ECP 2005 GEO 038026 EGN

EGN

UML representation of the Conceptual schema & documentation

Deliverable number	*D4.2 e*
Dissemination level	*Public*
Delivery date	*31st March 2009*
Status	*Final*
Author(s)	*EDINA, University of Edinburgh*

e*Content*plus

This project is funded under the *e*Content*plus* programme[1],
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

[1] OJ L 79, 24.3.2005, p. 1.

1 Introduction

1.1 Project Aim

The eContentplus matched funds EuroGeoNames (EGN) Project is developing a European Infrastructure for the exploitation and management of geographical names. In popular terminology, it is building a federation of distributed digital gazetteers[2] maintained by the relevant official national authorities (hencewith referred to as the National Mapping and Cadastral Agencies (NMCAs) for shorthand).

It is envisaged that a European gazetteer infrastructure will assist in a variety of domains and is predicated upon a range of formal Use Cases that have been identified as part of the project[3]. Fuller background on EGN and other deliverables are available from the EGN website at:

www.eurogeonames.com

1.2 Purpose of document

This document presents the UML representation of the EGN Conceptual schema (a map of concepts and their relationships). It represents the culmination of an iterative process of development under the auspices of the EGN Project. It has been produced by the EGN Consortia in league with its wider Reference Group (data providers) and within the context of INSPIRE best practice guidance.

Earlier versions of this schema (May 2007 and July 2008) were publicly published as EGN Deliverable D4.2. As noted in these predecessors, "the Conceptual schema should be considered informative rather than definitive at time of writing as subsequent developments within INSPIRE and during implementation may necessitate model revision". This document reflects the latest and *final* revision of the Conceptual Schema within the project funding period.

1.3 Target audience

This document is intended for those interested in the EuroGeoNames project including:

- The EGN Consortium
- The EGN Reference Group
- The EU Commission and eContentplus programme
- EU spatial data initiatives, specifically INSPIRE
- General geographic information consumers in Europe
- The Global gazetteer community
- The general interested lay reader

1.4 Status

This document is the final version as of March 2009.

---

[2] The term 'gazetteer' is used here as short hand for a more comprehensive compendium of geographical named features than is implied from the popular conception of a 'gazetteer' e.g. "A **gazetteer** is a geographical dictionary, an important reference for information about places and place-names..." [Wikipedia entry - http://en.wikipedia.org/wiki/Gazetteer]. The EGN project thus alludes to much richer content than might be casually inferred from the popular definition.

[3] The methodology employed by the project was informed by the findings and recommendations of the INSPIRE initiative and the RISE project of EuroGeographics. Specifically, a modified version the RISE Use case template has been used to document the formal geographical names Use Cases.

1.5 Change summary

Version Revision	Date	Author	Notes
0.1	12/11/07	JSR, AC, AS	Updates to reflect ISO expert comment
0.2	23/11/07	JSR, AC, AS	Added fuller textual overview post Paris Workshop and updated Gender data type
0.3	05/12/07	PGZ, JSP, SAB	Update UML Image, add UML basic description
0.4	19/12/07	AS, JSR	Modify Feature classification and UML
0.5	31/01/08	JSR	Minor changes to reflect changes to UML necessitated by light implementation issues
0.6	03/03/2009	JSR	First cut of final redraft and update.
0.7	25/03/09	PGZ, JSP, SAB; JSR	Revision of final redraft and update

1.6 Related documents

Title	Author	Date	Version
EGN Data Models Documentation (D4.1)	**EDINA**	**02/07**	**Final**
Conceptual Schema and documentation (D4.2)	**EDINA**	**31/05/07**	**Final (but subject to revision)**
Conceptual Schema and documentation (D4.2a)	**EDINA, BKG**	**14/08/2007**	**Final (but subject to revision)**
Conceptual Schema and documentation (D4.2b)	**EDINA, BKG**	**30/10/2007**	**Final (but subject to revision)**
Conceptual Schema and documentation (D4.2c)	**EDINA, BKG**	**31/01/2008**	**Final (but subject to revision)**
Conceptual Schema and documentation (D4.2d)	**EDINA, BKG**	**30/06/2008**	**Final (but subject to revision)**

1.7 Distribution

Public

1.8 Contact details

James S Reid

EDINA
University of Edinburgh
Causewayside House
160 Causewayside
Edinburgh EH9 1PR
E-mail: james.reid@ed.ac.uk
Tel: +44 (0)131 651 1383

2 A Note on Standards

The following standards have been referenced in the development of the EGN Conceptual Schema.

ISO19100 Series:

ISO 19107:2003 specifies conceptual schemas for describing the spatial characteristics of geographic features, and a set of spatial operations consistent with these schemas. This used for specifying geometries of footprints.
ISO 19108 – for temporal characteristics, especially EX_TemporalExtent
ISO 19109 - GI rules for application schema
ISO 19112 - Spatial referencing by geographic identifiers (subsumed within proposed changes to INSPIRE model)
ISO 19115 – Metadata (for service discovery level) and especially PT_FreeText
ISO 19119 - Services
ISO 19136 - GML Annex-E (GM_Object inheritance)
ISO/TS 19139 PT_FreeText from the conceptual model specified in ISO/TS 19139 is intended to be used as the data type for multilingual text and supplants the informative version of ISO 19115.

ISO 3166 is used for country codes.

The choice of language codes raises some issues. As no code list has been assessed as ideal in the context of INSPIRE, a pragmatic approach had to be found. The code lists under discussion were ISO 639-2 and ISO 639-3 in its current editions. Although ISO 639-3 is the most comprehensive list, the use of it within EGN would have led to the problem that no languages families/groups are allowed. This would,in turn, have caused a data model mapping problem for some NMCAs. Although ISO 639-2 seemed to solve the technical problem for EGN, specific information for some individual languages is scarificed. Crucially however, it supports at the most important languages (comprising minority languages) and satisfies most of the European NMCAs without re-engineering the data sources which are currently available.
In the INSPIRE data specification on geographical names (v2.0) ISO 639-2 is recommended for pragmatic reasons too (codes mandated in INSPIRE Implementing Rule on metadata, codes used in most national databases), as no other solution is fully tractable. Additionally, the INSPIRE Thematic Working Group (TWG) on Geographical Names strongly recommends to push ISO for a useful combination of the various versions of ISO 639.

ISO 8601 is used for representation of dates.

The International Phonetic Alphabet (IPA) provides the notational standard for the phonetic representation of places.

3 Notation

3.1 UML - Unified Modelling Language

The following paragraph gives an overview of the Unified Modelling Language (UML) notation the EGN data model is based on. The terms have been defined in ISO 19115 "Geographic information – Metadata" but are summarised here to aid in explanation for readers less familiar with UML notation.

The UML notations used in this EGN data model are:

- Association,
- Aggregation and
- Generalisation (see Figure 1).

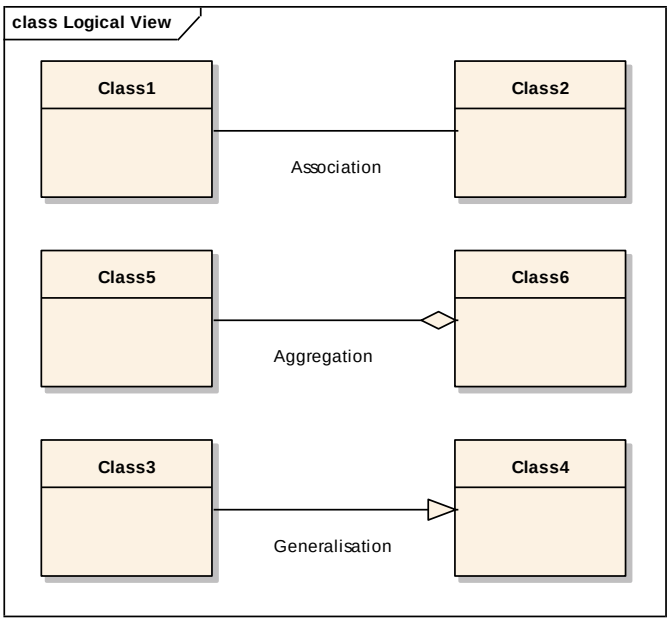


Figure 1: UML notation used in EGN data model

The three types have different semantics. An ordinary association shall be used to represent a general relationship between two classes. The direction of an association must be specified. If the direction is not specified, it is assumed to be a two-way association. If one-way associations are intended, the direction of the association can be marked by an arrow at the end of the line ("→").

The aggregation shall be used to create part-whole relationships between two classes. An aggregation is a relationship between two classes in which one of the classes plays the role of container and the other plays the role of a containee.

A generalisation is a relationship between a super class and the subclasses that may be substituted for it. The super class is the generalised class, while the subclasses are specified classes.

The cardinality defines the number of allowed classes or attribute instances. The lower and the upper limit are given. If the lower limit is defined with 0, then the class is optional (this corresponds to the textual description of the attributes given in 4.4 below e.g. "Required" = "False").

The most common cardinalities are:

1	The attribute occurs once
1 .. *	The attribute occurs once or several times
0 .. 1	The attribute occurs once or never
0 .. *	The attribute occurs never or several times

Cardinality is not only possible for attributes (see Figure 2) but also for association, aggregation and composition (see Figure 3).

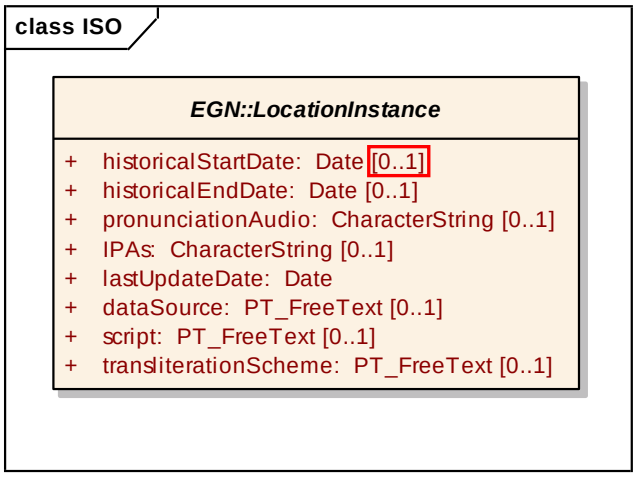


Figure 2: Cardinality for attributes

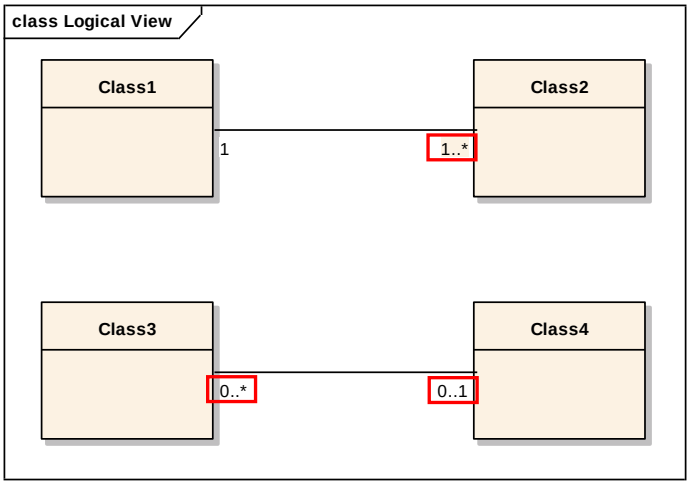


Figure 3: Cardinality for association

Readers wishing to know more about UML and specifically UML as used within the field of geographic information should consult: ISO TS 19103:2005 which provides rules and guidelines for the use of a conceptual schema language within the ISO geographic information standards. The chosen conceptual schema language is the Unified Modelling Language (UML).

4 UML Representation of the Conceptual schema and descriptive documentation

4.1 Descriptive Model Summary

The purpose of this section is to provide a very high level descriptive summary of the proposed data model. It is geared towards a broad audience who may have an interest in gazetteers but for whom the formal UML model is inaccessible.

A gazetteer as used here is a list of geographical names, postal codes or street addresses (or in ISO parlance, 'location instances') that contain information about places and geographical features, postal code areas or about addresses.
One location instance exists for each so-called 'geographic identifier'. Geographical names, postal codes or street addresses are considered to be geographic identifiers. Due to the fact that a geographic identifier has to be unique .it have to be concatenated with a specific code (see 4.4.3 SI_LocationInstance).

As for geographical names any named geographical entity that has a name referent e.g. Paris, River Thames, Atlas mountains, can be an entry in a gazetteer. Typically, such names have additional information:

- They are of some *feature type* such as city; mountain; road
- They have some *geographical footprint* such as the territory covered by a country or the course of a river.

Essentially each entry in the gazetteer (and whilst not a universal rule, in digital gazetteers this is frequently the case) must have a name, a feature type and a geographical 'footprint'. Additional information related to the name can also be recorded, as in the EGN case where attributes such as the official status of the name and the gender of the name are recorded.

In the model presented in section 4.2 there is descriptive information about the gazetteer itself (held in class SI_Gazetteer), which is seen to be an aggregation of location instances (name entries). Each location instance may have different geographical (spatial) footprints associated with it (such as a point coordinate for its location in space, a bounding box for an areal feature or the full complex geometry describing in detail the shape of the geographical entity in question (in the diagram this is alluded to in the ReferencableSpatialObject class).

Also associated with a location instance is a classification scheme (or catalogue) that describes the feature types. Thus, through the location instance we fulfil the stipulation of recording a name (which may be an endonym maintained at the NMCA database or an exonym maintained at the exonyms and other variant names database), feature type and spatial footprint for all entries in the gazetteer.

4.2 Model Detail

The following UML model shows the final EGN gazetteer model (March 2009). This is primarily based on the existing ISO 19112 standard with amendments made for

specifically EGN purposes and to reflect the known deficiencies of the existing ISO model as for example identified by the INSPIRE Drafting Team Data Specifications. Amongst others, the support of multilingualism as well as of multiple alternative geographic identifiers has been introduced (as a reference, please see document "D2.5 INSPIRE Generic Conceptual Model" in its latest version, June 2008, at the INSPIRE website:
http://inspire.jrc.it/reports/ImplementingRules/DataSpecifications/D2.5_v3.0.pdf).

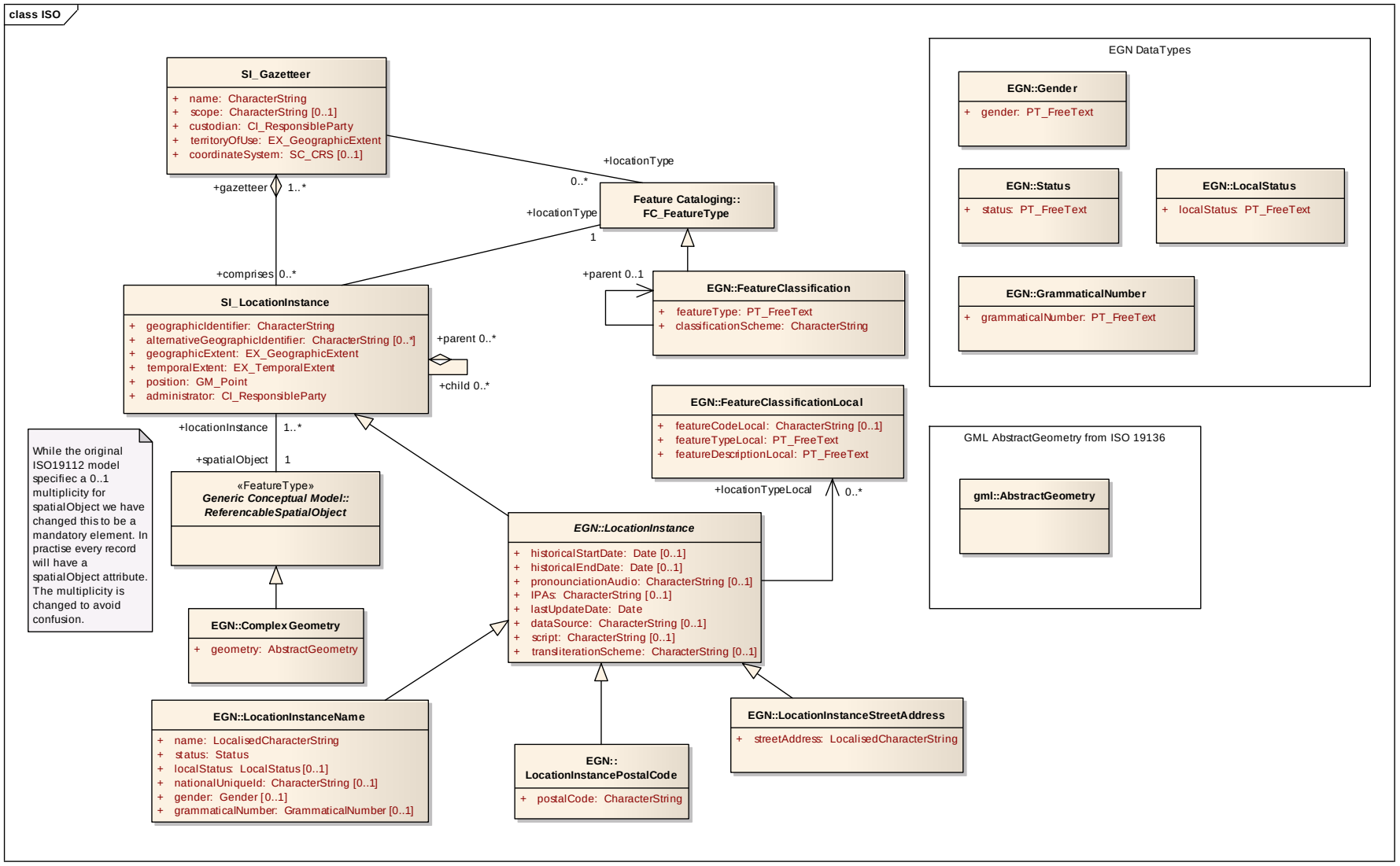
The final EGN UML model aspires to conform to the gazetteer requirements set by the INSPIRE Drafting Team as far as the EGN consortia have been able to ascertain what these are within the time constraints imposed by the project. This model has been further subject to external verification by both ISO and other gazetteer experts.

This final revision to the UML schema makes the following minor changes to the previously published model:

- Changed the type of LocationInstanceName.name to LocalisedCharacterString.

Changed the multiplicity of the PT_FreeText-valued attributes from [1..*] to [1].

4.3 UML Data model



4.4 Description of classes

4.4.1 PT_FreeText

Class as described in ISO 19139. This class implements the multilingualism. In EGN it is used as data type, which does not always require all attributes in the EGN classes described.Note it is worth pointing out that both PT_FreeText and LocalisedCharacterString are subtypes of CharacterString. As a result, an instance of PT_FreeText *or* LocalisedCharacterString may in principle also be the value of a property that is modelled with a value type of CharacterString in the UML model. Whether or not the PT_FreeText is needed or a simple (Localised)CharacterString is sufficient was decided on a case-by-case. As an example, the type of LocationInstanceName.name was changed from PT_FreeText (used in earlier model incarnations) to LocalisedCharacterString. This was based on the fact that all names in EGN have a language associated with them, but there is only one language represented in any given name. Therefore, LocalisedCharacterString is the more appropriate type to use as PT_FreeText does not add any further value.

As a general rule EGN does not use the Country attribute as this is catered for elsewhere in the EGN data model and was therefore deemed to be potentially confusing.

Attributes

Name		Type
languageCode		ISO639 (-2)
Attributes:	Fixed Length	
DataUpdatable:	True	
Required:	False	
UnicodeCompression:	True	
Description:	Language used for documenting a plain text. [ISO 19115]	
country		ISO3166 **NOTE** – this attribute will NOT be used in the EGN infrastructure.
Attributes:	Fixed Length	
DataUpdatable:	True	
Required:	False	
UnicodeCompression:	True	
Description:	usually a character string holding the ISO 3166 country identifier (fixed Length, 2 characters). But, the "countryCode" should not be used, whenever the language of a name is indicated, in order to avoid wrong relations between languages and countries. Hence, this attribute will not be used in class PT_FreeText within the EGN infrastructure. The "countryCode" is already satisfactorily included in the “INSPIRE_UID” and in the "geographicIdentifier", both can be easily identified by the system.	
characterSetCode		MD_CharacterSetCode
Attributes:	Variable Length	
DataUpdatable:	True	
Required:	False	
UnicodeCompression:	True	
Description:	full name of the ISO character-coding standard used for documenting a plain text. [ISO 19115]	
plainText		CharacterString
Attributes:	Variable Length	

DataUpdatable:	True
Required:	True
UnicodeCompression:	True
Description:	content of a free text metadata element. [ISO 19112]

4.4.2 SI_Gazetteer

A gazetteer is a directory of instances of location types in a spatial reference system.

Attributes

Name — **Type**

name — CharacterString

Attributes:	Variable Length
DataUpdatable:	True
Required:	True
UnicodeCompression:	True
Description:	Name of the gazetteer. Version date of the gazetteer is included in name [ISO 19112]. Any language can be used. For the EGN Central Service the value will be 'EuroGeoNames'.

scope — CharacterString

Attributes:	Variable Length
DataUpdatable:	True
Required:	False
UnicodeCompression:	True
Description:	Description of the location types contained in the gazetteer. Examples of scope are "streets of London" and "rivers of North America". [ISO 19112]. This dataset will be documented in English only.

custodian — Cl_ResponsibleParty

Attributes:	
DataUpdatable:	True
Required:	True
Description:	Name of the organisation responsible for maintenance of the gazetteer [ISO 19112]. Within EGN it is the name of the custodian of the gazetteer service. For EGN Local Services this might be equal to attribute 'administrator' in the class SI_LocationInstance, but not necessarily.

territoryOfUse — EX_GeographicExtent

Attributes:	
DataUpdatable:	True
Required:	True
Description:	Geographic domain covered by the gazetteer. Examples for geographic domain are for a gazetteer of rivers, "North America" and for a gazetteer of streets, "London".[ISO 19112]. In EGN this attribute will be used to provide the bounding box for the whole territory covered by the national gazetteer. It comprises all the bounding boxes of EX_GeographicExtent in the class SI_LocationInstance.

coordinateSystem — SC_CRS

Attributes:	
DataUpdatable:	True
Required:	False
Description:	Name of coordinate reference system used in the gazetteer for describing position [ISO 19112]. Within EGN the European terrestrial reference system ETRS89 will be used.

4.4.3 SI_LocationInstance

A LocationInstance is referred to a ReferencableSpatialObject. The latter one is an abstract representation of a real world phenomenon related to a specific location or geographic area. One LocationInstance exists for each geographic identifier.

The geographic identifier is unique and can be:
- a geographical name of the feature/spatial object

- a postal code
- a street address

concatenated with the two-letter ISO 3166 code (e.g. DE, NL) and a code, which can be a simple integer, a character string or a code generated e.g. according to the BASE36-system (e.g. "2YC67000B").
The name/postal code/street address will be separated from the code by "**;**" (semicolon). Apart from that, the countryCode will be separated from the code by "**.**" (point). The geographicIdentifier must be unique not only within its own national gazetteer but also across the EGN gazetteer.

Example: the same geographical name might apply for different spatial objects in one European country (or even in different European countries). E.g. in Germany more than one geographical name 'Neustadt' exists. Thus, the geographicIdentifier for the 'Neustadt' located in the state of 'Niedersachsen' would be different from the geographicIdentifier for 'Neustadt' located in the state of 'Baden-Württemberg'. 'Neustadt;DE.CV47110BN' and 'Neustadt;DE.UOP97800R' might be the geographic identifier for these two locations.

Example: in some European countries the postal code areas do not always match with the administrative areas/units. E.g. in Germany one postal code area covers a part of the administrative unit 'Alpenrod' and the administrative unit 'Nistertal'. Thus, the postal code area should be referred to a separate ReferencableSpatialObject. It is not equal to the ReferencableSpatialObjects the administrative units are linked with.

That means in the case of geographical names that – if more than one geographical name exists for the same spatial object – one or more LocationInstances refer to the same ReferencableSpatialObject.

Example: 'Bautzen' (German) and 'Budyšin' (Sorbian) are both official and equally treated names regarding their official status. These will create two different geographicIdentifier(s), and thus also two LocationInstances, but they refer to the same ReferencableSpatialObject.

The parent-child-relation of the LocationInstances is a self-join, i.e. the parent-child-relation enables that each extended class
- LocationInstancesName,
- LocationInstancesPostalCodes and/or
- LocationInstancesStreetAddress

is linked/associated with each other through its superordinated LocationInstance.

Example: Through the parent-child-relation a geographical name (parent) might be linked to a postal code (child) or to a street address (child). Very likely, many postal codes and many street addresses reference to one geographical name.

Attributes

Name		**Type**
geographicIdentifier		CharacterString
Attributes:	Variable Length	
DataUpdatable:	True	
Required:	True	
UnicodeCompression:	True	
Description:	The geographic identifier is unique and can be: - a geographical name of the spatial object - a postal code	

- a street address
concatenated with the two-letter ISO 3166 code (e.g. DE, NL) and a code generated e.g. according to the BASE36-system (e.g. “2YC67000B”). The name/postalcode/streetaddress will be separated from the code by ";" (semicolon). Apart from that, the countryCode will be separated from the BASE36-system by "." (point). The geographicIdentifier must be unique not only within its own national gazetteer but also across the EGN gazetteer.
Examples for the geographicIdentifier might be:
"Berlin;DE.2YC67000B", "Helsinki;FI.HJK675901", "Rhein;DE.CD4711XXX", "Rijn;NL.CD4711XXX", etc .

alternativeGeographicIdentifier — CharacterString

Attributes:	Variable Length
DataUpdatable:	True
Required:	False
UnicodeCompression:	True
Description:	This attribute will be used in the EGN infrastructure as a ‘helper’ attribute to assist in fast querying. All variant names, in particular exonyms, will be cross-related to the respective geographic identifier.

geographicExtent — EX_GeographicExtent

Attributes:	Variable Length
DataUpdatable:	True
Required:	True
UnicodeCompression:	True
Description:	The geographic extent shall be defined in one of the following ways: a) as a collection of smaller geographic features, for example the European Union, defined by its constituent countries; or b) by a bounding polygon, described either: as a closed set of boundary segments (each defined by one or more geographic features), for example a block defined by the bounding streets; or by a set of coordinates, for example, a land parcel defined by the coordinates of its boundary. Within EGN this attribute will be used for storing the bounding boxes for all spatial objects.

temporalExtent — EX_Temporal_Extent

Attributes:	
DataUpdatable:	True
Required:	True
Description:	The time period covered by the content of the dataset [ISO 19115]. Within EGN it is used for the date of the first storage in the EGN database of this name, postal code or street address, e.g. 2008-06-30.

position — GM_Point

Attributes:	
DataUpdatable:	True
Required:	True
Description:	GM_Point contains the central point of the associated spatial object. ETRS89 (epsg:4258) coordinates must be used. The point should be derived from the EGN::ComplexGeometry in the following way: point: The same point; line: A point on the middle of the line; polygon: The centroid (warning: this point could be outside of the polygon); multipoint: The centroid; multiline: The middle of the longest line; multipolygon: The centroid of the largest polygon (warning: this point could be outside of the polygon);

administrator — Cl_ResponsibleParty

Attributes:	
DataUpdatable:	True
Required:	True
Description:	Name of the organisation responsible for defining the characteristics of the location instance [ISO 19112]. Within EGN it is the name of the data provider. For EGN Local Services this might be equal to attribute ‘custodian’ in the class SI_Gazetteer, but not necessarily.

4.4.4 FC_FeatureType

FC_FeatureType provides the reference to a definition in a feature catalogue (or a feature concept dictionary once a model for these is available).

4.4.5 <<FeatureType>>ReferencableSpatialObject

The ReferencableSpatialObject is an abstract representation of a real world phenomenon related to a specific location or geographic area.

If more than one geographical name exists for the same spatial object, one or more LocationInstances refer to the same ReferencableSpatialObject.

Example: 'Bautzen' (German) and 'Budyšin' (Sorbian) are both official and equally treated names regarding their status. They will create two different geographicIdentifier, thus also two LocationInstances, but they refer to the same ReferencableSpatialObject.

A ReferencableSpatialObject might even comprise multiple complex geometries, but – according to the INSPIRE principles – only the most accurate geometry shall be provided to the EGN infrastructure.

The geographicIdentifier is uniquely linked/associated to the ReferencableSpatialObject through a "unique identifier" (within EGN it is called 'spatialObjectUID') which uniquely identifies the spatial object as demanded by INSPIRE.
This spatialObjectUID must be unique not only within its own gazetteer but also across all the gazetteers and across all INSPIRE-connected countries. According to the INSPIRE requirements, the spatialObjectUID has to be provided by the member countries/NMCAs.
The "unique identifier" (spatialObjectUID) will be composed of a namespace to identify the data source plus a local identifier, assigned by the data provider (this must be unique within the namespace). All namespaces will start with the two-letter ISO 3166 code (e.g. DE, NL).

Example: "DE.BKG.GN" may be the namespace used by BKG, Germany for spatial objects in their geographical names database. The local identifier can be generated according to the BASE36-system (e.g. "2XH50000A"). The result for one ReferencableSpatialObject might be "DE.BKG.GN.2XH50000A"

4.4.6 EGN::ComplexGeometry

This subtypes the abstract ReferencableSpatialObject class to provide a concrete manifestation of the abstract GML geometry type. In EGN it will be used to store the most detailed geometry for a feature and supplements the position and EX_GeographicExtent attributes in SI_LocationInstance.

4.4.7 Abstract class EGN::LocationInstance

This is an abstract class from which EGN::LocationInstanceName, EGN::LocationInstancePostalCode and EGN::LocationInstanceStreetAddress inherit.

Attributes

Name	Type

historicalStartDate — DateTime

Attributes:	
DataUpdatable:	True
Required:	False
Description:	Defines the start date of the lifetime of the geographical name, postal code or address. Within EGN this time stamp should be filled in with a start date only for historical names like 'Karl-Marx-Stadt' (valid for 1953 to 1990) for 'Chemnitz' (since 1990), e.g. 1953-01-01.

historicalEndDate — DateTime

Attributes:	
DataUpdatable:	True
Required:	False
Description:	Defines the end date of the lifetime of the geographical name, postal code or address. Within EGN this time stamp should be filled in with an end date only for historical names like 'Karl-Marx-Stadt' (valid for 1953 to 1990) for 'Chemnitz' (since 1990), e.g. 1990-10-03.

pronunciationAudio — CharacterString

Attributes:	Variable Length
DataUpdatable:	True
Required:	False
UnicodeCompression:	True
Description:	URL of audio file with the pronunciation of the name, postal code or street address

IPAs — CharacterString

Attributes:	Variable Length
DataUpdatable:	True
Required:	False
UnicodeCompression:	True
Description:	Pronunciation of the name, postal code or street address using the International Phonetic Alphabet (IPA).

lastUpdateDate — DateTime

Attributes:	
DataUpdatable:	True
Required:	True
Description:	The date of the last change of the recorded attributes of the name, postal code or street address.

dataSource — CharacterString

Attributes:	Variable Length
DataUpdatable:	True
Required:	False
UnicodeCompression:	True
Description:	This attribute is a character string describing the origins of the name, postal code or address. Considering that e.g. in Germany populated places are adopted by the commune level, this attribute was intended to be populated with something like "Register of communes of the Federal Republic of Germany", provided by the Federal Statistical Office in Germany.

script — CharacterString

Attributes:	Variable Length
DataUpdatable:	True
Required:	False
UnicodeCompression:	True
Description:	The script in which the name is rendered. . Some European languages/scripts covered by systems recommended by the United Nations are: Bulgarian, Greek, Macedonian Cyrillic, Russian and Serbian. Other languages/scripts can be: Armenian, Byelorussian, Georgian and Ukrainian. Regarding values for 'script' or 'transliterationScheme' the following examples for three location instances (names) related to the same spatial object could be stated:

Λευκωσία, language: Greek, characterEncoding: utf-8, script: Greek, transliterationScheme: e.g. ELOT 743
Lefkosia, language: Greek, characterEncoding: utf-8, script: Latin, transliterationScheme: ELOT 743
Lefkoşa, language: Turkish, characterEncoding: utf-8, script: Latin, transliterationScheme: none

transliterationScheme CharacterString

Attributes:	Variable Length
DataUpdatable:	True
Required:	False
UnicodeCompression:	True
Description:	This attribute is a character string describing the method of names conversion between different alphabetic scripts and syllabic scripts, in which each character or di-, tri- and tetragraph of the source script is represented in the target script in principle by one character or di-, tri- or tetragraph, or a diacritic, or a combination of these (source: UNGEGN). (further explanations see 'script' above)

4.4.8 EGN::LocationInstanceName

LocationInstanceName extends the LocationInstance class and comprises all geographical name-related attributes.

Attributes

Name **Type**

name LocalisedCharacterString

Attributes:	Variable Length
DataUpdatable:	True
Required:	True
UnicodeCompression:	True
Description:	This attribute is the geographical name of the spatial object. "plainText" and "languageCode" will be used. Example: Bautzen

status Status

Attributes:	
DataUpdatable:	True
Required:	True
Description:	This attribute is held in an EGN::Status object. Represents NMCA's view on the official status of name.

localStatus LocalStatus

Attributes:	Variable Length
DataUpdatable:	True
Required:	False
Description:	Represents NMCA's view on the local status of the name.

nationalUniqueId CharacterString

Attributes:	Variable Length
DataUpdatable:	True
Required:	False
UnicodeCompression:	True
Description:	This attribute holds the NMCA's national identifier of the spatial object or name.

gender Gender

Attributes:	
DataUpdatable:	True
Required:	False
Description:	This attribute is an entry in the EGN::Gender data type.

grammaticalNumber		GrammaticalNumber
Attributes:		
DataUpdatable:	True	
Required:	False	
Description:	This attribute is an entry in the EGN::GrammaticalNumber data type.	

4.4.9 EGN::LocationInstancePostalCode

The EGN::LocationInstancePostalCode class inherits from the abstract class EGN::LocationInstance and provides all postal codes for indirect referencing.

Attributes

Name		Type
postalCode		CharacterString
Attributes:	Variable Length	
DataUpdatable:	True	
Required:	True	
Description:	This attribute is a character string holding the postal code of the LocationInstance if available.	

4.4.10 EGN::LocationInstanceStreetAddress

The EGN::LocationInstanceStreetAddress class inherits from the abstract class EGN::LocationInstance and provides all the street names and house numbers for indirect referencing.

Attributes

Name		Type
streetAddress		LocalisedCharacterString
Attributes:	Variable Length	
DataUpdatable:	True	
Required:	True	
Description:	This attribute is holding one or more street addresses of the LocationInstance if available. “plainText” and “languageCode” will be used.	

4.4.11 EGN::FeatureClassification

The EGN conceptual schema allows for different feature classifications to be used. The EGN feature classification scheme will be satisfactory for the purposes to which it is required (essentially, query filtering). The EGN project uses the term ‘classification’ throughout to distinguish from the term ‘catalogue’ used in other contexts (such as ISO 19110) which implies a degree of formalism not required for the purposes of EGN. The purpose of the EGN feature classification is primarily to assist in query and discovery.

The EGN feature classification is structured in main and sub classes. Through the parent-relation the main classes and, indirectly, the sub classes (children) can be linked with each other. The NMCAs shall provide translations of terms of the EGN feature classification (main classes and sub-classes) in the respective 23 languages of the 15 participating countries and in English (default language).

Attributes

Name	Type

featureType — PT_FreeText

Attributes:	Variable Length
DataUpdatable:	True
Required:	True
Description:	This attribute is a pointer to an entry for a *feature type* in the EGN multilingual feature classification scheme. “plainText” and “languageCode” will be used e.g. ‘river’;’eng’ or ’rio’;’spa’.

classificationScheme — CharacterString

Attributes:	Variable Length
DataUpdatable:	False
Required:	True
Description:	This attribute is the name of the feature catalogue or dictionary e.g. ‘EGN Classification Issued 20080101’.

4.4.12 EGN::FeatureClassificationLocal

This class allows for the details of the local (NMCA specific) feature classification schemes to be recorded and provides end users with a mechanism for discovering finer grained feature typing for any specific geographical name in EGN. The NMCAs shall provide translations of local terms/information used in the EGN feature classification in all official languages spoken in the country (if possible) and in English (if possible). EGN::FeatureClassificationLocal has a 1..1 association with EGN::LocationInstance.

Attributes

Name — **Type**

featureCodeLocal — CharacterString

Attributes:	Variable Length
DataUpdatable:	True
Required:	False
Description:	This attribute is a string holding a local code identifier for the feature in the local NMCA feature classification scheme e.g. ‘3501’ in the German case which identifies ‘Railway station facility’ features.

featureTypeLocal — PT_FreeText

Attributes:	Variable Length
DataUpdatable:	True
Required:	True
Description:	This attribute is a string holding a pointer to the feature type in the local NMCA feature classification scheme e.g. ‘seasonally flooding tributary’;’eng’. “plainText” and “languageCode” will be used.

featureDescriptionLocal — PT_FreeText

Attributes:	Variable Length
DataUpdatable:	True
Required:	True
Description:	This attribute is a string holding a description of a feature in the individual NMCA supported languages. “plainText” and “languageCode” will be used.

4.4.13 EGN Data types - EGN::GrammaticalNumber

EGN::GrammaticalNumber provides an enumerated list of grammatical numbers. Associated with the class ‘PT_FreeText’ the grammatical numbers will be provided in individual languages. The first implementation within the 30 month funded project duration will support at least the 23 languages of the 15 Reference Group members (comprising the nationally officially recognized minority languages).

Attributes

Name — **Type**

grammaticalNumber	PT_FreeText
Attributes:	Variable Length
DataUpdatable:	True
Required:	True
Description:	An enumerated list comprising of: plural, singular and dual. “plainText” and “languageCode” will be used.

4.4.14 EGN Data types - EGN::Gender

EGN::Gender provides an enumerated list of gender. Associated with the class ‘PT_FreeText’ the gender will be provided in individual languages. The first implementation within the 30 month funded project duration will support at least the 23 languages of the 15 Reference Group members (comprising the nationally officially recognized minority languages).

Attributes

Name	Type
gender	PT_FreeText
Attributes:	Variable Length
DataUpdatable:	True
Required:	True
Description:	An enumerated list comprising of: neuter, masculine, feminine, common. “plainText” and “languageCode” will be used.

4.4.15 EGN Data types - EGN::Status

EGN::Status provides an enumerated list of status codes which will be referred either to the endonyms or to exonyms. Associated with the class ‘PT_FreeText’ the status attributes (‘official_endonym’ or ‘other_endonym’ and ‘standardized_exonym’ or ‘other_exonym’) will be provided in individual languages. The first implementation within the 30 month funded project duration will support at least the 23 languages of the 15 Reference Group members (comprising the nationally officially recognized minority languages).

Attributes

Name	Type
status	PT_FreeText
Attributes:	Variable Length
DataUpdatable:	True
Required:	True
Description:	An enumerated list comprising of: official_endonym, other_endonym standardized_exonym, other_exonym. “plainText” and “languageCode” will be used.

4.4.16 EGN Data types - EGN::LocalStatus

This data type was introduced to cater for instances where Status itself may have sub-national codes. By way of example, in Germany the names may have 1 out of 5 status attributes (‘amtlich’, ‘gebräuchlich’, ‘historisch’, ‘Kurzform’, etc.). In Finland only the different languages (Finnish, Swedish, North Saami, Inari Saami, Skolt Saami) may have 1 out of 20 different potential *local* statuses – the 5 languages may be unofficial or official and may be majority language in municipality or minority language in municipality. The localStatus attribute permits just such permutations for local status to be recorded. The NMCAs shall provide translations of local terms/information used in all official languages spoken in the country (if possible) and in English (if possible).

Attributes

Name	Type
localStatus	PT_FreeText

Attributes:	Variable Length
DataUpdatable:	True
Required:	True
Description:	The local enumerations for status e.g. in Finland only the different languages may have 1 out of 20 different MUNICIPAL status: ***unofficial/official* *language* *majority/minority language in that municipality:** In the Finnish database the following distinction is made: Language_Code 1 (1 = Finnish, 2 = Swedish, 3 = North Saami, 4 = Inari Saami, 5 = Skolt Saami) Language_Official_Status_Code (1 = official, 2 = unofficial) Language_Majority_Satus_Code (1 = Majority language in municipality, 2 = Minority language in municipality)

5 Examples to be used for explaining details of the EGN data model (D4.2e)

1) Output related to "Status" information

EGN::LocationInstanceName		ReferencableSpatialObject	SI_LocationInstance	EGN::Status		EGN::LocalStatus	
Name	Language code[4]	spatialObjectUID	geographicIdentifier		Translations[5]		Translations[6]
Alzenau i.Ufr.	ger	DE.BKG.GN.2XH5000A	Alzenau i.Ufr.;DE.CD4711XXX	official_endonym	amtlich (ger), xxx (fri), xxy (wen), xyx (dan), xxy (fra), etc.	amtlich (ger)	xxx (wen), xyv (fri), xyz (dan), official (eng)
Alzenau in Unterfranken	ger	DE.BKG.GN.2XH5000A	Alzenau in Unterfranken;DE.XBY124RZW	other_endonym	sonstige (ger), xxx (fri), xxy (wen), xyx (dan), xyy (fra), etc.	ausgeschrieben (ger)	Full form (eng)
Alzenau	ger	DE.BKG.GN.2XH5000A	Alzenau;DE.CZ789V413	other_endonym	sonstige (ger), xxx (fri), xxy (wen), xyx (dan), xyy (fra), etc.	gebräuchlich (ger)	in use (eng)
Alzenau	ger	DE.BKG.GN.6Y56R00C	Alzenau;DE.P68234HGJ	official_endonym	amtlich (ger), xxx (fri), xxy (wen), xyx (dan), xxy (fra), etc.	amtlich (ger)	xxx (wen), xyv (fri), xyz (dan), official (eng)
Helsinki	fin	FI.NLS.GNR.9XXGB20C	Helsinki;FI.HJK675901	official_endonym	xxx (fin), xxy (swe), amtlich (ger), xxx (fri), xxy (wen), xyx (dan), xxy (fra), etc.	xxx (fin)	Name in official majority language in the municipality (eng)
Helsingfors	swe	FI.NLS.GNR.9XXGB20C	Helsingfors;FI.095GH67XY	official_endonym	xxx (swe), xxy (fin), amtlich (ger), xxx (fri), xxy (wen), xyx (dan), xxy (fra), etc.	xxy (swe)	Name in official minority language in the municipality (eng)
Ελσίνκι	gre	FI.NLS.GNR.9XXGB20C	Exonym_ID	standardized_exonym	None	None	None
Lille	fra	FR.IGN.NGF.7XX5006B	Lille;FR.789001VBN	official_endonym	xxy (fra), amtlich (ger), xxx (fri), xxy (wen), xyx (dan), xxy (fra), etc.	officiel (fra)	xxx (???), official (eng)
Ryssel	nel	FR.IGN.NGF.7XX5006B	Exonym_ID	standardized_exonym	None	None	None

4 in the EGN infrastructure the three letter language code of ISO639-2 in its current version will be used.
5 Translations in 23 languages of the 15 participating countries
6 Translations in the official minority languages in the country (if possible) + English (if possible)

2) Output related to "feature type" information

EGN::LocationInstanceName		**ReferencableSpatialObject**	**SI_LocationInstance**		**EGN::featureClassification**		**EGN::FeatureClassificationLocal**	
Name	Language code[7]	spatialObjectUID	GeographicIdentifier	alternativeGeographicIdentifier		Translations[8]		Translations[9]
Rhein	ger	DE.BKG.GN.2XH5000A	Rhein;DE.CD4711XXX	Rijn;DE.CD4711XXX Rhin, et al.	Flowing water	Fließgewässer (ger), xxx (fri), xxy (wen), xyx (dan), xyz (spa), xxy (fra), etc.	Strom, Fluss, Bach (ger)	Stream, river, creek (eng)
Rijn	nel	NL.TDK.GN.45TZU00BV	Rijn;NL.CD4711XXX	Rhein;DE.CD4711XXX Rhin, et al.	Flowing water	xxx (nel), Fließgewässer (ger), xxx (fri), xxy (wen), xyx (dan), xyz (spa), xxy (fra), etc.	xxx (nel)	xyv (eng)
Rhin[10]	fra	NL.TDK.GN.45TZU00BV, DE.BKG.GN.2XH5000A	Exonym_ID	None	Flowing water	Fließgewässer (ger), xxx (fri), xxy (wen), xyx (dan), xyz (spa), xxy (fra), etc.	None	None
Bautzen	ger	DE.BKG.GN.5678WE00	Bautzen;DE.90MNI54AV	Budyšin;DE.98IJE000A	Administrative unit	Verwaltungseinheit (ger), xxx (fri), xxy (wen), xyx (dan), xyz (spa), xxy (fra), etc.	Gemeinde (ger)	xxx (fri), xxy (wen), xyv (dan), municipality (eng)
Budyšin	wen	DE.BKG.GN.5678WE00	Budyšin;DE.98IJE000A	Bautzen;DE.90MNI54AV	Administrative unit	Verwaltungseinheit (ger), xxx (fri), xxy (wen), xyx (dan), xyz (spa), xxy (fra), etc.	Gemeinde (ger)	xxx (fri), xxy (wen), xyv (dan), municipality (eng)

7 in the EGN infrastructure the three letter language code of ISO639-2 in its current version will be used.

8 Translations in 23 languages of the 15 participating countries

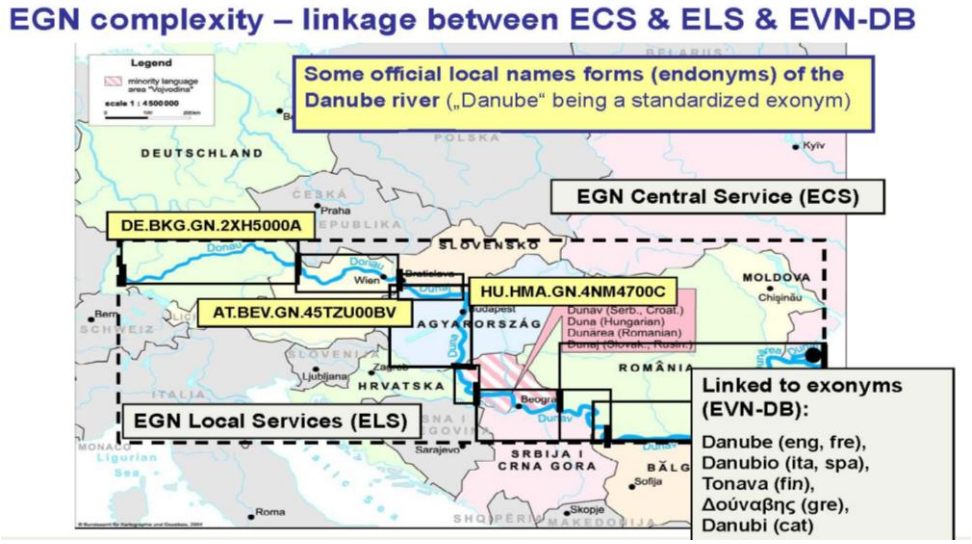
9 Translations in the official minority languages in the country (if possible) + English

10 Border crossing features have to be linked in the "database of exonyms & other variant names" through the spatialObjectUIDs

3) EXAMPLE: Danube as border-crossing spatial object

EuroGeoNames (EGN) solution: Link between EGN Central Service, the EGN Local Services and the Exonyms and other variant names database (EVN-DB)

3.1) The objective of EGN



A German user wants to get the information about the Danube river and starts his/her single inquiry with "Donau". He/she aims at getting information (all names and the geographic extent) about the complete spatial object (which may be a combination of 9 spatial objects from 9 national datasets).

3.2) EGN Local Services

Assuming that all EGN Local Services needed are running, the EGN Local Services provide the following information:

Country	SpatialObject_UID	Endonyms	geographicIdentifier	GeographicExtent
Germany	DE.BKG.GN.2XH5000A	**Donau**	Donau;DE.98673ABC	BoundingBoxDE
Austria	AT.BEV.GN.45TZU00BV	**Donau**	Donau;AT.786543C	BoundingBoxAT
Slovakia	SK.SMA.GN.87958377	**Dunaj**	Dunaj;SI.72468764	BoundingBoxSI
Hungary	HU.HMA.GN.4NM4700C	**Duna**	Duna;HU.21342315	BoundingBoxHU
Croatia	HR.HMA.GN.985463	**Dunav**	Dunav;HR.564838	BoundingBoxHR
Serbia	SZ.SMA.GN.9945344	**Dunav**	Dunav;SZ.ATRG778	BoundingBoxSZ
Bulgaria	BG.BMA.GN.33578788	**Дунав**	Dunav;BG.4238745	BoundingBoxBG
Bulgaria	BG.BMA.GN.33578788	**Dunav**	Dunav;BG.4238745	BoundingBoxBG
Romania	RO.RMA.GN.56TZHN8	**Dunărea**	Dunărea;RO.6364287	BoundingBoxRO
Moldava	MD.MMA.GN.85867987	**Dunărea**	Dunărea;MD.76ZZTH9	BoundingBoxMD
Ukraine	UA.xy	**Dunaj**	Dunay;UA.xy	BoundingBoxUA
Ukraine	UA.xy	**Дунай**	Dunay;UA.xy	BoundingBoxUA

One country/NMCA may provide more than one geographical name associated to the respective spatialObject_UID.
The linkage between the "national" pieces of the whole spatial object (border-crossing spatial objects) is done within the Exonyms and other variant names database – EVN-DB.
The EGN Central Service provides the respective national pieces from the EGN Local Services together with the information stored and maintained in the EVN-DB.

3.3) Relation to the Exonyms and other variant names database – EVN-DB

SpatialObject_UID	Endon.	eng	geog.Identifier1	fre	geog.Identifier2	[…]
DE.BKG.GN.2XH5000A	Donau	**Danube**	**Danube;EU.567493**	**Danub**	**Danub;EU.45637**	**dito**
AT.BEV.GN.45TZU00BV	Donau	**Danube**	**dito**	**dito**	**dito**	**dito**
SK.SMA.GN.87958377	Dunaj	**Danube**	**dito**	**dito**	**dito**	**dito**
HU.HMA.GN.4NM4700C	Duna	**Danube**	**dito**	**dito**	**dito**	**dito**
HR.HMA.GN.985463	Dunav	**Danube**	**dito**	**dito**	**dito**	**dito**
SZ.SMA.GN.9945344	Dunav	**Danube**	**dito**	**dito**	**dito**	**dito**
BG.BMA.GN.33578788	Dunav	**Danube**	**dito**	**dito**	**dito**	**dito**
RO.RMA.GN.56TZHN8	Dunărea	**Danube**	**dito**	**dito**	**dito**	**dito**
MD.MMA.GN.85867987	Dunărea	**Danube**	**dito**	**dito**	**dito**	**dito**
UA.xy	Dunav	**Danube**	**dito**	**dito**	**dito**	**dito**

The EVN_DB stores one set of exonyms and variant names [1..*] which can be associated to all (national) spatialObject_UIDs with cardinality [1..*].
As for the Danube river, 1 set of exonyms and variants are stored for 9 spatial objects – which are linked together through the EVN-DB .
The English exonym or variant name is always introduced if available.
Border-crossing spatial objects without associated exonyms are not linked within the EU-funded period.

3.4) Results provided through the EGN Central Service in combination with the EGN Reference Application (according to the EGN data model):

Endonym	geographicIdentifier	alternativeGeographicIdentifier
Donau	Donau;DE.98673ABC	Donau;AT.786543C, Dunaj;SK.72468764 Duna;HU.21342315 Dunav;HR.564838 Dunav;SZ.ATRG778 Dunav;BG.4238745 Dunărea;RO.6364287 Dunărea;MD.76ZZTH9 Dunaj; UA.xy Danube;EU.567493 Dunava;EU.45637 […]

The compiled boundingBox will not be calculated automatically, but the national parts will be provided together with their respective boundingBoxes!