



**UN-GGIM
EUROPE**

UNITED NATIONS
COMMITTEE OF EXPERTS ON
GLOBAL GEOSPATIAL
INFORMATION MANAGEMENT



Core Spatial Data Theme 'Geographical Names' Recommendation for Content

Working Group A - Deliverable of Task 1.b

Version 1.0 - 2017-11-10

Version History

Version number	Date	Modified by	Comments
1.0	2017-11-10	WG A	Consolidated draft, for review by geographic and statistic community

Content

1	Executive Summary	4
2	Foreword	5
2.1	Document purpose and structure	5
2.2	Core data context	6
2.3	Document objectives and principles	7
2.4	Abbreviations	8
2.5	Glossary	8
2.6	Reference documents	9
3	Overview	10
3.1	General scope	10
3.2	Use cases	10
4	Data content	11
4.1	Features types and attributes	11
4.2	Geographical extent.....	12
4.3	Levels of detail and data capture rules	13
4.4	Quality	15
5	Other recommendations	17
5.1	Coordinate Reference System (CRS)	17
5.2	Metadata	17
5.3	Delivery	18
6	Considerations for future	19
6.1	Improving the geometric representation.....	19
6.2	Importance of named places	19
6.3	Pronunciation of names	20
6.4	European classification of named places	20
7	Annex A: Relationship with INSPIRE	22
7.1	Data model	22
7.2	Other	24
8	Annex B: Methodology	26

1 Executive Summary

As the United Nations (U.N.) Millennium Development Goals (2000) era came to a conclusion with the end of the year, the U.N. announced the 2030 Agenda for Sustainable Development in September 2015, an ambitious, integrated, indivisible and transformational global agenda with 17 Sustainable Development Goals, 169 associated targets and 230 indicators promising to achieve sustainable development in its three dimensions – economic, social and environmental – in a balanced way. Geospatial data supports measuring, achieving and monitoring several if not all goals and targets set by the 2030 Agenda. The 2030 Agenda mentions the need for new data acquisition and integration approaches to improve the availability, quality, timeliness and disaggregation of data. Goal 17 explicitly emphasizes the need for developing capacities and partnerships. In this context the success of Agenda 2030 depends on senior administrators owning and leading the geospatial efforts in their respective countries.

Building on INSPIRE Directive and pertinent documentation and redirecting the focus on a cohesive Spatial Data Infrastructure without gaps in content and discrepancies in quality, stakeholders in Europe are working on geospatial standardization and increasing richness of data through Core Data Recommendation for Content that corresponds to the first phase of the WG A work program. Core Data is primarily meant for fulfilling the common user requirements related to SDGs in Member States and European Institutions.

Geographical names are necessary for two main use cases: as search criteria for geocoding gazetteer services and for mapping. In addition, the populated areas with their geographical names help to locate where people are, what is a basic starting point to analyse, achieve, or monitor many SDGs.

In order to avoid duplication of production efforts, the scope of core theme Geographical Names has been limited to the named places and names not present in other core themes; i.e. it focus on populated places and on some natural named places, such as orography or vegetation.

Its Recommendation for Content includes a two-step approach for its adoption by UN member states.

The first step pertains to **core recommendations** and was widely inspired by the results of the ELF project. The main core recommendations include the capture of geographical names on whole territory at large scale for national use and the provision of exonyms for international use, the focus on named populated areas – which are of key importance for many SDG related use cases, the definition of the key attributes of named places: geometry, classification, population indication, geographical name with its spelling and information on its language, nativeness, status and source.

The second step pertains to **good practices**. The main good practice is that named places should be captured with their “true” geometry, i.e. generally as a surface or multi-surface. This good practice encourages to move from a mere cartographic viewpoint to a wider topographic viewpoint, where the named place stands for a real world physical phenomenon, in order to enable user processing such as spatial analysis or cartographic selection.

2 Foreword

2.1 Document purpose and structure

2.1.1 Purpose

This document provides the main characteristics of core data for theme Geographical Names with focus on the recommendation for content. This document aims to help decision makers (from governments, data producers, national coordination bodies, etc.) to define their policy regarding the improvement of existing data and production of new geospatial data. It addresses digital data.

This document has annexes containing more detailed explanations targeting the technical people who will be in charge of implementing or adapting core data recommendations (e.g. for production purpose, as source of other standards, etc.).

2.1.2 Structure

The executive summary synthesizes the main conclusions of the Working Group A (WG A) process and results to develop the recommendation for content. It is meant mainly for high level decision makers.

The foreword reminds the general context of core data, the first step achieved by WG A (i.e. selecting core data themes), and it explains the general principles set by WG A to develop the recommendations for content of core data specifications for all selected themes.

The ‘recommendation for content’ document itself includes four chapters:

- Overview: it provides the general scope of the theme and describes the main use cases addressed;
- Data content: it provides the main characteristics of the recommended content, such as the list of core features and attributes (for vector data), as well as data capture and quality rules;
- Other recommendations: Coordinate Reference System, Metadata, Delivery;
- Considerations for future: this chapter addresses some key trends or significant user requirements that cannot be considered as core today but that might be considered in future.

The ‘recommendation for content’ document is meant for medium level decision makers. It is written in natural and not too technical language.

The technical explanations included in annexes describe the relationship between the recommendation for content and the corresponding INSPIRE specification, and contain any other appropriate information useful for this theme.

2.2 Core data context

2.2.1 Rationale for core data

The following background of harmonised pan-European data was identified.¹

Authoritative geospatial data are used to support both the implementation of public policies and the development of downstream services. Moreover, geospatial data are required to be homogenous to enable the implementation of public policies in a coherent and coordinated way among countries and at regional or global level. Likewise, significant opportunities exist if services developed by industry can be exploited without requiring country specific adaptation.

The INSPIRE Directive has set up the legal and technical framework for harmonisation of the existing data related to the themes in annexes I, II and III. INSPIRE specifications provide common data models that ensure a first step towards interoperability, however ensuring homogeneous content is outside their scope, as they contain no indication about levels of detail, very few recommendations about quality, and as most features and attributes are “voidable”, i.e. to be supplied if available or derivable at reasonable cost.

This background led the UN-GGIM: Europe Regional Committee to setup in 2014 the Working Group A on Core Data to deal with core data content and quality, production issues, funding and data availability.

Recommendations for content of core data will complement INSPIRE data specifications by defining the priorities on the core content that is encouraged to be made available in Europe in order to fulfil the main user requirements that are common to many countries, with focus on the SDG related ones.

Core data availability may be ensured either through upgrading of existing data when feasible or through production of new data when necessary.

2.2.2 Core data scope

In its first phase, WG A selected core data themes according to the following criteria: core data is the geospatial data that is the most useful, either directly or indirectly, to analyse, to achieve and to monitor the Sustainable Development Goals.

Among the 34 INSPIRE data themes, 14 have been considered as core including theme Geographical Names.

More information about the selection process and results may be found in document [‘Core Data Scope - Working Group A - First Deliverable of Task 1.a - Version 1.2’](http://un-ggim-europe.org/content/wg-a-core-data) on <http://un-ggim-europe.org/content/wg-a-core-data>

¹ Extract from the Report by the Preparatory Committee on the establishment of the UN-GGIM: Europe Regional Committee, European Commission Ref. Ares(2014)1491140 - 09/05/2014.

2.3 Document objectives and principles

2.3.1 Encouraging content availability

This deliverable provides recommendations for national governments and data producers, aiming to help them to define their priorities for enriching existing data or producing new data. This deliverable is meant mainly for data producers, however it defines the recommended result and target but not the production process.

2.3.2 Complementing INSPIRE

Core data specifications are built upon INSPIRE data specifications. On one hand, they often simplify INSPIRE by selecting core feature types and attributes and by restricting or clarifying the scope; On the other hand, they enrich INSPIRE by recommending specific levels of detail, quality rules and sometimes data model extensions. Besides, the INSPIRE common terminology is thoroughly used for naming core features and attributes.

Regarding the levels of detail, the ELF (European Location Framework) project terminology has been used. The ELF levels of detail are the following: Global, Regional, Master level 2, Master level 1, Master level 0. These terms are defined in the glossary.

Regarding delivery, core data may be supplied according to several ways. It is expected that, very often, the core data recommendations will be used to enrich and upgrade existing products. In this case, core data will be available through these improved products. Core data may also be delivered through INSPIRE conditions (specifications and services).

2.3.3 Status of core data recommendations

This document contains recommendations that are not legally binding. However, some recommendations are more important than others. This order is indicated as follows:

Core Recommendation X

It is first priority recommendation, considered as both necessary and achievable in principle. Ideally, it should encourage involved stakeholders to launch short-term actions (typically within a couple of years).

Core recommendations are usually addressing only technical aspects and are meant for the organisations in charge of producing this theme. The set of core recommendations defines the basic expectations on core data.

Good Practice X

It is second priority recommendation; if adopted, it will provide significant added value to core data; it indicates a relevant trend to be adopted as much as possible. It encourages involved stakeholders to take these recommendations into account in long term, if not possible in short term.

NOTE: some of these good practices may be quite easy to achieve and are already effective in some countries whereas some others may be more difficult to achieve. This is typically the case when these good practice recommendations involve other stakeholders in addition to the organisations in charge of producing this theme, and when they address not only technical aspects but also legal or organisational ones.

A “core data set” should contain the minimum data defined by the core recommendations (and ideally also by the good practices) of this deliverable but may of course contain more and/or better information.

2.4 Abbreviations

CRS	Coordinate Reference System
ELF	European Location Framework
EGM	EuroGlobalMap
EGN	EuroGeoNames
ERM	EuroRegionalMap
GN	Geographical Names
LAU	Local Administrative Units
SDG	Sustainable Development Goal
UNGEGN	United Nations Group of Experts on Geographical Names
UN-GGIM	United Nations initiative on Global Geospatial Information Management
WG A	(UN-GGIM: Europe) Working Group on Core data

2.5 Glossary

Endonym	Name for a geographical feature in an official or well-established language occurring in that area where the feature is located.
Exonym	Name used in a specific language for a geographic feature located outside the area where the language is widely spoken and differing in form from the respective endonym(s) in the area where the geographic feature is located.
Global	Level of detail defined by ELF: data to be used generally at scales between 1: 500 000 and 1: 1 000 000, i.e. mainly at international level
Master level 0	Level of detail defined by ELF: data to be used generally at scales larger than 1: 5 000; typically, data at cadastral map level, for local level actions.
Master level 1	Level of detail defined by ELF: data to be used generally at scales between 1: 5 000 and 1: 25 000; data for local level actions.
Master level 2	Level of detail defined by ELF: data to be used generally at scales between 1: 25 000 and 1: 100 000; data for regional (sub-national) actions.
Regional	Level of detail defined by ELF: data to be used generally at scales between 1: 100 000 and 1: 500 000; data for national or regional (European or cross-border) actions.
Script	A set of graphic symbols employed in writing or printing a particular language, differing from another set not only by typeface or font [UNGEGN Glossary 2007]

Script	A set of graphic symbols (for example, an alphabet) employed in writing the name, expressed using the four letters codes defined in ISO 15924 when applicable [INSPIRE TWG GN]
--------	--

2.6 Reference documents

INSPIRE Data Specification on GN– Technical Guidelines 3.1:

<http://inspire.ec.europa.eu/id/document/tg/gn>.

ELF Data Specification (chapter 5.3.6):

http://elfproject.eu/sites/default/files/ELF_DataSpecification_v0.12_20160328.pdf

3 Overview

3.1 General scope

Definition: Named places such as regions, localities, cities, suburbs, towns or settlements, or any geographical or topographical feature of public or historical interest [adapted from INSPIRE Directive 2007/2/EC].

Description: Geographical names are proper nouns applied to real world entities, i.e. to named places. In order to avoid duplication of production effort, core data scope is restricted to the named places and geographical names that are not present in other core data themes.

NOTE 1: for instance, generally speaking, named places include administrative units (already present in theme Administrative Units), rivers and lakes (already present in theme Hydrography) and streets (already present in theme Transport Networks). These named places should preferably be captured respectively in themes Administrative Units, Hydrography and Transport Networks.

NOTE 2: Theme GeographicalNames describes named places and their geographical names. Therefore, most of the following recommendations and good practices related to geographical names are relevant for all core data themes where geographical names are used.

NOTE 3: In practice, the general scope of core theme Geographical Names is restricting the scope of the INSPIRE theme Geographical Names but as INSPIRE, it addresses the multi-lingual, multi-scriptural context of Europe. A script is the set of graphic symbols (e.g. an alphabet) employed in writing or printing a particular language (cf glossary).

More detailed comparison with INSPIRE is available in the annex A of this document.

3.2 Use cases

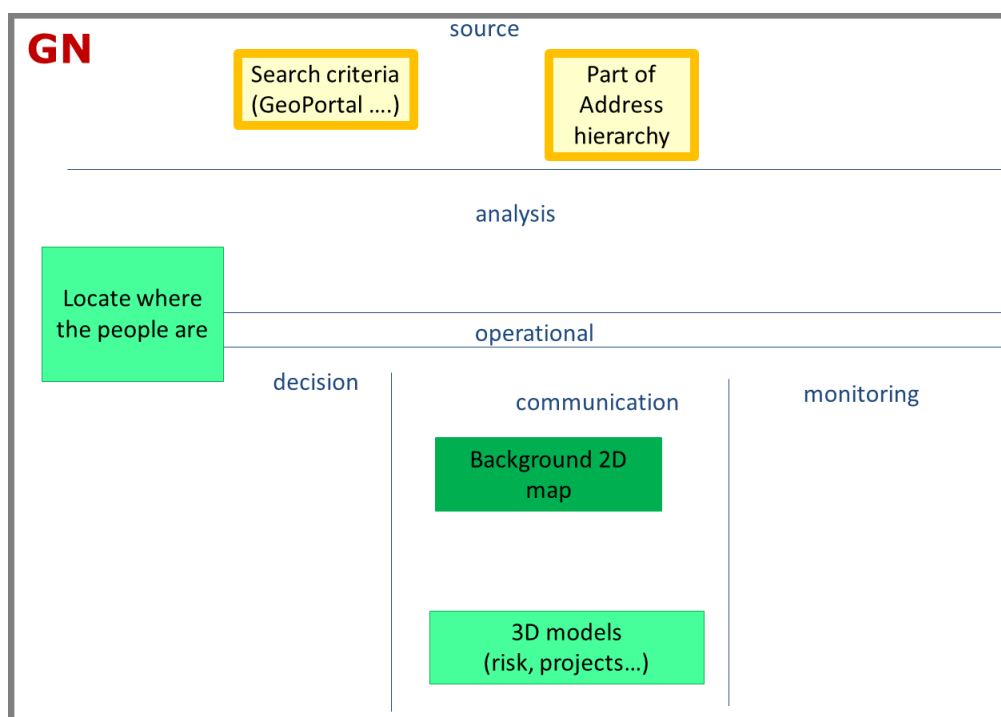


Figure 1: map of use cases for theme Geographical Names

Generally speaking, 'Geographical Names' data is mainly used for identifying and locating geographical features, i.e. as search criteria in gazetteer services, geo-portals, geo-catalogues etc. In other words, one of its main uses is for geocoding process: a geographic data set containing named places with their geometry and name(s) enables the transformation of an indirect location (e.g. a geographic name, such as a village name) into a direct location (e.g. a point with its coordinates, such as latitude and longitude).

Regarding core data, it is expected that this role of search criteria will be ensured by the geographical names included in various core themes, mainly Administrative Units, Hydrography, Transport Network and of course the theme Geographical Names itself.

The other general use case of this theme is for mapping. Geographical names are a key element of any kind of maps: no one would understand a map without geographical name; this states for any kind of graphical representation (background 2D maps, ortho-images, 3D models). Once again, this role will be ensured both by theme Geographical Names itself and by the names included in other core themes.

In addition, as theme Geographical Names includes populated areas, i.e. settlements, cities, etc., that cannot be found as such in any other core theme, this theme is of great interest for many SDG related use cases: in the analysis phase or for more operational purposes, such as raw locating where the people are (e.g. to ensure them accessibility to services or to assess the human pressure on environment or to coarsely assess the impact of a risk or pollution), the data on populated areas will be quite useful.

Theme 'Geographical Names' is part of the basic geographic equipment of a country and enables location of various phenomena on cartographic products at any scale.

4 Data content

4.1 Features types and attributes

Core Recommendation 1

Core data should include feature type Named Place with following attributes:

- **geometry**
- **unique and persistent identifier**
- **classification (type)**
- **population indication (for populated places)**
- **geographical name(s) with its spelling and with information on its language, status, nativeness and (if relevant) source.**

NOTE 1: the ELF project has extended the INSPIRE data model of theme Geographical Names, adding some attributes. Two of them are considered here as core information: population indication (for populated places) and the EGNtype for the classification of named places.

NOTE 2: the EGN (EuroGeoNames) initiative developed a two levels hierarchic European object/feature classification of named places; the EGNtype offers a good compromise between detailed information and harmonised information. Therefore, it is advised to use this EGNtype for

classification of core data on named places or at least to use a code list that can be easily matched with the EGN classification².

NOTE 3: the estimated population may be provided as a single number (e.g. 10 000 inhabitants) or as an interval (e.g. between 9000 and 10 000 inhabitants). This attribute aims to give an idea of the settlement importance for mapping purposes but is not suitable for reliable population analysis; reliable population analysis is expected to be carried out using administrative or statistical units and census data.

NOTE 4: the attributes describing the name (language, nativeness, status, source) should help users to decide on which name(s) are the most relevant to be displayed on a map. The information about “source” is relevant if some sources are considered as more reliable than others.

NOTE 5: nativeness of the name indicates if the name is an endonym (name used in the named place’s language) or an exonym (name used in another language than the named place’s one). For instance, “Londres” is the French exonym for “London” whereas “London” is the national English endonym.

NOTE 5: significant information is the script, i.e. the set of graphical symbols used to write the names. It is expected that in most datasets, the script will be the same for all the names; in this case, it is enough to document the script in the dataset metadata. In case of a dataset combining names with different scripts, it is advised to document the script for each name present in the dataset.

4.1.1 Geographical name

Core Recommendation 2

In feature type Named Place, the various geographical name(s) should be captured with their spelling in their national script(s).

4.1.2 Temporal aspects

Once features have been captured, it is recommended to keep them in the data base, even after their end in the real world

Good Practice 1

It is recommended to manage the history of Named Place features, using the mechanism provided by the INSPIRE data specifications: versioning and life-cycle attributes.

NOTE: the versioning and life-cycle attributes enable change-only updates.

4.2 Geographical extent

Core Recommendation 3

Core data on Geographical Names should be available on whole territory, both land and sea.

² Current Tables and their attributes of the EuroGeoNames database: http://www.eurogeographics.org/sites/default/files/EGN_DB_tables_attributes1.pdf or the Feature classification developed for the EuroGeoNames Project. It consists of 8 main classes and 27 sub-classes: <https://wiki.gdi-de.org/display/wgtdfg/F3.3+Reference+materials+and+classifications>

NOTE 1: in practice, land features should be given first priority (due to the populated named places with persons-related use cases) but named places in coastal and sea areas are also necessary.

NOTE 2: Open sea is not part of a national territory. Names do not have official status; however, they are of key interest and some arrangement has to be found among the international community in order to ensure the capture and delivery of data on these open sea named places.

4.3 Levels of detail and data capture rules

Theme Geographical Names is required at various levels of detail, from large scale (Master level 1) to Global. The data capture rules will mainly depend of course on the level of detail but also on the nativeness of the name, i.e. if the name is an endonym or an exonym.

4.3.1 Case of endonyms

Core Recommendation 4

Core named places should include first populated areas and then the natural named place of main interest that are not in other core themes (mainly landform), land cover).

NOTE 1: Populated places are of key importance for many SDG related use cases.

NOTE 2: the named places of main interest are the named places referring to Master level 1 density and to less detailed levels (Master level 2, Regional, Global). Micro-toponyms present only at Master level 0 are not considered as core.

Core Recommendation 5

Named places with their endonym(s) should be captured at large scale (Master level 1 level). Other levels of detail should be derived from the large scale core data.

NOTE 1: the derivation of less detailed data implies selection of named places (for Master level 2, Regional and Global) and generalisation of geometry (for Regional and Global level).

NOTE 2: the ELF project focussed on Master level 1, including also Regional and Global levels. Based on this experience, this deliverable is proposing capture or selection rules for these 3 levels of detail.

NOTE 3: data providers may offer their users Master level 2 or may let up to them the selection process.

Good Practice 2

Named places should be captured with their “true” geometry, i.e. generally as a surface or multi-surface.

NOTE 1: in real world, the named places that are not yet present in themes Transport or Hydrography generally have a surface extend. Only very few geographical features can be represented correctly by a point co-ordinate (e.g. sign posts, small monuments...). Therefore, it is advised to provide them as surface or multi-surface at any level of detail. For Regional and Global levels, the surface may have to be generalised but it should be kept in the GN database as surface to enable various processes (selection of named places for cartography, spatial analysis ...). If required, transformation into points may be done automatically, for mapping, by using GIS tools.

NOTE 2: in existing data, it often occurs that named places are represented by points, locating the name label centre; WG A advises to move from a cartographic viewpoint to a topographic viewpoint, where the named place represents a real world physical phenomenon.

NOTE 3: in the real world, many named places do not have a clear definition (e.g. big areas such as mountain range or sea). These named places may be represented with a very fuzzy geometry, for instance just the envelop of the current named place label on the map. In these cases, “true” geometry should be understood as the estimated extent of the named place.

NOTE 4: gazetteer services are usually providing information based on point or bounding box geometries. These simplified geometries may be automatically derived from the surface geometry recommended in Good Practice 2.

Good Practice 3

All the core named places that are of current use within a country should be captured at Master level 1.

NOTE 1: this recommendation implies that big areas such as mountain range or landscape have also to be captured at Master level 1, even if the related feature is not displayed or labelled on large scale maps.

Good Practice 4

Ideally, named places at Regional and Global levels should be automatically derived from Master level 1 data.

NOTE 1: Automatic selection and derivation of named places ensures consistency between various levels of detail and avoids parallel updating.

NOTE 2: selection criteria for Regional and Global levels have to be adapted to the context of each country, however some basic selection rules are proposed in following notes to promote better harmonisation in Europe.

NOTE 3: minimum selection should include the populated places that fulfil at least one of the following criteria

Regional	Global
Be residence of authority of an administrative unit at any level	Be residence of authority of an administrative unit level LAU1 or upper
Have more than 100 habitants	Have more than 1 000 habitants
Be landmark along the road network (especially along roads without number)	
Be of touristic, economic, political,..., interest	

Good Practice 5

To enable the selection of the populated places that are residence of authority, it is recommended to capture the association between the Administrative Units and the Named Places.

NOTE 1: Residence of authority is centre of national or local administration or i.e. the place from which the administrative place is administrated.

Minimum selection should also include other significant named places, such as mountain ranges, highlands, plains, valleys, gorges, peaks (in landform), forests (in land cover), seas, bays or fjords (in hydrography).

Minimum area size is fixed to 0.4 km² for Global level and to 0.04 km² for Regional level in ELF data specifications; these thresholds may be also used as general guidelines for selection criteria of core data. But of course, smaller size named places may be selected (e.g. peaks) if significant by other criteria (e.g. touristic interest).

Core Recommendation 6

All endonyms that are in current use within the country should be captured.

NOTE 1: this rule addresses the selection of names whereas the previous ones address the selection of named places.

NOTE 2: according to this recommendation, historical names are not considered as core.

4.3.2 Case of exonyms

Exonyms are used to make publications in foreign languages and are of key interest as search criteria in cross-border, pan-European or even global geocoding services. Typically, they are required for setting up a European gazetteer.

The exonyms that are the most relevant for core data are mostly addressing significant features, such as countries, big cities, long rivers, chains of mountains, seas, etc. Therefore, they are mainly related to Regional and Global levels.

Good Practice 6

Member states should store and manage as digital data the exonyms they are in current use in their national languages and should make them available for European or international gazetteers.

NOTE 1: regarding exonyms, the scope of theme Geographical Names is not restricted to named places not yet in another core theme but should include all the exonyms in current use.

NOTE 2: Easy access to exonyms will facilitate management of geographic information at continental and international levels and so will help to achieve or monitor some SDGs.

NOTE 3: This good practice implies that Member States should maintain data on names that are generally outside their territory, being located in foreign countries.

4.4 Quality

4.4.1 Completeness

Core Recommendation 7

At Master level 1, completeness should be ensured for populated places, i.e. all populated places should be provided ideally with all their names (endonyms) that are in current use.

NOTE 1: Completeness rate of 100% may be difficult to achieve, especially in countries with scattered habitat but the aim should be to maintain completeness of 95% or more of the populated places. The target includes all the populated places of interest for mapping or geocoding, such as cities, villages,

neighbourhoods, hamlets, isolated buildings. In other words, names of single buildings in urban areas having also a classical address (house number + street name) are considered of minor interest and are not in the target of this rule;

NOTE 2: at Regional and Global levels, some selection of populated places has to be done, based on the criteria proposed in previous chapter.

4.4.2 Thematic accuracy

Core Recommendation 8

Great care has to be taken when capturing the spelling of names, especially for populated places.

The names of Named Places (and especially of populated places) are used by many users and applications, e.g. they may be part of address data. Therefore, it is of key importance to have reliable names in core theme Geographical Names that may be used as reference data by all stakeholders.

More generally, it is advised to provide officially standardized name forms as it enables a consistent politically correct use of geographical names. This is more especially necessary for exonyms that aim to be used in cross-country context.

Good Practice 7

Exonyms should be provided in a standardized form.

NOTE 1: for endonyms, the main target is to ensure completeness. It is advised to provide all the names, even if not (yet) standardised. The status of the name (if official/standardised or not) should be documented.

Core Recommendation 9

The attribute “estimated population” of populated places has to be captured at relevant level of granularity.

NOTE 1: the estimated population may be provided as a single number or as an interval. However, to be useful, this interval should not be too wide! The purpose of this attribute is to give an estimate of the population (e.g. mean value +/- 20%) and not just to provide a coarse classification of populated places (e.g. [10 000, 50 000]).

4.4.3 Geometric accuracy

In real world, many named places do not have a clear definition therefore the accuracy of their boundaries is not always really meaningful.

Good Practice 8

It is recommended to document the geometry reliability, e.g. good, medium, undefined.

Good accuracy applies for entities well-delimited in real world (e.g. an island or a mountain pass) and captured with an accuracy of some meters in the database. Medium accuracy applies when the geometry is significant (i.e. aims to represent the geometry of the real-world entity) but fuzzy: for instance, populated places may be represented by different geometries, depending on the various possible ways to decide on their boundaries, with up to a pair of hundred metres of difference between these different representations. Undefined accuracy applies when the geometry does not

aim to represent the geometry of the real-world entity but just to provide a very rough location: for instance, a sea or a chain of mountains represented by a point.

5 Other recommendations

5.1 Coordinate Reference System (CRS)

5.1.1 Case of 2D data

Good Practice 9

Core data should be stored and managed in a CRS based on datum ETRS89 in areas within its geographical scope, either using geographic or projected coordinates.

NOTE 1: geographical scope of ETRS-89 excludes over-sea territories, such as Canary Islands or French Guyana or Madeira Islands and Azores Islands. In these cases, it is recommended to use a CRS based on ITRS (International Terrestrial Reference System).

NOTE 2: storing and managing data in CRS based on international datum facilitates the import of measures from modern sensors, ensures that data is managed in a well-maintained geodetic framework and of course, facilitates the export of data into international CRS (e.g. those mandated by INSPIRE).

NOTE 3: if core data at regional and global levels has to be provided as a single data set on an area including over-sea territories, it is recommended to use as CRS geographic coordinates with any realisation of the International Terrestrial Reference System (ITRS), known as International Terrestrial Reference Frame (ITRF). At small or medium scales, all ITRS realisations can be considered as equivalent, as deviations between them are negligible compared to data accuracy.

5.1.2 Case of 2,5D or 3D data

Not applicable for theme Geographical Names. It is expected that this theme will be provided as 2D data.

5.2 Metadata

Good Practice 10

Core data should be documented by metadata for discovery and evaluation, as stated in the INSPIRE Technical Guidelines for metadata and for interoperability.

NOTE: this is a legal obligation for the Member states belonging to the European Union. For the other countries, this is a way to make their data easily manageable by transnational users.

Good Practice 11

The attributes describing the name (language, status, source) should help users to decide on which name(s) are the most relevant to be used under given circumstances (language of publication, audience addressed, ...), e.g. to label a map. Their relevance differs from one country to the other and depends on the use cases supported. For this purpose, some explanation should be given about the values of these attributes.

NOTE 1: the best way to provide these explanations is to include them in a national named places data specification document. Ideally, the documentation should be available both in official language(s) of the country and in English.

NOTE 2: for instance, in INSPIRE, the status can take the values: official, standardised, historical and other. Typically, it would be of interest for users to understand the rules that decide if a name is “official” or “standardised”.

NOTE 3: regarding the languages, it is of interest to know the context of each language (official or unofficial, local or national, etc.) and about which (groups of) stakeholders in practice are using which language for named places.

NOTE 4: it is reminded that the information about script should be documented either as metadata if the script is the same for all names or as attribute of the name spelling if different scripts are used in the given dataset. This information will help to decide which spelling is the most adapted.

5.3 Delivery

It is expected that core data will be made available through improved existing products (or new products) or as INSPIRE data, and perhaps as specific core products (delivery issues still have to be investigated by the working group).

5.3.1 Combination with other themes

This document defines some minimum content for theme Geographical Names, aiming to avoid duplication of data capture. It is why the scope is limited to the names and named places that are not yet included in other core themes. However, there is generally big interest for a national gazetteer compiling all the geographical names within a country.

This gazetteer might be derived from core themes Geographical Names, Administrative Units, Addresses, Hydrography, etc.

Good Practice 12

It is advised to provide a national dataset combining all the geographical names from various themes and from various data producers. An alternative solution is to provide a geocoding service based on all the geographical names of the country.

5.3.2 Facilitating international uses

Good Practice 13

Core data should be made available according to the INSPIRE Technical Guidelines for interoperability, for metadata and for services.

NOTE: this is a legal obligation for the Member states belonging to the European Union. For the other countries, this is a way to make their data easily manageable by transnational users.

The good practice above applies to all core themes and ensures some minimum interoperability in Europe. However, regarding Geographical Names, there are remaining issues due to the multi-scriptural context and to management of exonyms.

For international use, the geographical names in Greek or Cyrillic alphabets may have to be romanised (transliterated or transcribed) in Latin script.

Good Practice 14

The concerned countries should make possible the transliteration system for their geographical names to be applied, either by providing directly the name spelling(s) in Latin script or by providing the transliteration system to be applied. If the transliteration system refers to the whole dataset, it may be provided as metadata.

NOTE: The transliteration should be conform with the systems approved by the UNGEGN (<http://www.eki.ee/wgrs/>)

The use cases related to set up a European or international gazetteer are particularly interested in the unambiguous combination of official endonyms (provided by public authorities, such as the National Mapping and Cadastre Agencies) linked to standardized exonyms and to other variant names (provided by language communities/names boards). This combination is a clear user requirement.

Good Practice 15

Provider of on Geographical Names data and services should cooperate in order to ensure a reliable combination of exonyms to the related national endonyms.

6 Considerations for future

6.1 Improving the geometric representation

Good practice 3 recommends to capture the “true” geometry of named places. However, it is recognized that this true geometry may be quite fuzzy for large areas named places, such as seas or mountain ranges. In a first step, it is suggested to capture just the envelop or bounding box around the named place label on the map: though very inaccurate, this surface representation will be much better than the current point representation, enabling more use cases.

However, this representation may be improved in future, typically by collaborative work: expert users, such as physical geographers, might propose more meaningful geometric representations. In addition, for the large areas named places crossing several countries, it will become possible to check if they match or not at the international boundaries. There is also place for progress.

6.2 Importance of named places

Mapping is among the main use cases of Named Places. Many users want to make their own maps, either as paper or as screen maps. Significant information is the importance of the named place that enable relevant selection criteria (at any scale) and relevant choice of the label font.

This importance may be provided based according to two main principles:

- Selection of the named place for maps at some given scale(s): cartographic viewpoint

- Estimation of the importance of the named place in the real-world: topographic, database viewpoint

The first option optimises the choice of named places for a given map, at a given scale but is very specific, both to a territory (and so difficult to harmonise across Europe) and to a scale or limited set of scales.

Therefore, there is an interest for a topographic view on named places focusing on the importance of the real world object, instead of a pure cartographic view. This document is promoting the capture of quantifiable criteria measuring the **importance of the named place in real world**, such as its area or its population (for populated places). However, these criteria are not the only ones, e.g. touristic interest also influences the importance, the historical or economic relevance of a named place.

The geographic community should encourage further research and knowledge exchange in order to define relevant objective criteria enabling a detailed hierarchy of named places. Based on these criteria, it would then be possible for cartographers to define common rules for a wide range of maps, at any scale and on any part of Europe.

6.3 Pronunciation of names

Pronunciation of names would enable lots of innovative applications, using the voice recognition. However, currently, there is almost no existing data on this topic.

Pronunciation of names may be provided as audio record or in International Phonetic Alphabet. First option is more directly usable but may be difficult to achieve; the second option is more feasible but some extra-work would be required to enable voice recognition.

Pronunciation of names is clearly not core data for today but it might become ambition of core data within a few years. Experiences of production and cost-benefit assessments of the two potential technical options should be encouraged.

6.4 European classification of named places

The benefits of the development of a feature type and classifications list for Europe are extensive, and include the ability for (naming) authorities to identify areas for increasing the scope of their geographical names data collection methods.

To understand geographical names data it is necessary to specify what type of feature is being referred to. This leads to a range of issues including how finely feature types are differentiated. When integrating data from multiple sources, different feature type classifications increase the complexity of any interpretation or searching process.

Up to now feature type categories for official geographical names data have been developed independently by different authorities. Other datasets that act as gazetteers, e.g. postcodes or census districts, also have an implicit feature type, typically applying to the whole data set. Thus, there may exist as many feature type classification schemes as there are geographical names data sources.

There are extensive issues to be resolved with regard to linking, comparing and aligning the multiple feature type lists available internationally. Two approaches to resolving the feature types issue are

described for the international level and may apply for Europe as well: the first is a proposal to develop a comprehensive feature type catalogue, the second is to develop a set of feature type classifications to which feature types from source geographical names data sources can be referenced and linked.

7 Annex A: Relationship with INSPIRE

7.1 Data model

The UML models provided in this annex are only graphical illustrations of the core recommendations and of the good practices present in this document.

For theme Geographical Names, the UML illustrations are coming from an extension of the INSPIRE data models.

The recommendations for content are represented by highlighting the selected attributes in the following way:

Core recommendation



Good practice



Core Recommendation 1

Core data should include feature type Named Place with following attributes:

- geometry
- unique and persistent identifier
- classification (type)
- population indication (for populated places)
- geographical name(s) with its spelling and with information on its language, status, nativeness, and (if relevant) source

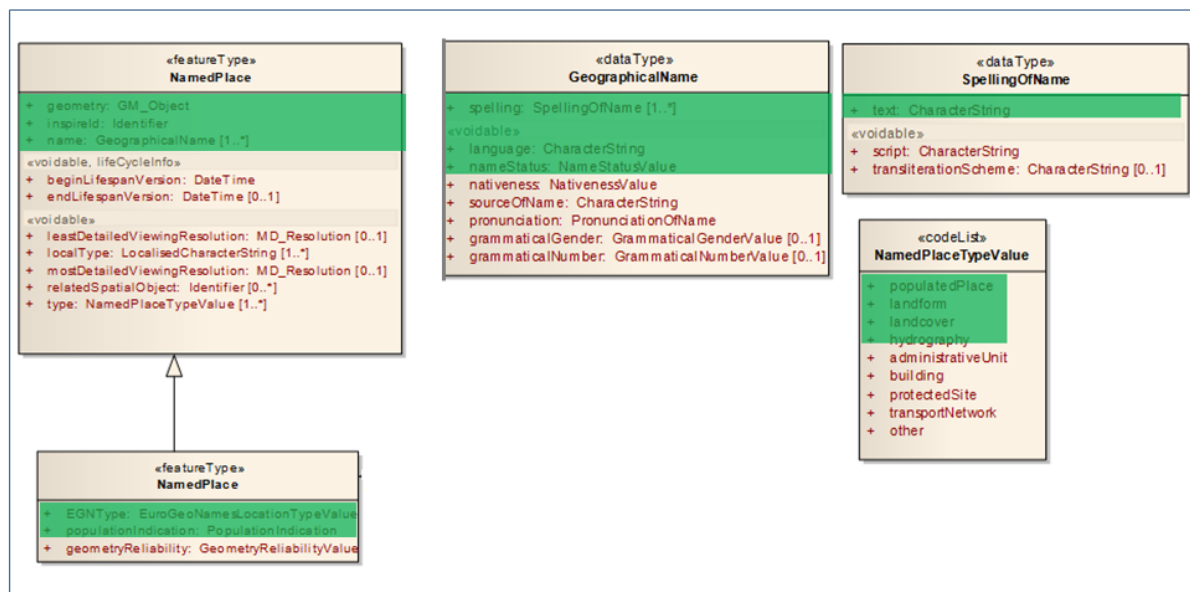


Figure 2: Extract and Extension from the INSPIRE data model for theme Geographical Names

WG A recommends the additional attributes (EGNType and populationIndication) of the child feature type "NamedPlace". These additional attributes are also included in the ELF data model.

The EGNType is the category of named places, according to the EGN (EuroGeoNames) project. It should be provided, using the possible values of the hierarchical code list EuroGeoNamesLocationTypeValue. It is available on <http://www.locationframework.eu/codelist/>

Good practice 1

It is recommended to manage the history of Named Place features, using the mechanism provided by the INSPIRE data specifications: versioning and life-cycle attributes.

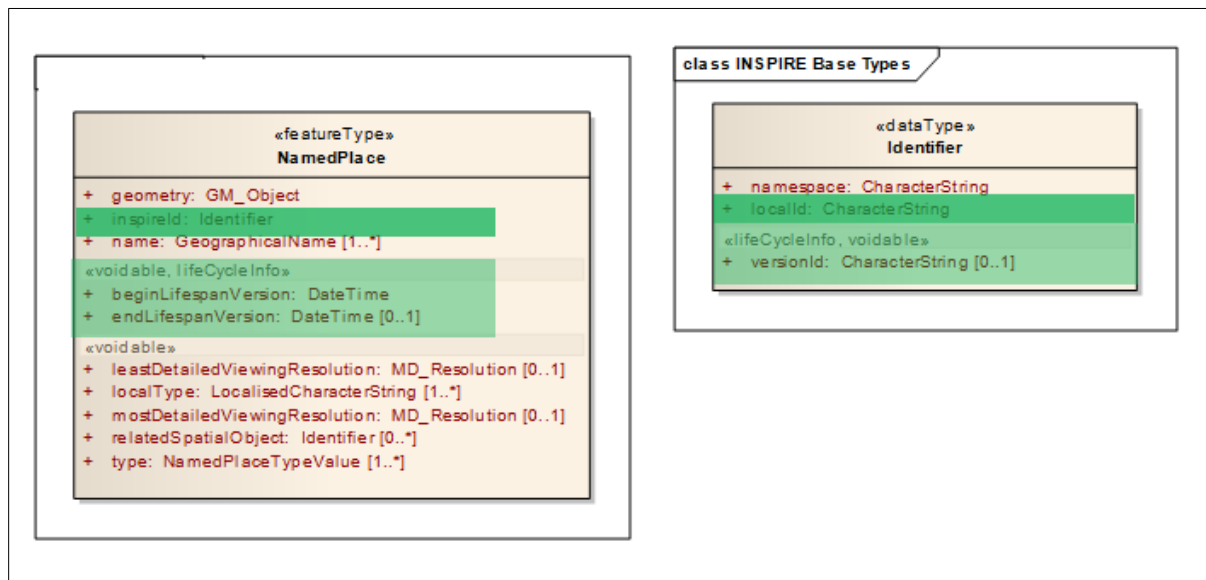


Figure 3: Temporal Aspects for theme Geographical Names

NOTE: in the data type “Identifier”, the namespace is generally a constant value that does not need to be stored at feature level.

Good Practice 5

To enable the selection of the populated places that are residence of authority, it is recommended to capture the association between the Administrative Units and the Named Places.

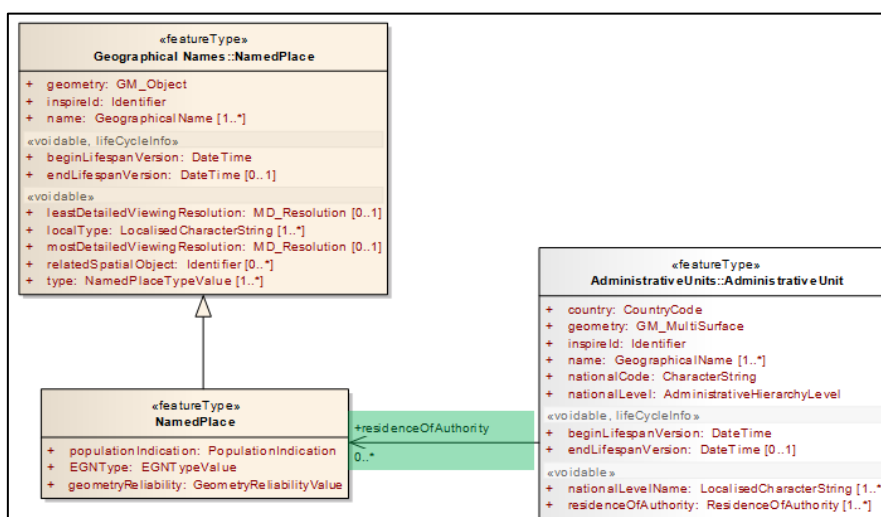


Figure 4: Named Places and Administrative Units

Good practice 8

It is recommended to document the geometry reliability, e.g. good, medium, undefined.

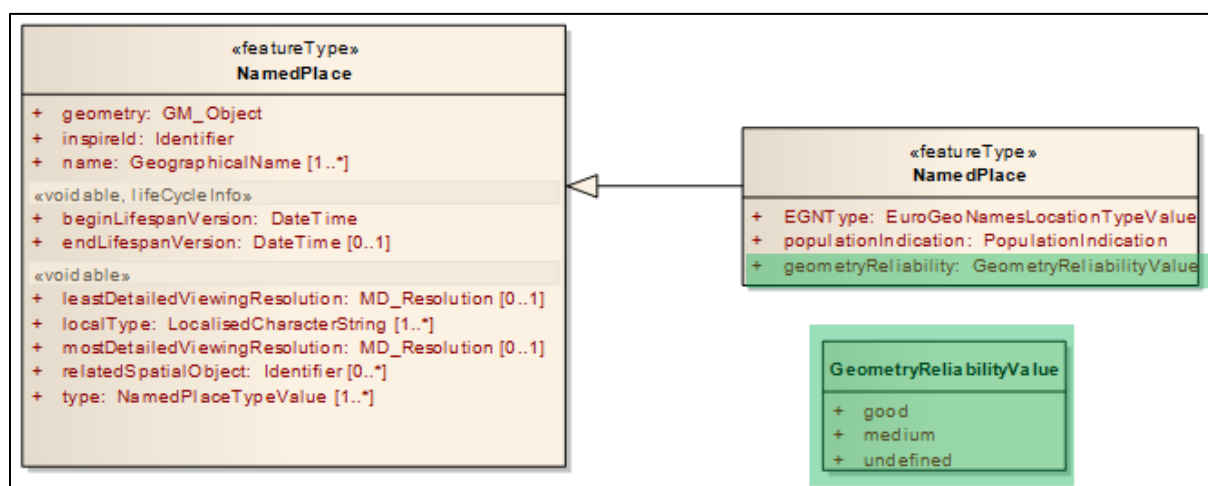


Figure 5: Geometry Reliability

Definitions and examples of values “good”, “medium” and “undefined” have been provided in chapter 4.4.3.

7.2 Other

7.2.1 Scope – Data capture

The INSPIRE theme GN includes any dataset with information on geographical names whereas core data theme GN is limited to the named places that are not captured in other core themes (Administrative Units, Hydrography, Transport ...). This is quite consistent with the INSPIRE principle “capture once, use many”: as core data is production oriented, useless duplication should be avoided but this should not prevent producers to supply more data for INSPIRE theme GN (as INSPIRE is delivery oriented).

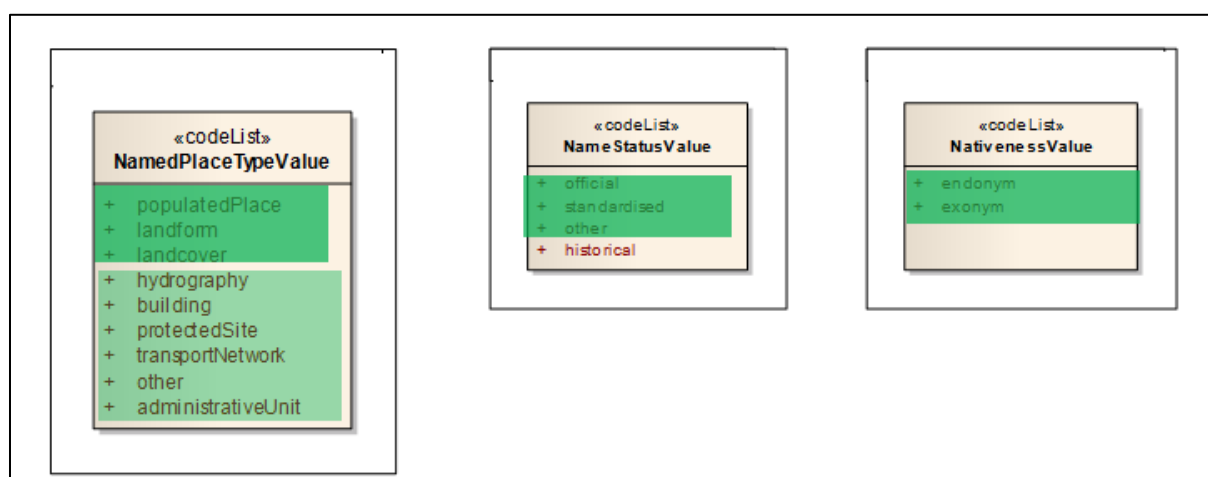


Figure 6: Scope of theme ‘Geographical Names’

NOTE 1: core data gives priority to populated places, landform and land cover (core recommendation 5) but other named places are also of interest. For instance, not all hydrography related named places are present in theme Hydrography (e.g. bay, glacier) and so, deserve to be captured in theme GeographicalNames.

NOTE 2: core data recommends to capture the names that are in current use within the country (core recommendation 6); therefore, historical names are not considered as core data. However, they may of course be captured, as optional co

In addition, INSPIRE focuses on the description of names rather than the description of spatial objects (with particular description of characteristics of names) whereas, the core data approach is more balanced: GN core data being dedicated not to name specialists but to SDG related use uses, the description of names is limited to two practical issues: capturing names in multi-lingual, multi-scriptural context and helping users to select the most relevant names to be displayed on a map.

7.2.2 Level of detail -accuracy

Whereas there is not any recommendation in INSPIRE about levels of detail or about accuracy, core data is recommending to be captured at Master level 1 and to be derived at least for the Regional and Global levels of detail.

In addition, core data is promoting better accuracy of GN data by recommending to capture the true geometry of named places, i.e. in general, to capture named places as surfaces rather than as point (what is the most frequent case currently).

8 Annex B: Methodology

Core data specifications have been elaborated based on one hand on user requirements (with focus on the ones related to SDG) and on the other hand on INSPIRE data specifications.

Core data specifications have been elaborated based on one hand on user requirements (with focus on the ones related to SDG) and on the other hand on INSPIRE data specifications.

The INSPIRE data specifications have benefited from the ELF experience: ELF developed services related to the two main use cases of theme GN:

- The ELF GeoLocator is a geocoding service, with GN being used as search criteria
- The ELF topographic BaseMap, with GN being used as main source of labels for the map, at various levels of zoom.

Based on the requirements of the topographic BaseMap, of the GeoLocator and on current content of European products (ERM, EGM), ELF also extended the INSPIRE data model with 3 attributes related to the EGNtype, the population of a named place and to the reference name.

In addition, ELF provided a detailed state-of-play of available data among the NMCA's that were partners of the project



Figure 7: Stepwise approach to core data

The ELF experience has helped the UN-GGIM: Europe WG A mainly in the selection of the core semantic content (attributes) and in the selection of levels of detail.

This has been completed by more specific investigation within a few National Mapping Agencies, discussion with name experts and of course, between WG A members.

For instance, the name experts explained that the ELF proposal for 'referenceName' cannot be supported as this approach is not compliant with resolutions and recommendations of the United Nations / UNGEGN: the designation 'referenceName' is politically incorrect and should not be used even not to indicate the name suitable for cartographic representation.

This document has also benefited from the wide experience of European products (ERM, EGM).