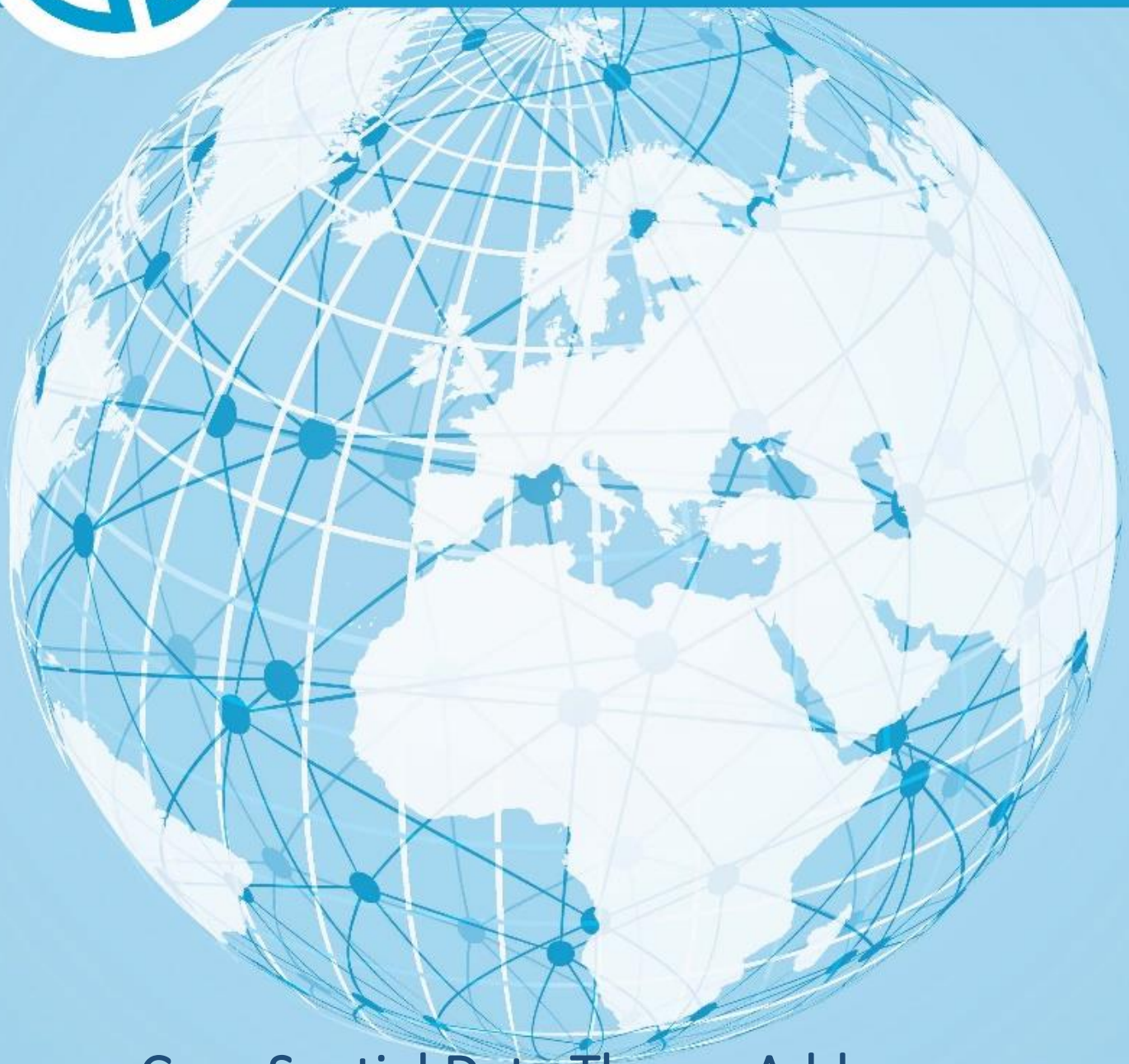
UN-GGIM EUROPE

UNITED NATIONS COMMITEE OF EXPERTS ON GLOBAL GEOSPATIAL INFORMATION MANAGEMENT



Core Spatial Data Theme Addresses Recommendation for Content

Working Group A - Deliverable of Task 1.b

Version 1.1 - 2019-08-26

Version History

Version number	Date	Modified by	Comments
1.0	April 2017	WG A	Consolidated draft, for review by geographic and statistical community
1.1	May 2022	WG A	Comments from geographic and statistic community taken into account Definitive deliverable

Warning: in the following parts of this document, the paragraphs written in grey e.g. "This document has annexes containing more detailed explanations " are common to all core spatial data themes; they aim to provide context and objectives of core data. The paragraphs written in black are specific to core spatial data theme Addresses.

Content

1 Executive Summary

In September 2015 the countries of the United Nations adopted the 2030 Agenda for Sustainable Development; a set of goals to end poverty, protect the planet, and ensure prosperity for all as part of a new sustainable development agenda. Each goal has specific targets to be achieved over the next 15 years. The 17 Sustainable Development Goals (SDGs) of the 2030 Agenda are supported by 169 targets and 230 indicators.

Geospatial data supports the measuring, achieving and monitoring of many of the goals and targets set by the 2030 Agenda. The 2030 Agenda demands 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 the 2030 Agenda depends on senior administrators owning and leading the geospatial efforts in their respective countries.

In Europe, building on the INSPIRE Directive redirecting the focus on a cohesive spatial data infrastructure without gaps in content and discrepancies in quality, stakeholders are working on geospatial standardization and increasing richness of data through Core Data Recommendations for Content that correspond to the first phase of WGA work program. Core data is primarily meant for fulfilling the common user requirements related to SDGs in Member States and European institutions.

Achievement of the SDGs requires knowledge of the location of people, buildings and events as well as efficient government administration. Address data underpins government administration at all levels, and is crucial for ensuring access to utility services. It also enables effective communication with citizens: informing them of policies applying to them, notifying them of events or incidents affecting them and supporting the carrying out of social surveys. For instance, addresses are particularly important for census purposes, to verify complete coverage of the population. In many analytical uses address data plays a crucial role in linking the records of otherwise-unrelated datasets. A geocoded address dataset enables spatial analysis which is of benefit not only to government actors but also to the private sector for purposes such as retail location planning or logistics. Thus addresses have significant economic value.

Core Spatial Data Theme “Addresses” Recommendation for Content includes a two-step approach for its adoption by UN member states: a) Core recommendations, such as full geographic coverage of the territory, a specified minimum set of attributes for each address, a definition of ‘building’ as the core addressable unit, and quality requirements about completeness or correct spelling of address components; and b) good practices, for example allocating valid addresses to vacant plots and buildings under construction and recommending the creation of “true addresses” with street names and house numbers in areas where address locator is missing.

2 Foreword

2.1 Document purpose and structure

2.1.1 Purpose

This document provides the main characteristics of core data for theme Addresses 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: e.g. 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 Addresses.

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' 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 follow:

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

AD	INSPIRE theme Addresses
CRS	Coordinate Reference System
EC	European Commission
ELF	European Location Framework
INSPIRE	Infrastructure for Spatial Information in Europe (Directive 2007/2/EC)
ISA	Interoperability solutions for public administrations, businesses and citizens
OGC	Open Geospatial Consortium
SDG	Sustainable Development Goal
UNECE	United Nations Economic Commission for Europe
UN-GGIM	United Nations initiative on Global Geospatial Information Management
WG A	(UN-GGIM: Europe) Working Group on Core data

2.5 Glossary

2.5.1 Levels of detail

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
Regional	Level of detail defined by ELF: data to be used generally at scales between 1: 100 000 and 1: 500 000; data mainly for national or regional (European or cross-border) 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 mainly for regional (sub-national) 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 mainly for local level actions.
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, mainly for local level actions.

NOTE: the above definitions are indicative; in practice, detailed data (Master levels) may also be required also by national, European or international users.

2.6 Reference documents

INSPIRE Data Specification on Addresses– Technical Guidelines 3.1:
http://inspire.ec.europa.eu/id/document/tg/ad

Core spatial data theme ‚Geographical Names' – Recommendation for content – Final version 1.1

https://un-ggim-europe.org/wp-content/uploads/2018/11/UN-GGIM-Europe_WGA_Recommandation_Content-GN-v1.1.pdf (document to be uploaded)

3 Overview

3.1 General scope

Definition: Location of properties based on address identifiers, usually by road name, house number, postal code [INSPIRE Directive 2007/2/EC, Annex 1].

NOTE 1: The term “address” may be confusing as it may define different concepts:

- Address in the real world: addresses in current use, e.g. materialised in the streets or used in mails.

Figure 1: Illustrations about real-world addresses

- Address in text only databases: many stakeholders, such as Post Offices, Taxation Offices, Statistical Institutes, utilities network managers, business companies have text only data on addresses: this data includes the address components (house number, road name, postal code, etc) sometimes associated with the name of the physical or moral person living or working at this address.

Feature of interest			
Address	**Attribute 1**	**Attribute 2**	**.....**
75004 Paris - 1, Avenue de Rivoli	xxxx	xxxx	
75004 Paris - 2, Avenue de Rivoli	xxxx	xxxx	
75004 Paris - 3, Avenue de Rivoli	xxxx	xxxx	
75004 Paris - 4 Avenue de Rivoli	xxxx	xxxx	
75004 Paris - 5, Avenue de Rivoli	xxxx	xxxx	
75004 Paris - 6, Avenue de Rivoli	xxxx	xxxx	
75004 Paris - 7, Avenue de Rivoli	xxxx	xxxx	
75004 Paris - 8, Avenue de Rivoli	xxxx	xxxx	
75004 Paris - 9, Avenue de Rivoli	xxxx	xxxx	
75004 Paris - 10, Avenue de Rivoli	xxxx	xxxx	

Figure 2: Illustration about text only Address database

- Address in geographic databases also called geocoded address databases: the address data includes not only the address components (house number, rod name, postal code, etc) but also the address location, i.e. a point geometry defined by a set of coordinates in a given Coordinate Reference System. Geographic addresses may be displayed on a map and are generally produced by National Mapping or Cadastral Agencies.

Figure 3: illustrations about geographic addresses

This document focuses on the third case, i.e. on the geographical representation of addresses.

NOTE 2: As general principle, core addressable objects are buildings. More details are provided in chapter 4.5 on Data capture.

NOTE 3: The scope of core data is very close to the scope of the INSPIRE theme Addresses. More detailed comparison with INSPIRE is available in Annex A.

3.2 Use cases

The most obvious use of addresses is the physical delivery of mail, but the power of address data in the digital age lies in its geocoding capability. A great deal of information is linked to addresses, and a geocoded address database allows such information to be linked to a physical location. Thus address data can add significant economic value, for example in marketing and logistics, contributing for instance to SDG8 (Decent work and economic growth).

Address data underpins government administration at all levels, and good administration is a prerequisite for the achievement of the SDGs. It supports the provision of services and enables effective communication with citizens: informing them of policies applying to them, notifying them of events or incidents affecting them and supporting the carrying out of social surveys.

The analysis carried out by WG A use case focuses on the role of addresses in locating people; locating buildings; locating services; and locating events. Addresses can also act as a proxy for the density of population or buildings.

Principle 1 of the Global Statistical Geospatial Framework[2], being developed under the auspices of UN-GGIM, emphasises the importance of geocoded addresses as a component of a point-based statistical framework, and address is one of the candidate themes for point-based statistics recommended by UN-GGIM: Europe WG B. Geocoded addresses may form the basis of a geospatial statistical system, or they may be used as auxiliary information to enrich statistical data. Addresses are particularly important for census purposes, to verify complete coverage of the population as well as for analytical purposes.

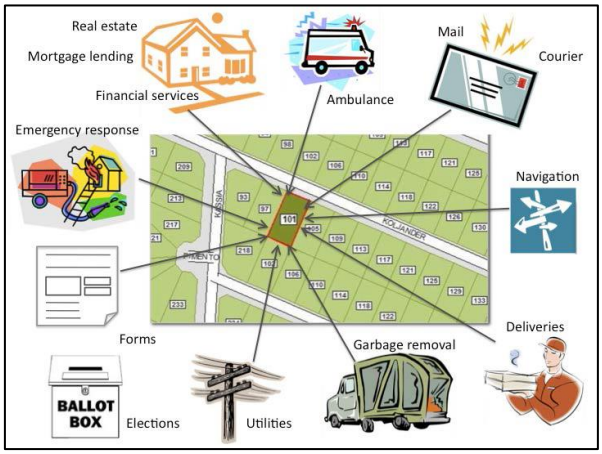


Figure 4: Use case illustration (Source: Review summary of ISO Project 19160, Addressing)

[2] http://ggim.un.org/meetings/GGIM-committee/8th-Session/documents/Global-Statistical-Geospatial-Framework-July-2018.pdf

A systems view of the use cases reduces the wide variety of applications of address data to three broad categories:

- **Search** for addresses
- **Navigation** to address (enabled by good links; see Transport Network theme)
- **Analysis** of addresses and of other data using address as a link.

4 Data content

4.1 Features types and attributes

Core Recommendation 1
Core data should comprise feature type Address with at least the following attributes: one two-dimensional geographic position, one locator (e.g. number or name) if available, and such other address components as are in current use. A unique and persistent identifier is also required.

NOTE 1: The terminology here is taken from the INSPIRE data specification. A geographic position locates the address spatially. A locator is a number and/or name assigned to a building or building unit to distinguish an address from its neighbours, in practice, it is very often a house number. An address is associated with a variable number of address components, of which there are four subclasses: administrative unit name, address area name, thoroughfare name (e.g. street, road) and postal descriptor (postal code). Address components may be used in any combination.

NOTE 2: If theme Addresses is the main source for thoroughfare names (or for address area names), these names should be captured according the recommendations for content of core spatial theme Geographical Names; this entails for instance that the name should be documented by its source, status, nativeness and source (if relevant).

The INSPIRE specification requires that an address should have at least one locator and one geographic position. It is recognised that in some countries this requirement is not currently met; for instance, in remote villages there may be no street names and no house numbers so the address is given only at village level (address area name). As a consequence, several buildings may share the same address. This lack of 1:1 association between building and precise location causes practical problems for emergency services, for utility management and even for everyday life – for example in locating an address by satellite navigation.

Good Practice 1
All core addressable objects should be provided with a unique address enabling their unambiguous location, i.e. an address with a locator and geographic position.

NOTE: This will generally imply that authorities should create, for all core addressable objects, detailed addresses (with street names and building numbers or names) in areas where they are missing.

Good Practice 2
Geographic position should be further refined by including the geometry specification attribute, which describes the type of spatial object used to derive the position. Examples from the INSPIRE code list include: building, entrance, parcel, postal delivery point, postal descriptor and administrative unit. Wherever possible, building or entrance should be used, for reasons of precision.

There would be a strong interest, e.g. for statistical purposes, to get the geographic extent of the address components, i.e. to get the geometry of the administrative unit, of the address area, of the thoroughfare and of the postal area. This is already the case for administrative units and for thoroughfares: both the INSPIRE data specifications and the UN-GGIM: Europe recommendations for content include the geometry of such features (themes Administrative Units, Transport Networks and Hydrography) and this is current practice from data providers to capture such data. Address areas are populated places that are present in INSPIRE and in core data theme Geographical Names. However, the current practice is generally to capture them only as points. The recommendations for content of core theme Geographical Names promote as good practice the capture of a realistic geometry of populated places, i.e. as a surface (even if approximate) rather than as a point.

As a conclusion, data about geographic extent of the address components are often missing about postal codes; this is rationale for the following recommendation.

Good Practice 3
The geographic extent of postal codes should be captured, by its geometry and by specifying what the postal code zone represents.

4.1.1 Temporal aspects

Core Recommendation 2
Current, valid addresses are considered as core data.

NOTE 1: In other words, efforts to capture features of the past (obsolete, destroyed, disused) are not considered as a priority.

NOTE 2: Core data being minimum data, a data producer may of course also capture features of the past as additional data; in these cases, it is advised to document the temporal attribute (statusValue) in order to make distinction between past and current features.

However, once features have been captured, it is recommended to keep them in the data base, even after their end in the real world. This can be particularly important for the statistical community and therefore is included here as good practice.

Good Practice 4
It is recommended to manage the history of addresses, using the mechanism provided by the INSPIRE data specifications: versioning and life-cycle attributes.

NOTE 1: The versioning and life-cycle attributes enable change-only updates; they also enable to retrieve the status of geographic AD data, at any time of the past (since the adoption of these mechanisms).

NOTE 2: It is recognised that the process to provide change-only update may be complicated, especially in case the address components are considered as feature types (as it is the case in the INSPIRE data model). However, managing versions of objects and capturing life-cycle are necessary preliminary steps.

NOTE 3: The above Core recommendation and Good practice may look contradictory but in fact they are not. Let us imagine a data producer deciding to implement the core recommendations and good practices of this deliverable from 2022:

- In a first step, according to the above Core recommendation, first priority is to capture the addresses that are valid (in 2022), as they are both the most useful and the easiest to be captured. For instance, capturing features from the past would require significant efforts for limited benefits.
- In a second step, for instance in 2025, a given address disappears in the real-world; the related feature – already captured in 2022 – should be kept in the database as "deprecated", which is documented by the life-cycle attributes of INSPIRE. This may be done quite easily through proper database management.

4.2 Levels of detail

Core theme Addresses corresponds to Master Level 0 data.

NOTE: Levels of detail are defined by a scale range in the glossary.

4.3 Geographical extent

Core Recommendation 3
Core address data should cover the full geographic extent of the territory.

NOTE: Countries are encouraged to provide full geographical coverage in order to permit maximum use of address data to support the SDGs across the whole territory.

Good Practice 5
Core address data should also encompass isolated buildings and buildings in remote areas

4.4 Data capture

It is reminded that a core data set shall contain the minimum data in content and in quality defined by the core recommendations but may include more or better information where available.

4.4.1 Core addressable objects

Core Recommendation 4
The core addressable object is a building, i.e. a permanent construction, intended or used for the shelter of people, having at least one entrance from publicly-accessible space.

NOTE 1: A building could be a house (whether or not attached to its neighbours), office, factory, retail unit, leisure centre, block of flats/apartments, place of worship, etc.

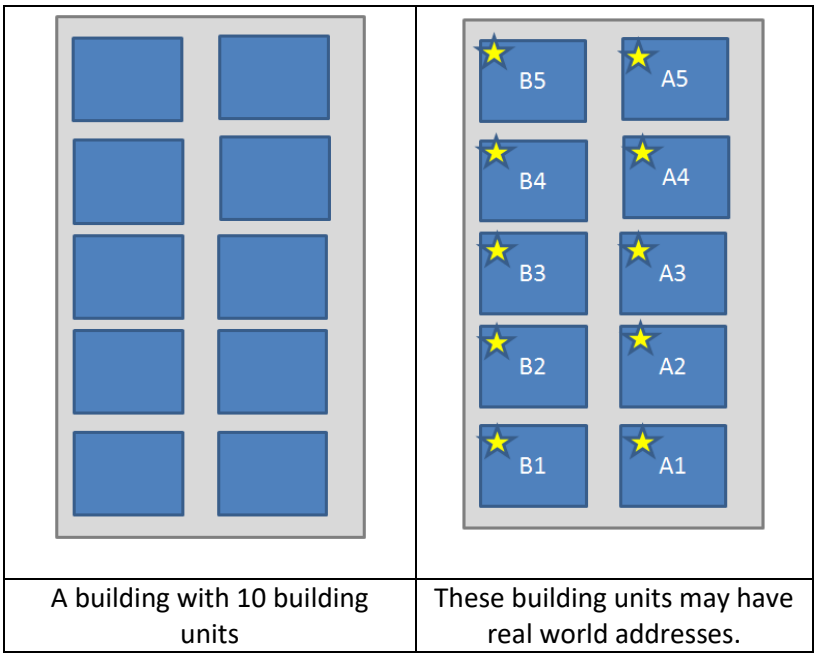
NOTE 2: Structures ancillary to human habitation are not considered as core address data; into this category would fall garden summer houses, workshops, toilet blocks and the like.

NOTE 3: In practice governments and local jurisdictions may include a wide range of other entities within the scope of their address data; examples being moorings, electricity substations, car parks, street furniture and so on. The capture of these addressable objects, though of interest, is not considered as a priority because this kind of addressable objects has minor importance for the SDGs.

4.4.2 Case of building units

Many buildings are composed of building units, each having its own entrance from shared space; this can include apartments and individual shops within an enclosed shopping centre or mall.

4.4.2.1 Real world addresses



A building with 10 building units	These building units may have real world addresses.

Figure 5: Vertical view (building façade) of a building and its building units

The existence and status of real world addresses on building units varies according to Member States: in some countries, building units are systematically identified by their own address, these addresses being materialised as for buildings whereas real world addresses on building units are missing, badly defined or poorly used in other countries.

4.4.2.2 Geographic addresses

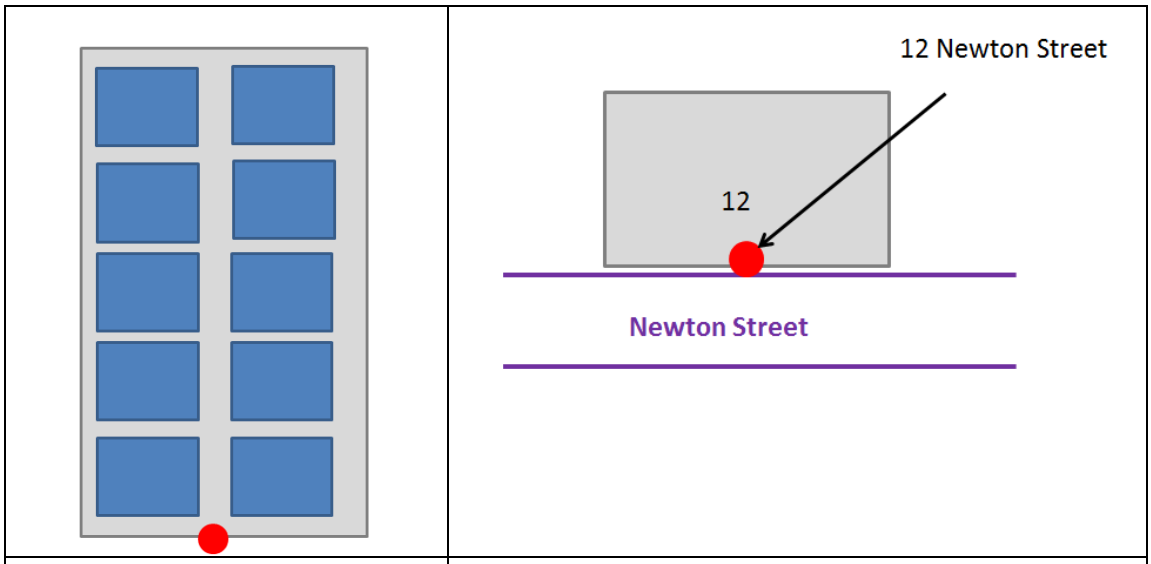
There are various ways in which the responsible authority may deal with addresses in the case of multi-unit buildings:

a) real-world addresses exist only at building level;

b) real-world addresses exist only at building units level.

c) real-world addresses exist both at the building and building unit level.

4.4.2.2.1 Case a): AD at building level



Vertical view (building façade)	View on the map

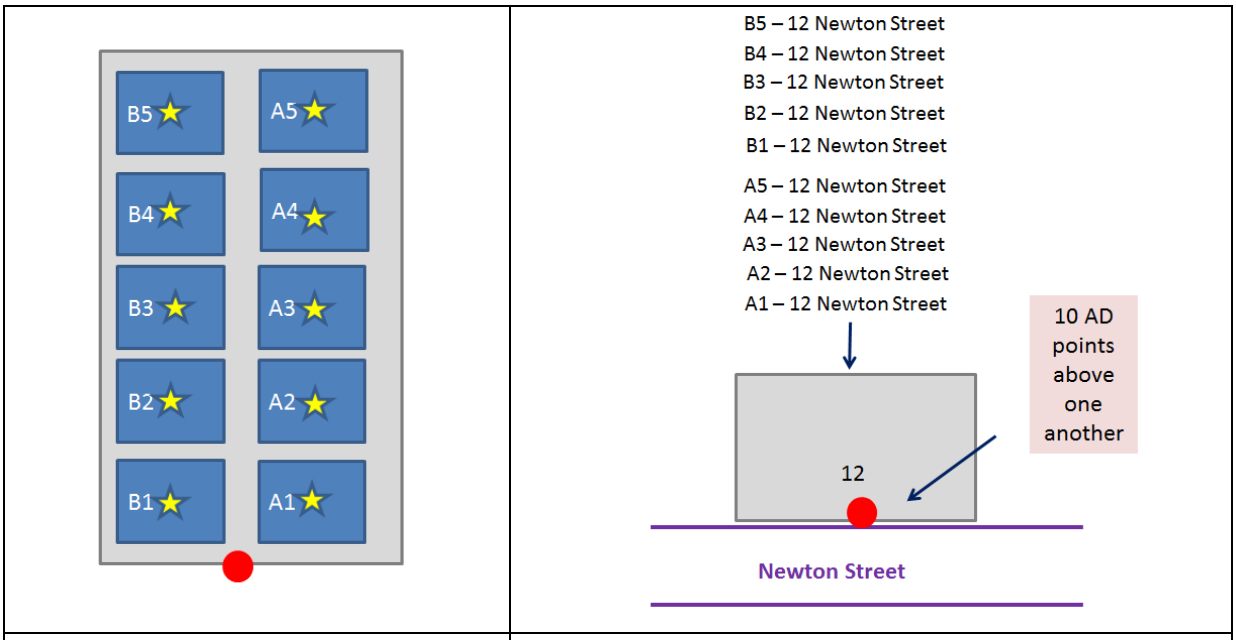
Figure 6: Illustration of case a) with the geographic data consisting in a single point (ideally) located at the building entrance

NOTE 1: Capturing geographic address at building level is enough to fulfil most of the use cases of Address theme, especially the geocoding one. This case is covered by core recommendation 4.

4.4.2.2.2 Case b): AD at building unit level only

Core Recommendation 5
In case b), where there is no address at building level, the geographic addresses of building units should be included as core data.

NOTE 1: Geographic addresses at building unit level are obviously required if real-world addresses at building level don't exist.



Vertical view (building façade)	View on the map

Figure 7: illustration of addresses at building unit level

NOTE 2: Addresses being captured (until now) as 2D data, the geographic location of the building units address will in most case be similar to the location of the building, as shown in previous illustration (based on the hypothesis the address is captured at building entrance).

4.4.2.2.3 Case c): AD both at building and building unit

If real-world addresses exist at building unit, it is advised to collect and record them (at least) in text only database. For instance, flats and apartments are core topics in UNECE recommendations for 2020 round of Population and housing censuses and their registration in Dwelling register is strongly promoted.

In case real-world addresses exist both at building and at building unit level, it is up to each Member State to decide if it is worth to capture or not the geographic location of building units. Location at building level is enough for most of the AD use cases(see figure 7 and following NOTE); however location of building units addresses may have some interest as it would enable theme Addresses to be used as a proxy for population.

In case, geographic data on building units addresses is captured, it is advised to supply it together with Address core data.

4.4.3 Buildings under construction

When it is planned to construct a new building, its address would be useful (e.g. for administrative processes or in case of accident during the construction).

Good Practice 6
Vacant building plots and sites under construction should be allocated addresses as soon as possible.

4.5 Quality

4.5.1 Completeness

INSPIRE does not require national administrations to fill gaps in data availability, but the achievement of the SDGs depends on having adequate data for all parts of the territory.

Core Recommendation 6
Compliant addresses should be available for all addressable objects meeting the criteria of the description in 4.5, on the basis of continuous updating (rather than periodic refreshing of the database).

NOTE: It is recognised that 100% completeness is not achievable in a world where addresses are changing daily, but the minimum aim should be to maintain completeness of 95% or more of core addressable units. A target of 98% might be achievable in most countries.

4.5.2 Location accuracy

Good Practice 7
Coordinates of an address should be accurate to within 10 metres of the true position of the building centroid or entrance, where possible.

NOTE: Relatively high location accuracy is desirable to ensure that individual addresses are readily distinguished within the immediate locality. Particular care is required to locate the address on the correct side of the street.

4.5.3 Thematic accuracy

Core Recommendation 7
Steps must be taken to ensure that address components are recorded consistently, especially in respect of the spelling of address components.

NOTE 1: Geocoding gives good results only if there is common semantics (i.e. mainly same spelling of street names and of address area names) between the file to be geocoded and the geographic address dataset(s). To achieve the best way is to have a single register of addresses, ideally at national level, that may be used as reference data by all public bodies and other stakeholders to capture or to check the addresses of their features of interest (e.g. business register, person register, taxpayer register, customer register, etc.). Other technical approaches, such as point-of-entry data validation, may be helpful in improving thematic consistency.

NOTE 2: Address registration should preferably be standardized and uniform, with respect to spelling, use of spaces, diacritics, capital letters, address additions, etc.

5 Other recommendations

5.1 Coordinate Reference System (CRS)

5.1.1 Case of 2D data

Good Practice 8
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).

5.1.2 Case of 2.5D or 3D data

Only two-dimensional positions are required as core data; the third dimension should be represented if required as a locator, e.g. Floor 3.

5.2 Metadata

Good Practice 9
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 an INSPIRE recommendation (only the INSPIRE Implementing Rules are legally binding for the Member States belonging to the European Union, but the Technical Guidelines are considered necessary to make the European Spatial Data Infrastructure work in practice). For other countries, this is a way to make their data easily manageable by transnational users.

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).

Good Practice 10
Core data corresponding to INSPIRE theme AD should be made available according to the INSPIRE Technical Guidelines for interoperability, for metadata and for services.

NOTE: This is an INSPIRE recommendation (only the INSPIRE Implementing Rules are legally binding for the Member states belonging to the European Union, but the Technical Guidelines are considered necessary to make the European Spatial Data Infrastructure work in practice). For the other countries, this is a way to make their data easily manageable by transnational users.

Good Practice 11
It may be helpful to provide the address semantics in the form of a simple text string as an additional attribute.

NOTE 1: The address semantic would be the concatenation of the address components, ordered according the way these components are usually used by national or local practice in real-life (e.g. locator such as house number + thoroughfare name + postal code + administrative unit name).

NOTE 2: Providing addresses with the address semantics as direct attribute would make the Address product more user-friendly, e.g. facilitating consultation.

6 Considerations for future

6.1.1 Historical addresses

There may be a requirement to capture addresses that have been superseded. Typically these may be required to check identity documents or to geocode business registers; they may also have been used to define regulated areas in legal documents.

Capturing addresses from the past raises many issues: is it achievable, and how?; how far back in the past?; are the benefits worth the costs? The increasingly widespread digitisation of old documents, together with crowdsourcing and machine learning, will tend to reduce costs over time. It may be helpful to share experiences and to encourage research activities on these topics.

6.1.2 Linked data

In order to maximise the usefulness of core address data, it may be advisable to publish it as linked data. However, as this technology is still relatively new, more experience and more feedback on costs and benefits of such practice would be useful to support a potential future recommendation.

6.1.3 A common unique and persistent identifier for addressable objects

This deliverable recommends a unique and persistent identifier for each address feature, i.e. an identifier unique in the geographic Address database.

The GEOSTAT 2 project (from the statistical community) is more demanding as it recommends a unique identifier that would be common to all address stakeholders. This common unique and persistent identifier would enable more reliable geocoding than the simple semantic matching. The principle would be to have a unique identifier of the real-world object.

This requires national co-ordination at a high level, but is achieved in some countries. Here again, it may be helpful to share experience internationally. Some of the obstacles may be organisational (how to persuade the various address stakeholders to work together?); others technical (how to construct such common identifiers? which common life-cycle rules?); and others financial (what are the costs? what are the benefits?).

A first step could be to decide on this common identification at least between the National Statistical Office and to the geographic Address data producer.

Research activities and knowledge exchange should be encouraged on this topic to get better understanding of the current state of the art, and the advantages and difficulties of such a proposal.

6.1.4 Addresses in complex infrastructures

The “recommendation for content” for core data theme Address reflects the current address system, mainly based on street names and house numbers. Core data is expected to be efficient to fulfil the use case requirements for “simple” buildings such as single house or blocks of flats. However, the current address system may be not enough to locate a feature of interest in complex built infrastructures, such as airport, railway station, shopping centres.

An OGC Working Group is dealing with Mobile Location Service, trying to define relevant address or index system for localisation in such complex infrastructures.

The conclusions of this OGC working group may have to be considered to define the core data of theme Address in a few years.

The notion of parent/child relationship between Addresses, as used in the INSPIRE data model, may also help to design the addressing systems of such complex infrastructures.

6.1.5 The third dimension

Where building units are allocated unique addresses, it becomes desirable to geocode in three dimensions so that each unit has a unique location. This is not yet considered a requirement, but future iterations of this document may be expected to make recommendations on this point.

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.

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

Core recommendation

Good practice

7.1.1 Comparison between Core Data and INSPIRE content

Core Recommendation 1

Core data should comprise feature type Address with at least the following attributes: one two dimensional geographic position, one locator (e.g. number or name) if available, and such other address components as are in current use. A unique and persistent identifier is also required.

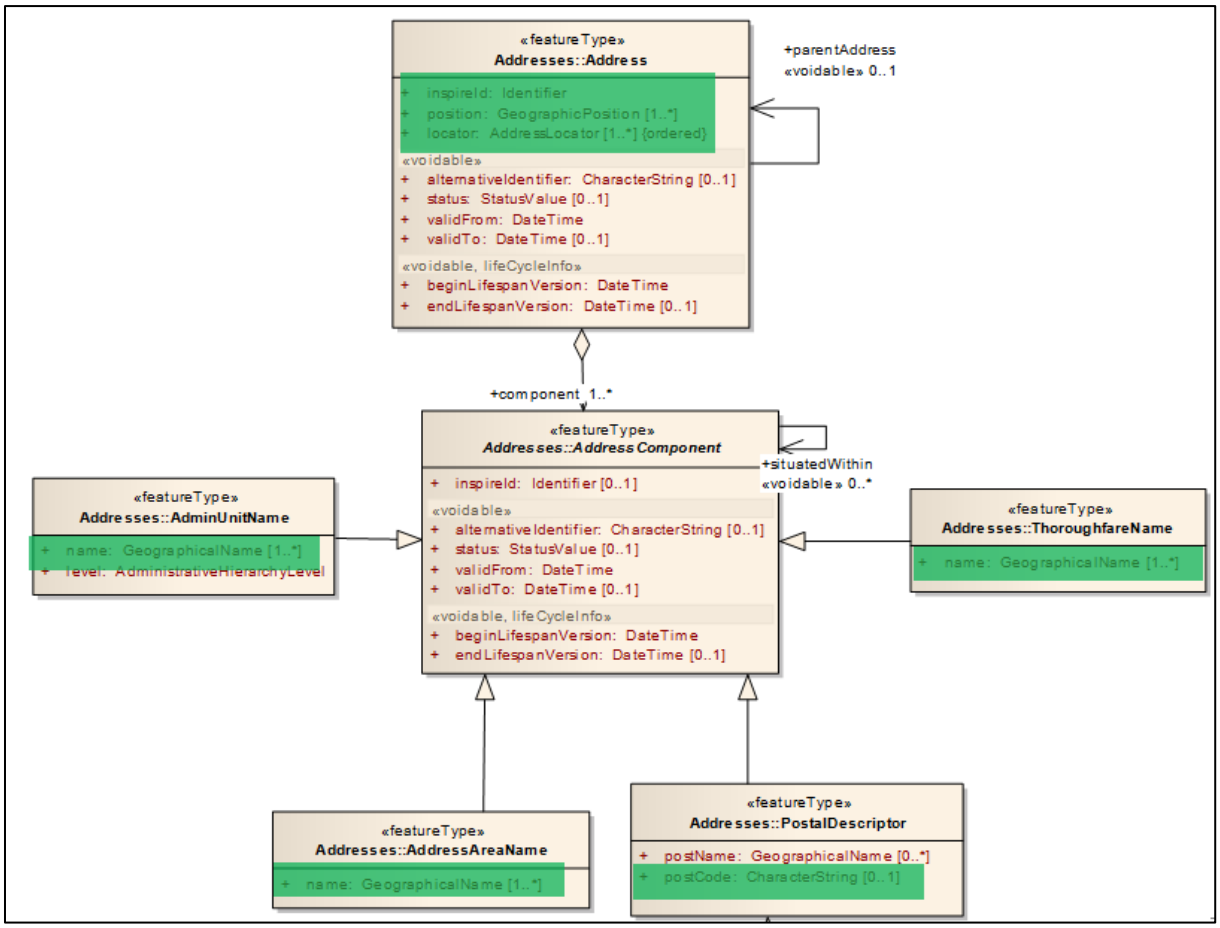


Figure 8: core content from INSPIRE

NOTE 1: Each of the four subclasses of address component is voidable, as not all are relevant to all addresses, but sufficient components are needed to provide unambiguous identification at country level.

NOTE 2: In INSPIRE, there is the association between parent and child addresses, the parent address being the main address and the child address being a sub-address.

Regarding core data, in case of multi-unit buildings, core addresses are the parent addresses whereas sub-addresses are optional

- If parent address at building level, the building parent address is core and the children addresses at building unit level are optional (core recommendation 4)
- If addresses only at building unit level, these addresses are the main addresses and should be included as core (Good practice 6)

Good practice 3

The geographic extent of postal codes should be captured, by its geometry and by specifying what the postal code zone represents.

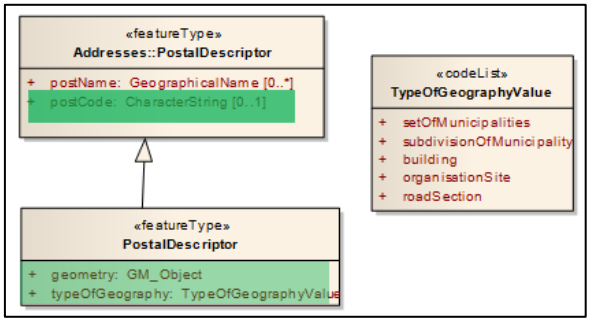


Figure 9: capturing the extent of postal codes

NOTE: the code list provided in this illustration is just an example; it has to be adapted to the practices of each Member State. In case, it is the same value in whole country, it may be documented as metadata on whole dataset rather than as an attribute on each individual feature.

Good practice 4

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

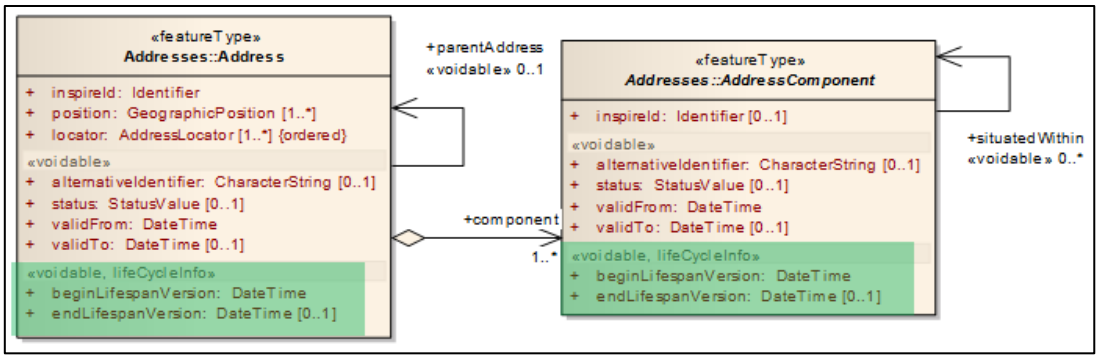


Figure 10: INSPIRE life-cycle attributes

Good practice 11

It may be helpful to provide the address semantics in the form of a simple text string as an additional attribute.

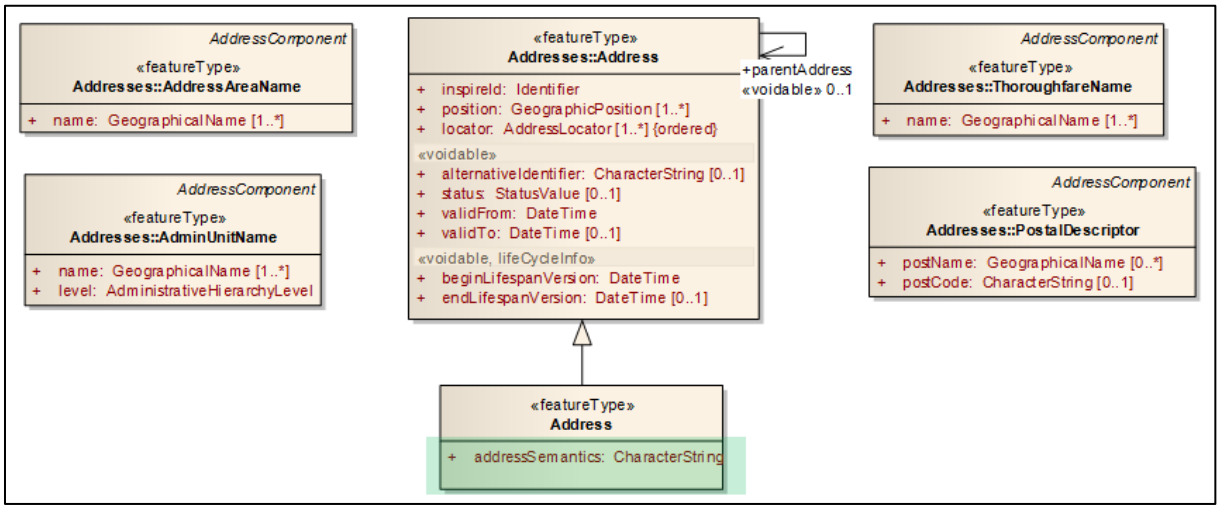


Figure 11: The address semantic as simple attribute

NOTE: INSPIRE has also recognizes the interest of the Address being provided as attribute by providing the data type AddressRepresentation. In INSPIRE, this AddressRepresentation is a datatype, i.e. a complex attribute composed of the main information coming from the address components; this is quite suitable and adapted to the various addressing systems in Europe. Core data is recommended a simple attribute, with the address components being ordered according the national (or local) practice.

7.1.2 Alternative data model

An alternative solution is to consider the address components as attributes directly carried by the Address features. This alternative model would generally be considered as simpler by users but it would have the following drawbacks:

- It is quite less flexible as number of components is limited ; in practice, it may work if the real-world addressing system is relatively simple and homogeneous within the country
- It implies to store lots of redundant information, as each component is duplicated for all the addresses using it.

It is reminded that the modelling approach is up to each Member State. The current document focuses on the minimum content to be captured.

7.2 Other topics

7.2.1 Scope – Data capture

The scope of core data is globally very similar to the scope of INSPIRE theme AD. However, the core data capture recommendations are defining buildings as main addressable objects whereas INSPIRE is more generic. Typically, places that are not devoted to shelter human beings (such as moorings, electricity substations, car parks, street furniture) and places that are smaller than buildings (building units) are not considered as core addressable objects.

7.2.2 Quality

INSPIRE doesn’t define minimum quality recommendations whereas core data is promoting good geometric and thematic accuracy and regular updates.

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.

8.1 Preliminary analysis

The work began with an analysis of the INSPIRE data specifications and a comparison between the INSPIRE context (focus on environmental policy and cross-border interoperability) and the core data context (focus on the Sustainable Development Goals).

This work was supplemented by examination of ISO 19160 Part 1 and the ISA Core Vocabulary for addresses, as well as study of the work of the Universal Postal Union.

This first analysis conducted to the following driving principles:

- The address systems are quite various in Europe and so, the flexible INSPIRE model is quite adapted
- The INSPIRE scope includes a wide range of addressable objects ; for core data, it would be useful to define priorities
- INSPIRE (being based on harmonised delivery of existing data) doesn't provide any quality recommendations whereas strong user requirements were identified (mainly about geometric or thematic accuracy and about update frequency).

8.2 Next steps

The key topic of AD scope was deeply investigated and debated within WG A. For instance, questions were added to a survey of the members of EuroGeographics' Cadastre and Land Registry Knowledge Exchange Network, to ascertain the current state of play in the countries of Europe, mainly regarding addressable objects.

Requirements of the statistical community were taken into account, for example by reviewing the Geostat 3 project reports.

The document "Priority Geospatial Datasets for the European Commission » identified Postal Code as a core reference dataset. In order to take this requirement into account, WG A decided to enrich the AD data model with the geographic extent of the postal codes.

Draft proposals were shared with WG A members at several workshops, leading to further development and refinement of the proposals. Eventually, the document was submitted to the review of the geo-statistical community.