Senones	Sénonais	Sens	Sénonais
Cenomani	Maine < (pagus) Cilimanicus	Le Mans < Ceromannos < civitas Cenomanorum	Manceaux
Diablinti		Jublains < civitas Diablintum	Diablintes
Abrincatui	Avranchin	Avranches < Abrincas	Avranchinais
Curiosolitae		Corseul	Curiosolites
Redones		Rennes < Redonas	Rennais
Veneti		Vannes < civitas Venetum	Vannetais
Andes/Andecavi	Anjou < Andecavum	Angers <	Angevins
Namnetes		Nantes < Portus Namnetum	Nantais
Turones	Touraine < civitas Turonensis	Tours	Tourangeaux
Pictavi	Poitou	Poitiers < Pictavis	Poitevins
Lingones		Langres	Lingons

Gaulish tribe	Derived regional name	Derived city name	Demonym of modern region/city name
Bituriges	Berry	Bourges	Berrichons (region); Berruyers (city)
Ambarri		Ambérieu < Ambariacus	Ambarrois
Arverni		Auvergne < Arvernia	Auvergnats
Santones	Saintonge	Saintes < (Mediolanum) Santonum	Saintongeais (region); Saintais (city)
Lemovices	Limousin	Limoges < Lemovicas < civitas Lemovicum	Limousins (region); Limougeauds(city)
Medulli	Médoc < Medulicus		Médoquins/Méduliens
Petrocorii	Périgord	Périgueux	Périgordins (region); Périgourdins (city)
Boii	Buch	Biganos < Boios	
Vasates	Bazadais < Vasadés < (pagus) Basatensis	Bazas < Vasats [Gascon] < civitas Basatica	Bazadais / Vasadés [Occitan]
Cadurci	Quercy < Carcin [Occitan] < Cadurcinum	Cahors < Cadurcum < Divona Cadurcorum	Cadurciens
Vellavi	Velay		Vellaves
Sotiates		Sos	Sotiates
Gabali	Gévaudan	Javols < ad Gabalum	Gévaudanais (region); Javolais (city)
Ruteni	Rouergue < (pagus) Rutenicus	Rodez < Ruteni < civitas Rutenorum	Rouergat (region); Ruthénois (city)
Elusates	Eauzan	Eauze < Elusa	Élusates
Tarusates/Aturenses	Tursan < pagus Aturensis	Aire-sur-l'Adour < Atura	Aturins
Auscii		Auch	Auscitains
Beneharni	Béarn		Béarnais
Suburates	Zuberoa [Basque, = Soule]		Zuberotar [Basque]
Bigerri	Bigorre		Bigourdans
Consoranni	Couserans < Consoranni		Couserannais
Caturiges		Chorges < Cadorgas	Caturiges
Quariates	Queyras		
Brigiani		Briançon < Brigantio	Briançonnais
Vergunni		Vergons	Vergonais
Reii		Riez < Reii (Apollinares)	Réiens
Sentii		Senez < civitas	Seneziens
Vordenses		Gordes	Gordiens
Tricastini		St-Paul-Trois-Châteaux < St Pau de Tricastin [Occitan] < Noviomagus Tricastinorum	Tricastins

Chapter 25 **Place Names and Cultural Heritage in an Archipelagic Country**

Multamia RMT Lauder and Allan F Lauder

25.1 Introduction

This chapter looks at the issue of the relationship between toponymy and cultural heritage in the context of an Asian archipelagic nation. The chapter is divided into two parts.

Part one describes key features of cultural heritage and the links with toponymical studies and also provides geographical, historical, linguistic and cultural information on insular Southeast Asia with a focus on Indonesia. There we find a high level of diversity in terms of language, culture and ideology. This has affected the development and use of toponyms in the country.

Part two gives a description of a number of place names. These examples represent place names for natural features, man-made features and administrative features. They all can be seen to possess cultural heritage status in some way. Different types of data are needed to properly understand their meaning.

25.2 Background

Cultural Heritage and Toponymy

Interest in the relationship between cultural heritage and toponymy has grown in UNGEGN since its establishment in 1960. In 2002, a working group was established to address the promotion and use of indigenous, minority and regional language names. In 2012, this working group was changed to a working group for the issue of geographical names as cultural heritage (Kerfoot, 2015, Watt, 2015). Contributions to the discussion of culture in UNGEGN are to be found in various documents resulting from the ten UNGEGN Conferences from 1967 to 2012 and other publications including Kadmon (2000), UNGEGN (2006) and Helleland (2006). The field of toponymy has experienced a major transformation over the course of the last 20 years. A variety of new thematic concerns have been explored, and there is now a far greater recognition that toponymical research should be firmly grounded in an explicit engagement with critical theories of space, place, and landscape (Rose-Redwood et al., 2010: 458).

Culture covers a wide range of phenomena. In cultural anthropology two aspects of culture are distinguished. The first is ways of thinking, ideas and values. The second is behaviors, practices and skills (Eller, 2016: 21, Haviland et al., 2011: 27). The 'classic' qualities of culture are that it is learned, not innate; it is a set of norms shared by a community or group; it is a system we use to structure the world and give meaning to it; its parts function together as an integrated whole; it helps us to adapt to the world and thrive; it is in a state of constant change (Nanda and Warms, 2012). Culture can also be defined in terms of its mobility, its ability to move into new geographical or social spaces; is produced through human cognition and action in specific locations.

Place names are names that reflect language-users' perception of and relation with their environment. Place names are 'cultural artifacts' produced by complex interactions between mind, culture and environment (Taylor, 2016: 70). Most place names also have a historical dimension. They were coined at some time in the past and reflect the geographical, cultural, cognitive and attitudinal aspects of human life at a former point of time.

Cultural heritage is the cultural legacy of past generations. This can include tangible products of heritage such as built structures, tools and fabricated items. It can also include intangible heritage such as folklore, oral history, traditions, language, and indigenous knowledge. Natural heritage, meanwhile refers to works of nature with cultural value, with which humans interact meaningfully. Cultural heritage is a product of human action or cognition which is thought worthy of preservation for the benefit of others and of future generations.

The rationale behind the idea that cultural heritage has value and is worth preserving is based in the idea that previous generations have produced ideas, knowledge, traditions, artifacts and built structures that have value for us as cultural achievements. Heritage helps us to learn from the past by appreciating the development of civilization, of science, of technology and of literature and the arts. Heritage preservation is about remembering, understanding and admiring. Three reasons have been put forward for the link between geographical names and a nation's cultural heritage. First, geographical names provide information about natural and cultural conditions at the time they were coined and thus represent a memory of places, and this can act as historical documentation. Second, they are part of the local language and history. Third, they represent a link between communities and their environment (Helleland, 2006: 121).

The study of cultural heritage is highly multi-disciplinary. Dealing with tangible cultural heritage, such as the discovery and preservation of cultural artefacts, buildings and sites is the business of archaeology, history, heritage management, and heritage tourism. Intangible cultural heritage, such as oral traditions and endangered indigenous languages, can be studied by analyzing inscriptions or written historical sources. Natural heritage meanwhile is studied in geography, ecology, geo-diversity and eco-tourism. Cultural heritage also involves the study of law with respect to the establishment and enforcement of rights, and of conflict studies and criminology for the illegal trading and destruction of cultural heritage.

UNESCO has played an important role in the definition of what cultural heritage is and how it should be preserved. The 2003 UNESCO Convention on the Safeguarding of the Intangible Cultural Heritage (ICHC) has been seen as a significant departure from previous conventions as the intangible aspects of cultural heritage are given precedence over its physical manifestations such as movable cultural property, monuments and sites (Blake, 2009: 45-46). It has also offered some protection for the endangered languages of small indigenous communities and their oral traditions, arts, social practices and traditions. All of these are part of their intangible cultural heritage and relevant to toponymy (Marrie, 2009: 169-170).

25.3 History of Indonesia as an Archipelagic and Multicultural Nation

The peopling of the archipelago, the arrival of modern day humans, got underway 50,000 years ago. The descendants of these early people are today mostly found in the east of the country. However, the majority of today's population are descended from a group of people known as the Austronesians, whose homeland was in Taiwan around 5,000 years ago (Bellwood et al., 2006). They possessed seafaring skills that allowed them to move from island to island and they spread throughout the region. These peoples spoke an ancestral form of the many languages spoken today. As many as 719 languages have been attested in Indonesia and about two thirds of all of these share a common ancestral language and belong to the Austronesian language family, the remainder being part of the Papuan language family (Lauder and Lauder, 2015).

The history of East and Southeast Asia shows that people were travelling by sea for exploration, for trade and diplomacy as far back as the Greco-Roman period when spices and other goods were shipped as far as the present day Middle East. The different routes were part of the Sea Spice Route or Maritime Silk Road which existed in parallel with the Silk Road. The early Indonesians were engaged in trade and were part of the maritime spice trade routes that linked China, Southeast Asia, India, Africa, and the Middle East. Cosmopolitan settlements sprung up around the coastal areas of the South China Sea, Java Sea and Malacca Strait. These coastal ports were the hubs of the trade networks connecting islands and continents from as early as the first century. Over time, the region submitted to the unifying force of a series of kingdoms.

Among these kingdoms were the Hindu-Buddhist Srivijaya (600-1290 AD) which controlled the trading routes of the Malacca Strait and Java Sea and the Majapahit kingdom which was founded in Java and lasted from the late 13th century to around 1500 and territory extended from the southern Malay Peninsula, to Borneo, Sumatra and Bali.

The Indonesian archipelago was open to the inflows of new cultural influences. These were brought by the foreigners who stayed in coastal ports, and over time they were adopted, and spread. But they were also absorbed, nurtured and transformed through contacts with existing local cultures. This process, when seen over centuries, resulted in complex layers of different cultures. The result was often hybrid, recognizable but unique. In addition, it led to the blurring of cultural and geographic boundaries.

With the arrival of the Europeans in the 16th century, the existing shipping networks decreased in importance as the Portuguese and Dutch took over the trading routes, monopolizing the profits from the sale of spices and other products to Europe. As a result, the previously prosperous coastal towns become impoverished and the Indonesians lost their memory of being sea-going people (Tanudirjo, 2006). From the 17th until the 20th century the influence of the Dutch - first as trading company, later as colonial power, increased. During their hegemony, the upper echelons of society were schooled in Dutch and traces of Dutch can be found in place names. During WWII, the Japanese took over the country. The Japanese did not leave much of a trace linguistically. Today, global English dominates and is permeating many aspects of modern life, including place names.

Since its independence in 1945, the country has a language policy that promotes a single unifying language, the national language Indonesian (Bahasa Indonesia) but which also supports the maintenance and preservation of the regional indigenous languages. While the worldwide trend for indigenous languages is the existential threat from globalization and modernity, leading to loss of speakers of small languages, local languages in Indonesia still form part of the tapestry of diversity that exists alongside the national language and foreign languages. The language situation plays an important part in the unique situation for place names in the country.

25.4 Place Names

This section provides a number of examples of significant Indonesian place names. Examples for three types of geographical names are given: Names of natural features, built features and administrative areas. All of these can be said to represent Indonesian cultural heritage. In the case of Borobudur, this is also a heritage site recognized by UNESCO.

All of them have names that were coined at some point in the past and which may have experienced changes over time. In order to study such names in a nation with a complex culture that has been built in layers from influences from different cultures and languages, the methodology must first of all consider historical sources. The field of historical geography lies at the intersection of geography and history. It studies the way geographical phenomena have changed over time. A proper understanding of historical geography is best based on a review of earlier sources Starting point usually is the physical environment, as created by geology, botany and climate, influenced by social, cultural, and technological factors as well as by people's personalities and individual choices.

Their study requires careful consideration of historical, geographic, political, and cultural dimensions. It is necessary to consider the textual sources that may be available. In this case, they include texts in classical Javanese, classical Chinese, old Portuguese and old Dutch, as well as texts in a number of other foreign languages, including English and indigenous languages such as Sundanese, Balinese, and Malay. These need to be investigated in order to identify and then describe and explain the different toponyms which are the object of the study.

The methodology for a multidisciplinary study of toponymical cultural heritage based on various written sources might consider the following:

1. Identify references to old maps, manuscripts and stone inscriptions, history, folk tales, myths, in the search for new data;
2. Explore descriptions of historical activities in old manuscripts and stone inscriptions and use these to clarify and expand on existing descriptions;
3. Interpret and explain activities and affairs based on the textual and archaeological data, in particular attempting to throw light on the network and finding out how they bind together different geographical and political centers into a larger civilizational entity; and make connections between these findings and the meaning.

Written sources from the past include inscriptions on stone or other material, and manuscripts or codex. These may contain place names and descriptions of them, information on where they were located and on activities or events that occurred there in the past. Place names in manuscripts may refer to places that still exist or whose location still can be traced. They may also include places where the area has changed leaving little physical evidence of that place in the present, and thus provide clues for academic work to rediscover those places.

There are relatively few authoritative toponymical studies in Indonesia. However, one important example of the study of place names in ancient Javanese inscriptions is the monumental onomastics directory of Javanese epigraphy by the Frenchman Louis-Charles Damais. This systematically presents in its 1025 pages the findings of a study that took about 11 years (1955-1966) on anthroponyms and toponyms for geographical features, building names or administrative areas (Damais, 1970).

Textual evidence can often be profitably augmented or verified by ~~site~~ visits to historical or archaeological sites. Some examples of place name studies in Indonesia are given here.

25.5 Geographical Names for Natural Features

Gunung Krakatau / Krakatoa Mountain

The name Krakatoa is known around the world due to the massive volcanic eruption in 1883 (Winchester, 2003). The island and the volcano on it are situated in the Sunda Strait between Java and Sumatra. The earliest mention of it is in an old Javanese manuscript Pustaka Raja, The Book of Kings, which describes a volcanic eruption on a Mount Batuwara in 416 AD, which is almost certainly a reference to Krakatoa. Other eruptions took place in 535 AD and 1620.

Since 1927, a new volcano has become active and a volcanic cone has been growing. This is known as Anak Krakatoa, lit. Child of Krakatoa.

Although there are earlier descriptions of an island in

Figure 25-1 Anak Krakatau

the Sunda Strait with a “pointed mountain,” the earliest mention of Krakatoa by name in the western world was on a 1611 map by Lucas Janszoon Waghenaer, who labeled the island “Pulo Carcata” (*pulo* is the Sundanese word for “island”). The first known appearance of the spelling *Krakatau* was by Wouter Schouten, who passed by “the high tree-covered island of Krakatau” in October 1658 (Winchester, 2003: 27). Krakatau could have its origins in Sanskrit *karka* or *karkata* or *karkataka*, meaning “lobster” or “crab”. *Rakata* also means “crab” in old Javanese. This is considered the most likely origin. The Smithsonian Institution's Global Volcanism Program cites the local name, *Krakatau,* as the correct name, but says that *Krakatoa* is often employed. While *Krakatoa* is more common in the English-speaking world, *Krakatau* tends to be favored by geologists.

Pulau Onrust/ Onrust Island

Onrust Island is one of a chain of 110 islands called the Thousand Islands (Kepulauan Seribu) which stretch 45 km north from the West Jakarta Bay into the Java Sea. Onrust Island is 3,5 km^2 in size and is located 14 km from the Jakarta coast. It is important historically and culturally because it played an important role for the Dutch as part of their defenses of the city of Batavia (now Jakarta). The Dutch East India Company got permission in 1610 from the Prince of Jayakarta to build a naval base there. Its later Malay name Pulau Kapal (Ship Island) was given because of this function; it appeared for the first time alongside the name Onrust (“unrest”), given by the Dutch, on a 1761 chart of Batavia Bay by Gerrit de Haan.

After being destroyed by the British in the Napoleonic wars, the Dutch rebuilt Onrust Island as a place where

Figure 25-2 Model of Shipyard on Onrust Island

ships owned by the Dutch East Indies Government and private sector could be repaired. Because of the eruption of Krakatoa in 1883 and the construction of a new port in Tanjung Priok in Batavia, larger and better equipped for ship repair and ship building, Onrust Island went into decline.

After this, it was used as a quarantine station and as a hospital for sufferers of tuberculosis (TB). Between 1911 and 1933, it was also used to quarantine people who planned to go to Mecca to perform their Hajj. In 1972, when Ali Sadikin was Governor of Jakarta, the island was identified as a protected historical site. Indonesians continue to name the island Onrust to remind us of its cultural and historical value.

Figure 25-3 Old Fort on Onrust Island

25.6 Geographical Names for Built Features

Jalan Malioboro / Malioboro Street

Malioboro, is a street in Yogyakarta, an area with the status of a Special Province, which is situated in central

Java. Its special status rests on it being to this day the seat of the Mataram Sultanate where the present Sultan also holds the position of Provincial Governor.

There are a number of hypotheses about the origins of the name Malioboro. The first of these is that it derives from the name of John Churchill, First Duke of Marlborough (1650-1722), whose victories in Europe on behalf of the joint British/Dutch forces against the French were thus commemorated. Carey (2015) dismisses this. He points out that Malioboro is a main thoroughfare used for ceremonial processions by the royal family on special days. On such occasions, Malioboro Street is decorated with flower garlands according to the traditions of the royal house. Carey claims that Malioboro is most likely to have derived from the Sanskrit word *mâlyabhârin* which means 'decorated with flower garlands'.

Another theory of the etymology of Malioboro can be found in Suratminto (2016). Mataram was an agrarian kingdom. The population were not expert in trade. The Sultan, concerned that his people should not fall behind the times, encouraged them to embrace the practice of trading. The Indonesian word 'berdagang' (trade) is 'boro' in Javanese. The Sultan gave the following instruction in Javanese to his people, *'sira kabèh sing manggon nèng kéné* ***maliha boro*** *ing pasar gedhé iki kang* ***bèr ing raharja****.'* which can be translated as 'all of you staying here, shift from farming to trading in this large market full of benefits'. In this way, the Sultan's use of the word '*maliha boro' means* 'shift from farming to trading' and this word experienced a sound change to become 'malioboro'. The street took this name as it was a busy main road on either side of which were shops and traders. In addition, traditional markets nearby Malioboro were widely known and still are in operation to this day. The biggest of these, Pasar Beringharjo (Beringharjo Market) also takes its name from something said by the Sultan '*bèr ing raharja*' which means 'full of prosperity' (penuh dengan kesejahteran).

Figure 25-4 Malioboro Street (with the street name sign also in Javanese script)

Candi Borobudur / Borobudur Temple

Candi Borobudur (Borobudur Temple) is a 9th-century Mahayana Buddhist temple in Magelang, Central Java, Indonesia, as well as the world's largest Buddhist temple. Built during the reign of the Sailendra Dynasty, the temple was designed in Javanese Buddhist architecture, which blends the Indonesian indigenous cult of ancestor worship and the Buddhist concept of attaining Nirvana. The monument is both a shrine to the Lord Buddha and a place for Buddhist pilgrimage. Borobudur has the largest and most complete ensemble of Buddhist reliefs in the world. The monument was listed as a UNESCO World Heritage Site.

The Borobudur temple was mentioned in the *Karangtengah* inscription dated 824 and *Tri Tepusan* inscription, dated 842. The only old Javanese manuscript that refers to a monument called *Budur* as a holy

Buddhist sanctuary is *Nagarakertagama*, written by Mpu Prapanca, a Buddhist scholar of the Majapahit court, in 1365 (Moens, 1951: 326-386). In the form of a poetic eulogy for the monarch Hayam Wuruk, it provides an excellent description of palaces and temples and ceremonial observances.

Borobudur is Indonesia's single most visited tourist attraction. However, Borobudur is still used for religious ceremonies. Once a year, Buddhists in Indonesia celebrate Vesak at the monument.

The origins of the name *Borobudur,* however, are unclear. The original names of most ancient Indonesian temples are no longer known (Soekmono, 1976: 13, 2005). The name Borobudur first appears as 'Boro Bodo' in a manuscript map of Java by J.A. Dubois (Kaart van eiland Java, 1811. Manuscript, ARA, MIKO 73); this map was later engraved and printed for Sir Thomas Raffles' book on Javanese history published in 1817 (Raffles, 1965) in which he likewise refers to the 'Ruins of Boro Bodo'. There are no older documents suggesting the same name (Soekmono, 1976: 13).

Another possible etymology suggests that *Borobudur* is a corrupted simplified local Javanese pronunciation of *Biara Beduhur* written in Sanskrit as *Vihara Buddha Uhr*. The term *Buddha-Uhr* could mean "the city of Buddhas", while another possible term *Beduhur* is probably an Old Javanese term, still surviving today in Balinese, referring to "a high place". The morphology is constructed from the stem *dhuhur* or *luhur* (high). This would mean that Borobudur means vihara (meeting place of monks) of the Buddha located on a high place or on a hill.

Figure 25-5 Borobudur temple

25.7 Geographical Names for Administrative Features

Kota Cirebon / Cirebon City

Cirebon is a port city located on the north coast of Java nearly 300 km east of Jakarta. The city today has a population of around 300,000 people who are made up of Javanese, Sundanese and Chinese ethnolinguistic groups. It is also the seat of an Islamic Sultanate that goes back to the 15th century. There are in fact two palaces (*keraton*) today. These are the Keraton Kasepuhan Cirebon and Keraton Kanoman Cirebon. Cirebon has a long history of being a major port and center for trade, shipping and the spread of Islam in Java.

We know from archaeological evidence from a sunken ship, that the name Cirebon goes back to at least the tenth century. The names *Cĕrbon* and *Charabao* appear on fourteenth century maps written in Javanese and on fifteenth century maps in Latin. We can conclude that these place names were known at that time as far away as Europe and that *Cirebon* was also an international harbor during the period from the tenth to the fourteenth centuries (Ricklefs, 2001: 104). During Cirebon's golden age, in the fifteenth century, it was one

Figure 25-6 Keraton Kasepuhan

Figure 25-7 Inside the Keraton Kasepuhan Cirebon

of the most important ports of call on the sea routes for the spice trade and maritime Silk Road.
Some work has been done to determine the etymology and the meaning of the place name Cirebon. The name appears with variant spellings in a number of old manuscripts. There is more than one explanation of its origins and we present some of them here. The toponym Cirebon village *(Dukuh Cirebon)* appears in the text *Babad Tanah Sunda* (1447). Also, the name appears with the variant spelling *Cheroboam* in a Portuguese text by Tomé Pires called *Summa Oriental* (1513–1515) (Pires and Rodrigues, 1990). In the text *Sejarah Cirebon*, it is explained that the meaning of the village named *Cairebonan* is derived from *cai/air rebon* or shrimp paste broth, possibly referring to a cooked dish consisting of small shrimp (*rebon*) boiled in a broth and concentrated (*air rebon yang diberi bumbu petis*). The name *Cirebon* appears in the manuscript *Carita Purwaka* (1720 M)[1] which suggests that the name evolved through a process of sound change from *Carbon* and then *Caruban*. An example of a Chinese exonym for Cirebon can be found in the book *Shun-fêng hsiang-sung* (A good wind as a companion) which describes sea voyages between the fifteenth and seventeenth centuries from China to *Shun-t'a* (Sunda) with one going via *Che-li-wen* (Cirebon).

Kota Jakarta / The City of Jakarta

The city of Jakarta is the country's Special Capital Region and is the country's political, economic and cultural center. It is located on the northwest coast of Java Island. It has, over the years, had a number of names, namely *Sunda Kelapa*, *Jayakarta* and *Batavia*.

Its original name was *Sunda Kalapa*. The etymology of this name comes from the Malay word for coconut 'kelapa'. The area was a major port in the kingdom of Sunda Pajajaran where the area's name took after the coconut trees that lined the shore. Tomé Pires called it *Cunda Calapa* or *Sunda Kalapa* and wrote in 1515 that 'the Port of Kalapa belongs to the Sunda kingdom'. In 1527, Sunda Kalapa was captured by troops of the Demak Sultanate and the port fell under its ruler. Sunda Kelapa was renamed Jayakarta meaning 'victory' or 'absolute wellbeing'. The city name is recorded in numerous documents with variation in spelling: *Jayakarta*, *Jaketra*, *Jakerta*, and *Jakarta*.

The Netherlands recognized the economic potential of trade through Jayakarta, in particular its trade in pepper and other spices which they wanted to gain a monopoly on. In 1602, the ruler of Jayakarta, Prince Wijayakrama encouraged international trading activities. He gave the British and the Dutch permission to establish a trading station on the west bank of the Ciliwung River estuary. Many Chinese merchants built houses around the fort. Jayakarta developed into a busy port city.

Figure 25-8 Jakarta today

[1] This date comes from the appendix in *Pelabuhan Muara Jati Cirebon*.

The Dutch were determined to dominate the trade and eventually, Jayakarta fell into the hands of the Dutch

Figure 25-9 Portrait of Jan Pieterszoon Coen, founder of the city of Batavia

East India company (VOC). When the Dutch Governor General J. Coen (see figure 25-9) was in charge of the city, on 30th May 1619, he ordered the construction of a fort and a new urban area which he first wanted to name Hoorn, after his home town in the Netherlands. The board of the VOC, called the Council of Seventeen '*De Heeren Zeventien'* which sat in Amsterdam, however renamed Jayakarta Batavia, after the ancestors of all the Dutch, the ancient Germanic tribe of the Batavi. In the 18th century the city was described as beautiful and became known as 'The Queen of the East' or, because of its many canals dug by the Dutch, 'The Venice of the East' *Venetië in het Oosten*'.

Figure 25-10 Former Town Hall (Stadhuis) of Batavia/Jakarta

During the Japanese occupation in WWII, the name Batavia was changed back to Jayakarta or Jakarta through a Government News Announcement of the Japanese Military Government Kang Po No. 9 December 8, 1942 (Ruchiat, 2011). Up till the present, Jakarta has been the name of Indonesia's capital.

25.8 Conclusion

It can be seen from the few examples of toponyms given in this chapter that cultural heritage is not a monolithic concept. While Borobudur is recognized as a world cultural heritage site, the natural feature Krakatoa is not. However, it has made a strong impression on the minds of all who know the story of its great eruption and the effect on people still resonates today. In the sense that cultural heritage can include landscapes that interact with human perceptions, Krakatoa certainly qualifies.

The names of built features such as Malioboro Street and Borobudur both have complex histories involving different cultures and languages, but Malioboro, a street name in Yogyakarta, is fascinating because of the different possible hypotheses about its origin.

Names of administrative features are especially shaped by the periods of history that they were found in and in the kinds of data used to understand them. Archaeological evidence is strengthened greatly by the existence of textual evidence from different countries including travel journals and reports.

This chapter has focused on cultural heritage from a historical, geographical perspective, and drawn its data from archaeological or textual sources. However, in Indonesia there is another large repository of examples that there was no space to mention, that of the creation of place names for natural features from many of the country's 719 indigenous regional languages. Many of these are of value because they encapsulate local cultures and world views but many are endangered and as such also in need of protection. They are important for toponymy because they provide a huge source of local names for natural geographical features. The method of studying the larger ones can turn to textual sources, but the majority of those languages with less than a million speakers are oral languages without a script and so the only way to study the toponyms in them is to do anthropological, linguistic, or ethnographic field work.

25.9 References

-Bellwood, Peter S., Fox, James J., and Tryon, Darrel eds. (2006) *The Austronesians: Historical and Comparative Perspectives*. *Comparative Austronesian Series*. Canberra: ANU E Press.

-Blake, Janet (2009) UNESCO's 2003 Convention on Intangible Cultural Heritage: The implications of community involvement in 'safeguarding'. In *Intangible Heritage*. eds. Laurajane Smith and Natsuko Akagawa, 45-73. London & New York: Routledge.

-Carey, Peter ed. 2015. *Usul Nama Yogyakarta dan Malioboro [Origins of the Names Yogyakarta and Malioboro]*. Jakarta: Komunitas Bambu.

-Damais, Louis-Charles (1970) *Répertoire onomastique de l'épigraphie Javanaise (jusqu'à pu sindok 'Sri Isanawikrama Dharmmotungadewa): étude d'épigraphie indonésienne*. Publications de l'École française d'Extrême-Orient. Paris: École française d'Extrême-Orient.

-Eller, Jack David (2016) *Cultural Anthropology: Global Forces, Local Lives*. 3rd Edition. New York: Routledge.

-Haviland, William A., Prins, Harald E. L., McBride, Bunny, and Walrath, Dana (2011) *Cultural Anthropology: The Human Challenge*. Thirteenth Edition. Belmont, CA: Wadsworth, Cengage Learning.

-Helleland, Botolv. 2006. The Social and Cultural Values of Geographical Names. In *Manual for the National Standardization of Geographical Names*. ed. United Nations Group of Experts on Geographical Names, 121-128. New York: United Nations.

-Kadmon, Naftali. 2000. *Toponymy: the Lore, Laws, and Language of Geographical Names*. New York: Vantage Press.

-Kerfoot, Helen. 2015. Tracing UNGEGN's Evolving Interest in Geographical Names as Cultural Heritage. In *Geographical Names as Cultural Heritage: Proceedings of the International Symposium on Toponymy, Seoul, 7-9 November 2014*. ed. Sungjae Choo, 15-39. Seoul, Korea: Kyung Hee University Press.

-Lauder, Multamia RMT, and Lauder, Allan F (2015) Keragaman dan Kepunahan Bahasa di Dunia Melanesia. In *Diaspora Melanesia di Nusantara*. eds. Taufik Abdullah and Mukhlis PaEni, 189-225. Jakarta: Direktorat Sejarah dan Nilai Budaya, Direktorat Jenderal Kebudayaan, Kementrian Pendidikan dan Kebudayaan.

-Marrie, Henrietta (2009) The UNESCO Convention for the Safeguarding of the Intangible Cultural Heritage and the protection and maintenance of the intangible cultural heritage of Indigenous peoples. In *Intangible Heritage*. eds. Laurajane Smith and Natsuko Akagawa, 169-192. London & New York: Routledge.

-Moens, J. L. (1951) *Barabudur, Mendut en Pawon en hun onderlinge samenhang [Barabudur, Mendut and Pawon and their mutual relationship]*. Tijdschrift voor de Indische Taal-, Land- en Volkenkunde: Het Bataviaasch Genootschap van Kunsten en Wetenschappen.

-Nanda, Serena, and Warms, Richard L. (2012) *Culture Counts: A Concise Introduction to Cultural Anthropology*. 2nd Edition. Belmont, CA: Cengage Learning/Wadsworth.

-Pires, Tomé, and Rodrigues, Francisco (1990) *The Suma oriental of Tomé Pires* [1515] *and the Book of Francisco Rodrigues*. ed. Armando Cortesao. New Delhi: Asian Educational Services.

-Raffles, Thomas Stamford (1965) *The History of Java*. [1817]. Oxford in Asia historical reprints. Kuala Lumpur & New York: Oxford University Press.

-Ricklefs, Merle C. (2001) *A History of Modern Indonesia since c. 1200*. Third Edition. [1981, 1993]. Basingstoke, Hampshire: Palgrave Macmillan.

-Rose-Redwood, Reuben, Alderman, Derek, and Azaryahu, Maoz.(2010) Geographies of toponymic inscription: new directions in critical place-name studies. *Progress in Human Geography*, 34(4), 453-470.

-Ruchiat, Rachmat (2011) *Asal-Usul Nama Tempat di Jakarta [Origins of Place Names in Jakarta]*. Jakarta: Masup.

-Soekmono, R. (1976) *Chandi Borobudur: A Monument of Mankind*. Assen and Paris: Van Gorcum & The UNESCO Press.

-Soekmono, R. (2005) *Candi: Fungsi dan Pengertiannya*. Jakarta: Jendela Pustaka.

-Tanudirjo, Daud Aris (2006) The Dispersal of the Austronesian-Speaking People and the Ethnogenesis of the Indonesian People. In *Austronesian Diaspora and the Ethnogenesis of People in the Indonesian Archipelago*. eds. Truman Simanjuntak, Ingrid H.E. Pojoh and Mohammad Hisyam, 83-98. Jakarta: Indonesian Institute of Sciences (LIPI) Press.

-Taylor, Simon (2016) Methodologies in Place-Name Research. In *The Oxford Handbook of Names and Naming*. ed. Carole Hough, 69-86. Oxford: Oxford University Press.

-UNGEGN (2006) *Manual for the National Standardization of Geographical Names*. United Nations Group of Experts on Geographical Names. New York: United Nations.

-Watt, William (2015) Getting the Names Right: Geographical Names as Cultural Heritage and Activities

of UNGEGN Working Group. In *Geographical Names as Cultural Heritage: Proceedings of the International Symposium on Toponymy, Seoul, 7-9 November 2014*. ed. Sungjae Choo, 40-48. Seoul, Korea: Kyung Hee University Press.
-Winchester, Simon (2003) *Krakatoa: The Day the World Exploded, August 27, 1883*. London: Penguin Books.

Chapter 26 **Etymology and the historic study of geographical names**

Isolde Hausner

26.1 What is etymology?

Etymology is known as the study of the origins, the „true sense“and meaning of words. Etymologies answer questions about historical relationships of and among languages and in that sense, they address additionally to the linguistic issues the full historical dimension of toponyms. Place names or toponyms (or geographical names [see Glossary no. 224, 216 and 339]) are the general terms for any place or geographical entity and as such are the research objects of onomastics, which is known as a subbranch of linguistics. Geographical names are defined as 'names applied to a feature on Earth' and they are closely connected to toponyms, which are defined as 'comprehensive term for geographical names and extraterrestrial names' (see Glossary no.216 and 339).
In general, a geographical name is the proper name (a specific word, combination of words, or expression) used consistently in language to refer to a particular place, feature or area having a recognizable identity on the surface of the Earth (Manual p. 9). Toponyms are constituent parts of cartography and topography.
It needed some centuries foretymology to develop as a scientific discipline from the early attempts of etymologies in Platons dialogue *Kratylos* (see www.christianlehmann.eu [Schriftenverzeichnis → Etymologie; accessed 10-01-2016) and by the Greek Stoics until the present day. Etymology in a broad sense can be understood „… *as historical research that aims at investigating the origin of a word (lexical etymology) or structure (structural etymology)…“* (Bibliographical note to Mailhammer, ONB Vienna [online catalogue]; see also Mailhammer p. 3f.). And with the focus on place names one can add, that etymology is moreover a scientific method, which opens an interface to the large variety of human actions in time and space, because toponyms own a special position within linguistics and exert manifold functions expressed through their lexical, structural and historical peculiarities.
„There are two basic etymological configurations. Either a word is native in the sense that it was created at a chronologically earlier stage of a language and handed down from one generation to the next, or it came into a language from another through some form of language contact“ (Mailhammer p. 9).

Modern etymology regards the diachronic as well as the synchronic aspect of etymological analysis. Synchrony and diachrony are two opposing but at the same time complementary perspectives on linguistic phenomena, that makes a distinction between a sheer synchronous or diachronous perspective impossible. Synchrony refers to the static stage of a language (word formation, sound laws etc.), while diachrony is affiliated with the history of words (name history, language changes etc.). Etymologies of toponyms deliver informations on the language, relationship, phonological and semantic structure, the history of words and in regarding all these factors, and transcending the limits of linguistic research, it opens a window to a whole cosmos of human settlement activity (See Bergmann, H. et al. [2009]). The socio-cultural aspect cannot be left out of consideration and it is commonly agreed, that toponyms belong to the immaterial cultural heritage of mankind. Other than with common nouns the toponyms (like all names) own a third dimension, that is the identifying quality and vice versa the named features enable us to take a look at the times when the places were named, they give us insight into the original intentions why this name was given, under which socio-cultural conditions a social community coined this place name.
Linguistic science came into being in Europe and so Europe has a long tradition in etymological research. Most of the significant etymologists were and are native speakers of Indo-European languages. The 19th century brought a rich increase of linguistic knowledge which inspired the etymological approach to words and nouns and by the end of that century etymological dictionaries for most of the Indo-European language families were published. In the 20th century, especially in the decades after the Second World War the methods and theories of modern etymology were developed, which opened a new era to the discipline of etymology. Including the social factor and merging it with the linguistic details opened a new dimension for the understanding of the nature of names.
Geographical names have and had in all the times a broad cultural significance. When we speak of toponyms and especially of toponyms in cartography, their overall significance lies in their multiple functions (address function, identification, language, social, economic and cultural relevance, the whole system of the mental comprehension of natural and cultural environment of peoples) which form the multi-faceted cultural contexts around the name-giving act of toponyms. Names are in the present and were in the past coined by a social community with a specific cultural background or

worldview. A farming society, for example, coins names that differ from those of an industrial community; the natural environment is reproduced in the adjacent toponyms (Mont Blanc 'white mountain'), other geographical features as, for instance, seas bear names from the worldview of the Greek mythology (Atlantic Ocean 'Sea of Atlas').

26.2 Etymology as the basis for elaborating a standard form of geographical names

Onomastics as a subbranch of linguistics is the study of the origin, history and use of proper names. Toponymy is one of the most important branches of onomastics and is understood as the study of place names. Factors such as the meaning of the elements out of which the toponyms were created, encompassing the topography, geology and ecology of the places bearing the names, the general and local history and the surrounding natural features are integral components that shape and mirror the history and culture of a region, of a state.

26.2.1 The etymological approach to analyze toponyms, to compile an etymology, comprises linguistic and extralinguistic factors. One must be aware that the extralinguistic factors outweigh the linguistic ones. The standardization of existing and the coining of new geographical names are highly dependent on political factors, such as official language(s), the authorization of traditional or past colonial name forms etc., be it a traditional spelling or naming, or the authorization of a particular language, of minority languages (additionally or stand-alone) or a special dialect. In the case of abandoned toponyms, e.g. for historical sites or excavations, for religious places etc. one is well advised to research the history around that particular toponym and on this basis reconstruct an orthographically and phonologically correct standard form, which can undergo the official act of standardization by the authorities. Specialist knowledge is required to analyze the origins and meaning of toponyms, tracing their linguistic and extralinguistic history back in order to determine which language and which social group they are from.

The etymological research concentrates on the following linguistic (I) and extralinguistic (II) factors:

26.2.1.1 Linguistic factors and methods in etymological research

I: language, historical records, word formation, phonetics, vowel and consonant alternations, meaning, semantics, naming motives, dialect pronunciation and, not less important, the folk etymology if there exists one.

The first step in the etymological processing of toponyms consists of the language survey: what is or are the present language(s) of a state, what former languages were in use and what is or are the authorized language/s. The next task is to consult the archival sources as well as old maps to set up a data set of historical records for a specific toponym from its first recorded entry until the present, whereby the first historical record is given particular attention as regards the spelling and language stage.

Source critics concerning the time horizon at which a text or a map was produced, and concerning the commissioners in whose service the authors (especially for maps) stood, complete the diachronic process. If an author is from outside and not familiar with the regional language(s), orthographic specifics from other languages can influence the spelling. Archival sources provide us the earliest written forms of a toponym and the etymologist must have a good knowledge of the historic, palaeographic and linguistic peculiarities to handle this challenging task. A discontinuity and an interruption in the sequence of the historical records or the occurrence of a completely new name are indications of extralinguistic processes, be they a population shift, the peaceful immigration of other populations, the takeover of the reign by colonial powers or the devastation of home grounds resulting from natural disasters or wars; a potential later resettlement can create completely new names based on new motives.

Toponyms very often preserve old spellings, dialectal forms and an antiquated vocabulary. Applied rules of sound changes from the first recording until the present lead to the current spelling of the researched toponym. The Austrian city *Inns<u>bruck</u>* e.g. shows in its generic element *-bruck* the old Bavarian dialect form which was standardized instead of the standard language German *-brücke* ('bridge over the river Inn'). It proves, that geographicalnames must not be subjected to the same orthographical and grammatical rules as common words. They very often preserve antiquated spellings. One must also regard that the meaning of the word referred to by the name can have undergone a semantic change in the course of the centuries. In present times frequently occurring toponyms may be equipped with distinctive adjuncts for an unambiguous localization (see the example below).

The dialectal pronunciation explored in the course of the field collection of the toponyms can support the plausibility of the etymology or may give suggestions for a completely different origin of a place name. Moreover, fieldwork reveals which names are in use – written or orally - and if there are namevariants besides the official ones. Local informers should be carefully selected in regard to a longtime established dwelling place in that region, because only they know the correct local pronunciation, potential names variants and myth and legend associated with these names.

In the case of a contact etymology, that is when peoples from other countries with other languages immigrate to a country, the two languages can merge or one language may prevail over the other. The substitution rules from one language to the other, that is from the source language to the target language, are applied, vowel and consonant substitutions can be evidence of the period when the immigration took place. An example of these processes sis the Austrian place- and river name *Pielach,* originally of Slavonic origin to Slav.**běla* 'white' (reference to the quality of the water). This name indicates two linguistic processes relevant for the transfer of data from Slavonic to German: (1) the initial Slavonic *b-* came into German (= the South Bavarian dialect as part of the German language) as *b-,* which was only possible up to the eighth century according to the substitution rules (from the 9th to the middle of the 11th century it would have been substituted as *v-/f-* as it can be seen in the Lower Austrian hydronym *Fladnitz* attributed to the same Slavonic stem *běla*). In the following it underwent the High German Consonant Shift *b-* > *p-* in the second half of the eighth century and after that the German sound changes. (2) The Slavonic etymon **běl-* no longer took part in the Slavonic liquid metathesis > *blē-* of the ninth century. These sound processes allow conclusions for the time horizon, when Slavonic settlement in that specific region of Austria ceased and was gradually replaced by Bavarian colonization.

The naming motives on the other hand are more or less found in the extralinguistic sphere and are better explained below (under item 32.2.1.2).
Frequently one is confronted with a folk etymology. It is the name of the process of making obscure or opaque toponyms transparent by phonetic adaptation or by sound assimilation by the local population, giving the name a new meaning, which has nothing to do with the original name, but relies only on phonetic similarities or a reinterpretation of the genuine name. The difference between etymology and folk etymology lies in the transformation of an unknown word by analogy or assimilation and a name, which has been modiefied in such a way cannot be traced back to its original stem by linguistic recursion (see the example of *Mailand* below) These modified names loose their synchronous and diachronous quality as authentic cultural-historical testimony.
We all treasure the identification function of toponyms, and this might be the background for their retention of very conservative forms in spelling, grammar and often also in vocabulary that are no longer understood by the local residents. This quality makes them favourite objects for a popular explanation, generally known as folk etymology (e.g. the German exonym *Mailand:* Latin *Mediolanum* > Middle-High-German. *Mîlan* > New-High-German *Mailand*, where the old Latin – element – *[p]lanum* 'plain' (*Medio-[p]lanum* 'situated in the middle of the plain [of the river *Po*]) was not understood and was substituted by the -German generic element –*land* 'land', which made the generic element understandable for a German speaking community). This new generic element connotes a secondary motivation, which was created by assimilation to the German lexicon and is far away from the original meaning of that name.

26.2.1.2 Extralinguistic factors in etymological research
The extralinguistic aspects comprise topography, geology, ecology, and the local as well as the general history of a country. They play a prominent role in the analysis of the toponyms and give important contributions to the culture of a specific region. „Going back to the roots"means also to ascertain the naming motive of a particular toponym. The naming motives belong to the extralinguistic sphere of toponyms and are manifold. They reveal in most cases a further quality of toponyms, namely the culture-specific testimony of the economic, ecological and natural preconditions in the earliest times of human habitation. The etymologist is well advised to contact experts from various disciplines, to find the meaning of handicraft or farming terms, or technical terms from mining or geology, biology and botany, navigation or religion. Sometimes toponyms may also refer to regional myths and legends.

26.3 Place names and their manifold etymological histories

A **transparent etymology** exists, when the linguistic element or elements composing the toponym, be it a simplex or a composite toponym, can clearly be understood. An example for such a case is e.g. the toponym *Amsterdam* 'dam in the river Amstel'.

The **native** / **continuous etymology** is the case, when the toponym shows a continuous recording (regarding the regular vowel and consonant changes in the spellings in the single language periods) through the centuries: for example, *Paris* derives its name from the Celtic ethnic group of the *Parisii*, living on the banks of the river *Seine* with their capital city *Lutetia* (affiliated to the Indo-European stem **lut-* 'mud, swamp') , in Roman times called *Lutetia Parisiorum*, which resulted in the present name *Paris*, omitting the first component. *München*, the capital city of Bavaria/Germany derives its name from the - Old-High-German word *munih* 'monk' (< Latin *monacus*), the first record dates from 1158 *apud Munichen* > *München* (*u* > *ü* by metaphony).

26.3.1 Name change as a result of migration, immigration or population shift:
Many places in Europe look back at a long history. The Roman Empire left behind numerous place names, but not all of them found a continuation after the decline of the Roman Empire which happened about the 5th century. One of the examples with a broken historical names tradition is the capital of Austria, *Vienna*.

The Roman name was *Vindobona*, a compund with the generic element *-bona* 'estate' and the Celto-Roman personal name *Vindos*; it became extinct after the withdrawal of the Romans and only in the 9th century the new name recorded in a historical source from 881 (copy 12 th century) *ad Wēniam* > *Wien* of Germanic origin came up; it was given after the name of the river *Wien* (Old-High-German *witu* 'wood, timber'), which has its source in the Vienna Woods and flows through the city. The name was probably transferred as the name for the city by the Langobards, who settled for a short time in that region.

Another example is the Croatian city now named *Dubrovnik*, which had originally in the 7th century the name *Ragusa*, when the inhabitants of *Cavtat* fled to the nearby coastal rock where now Dubrovnik is situated. It is uncertain where the name *Ragusa* came from, it is presumably a pre-Roman name, probably even belonging to a so-called Illyrian stratum; in Roman times, it was recorded in the historical sources as *Ragusium* (Schramm p. 75 and p. 331 ff.). Due to historical testimonies from the 12 th century the inhabitants assimilated with the -Slavs residing nearby and the city obtained the new name *Dubrovnik,* derived from Slavonic **dǫ bъ* 'oak'. The naming motive was its position below the mountain *Srđ*, which was in the past forested with oak trees.

Large regions or long rivers can have **sectional names**, like the river *Brahmaputra*, which flows through China, India and Bangladesh and is consecutively named *Tsangpo* ('river flowing through the western Tibetan province Tsang'), *Brahmaputra* ('son of Brahma'), *Jamuna* (vague etymology, perhaps *Yamuna,* the daughter of the sun god Surya in Hindu mythology) and *Meghna,* all with different etymologies. On the other hand, Europe`s second-longest river the *Danube*, flows through ten countries and has nowadays only one name, derived from the Roman name *Danuvius* and adapted to the modern languages of the individual countries. But in the past it had two different names for different sections of the river: *Ister* (for the lower courseIndo-European stem **oi̯s-/*is-* 'fast flowing') and *Danube* (Indo-European **dah*$_2$- 'flowing'). The extinct name *Ister* referred to the gorge at the Iron Gate on the border between Serbia and Romania, whereas the name in the upper and middle course reflects its rather moderate flow character.

26.3.2 Name changes as a result of political changes:
Wars always change the world map and old names disappear while new names arise. The change of the geographical landscape, visible in the substitution of previous traditional names by new toponyms, is often the result of a newly established state or of a new political regime. Toponyms in their function as identity markers are a medium to strengthen the power of the new government. For example, the German city *Chemnitz* derives its name from the Slavonic name of the small river which flows through the city ('river rich of stones') and thus communicates knowledge about the geology of that region. During the existence of the German Democratic Republic ([GDR] with a communist government) from 1953 to 1990 the old name was replaced by *Karl-Marx-Stadt* (Karl Marx, the founder of communism was born in Chemnitz). After the demise of the GDR the population voted for a return of the old name.

Apparently **arbitrary name changes** can also result from politically motivated reasons**:** the capital of Kazakhstan

Astana (meaning 'capital city'), was founded as *Akmolinsk* (1830-1961), from 1961 to 1991 it was named Zelinograd (meaning in Russian: 'new land city'), between 1992 and 1998 it was known as *Aqmoly* (1992–1998, meaning 'white grave'; Wikipedia accessed 2015-10-01), in 1992 it was renamed as *Aqmola*, and finally in 1998 it received its present official new name *Astana.*

Coining a new name can take place, when in the course of political realignments new states arise or when new settlement areas are created. The capital of Pakistan *Islamabad* (Urdu اسلام آباد 'residence of Islam'; Wikipedia accessed 2015-10-01) was founded in the 1960s; this new composite name refers in its specific element to the religion 'Islam' while its generic element *-ābād* means 'inhabited place; city, town'. This type of name fits into the culture-specific scheme of place names in the subcontinent, like e.g. *Hyderabad***.**

Another example is the new capital of Myanmar, when in 2006 the government decided to found a new capital city with the name *Naypyitaw* (generally translated as 'royal capital'; Wikipedia accessed 2015-10-01). Its official name was announced on 27 March 2006, spelled *Nay Pyi Taw* (sometimes also as *Naypyidaw*). Failing an official Romanization system in Myanmar there are **Variant spellings of names** in use, e.g. *Bagan* (variant *Pagan)*, *Irrawaddy* (river), *Ayeyawady* and *Ayeya(r)wady* (region), *Yangon* (variant *Rangoon*).

Former colonial names and the **reinstatement of indigenous names:** The generally accepted political tendency nowadays is to establish a programme for the revitalization of the old names, to collect, etymologize and officialize them (see Australia, Canada)**.** One can remark, that the global usage of the reinstated names gains credence. *Mount Everest* (named after a British surveyor of India) is a good example: its Tibetan Pinyin name is *Qomolangma* ('Holy Mother' in Tibetan [Wikipedia 'Mount Everest' accessed 3-11-2015) and this spelling was standardized by the Chinese authorities, romanized as *Chomolungma, Qomolangma Feng.*

Mount Everest's north face from the Tibetan plateau (Wikipedia, accessed 3/10/2015)

The same took place very recently, when in 2015 *Mount McKinley* (named after an American presidential candidate), the highest mountain in North America, was renamed officially with its old local name *Denali* (meaning 'high, tall' in the local Koyukon Athabaskan Indian language), which has always been used for centuries by the indigenous people in Alaska.

Another good example for the reinstatement of an old name is mount *Uluru* in Australia. Located in South Australia, which allows double naming, the feature is now officially referred to as *Uluru / Ayers Rock.*

Colonial names are a special chapter in the historic study of place names. They can be found all over the globe, **and** were introduced by the Colonial powers; they replaced the local place names or are new names. The histories of colonized regions are mostly well documented and so the naming motives are well known. The striking characteristics of colonial place names are that they contain names of persons, which are associated with the colonial histories or that the colonizers named their new territories or settlements after a place in their home country. Tracing back the etymologies of the colonial place names, one learns that in the early stages of colonialism the names were not stable, underwent many changes, dependent on the alternate occupiers, for example: Nieuw Amsterdam (1625-1664) – New York (1664-1673) – Nieuw Oranje (1673-1674) – New York (1664-).

Namibia, formerly known as German South-West Africa, then South West Africa, a German colony between 1884 and 1919 holds a number of toponyms of Germanic origins: *Lüderitz* (since 2013! Nami-*Nüs* [with click consonants] in the Khoisan language), founded in 1884, was named after the tobacconist *Franz Adolf Eduard Lüderitz* from Bremen, one of the founders of this first German colony in Africa. Another city in this territory with a German name is *Schuckmannsburg* (founded 1909)*,* named after Bruno von Schuckmann, the imperial governor of the colony in the years 1907 until 1910. With the administrative reform in 2013 it was renamed *Luhonono.*

The Democratic Republic of the Congo, formerly known as Belgian Congo, then Zaïre (1971-1997) is another example for a broad inventory of colonial names. In this region, we are sometimes confronted with names coined on a French language background: e.g. *Léopoldville* (in 1966 renamed *Kinshasa*), named in honour of King Leopold II of Belgium; Élisabethville (Belgian Dutch *Elisabethstad*, named after Queen Elisabeth, wife of King Albert I of Belgium) → from 1960 on *Lumumbashi; Stanleyville* (Belgian Dutch *Stanleystad*, named after the Welsh explorer Sir Henry Morton Stanley) → *Kisangani.*

26.4 References

-Durkin, Philip (2009): The Oxford Guide to Etymology. Oxford University Press.

-Geographical Names as a Part of the Cultural Heritage (2009). Wiener Schriften zur Geographie und Kartographie, Bd. 18. Ed. by BERGMANN, H., CHEETHAM, C., HAUSNER, I. and JORDAN, P.

-GLOSSARY of terms for the standardization of geographical names. Department of Economic and Social Affairs Statistics Division United Nations Group of Experts on Geographical Names ST/ESA/STAT/SER.M/85. United Nations, 2002. http://unstats.un.org/unsd/geoinfo/UNGEGN/publications.html

-Lehmann, Christian: see www.christianlehmann.eu → Schriftenverzeichnis

-MAILHAMMER, Robert (ed., 2013): Lexical and structural etymology. Beyond word histories. *Studies in Language Change*. Edited by Cynthia ALLEN, Harold KOCH, Malcolm ROSS, Vol. 11.

-MANUAL for the national standardization of geographical names / United Nations Group of Experts on Geographical Names. http://unstats.un.org/unsd/geoinfo/UNGEGN/publications.html

-ONB: Österreichische Nationalbiblithek (Austrian National Library)

-Schramm, Gottfried (1981): Eroberer und Eingesessene. Geographische Lehnnamen als Zeugen der Geschichte Südosteuropas im ersten Jahrtausend n. Chr. Anton Hiersemann Stuttgart.

-Seebold, Elmar (1981): Etymologie. Eine Einführung am Beispiel der deutschen Sprache. Beck`sche Elementarbücher.

Section 10 Toponymic research and documentation

Chapter 27 **Some aspects of cartographic and linguistic toponymic sources**

Hubert Bergmann

27.1 Introduction

Why would someone want to consult such sources? In order to find the official spelling of the name of a topographic object (its standardised form), its meaning, its derived versions (names of its inhabitants, adverbs derived from it. Or they might want to know how a toponym behaves in a text – is it declined according to its case (such as a locative); are there different forms for: to Wien, from Wien or at Wien (or suffixes and prefixes might be added*);* is it singular, dualis (Bahrain = the 2 seas) or plural, its gender can be relevant ('Wien and her churches'?). Then the definite or indefinite versions of the name might be relevant, and where to find which ones should be entered on the map

The range of sophisticated toponymic sources, be it print publications or online material, is very wide. Basically, they can be divided into two groups: The first group is composed of publications – for the main part gazetteers – that are the result of cartographic documentation or aim at the documentation of toponyms for reasons of standardization. Often these two aims (cartographic representation and standardization) are interlinked, because in many (but by far not all) countries the state authorities in charge of cartography are also responsible for the standardization of geographical names. The second group of sources comprises publications or online material that was elaborated mainly out of an academic, mostly linguistic, interest in place names and focuses on toponymy as a sub discipline of linguistics. These two groups of publications basically focus on different aspects of place names though they sometimes overlap.

Apart from such sophisticated toponymic sources a lot of other linguistic publications may also contain information that can be useful when dealing with toponyms. Dictionaries for example may contain toponyms as well. With regard to several aspects of geographical names they may even be the first source of information. Even grammars may be of relevance for the study of place names, because toponyms may follow specific rules when it comes to the grammatical system of a language. Apart from aspects of spelling and grammar, also the phonetic side of names (i.e. their pronunciation) and the respective sources of reference have to be taken into account. Finally, a lot of toponymic information has been or is still being collected in connection with field research for different purposes by representatives of different disciplines. This material, which is often not published, may also be useful for anybody interested in place names.

27.2 Toponymic publications for cartographic documentation or standardization purposes

Many state bodies that deal with the standardization of geographical names have published gazetteers of standardized place names. Within the frame of United Nations Group of Experts on Geographical Names (UNGEGN), a Working Group on Toponymic Data Files and Gazetteers was established which "addresses various issues of toponymic data processing, including the creation, maintenance and outputs of toponymic data files (including gazetteer production) and aspects of toponymic data exchange formats and standards."[1] Gazetteers or databases of this kind primarily focus on the correct spelling of names and their categorization and localization, although also other aspects may be taken into consideration (correct standard pronunciation, administrative affiliation etc.). These publications are therefor useful for finding the official spelling of the name of a topographic object, or – in other words – its standardised form. An example for a print publication of this kind is the "Gazetteer of Austria", published by Joseph Breu in 1975. Its entries contain the following information: Standard German pronunciation according to the International Phonetic Alphabet (IPA), description of the feature (topographic category, description of the location in words, administrative affiliation, geographical coordinates, indication of the respective sheet number of the official map) and – if necessary – grammatical information.[2]

Nowadays it is mainly online databases that have replaced printed gazetteers. As an example the New Zealand Gazetteer provided by Land Information New Zealand (LINZ) can be mentioned (see chapter 13); it contains "all official names for features within the New Zealand Geographic Board's coverage" and "other types of unofficial names, such as recorded and collected names".[3] There along with each place name the following information is provided: status of the respective name (official or non-official) plus reference information, administrative affiliation, feature type plus description of position and exact coordinates. The respective name is also shown on a map.

As I have already mentioned in many countries state cartographic services are also responsible for the standardization of geographical names. In these cases, their cartographic products may contain standardized place names and therefor assume the character of a gazetteer. In Austria, for example, there are no standardized special gazetteers for names of rivers, brooks, mountains etc., of which the above-mentioned gazetteer by Breu contains only the major features. The products of the Federal Mapping Agency (BEV) can be regarded though as de facto standardized names. As there are also no place name gazetteers published any more in this country, the maps of the BEV (available online and searchable[4]) therefor assume the role of a work of reference when it comes to standardized place names.

27.3 Toponymic publications for academic (linguistic) purposes

Place names have attracted the attention of linguists since the very beginning of modern philology as a discipline. Their predominantly historical-etymological interest in toponyms has different reasons: Place names conserve elements of the lexicon that are not used any more in the modern varieties of languages. Furthermore, in many areas of the world place names are the last relics of languages that were once spoken in the respective regions, but which – for whatever reason – have died out. The way these names have been taken over and integrated into the linguistic system in turn may reveal interesting details about the history of the language that prevails today. These are just a few reasons why onomastics have developed into a branch of its own within philology. There are also other disciplines though that are interested in the results of place name research. The etymology of a name may tell interesting stories about the history of a place and its environment, that's why toponomastics are also relevant for historians and archaeologists, for example. And finally names whose etymology refers to environmental incidents like floods or avalanches or to soil conditions can be used as indicators for potential risks, especially in an era of rapid climate change.

As the historical-etymological dimension is a central – but not by far the only – element of onomastic research, the collection of the earliest traceable written records of a place name is of crucial importance for the documentation and interpretation of place names. Why? Simply because names may be subject to substantial changes in both pronunciation and spelling over the centuries. In many cases in the German-speaking area, for example, the initial "meaning" of a place name got lost, either through language change or innovative processes within the German language. In the course of time people tried to remotivate the name and thus modified it. The village name *Fraunhofen* in Upper Austria for example is today associated by most speakers of German with the nouns *Frauen* 'women' and *Hof* 'farm'. The oldest record of this name dating from the year 1342 (*Vronhof*) though shows, that it is actually not a 'women's farm', but a 'feudal manor', for the first element of this composed name reflects Middle High German *vrôn* 'command'.[5] Depending from where we are in the world, these records may date back several hundreds or even thousands of years – or only a few decades. Even in areas where we have place names already mentioned in mediaeval documents, the density of historical records may vary from one category of place names to another. In Central Europe for example names of towns or villages or major rivers tend to appear earlier in written documents than field names or names of mountains or brooks, for example. The above-mentioned relevance of historical records is also mirrored in the structure of scholarly toponomastic reference books, of which they are often an essential component. Furthermore, such publications often contain the current phonetical realization of the names, which may substantially differ from the standard phonetical form due to the presence of dialect variations. The historical-etymological dimension of place names may also be of relevance for standardization purposes, e.g. if it is used as an argument for (or against) the change of the official form of a place name.

The classical format of academic literature on place names that covers larger areas is the place name dictionary. It is often structured according to alphabetical order, but also other order criteria (e.g. geographical position or administrative affiliation) may be followed. Nowadays several publications of this type are also available online, either simply as pdf-files (sometimes searchable) or as more sophisticated online publications, based on databases. As examples for place name dictionaries available as (searchable) pdf files the 1st edition of P. E. Raper's "Dictionary of South African Place Names" can be mentioned[6], or – as an older example – the "Place Names Dictionary of Halland", a Swedish province.[7] As a good example for a (printed) multivolume place name dictionary transferred into a database format and now also available online (including mapping of the names) is the "Place name dictionary of Liechtenstein" ("Liechtensteiner Namenbuch")[8]. It covers more or less the complete toponomastic landscape of this country.

Apart from place name dictionaries we find a broad variety of toponymic publications that deal with all possible aspects of place names. Due to the nature of the subject, these publications often also contain maps. Some scientific publications even focus on the mapping (and thus visualization) of their toponymic content. An example for such a publication could be the "Typological atlas of Old Sorbian place names" ("Atlas altsorbischer Ortsnamentypen")[9], analyzing a certain area in East Germany and a smaller adjacent Region in Poland.

To get an overview of publications on toponomastics, bibliographies are a useful tool. The International Council of Onomastic Studies (ICOS) has been drawing up an onomastic bibliography for many decades (1950-1996), published in its periodical “Onoma”. In recent years a database version of this bibliography has been created.[10] Publications on onomastics may also be included in bibliographies of other disciplines. We find for example a section “Onomastik/Onomastique” within the “Romanische Bibliographie Online Datenbank 1965–2013”, an online bibliography of Romance Studies published by the German publishing house De Gruyter (which is not free of charge, though). This section is again subdivided in different subsections, for example in “Toponomastik/Toponomastique”.[11]

27.4 Other linguistic publications relevant for the study of place names

Apart from place name dictionaries or other literature that focuses on toponomy, there are other linguistic works that may provide useful information with regards to toponymy. Classical language dictionaries for example, be it standard language dictionaries or dictionaries for non-standard varieties (dialects, sociolects, etc.) very often contain toponyms as well. The respective approaches vary from publication to publication, in any event it can be useful to consult the foreword or other outline texts where criteria for the inclusion of toponyms may be described (or not). There are several aspects of geographical names for which dictionaries may even be the first source of information. One example is the adjectival forms of toponyms or the names of residents or natives of a particular place (demonyms). There are languages in which these derivations follow more or less strict rules, e.g. German: *Wien* (i.e. Vienna) – *Wiener* (adjective) – *Wiener*, *Wienerin* (demonym, masculine and female form). In many languages though these forms are not predictable because of competing derivation models, e.g. in Italian: *Milano* (i.e. Milan) – *milanese* (adjective) – *milanese* (demonym, masculine and female form), but *Napoli* (Naples) – *napoletano* (adjective) – *napoletano* (demonym, masculine and female form). As place name gazetteers, sometimes do not show this kind of information, it can be very useful to consult a dictionary, in the case of demonyms of Italian towns for example Lo Zingarelli in its 2010 edition. There we not only find *milanese* and *napoletano* as lemmas but also the sections “Inhabitants of Italy” (i.e. demonyms of Italian towns and regions) and “Inhabitants of the world” as an appendix at the end of the volume.[12]

As place names may assume special roles in the system of a language and “behave” different from appellatives, they are very often also taken into account by grammars. If we take a popular grammar of Standard German, for example,[13] and look up “geographical names” (in German “geographische Namen”) in the index, we see that they are considered under the following sections of the grammar: semantic description of nouns, declination of names, pluralia tantum (cf. geographical names used only in the plural form, e.g. *Alpen*, *Kurilen*, *Niederlande* in German or *Alps*, *Kuril Islands*, *Netherlands* in English) and singularia tantum (a category to which practically all geographical names used in the singular form belong, e.g. *China* in German or English), use of the definitive article and the zero article (cf. the German name for Turkey, which must be used with the definitive article only [*die Türkei*], whereas other geographical names are used without any article [eg. *China*], the latter phenomenon being called “zero article” in grammars of the German language), particularities in connection with the use of articles.

An important aspect of the linguistic nature of toponyms, apart from spelling and grammar, is their pronunciation, their phonetic form. In some languages, the spelling of place names may be of a very conservative nature and follow earlier orthographic principles. The correct pronunciation may therefore not be predictable. This is for example the case with the German town of *Duisburg*, where the digraph *<ui>* is pronounced as [yː], which is against the orthoepic principles (i.e. rules of pronunciation) of (today’s) Standard German. Place name gazetteers may contain this kind of information, but also dictionaries[14] (often only of major or commonly known objects though). For many languages, special pronunciation dictionaries (print editions with phonetic transcription or online versions with integrated audio files) exist, which may also contain toponyms. Last but not least the “Placenames Database of Ireland” can be mentioned as an excellent example for the successful inclusion of an existing specific toponymic database containing pronunciation audio files in a wider online geographical names database, linked to online maps, scans of unpublished toponymic material, a “Glossary of words commonly found in Irish place names”, biographical resources (“Irish-Language Biographies”) etc.

27.5 Unpublished material relevant for the study of place names

Apart from publications or online sources also unpublished material can be very useful for the study of toponymic issues. In connection with linguistic field research (for example in connection with a dictionary project) more or less as a huge amount of toponymic data may be collected. If we take a closer look at some questionnaires for such dictionary projects we may find questions which directly aim at collecting toponyms. Number C5 of the questionnaire of the “Dictionary of

American regional English" for example reads as follows: "What are the names of some of the lakes and ponds around here?"[15]

Also ethnological research projects may implicitly help to document toponyms. The material of the "German Ethnological Atlas" ("Atlas der deutschen Volkskunde") for example contains also a lot of information sent back by the thousands of persons who filled out the respective questionnaires in the 1930s. A few questions even directly address toponyms, for example question 167b, where informants are asked if a specific road or path is used by the funeral procession on its way to the cemetery and what the name of this road or path is.[16]

Even if these toponyms are often not in the focus of the projects by which they were collected, it would be a great benefit also for onomastics to have the respective unpublished and often not analysed material at least digitized and available online. This is for example already the case with the "German Language Atlas" ("Deutscher Sprachatlas"), a huge long-term project started in the 19th century and aiming at mapping dialect phenomena of German. In a certain phase of the collection process (1880s until 1930s) informants were also asked to write down the name of their village in the local dialect form. Due to this initiative, there are thousands of questionnaires available from Southern Germany, Luxemburg, Switzerland, Austria, Southern Tyrol and other regions containing this kind of information. Since a few years the scans of the respective questionnaires are now online and accessible via a catalogue.[17]

27.6 Endnotes

1 See http://unstats.un.org/unsd/geoinfo/UNGEGN/wg2.html, accessed 5 December 2015.
2 See Breu, Josef: Geographisches Namenbuch Österreichs, bearbeitet nach den Empfehlungen der Vereinten Nationen. Wien: Verlag der Österreichischen Akademie der Wissenschaften, 1975 (= Veröffentlichungen des Instituts für Kartographie, Forschungen zur theoretischen Kartographie 3).
3 See http://www.linz.govt.nz/regulatory/place-names/find-place-name/new-zealand-gazetteer-place-names, accessed 5 December 2015.
4 See www.austrianmap.at, accessed 5 December 2015.
5 All this is illustrated in the respective article of the "Dictionary of Upper Austrian Place Names": Wiesinger, Peter (ed.), Ortsnamenbuch des Landes Oberösterreich, vol. 3, 1994, p. 64f.
6 See https://archive.org/details/DictionaryOfSouthernAfricanPlaceNames, accessed 5 December 2015. It should be mentioned though that in the meantime newer editions of this work have been published, cf. Raper, P. E.: New dictionary of South African place names. Johannesburg: Ball, 2004.
7 Ortnamnen i Hallands län. 3 vols. Uppsala: Lundequist, 1948.
8 Stricker, Hans et. al: Liechtensteiner Namenbuch 1: Ortsnamen. 6 vols. Vaduz: Historischer Verein für das Fürstentum Liechtenstein, 1999; for the online version cf. http://historischerverein.li/taetigkeiten/projekte/abgeschlossene-projekte/liechtensteiner-namenbuch, accessed 5 December 2015.
9 Eichler, Ernst (ed.): Atlas altsorbischer Ortsnamentypen. Studien zu toponymischen Arealen des altsorbischen Gebietes im westslawischen Sprachraum. 5 vols. Leipzig: Verlag der Sächsischen Akademie der Wissenschaften, 2000-2004.
10 Cf. International Bibliography of Onomastic Sciences, http://www.icosbibliography.net/cob/, accessed 5 December 2015.
11 See http://www.degruyter.com/databasecontent?dbid=rom&dbsource=%2Fdb%2From, accessed 5 December 2015.
12 Lo Zingarelli 2010. Vocabolario della lingua italiana. 2010 reprint of the 12th ed. Bologna: Zanichelli, 2009, cf. pp. 1388, 1455, 2634ff.
13 Gerhard Helbig / Joachim Buscha: Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht. 10th ed. Leipzig: Verlag Enzyklopädie, 1987, cf. p. 723.
14 Cf. the entry *Duisburg* in the online version of the most popular dictionary of the German language, the "Duden", which also contains an audio file with the Standard German pronunciation of this place name: http://www.duden.de/rechtschreibung/Duisburg, accessed 5 December 2015.
15 Cf. http://www.daredictionary.com/page/questionnaire/#Topography, accessed 5 December 2015.
16 M. Zender (ed.): Atlas der deutschen Volkskunde. Neue Folge. Vol. 1. Marburg an der Lahn, 1958-1979, p. 28.
17 See http://regionalsprache.de/Wenkerbogen/Katalog.aspx, accessed 5 December 2015.

Section 11 Cartographic aspects

Chapter 28 **Updating Geographical Names and assessing the optimal Names density in Topomaps**

Helmut Zierhut

28.1 Introduction

Just as in reality the landscape changes over time, so the graphical map content has to be adapted and updated. Changes in the landscape lead to changes in the geographical names or to new geographical names, and these have to be taken into account and to be corrected in the map.

Depending on the nature of its landscape, an area will contain more or less named features. Depending on the scale of the map it will not always be possible to render the names of all these features on the map. This is because we have to be aware of the fact that each additional name on the map will not only bring information, but will also lose information, as it will also conceal graphical map content.

Chapter 28 will deal with updating the geographical names and acquiring a meaningful names density on modern topographical maps.

28.2 Updating the map content

Simultaneously with the systematic updating of the graphical map contents (in Austria every six years), the geographical names in the map will be checked as well, and be corrected if necessary. This updating will be effectuated by trained topographers of the national mapping agency. All changes, additions and deletions in the toponymy will be submitted for approval to specialists in the authorized geographical names commissions (in Austria these exist for each federal state). If necessary, the submitted names will be corrected by them. A new name can only be inserted in the map after approval by the commission.

The total surface area of Austria is covered by 191 map sheets at the scale 1:50 000 (see figure 28-1). Each map sheet extends for 12`x20` and covers an area of about 520km².

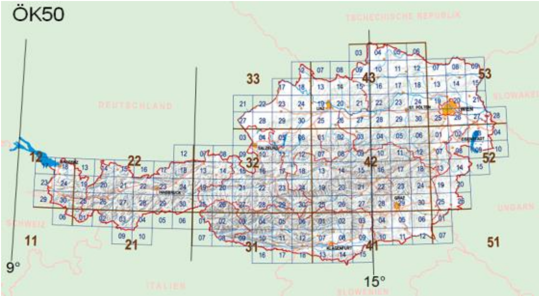


Figure 28-1 Coverage of Austria by the 1:50 000 topographic map series.

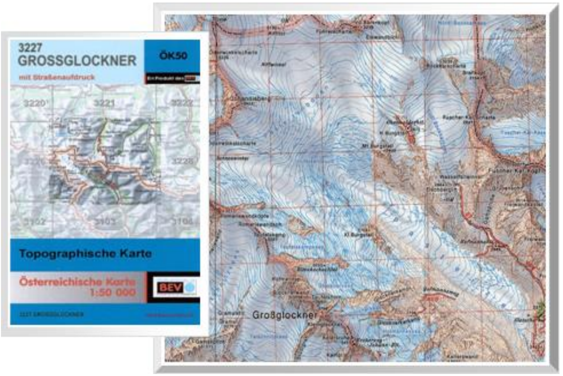


Figure 28-2 Title page of the 1:50 000 Austrian topomaps

The title page of each map sheet (see figure 28-2) shows an overview of the extent covered by that sheet with parts of the adjoining sheets.

28.3 The optimal names density

Depending on the nature and structure of the landscape – whether it is a plain, low or high mountain range, whether it is a settles or non-settled area - the number of named features will vary as well as the number of named features worth to be inserted in the map. This is because we have to be aware of the fact that each additional name on the map will not only bring information, but will also lose information, as it will also conceal graphical map content.

A topographical map should describe the important terrain features and assist map users to orient themselves in the terrain.

To allow for that purpose, internal workbooks of the Austrian national mapping agency require that each map sheet should contain at least 700 geographical names, and at most 1000.

28.4 Categories of Geographical Names

A distinction is being made between settlement names and non-settlement names.

28.4.1 Settlement names

A settlement name, or place name proper or oikonym, (from Greek οἶκος ‘house‘), is the name of a settlement. This could be a town (like Zell am See in figure 28-3) or a village, a single farm, a hamlet or a dispersed settlement (like Schmitten in figure 28-3).

Figure 28-3 Types of settlements

Usually, for the orthography of a settlement name, a local authority (e.g. a municipality) is responsible. For the correct writing of the names of single objects like private farms the owner would be responsible.

28.4.2 Non-settlement names

Non-settlement names generally describe the general topographical features of the Earth surface. These could be extensive areas like mountain ranges, valleys, or administrative areas but also point features like mountain summits, mountain passes or saddles.

28.4.3 Type sizes and type faces to be used on topographic maps on the scale 1:50 000

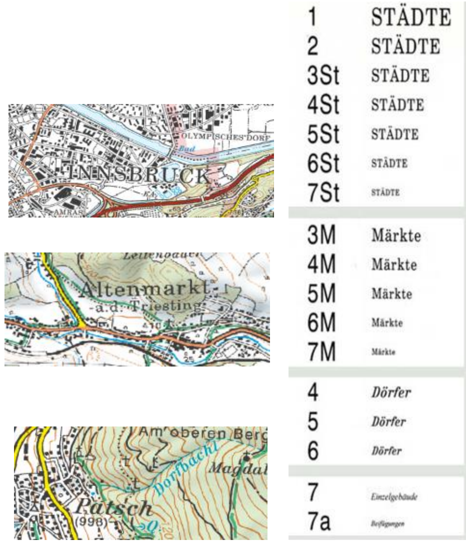
The examples in figure 28-4 show the diversity in type sizes and type faces in use for the Austrian 1:50 000 map series.

Type size and type face will vary depending on the structure, size, number of inhabitants and administrative status of the settlement.

Towns,

market towns,

Villages,



Individual buildings and extensions

Figure 28-4 Types of settlements and matching type sizes and faces

A locality or area without a place of residence will be indicated on the map by non-settlement names, such as names of mountain ranges or mountain tops, hydrographical names or landscape names. The examples in figure 28-5 show the variety in type sizes and type faces in use for indicating the names of mountain ranges and summits, landscapes and rivers.

areas

hydro graphy

ranges,

summits

Figure 28-5 Non-settlement features and the representation of their names on the Austrian 1:50 000 map

28.4.4 Geographical Names in areas with minority languages

In some Austrian areas, next to the German language, Croat, Slovene and Hungarian also have official status. According to the appropriate laws and regulations, nameplates for settlements and streets have to be rendered bilingually, in German and the respective minority language. In Austria, these bilingual areas can be found around and south of Eisenstadt and around Klagenfurt.

Figure 28-6 – Minority language areas in Austria

Figure 28-7 Slovene names rendered in parentheses after the German names

In the Austrian topographic maps 1:50 000 und 1:250 000, officially recognised minority names are rendered in parentheses, after their German language name versions, as can be seen in figure 28-7.

28.5 Abbreviations

As mentioned in section 28.3, an optimal names density is required in order to safeguard map reading and interpretation possibilities.

Figure 28-8: list of abbreviations used

Bhf., -bhf.	Bahnhof, -bahnhof
Fh.	Forsthaus
Kls., -kls.	Kloster, -kloster
Q.	Quelle

Generic concepts that occur frequently, like railway station, foresters' residences, monasteries, wells, etc. can therefore be abbreviated and thus have a less strong impact on the map image.

In order to ease the use of abbreviations, as well as to standardize them, a list of abbreviations is often included in the map margin (see figure 28-8).

28.6 Geographical Names in adjoining foreign countries

Settlements beyond the state border, that used to have close relationships with the German-speaking population, and of which the German name versions still are well-known and used in Austria, would be rendered by both their official foreign name and by their German name (in parentheses) on official Austrian topographic maps 1:50 000, 1:200 000, 1:250 000 und 1:500 000. This is shown in figure 28-9, where settlement names in adjoining Czechia are rendered bilingually.

Figure 28-9 Austrian Czech border area as rendered on Austrian maps 1:50 000

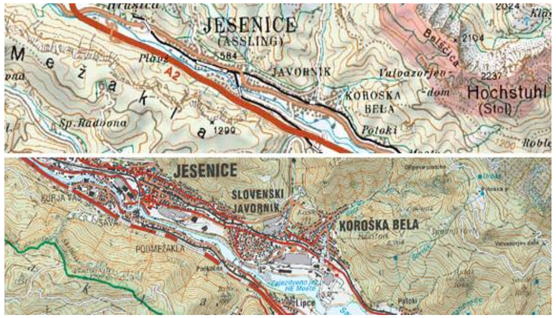


Figure 28-10 Top: Austrian map 1:200 000 with the Slovene settlement Jesenice (Assling). Below: Slovenian map 1:50 000 without German name variants

In the new civilian-military Austrian topographic maps, the cartographic contents of the neighbouring states rendered on them are no longer processed, drawn and updated by Austria. Instead, the map content for these foreign areas is based on the updated original databases of these adjoining states. As can be seen in figure 28-10, by comparing the two maps, these foreign cartographic databases do not contain German-language name variants for settlement names.

I hope that with this information about cartographic procedures in Austrian official mapping I have provided some incentives for the production and processing of cartographic products.

28.7 References

-Bundesamt für Eich- und Vermessungswesen, Wien (2000). Technische Weisungen zum Zeichenschlüssel, (interner Arbeitsbehelf für Topographen und Kartographen)

-Bundesamt für Eich- und Vermessungswesen, Wien (2007). QM-Verfahrensanweisungen, Österreichische Karte 1:50 000, 1:250 000. Aktualisierung der Kartographischen Modelle KM50, KM250.

Section 11: Cartographic aspects

Chapter 29 **Dealing with areal names on adjoining map sheets; multiple naming**

Helmut Zierhut

29.1 Introduction

Depending on the map scale, every topographic map only portrays a limited part of the Earth surface. Although the objects rendered on the map have a limited extent as well, it will frequently be the case that they surpass the map margins and continue on the next map sheet. This will be the case especially for features with a larger extent such as administrative areas, mountain ranges or valleys, but larger lakes and rivers might require more map sheets as well for their portrayal. Consequently, they should be named on each of these sheets. The following sections will show how to go about map lettering close to the map margins.

29.2 Technical methods of map lettering

29.2.1 Analogue map production

Until the end of the 20th century, maps were produced using analogue techniques, and the unit of production always was one single map sheet. The cartographer who had to effectuate the map lettering, tried to do it in such a way that the extent of a feature on the map sheet would be visualized optimally by the size, spacing and extent of the lettering within that map sheet. But this could result on neighbouring map sheets in map names in sizes that did not reflect the actual extent of the feature to be portrayed in reality.

If only a small part of the feature would be located on one of the sheets that were to be prepared, then the size, spacing and extent of the name would only reflect the (clipped) area of the feature portrayed on that particular map sheet. The true size of the feature could not be deduced from the lettering only.

29.2.2 Digital Map production

When digital map production techniques were introduced, cartographers updating the map contents were no longer bound to specific sheet lines. The new data would be incorporated in a „cartographic model" without any sheet lines or subdivisions. When providing the map files for printing, any sheet lines could be chosen that were thought appropriate.

So, cartographers, when lettering the maps, no longer took account of the sheet lines that would define the mapped area in the final printed map product. The geographical names of physical or administrative areas or mountain would be located on the map in such a way that through their size and spacing they would optimally visualize the extent of the mapped feature. Repeating the name of the object in the "cartographic model" no longer was necessary.

29.2.3 Processing names for map sheet production

When a map sheet is produced from a „Cartographic model" without sheet lines, then some names may be cut off , to continue on the adjoining map sheet. If the map user would only dispose over one of the two adjoining map sheets, then he would only be able to perceive part of the object name, and thus be unable to assess or even guess the complete name of the feature. This impossibility forces the cartographer to add to every map sheet the missing part of the cut-off names in the map border.

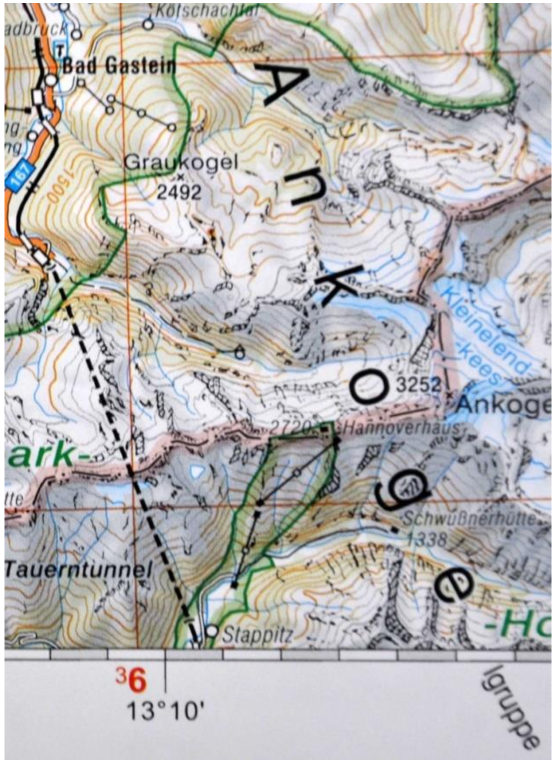


Figure 29-1 Map lettering running out of the map into the border

In figure 29-1 one can see the example of how part of the name „Ankogelgruppe" has been delegated to the map border, in a smaller type size.

The cut-off and therefore missing part of each geographical name will be continued in the map border in a smaller type size, when editing the map border.

A similar procedure is chosen when the first part of a geographical name is cut off, and it is only the remaining part of the name that is incorporated in the map sheet. In figure 29-2 an example is given of the way in which the name „Samnaungruppe“ is being dealt with. Here, the first part of the geographical name is incorporated in a smaller type size in the map border. As this mountain range has a southwest-northeasterly direction, the name

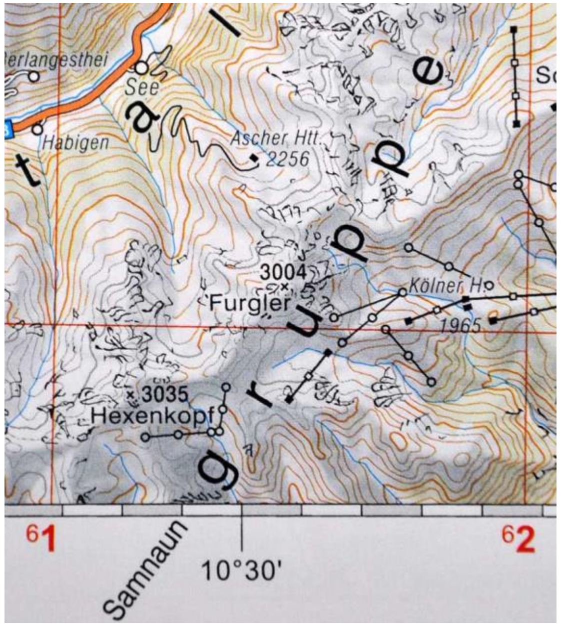


Figure 29-2 Map lettering running into the map from the border

would start in the adjoining map sheet south of the present one. This method of completing the name in the map border is called ‘lettering running into a map from the border’ or ‘lettering running out of the map into the border’.

Figure 29-3 shows, on the example of the name „Weinsberger Wald“ the possibility that one and the same map name can run both in and out of the map border.

Figure 29-3 Weinsberger Wald

29.3 Map sheet lines with overlap

The map sheet lines on Austrian topographic map series are in principal formed by meridians and parallels. However, every map sheet extends over these sheet lines, by an overlap zone of about 1 km wide.

In figure 29-4 the sheet lines of *15° East of Greenwich* and *48°24´ have been designated by* purple lines. As can be seen clearly, the printed map sheet extends over these sheet lines and shows a considerably larger area.

Figure 29-4 Relation between the sheet lines and the map border, resulting in an overlap with the adjoining sheets.

The reason for opting for this overlap is the following:

In the first place, this overlap zone simplifies the use of the map close to the map border. In most cases it won’t necessitate the map user to switch from one map to the adjoining map.

Secondly, having an overlap zone also eases the cartographer’s task. Although geographical names will still be transected by sheet lines, only a very few will be transected by the map margins. The few names for areas that extend beyond both the sheet lines and the map margin will be provided with lettering running into or out of the map border, as shown above.

I hope that with this information about cartographic procedures in Austrian official mapping I have provided some incentives for the production and processing of cartographic products.

29.4 References

-Bundesamt für Eich- und Vermessungswesen, Wien (2000). Technische Weisungen zum Zeichenschlüssel, (internal resource for Topographers and Cartographers)

-Bundesamt für Eich- und Vermessungswesen, Wien (2007). QM-Verfahrensanweisungen, Österreichische Karte 1:50 000, 1:250 000. Aktualisierung der Kartographischen Modelle KM50, KM250.

Chapter 30 **The Canadian Geographical Names Data Base – an example of a national system combining data from different jurisdictions**

Helen Kerfoot[1], Kristina Kwiatkowski[2], Heather Ross[3]

30.1 Introduction

Where is *Flin Flon*? Are *St. Lawrence River* and *fleuve Saint-Laurent* both officially recognized names? What was the formerly approved name for *Iqaluit*? In Canada, geographical names constitute a significant part of culture, heritage and identity (Figure 30-1). They are formalized through Canada's national naming authority, the Geographical Names Board of Canada (GNBC), and stored in the national toponymic database, the Canadian Geographical Names Data Base (CGNDB).

The need for toponymic databases

Today toponyms can be gathered into municipal, provincial, national and regional databases to make them accessible, often through Internet-based Spatial Data Infrastructures, to the widest possible audience. A toponymic database can serve many purposes, particularly if linked to other spatial data, as a name is an intuitive entry point to search for associated information.

Figure 30-1 Road sign in New Brunswick, including a bilingual (French/English) town name; Village signage in Saskatchewan; Street signs in Lytton, British Columbia, showing English and Nłeʔkepmxcín (or Thompson) language

In addressing data needs to analyze complex physical and cultural associations, our focus must be on the toponymic data itself, its attributes, storage, and accessibility so that correct geographical names are widely available. For a national toponymic database to be of optimum use, it is necessary to consider such questions as: what geographical names and what attributes should be included; how will data derived from different sources be pulled together; how will data quality (e.g. accuracy, consistency across the country, completeness) be planned and achieved; how will the records be kept up to date (and what does that mean); how will the toponymic data be accessed and linked to other spatial data.

Distributed databases

In some countries, the authority for approving geographical names lies not with a single national committee or names board, but with boards/authorities at the first level administrative unit (e.g. province, state). In these cases, a national database must take into account the harmonization of records from the different authorities, the process(es) for updating the data, and a framework

[1] Former Emeritus Scientist, Natural Resources Canada; Chair, United Nations Group of Experts on Geographical Names, 2002-2012

[2] Toponymy Specialist, Natural Resources Canada, Geographical Names Board of Canada Secretariat

[3] Toponymist, Natural Resources Canada, Geographical Names Board of Canada Secretariat

for expanding, modernizing and rationalizing the system, as needed.

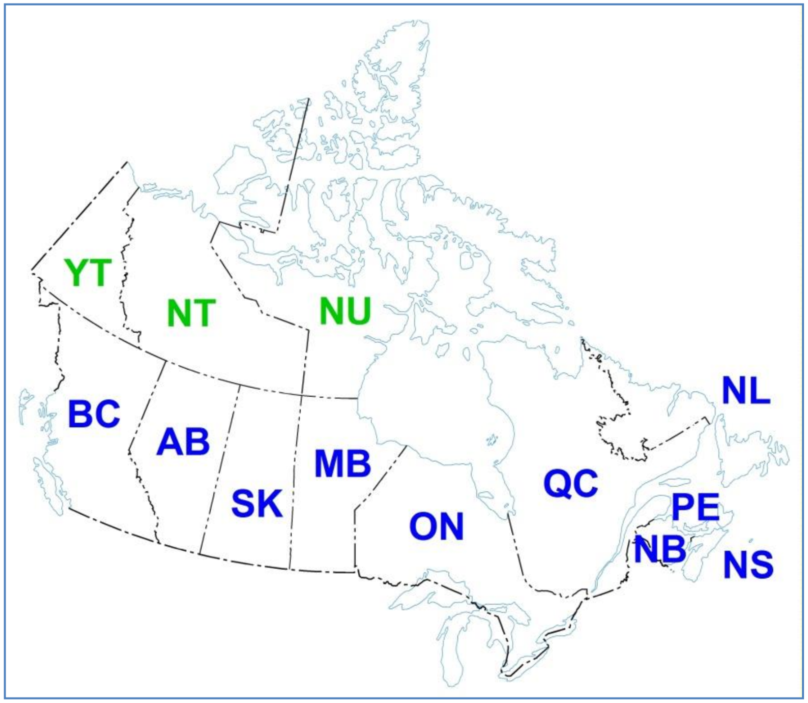


Figure 30-2 Provinces and territories of Canada

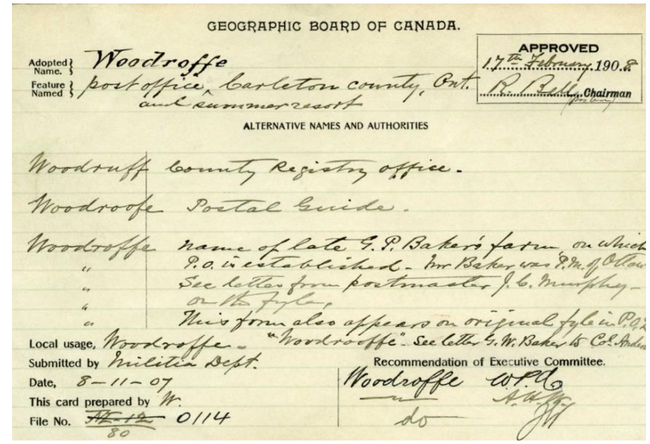
- *Provinces (blue): BC British Columbia; AB Alberta; SK Saskatchewan; MB Manitoba; ON Ontario; QC Quebec; NB New Brunswick; PE Prince Edward Island; NS Nova Scotia; NL Newfoundland and Labrador.*
- *Territories (green): YT Yukon; NT Northwest Territories; NU Nunavut*

As an example of a national toponymic database that includes data from various sources, we will elaborate on the development of the Canadian Geographical Names Data Base (CGNDB) in which toponyms from Canada's 10 provinces and 3 territories (Figure 30-2) are combined, updated regularly and made generally available through the Internet. The CGNDB, unlike some national databases, is not created for a particular map scale, but is the authoritative database of Canada's toponyms for use by governments, industry, academia and the public.

30.2 Early records of Canada's national names authority

The Geographic Board of Canada was initially created by an Order in Council of the Government of Canada in 1897. At that time the Board Secretariat in Ottawa started keeping card records for the names approved by the Board for places and features across the country. Within a year, provincial representatives were included in the Board to provide advice on name decisions. Not until 1912 did a province (Quebec) create its own names board.

Although the early cards were fairly consistently compiled, the layout of the card, the amount of data recorded and the legibility varied considerably over time (Figures 30-3 and 30-4). However, the cards were a suitable source of data for national names lists, regular reports of officially recognized toponyms, and for compilation of gazetteer volumes for each individual province and territory.


GEOGRAPHIC BOARD OF CANADA.

Adopted Name. Woodroffe

Feature Named post office, Carleton county, Ont. and summer resort

APPROVED 17th February 1908
R. Bell Chairman

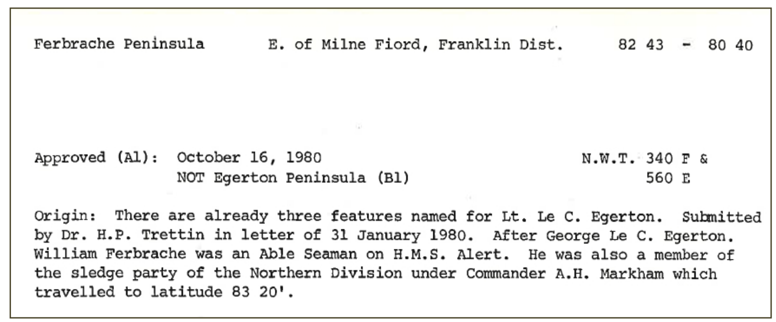
ALTERNATIVE NAMES AND AUTHORITIES

Woodruff County Registry office.
Woodroofe Postal Guide.
Woodroffe name of late G. P. Baker's farm on which P.O. is established. Mr Baker was P.M. of Ottawa
See letter from postmaster J. C. Murphy on the fyle.
This form also appears on original fyle in P.O.

Local usage, Woodroffe. "Woodroffe". See letter G. W. Baker to Col. Anderson
Submitted by Militia Dept.
Date, 8-11-07
This card prepared by W.
File No. 0114

Recommendation of Executive Committee.
Woodroffe

Figure 30-3 Example of a 1908 hand-written card from Board records (note: no coordinates were included in these early records)


Ferbrache Peninsula E. of Milne Fiord, Franklin Dist. 82 43 - 80 40

Approved (A1): October 16, 1980 N.W.T. 340 F &
NOT Egerton Peninsula (B1) 560 E

Origin: There are already three features named for Lt. Le C. Egerton. Submitted by Dr. H.P. Trettin in letter of 31 January 1980. After George Le C. Egerton. William Ferbrache was an Able Seaman on H.M.S. Alert. He was also a member of the sledge party of the Northern Division under Commander A.H. Markham which travelled to latitude 83 20'.

Figure 30-4 Example of a 1980 card created before the national database became fully digital

Changes in jurisdictional responsibility for geographical naming

By the 1960s, the provinces had all taken over the authority for making the decisions on geographical names within their own jurisdictions, and in the 1970s, the territories also took on these decision-making responsibilities. However, card records, based on decisions from the jurisdictions, were still kept for the whole country by the Board Secretariat.

Today, representatives of provincial, territorial, and federal naming authorities, together with a Chairperson and advisors, constitute the Geographical Names Board of Canada (GNBC), the national coordinating body responsible for geographical naming activities in Canada. The GNBC members supply their toponymic decisions to the national Canadian Geographical Names Data Base maintained by Natural Resources Canada (NRCan), a department of the Government of Canada. Figures 30-5 and 30-6 compare the composition of the Board in its early years with 2016.

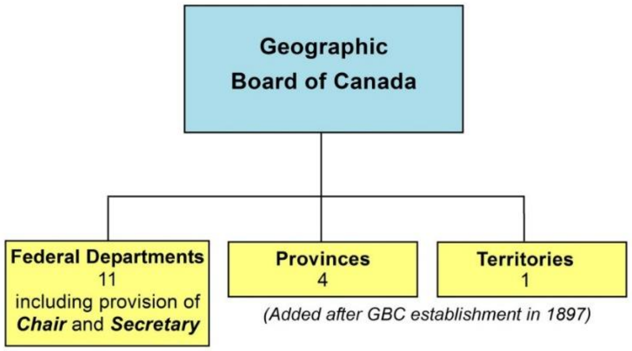


Figure 30-5 Composition of the Geographic Board of Canada, 1900

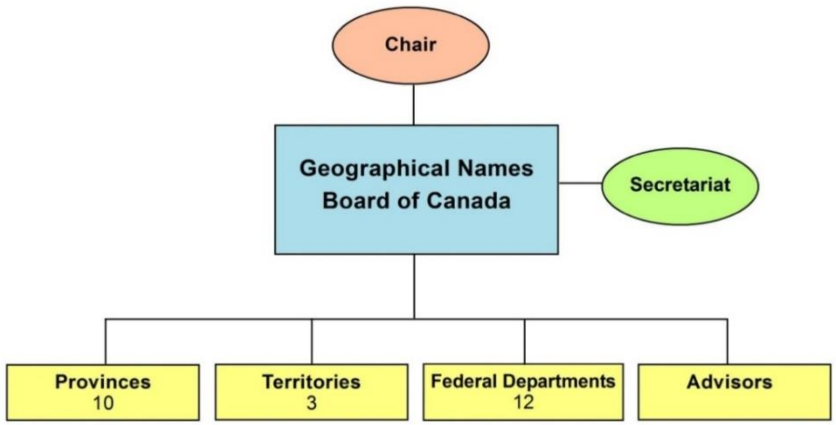


Figure 30-6 Composition of the Geographical Names Board of Canada, 2016

30.3 Establishing Canada's national geographical names database

The late 1970s saw the development of a digital database from the central card records, to increase efficiency of gazetteer production and names compilation for federal maps. Discussions led to a structural framework more standardized than the rather free-wheeling presentation of material on the card records. Among the most important steps were the development of suitable data fields (e.g. name, feature type, coordinates, etc.); the creation of status codes (for example, "A" status categories for official names, "B" status categories for unofficial names); the coordination of sets of generics / feature type designations that suited each jurisdiction and Canada as a whole. A set of "core fields"[4], necessary for all name records, as well as optional data fields (e.g. unofficial variant names; historical/origin data) were established for consistency across the country's toponymic records.

In the early years of the national database, provinces and territories did not have their own

---

[4] Core fields were considered as: geographical name; province/territory; status of name; a cross-reference to an official name, if a name had changed; date of decision; type of feature/place; latitude and longitude; National Topographic System map at 1: 50,000 scale; sub-unit of province/territory; narrative of location; graphic representation showing limits of named feature/place.

databases. They provided paper decisions in various formats, containing the basic information on new names, changed names, and rescinded names. The data was entered by Secretariat staff, who consulted the data providers whenever there were questions. Federal departments and agencies responsible for administrative entities (such as national parks and military reserves) also worked with the Secretariat to add their names records to the CGNDB. As well, the Canadian Hydrographic Service contributed undersea feature names to the national data set.

By 1982 the national database was operational, containing some 350,000 approved names and over 100,000 unofficial names.

Database users' manual

Since the 1980s a detailed users' manual for the CGNDB has been compiled and distributed to those who supply data to the national database. As part of the manual's content, lists of codes, with their associated terms and definitions, were reviewed and approved by all GNBC provincial, territorial and federal contributors of geographical names data. To establish standards, the code lists in the manual provided definitions of field contents, and how the data should be entered. New codes were created as required, and updates of the manual have been distributed regularly (see Figure 30-7).

As technology developed, some provinces and territories created their own names databases to meet their specific needs. For some, this was for a mapping program, for others to store cultural and historical information on names. However, all agreed to feed their data into the national database, the CGNDB.

TITLE	SECTION
Name information	***1.0***
Feature name	**1.1**
Name key	**1.2**
Cross-reference	**1.3**
Related names	**1.4**
Accented characters	**1.5**
Names with symbols or numbers	**1.6**
Modified extended Roman alphabet characters	**1.7**
Syllabics	**1.8**
Datum	**1.9**
Record identifiers	***2.0***
CGNDB key	**2.1**
Feature identifier	**2.2**
Map Information	***3.0***
Gazetteer map	**3.1**
Map sheet information	**3.2**

TITLE	SECTION
Location information	***4.0***
Region code	**4.1**
Coordinates	**4.2**
Location information fields	**4.3**
Geographic location	**4.4**
Administrative location	**4.5**
Table of location information	**4.6**
Location narrative	**4.7**
Headwaters information	**4.8**
UTM grid	**4.9**
Datum	**4.10**
Attribute information	***5.0***
Status code	**5.1**
Obscure	**5.2**
Decision date	**5.3**
Meeting date	**5.4**
CGNDB update	**5.5**
Boundary flag	**5.6**
National Park code	**5.7**
Generic code	**5.8**

TITLE	SECTION
Appendix	***6.0***
Terminology	**6.1**
Index	**6.2**
Data input sheet	**6.3**
Sample CGNDB screen	**6.4**
Core data fields	**6.5**

Figure 30-7 Core content for the CGNDB Users' Manual

Key attributes (fields) used in geographical names records

In order to facilitate data transfers, and matching of provincial and national records, each record is given a unique identifier when it is created. This five-letter code, called the CGNDB key, identifies the record for its entire lifespan, whatever changes occur to the data within the record. Names records continue to be stored based on province or territory, and the first letter of this unique identifier indicates the jurisdiction. Additionally, a code identifying the province or territory is included in each name record.

In each record, feature type is identified by a four-character numerical code, called the generic code (see Figure 30-8). There are over 1000 generic codes, with their associated generic terms, which fall into 11 categories, such as Populated Places or Water Features. Some categories are also divided into sub-categories. The Water Features category has seven sub-categories, including Flowing

Freshwater, Standing Water, and Tidal Water (Figure 30-9).

Field Name	Example: Ottawa
CGNDB Key	FEOLW
Related Key	FEGRN
Generic Code	0121
Feature Identifier	5997e90fc6ce11d892e2080020a0f4c9

Figure 30-8 Examples of some CGNDB codes

Origin information exists for many records across Canada, but not for all. A wide variety of information is contained in this field: administrative detail such as approval dates, incorporation dates, or files where decision information is located; historical information or the reason for naming.

Latitude and longitude for a name record were originally stored in degrees and minutes. As mapping at larger scales extended across the country, allowing more precision, coordinates were upgraded to degrees, minutes and seconds. In contrast to national databases in some countries, coordinate values in the CGNDB represent the location of the named feature on the ground, not necessarily the positioning of text on particular map scales.

Extract of *CGNDB Users' Manual*

WATER FEATURES	GENERIC CODES
Flowing freshwater	0600 - 0653
Features on flowing water	0700 - 0799
Standing water surrounded by land	0951 - 0997
Water sources	1150 - 1158
Standing water connected to two or more bodies of water	1200 - 1250
Tidal water features	1352 - 1358
Shoreline water features	1400 - 1465

N.B. Numbers in ranges indicate generic codes which currently exist. Space remains in each category to add new generics if required.

Figure 30-9 Generic codes showing ranges for generics in use for water feature toponyms, which constitute more than 50% of the CGNDB records

Currently, to accommodate the varied needs of data users, latitude and longitude are displayed in online query results in both degrees, minutes and seconds (DMS) and decimal degrees. Over 90% of DMS coordinates are precise to the second (exceptions are mainly for very large features, where such precision is of limited value). To ensure point coordinate accuracy and consistency, CGNDB data conforms to the NAD 83 geodetic datum. Today most data providers use NAD 83, but for those who do not the data is converted to this datum before being loaded into the CGNDB.

Canadian toponyms on the web

In 1994, the Geographical Names of Canada website was launched. The name query tool allowed users to query the national database for all official names, as well as formerly official names across the country. Only the core fields were available, and there was a limit on the number of allowable records which could be returned per query. The name query quickly became very popular and was widely used by government, business, and the educational sector. National and regional files were made available, greatly increasing the number of clients, and enabling users to update their data more frequently.

30.4 Current status of the CGNDB

Improved data model

In 2013, a need was identified to update the CGNDB data model to comply with broader data management requirements, as well as a need for individual jurisdictions to upload their data directly into the national database. By 2015, the data model for the CGNDB evolved from an attribute-based model to a geospatial-based model. The new data model was designed using a relational ISO standard model and vastly improves the functionality and interoperability of the national database. The previous data model contained one large table with over 90 fields, whereas, the new data model contains 45 tables, and is capable of handling spatial and relational data, so enabling relationships between toponyms and spatial delineations of the named features. There are currently over 133,000 spatial delineations contained in the CGNDB, predominantly for hydrographic features such as rivers and lakes (Figure 30-10). The new model can handle multiple formats such as decision documents, shapefile delineations, and sound files. It also enables better support for data validation, database monitoring and statistical reporting. The new data model also better serves the needs of the provincial and territorial jurisdictions of the GNBC; the new data model supports toponym data entry and editing by each jurisdiction.

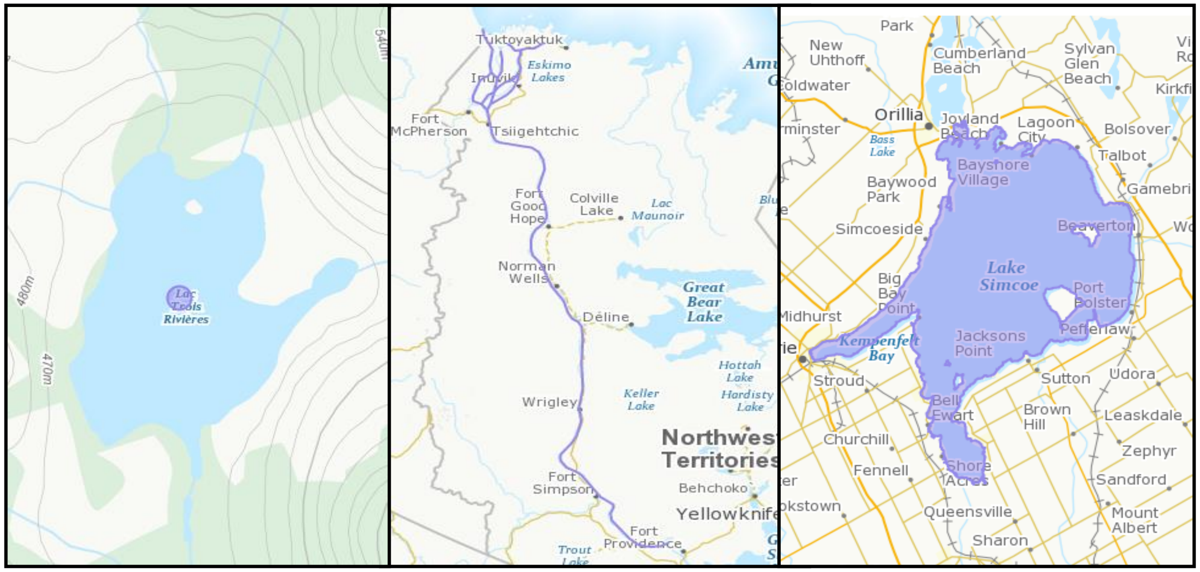


Figure 30-10 Example of spatial delineations for named features: point, line, polygon

Feature identifiers

Key to the transformation of the database from attribute-based to geospatial-based was the inclusion of Feature Identifiers (FIDs). FIDs uniquely identify each named feature; they are implemented as a Universal Unique Identifier containing 32 alphanumeric characters. FIDs remain associated with a feature regardless of any future name changes[5]. FIDs allow for retrieval of all the names with the same spatial extent. The FID is auto-generated for new entries into the CGNDB and is used throughout NRCan for geospatial work flows. In general, if there is a significant change in the spatial extent of the feature, a new FID will be assigned. For example, if a lake were to become two separate lakes, two new FIDs would be assigned, one to each new lake. Criteria are currently being developed to define standard procedures to handle more complex cases, such as successive municipal amalgamations which involve multiple changes to polygons, although the names remain the same.

In 2016, on behalf of the provinces, territories, and federal members of the GNBC, NRCan completed the work to add FIDs to all toponyms in the CGNDB. The process was semi-automated. For example, a related official and former official name that have the same generic definition were automatically assigned the same FID. To handle records that did not meet the automatic assignment requirements, a GNBC working group discussed various scenarios and created a set of FID assignment rules. A thorough examination of the historical relationships and spatial extent of the records was carried out by NRCan's geospatial technicians in consultation with the provincial and territorial naming authorities to ensure that the FIDs were correctly assigned.

Geographical Names Web Application (GNApp-II)

---

[5] Prior to the use of FIDs, geographical name relationships were handled using a related key (i.e. the unique key of the related name record). The Related Key follows the "history of the name", meaning it follows the various name changes of the feature over time, but does not necessarily track the exact spatial extent of the named place.

NRCan has also updated a web-based application to support the development of the new data model and facilitate queries and edits to the CGNDB. The new application was launched in February 2016 and supports improved interaction of the GNBC members with the national database (Figures 30-11 and 30-12). The new application was developed with the input of GNBC members through extensive requirements gathering and usability testing. The application offers an improved display and searching functionality, as well as a map visualizer. The GNBC naming authorities can now attach name decisions to database records, as well as upload spatial delineations in shapefile format.

By giving the naming authorities of the GNBC direct access to the national database through GNApp-II, NRCan can ensure that the most up to date and accurate names information is contained in the CGNDB. For those provinces and territories that maintain their own jurisdictional databases, an improved batch upload process is being developed. This will involve the jurisdiction submitting the updated names from their database into GNApp-II using a standardized template to ensure efficient data entry into the CGNDB. Currently, jurisdictions that have not adopted the use of GNApp-II submit their new name decisions in batch format, typically through a spreadsheet. The GNBC Secretariat and NRCan's database team clear the data through a validation procedure before adding the records to the CGNDB. Records that do not clear the validation are examined and discussed with the jurisdiction.

Some GNBC naming jurisdictions have their own application programming interface (API) service which allows NRCan's database team to easily access jurisdictional records and update them in the CGNDB. To improve the currency of the CGNDB and its compatibility with jurisdictional databases, NRCan will in future investigate developing an application to query jurisdictional APIs and alert the database team of any updates. In this way data could be fetched as it becomes available.

The data flow process from the various sources into the CGNDB is shown in Figure 30-13.

Figure 30-11 Screen shot of GNApp-II data entry options

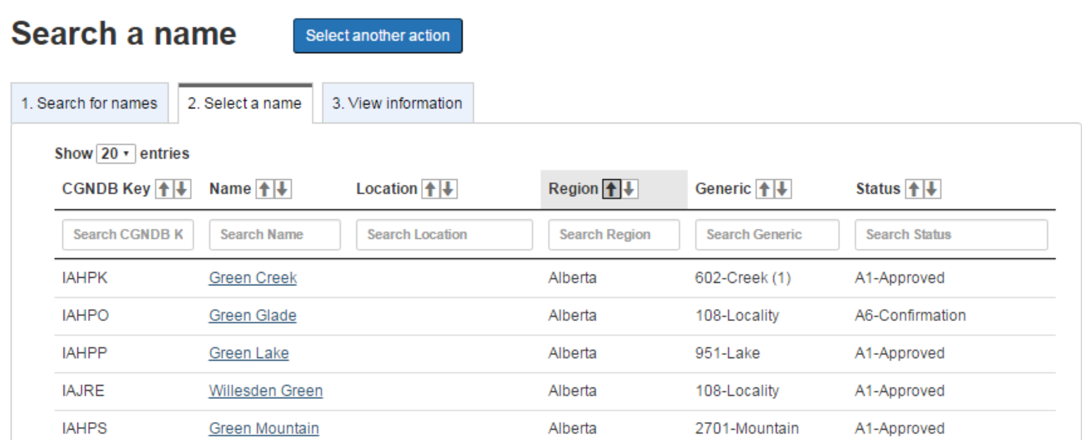


Figure 30-12 Example of results of "Search a name"

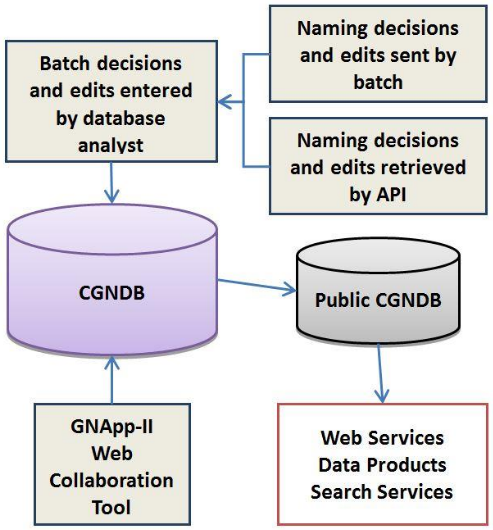


Figure 30-13 Data flow process

For jurisdictions with their own databases, the GNBC Secretariat performs periodic data reconciliations. This is undertaken when the jurisdiction sends the Secretariat a copy of their database and the records are compared with those in the CGNDB. It is a systematic way to update the CGNDB and ensure that the national database contains the same information as provincial and territorial databases.

Some challenges faced with multiple sources of data

Issues that may be encountered when dealing with geographical names data from multiple jurisdictions include:

- standardization requirements
- handling of features that span jurisdictional boundaries
- quality of the data
- frequency of data updates

As mentioned earlier, each Canadian jurisdiction has different needs and requirements for their names database, and some rely solely on the national CGNDB as their database. Some jurisdictions have wider mandates than others, for example responsibility for street names or tourist route names. To maintain consistency and standardization across all jurisdictions, only records of agreed upon feature types can be entered into the national database.

Canada contains many geographical features that cross provincial and territorial boundaries. When a feature crosses a boundary, both jurisdictions involved will make a decision on the name of the feature, and whenever possible, try to reach an agreement so that the feature has the same name in both jurisdictions. These cross-jurisdictional features are assigned an attribute called a border flag to notify the user that a feature crosses into an adjacent jurisdiction. For every province or territory in which the feature is located, there is a separate record in the database. However, each record will show the same feature identifier.

Another issue faced with combining data from multiple jurisdictions is the varying quality of the data. Validation rules are established and data reconciliations catch the obvious errors. Nevertheless, in order to have consistent and standardized data at the national level, clear data entry guidelines must be in place and cooperative efforts must be undertaken to resolve inconsistencies between the national and jurisdictional databases, as they arise.

The frequency of updating records in the CGNDB is often reviewed to meet user needs for data currency. Collaboration is required between the names jurisdictions, the GNBC Secretariat, and NRCan's database team to aim for timely and consistently updated CGNDB data, and to ensure that the updated data is made available in the public database in a timely manner.

Jurisdiction	Place Name	Syllabic Form	ISO 639-3 Language
Nunavut	Nanuqsiutitalik	ᓇᓄᖅᓯᐅᑎᑕᓕᒃ	Inuktitut-Eastern Canadian
Northwest Territories	Behchokǫ̀		Dogrib
Yukon	Nu Détsyáwa Mǟn		Northern Tutchone
Nunavut	Aggijjat	ᐊᒡᒋᔾᔭᑦ	Inuktitut-Eastern Canadian
Northwest Territories	Igyųų Yik Thidye'		Kutchin-Gwich'in (Loucheux)
Yukon	Nį̀į'į̀į' Lake		Upper Tanana

Figure 30-14 Names of Indigenous languages, showing examples of extended Roman alphabet characters and Inuktitut syllabics

Indigenous languages in the Canadian Geographical Names Data Base

The CGNDB enables naming authorities of the GNBC to indicate the language of the toponym. The language may be defined from a standardized ISO list that contains 74 languages relevant to Canada's Indigenous Peoples, as well as English and French. All provinces and territories have official name records with Indigenous origins. Most are written in Roman alphabet characters that are consistent with English and French. However, the use of the UTF-8 standard encoding in the CGNDB allows representation of special characters of the extended Roman alphabet used in geographical names in Canada (for example, ầ , ậ', used in toponyms in the Yukon).

If Inuktitut (an official language of Nunavut) is selected as the language of the toponym in GNApp-II, the application will automatically convert between the Romanized form of the name and the name written in Inuktitut syllabics. Figure 30-14 highlights some examples of the special characters that the CGNDB can handle.

Publicly available data from the CGNDB

Geographical names in Canada can be accessed using a web-based search tool and through downloadable data products supported by NRCan. The data contains official names of geographical features, including populated places and undersea features.

Web-based toponymic queries can be based on the name, feature type, province/territory, coordinates, rectangular area, or unique identifiers. Users may also search for names containing characters particular to Indigenous languages of Canada (Inuktitut syllabics or extended Roman alphabet characters). A query returns the feature type, region, unique identifiers of both the name and the feature, latitude and longitude, the date when the name was approved or changed status, and (if available) a spatial delineation of the name's application overlaid on a base map in a web map viewer (Figure 30-15). CGNDB records can be accessed at: http://www.nrcan.gc.ca/earth-sciences/geography/place-names/search/9170

An updated and standardized suite of geographical names digital products can be downloaded in Shapefile, KML, GML, and CSV formats from NRCan's GeoGratis portal (http://geogratis.gc.ca/) as well as from the Government of Canada Open Data Portal (http://open.canada.ca/en/open-data).

In addition, NRCan offers an Application Programming Interface (API) as a means of public access to the CGNDB for customized searches.

In some instances the feature boundary may not align with the base map due to the scale and datum at which the feature was collected.

Name	**Sirmilik National Park of Canada**
Language	English
Other official name	Parc national du Canada Sirmilik
Key	OAUQP
Status	Official
Feature Type	Conservation Area
Feature Generic	National Park
Location	
Province/Territory	Nunavut
Latitude - Longitude (DMS)	72° 50' 0" N, 80° 35' 0" W
Latitude - Longitude (decimal)	72.833333, -80.583333
Map	048A16
Relevance at Scale	1: 5 000 000
Toponymic Feature ID	8fde721fba3611d892e2080020a0f4c9
Decision Date	2001-01-12
Source	Parks Canada

Figure 30-15 Result of a public web search for Sirmilik National Park of Canada

Some uses of the CGNDB

When the database was first created, it was invaluable for the production of printed gazetteers and for map production, streamlining the creation of both products. For maps, a names list which previously was made manually from earlier editions could be created automatically, saving considerable time. Conversely, gazetteer and map publishing were important drivers for quality control of database records.

Today the CGNDB serves a broad set of needs and the national data is included in many applications (e.g. GPS and smart phone applications). Museum curators labelling specimens collected by scientists use the name query to verify place names used in field reports, and to locate the sites. Couriers and trucking companies create lists of populated places to find locations not on road maps. Students and teachers use the geographical names data and a radius tool in combination with educational modules to improve their knowledge of geography and history, and to research the connections between place names, culture and heritage. Figure 30-16 shows some uses of CGNDB data.

The CGNDB continues to be a valuable resource for a wide variety of research. Genealogists and historians use it to find names mentioned in historic texts or family documents. It is possible to find the

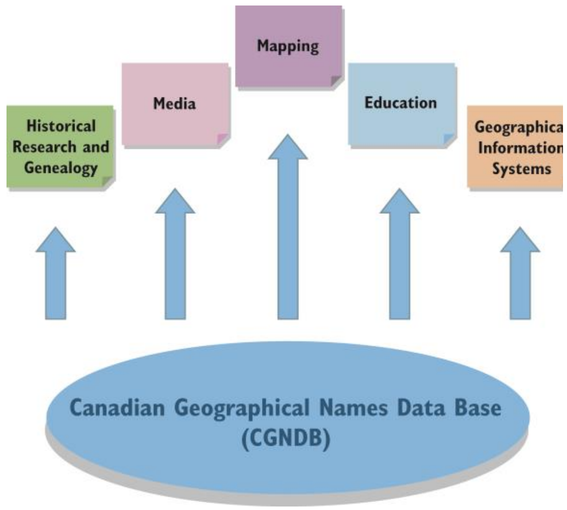


Figure 30-16 Some of the many uses for CGNDB data

longest, shortest or most common names in a province/territory, region, or all of Canada, as well as extracting data for names associated with specific feature types. The CGNDB is used for data analysis, and the collection of statistics. For example, the GNBC has studied the density of official names across the country to identify gaps in fieldwork, and to suggest possible areas for future research.

Although the CGNDB is not designed solely for map production, government maps at various scales use official names from the CGNDB (Figure 30-17). Geographical names data also enhances the value of other datasets. The names provide an easily recognizable frame of reference for thematic maps and help to put other data in context for users.

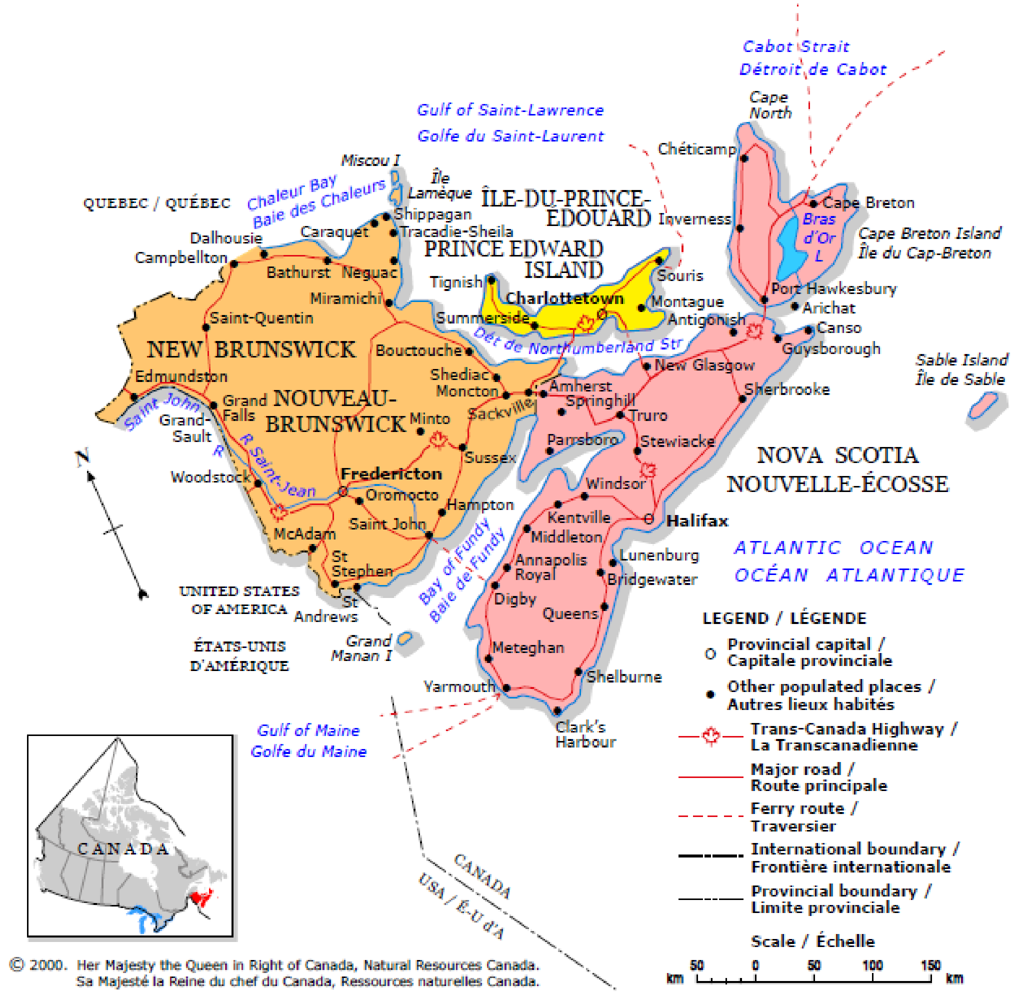


Figure 30-17 A map of Canada's Maritime Provinces, showing official names, including those which are approved in both of Canada's official languages (English and French)

30.5 Further thoughts on multiple source toponymic databases

For a national toponymic database to be of optimum value, the data needs to be accurate, consistent, authoritative and up to date. For databases, such as the CGNDB, where information is combined from multiple jurisdictions, challenges exist in coordinating data that may have been gathered under differing conditions and by various processes. Add to this the different age of records which are accumulated in a database, and it is clear that for consistency, standards must be set and maintained for data fields and their contents. For example:

- an "approved" name entry should use the combination of upper and lower case letters, numerals, diacritics and punctuation, as prescribed by the names authority, and conform to a standard character set, such as UTF-8
- codes for categories of feature types must cover all data contributed to any national database, and should conform to appropriate database modelling standards
- treatment of cross-references (e.g. former names, other language versions, informal variant names) must be consistent
- using unique identifiers for individual name records and feature identifiers for spatial geometries will facilitate linking geographic extents with names and their attributes
- status options for names must be agreed upon by the data providers
- either a regular update schedule, or close monitoring and timely processing of name decisions and updates, is necessary to ensure that national databases remain current

For a successful database, continued discussion and cooperation is essential between those involved with the supply and maintenance of the names data. This collaboration ensures the needs of all parties are considered and the system can be updated to meet technical changes and evolving requirements to access and use the toponymic data.

In future, it is likely that data from municipal governments (such as street and building names) or from research files (e.g. linguistic or historical information on toponyms) will be linked to national systems through location and feature identity. Additionally, the content of geographical names databases is increasingly being associated with other aspects of national (or international) geospatial data, allowing names to be used in conjunction with topographic data, climate data, census data, and so on. For these links and uses to be effective, adherence to well-developed standards for maintaining the national toponymic database cannot be underestimated.

For some, the verification of data being entered (perhaps by several different methods) into the national database will be complicated by multiple language records, by different mandates of the data suppliers, by lack of human and financial resources, or by technical issues. Nevertheless, whether a national database is simple or complex, the goal remains the same – to provide for global users a reliable and accessible view into the toponyms that form a significant part of every country's history and culture.

30.6 References

Listed by date; documents are available on the UNGEGN website (http://unstats.un.org/unsd/geoinfo/UNGEGN/)

Munro, M.R. (1982). *Automation of Canada's National Toponymic Data Base: 1977 to 1982.* E/CONF.74/L.34. Fourth United Nations Conference on the Standardization of Geographical Names, 1982.

Revie, Peter and Kerfoot, Helen (1998). *Canadian Geographical Names Data Base.* E/CONF.91/L.33. Seventh United Nations Conference on the Standardization of Geographical Names, 1998.

Geographical Names Section, Natural Resources Canada (2002). *The Canadian Geographical Names Data Base.* E/CONF.94/INF.22. Eighth United Nations Conference on the Standardization of Geographical Names, 2002.

Ross, Heather (2007). *The Canadian Geographical Names Service in 2007.* E/CONF.98/109/Add.1/EN. Ninth United Nations Conference on the Standardization of Geographical Names, 2007.

Kwiatkowski, Kristina (2016). *Advances in spatial data management in Canada's national geographical names database.* W.P. 71/9. Twenty-Ninth Session of the United Nations Group of Experts on Geographical Names, 2016.

Section 13 Media training

Chapter 31 **Media cartography**

Ferjan Ormeling

31.1 Introduction

When maps have to be inserted in newspaper articles, or have to be shown in television broadcasts, they have to answer special requirements. These are of a **graphical nature** – because they often will be printed on coarse-grained paper in a high-speed printing process, with inferior inks - and of a **conceptual nature**, because of restrictions in size and in the time available for viewing the image – so the graphical contents have to be reduced to the essential spatial aspects of the message, and the same goes for the accompanying toponyms. The message will be about some theme, and everything irrelevant for that theme should be omitted, with the exception of those features (boundaries, communication lines, extent of built-up areas, with their names) that are deemed necessary for orientation.

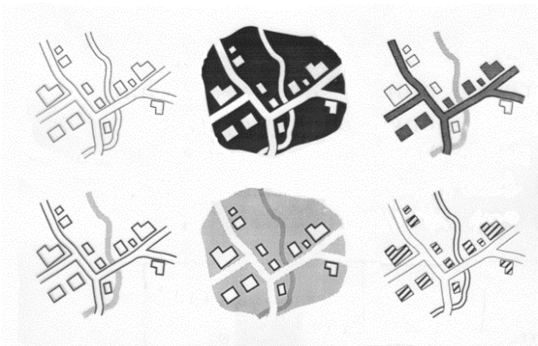


Figure 31-1 Different forms of graphical emphasis

31.2 Graphical requirements

In order to get some idea of the graphical requirements for media maps, we can refer to the contrast that is necessary to let the items mentioned in the message stand out. An example of the techniques to be used is shown in figure 31-1. It shows that graphical emphasis determines the relationships perceived between groups of symbols. The darker elements will be perceived as the map subject, and the lighter ones as background.

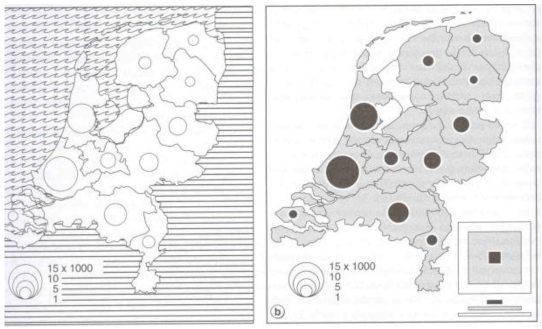


Figure 31-2 Emphasizing the map subject on the map

So, this is the first rule our media maps have to answer: through graphical emphasis the map theme or subject has to be highlighted. This is followed in figure 31-2. Here the number of people working in the IT sector in the Netherlands is shown for 1980, by proportional circles. In the map at the left, foreign parts and the sea are more conspicuous than the proportional circles, and this has been redressed at right. Note that the legend has not been coloured in, as this would draw the attention to the southwest undeservedly.

In traditional newspaper maps, from the second half of the 20th century, the use of pre-printed screens allowed cartographers to differentiate between the subject area and the non-subject area on the map, thus highlighting the location of the subject area in its regional setting. Figure 31-3 shows the example of Kosovo in its regional setting (Yugoslavia), with the non-subject area coloured in by a dark screen. Albania, as an area related to Kosovo because of its language, gets an intermediate screen.

Figure 31-3 Carpress newspaper map agency: Situation of Kosovo in Yugoslavia 12-2-1991

Other graphical aspect of our media maps would be the minimum dimensions of the symbols we were to use on the maps. Even if today we can click on a symbol to see better what is represents, for a proper idea of the distribution of a phenomenon the symbols that represent it should be perceived clearly. Tests have been done, using charts similar in concept to ophthalmologist charts (Snellen charts, developed in the Netherlands in 1862), on which symbols or figures (called optotypes) are displayed in varying sizes, being filled in colour or in black or left unfilled (see figure 314).

When viewed at specific distances, like the average distance to a newspaper or to a television set, the percentage of viewers who would be able to discern correctly between the symbols would be decisive for accepting symbols for specific news media.

Figure 31-4 Set of optotypes on an ophthalmologist chart (from Kolacny 1969)

As a result of perception tests done with such ophthalmologist charts, tables can be developed that show the minimum dimensions of symbols to be used for media at specific reading distances. In the table in figure 31-5, these viewing or reading distances are for atlas use (40 cm), for hand maps (viewed at a distance of 1.5m) and for wall charts (10m).

The tests were based on the phenomenon of the (perceived) viewing angle, the relation between the size of the object or feature to be seen (the optotype) and its distance from the viewer (see figure 31-6). It is expressed as the angle ϑ = S/d (S=size of the object and d = its distance from the eye). When the viewing angle becomes too small it gets impossible for the viewer

Symbol	The smallest possible dimensions for the symbols in mm. for the map reading from the distance:		
	d = 40 cm	d = 1,5 m	d = 10 m
■ a	a = 1,0	3,8	17,5
□ a	a = 1,0	3,8	17,5
▲ v	v = 1,0	3,8	17,5
△ v	v = 1,0	3,8	17,5
▲ a v	a = 0,7 v = 1,3	2,6 4,9	15,1 25,0
△ a v	a = 0,7 v = 1,3	2,6 4,9	15,1 25,0
● d	d = 1,0	3,8	17,5
○ d	d = 1,0	3,8	17,5
⬒ a b	a = 0,8 b = 1,6	3,0 6,0	17,5 35,0
⬯ a b	a = 1,4 b = 0,7	5,2 2,6	30,2 15,1
M a b	a = 1,1 b = 1,1	4,1 4,1	20,1 20,1
M a b	a = 1,1 b = 1,1	4,1 4,1	20,1 20,1
✳ d	d = 1,0	3,8	17,5

Figure 31-5 The smallest possible dimensions of map symbols differentiated by shape and colour. Results of tests by Kolacny (1969) into the legibility of map symbols for different media.

to discern between an o and a c. for example.

Typeface

The typeface used has to be adapted to the medium as well – especially when viewed on small screens. Typefaces that use thick lines and thin serifs are out, as the serifs may cause letters to appear merged. The letters may not be too bold either, because of the danger of filling up the white areas in letters like w, m or p. The conventional wisdom is that on maps

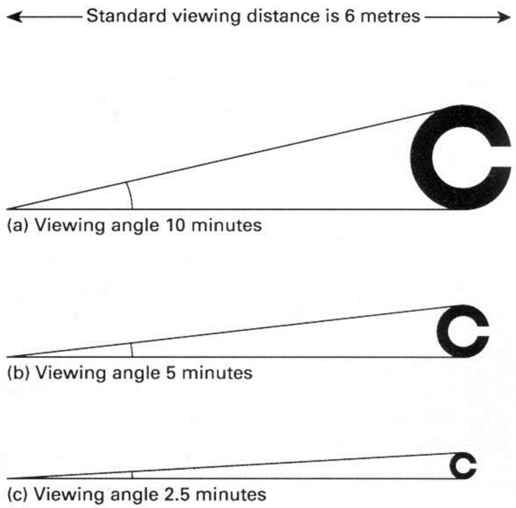


Figure 31-6 The (perceived) viewing angle concept.

lower case letters will do better than capitals, because ascenders and descenders improve recognition of letters (see also figure 31-7). Sans-serif typefaces like Univers with their open characters and wide angles are well-suited for most media.

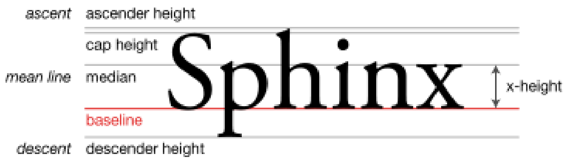


Figure 31-7 Ascenders and descenders improve recognition (Max Naylor 2007)

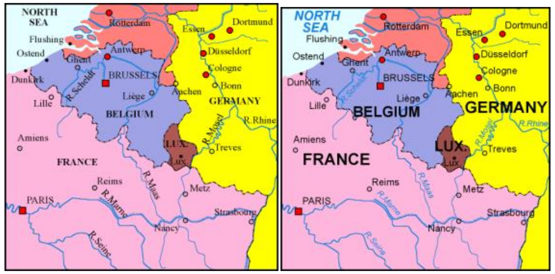


Figure 31-8: Differences in spacing of the names

By spacing the letters of a name and, of course, also by changing its orientation when the object to be named has a large non-horizontal extension (such as Chile for example) the extent of the named object should be indicated clearly by spacing the letters in the name. Thus, in figure 31-8 at left the country names characterize the respective countries insufficiently; at right this has been improved.

31.3 Communication aspects

The editors of the newspaper or TV news journal that includes maps in their papers or broadcasts want to make sure that their audience can handle these maps – abstract representations of reality at best. They have, therefore to match the previous knowledge of that audience – gained at school by using school atlases and looking at wall charts. So, this previous knowledge of the users has to be taken into account by the cartographers (see figure 31-9).

The geographical names on newspaper maps are the best link between these maps and the article in the newspaper. In these newspaper maps people are confronted with geographical features that are unfamiliar to them (such as the locations of earthquakes, tsunamis, battles, railway accidents, etc.). In order to be understood, the location of these new features have to be linked to map features people already know. Readers would be familiar with the names incorporated in the school atlases they used, and there is a good chance that these names would have been exonyms. Of course, the media also have an educational function, and that is why they should make the audience also familiar with the endonym. It is for the journal's editor to decide which course to take here.

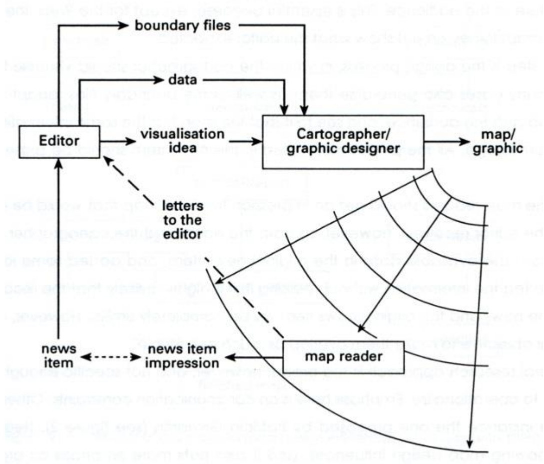


Figure 31-9 Model of the organisational procedure of spatial information transfer through the media (Ormeling 1997)

31.4 Conceptual aspects

For diapositive text slides we used to have the rule that because of the restricted time these would be on display, the number of words per line shouldn't exceed 7 or 8, and there should not be more than 7 or 8 lines either.

Marginal information
As maps in the media, especially when they turn out to be successful, tend to get detached from the presentation they belong to, it is essential that key marginal information is combined with the map display, such as the map title, the map scale (a graphic scale, that will be enlarged or reduced together with the map) and the legend (see figure 31-10). Of course, actually the name of the presentation and the producer should also remain linked to it, but we should already be happy when at least scale, legend and title have been preserved. The title will decide whether the map will be looked at all, and so it must be concise and informative, with mention of the geographical area concerned, the theme mapped and the year for which the data were collected: "Unemployment in Brittany in 1990" would be an example of a good title (see figure 31-10). Additional information, like the units in which the data have been measured or the nature of the enumeration areas, can be added in a subtitle.

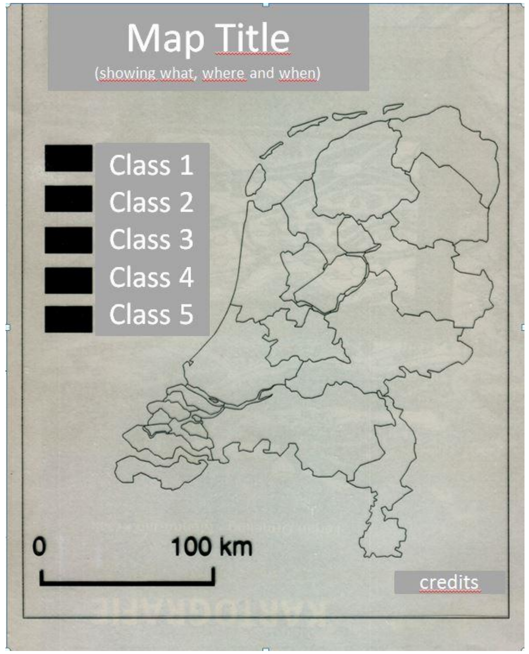


Figure 31-10 Essential marginal information on maps

31.5 Toponymic aspects

The maps should be adapted to the intended audience. As much as UNGEGN is trying to reduce the use of exonyms, many are still retained as they are felt to be part of the respective languages.

But apart from the choice between exonym and endonym, one should never subject one's audience to the use of exonyms in other languages. It would be foolish to use the English (or Dutch) exonym Ceylon for a French audience that would be familiar with the French exonym Ceylan for the name of the island on which the state Sri Lanka is situated (see figure 31-11).

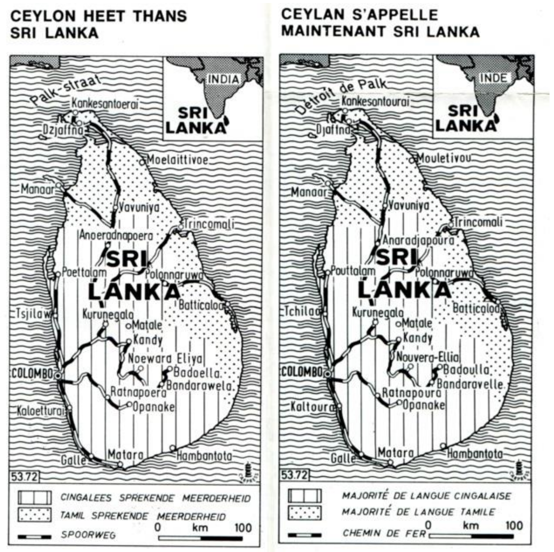


Figure 31-11 Dutch and French versions of a newspaper map produced by Carpress, Brussels, in 1972.

Generic terms should also be translated into the language of the intended audience. In figure 31-11 the name Palk Strait has been translated into Dutch *Palk-straat* for a Dutch newspaper (at left), and to *Détroit de Palk* for a French newspaper (at right). One should realize that there are also differences in the use of generics: on Dutch maps, the generics for river or island would not be shown, as these are thought to be obvious, because of the location of the toponyms on the map, while on French maps these generics would be included.

In figure 31-11, *India* is called *India* in Dutch as well, but *Inde* in French. If the map depicts a country using a non-Roman writing system (as is the case for Sri Lanka), then a Romanization system should be used which is either adapted to the language of the audience, or is an official UNGEGN-accepted conversion system. In the case of Sri Lanka, no UNGEGN-accepted conversion system from Singhalese into the Roman alphabet exists, and therefore on the map at left in figure 31-11, a transcription is used that allows for a correct pronunciation in Dutch, and in the map at right a correct pronunciation in French:

English	Dutch	French
Jaffna	Dzjaffna	Djaffna
Puttalam	Poettalam	Pouttalam
Mulaittivu	Moelaittivoe	Mouletivou
Chilaw	Tsjilaw	Tchilao
Badulla	Badoella	Badoulla

Box 1 Transcriptions of Singhalese names in figure 31-11 into different languages using the Roman alphabet

These considerations have been expressed by Stani-Fertl (1997) in the following guidelines:

1) all geographical names must either be in the language of the audience or in that of the country where the named feature is situated

2) if exonyms are to be used, they should be current

3) if no exonyms exist for a feature in a country not using the Roman alphabet, for an audience using that alphabet an official (UNGEGN-approved) conversion system should be used

The use of exonyms on media maps also depends on the name category: names of countries, seas, straits, mountain ranges usually have exonyms, while for the names of cities or administrative areas endonyms would be used.

31.6 Political aspects

Many boundaries in the world are contested. So it would be wise to check whether the audience the media maps are produced for, agree with their delineation. The same goes for some names – especially sea names. For some audiences, a sea name using the name of an adjoining country would just be considered as a neutral name, for others using this name might involve the recognition of territorial claims over that sea area.

When the Netherlands and Belgium separated in 1830, some provinces, like Brabant and Limburg were split up, causing the emergence of similarly named provinces both north and south of the new border, and this has never caused any problems. Elsewhere the emergence of similarly named areas on both sides of the state boundary might be felt as creating political advocacy for claims (irredenta).

On maps produced by the United Nations Cartography Section, or Geospatial Information Section as it is now called, only those boundaries are represented that are recognised by all parties concerned. These would be *de jure* boundaries, as opposed to *de facto* boundaries. A special case of the latter is the administrative boundary between Egypt and Sudan, which apparently has been demarcated by mutual consent without having been codified by a boundary treaty. Here follow some other examples of the various ways the UN renders boundaries:

- In the border conflict between India and China the UN recognises neither of the two claims, and consequently both the Indian- and the Chinese –claimed borderline have been demarcated on UN maps of these countries. Still there would be a proviso on these maps stating: *The boundaries and names shown and the designations used on this map do not imply official endorsement or acceptance by the United Nations*.
- Western Sahara is considered an independent country by the United Nations, and on UN maps of the area it would be rendered as such, even if the berm constructed by the Moroccan forces for protection against Saharan forces is indicated on UN maps of the Western Sahara.
- On maps of Sudan and South Sudan occurs the following statement: *Final boundary between the Republic of Sudan and the Republic of South Sudan has not yet been determined. Final status of the Abyei area is not yet determined.*
- In other places, the provisional character of the border line is indicated on UN maps by a dashed line (such as in the Ogaden area of eastern Ethiopia).
- In Cyprus, the cease-fire lines of the Turkish forces and of the Cypriot National Guard are marked.

Figure 31-12 Map of Israel, produced by UN Geospatial Information Section (no.3584 Rev 2, produced January 2004)

- For Israel, the 1949 Armistice demarcation line is still used on UN maps to demarcate Israel, although the UNIFIL and UNDOF operational boundaries are indicated as well on UN large-scale maps (see figure 31-12).
- For Georgia, Abkhazia and South Ossetia, according to UN maps, are regarded as integral parts of Georgia.

UNGEGN does not rule on individual names. However, in the United Nations some names have been voted upon by other UN bodies, and have thus been codified, either short-term or long-term. Among them are the sea names *Sea of Japan, Persian Gulf*, and the country names *Côte d'Ivoire, Myanmar* and *The Former Yugoslav Republic of Macedonia*, for use in international maps. The same is valid for the name pair Falkland Islands (Malvinas), so decided by the UN General Assembly in the 1960s. So in documents and maps for which the UN bears responsibility, these names have to be used. In other cases, the UN Geospatial Information Section has used its own authority to render on the maps name pairs, like Lake Malawi/Lake Nyasa. UNGEGN produces rules and regulations about the use of geographical names, expressed in resolutions accepted by ECOSOC. Those are binding for international use by UN, and we hope for its member states as well. But UNGEGN cannot oblige individual countries to use any particular geographical names for national use. This is something one has to bear in mind when producing media maps for strictly national consumption.

31.7 References

Kolačny, Antonin (1969) Utilitarian cartography – the road towards the optimal effect of cartographic information. ICA Working Group on Cartographic Information, Prague.

Naylor, Max (2007) *Typography Line Terms*. Own work. Licensed under Public Domain via Commons – Wikipedia.

F.J. Ormeling - Mass media maps as objects of cartographic thought and theory. p.22-33, W. Scharfe e.a. - International Conference on Mass Media Maps - Proceedings. Berliner Geowissenschaftliche Abhandlungen Reihe C, Kartographie, Band 16 (1997).

Stani-Fertl, Roman (1997) Geographical names and mass media maps. Pp 173-178 in W. Scharfe (ed.) International Conference on mass media maps-proceedings. Berliner Geowissenschaftliche Abhandlungen Reihe C – Kartographie-Band 16. Berlin: Freie Universität-Technische Universität-Technische Fachhochschule.

Vujakovic, P. (2014) *'The state as a 'power container': the role of news media cartography in contemporary geopolitical discourse.* The Cartographic Journal, 51 (1). pp. 11-24. ISSN 0008-7041.

The Carpress cartographic News Agency was operational from 1950-1990, based in Brussels, Belgium.

Section 13 Media training

Chapter 32 **Exercise in Producing Media maps**

Ferjan Ormeling

32.1 Introduction

Media maps are produced in order to illustrate news items that have a spatial connotation. Figure 32-1 shows that for a news item something of a particular nature happens at a specific place and time to a subject, because of some internal or external development, and that this is supposed to have led to specific consequences.

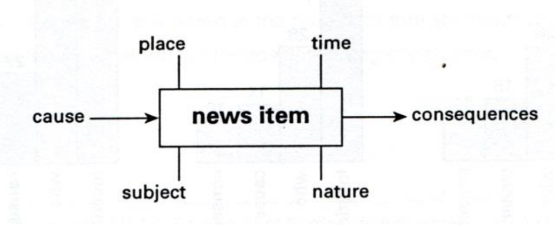


Figure 32-1 Model of the structure of a news item or news event

The map would be relevant in trying to show why the event happened there, as well as the consequences of the event happening there (a strike, traffic jam, earthquake, missile theft, famine, epidemic, etc). The map would portray the relevant spatial relationships, and in order to realize that, it has to show connections between the location of the event and places known to the audience.

It is no use, when visualizing an event that happened in Zevendijke in the central Netherlands, to show that village against its municipal background – it has to be shown against the background of well-known places like Amsterdam, Rotterdam or The Hague (the seat of the government of the Netherlands, 's-Gravenhage or Den Haag, is known internationally as The Hague).

Figure 32-2 Relevant locator map

But location alone is not enough. It is a well-known aspect of media graphics (Findahl and Hoyer, 1971-1985) that in newspapers or television news broadcasts, news items are better remembered when supported by visual evidence such as maps. And a map which does not only indicate the place of the event, but also highlights the (nature of the) event itself again in a caption on the map, leads to even better results (figure 32-3 C)

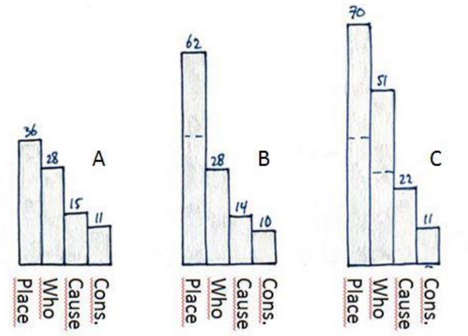


Figure 32-3 The role of the map in remembering news items from a newspaper: The graph shows the percentage of people able to recollect the place, persons involved, cause and consequences of news items: A) (left) without a map; B) with a map added; C) with a map added which also refers to the event in a text block or caption mentioning it.

Figure 32-4 Locator map with event caption added, and inset to show the location of the main map

Of course, the map has to be attuned to the article describing the news item: the spelling of geographical names in the text and in the map must be similar, and all places mentioned in the text must be shown on the map.

32.2 Map function

What functions do maps have in the media and how do they answer the relevant requirements for these functions?

Most maps in the media have an orientation function. So, they act as Locator maps. Locator maps inform about the location of a specific event or process.

Geographers know that maps will also inform about site and situation, and therefore provide spatial explanations. The site of Rotterdam is on the North Sea at a location where the tidal difference in water level is the least; the situation of its port is its proximity to one of the largest industrial areas of the world, connected to it by way of a river as an inexpensive carrier of mass goods. So, site and situation provide the explanatory function of media maps. But it needs to be sustained through symbols or texts in order to be effective.

Explanatory function and orientation function are not always easy to discern between. Take the case of the Falklands. In a media map of South America, the Falklands are so close to the mainland, that it seems callous of the United Kingdom not to render them to Argentina. But when one would add Britain at the same scale, it would enable one to become aware of the distances involved: Port Stanley is as far from the Argentine coast as Iceland is from Britain - and one would not think of claiming Iceland for Britain or vice versa on the basis of the distance between them! (from Ormeling 1997). See figure 32-5.

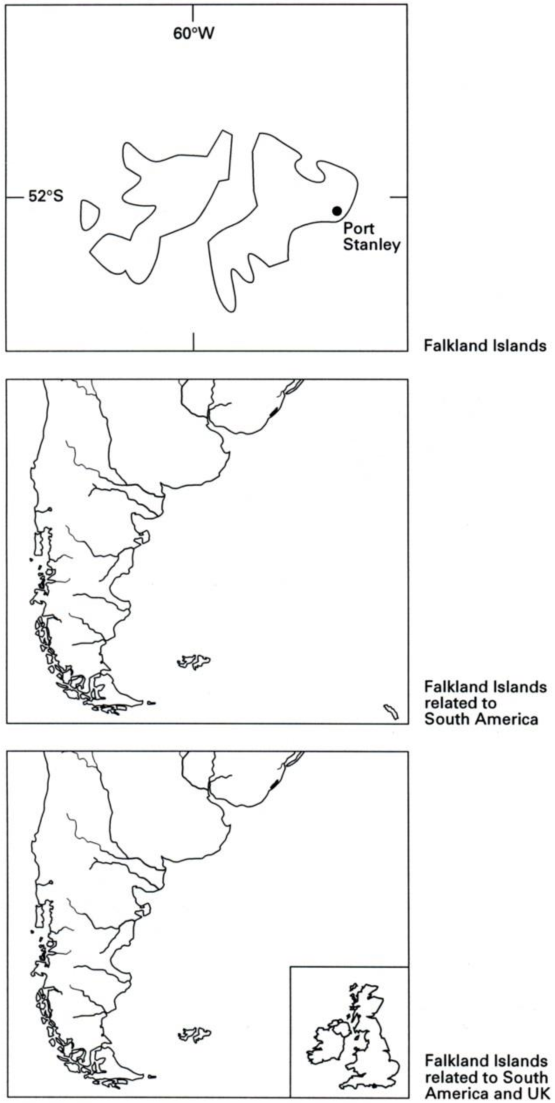


Figure 32-5: the influence of map scale (Ormeling 1997)

Explanatory aspects can both be based on things that occur at the same place (this is called vertical explanation), or on things at different places that influence each other (horizontal explanation). The examples of the Falklands is a case of horizontal explanation. Showing the proposed route of a new highway relative to subsoil characteristics, is an example of vertical explanation.

c) More mundane is the **decorative function** or stimulus-function of a map: something graphic that looks nice and therefore beckons the reader to get at a specific text.

d) Finally, there would be the **logo-function**: a simplified map that would show the readers by its outline what geographical region the news text highlighted by it, would refer to.

e) Answering a) the orientation function, would call for the provision of means that linked the area discussed to the readers' cognitive map (i.e. the ideas about space or about their environment the readers have in their minds): either by providing a shape that they would know or by providing the location of the area relatively to areas the readers would know. Showing the Falkland Islands on their own with identification of their longitude and latitude would not say anything to most readers; by showing them at a smaller scale that would also allow for (part of) South America to be rendered, would improve orientation possibilities considerably. In an Internet environment or, in future, in an electronic newspaper environment, it would be feasible right now to have the possibility to ask for extensive additional orientation possibilities (e.g. showing the map on which both the viewer's home country and the area discussed would be rendered.

To what degree are **comparison maps** that compare the size of a specific area to the size of an area known to the viewer also necessary? These comparison maps are devices that are used in school atlases. They would show, in some corner of a foreign land, (part of) their home country, in order to be able to assess the distances or surface areas involved, and get a proper idea of their magnitudes.
Part of the explanatory function dealt with on the previous page, is to prevent the audience from taking the wrong conclusions from the map.
In figure 32- 6 a map of the number of abortions in France is shown, per department. This is done by relating the numbers of abortions per department to the number of life births in the same department. As the availability of abortion clinics in every department would be an important issue here, a warning is included: the observed differences between departments will partly have been caused by insufficiency of abortion facilities.

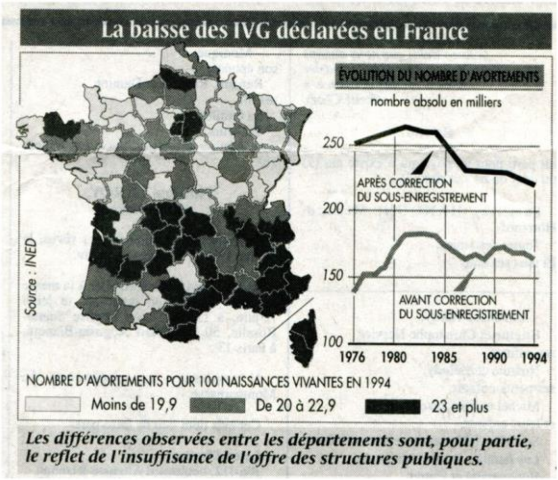


Figure 32-6 The decrease in declared abortions in France (Le Monde 11/6/97). The bottom line reads: The observed differences between the départements are partly due to insufficient supply or capacity of public infrastructure.

32.3 Exercise aspects

What is needed for an exercise is, of course, a suitable news item – in our case a news item with a location that can be mapped where something tangible happened: where pictures were stolen, statesmen met, treaties were signed, trains collided, battles were fought, etc.

Place

Next, a base map would be needed om which the location can be pinpointed (the site), but which also allows for reconstructing the relevant spatial aspects of the site (the situation): what were the getaway routes, what nearby airport did the statesmen fly in to, where did the trains come from, or which was the tactical situation on the battle front? So apart from a (topographical) base map, a railway map or tactical terrain map might be needed as well for documentation, in order to be able to include all relevant data on the newspaper map. As was shown in the Falkland example, the scale of the map is relevant as well.

Time

One of the aspects of the news item to be supported by a map is time (see figure 32-1). Of course, a newspaper map is static, and for television news broadcasts, where it would be possible to produce dynamic maps, there usually wouldn't be time enough to produce (and show) the movements that led to or followed from the reported news event. Only in exceptional situations that occur and are reported on continuously, like the weather, is it worthwhile to set up a system that allows for dynamic maps, for instance for showing the movement of weather fronts.

But there are other aspects of time, we also have to reckon with, using static maps: when for instance a war situation is being reported on, and on consecutive days the location of the front is to be shown, and thus the map becomes part of a series, we have to take care that we use the same scale, the same shades for each of the protagonists, and the same symbology.

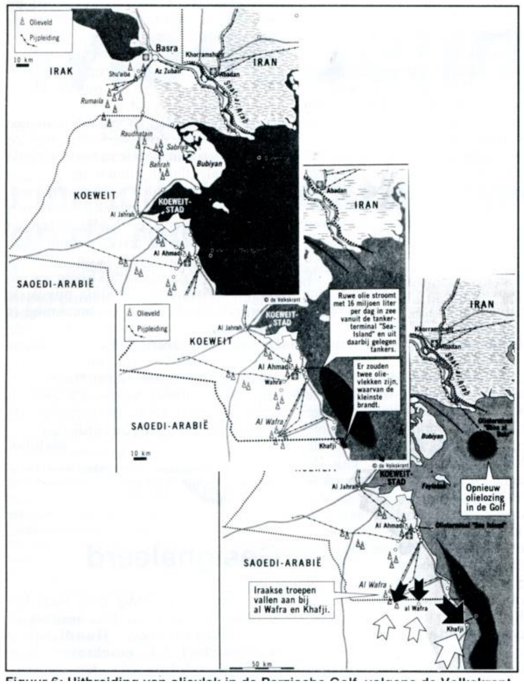


Figure 32-7 series of oil spill maps from newspapers on consecutive days, on the same scale and with the same legend.

Subject.

There frequently is an antithesis in the news item: the cops and the robbers, the two parties in a treaty, the belligerents, etc. In comic strips the hero is always dressed in white and the knave in black; the hero comes into the drawing from the right, unlike his sinister adversary. Although these kinds of clues might not be used in newspaper maps, at least one should think about the way in which the opposite parties, adversaries, etc are being portrayed.

Nature

Monmonier (1989) proposes to discern amongst the following themes for media maps:

1. crime
2. physical planning
3. tourism
4. politics/elections
5. military conflicts/terrorism
6. defense issues/geopolitics
7. (development of) resources and their transport
8. accidents, natural disasters
9. science, research, ecological disasters
10 education, history, demography.

This seems to be a useful categorization (though it seems not to account for weather maps?) which can be used in a number of ways. The development of mass media maps over time would be an example. There seems to be a trend the last 50 years towards more diversification amongst mass media maps. First, only weather maps and maps of disasters and military conflicts were incorporated in newspapers. In conjunction with the rise of thematic maps, socio-economic maps and elections maps were being added gradually as well as maps on crime and physical planning, while the incorporation of science and tourist maps seems a rather recent addition.

This differentiation into, first map functions, and now map themes, is included here, because those that do the following exercise might want to adapt their design to the function or theme of the map. The nature of the event/theme can be expressed by the style of the media maps

32.4 Exercise

News item

The exercise is about the following fictitious news item: a crash on Tuesday May 2015 between an early morning passenger train going north from Meppel (departure Meppel 7.30) to Leeuwarden and a freight train loaded with crude oil, coming southwest from the Eastern Drenthe oilfields, via Meppel to the refineries in Rotterdam, which apparently ran through a red light. The two trains collided at the junction about 1 km north from the passenger station in Meppel, in the countryside next to a canal. Part of the passenger train derailed and was left hanging above the canal. Injured casualties were brought to Meppel hospital, but as that had limited capacity, ambulances also went for Hoogeveen and Zwolle hospitals. Because of the fire that had spread, there were many victims covered in burns, and for them a helicopter airlift to specialised hospitals in Amsterdam and Beverwijk was organised.

As the accident happened during rush hour, a total chaos and monumental traffic jam ensued which provided additional difficulties for the ambulances. At this moment (13.00 hours) 44 casualties and 240 wounded have been reported. There was a sizeable contingent of British horsemen aboard the train, that had spent the night in Zwolle and Meppel hotels, on their way to Heerenveen in Frisia province in order to participate in the Military there. Because of the limited hotel capacity in Heerenveen, Military participants had had to find accommodation some 30k away from the racecourse.

Assignment

The exercise is about designing a media map to support the newspaper article about this event for a British newspaper. Below you will find a site map, an overview map of the situation, a map showing the geographical relation of the Netherlands to Britain, and, finally, a possible solution to the assignment

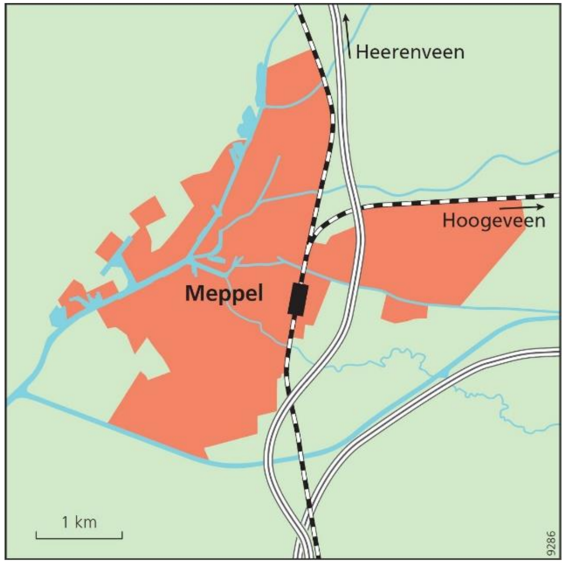


Figure 32-8 site map, showing the fatal bifurcation in the rail network, the built-up area of Meppel (30 000 inhabitants), the motorways and canals

What the maps in figures 32-8 and 9 do not show is the soggy character of the area where the trains collided. The canal drains the Drenthe plateau to the northeast of Meppel, and during the spring rains the canal is frequently overflowing, flooding the adjacent area, except for the rail tracks that run on an embankment

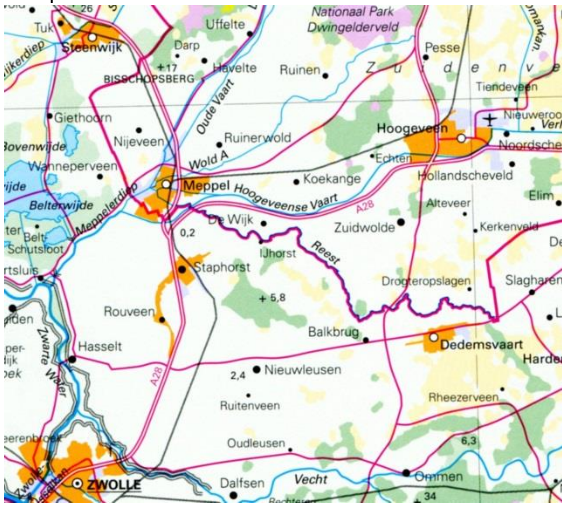


Figure 32-9 Situation map

Figure 32-10 Map showing the relation to Britain

Figure 32-11 Detail of the rail network map of the Netherlands

Figure 32-12 Your map assignment

32.5 References

Findahl, Olle and Hoyer, Birgitta (1975) Effect of additional verbal information on retention of a radio news programme. Journalism Quarterly 52 (3) 493-99

Findahl, Olle and Hoyer, Birgitta (1985) Some characteristics of news memory and comprehension. Journal of broadcasting and electronic media 29, 379-396

Findahl, Olle (1981) The effect of visual illustrations upon perception and retention of news programmes. Communications 7, 151-167

Monmonier, M.J. (1989) Maps with the news. The development of American journalistic cartography. Chicago: University of Chicago Press.

Ormeling, F.J. Mass Media Maps as Objects of Cartographic Thought and Theory. *Proceedings International Conference on Mass Media Maps. Berliner Geowissenschaftliche Abhandlungen Reihe C, Kartographie, Band 16,* ed. W. Scharfe (ed), 22-33. Berlin: Freie Universität, 1997

Chapter 32 **Fieldwork Interviews**

Élisabeth Calvarin

33.1 Fieldwork preparation

Preparations for the collection of geographical names make it necessary to anticipate possible difficulties that could manifest themselves in the field and to try to devise adequate solutions in advance, in order to avoid unnecessary efforts and loss of time.

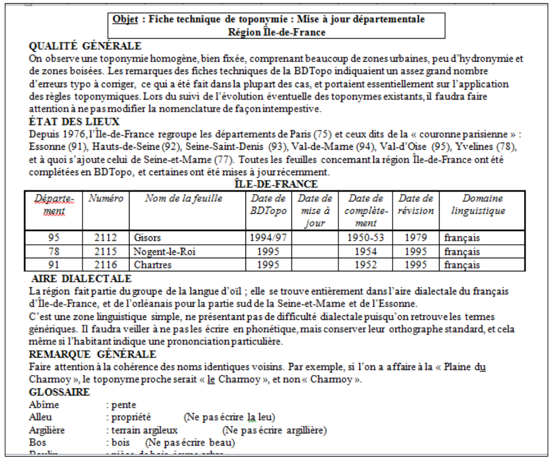
33.2 At the office, selection of the itinerary

Before the whole undertaking, one should start at the office with the selection of the itinerary. It would be good practice to opt for a test-route first, selecting an area with a variation in geographical objects: administrative centres, dispersed or concentrated population, varied relief, permanent rivers or wadis, construction works (wells, dams, bridges), forests, cattle-breeding and agricultural areas, many hamlets, manufacturing plants, schools, cultural centres, etc…

Figure 33-1 Itinerary selected on the basis of a map 1: 200 000 of Ouagadougou (Burkina Faso) in 2008

33.3 Specific preparatory information needed for meeting local authorities

It is always advantageous for those in charge of the fieldwork operations to contact the names bureau or the toponymists in their agency. Those may have prepared toponymic card-index systems, technical reports, or guidelines regarding the languages spoken in the fieldwork region, linguistic and social influences exercised there, sub regions for which the orthography of place name families should be harmonised, or even overviews of the mistakes that have been made in the past.



Objet : Fiche technique de toponymie : Mise à jour départementale
Région Île-de-France

QUALITÉ GÉNÉRALE
On observe une toponymie homogène, bien fixée, comprenant beaucoup de zones urbaines, peu d'hydronymie et de zones boisées. Les remarques des fiches techniques de la BDTopo indiquaient un assez grand nombre d'erreurs typo à corriger, ce qui a été fait dans la plupart des cas, et portaient essentiellement sur l'application des règles toponymiques. Lors du suivi de l'évolution éventuelle des toponymes existants, il faudra faire attention à ne pas modifier la nomenclature de façon intempestive.

ÉTAT DES LIEUX
Depuis 1976, l'Île-de-France regroupe les départements de Paris (75) et ceux dits de la « couronne parisienne » : Essonne (91), Hauts-de-Seine (92), Seine-Saint-Denis (93), Val-de-Marne (94), Val-d'Oise (95), Yvelines (78), et à quoi s'ajoute celui de Seine-et-Marne (77). Toutes les feuilles concernant la région Île-de-France ont été complétées en BDTopo, et certaines ont été mises à jour récemment.

ÎLE-DE-FRANCE

Départe-ment	*Numéro*	*Nom de la feuille*	*Date de BDTopo*	*Date de mise à jour*	*Date de complète-ment*	*Date de révision*	*Domaine linguistique*
95	2112	Gisors	1994/97		1950-53	1979	français
78	2115	Nogent-le-Roi	1995		1954	1995	français
91	2116	Chartres	1995		1952	1995	français

AIRE DIALECTALE
La région fait partie du groupe de la langue d'oïl ; elle se trouve entièrement dans l'aire dialectale du français d'Île-de-France, et de l'orléanais pour la partie sud de la Seine-et-Marne et de l'Essonne.
C'est une zone linguistique simple, ne présentant pas de difficulté dialectale puisqu'on retrouve les termes génériques. Il faudra veiller à ne pas les écrire en phonétique, mais conserver leur orthographe standard, et cela même si l'habitant indique une prononciation particulière.

REMARQUE GÉNÉRALE
Faire attention à la cohérence des noms identiques voisins. Par exemple, si l'on a affaire à la « Plaine du Charmoy », le toponyme proche serait « le Charmoy », et non « Charmoy ».

GLOSSAIRE
Abîme : pente
Alleu : propriété (Ne pas écrire la leu)
Argilière : terrain argileux (Ne pas écrire argillière)
Bos : bois (Ne pas écrire beau)

Figure 33-2 Abstract of a toponymic card prepared by the Names bureau of IGN, the French national topographic mapping agency.

Moreover, it is always essential for those in charge of the names collecting fieldwork, to contact local authorities, to advise them of their coming, in order to inform them of the reasons for their visit and about the nature of the work required. The local authorities thus would also be asked to support the names collection work with their local knowledge and historical expertise.

33.4 Quality check

It is always worthwhile to take stock of the state of the toponymy of the region concerned, on the basis of the collected documentation, in order to be able to estimate the time needed to complete the work.

The names bureau may assess the quality of the existing toponymy on the basis of the principles adopted (standardisation rules, transcription, transliteration, use of glossaries). By doing so, place names may be judged correct, muddled, ready for improvement or for correction after verification.

33.5 Names density

It would be just as important to discuss in advance the required average names density and the insertion or positioning of the selected names on maps of a given scale.

The number of names inserted would vary according to the nature of the operations (depending on the kind of the terrain and the legibility). The following numbers might give an idea of good practice:

On the average, we need 4 names per km2, that is between 420 and 550 names on a standard map sheet at the scale 1:10 000, between 660 and 780 names on a standard map sheet at the scale 1:20 000, and from 1800 to 2000 names on maps at the scale 1:25 000.

But it would be wise to recognise that desert areas cannot contain but a very limited number of names and that highly developed areas would contain many more names. In the latter, legibility will impose restrictions, depending on the scale and (special) purpose of the map in question.

33.6 Knowledge of administrative unit lists

As it is our role to formulate general principles, we have to make sense of the country, that is, we have to find out its administrative structure and define its permanent units, made up from former or transitory units.

- 1 <u>Structure of the country</u>: we have to get to know the first and second order administrative divisions, and next the suitable administrative unit under which we can group the toponyms we have collected. It would be useful to write down for every administrative unit the local words that signify « family » (household, house, and yard), « clan, line », etc. in order to render these social constructs on the map and include them in the data base, taking account of this immaterial cultural heritage.

As an example, we take Tunisia: here the 24 first order administrative units (governorats) have been subdivided into 264 second-order units (délégations or municipalities). The smallest administrative division is the ward or *imada*, of which in December 2011 there were 2073.

- 2 <u>define the objects to register</u>: record all permanent constructed objects (such as towns, villages, local government buildings, etc.) and render them differently from former or temporary objects (former villages, dispersed buildings, etc.). – without forgetting the latter.

A complete list of administrative toponyms can only be arrived at with the help of local government officials or dignitaries that have been designated by those in charge. It would be essential to contact those in charge of local administration in order to avoid mistakes, like the insertion of non-existing administrative names (like river names), imaginary or invented names.

33.7 Preparation of the names inventory forms and of the map of the fieldwork

The objective of the toponymic fieldwork is to collect for a given area the toponyms in current use (checking whether former toponyms still are used, or new toponyms created), checking the disappearance or re-emergence of names, or whether the reference area of the names has changed, in order to create or to update toponymical data bases.

The results of this work will be written down both in names inventory forms (in French: cahier de terrain, cahier d'inventaire or État Justificatif des Noms [EJN]), and on maps of the fieldwork area.

These two documents (name forms and maps) together should contain all the necessary information and documentation regarding the current use (authenticity) of the toponyms for a well-defined area.

In order to be understood (by both the map reader and the inhabitants), it should be one's goal to let the toponyms remain as close as possible to local usage: all additions or modifications of toponyms in order to render their orthography more correct, logical, or in line with usage elsewhere should be introduced with the help of clear standardisation principles, after having consulted local users.

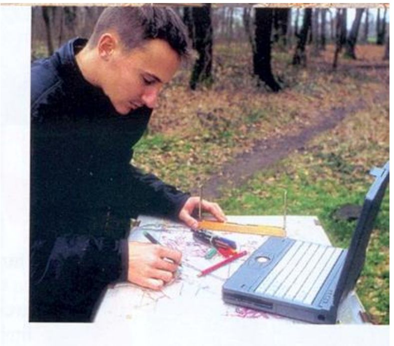


Figure 33-3 An operator is collecting toponyms, with the help of his map and his digital names inventory forms.

It is the goal to establish a names inventory form which, corresponding with the fieldwork area map, will allow the names bureau or the person responsible for the toponymy to determine, under optimal conditions, the graphic form of the names that will be incorporated in the data base.

The names inventory forms have to be filled in as follows:

All the place names listed during the field work have to be correctly located, in their final orthography in the names inventory forms, supported by the map (which is also named in French « minute de complètement »).

Whatever the format of the names inventory form (paper or digital), it has been established in order to legitimise the orthographies adopted. It should contain all the information the operator has been able to collect during the fieldwork on every name. It should at least contain the following sections: the map title, the names contained in the maps already published, administrative names (for census areas or the land cadastre), the names according to the local inhabitants and according to name signs, the local pronunciation, former spellings, the etymology, the meaning of the name, the nature of the named object, and finally the orthography of the name as proposed by the operator.

In the names inventory form, fields have to be filled in. Some of them can be filled in beforehand, at the office; others should be filled in in the terrain. The last section, showing the final orthography as decided by the Toponymy commission or by the person responsible for the toponymy, has to be filled in after the fieldwork.

33.8 Types of names inventory forms

The names inventory forms should contain as many pages as is necessary. Its pages will be divided into different sections.

First, names should be grouped under administrative headings, with a clear title (in France this could be a préfecture or a municipality; in Tunisia a first order administrative unit (gouvernorat), and under the number of the map sheet concerned, according to the numbering system in the sheet index.

GPS geographical coordinates. Both the latitude and the longitude of the named object have to be assessed accurately; this may be done in 10m units.

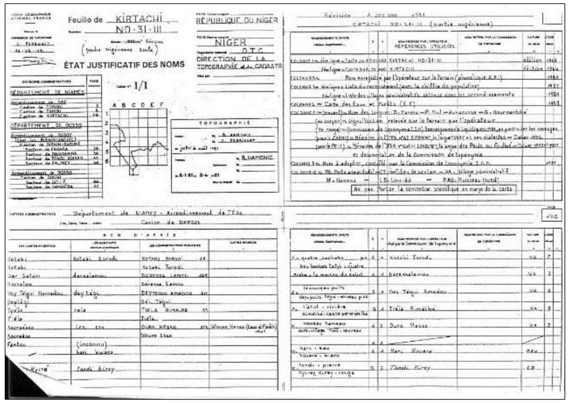


Figure 33-4 Example of some pages from a names inventory form, established in Niger in 1981

Maps already published. Here it is about the name as written on the previous edition of the map. Should this not be available, then one should have recourse to old maps, of which the title, date, scale and publisher should be mentioned.

Nature of the named object. In this section, the nature of the named object should be specified: does the name refer to the principal town, to the parish, to permanent or temporary villages, to hamlets that may be inhabited or not, to rivers ending in the sea or in other river, watercourses, lakes, ponds, to mountains, hills or valleys, passes, forests or woods, cultivated fields or conspicuous trees, etc. To guide this process, one has to compose a summarizing table on the nature of the detail (see figure 33-5) to which one may refer, and to which one has to conform to.

Importance. Regardless whether it concerns physical or man-made objects, every object is more or less important in relation to nearby objects of the same kind.

Starting from the nature of the named object, one should stick to the criteria for object importance as noted in the summarizing table in figure 33-5.

	Very important	Important	Less important
Inhabited places	1.Administrative names(provinces, municipalities, etc)	7.Hamlets, wards, castles,misc.institu - tions, abbeys	13.Gamekeepers lodge, farms
Orography	2.Mountain tops, ranges, passes, hills, islands, capes	8.Table lands,rocks ,coombs and glens, mountain ridges	14.Caves, beaches
Hydrography	3.Rivers, lakes, canals, glaciers, bays, gulfs, deltas	9.Ponds, marshes, brooks, torrents, (water)falls	15.Wells, fountains, pools
Field names	4. National parks, forest reserves, woods	10.Non-inhabited physical regions of average importanc	16.Field names of minor importance, remarkable trees
Infrastructure	5.Motorways, airports, dams, tunnels	11.Highways, viaducts, bridges, approaches,exits	17.Roman roads, ways, ski lifts
Miscellaneous	6.	12.ruins, prehis-toric monuments, stadiums	18.Monuments, statues, barns, cabins

Figure 33-5 Summarizing matrix table showing the nature and the relative importance of the named objects. The fields 1-18 are ranged from higher to lower importance

Name according to cadastral or administration documents or other sources. In principle, these are the names contained on the cadastral plan. In case there are differences between the names on the cadastral plans and the ledgers, this should be mentioned. Whenever possible, one should also try to collect the name variants used by the postal authorities, by foresters, on road signs, in village guides, on village lists and in gazetteers…In this list one could also mention the names as used on information panels, or derived from local history experts.

The inhabitants. One should rather consult locals that have lived in the area all their life than recently settled persons. The best indications are usually given by those that are most familiar with the area.

Miscellaneous. One cannot do without this section. Here one should be able to find the meaning of the name, its former and local variants, the local pronunciation, and is supposed etymology. When neither meaning nor etymology are given, one should write down the current nature of the named object, the aspect of its slopes, the kind of relief forms as well as remarkable objects too be found there: rocks, springs and quarries or its former nature, such as abandoned villages. Just one of those details might explain the meaning of the name as it has been handed down.

Name proposed by the operator. Here the name has to be entered in the orthography in which it should be inserted on the map and in the data base. Current practice in France is to insert here the name as given in the field « principal name version », « name in French », « variant name » or « local nom ».

Name retained by the names bureau. This section is left for the toponymic authorities or the names bureau to fill in for every name they have the final decision on.

33.9 Specific questions to put when meeting local authorities or inhabitants in the terrain

How to put questions?

One has to make sure to what geographical objects the names are linked. Frequently the local population attributes them to terrestrial objects rather than to rivers, or only to rivers or to some characteristic locations; names may be applied to a mountain or only to one of its slopes. Each of the main slopes or aspects of a mountain might have a different name, different again from the name of its summit. Are these the names used in their own language, when amongst themselves?

Wherever the fieldwork takes place, one should never forget that human evidence is something fragile and uncertain. In oral inquiries, a confirmation cannot be regarded as valid unless it is corroborated from another, different source. This is a guideline one should always adhere to. And in order to validate such corroborations one should make sure that they are given spontaneously.

Regarding interview techniques, it is almost impossible to interview people regarding the names of places and designated points, without generating possible confusion; it has to be done with sensitivity. The way in which we pose questions influences in a major way the spontaneity and the validity of the answers. The question should be completely open and should leave the spokespersons completely free in answering them, thus avoiding leading questions suggesting specific answers. « How this name should be written? » is a question well-posed. One should avoid questions like « Does one write it like this? ».

1st point to deal with: identifying the place. One would ask « How do you call the place where we are? » Depending on the answer, the question could be completed by asking « What is its name in the administration language and what is it in the regional language » « How is this place named on the cadastral plan, and what is its permanent name? » One should record the pronunciation accurately, subsequently one could enlarge the scope of the investigations in order to test the names collected later against each other: « What is the name of the wood we see from here in that direction » « What is the name of the locality situated immediately north of where we are now? » To support this question, we may compare it with a copy of the cadastral plan, or other administrative documents that were taken along.

2nd point: the meaning of the names. « What is the meaning of that name? ». «To what events in local history does it relate? » « What legends does it refer to? » One should also write down the relationship with the terrain forms, the geographical location (situation), the nature of the soil, or whether there are any rocks, remarkable wells or ruins nearby…

3ed point: the orthography. Here one should take care not to make any suggestion regarding the spelling, in order to avoid that the spokespersons might find it better than the one they use themselves, and therefore could be tempted to substitute them. « How do you write that name? », and subsequently « Is the name correctly written on the cadastral plan? » « Did you also see that name written differently? » and if so: « Where? »

4th point: the pronunciation. It frequently transpires that a name can be pronounced in different ways according to the spokespersons. Of course, one has either to record the pronunciation with the help of the international phonetic (API) or with the Roman alphabet as its letters is pronounced in the country concerned

One should only pose simple, direct and precise questions, and deal successively with the different points they are bearing on. One should adapt the conversation to the level of the spokespersons in order not to ask them something beyond them, which would put them in a difficult position, as the necessity could always emerge that the spokesmen would have to be interviewed again.

Finally, it would be local cultured persons (intellectuals, teachers…) that can shed light on the etymology of the place names collected, former orthographies, different possible or conjectured meanings, social histories, the credibility of ancient legends, and the validity of the information already received elsewhere.

Figure 33-6 Collection of toponyms during fieldwork at Ghar el Melh (Tunisia) in 2015

Never believe that this multitude of precautions is redundant or superfluous. All information collected can be required later when the names bureau has to express its opinion regarding the difficult cases that would be submitted to it, as every piece of information could enlighten it.

33.10 References

Deslandes, M., *Toponymie,* Fascicule I, École nationale des sciences géographiques (ENSG-IGN), Paris (1963) revu Calvarin, Élisabeth (2007)

Institut de géographie du Burkina (IGB) *Atelier de sensibilisation sur les noms de lieux*, Ouagadougou (2008)

Centre national de cartographie et de télédétection (CNCT), *Workshop on Geographical Names*, Tunis (2015)