INSPIRE
Infrastructure for Spatial Information in Europe

Drafting Team "Data Specifications" Methodology for the development of data specifications

Title	Drafting Team "Data Specifications" – deliverable D2.6: Methodology for the development of data specifications
Status	Baseline version
Creator	INSPIRE Drafting Team "Data Specifications"
Date	2008-06-20
Subject	Methodology for the development of data specifications
Publisher	INSPIRE Drafting Team "Data Specifications"
Type	Text
Description	Proposed methodology for the development of INSPIRE data specifications for the spatial data themes as specified in the Annexes of the INSPIRE Directive
Contributor	Members of the INSPIRE Drafting Team "Data Specifications", INSPIRE Spatial Data Interest Communities & Legally Mandated Organisations, INSPIRE Consolidation Teams and other Drafting Teams
Format	MS Word (doc), pdf
Source	Drafting Team “Data Specifications”
Rights	Public
Identifier	D2.6_v3.0.doc, D2.6_v3.0.pdf
Language	En
Relation	n/a
Coverage	Project duration

Table of contents

Foreword ... 5

Introduction ... 7

1 Scope ... 11

2 Normative references ... 11

3 Terms and abbreviations ... 12

3.1 Terms ... 12
3.2 Abbreviations ... 18
3.3 Verbal forms for the expression of provisions ... 20
3.4 References within the document ... 20

4 Background and principles ... 22

4.1 Requirements as stated in the INSPIRE Directive ... 22
4.1.1 Articles of the Directive ... 22
4.1.2 Recitals in the Directive ... 25
4.2 Data interoperability components ... 25

5 General aspects ... 27

5.1 Principles ... 27
5.2 The development process in the INSPIRE Work Programme ... 27
5.3 Roles in a Thematic Working Group ... 28
5.4 Maintaining the INSPIRE data specifications ... 29

6 Steps in the development of INSPIRE data specifications ... 31

6.1 Overview ... 31
6.2 Results ... 34
6.3 Step CT-TWG-1: Use case development ... 35
6.3.1 Responsible party ... 35
6.3.2 Step results ... 35
6.3.3 Exit conditions ... 35
6.3.4 Resources ... 35
6.3.5 Additional information ... 35
6.3.6 Detailed description ... 36
6.4 Step TWG-2: Identification of user requirements and spatial object types ... 37
6.4.1 Responsible party ... 37
6.4.2 Step results ... 37
6.4.3 Exit conditions ... 37
6.4.4 Resources ... 37
6.4.5 INSPIRE data specification template ... 37
6.4.6 Additional information ... 37
6.4.7 Detailed description ... 37
6.5 Step TWG-3: As-is analysis ... 39
6.5.1 Responsible party ... 39
6.5.2 Step results ... 39
6.5.3 Exit conditions ... 39
6.5.4 Resources ... 39
6.5.5 Additional information ... 39
6.5.6 Detailed description ... 39
6.6 Step TWG-4: Gap analysis ... 40
6.6.1 Responsible party ... 40
6.6.2 Step results ... 40
6.6.3 Exit conditions ... 40
6.6.4 Resources ... 40

6.6.5 Additional information 40
6.6.6 Detailed description 40
6.7 Step TWG-5: Data specification development 42
6.7.1 Responsible party 42
6.7.2 Step results 42
6.7.3 Exit conditions 42
6.7.4 Resources 42
6.7.5 Additional information 42
6.7.6 Detailed description 42
6.8 Step CT-6: Implementation, test and validation 43
6.8.1 Responsible party 43
6.8.2 Step results 43
6.8.3 Exit conditions 44
6.8.4 Resources 44
6.8.5 Additional information 44
6.8.6 Detailed description 44
6.9 Step CT-TWG-7: Cost-benefit considerations 45
6.9.1 Responsible party 45
6.9.2 Step results 45
6.9.3 Exit conditions 45
6.9.4 Resources 45
6.9.5 Additional information 45
6.9.6 Detailed description 45

7 Recommendations for INSPIRE data specifications 46

7.1 General 46
7.2 Identification information 46
7.3 Content and structure 47
7.4 Reference system 47
7.5 Data quality 47
7.6 Data capture information (optional) 48
7.7 Maintenance information (optional) 48
7.8 Portrayal information (optional) 48
7.9 Delivery 48
7.10 Additional information (optional) 48
7.11 Metadata 49
7.12 Service specifications 49

Annex A (informative) Harmonisation Guidelines 50

A.1 Introduction 50
A.2 INSPIRE principles 51
A.3 Terminology 51
A.4 Reference model 51
A.5 Rules for application schemas 52
A.6 Spatial and temporal aspects 55
A.7 Multilingual text 58
A.8 Coordinate referencing – units of measurements 59
A.9 Object referencing modelling 59
A.10 Data transformation model/guidelines 59
A.11 Portrayal 62
A.12 Identifier management 63
A.13 Registers and Registries 69
A.14 Metadata 71
A.15 Maintenance 72
A.16 Quality 73
A.17 Data transfer 75
A.18 Consistency between data 75
A.19 Derived reporting/multiple representation 77
A.20 Data capture 79
A.21 Conformance 82

Annex B (informative) Management of connections at international boundaries 83

B.1 Introduction 83
B.2 Which spatial objects have to be matched? 83
B.2.1 Vector data 83
B.2.2 Coverage data 86
B.3 How to match spatial objects? 87
B.3.1 Geometry edge-matching: 88
B.3.2 Creation of connecting spatial objects: 91
B.4 How to match attributes? 92
B.4.1 Matching rule for duplicated spatial objects located on the national boundary 92
B.4.2 Matching rule for spatial objects crossing the national boundary 92
B.5 Which national boundaries? 93
B.6 Organisational point of view 94

Annex C (informative) Data specification document template and example 95

Annex D (informative) Example application schemas in UML 96

D.1 General remarks 96
D.2 Application schema example: "Cadastral parcels" 97
D.3 Application schema example: "Elevation" 98
D.4 Application schema example: "Meteorology" 98
D.4.1 Synoptic Observations 98
D.4.2 Analyses 101
D.5 Application schema example: "Geology" 103
D.5.1 Mapped Feature 103
D.5.2 Geologic Unit 104

Annex E (informative) Use case template 106

Annex F (informative) Checklist for data interoperability 108

F.1 General remarks 108
F.2 Overview 108
F.3 The checklist 108

Annex G (informative) Tools 121

G.1 UML model 121

Bibliography 122

Foreword

INSPIRE is a Directive proposed by the European Commission in July 2004 setting the legal framework for the establishment of the Infrastructure for Spatial Information in the European Community, for the purposes of Community environmental policies and policies or activities which may have an impact on the environment.

INSPIRE will be based on the infrastructures for spatial information that are created and maintained by the Member States. The components of those infrastructures include: metadata, spatial data themes (as described in Annexes I, II, III of the Directive), spatial data services; network services and technologies; agreements on data and service sharing, access and use; coordination and monitoring mechanisms, processes and procedures.

The guiding principles of INSPIRE are that the infrastructures for spatial information in the Member States will be designed to ensure that spatial data are stored, made available and maintained at the most appropriate level; that it is possible to combine spatial data and services from different sources across the Community in a consistent way and share them between several users and applications; that it is possible for spatial data collected at one level of public authority to be shared between all the different levels of public authorities; that spatial data and services are made available under conditions that do not restrict their extensive use; that it is easy to discover available spatial data, to evaluate their fitness for purpose and to know the conditions applicable to their use.

The text of the INSPIRE Directive is available from the INSPIRE web site (http://www.ec-gis.org/inspire). The Directive identifies what needs to be achieved, and Member States have two years from the date of adoption to bring into force national legislation, regulations, and administrative procedures that define how the agreed objectives will be met taking into account the specific situation of each Member State. To ensure that the spatial data infrastructures of the Member States are compatible and usable in a Community and transboundary context, the Directive requires that common Implementing Rules (IR) are adopted in a number of specific areas. Implementing Rules are adopted as Commission Decisions, and are binding in their entirety. The Commission is assisted in the process of adopting such rules by a regulatory committee composed by representatives of the Member States and European Parliament[1]. The committee is chaired by a representative of the Commission (this is known as the Comitology procedure). The committee will be established within three months from the entry in force of the Directive.

The IR will be shaped in their legal structure and form by the Commission legal services on the basis of technical documents prepared by especially convened Drafting Teams, for each of the main components of INSPIRE: metadata, data specifications, network services, data and service sharing, and monitoring procedures. For data specifications, the technical documents for each spatial data theme will be prepared by especially convened Thematic Working Groups.

This document represents a contribution of the Data Specification Drafting Team.

An earlier version of this document (version 2.0) was published on the INSPIRE web site for public view and commenting by registered SDICs and LMOs. 1148 comments were received and resolved to produce this version. The comment resolution process included a workshop with representatives of SDICs and LMOs. Based on the discussions, the Drafting Team “Data Specifications” proposed comment resolutions that were reviewed by the Consolidation Team. The table containing the comments and the resolution is available on the INSPIRE web-site (http://www.ec-gis.org/inspire/reports/ImplementingRules/DataSpecifications/D2.6_Comments-Resolutions-20062008.pdf).
This baseline version (version 3.0) is published on the INSPIRE web site and will be used by the Thematic Working Groups to prepare the data specifications for the IR for the interoperability for spatial data sets and services. It is expected that this methodology will be updated during the data specification development process, if requirements for changes are identified.

---

[1] The implementing rules are formally adopted through the comitology procedure that has been amended by Council Decision of 17 July 2006 (2006/512/EC). Under the new regulation, the Parliament and the Council are on equal footing for all comitology procedures related to co-decision acts. As a consequence, all measures must be ratified by all three institutions to come into force.

It is important to note that this document is not a draft Implementing Rule, but a document that assists in the development of the thematic data specifications that will eventually become Implementing Rules.

This document is based on the results from the projects “Reference Specifications for Europe (RISE)[2]” and “Marine Overlays on Topography for Annex II Valuation and Exploitation (MOTIIVE)[3]”. These projects were co-funded by the European Commission within the 6th Framework Programme for research (2002-2006).

The document will be publicly available as a ‘non-paper’, as it does not represent an official position of the Commission and as such cannot be invoked in the context of legal procedures.

[2] http://www.eurogeographics.org/eng/03_RISE.asp
[3] http://www.motiive.net/

Introduction

This document contains the baseline version of the methodology for the development of INSPIRE data specifications (document identifier: D2.6).

One of the main tasks of the INSPIRE programme is to enable the interoperability and, where practicable, harmonisation of spatial data sets and services within Europe. Here, it is important to note that interoperability has to go beyond any particular community, but take the various cross-community information needs into account. If one takes a look at the huge difference in the scope of the different themes (from reference systems to hydrography and from cadastral parcel to atmospheric conditions), the question does arise about the specific requirements of and for interoperability and harmonisation of the geographic information. These were also the questions faced by the Drafting Team "Data Specification" and one of the contributions of the Drafting Team is the identification of a set of *interoperability components*, which make the concepts of interoperability and harmonisation more tangible. Examples of interoperability components addressed in this document are: rules for application schemas, coordinate referencing and units model, identifier management, multi-lingual text and cultural adaptability, object referencing modelling, multiple representations (levels of detail) and consistency, and more. All these components do apply to (nearly) all themes identified within INSPIRE and this document together with the Generic Conceptual Model describes approaches to these shared components. Using this framework across the different themes will therefore result in a first level of interoperability.

It is important to note that “interoperability” is understood as providing access to spatial data sets as specified in Article 4 of the Directive through network services in a representation that allows for combining them with other such spatial data sets in a coherent way. This includes agreements about the different interoperability components. In other words, by enabling interoperability data can be used coherently, independent of whether the existing data set is actually changed (harmonised) or “just” transformed by a download service for publication in INSPIRE depending on the approach taken by the Member State. It is expected that these agreements will be based on existing data interoperability or harmonisation activities, whenever feasible and in-line with the environmental requirements.

The starting point for the development of INSPIRE data specifications is the input delivered by the LMOs and SDICs with their reference material and domain knowledge. Further and more specifically the foundation is formed by the internationally accepted standards reflecting the collective state-of-the-art knowledge (such as the reference model described in ISO 19101).

The individual themes (as defined in the Annexes I, II and III of the Directive and refined in document D2.3 'Definition of Annex Themes and Scope') will be modelled based on document D2.5 ‘Generic Conceptual Model’. This document specifies the process. The result are data product specifications for the individual themes, i.e. conceptual information models that describe the relevant classes, their attributes, relationships, constraints, and possibly also operations as well as other appropriate information like data capturing information or data quality requirements. Care has to be taken that common or shared spatial object types relevant in multiple themes are identified and modelled in a consistent manner. This could then be considered a second level of interoperability: agreement on the shared (formal) semantics between the different themes. Note that the spatial characteristics of a spatial object will be represented by vector geometries, coverage functions and/or references to gazetteer entries.

This document specifies how individual INSPIRE spatial data themes will be modelled based on the user requirements, the INSPIRE Generic Conceptual Model (document D2.5) and the relevant international standards. It provides a process model and tools to assist in the process.

How the geographic information will actually be encoded for the transfer process will be described in document D2.7 ‘Guidelines for the encoding of spatial data’ (the third level of geographic information interoperability).

This methodology is applicable for INSPIRE data specifications. It is not required that it will be applied for the modelling of data specifications at the national level. What is important is that each Member State is able to transform existing data sets to the INSPIRE data specifications and publish the transformed data via network services. On the other hand, this methodology is expected to influence modelling activities at the national level, because it adds value to the national spatial data infrastructure and simplifies synchronisation with the INSPIRE data specifications.

Besides the documents D2.3, D2.5, D2.7 and D2.8.m.n[4], this document is also related to other INSPIRE documents and registers:

- The terms used in this document are drawn from the "INSPIRE Glossary".
- INSPIRE application schemas will be based on the Generic Conceptual Model and maintained in the "Consolidated INSPIRE UML model" that also includes the external schemas, for example, the harmonised model of the ISO 19100 series published by ISO/TC 211. INSPIRE application schemas will be developed for every theme listed in the annexes of the INSPIRE Directive.
- The "INSPIRE Feature Concept Dictionary Register" is used to manage the names, definitions and descriptions of all spatial object types used in INSPIRE application schemas. In the future, the register may be extended to manage properties, too.
- Other registers include a coordinate reference system register, a feature catalogue register and a code list register.
- The implementing rule on metadata and associated guidelines.
- The implementing rules on network services and associated guidelines

Figure 1 below illustrates relationships from the point of view of the data specifications. The boxes denote INSPIRE Implementing Rule documents or supporting documents, the cylinders registries. The arrows denote dependencies, the areas with dashed boundaries denote areas of responsibility.

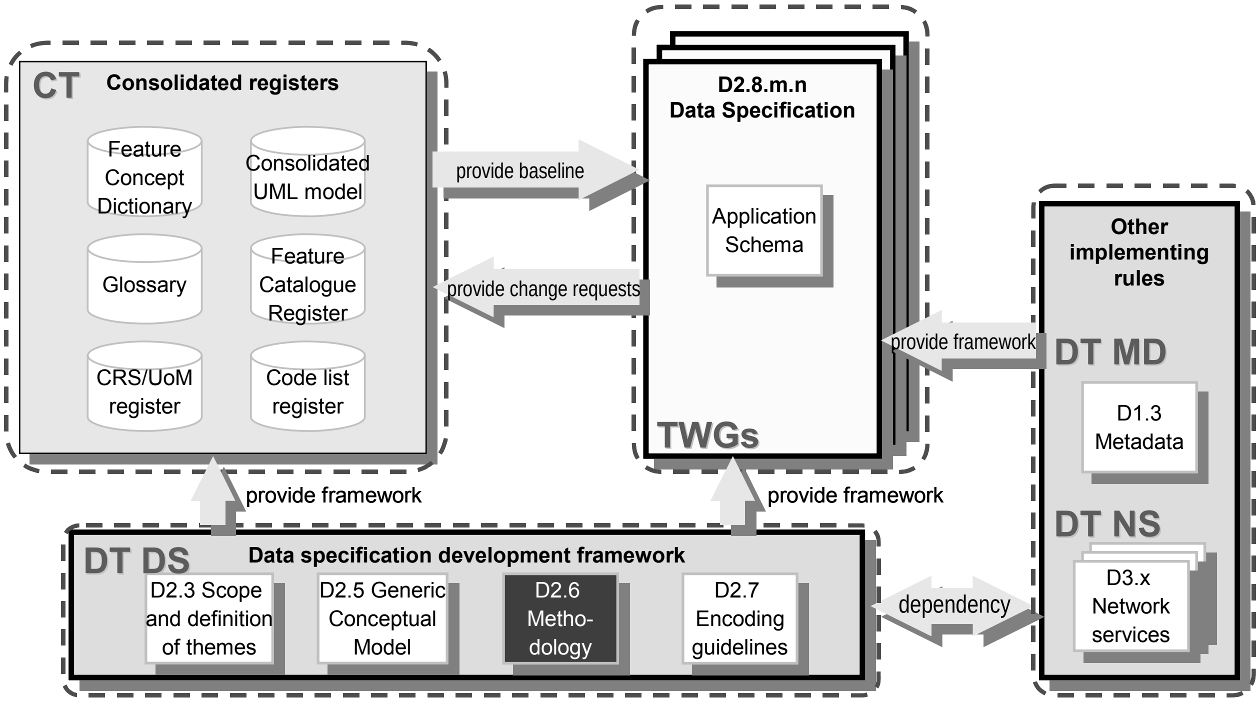


Figure 1 – The Generic Conceptual Model as part of the data specification development framework

Since the conceptual modelling framework of INSPIRE is based in the ISO 19100 series of International Standards, in-depth knowledge about this series is required in every team developing an INSPIRE data specification.

The intent of the methodology can be illustrated as follows:

A simplified view to the processing of data today is shown in the following figure. In most cases, each Member State uses input data according to different, often undocumented or ill-documented data

[4] "m" is the number of the annex and "m" the sequential number of the theme within the annex.

specifications and uses different methods to process the input data to produce more or less similar information relevant for policies within the Community.

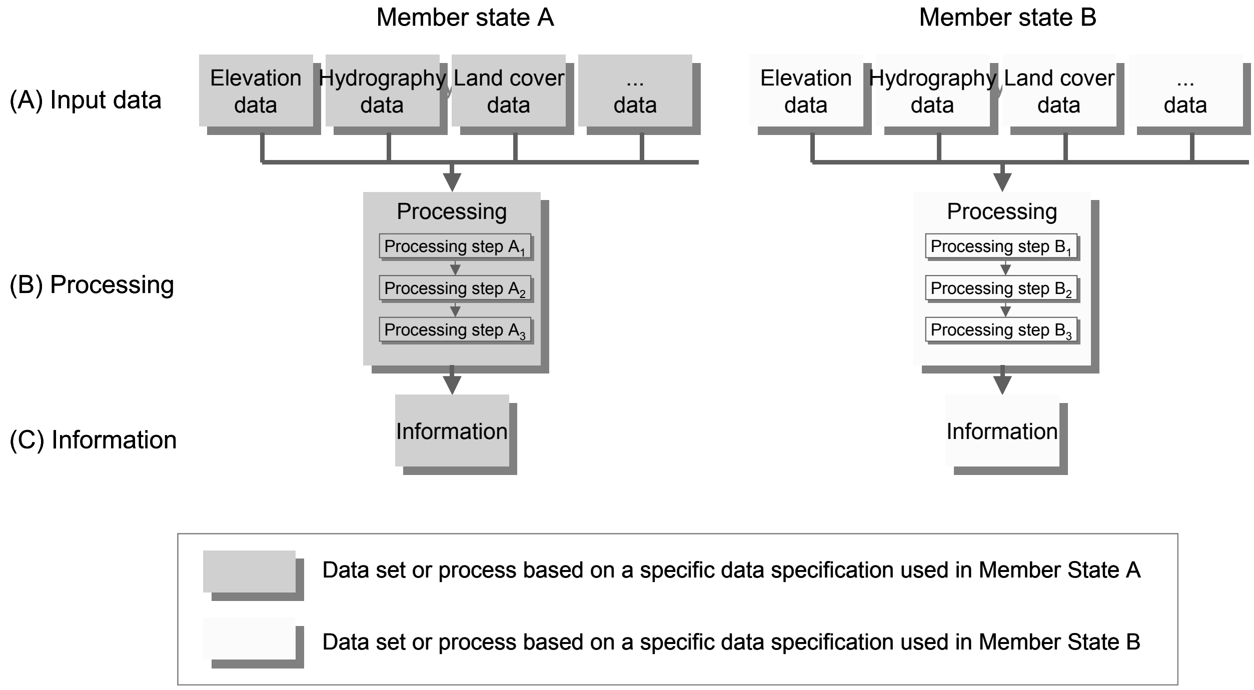


Figure 2 – Current situation: Data stovepipes

The methodology described in this document aims at interoperability for data in INSPIRE. Therefore it is necessary to understand the user requirements. Based on the user requirements, it focuses on the development of harmonised data specifications for the input data so that all input data from the different Member States follows the same data specifications – and in principle could use the same processing steps to derive the information although this harmonisation step is out of scope for this document.

The input data in the Member States and their maintenance procedures will typically be more-or-less the same prior to INSPIRE, but in addition the data will be provided by the network services of the Member States following the harmonised data specifications.

It is also worth noting that the methodology includes provisions to develop theme specific data specifications, e.g. data specifications for transport networks, incrementally so that new user requirements from a new application can be used to amend the existing data specification.

Figure 2 shows the target situation, where INSPIRE-conformant data specifications are to be applied in the highlighted area. These data specifications shall be developed using the proposed methodology of this document:

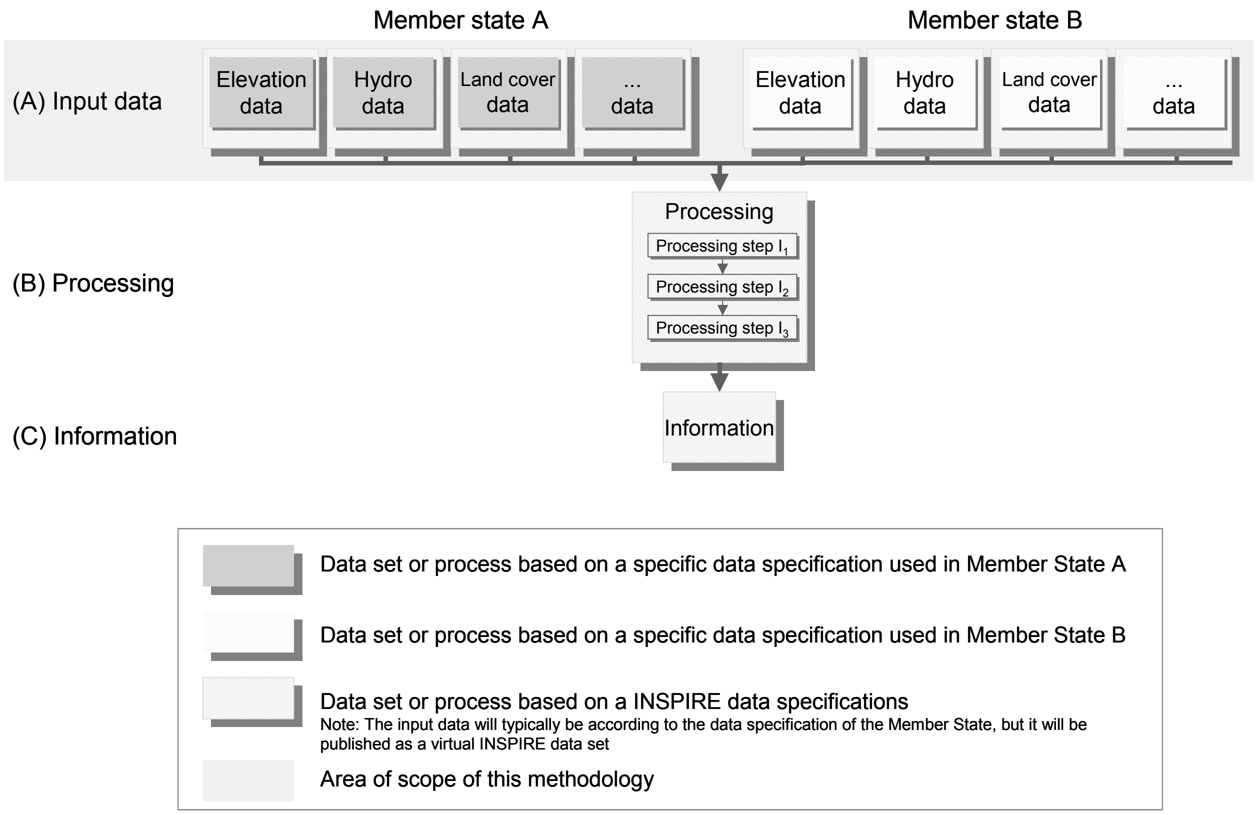


Figure 3 – Target situation: Harmonised data views, eliminating data stovepipes

In general it is important to keep in mind that INSPIRE does not foresee any arrangements in the Implementing Rules to update/manipulate spatial data.

1 Scope

The objective of the document is to facilitate the process of creating harmonised data specifications for the INSPIRE themes (as defined in the Annexes I, II and III of the Directive and refined in D2.3 'Definition of Annex Themes and Scope'). The individual Annex themes will be modelled based on the user requirements, the INSPIRE Generic Conceptual Model and the relevant international and industrial standards that are referenced. In this document the process is specified how this should be achieved. The result will then be data product specifications for the individual themes, i.e. conceptual information models that describe the relevant classes, their attributes, relationships and constraints, as well as other information as appropriate like data capturing information or data quality requirements.

Clause 4 provides an overview reflecting the articles and recitals in the Directive that are relevant to this document.

Clause 5 summarises the organisational aspects of the data specification development process in INSPIRE.

Clause 6 introduces a general step-wise methodology to develop harmonised data specifications.

Clause 7 lists recommendations for the content of INSPIRE data specifications.

The Annexes A and B provide guidelines and best practise for data interoperability. The two Annexes are structured according to the harmonisation components identified in D2.5 Generic Conceptual Model.

Annex C provides a template for the INSPIRE data specification.

NOTE This template has been moved to a separate document.

Annex D provides some examples to illustrate application schema modelling in UML.

Annex E holds a template for the description of a use case.

Annex F provides a checklist that may assist Thematic Working Groups in identifying the relevant data interoperability aspects.

Annex G describes tools that may be used in the process of specification development.

NOTE This document complements these other documents:

- deliverable D2.5 specifying the Generic Conceptual Model for INSPIRE data specifications,
- deliverable D2.7 providing rules for the encoding of spatial data.

2 Normative references

ISO/TS 19103:2005, Geographic Information – Conceptual Schema Language

EN ISO 19107:2005, Geographic Information – Spatial Schema

EN ISO 19108:2005, Geographic Information – Temporal Schema

ISO 19108:2002/Cor 1:2006, Geographic Information – Temporal Schema, Technical Corrigendum 1

EN ISO 19109:2005, Geographic Information – Rules for Application Schemas

EN ISO 19110:--[5], Geographic Information – Methodology for feature cataloguing

[5] to be published, an amendment to EN ISO 19110:2006 is currently at Committee Draft stage

EN ISO 19111:2007, Geographic Information – Spatial referencing by coordinates

EN ISO 19113:2005, Geographic Information – Quality principles

EN ISO 19114:2005, Geographic Information – Quality evaluation procedures

EN ISO 19114:2005/AC:2006, Geographic Information – Quality evaluation procedures, Technical Corrigendum 1

EN ISO 19115:2005, Geographic Information – Metadata

ISO 19115/Cor.1:2006, Geographic Information – Metadata, Technical Corrigendum 1

EN ISO 19119:2006, Geographic Information – Services

EN ISO 19123:2007, Geographic Information – Schema for coverage geometry and functions

ISO 19126:--[6], Geographic Information – Feature Concept Dictionaries and Registers

ISO 19131:2007, Geographic Information – Data Product Specification

EN ISO 19135:2007, Geographic Information – Procedures for item registration

ISO 19136:2007, Geographic Information – Geography Markup Language

ISO/TS 19138:2006, Geographic Information – Data quality measures

ISO/TS 19139:2007, Geographic Information – Metadata – XML schema implementation

ISO/IEC 19501:2005, Information technology — Open Distributed Processing — Unified Modelling Language (UML) Version 1.4.2

INSPIRE Glossary, (to be published http://www.ec-gis.org/inspire/ds/)

INSPIRE DS-D2.5, Generic Conceptual Model, v3.0

Terms of Reference for developing Implementing Rules laying down technical arrangements for interoperability and harmonisation of spatial data sets, July 2007

3 Terms and abbreviations

3.1 Terms

The terms in this sub-clause are taken from the “Glossary of Generic Geographic Information Terms in Europe” that specifies the terminology used in the INSPIRE Implementing Rule documents. The glossary is managed as a register in accordance with ISO 19135. It will be published on the INSPIRE website: http://www.ec-gis.org/inspire/ds/.

actor

An *actor* specifies a role played by a user or any other system that interacts with the system under consideration (UML 2.1.2 Superstructure, Clause 16)

application schema
conceptual schema for data required by one or more applications [ISO 19101]

class

[6] to be published, currently at Committee Draft stage

description of a set of **objects** that share the same properties, constraints, and semantics [UML 2.1.2 - modified]

codelist
value domain including a code for each permissible value [ISO 19136]

conceptual model
model that defines concepts of a universe of discourse [ISO 19101]

conceptual schema
formal description of a **conceptual model** [ISO 19101]

EXAMPLE ISO 19107 contains a formal description of geometrical and topological concepts using the conceptual schema language UML.

coordinate reference system
systems for uniquely referencing spatial information in space as a set of coordinates (x,y,z)
and/or latitude and longitude and height, based on a geodetic horizontal and vertical datum [INSPIRE Directive]

NOTE 1 ISO 19111 defines coordinate reference system as a coordinate system that is related to the real world by a datum.

EXAMPLE 1 A national coordinate system with the datum ETRS89.

NOTE 2 There is an ISO work item to provide an addendum 19111-2 to define parametric CRS.

EXAMPLE 2 The ICAO standard atmosphere; or ISO 2533:1975 which uses a pressure as a coordinate.

NOTE 3 Although the definition in the INSPIRE Directive is strictly seen restricted to spatial reference systems, temporal reference systems are understood as covered by the term coordinate reference systems as well, because temporal information has to be associated with a reference system just like spatial geometries. ISO 19111 also recognises temporal reference systems explicitly.

EXAMPLE 3 The Gregorian calendar is a temporal reference system.

coverage
spatial object that acts as a function to return values from its range for any direct position within its spatial, temporal or spatiotemporal domain [ISO 19123 - modified]

EXAMPLE Orthoimage, digital elevation model (as grid or TIN), point grids etc

data interoperability component
individual aspect that will be addressed to support the **interoperability** of **spatial data sets** [DS-D2.5]

EXAMPLE Rules for application schemas, identifier management, terminology, etc. are examples of the components.

data interoperability process
process of developing **harmonised data product specifications** and implementing the necessary arrangements to transform **spatial data** into **interoperable spatial data** [DS-D2.5]

NOTE Two general options exist: The reference version of the spatial data set may either be changed/restructured itself (“harmonised”) or it may be kept as-is and the transformation may occur on-the-fly whenever a spatial data service operating on the spatial data set is invoked. In cases where the location of a spatial object has to be changed to comply with Article 10 (2), it is expected that the location information in the reference version of the spatial data set is updated to reflect the mutual consent.

data product
data set or data set series that conforms to a **data product specification** [ISO 19131]

data product specification
detailed description of a data set or data set series together with additional information that will enable it to be created, supplied to and used by another party [ISO 19131]

data set
identifiable collection of data [ISO 19115]

data set series
collection of data sets sharing the same product specification [ISO 19115]

data specification
(used as a synonym to data product specification)

NOTE If the context is unambiguous, "data specification" is often used instead of "INSPIRE data specification" to improve readability.

entity
real-world phenomenon [DS-D2.5]

enumeration
data type whose values are enumeration literals [UML 2.1.2 - modified]

ESDI
European spatial data infrastructure as built based on the INSPIRE Directive [DS-D2.5]

NOTE The ESDI is expected to include, for example, additional content beyond the data provided by those that are legally mandated to do so according to the directive.

endonym
name of a **spatial object** in one of the languages occurring in that area where the **spatial object** is situated [UNGEGN Glossary of Terminology - modified]

exonym
name used in a specific language for a **spatial object** situated outside the area where that language is spoken, and differing in its form from the name used in an official or well-established language of that area where the **spatial object** is located [UNGEGN Glossary of Terminology - modified]

external object identifier
unique object identifier which is published by the responsible body, which may be used by external applications to reference the **spatial object** [DS-D2.5]

feature
abstraction of real world phenomena [ISO 19101]

NOTE The term "(geographic) feature" as used in the ISO 19100 series of International Standards, in other specifications like IHO S-57, and in this document is synonymously with **spatial object** as used in this document. Unfortunately "spatial object" is also used in the ISO 19100 series of International Standards, however with a different meaning: a spatial object in the ISO 19100 series is a spatial geometry or topology.

feature catalogue
catalogue(s) containing definitions and descriptions of the **spatial object types**, their attributes and associated components occurring in one or more **spatial data sets**, together with any operations that may be applied [ISO 19110 – modified]

feature concept

concept that may be specified in detail as one or more **spatial object types** [ISO/DIS 19126 – modified]

EXAMPLE The feature concept 'road' may be used to specify several different spatial object types, each with a different set of properties appropriate for a particular application. For a travel planning application, it might have a limited set of attributes such as name, route number, location and number of lanes, while for a maintenance application it might have an extensive set of attributes detailing the structure and composition of each of the layers of material of which it is composed.

feature concept dictionary
dictionary that contains definitions of and related descriptive information about concepts that may be specified in detail in a **feature catalogue** [ISO/DIS 19126]

feature-related concepts
abstract specification of the semantics of a property of a spatial object type [DS-D2.5]

gazetteer
directory of instances of a class or classes of features containing some information regarding position [ISO 19112]

NOTE A gazetteer can be considered as a geographical index or directory.

general feature model
meta-model for **spatial object** types and their property types specified by ISO 19109 [DS-D2.5]

geographic identifier
spatial reference in the form of a label or code that identifies a location [ISO 19112]

EXAMPLE 1 Place names: Paris, Rhine, Mont Blanc

EXAMPLE 2 Postal codes: 53115, 01009, SW1, IV19 1PZ

harmonised data product specifications
set of **data product specifications** that support the provision of access to **interoperable spatial data** through spatial data services in a representation that allows for combining it with other **interoperable spatial data** in a coherent way [DS-D2.5]

NOTE 1 The harmonised data product specifications will be based on the data interoperability components.

NOTE 2 The harmonised data product specification is not intended to replace or deprecate existing data specifications that are currently in use.

homologous spatial objects
set of **spatial objects** that correspond to the same real-world phenomenon, but are described by different information according to the different levels of details or point of views [DS-D2.5]

identifier
linguistically independent sequence of characters capable of uniquely and permanently identifying that with which it is associated [ISO 19135]

INSPIRE application schema
application schema specified in an INSPIRE data specification [DS-D2.5]

INSPIRE data specification
harmonised data product specification for a **theme** adopted as an Implementing Rule [DS-D2.5]

internal object identifier
unique object identifier which is used internally and is not intended to be used to identify or reference the **spatial object** by external applications [DS-D2.5]

interoperability
possibility for spatial data sets to be combined, and for services to interact, without repetitive manual intervention, in such a way that the result is coherent and the added value of the data sets and services is enhanced [INSPIRE Directive]

interoperable spatial data
spatial data conformant to the **harmonised data product specifications** [DS-D2.5]

level of detail
quantity of information that portrays the real world

NOTE The concept comprises data capturing rules of spatial object types, the accuracy and the types of geometries, and other aspects of a data specification. In particular, it is related to the notions of scale and resolution.

metadata
information describing **spatial data sets** and spatial data services and making it possible to discover, inventory and use them [INSPIRE Directive]

NOTE A more general definition provided by ISO 19115 is "data about data"

multilingual
in or using several languages [Oxford Dictionary]

multiple representation
representation of the relationship between **homologous spatial objects** [DS-D2.5]

object
in this document used synonymous with ***spatial object***

object referencing
consistent method of referencing **spatial data** to location using existing **spatial objects** [DS-D2.5]

ontology
representation of a set of concepts within a domain and the relationships between those concepts [Wikipedia]

profile
set of one or more base standards or subsets of base standards, and, where applicable, the identification of chosen clauses, classes, options and parameters of those base standards, that are necessary for accomplishing a particular function [ISO 19106]

NOTE A profile is derived from base standards so that by definition, conformance to a profile is conformance to the base standards from which it is derived.

reference data
spatial objects that are used to provide location information in **object referencing** [DS-D2.5]

NOTE Typical reference data are topographic or cadastral data.

reference model
architectural framework for a specific context, e.g. an application or an information infrastructure [DS-D2.5]

EXAMPLE ISO 19101 and the OGC Reference Model are reference models

register
set of files containing identifiers assigned to items with descriptions of the associated items [ISO 19135]

registry
information system on which a **register** is maintained [ISO 19135]

resolution
Resolution expresses the size of the smallest object in a spatial data set that can be described. It refers to the amount of detail that can be discerned. It is also known as granularity. Resolution is also limited because geo-spatial databases are intentionally generalised. Resolution affects the degree to which a database is suitable for a specific application. [INSPIRE Position paper Reference Data and Metadata Position paper v4.2]

NOTE In a regular gridded coverage (e.g. raster data in 2D grids), resolution expresses the area in the real world that is represented by the size of a raster cell.

scale
The relation between the dimensions of features on a map and the geographic objects they represent on the earth, commonly expressed as a fraction or a ratio. A map scale of 1/100,000 or 1:100,000 means that one unit of measure on the map equals 100,000 of the same unit on the Earth. [INSPIRE Position paper Reference Data and Metadata Position paper v4.2]

NOTE A data set has no real scale as the user may choose at which scale he or she wants to display it. Nevertheless, it is usually not appropriate to display a data set at an arbitrary scale, so typically some scale information is associated with a data set (or its data product specification). This may be a scale range or a medium scale (i.e. it is relevant to use and display the data set around this “medium scale”).

spatial data
data with a direct or indirect reference to a specific location or geographic area [INSPIRE Directive]

NOTE The use of the word “spatial” in INSPIRE is unfortunate as in the everyday language its meaning goes beyond the meaning of “geographic” – which is considered by the Drafting Team as the intended scope – and includes subjects such as medical images, molecules, or other planets to name a few. However, since the term is used as a synonym for geographic in the draft Directive, this document uses the term “spatial data” as a synonym for the term “geographic data” used by the ISO 19100 series of International Standards and which is defined as “data with implicit or explicit reference to a location relative to the Earth.”

spatial data set
identifiable collection of spatial data [INSPIRE Directive]

spatial object
abstract representation of a real-world phenomenon related to a specific location or geographical area [INSPIRE Directive]

NOTE It should be noted that the term has a different meaning in the ISO 19100 series. It is also synonymous with "(geographic) feature" as used in the ISO 19100 series.

spatial object type
classification of **spatial object**s [DS-D2.5]

EXAMPLE Cadastral parcel, road segment or river basin are all examples of potential spatial object types.

NOTE In the conceptual schema language UML a spatial object type will be described by a class with stereotype <<FeatureType>>.

spatial reference system
system for identifying position in the real world [ISO 19112]

NOTE Spatial reference systems do not necessarily use coordinates to identify a position.

EXAMPLE Geographic coordinates describing positions on the Earth surface (coordinate reference system), linear measurements along a river centreline from the intersection of a bridge (linear reference system), postal codes identifying the extent of postal zones (gazetteer).

spatial schema
conceptual schema of spatial geometries and topologies to be used in an **application schema** [DS-D2.5]

temporal reference system
reference system against which time is measured [ISO 19108]

temporal schema
conceptual schema of temporal geometries and topologies to be used in an **application schema** [DS-D2.5]

thematic identifier
descriptive **unique object identifier** applied to **spatial objects** in a defined information **theme** [DS-D2.5]

EXAMPLE an administrative code for administrative area objects in the administrative units theme, a parcel code for parcel objects in a cadastral theme

theme
grouping of **spatial data** according to Annex I, II and III of the INSPIRE Directive [DS-D2.5]

transformation
An act, process, or instance of change in structure, appearance, or character [Webster]

NOTE In the context of INSPIRE two types of transformations are in particular relevant. The transformation of coordinates from one reference system to another and the transformation of a query or data instance from one application schema to another.

unique object identifier
identifier associated with a **spatial object** [DS-D2.5]

unit
defined quantity in which dimensioned parameters are expressed [ISO 19111]

universe of discourse
view of the real or hypothetical world that includes everything of interest [ISO 19101]

use case
A *use case* defines a goal-oriented set of interactions between actors and the system under consideration.

3.2 Abbreviations

AAA	AFIS – ALKIS – ATKIS
AFIS	Amtliches Festpunkt-Informationssystem)
AIS	Automatic Identification System
ALKIS	Amtliches Liegenschaftskataster-Informationssystem
ATKIS	Amtliches Topographisch-Kartographisches Informationssystem
ATOVS	Advanced TIROS Operational Vertical Sounder
BKG	Bundesamt für Kartographie und Geodäsie
BUFR	Binary Universal Form for the Representation of meteorological data
CORINE	Coordination of Information on the Environment
CRS	Coordinate Reference System

CT	INSPIRE Consolidation Team
DEM	Digital Elevation Model
DFDD	DGIWG (Digital Geospatial Information Working Group) Feature Data Dictionary
DGIWG	Digital Geospatial Information Working Group
DG	Directorate General
DS	Data Specification
DT DS	INSPIRE Drafting Team Data Specifications
DTM	Digital Terrain Model
EA	Enterprise Architect
EC	European Commission
EEA	European Environment Agency
EIONET	European Environment Information and Observation Network
ENV	Environment
ESDI	European Spatial Data Infrastructure
EGM	EuroGlobalMap
EPSG	European Petroleum Survey Group
ERM	EuroRegionalMap
ETRS89	European Terrestrial Reference System 89
EU	European Union
EUROSTAT	Statistical Office of the European Communities
EVRF2000	European Vertical Reference Frame 2000
ESDI	European Spatial Data Infrastructure
GDAL	Geospatial Data Abstraction Library
GDF	Geographic Data Format
GI	Geographic Information
GIMODIG	Geospatial info-mobility service by real-time data-integration and generalisation
GIS	Geographic Information System
GMES	Global Monitoring for Environment and Security
GML	Geography Markup Language
GRIB	Grid-point values expressed In Binary form
HO	Hydrographic Office
HTML	Hypertext Markup Language
ICS	International Commission of Stratigraphy
IGN France	Institut Géographique National, France
IGN Belgium	Institut Géographique National, Belgium
INSPIRE	INfrastructure for SPatial InfoRmation in Europe
IR	Implementing Rules
ISO	International Organization for Standardization
ISO/DIS	Draft International Standard
ISO/TR	ISO Technical Report
ISO/TS	ISO Technical Specification
IST	Information Society Technologies programme
ITS	Intelligent Transport Systems
IUGS	International Union of Geological Sciences
JRC	Joint Research Centre
LMO	Legally Mandated Organisation
LoD	Level of Detail
MOTIIVE	Marine Overlays on Topography for Annex II Valuation and Exploitation
MRDB	Multiple Representation Databases
NACSN	North American Code of Stratigraphic nomenclature
NADM	North American Data Model
NMA	National Mapping Agency
NMCA	National Mapping and Cadastral Agency
NUTS	Nomenclature des Unités Territoriales Statistiques
NVDB	National Road Database (in Sweden)
OCL	Object Constraint Language
OGC	Open Geospatial Consortium
ORM	Object-Relational Mapping or OpenGIS Reference Model
OS	Ordnance Survey
PDF	Adobe Portable Document Format

RADEF	Road Administration Data Exchange Format
RISE	Reference Information Specifications for Europe
RM-ODP	Reference Model – Open Distributed Processing (ISO)
SDIC	Spatial Data Interest Community
SDIGER	Spatial Data Infrastructure to support WFD information access for Adour-Garonne and Ebro River basins
TIN	Triangular Irregular Network
TIROS	Television Infrared Observation Satellite
TOID	Tophography Object Identifier
TWG	INSPIRE Thematic Working Group
UK	United Kingdom
UML	Unified Modelling Language
WCS	Web Coverage Service
WFD	Water Framework Directive
WFS	Web Feature Service
WGS84	World Geodetic System 84
WMO	World Meteorological Organisation
XML	eXtensible Markup Language
XSLT	eXtensible Stylesheet Language - Transofrmations

3.3 Verbal forms for the expression of provisions

In accordance with the ISO rules for drafting, the following verbal forms shall be interpreted in the given way:

- "shall" / "shall not": a requirement, mandatory for every data specification
- "should" / "should not": a recommendation, but an alternative approach may be chosen for a specific case if there are reasons to do so
- "may" / "need not": a permission

To make it easier to identify the mandatory requirements and the recommendations for INSPIRE data specifications in the text, they are highlighted and numbered.

Requirements are shown using this style.

NOTE This document does not spell out requirements.

Recommendations are shown using this style.

NOTE This document identifies a number of required support actions by the European Commission in order to establish the necessary operational infrastructure. These are not stated as requirements (although they technically are requirements, but not for INSPIRE data specifications) and are consequently not highlighted in the style shown above to avoid confusion.

3.4 References within the document

In accordance with the ISO rules for drafting, references to highest level of the document structure include the word "Clause" (or "Annex" in case of an annex).

EXAMPLE "Clause 2", "Annex A"

References to lower levels within the document structure are given without this qualifier.

EXAMPLE 7.1, 7.1.8.4, A.1

References to ISO standards are given without the full title.

EXAMPLE "ISO 19101" instead of "ISO 19101 – Geographic Information – Reference Model" or "ISO 19101 (Reference Model)"

4 Background and principles

4.1 Requirements as stated in the INSPIRE Directive

4.1.1 Articles of the Directive

4.1.1.1 General remarks

This sub-clause provides an overview of the articles in the Directive which are addressed by this proposal and describes how they are addressed. To make this sub-clause easier to read, the articles, in particular from Chapter III "Interoperability of spatial data sets and services", are repeated in the text in italics.

4.1.1.2 Article 4

This Directive does not require collection of new spatial data

Article 4(4) is addressed by the "as-is analysis" step in the data specification development.

4.1.1.3 Article 7(1)

Implementing rules laying down technical arrangements for the interoperability and, where practicable, harmonisation of spatial data sets and services, designed to amend nonessential elements of this Directive by supplementing it, shall be adopted in accordance with the regulatory procedure with scrutiny referred to in Article 22(3). Relevant user requirements, existing initiatives and international standards for the harmonisation of spatial data sets, as well as feasibility and cost-benefit considerations shall be taken into account in the development of the implementing rules. Where organisations established under international law have adopted relevant standards to ensure interoperability or harmonisation of spatial data sets and services, these standards shall be integrated, and the existing technical means shall be referred to, if appropriate, in the implementing rules mentioned in this paragraph.

Article 7(1) is mainly addressed by D2.5. The methodology described in this document particularly aims at defining interoperable and, where practical, harmonised specifications, for example, by a centralised feature concept dictionary or by taking all known and stated user requirements into account.

In addition, the European Commission and the SDICs / LMOs are expected to provide input on the requirements that have to be taken into account in the development process. The consolidation of the set of user requirements is a first activity in the data specification development process.

The use of relevant standards adopted by organisations established under international law, if appropriate, is addressed by the specification development process described in the work programme for the transposition phase of INSPIRE.

4.1.1.4 Article 7(2)

As a basis for developing the implementing rules provided for in paragraph 1, the Commission shall undertake analyses to ensure that the rules are feasible and proportionate in terms of their likely costs and benefits and shall share the results of such analyses with the committee referred to in Article 22(1). Member States shall, on request, provide the Commission with the information necessary to enable it to undertake such analyses.

Article 7(2) will be addressed by the Commission, but it should be added that

- the methodology requires that feasibility and benefits with respect to identified user requirements have to be taken into account in the development of INSPIRE data specifications,
- the Drafting Team has proposed to validate the proposal in parallel to the consultation process by testing candidate specifications in one or more real-world pilot projects,

- the review of documents is also an opportunity for stakeholders to raise any concerns about feasibility or proportionality of any decisions.

4.1.1.5 Article 7(3)

Member States shall ensure that all newly collected and extensively restructured spatial data sets and the corresponding spatial data services are available in conformity with the implementing rules referred to in paragraph 1 within two years of their adoption, and that other spatial data sets and services still in use are available in conformity with the implementing rules within seven years of their adoption. Spatial data sets shall be made available in conformity with the implementing rules either through the adaptation of existing spatial data sets or through the transformation services referred to point (d) of Article 11(1).

Article 7(3) is not addressed by this document. However, the methodology takes the requirement "either through the adaptation of existing spatial data sets or through the transformation services referred to point (d) of Article 11(1)" into account.

4.1.1.6 Article 7(4)

Implementing rules referred to in paragraph 1 shall cover the definition and classification of spatial objects relevant to spatial data sets related to the themes listed in Annex I, II or III and the way in which those spatial data are geo-referenced.

Article 7(4) is mainly addressed by D2.5. The methodology described in this document particularly aims at defining interoperable and, where practical, harmonised specifications, for example, by a centralised feature data dictionary or by taking all known and stated user requirements into account.

4.1.1.7 Article 7(5)

Representatives of Member States at national, regional and local level as well as other natural or legal persons with an interest in the spatial data concerned by virtue of their role in the infrastructure for spatial information, including users, producers, added value service providers or any coordinating body shall be given the opportunity to participate in preparatory discussions on the content of the implementing rules referred to in paragraph 1, prior to consideration by the Committee referred to in Article 22(1).

Article 7(5) is addressed by the drafting and consultation process involving SDICs and LMOs.

4.1.1.8 Article 8(1)

In the case of spatial data sets corresponding to one or more of the themes listed in Annex I or II, the implementing rules provided for in Article 7(1) shall meet the conditions laid down in paragraphs 2, 3 and 4 of this Article.

See comments in paragraphs 2, 3 and 4 below.

4.1.1.9 Article 8(2)

The implementing rules shall address the following aspects of spatial data:
(a) a common framework for the unique identification of spatial objects, to which identifiers under national systems can be mapped in order to ensure interoperability between them;
(b) the relationship between spatial objects;
(c) the key attributes and the corresponding multilingual thesauri commonly required for policies which may have an impact on the environment;
(d) information on the temporal dimension of the data;
(e) updates of the data.

According to the Directive, the aspects listed under Article 8(2) only apply to Annex I or II themes. The specifications of spatial object types in Annex III would be restricted to spatial properties only, such as spatial object type and geometry, including the reference system.

The INSPIRE Consolidation Team has confirmed that Thematic Working Groups developing specifications for Annex III themes should follow a broader goal (see below):

"The INSPIRE Consolidation Team recognising the importance of specifying / harmonising Annex III data of the INSPIRE Directive recommends the following approach to the Drafting Teams and the Thematic Working Groups: Specifications for Annex III themes can go beyond the objectives of Article 7(4) if the identified user needs require doing so. The European Commission will put forward such extended specifications to the INSPIRE Committee based on:

1. the content of relevant candidate specifications,
2. the outcome of the analysis of the user requirements to be performed by each Thematic Working Group, and
3. the comments received in frame of consultation with the Spatial Interest Communities and the Legally Mandated Organisation.

The extended specifications can contain any of the aspects listed in article 8(2). This measure recognises the fact that some Annex III themes cannot be treated in a meaningful manner without regarding the aspects listed in article 8(2)."

Article 8(2)(a-d) is not addressed by this document, but by D2.5 and the future INSPIRE data specifications.

Article 8(2)(e) is addressed by D2.7 in conjunction with the implementing rules on network services.

4.1.1.10 Article 8(3)

The implementing rules shall be designed to ensure consistency between items of information which refer to the same location or between items of information which refer to the same object represented at different scales.

The currently available technologies and practices in the Member States do not allow for automatically ensuring the spatial and thematic consistency between different but related spatial objects. The use of object referencing and the guidelines on data consistency and multiple representation as described in the INSPIRE Generic Conceptual Model and in this document can help with establishing and maintaining consistency between data, but in any case organisational measures will be required to provide consistent data as part of INSPIRE.

4.1.1.11 Article 8(4)

The implementing rules shall be designed to ensure that information derived from different spatial data sets is comparable as regards the aspects referred to in Article 7(4) and in paragraph 2 of this Article.

Article 8(4) is addressed by the data interoperablity components described in 4.2 and the uniform use of ISO 19131 (Data product specification) for the individual INSPIRE data specifications.

4.1.1.12 Article 9

The implementing rules provided for in Article 7(1) shall be adopted in accordance with the following timetable:
(a) no later than (two years following the date of entry into force of this Directive) in the case of the spatial data sets corresponding to the themes listed in Annex I;
(b) no later than (five years following the date of entry into force of this Directive) in the case of the spatial data sets corresponding to the themes listed in Annex II or III.

Article 9 is not addressed by this document, but by the INSPIRE work programme.

4.1.1.13 Article 10(1)

Member States shall ensure that any information, including data, codes and technical classifications, needed for compliance with the implementing rules provided for in Article 7(1) is made available to public authorities or third parties in accordance with conditions that do not restrict its use for that purpose.

Article 10(1) is not addressed by this document, but by the implementing rules on data sharing.

4.1.1.14 Article 10(2)

In order to ensure that spatial data relating to a geographical feature, the location of which spans the frontier between two or more Member States, are coherent, Member States shall, where appropriate, decide by mutual consent on the depiction and position of such common features.

Organisational measures will be required to provide data that is spatially referenced in a consistent way. The guidelines on data consistency and edge matching in this document can support establishing and maintaining the location of spatial objects that span frontiers between Member States.

4.1.2 Recitals in the Directive

Of the 35 recitals of the Directive, recital (6) is partially relevant for the technical specification of implementing rules on data specifications:

"The infrastructures for spatial information in the Member States should be designed to ensure

- *that spatial data are stored, made available and maintained at the most appropriate level;*
- *that it is possible to combine spatial data from different sources across the Community in a consistent way and share them between several users and applications;*
- *that it is possible for spatial data collected at one level of public authority to be shared between other public authorities […]."*

Also, recital (16) and (28) state that *"implementing rules should be based, where possible, on international standards […]"* and that *"in order to benefit from the state of the art and actual experience of information infrastructures, it is appropriate that the measures necessary for the implementation of this Directive should be supported by international standards and standards adopted by European standardisation bodies.*"

4.2 Data interoperability components

The work on INSPIRE data specifications is based on a framework that identifies the components relevant to the interoperability and harmonisation of data. These components are introduced and described in the Generic Conceptual Model, sub-clause 4.3.

The following figure provides an overview over the components relevant for data interoperability. The different components cover different aspects that need to be addressed in the process. For each of the components, a separate clause in document D2.5 specifies how this component is addressed in the Generic Conceptual Model.

Some components reference this document, D2.6. This methodology does not specify a full process for the development of fully harmonised cross-domain data specifications. It has to be understood, that for INSPIRE a collaborative approach between the groups involved in the data specification development is crucial. Additional research may be needed in the future to specify such a process, also taking the experiences in the development of INSPIRE data specifications into account.

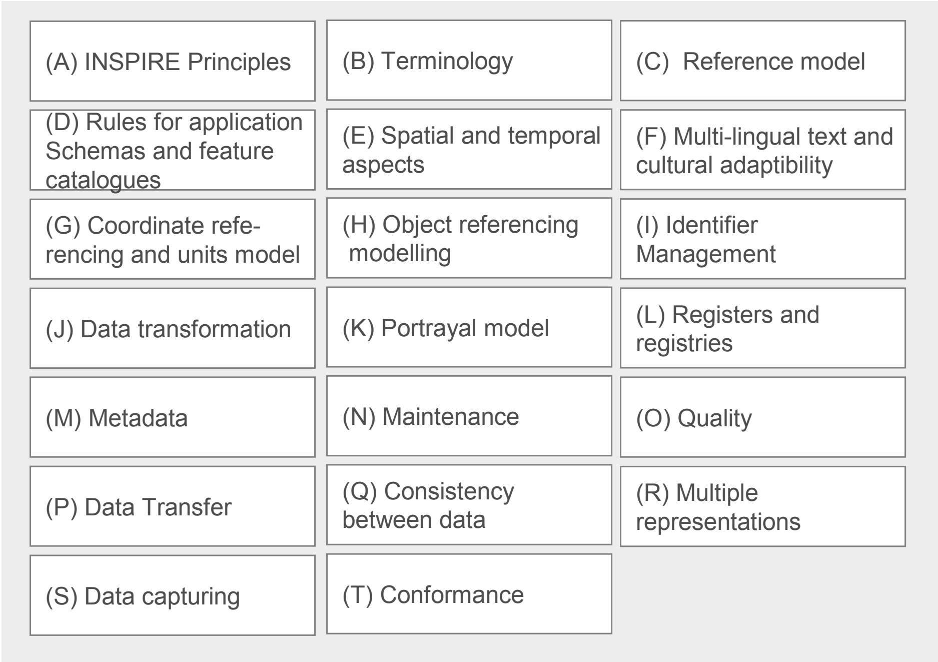


Figure 4 - Data interoperability components

5 General aspects

5.1 Principles

This methodology aims at a predictable and repeatable development process model. It is based on a iterative approach for incrementally growing INSPIRE's degree of definition and implementation based on user requirements in combination with a set of milestones to engage stakeholder commitment and bring about feasible and mutually satisfactory system solutions.

A process model answers two main questions: What should we do next and how long should we continue doing it? The answers to these questions vary depending on the context, the variation is driven by considerations about needs, risks and feasibility. This methodology emphasises the importance of having all of the success-critical stakeholders participate concurrently in defining and executing INSPIRE's processes.

In the INSPIRE data specification development, each cycle in the development process is defined by the steps listed in Clause 6.

In this, it is important to be aware of the following principles derived from the INSPIRE Directive:

- The series of INSPIRE data specifications will be structured according to the 34 spatial data themes defined in the annexes of the Directive.
- The development of data specification will start from environmental use cases – which in most cases will involve data from several themes.
- Since INSPIRE will be built on existing spatial data sets, the existing data in the Member States will be an important factor in the scoping of the INSPIRE data specifications in addition to the environmental use cases.

It is expected that INSPIRE data specifications for multiple themes will be developed in parallel to support cross-theme harmonisation of the data specifications. For example, the data specifications for all themes in Annex I of the INSPIRE Directive will be developed in parallel.

In the data specification development process, differences in starting points will need to be taken into account. For example:

- The Water Framework Directive provides relatively detailed data requirements while the Noise Directive is relatively unspecific about the contents of "noise maps".
- For some themes, international standards may already exist, in which case they form a basis for INSPIRE data specifications.
- In some themes, no data specification is available on a European level (only national specifications), in some cases there may already be a single European specification while yet in others there may be multiple (e.g. land cover classifications: CORINE and ISO 19144-2).

In a growing ESDI it is expected that increasingly existing data specifications for the different themes listed in the INSPIRE Annexes will emerge and be adopted in INSPIRE or amended to address new user requirements.

5.2 The development process in the INSPIRE Work Programme

The INSPIRE Work Programme for the Transposition Phase 2007-2009 outlines the general process, the organisational structure and the responsibilities for the development of data specifications for the Annex themes. It defines three scenarios which are characterised by their initial conditions:

Scenario 1 applies when a community has already agreed data specifications that have found widespread use. This community - as an SDIC or LMO - may submit these specifications as candidate draft INSPIRE implementing rules. The candidate specifications shall be verified against the recognised user requirements and checked for compliance with the Generic Conceptual Model. In case of minor adjustments these may be fixed by the proposing SDIC or LMO.

Scenario 2 applies when sufficient (raw) materials are at hand but no agreed data specifications are available, and designated user communities - as an SDIC or LMO - have expressed their interest in contributing to the work. In this case the CT will select SDICs and LMOs that delegate experts to a Thematic Working Group (TWG).

Scenario 3 applies when the material submitted to a theme is insufficient or completely missing. In this case a TWG will be formed from experts selected by the CT.

According to the INSPIRE Work Programme, the process comprises:

The analysis of reference material submitted by the SDICs and LMOs, paying special attention to the existing standards, the specifications established by organisations under international law, and the recognised user requirements.

Based on the results of the previous step, the TWG will develop a (candidate) data specification. This process can be skipped in scenario 1.

The candidate specification will be validated against the user requirements and checked for compliance with the Generic Conceptual Model.

The candidate specification will be published for review within the SDICs and LMOs, and will be tested. The cost-benefit aspects of the implementation will be assessed. The TWG will incorporate the results in a revision.

This process uses the step-wise approach detailed in Clause 6 and implements it in the organisational framework of INSPIRE.

5.3 Roles in a Thematic Working Group

The process of developing an INSPIRE data specification in a Thematic Working Group will typically involve the following areas of expertise:

- Domain expertise: expertise about the thematic domain and the data to be used in the application;
- GI expertise: expertise about geographic information specifications (ISO 19100 series, OGC standards) and information modelling; it is important to note that this includes the network services used to provide access to the spatial data sets as these are instrumental to harmonised data;
- INSPIRE expertise: expertise about the Generic Conceptual Model, data encoding guidelines, the INSPIRE architecture including the service architecture, the methodology to develop INSPIRE data specifications (this document) and other INSPIRE documents;
- Software expertise: expertise about implementation and deployment aspects of the relevant specifications

Depending on the complexity of a data theme, a person may cover several areas of expertise, or several persons may be required to cover a single area of expertise.

According to the "Terms of Reference for developing Implementing Rules laying down technical arrangements for interoperability and harmonisation of spatial data sets", this expertise will be provided by persons taking the following roles:

- Domain expert: provides expertise about the thematic domain and the data to be used in the application;
- Facilitator: manages the specification process and ensures that all data interoperability requirements (see the checklist) are identified and adequately addressed and the methodology is followed.
- Editor: is responsible for documenting the data specifications, the UML application schema and other elements in the data specification process.

The responsibilities of the various actors in the INSPIRE specification process are detailed in the "Terms of Reference for developing Implementing Rules laying down technical arrangements for interoperability and harmonisation of spatial data sets". This document relates the roles above with the organisations that participate in the process.

5.4 Maintaining the INSPIRE data specifications

This document specifies the process of developing INSPIRE data specifications. It has not yet been decided by the European Commission how INSPIRE data specifications will be turned into proposals for implementing rules and supporting guidance documents or standards.

INSPIRE is an information infrastructure and as such will – in the future – need clear and transparent maintenance mechanisms. Part of this is the maintenance of the INSPIRE data specifications once they have been established and adopted.

The European legal framework specifies the procedures for the adoption of new revisions of Implementing Rules, so in this subclause only proposed procedures and guidelines for the development of proposals for revisions are discussed.

In principle, the same process as described in this document can be applied as the methodology is established as an iterative process. Changes in requirements or as-is-situations may trigger a revision of the data specification applying the methodology including testing and validation.

However, a few aspects need to be pointed out:

- Once INSPIRE data specifications have been established, changes to them need to take backwards compatibility into account in order to avoid compromising the interoperability. Therefore, it is planned to prepare a document specifying patterns how INSPIRE data specifications may change.
- The current Work Programme and the Terms of Reference for the development of Data Specifications are limited to the first version of the Implementing Rules for interoperability of spatial data sets. It is recommended that the CT establishes a process involving stakeholders for any revisions of INSPIRE data specifications. The process should be kept as simple as possible and should adapt to different levels of changes, from a corrigendum to a significant revision.
- As the adoption of an Implementing Rule is a complex process, it is recommended that the Implementing Rule will consist of core requirements only and other aspects are specified outside of the adopted Implementing Rule so that they may be amended in a less complex procedure.
  - o A key aspect in this direction is the establishment of ISO 19135 conformant registers for the components of the data specifications as specified in the Generic Conceptual Model. The CT should consider a way to formulate the Implementing Rules that allows for changes of the registers without changing the Implementing Rules themselves.
  - o In order to support the goals of INSPIRE the implementation of the procedures for the maintenance of the registers should involve SDICs / LMOs at least to the same extent as they are involved in the development of the Implementing Rules.

  EXAMPLE It should be possible to add a new optional attribute to a spatial object type or a new value to a code list without requiring a change in an Implementing Rule.
- The maintenance of the registers specified in the Generic Conceptual Model needs to be organised according to ISO 19135. Table 1 below describes a possible distribution of the responsibilities for the INSPIRE Feature Concept Dictionary register according to the roles defined in ISO 19135. This information also needs to be detailed significantly to clearly specify, for example, the decision making process in the control body. Also, the responsibilities below are just a proposal. CT is preparing a document on the set-up and management of the registers during the data specification phase.

Table 1 – INSPIRE Feature Concept Dictionary register responsibilities

Roles	**Responsible party**
Register Owner	European Commission or INSPIRE Committee

Submitting Organisations	All Thematic Working Groups
Control Body	Group with the following members: - representative of DG JRC - representative of DG ENV - representative of DG EUROSTAT - representative of DT DS - representative of EIONET/EEA - representative of each TWG
Register Manager	representative of DG JRC
Registry Manager	representative of DG JRC

NOTE The future maintenance of the INSPIRE data specifications has not yet been discussed in detail within the Consolidation Team or the Drafting Team Data Specifications and is an open issue and will – to some extent – also depend on how the deliverables of the TWGs will be converted into Implementing Rules.

6 Steps in the development of INSPIRE data specifications

6.1 Overview

The steps in the development of an INSPIRE data specification for a theme are described in Table 2 below, including the responsible party. Steps starting with CT are the responsibility of the Consolidation Team, steps starting with TWG the responsibility of the Thematic Working Group associated with the theme. Steps starting with CT-TWG share the responsibility between CT and TWG. It should be noted however that the steps described here represent the theoretical methodology in an ideal situation. The current activities of the TWG's diverge in certain aspects from this ideal situation. This might have an impact on the proposed methodology that will be assessed during the framework document review based on the experiences of the Annex I TWGs

The remaining sub-clauses in this clause describe each step in more detail.

Table 2 – Steps in the development of INSPIRE data specifications - overview

Id	Step	Description
CT-TWG-1	Use case development	The use cases and application scenarios for the environmental policies to be supported by the INSPIRE data specifications in this development cycle are identified. The use cases are described in sufficient detail to clarify the requirements regarding the data from the spatial data themes. These use cases may be described according to the template for use case descriptions provided in Annex E.
TWG-2	Identification of user requirements and spatial object types	The theme-specific requirements regarding data are extracted from the use cases and application scenarios. This includes the identification of the required levels of detail, too. The result is a description of the relevant universe of discourse for the theme. A key result of this step is a candidate list of spatial object types with definitions and descriptions. These candidate concepts of spatial object types are shared and harmonised across the different themes – and TWGs – using the INSPIRE Feature Concept Dictionary Register (see Generic Conceptual Model 9.3). It is recommended to capture the understanding of the scope of the theme in this step as a "first cut" data specification comprising the spatial object types and their main properties and dependencies as well as other important information about the theme.
TWG-3	As-is analysis	An as-is analysis of the current situation regarding spatial data sets for the theme is carried out in parallel to TWG-2. A checklist (see Annex F) is provided to assist in identifying the relevant data interoperability aspects. The result provides the basis for the next step, the gap analysis.
TWG-4	Gap analysis	The gap analysis identifies user requirements that cannot be met by the current data offerings. For each gap, a data interoperability approach – which may also include a conclusion that specific user requirements cannot be met – will be identified and agreed upon.

TWG-5	Data specification development	This approach is then documented as one or more application schemas for the theme specifying the spatial object types with their properties, range of valid property values and constraints. See the Generic Conceptual Model Clause 9 for details on INSPIRE application schemas. The data specification itself will be documented according to ISO 19131, the International Standard specifying the contents of data product specifications in the field of geographic information. See the Generic Conceptual Model 8.3. The application schema, accompanied by a corresponding feature catalogue derived from the application schema, constitutes the core component of the data specification.
CT-6	Implementation, test and validation	These results will be reviewed by the stakeholders and tested within one or more pilots under real world conditions using the use cases developed in CT-TWG-1 to test the proposed specification for consistency, completeness, feasibility and implementability.
CT-TWG-7	Cost-benefit considerations	Incremental costs and benefits of the data interoperability and harmonisation efforts will be tracked and documented.

The results of every step may require an update in the results of the previous steps. This is discussed in more detail in each step.

This approach is intended to enable the identification of issues, relevant for data interoperability, as early as possible in the process. In particular, it is the goal to identify as many issues as possible before implementation. As data interoperability is a complex issue, it is not realistic to expect that every issue is caught early in the process, so implementing and testing is an integral part of the process, too.

In general, the steps are not carried out sequentially, but with a considerable overlap to allow for rapid feedback. The following diagram (Figure 5) illustrates roughly the flow of the development process. Note that feedback is intended to be propagated back to previous steps in every stage, where appropriate. The “feedback loops” include the formal review processes foreseen by the INSPIRE Work Programme (review by CT/DT DS/EIONET as well as the SDIC/LMO consultation), too.

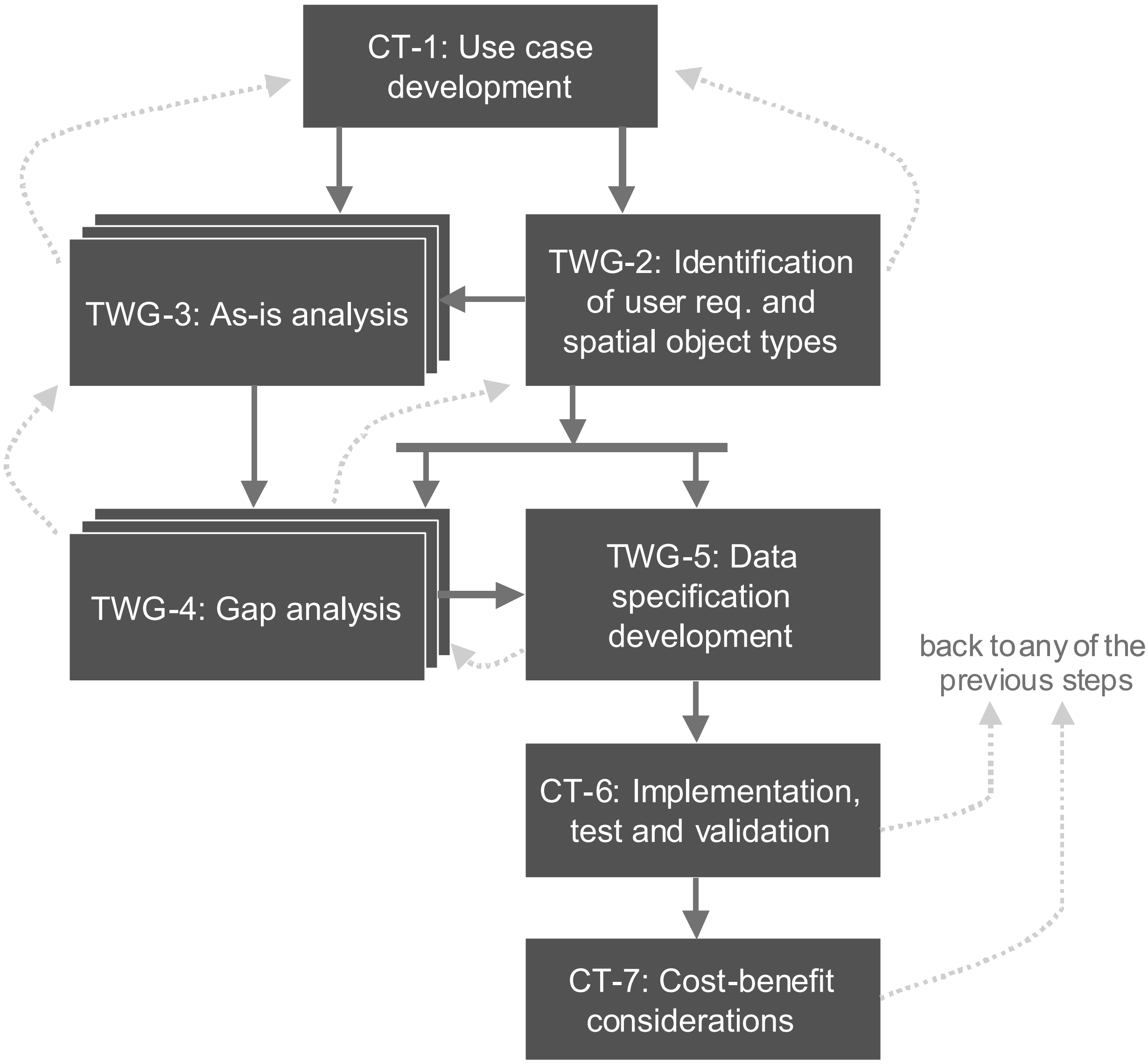


Figure 5 – The process of developing INSPIRE data specifications

The approach taken by this methodology is to offer "tools" that are intended to be adapted as appropriate in different contexts. For example, there is no specific mandatory process for how to create and analyse use cases, and turn them into an application schema for the data interoperability process. This recognises the fact that in practice different approaches are being used successfully and it does not seem appropriate to prescribe a particular approach and disallow others.

The tools are:

- a use case template (Annex E) that contains information typically provided in use case descriptions;
- a checklist for data interoperability aspects (Annex F), i.e. aspects that need to be understood in the context of the use case to address the interoperability issues of the use case; this checklist is based on the data interoperability components discussed in the clause 4.2;
- a Generic Conceptual Model (document D2.5), an INSPIRE Feature Concept Dictionary Register, an INSPIRE Glossary common to all INSPIRE data specifications and the integration of all application schemas into a consolidated UML model; see the Generic Conceptual Model for more details;
- The “INSPIRE data specifications Cost Benefit Considerations” document (to be published on http://www.ec-gis.org/inspire/ds/)

Maintaining a INSPIRE Feature Concept Dictionary Register and a consolidated UML model is intended to help in keeping a harmonised “view” to all of the information available in INSPIRE.

6.2 Results

Table 3 provides an overview over the results of the steps introduced above that lead to an INSPIRE data specification. The individual steps are explained in more detail in the following sub-clauses.

Table 3 – Steps in the development of INSPIRE data specifications - results

Id	Step	Results of the step	Reference material and clauses of this document
CT-TWG-1	Use case development	- Use case description - Updated glossary	- 6.3 - UML 2.1.2 Superstructure, Clause 16 - INSPIRE Glossary
TWG-2	Identification of user requirements and spatial object types	- List of spatial object types as proposals for entries in the INSPIRE Feature Concept Dictionary Register - “First cut” data specification - List of requirements (structured according to the data interoperability components described in 4.2)	- 6.4 - D2.3 Description and scope of themes - Generic Conceptual Model - ISO 19131 - INSPIRE Feature Concept Dictionary Register - INSPIRE data specification template - Consolidated INSPIRE UML Model
TWG-3	As-is analysis	- Description of the current situation with respect to the data interoperability components (in principle, for every source data set)	- 6.5 - Annex F
TWG-4	Gap analysis	- Description of data interoperability issues derived from the identified user requirements and taking the as-is analysis into account (in principle, for every source data set) - Choice of harmonisation approach - Updated as-is analyses (additional data sources that have been identified) or updated/reduced user requirements to reduce the identified gaps	- 6.6 - Annex A - Annex F

TWG-5	Data specification development	- Data specification (per spatial data theme) with clauses specified in ISO 19131 (including application schema in UML as well as the corresponding feature catalogue and GML application schema) - Updated Consolidated INSPIRE UML model - Updated INSPIRE Feature Concept Dictionary Register - Updated glossary	- 6.7 - Annex A - Generic Conceptual Model - ISO/TS 19103 - ISO 19109 - ISO 19110 - ISO 19126 - ISO 19131 - INSPIRE data specification template - Consolidated INSPIRE UML Model - INSPIRE Feature Concept Dictionary Register - INSPIRE Glossary - ISO 19136 (GML) and other data encodings
CT-6	Implementation, test and validation	- Implementation of the application using services and the data specification - Test report	- 6.8 - INSPIRE download service - Spatial data services
CT-TWG-7	Cost-benefit considerations	- Cost-benefit considerations	- 6.9 - INSPIRE data specifications Cost Benefit Considerations"

NOTE The analysis of the as-is situation typically should be carried-out separately per source data set. One particular difficulty in this step will often be that the existing data will not be properly documented and thus some reverse-engineering of the specification corresponding to the data or some other activities to compile a documentation of the data may be required.

6.3 Step CT-TWG-1: Use case development

6.3.1 Responsible party

- Consolidation Team/Thematic Working Groups

6.3.2 Step results

- Use case descriptions
- Change proposals for the INSPIRE Glossary[7]

6.3.3 Exit conditions

- CT internal review considers the available use case descriptions complete and mature for data specification development → TWG-2, TWG-3

6.3.4 Resources

- Use case template (Annex E)
- INSPIRE Glossary

6.3.5 Additional information

- UML 2.1.2 Superstructure, Clause 16

[7] The terms in 3.1 are part of the glossary. However, beside these general terms, the glossary will eventually include also theme specific terms and concepts that are not defined in the INSPIRE Feature Concept Dictionary Register or in code lists.

6.3.6 Detailed description

A *use case* defines a goal-oriented set of interactions between actors and the system under consideration. An *actor* specifies a role played by a user or any other system that interacts with the system under consideration (UML 2.1.2 Superstructure, Clause 16).

Use cases are used in this methodology because they are an accepted and commonly used mechanism to identify and document user requirements. As a result, use case descriptions comprised of use case diagrams and textual descriptions (a template is provided in Annex E) will be created to support the data specification development. They document the user requirements against which the specifications need to be built.

A use case is initiated by a user with a particular goal in mind, and completes successfully when that goal is satisfied. It describes the sequence of interactions between actors and the system necessary to deliver the service that satisfies the goal.

Generally, use case steps are written in an easy-to-understand structured narrative using the vocabulary of the domain. This is engaging for users who can easily follow and validate the use cases, and the accessibility encourages users to be actively involved in defining the user requirements.

The use case describes the user requirements, in this methodology the focus is on user requirements relevant to an INSPIRE data specification.

The identification and provision of the relevant use cases and application scenarios for the INSPIRE data specifications is within the responsibility of the Consolidation Team, supported by SDICs and LMOs.

The major sources for the relevant use cases and application scenarios are:

- the community environmental policies, or policies which may have an impact on the environment,
- the reference material submitted by the SDICs and LMOs,
- the results of user requirements survey of the Consolidation Team,
- the DT DS deliverable D2.3 “Definition of Annex Themes and Scope” together with the comments on it from the SDIC/LMO review,
- studies by JRC and EUROSTAT,
- EU-funded initiatives and projects.

New terms used in the use cases shall be added to the INSPIRE Glossary. Where changes to existing terms and definitions are considered important, they may be proposed.

A use case methodology should be applied. The use case template along with guidelines on its use may help in this step, but their application is optional as long as a result of the step the user requirements can be identified.

Unlike the other steps, this step should ideally not be executed per theme. As the development of data specifications will start from environmental use cases which typically contain data from several Annex themes, the use cases need to be known first. The respective set of use cases must be clearly identified, before the development of the first batch of INSPIRE data specifications can start. Otherwise, the development will not be coherent across the different themes. However, use cases can also be identified and elaborated by the thematic experts present in the TWG’s, taking also within-theme cases into account.

6.4 Step TWG-2: Identification of user requirements and spatial object types

6.4.1 Responsible party

- Thematic Working Group

6.4.2 Step results

- Change proposals for the INSPIRE Feature Concept Dictionary Register
- "First cut" of the INSPIRE data specification
- List of requirements

6.4.3 Exit conditions

- TWG internal review considers the step results sufficiently complete and mature for the gap analysis → TWG-4 and TWG-5
- CT review (with advice from EIONET and DT DS) considers documentation of user requirements complete → TWG-4 and TWG-5
  - o any issues raised during the review will be resolved by TWG and CT
- TWG considers the use cases descriptions insufficient → CT-TWG-1
  - o if no improvement of the use case descriptions is feasible, the as-is situation will be used to cover any gaps in the use case descriptions

6.4.4 Resources

- Use case descriptions (CT-TWG-1)
- Checklist (Annex F)
- D2.3 Description and scope of themes
- INSPIRE Feature Concept Dictionary Register

6.4.5 INSPIRE data specification template

- Consolidated INSPIRE UML Model

6.4.6 Additional information

- Generic Conceptual Model
- ISO 19131

6.4.7 Detailed description

The results from the use case development constitute the user requirements for the INSPIRE data specification to be developed.

The TWG shall consolidate the theme-specific user requirements documented in the use cases. They may reject user requirements with justification, and they may propose additional user requirements to the CT.

Taking into account Articles 8(2) and 8(4) of the INSPIRE Directive and the requirements spelt out in the Generic Conceptual Model, the TWG should in particular identify user requirements on

- the data content (e.g. spatial object types, attributes),
- the level of detail,
- relationships between spatial objects,
- data consistency,
- updating and the temporal dimension of the data,
- unique identifiers,
- metadata for evaluation,
- data quality.

NOTE No separate identification of requirements of metadata for use is listed as the data product specifications themselves comprise the metadata for use (Ref Implementing Rules on Metadata v2.0, Clause 8).

In the process, the TWG shall analyse the comments from SDICs and LMOs on D2.3 v2.0 which are specific to their theme(s) and which have not yet been resolved. Every comment submitted by an SDIC/LMO shall be duly addressed by the TWG (in this or in the later steps).

The checklist along with guidelines on its use (see Annex F) may help in identifying the specific requirements, but their application is optional as long as a result of the step the user requirements are identified and all data interoperability components have been considered. The goal of the checklist is to assist during a discussion between domain experts and the editor. The documentation of the result of this discussion is a documentation of the user requirements.

Where appropriate, the use case descriptions may be amended to include the additional details identified during this step.

Based upon the user requirements and the checklist, the editor will have a fairly good understanding of the spatial object types and associated attributes, constraints and association as well as other relevant information like coordinate reference systems, metadata, etc. involved, and this information will be the basis for the further analysis.

As a result, the user requirements that relate to data shall be transformed by the editor to a first-cut of a data specification for the spatial data theme – including one or more first-cut application schemas for the theme to facilitate the discussion in the subsequent steps.

If a potential candidate data specification for the relevant data theme already exists, an initial analysis shall be carried out to determine, if the existing specification should be used as basis and be amended or if the user requirements warrant the creation of a new data specification. Whenever possible, the reuse of an existing data specification should be considered, in particular, if an internationally standardised data specification exists.

The analysis of the user requirements should result in a documentation of the user requirements (e.g. checklist, amended use case descriptions). The TWG will deliver the documentation to the CT for review, with advice from EIONET and DT DS. The validated use-cases will be then used for the specification development process.

Recognising the fact that different approaches are being used successfully in practice, no specific formal process model regarding the conversion of the use cases into spatial object types (i.e. identifying classes in terms of object-oriented analysis and design) or an application schema is mandatory for the data interoperability process.

An example to aid the identification of the relevant classes/spatial object types for the first-cut application schemas is to consider the “nouns” representing real-world phenomena in user requirements and use case description as candidates for spatial object types. After eliminating duplicates and synonyms as well as nouns that describe concepts that are not spatial object types, the remaining list of nouns can be used as an initial list of spatial object types. By analysing the user requirements and the use case descriptions in more detail, the properties (i.e. attributes and relationships with other types) as well as the constraints will be specified for the individual spatial object types.

6.5 Step TWG-3: As-is analysis

6.5.1 Responsible party

- Thematic Working Group

6.5.2 Step results

- Description of the current situation with respect to the data interoperability components (in principle, for every source data set)

6.5.3 Exit conditions

- TWG internal review considers the step results sufficiently complete and mature for the gap analysis → TWG-4

6.5.4 Resources

- Use case descriptions (CT-TWG-1)
- Requirements (TWG-2)
- Checklist (Annex F)

6.5.5 Additional information

n/a

6.5.6 Detailed description

In parallel to the identification of the user requirements and the documentation of a “first-cut” data specification, an as-is analysis is executed. This step analyses the information in the reference material submitted for a particular data theme and existing data interoperability and standardisation initiatives. It is expected that in addition to spatial data sets with a geometry per spatial object, there will be data in various kinds of formats, from PDF documents to databases, that is relevant, too, and which is spatially referenced by geographic identifiers (e.g., place names, river identifiers, etc.). Also, data for a single data theme will in general be spread over multiple data sets and multiple data specifications in each Member State.

As in step TWG-2, the checklist along with guidelines on its use (see Annex F) may help in the analysis, but its application is optional as long as the analysis considers all data interoperability components. The goal of the checklist is to assist during a discussion between the domain experts and the editor. Rows of the checklist that are not applicable for a theme do not have to be addressed and may be removed. The documentation of the result of this discussion is a documentation of the current situation.

In most situations, the existing source material is an important part of the use case in general, when a user defining his/her requirements is well aware of existing data. In other situations, the existing source material is different from country to country and requires a search for potential input data based on the descriptions in the use case scenarios.

In general, it is not required to apply the “tools” and standards that are used to describe the harmonised data specifications to the management of existing data sets. If the available documentation of the existing data includes information about, for example, the spatial object types and their properties there is no need to recapture this information in a data specification according to ISO 19131.

However, sometimes the input data will not be documented sufficiently. This methodology does not provide specific rules for how to deal with this fact, but if documentation has to be created for the data set, it is recommended to use the same technologies for the input data specification as used for the harmonised data specification. It is important to note that such documentation is not an end result of the data specification process as such, but would only be an intermediate result to support the gap analysis.

Also, complete English versions of the national/regional feature catalogues are in general not available (sometimes feature and attribute names are translated, but not the descriptive text). Therefore, in principle, local experts from all Member States should participate in the data specification process. The GIMODIG project[8], for example, recommends that all parts of the reference feature catalogues / product specifications that are relevant for the harmonised data product specification should be translated into English. This would require the involvement of a large number of high-profile experts (the expert must know both the language and the domain). In practice, TWGs will consist of about five to eight persons, i.e. previous work and experience of European-wide SDICs needs to be leveraged where possible to ensure that the process can be managed. Support from the CT (translation of documents, inviting experts that can consult on specific topics, organising surveys, etc.) and the SDIC/LMO submitting reference material will be required where TWGs alone will not be able to represent the situation in all Member States.

6.6 Step TWG-4: Gap analysis

6.6.1 Responsible party

- Thematic Working Group

6.6.2 Step results

- Description of data interoperability issues derived from the identified user requirements and taking the as-is analysis into account (in principle, for every source data set)
- Choice of harmonisation approach
- Analysis of reference materials

6.6.3 Exit conditions

- TWG identifies additional source data sets / specifications → TWG-3
- TWG identifies gaps where bridging is not considered feasible under cost-benefit considerations by the TWG → TWG-2 or CT-TWG-1 depending on the significance of the gap
- TWG internal review considers the step results sufficiently complete and mature for the data specification development → TWG-5

6.6.4 Resources

- Use case descriptions (CT-TWG-1)
- Requirements (TWG-2)
- As-is analysis (TWG-3)
- Checklist (Annex F)

6.6.5 Additional information

- Examples of harmonisation approaches (Annexes A and B)

6.6.6 Detailed description

This analysis compares the results of each as-is analysis with the first-cut data specification, and evaluates, if the identified source material is sufficient to fulfil the data specification. It also identifies how to extract information from these data sources into the application schema.

Existing data will in almost all cases be structured in a way that differs from an INSPIRE data specification. Since the data in a spatial data set will often be collected for a local, regional or national

8 The project "Geospatial Info-Mobility Service by Real-Time Data-Integration and Generalisation" (GIMODIG) was funded from the European Union via the Information Society Technologies (IST) programme. The objective of the project was to develop methods for delivering geospatial data to a mobile user by means of real-time data-integration and generalisation. The project aimed at a seamless data service infrastructure providing access, through a common interface, to topographic geo-databases maintained by the national mapping agencies. See http://gimodig.fgi.fi/.

task, a transformation of data to an INSPIRE data specification will often involve transformations for several reasons:

- due to different encodings
- due to different levels of detail as the view required by an INSPIRE user may be substantially different from the view of the existing applications
- due to different coordinate reference systems
- due to different terminologies and concepts
- due to different languages
- due to inconsistencies with other INSPIRE data (for example along or across national borders, themes, sectors or at different resolutions)

As in steps TWG-2 and TWG-3, the checklist along with guidelines on its use (see Annex F) may help in the analysis, but its application is optional as long as the analysis considers all data interoperability components. The goal of the checklist is to assist during a discussion between the domain experts and the editor. The documentation of the result of this discussion is a documentation of the identified gaps.

Most of the gaps will be solved by extracting information from the data sources by transformation services in accordance with the INSPIRE principles. I.e., the INSPIRE data specification should be designed to ensure easy transformation between existing data and the data specification. Annexes A and B deal with harmonisation approaches and gives some ways to facilitate this transformation.

Nevertheless, some gaps will need greater efforts than transformation services. If the identified source material is not sufficient, a new as-is analysis or a change of the user requirements may be required.

Alternatively and if there is consensus among all stakeholders, the provision of new data may be considered, too. Provision of new data may be the capture of a new data set, but will more typically mean amending existing data (e.g. with new codes or classifications) or processing existing data (e.g. to derive the centre line of rivers or other generalisation algorithms). In this case, some cost considerations have to be taken into account to select the most appropriate solution.

In general, the gaps and decisions about harmonisation approaches have to be carefully documented and referenced; in case of competing solutions the choice should be reasoned in technical and cost-benefit terms. In this process, the TWG has to consider the following aspects:

- The recitals and articles from the Directive, in particular:
  - Recital (16): "Implementing rules ….. should not result in excessive costs for Member States"
  - Article 4(2): "This Directive does not require collection of new spatial data"
  - Article 7(3): "Spatial data sets shall be made available in conformity with the implementing rules either through the adaptation of existing spatial data sets or through the transformation services"
- Implementation aspects: Implementation of the proposed specifications shall be feasible within the timeline spelled out by the Directive, i.e. two years after the adoption of the Implementing Rules for all newly collected and extensively restructured spatial data sets, and seven years for other data sets.

The TWG will document the results of the as-is analysis and the gap analysis in the deliverable ‘Analysis of reference materials’, which according to the Terms of Reference is to be provided to the CT.

6.7 Step TWG-5: Data specification development

6.7.1 Responsible party

- Thematic Working Group

6.7.2 Step results

- Candidate INSPIRE data specification
- Change proposal for the Consolidated INSPIRE UML Model
- Change proposals for the INSPIRE Feature Concept Dictionary Register
- Change proposals for the INSPIRE Glossary

6.7.3 Exit conditions

- TWG identifies new gaps → TWG-4
- TWG internal review considers the step results sufficiently complete and mature for review and testing → CT-6

6.7.4 Resources

- Requirements (TWG-2)
- As-is analysis (TWG-3)
- Gap analysis (TWG-4)
- Harmonisation approaches (TWG-4)
- INSPIRE data specification template
- First-cut data specification (TWG-2)
- Consolidated INSPIRE UML Model
- INSPIRE Feature Concept Dictionary Register
- INSPIRE Glossary

6.7.5 Additional information

- Generic Conceptual Model
- Examples of harmonisation approaches (Annexes A and B)
- ISO/TS 19103
- ISO 19109
- ISO 19110
- ISO 19126
- ISO 19131
- ISO 19136 (GML) and other data encodings

6.7.6 Detailed description

The result of the checklist in Annex F (if used) is an important starting point for creating an INSPIRE data specification. But it is also clear that the result of the checklist itself is not detailed enough and that further work must be conducted in close cooperation between one of more domain experts and an editor. However, the result of the checklist is intended assist in the discussions, and ease the creation of the INSPIRE data specification.

Figure 7 shows the sequence from a data product specification that specifies a data product to its implementation as a data set described by metadata.

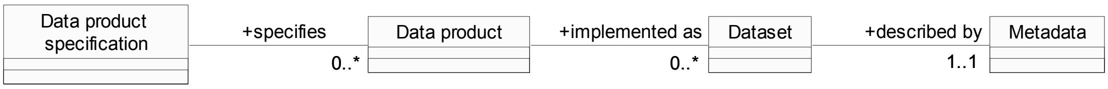


Figure 6 – Data product specification and its relation to data sets (ISO 19131, Figure B.1)

Based on the harmonisation approaches identified in step TWG-4, the first-cut data specification developed in step TWG-2 will be improved in a way that the information in the data specification can be extracted from sources at the Member States either through the adaptation of existing spatial data sets or through transformation services.

A core component of every INSPIRE data specification is the application schema. An application schema is a conceptual schema for data required by one or more applications. This application schema describes the conceptual model for the data that is supposed to fulfil the identified user requirements.

The development of the application schema is discussed in more detail in the Generic Conceptual Model clause 9.

The specification of the valid encoding(s) of spatial data for the theme is an integral part of each INSPIRE data specification. This topic is not addressed in this document in more detail, but will be addressed in the guidelines for the encoding of data (document D2.7).

Recommendations regarding the content of the data specifications are spelt out in Clause 7, with reference to the Generic Conceptual Model and the ISO 19100 series of standards.

Feasibility is a major concern of the INSPIRE Directive. Whatever candidate INSPIRE data specification a TWG proposes, the requirements in the specification have to be justified by user requirements on one hand and have to be feasible within the measures of the Directive on the other hand.
Before submitting the candidate INSPIRE data specification to the CT, the TWG has to verify conformance with the Generic Conceptual Model.

The description of the specification process and the decisions taken during this process will be published together with the candidate INSPIRE data specification to make the development process transparent to stakeholders.

6.8 Step CT-6: Implementation, test and validation

6.8.1 Responsible party

- Consolidation Team

6.8.2 Step results

This depends on the version of the Candidate INSPIRE data specification.

Version 1.x:
- Review comments

Version 2.x:
- Review comments
- Pilot implementation(s)
- Test report
- Candidate INSPIRE data specification v3.0

6.8.3 Exit conditions

This depends on the version of the Candidate INSPIRE data specification.

Version 1.x:

- CT, EIONET, DT DS identify issues during review → CT-TWG-1, TWG-2, TWG-3, TWG-4, or TWG-5 depending on the issue identified
- Otherwise → Version 2.0

Version 2.0:

- SDIC/LMO review identifies issues during review → CT-TWG-1, TWG-2, TWG-3, TWG-4, or TWG-5 depending on the issue identified

Version 2.x (including x=0):

- Pilot implementations identify issues during review → CT-TWG-1, TWG-2, TWG-3, TWG-4, or TWG-5 depending on the issue identified
- SDIC/LMO review and pilot testing does not raise any critical issues → CT-TWG-7

6.8.4 Resources

- Candidate INSPIRE data specification (TWG-5)
- Analysis of reference materials (TWG-4)
- Consolidated INSPIRE UML Model
- INSPIRE Feature Concept Dictionary Register
- INSPIRE Glossary

6.8.5 Additional information

- Generic Conceptual Model
- Implementing rule on download services

6.8.6 Detailed description

According to the INSPIRE Work Programme, draft INSPIRE data specifications created by a TWG will be evaluated first by the CT, the DT DS and the EIONET. After addressing all comments raised during this evaluation, the specifications will be reviewed by SDICs and LMOs.

NOTE The review process and the tasks of the actors are detailed in the Terms of Reference

In addition to the document review, every candidate INSPIRE data specification developed according to the description in the previous sub-clauses will be tested in a pilot under real world conditions.

In the test, a representative part of the relevant use cases should be implemented in conformance with the INSPIRE data specification. The use cases shall then be executed with reasonable variations and real-world data.

The pilot implementation of a use case shall comprise:

- a representative set of Member States providing spatial data according to the candidate INSPIRE data specification via download services for one or more themes depending on the use case;
- an application implementing the process described in the use case and accessing all spatial data from the download services.

In a successful test

- the spatial data from the Member States will all conform to the candidate INSPIRE data specifications;
- the application will access the data from the various download services;
- the application will be able to use the data from the various Member States without repetitive manual intervention (schema transformation), in such a way that the result is coherent;
- the application will be able to perform all necessary actions to execute the use case.

The candidate INSPIRE data specifications are revised by the responsible TWG based on the results of the tests.

The final INSPIRE data specifications are then brought forward to the relevant groups in INSPIRE for appropriate consideration, feedback and – eventually – adoption.

6.9 Step CT-TWG-7: Cost-benefit considerations

6.9.1 Responsible party

- Consolidation Team/Thematic Working Groups

6.9.2 Step results

- Cost-benefit considerations reported

6.9.3 Exit conditions

- n/a

6.9.4 Resources

- Candidate INSPIRE data specification (TWG-6)
- INSPIRE data specifications Cost Benefit Considerations"
- ToR Testing

6.9.5 Additional information

- n/a

6.9.6 Detailed description

A detailed description is being elaborated in the "INSPIRE data specifications Cost Benefit Considerations"" document, to be published on http://www.ec-gis.org/inspire/ds/.

7 Recommendations for INSPIRE data specifications

7.1 General

This clause lists recommendations for the content of INSPIRE data specifications. It is structured according to the structure of statements in a data product specification as given by ISO 19131.

Recommendation 1
Each Annex theme should be described by one INSPIRE data specification.

Current practise may be different, as many data producers create individual specifications for data sets with similar content but different level of detail. Taking into account that INSPIRE aims at interoperability, it is strongly recommended to establish harmonised application schemas independent from scale and resolution, where feasible.

Recommendation 2
For INSPIRE data specifications, it is recommended that only one general “specification scope” according to ISO 19131 is applied (the default scope). If needed, specific specification scopes may be identified in the textual descriptions, so that individual specification scopes according to ISO 19131 can be derived and transposed in a formal notation if required.

In case an INSPIRE data specification has different requirements on different parts of a theme, it may wish to distinguish them. ISO 19131 introduces the concept of “specification scopes” to divide a data product specification into partitions. Specification scopes can be defined by a list of feature types, by spatial extent, by temporal extent, by a set of coverages or other reference. While referring to a specification scope the parts of the data specification can either inherit or override the general case, e.g. for the definition of data quality. This concept of different scopes within a single data specification helps to maintain the consistency between the partitions, but it makes the specifications complex and difficult to read and handle.

7.2 Identification information

This statement contains general information such as the title of the specification, a list of the main theme(s) and the geographic extent of the data.

The identification information allows for the description of spatial resolution. Resolution is an important criterion for the usefulness of data. Due to the heterogeneity of data sources in Europe it may be difficult in an INSPIRE data specification to determine resolution in exact figures. The Drafting Team Data Specifications considered the approach from the position papers with four general classes ‘European level’, ‘National level‘, ‘Regional level’ and ‘Local level’ (INSPIRE Position paper Reference Data and Metadata Position paper v4.2). As it turned out that these classes may be used incoherently, the Drafting Team decided to not prescribe a common system for the classification of resolution, but leave it to the TWG to identify an appropriate system for their domain.

Recommendation 3
In cases where the TWG discovers that they cannot determine resolution in exact figures due to the heterogeneity of use cases and/or national data sources, they should identify the relevant level(s) of detail, link it with a measure of scale/resolution and provide that information with the specification.

NOTE: Examples are provided in annex A.18.

An overview matrix for the level of details with their scale/resolution per theme may be created for information purposes.

7.3 Content and structure

This statement contains the application schema requirements according to the Generic Conceptual Model.

NOTE Provision of application schemas is mandatory for all INSPIRE data specifications, including those specifications that describe coverage data and/or measurements and observations. An example for an application schema describing coverage data is provided in D.3 (Elevation). An example for an application schema describing observations is provided in D.4 (Meteorology) as well as D.5 (Geology).

Recommendation 4
To underpin the requirements on consistency between spatial data sets – including edge matching along national boundaries – it is recommended that examples are given as part of the INSPIRE data specification.

7.4 Reference system

This statement contains the reference system requirements for data of the theme according to the Generic Conceptual Model, Clause 12.

It is quite obvious that applications may have requirements regarding the reference systems that may, should or shall be applied. These will often differ from the national or local systems that are normally applied in a spatial data set. Most of these requirements may be fulfilled by a coordinate transformation service (a service offering coordinate conversions between two coordinate reference systems according to ISO 19111).

Recommendation 5
In specifying one or more coordinate reference systems to be used for data of the theme, the impact of coordinate transformation on other aspects of the data specification such as positional accuracy and metadata should be taken into account.

Another aspect of reference systems that has to be considered are temporal reference systems. The widespread application of computers and geographic information systems has led to the increased analysis of spatial data within multiple disciplines. Geographic information is not confined to a three-dimensional spatial domain. Many geographic information systems require data with temporal characteristics.

7.5 Data quality

ISO 19131 requires a data specification to cover the data quality elements and data quality sub-elements defined in ISO 19113. Those quality elements are:

- Completeness
- Logical Consistency
- Positional Accuracy
- Temporal Accuracy
- Thematic Accuracy

Apart from logical consistency (see the Generic Conceptual Model clause 20), the INSPIRE Directive does not spell out requirements for data quality. However, it is obvious that users require information on the quality of data sets to assess whether the data sets are useful for them or not. For that purpose, the standard ISO 19113 identifies quality elements and sub-elements. ISO 19131 requires all quality elements to be addressed; even if only to state that a specific data quality element or quality sub-element is not applicable. The Technical Specification ISO/TS 19138 defines a set of data quality measures that can be used when reporting data quality for the sub-elements in ISO 19113.

Recommendation 6
An INSPIRE data specification should not prescribe minimum data quality requirements. However, It may recommend minimum data quality requirements when justified by the user requirements. In this case the specification should introduce conformity levels to be reported with the metadata

Minimum quality requirements may restrict data from being published or may force data providers to capture additional data. This would contradict with the INSPIRE principles. The intent is to make existing data available in INSPIRE in a way so that users can assess whether the information is useful for them or not. For that reason it is recommended instead to report the data quality with the metadata.

7.6 Data capture information (optional)

Where this statement is included, it contains a data capture statement, i.e. a general description of the sources, processes and selection criteria to be used for the various spatial object types.

This statement, if provided, is simply a textual description and does not provide for / require any substructure in the information to be provided in the data specification.

Clause 24 of the Generic Conceptual Model spells out requirements for the data capturing rules in INSPIRE data specifications.

7.7 Maintenance information (optional)

Where this statement is included, it describes the principles and criteria applied in the maintenance of the data once it has been captured. This includes the maintenance and update frequency (frequency with which changes and additions are made to the data product).

This statement, if provided, is simply a textual description and does not provide for / require any substructure in the information to be provided in the data specification.

Clause 19 of the Generic Conceptual Model spells out requirements for the description of maintenance in INSPIRE data specifications.

7.8 Portrayal information (optional)

Recommendation 7
If a generally accepted portrayal rule for spatial object types of the themes exist, these should be referenced from the data specification and a mapping from the spatial object types in the existing portrayal specification to the spatial object types in the INSPIRE data specification should be provided.

The draft implementing rule on view services spells out requirements for the description of map layers.

7.9 Delivery

ISO 19131 introduces two components of delivery information: the delivery medium and the delivery format.

The draft implementing rule on download services and the guidelines for the encoding of data spell out requirements for the delivery of spatial data.

7.10 Additional information (optional)

This statement covers all kind of requirements that are not predefined in ISO 19131 or underlying standards. The analysis of the user requirements, in particular from use cases outside of the spatial domain, may point to additional categories of information that should be covered in INSPIRE data specifications.

7.11 Metadata

The implementing rules on metadata concentrate on metadata for discovery and for first level evaluation of a data set or data series as required by the Directive. The implementing rule specifies a core set of mandatory elements for that purpose.

Recommendation 8
The implementing rule on Metadata leaves it open to each INSPIRE data specification to define which metadata elements for evaluation should be made mandatory or mandatory by condition based on theme-specific requirements and practises, in addition to the set of mandatory elements. If any additional metadata elements are required based on user requirements, they should be specified as part of the INSPIRE data specification.

Clause 18 of the Generic Conceptual Model spells out requirements for the description of metadata in INSPIRE data specifications.

Metadata for use is comprised of the INSPIRE data specifications and the information published in the INSPIRE registers.

Recommendation 9
In cases where an INSPIRE data specification offers options, metadata elements should be identified in the specification so that a data provider is able to indicate the chosen options in the data set metadata.

NOTE If it turns out that there are common metadata elements used by multiple INSPIRE data specifications, these will be moved to the Generic Conceptual Model or into an amended implementing rule on metadata. The latter option is the preferred option.

7.12 Service specifications

Services will be specified in the implementing rules drafted by the Drafting Team Network Services. Additional, complementary spatial data services may be specified including the definition of service metadata for publication in discovery services.

Annex A
(informative)

Harmonisation Guidelines

A.1 Introduction

Two aspects of harmonisation need to be distinguished:

- a common process and methodology of developing data specifications in order to have an harmonised conceptual schema for all the themes involved in INSPIRE (this is addressed by D2.5 as well as by D2.6);
- for every individual data specification a conceptual schema needs to be designed that is capable of representing data from the various data sources and providers that need to provide the content for the download services. It is this aspect which is addressed in this chapter.

The objective of this annex is to give recommendations and examples of good practices, when defining harmonised specifications. As INSPIRE is about existing data, the harmonised specification for a given product must not only meet the user requirements but they must also be designed in a way which make easy or at least possible the matching between existing data sources and the harmonised specification.

These examples come mainly from:
- EuroGeographics projects (RISE, ERM, EuroRoads, …)
- international organisations (IUGS, …)
- any organisation having to harmonise data (e.g. WFD working group).

The recommendations and examples given in this Annex are structured according to the data interoperability components identified in D2.5.

The aim of this Annex is to propose and illustrate harmonisation approaches that may help to solve the gaps between requirements and existing data. The examples are not representative and are intended to illustrate specific aspects only.

A.2 INSPIRE principles

One of the main challenges regarding the design constraints on the development of harmonised European data specifications is to find the right balance between a simple, easy to implement/use solution and a complex, powerful solution depending on the needs and feasibility constraints. See the figure below.

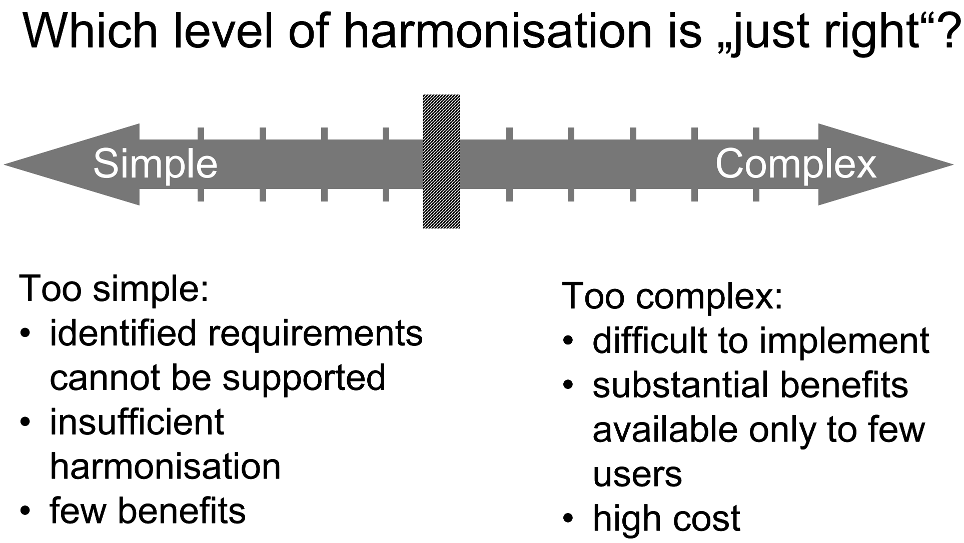


Figure 7

Some of the INSPIRE principles (see data interoperability component A) give overall guidance on the design constraints by stating:

- Data should be collected once and maintained at the level where this can be done most effectively
- It must be possible to combine seamlessly spatial information from different sources across Europe and share it between many users and applications
- It must be possible for information collected at one level to be shared between all the different levels, e.g. detailed for detailed investigations, general for strategic purposes

NOTE: however, for some themes, there may be only complex applications or very simple applications, for instance when users for the data may cluster strongly into expert and end-user (so, there are few users in the middle ground). In this case, it may be difficult to find an “average” solution.

A.3 Terminology

Terms from the INSPIRE Glossary have to be used when developing the thematic specifications. See D2.5.

A.4 Reference model

See requirements and recommendations in the Generic Conceptual Model (D2.5).

A.5 Rules for application schemas

Recommendation 10
To adapt to various existing data, the common application schema should have some flexibility.

This flexibility will depend on the user requirements:
- Strong user requirements, user requirements well-known: little flexibility
- Weak user requirements, user requirements not known in detail: more flexibility.

There are several manners to ensure flexibility in the application schema:
- by constraints on features and properties (attributes and association roles):
    - as INSPIRE is based on existing data, feature types and most attributes should be considered as optional
    - an attribute may be considered has mandatory if the feature would be meaningless without it, e.g. a geographical name without a name (see recommendation 5 in D2.5)
- by allowing flexible list of possible values for an attribute
- by offering options for representing the same real-world phenomena.

Recommendation 11
When options are offered, recommendations about the usage of the different options should be provided. If one should be privileged, indicate it.

Remark:
A flexible list of possible values for an attribute may be achieved through 2 ways:

- by defining a set of possible values in the common conceptual schema and the data provider may choose only a subset (see example 1)
- by defining a core set of possible values in the common conceptual schema and the data provider may extend this set (see example 2).

However, extensible code lists may be difficult to implement and to use. TWG must find a balance between flexibility and requirements for implementation.

Example 1: Attribute “FunctionalRoadClass” in EuroRoadS

ER_FunctionalRoadClass		
Definition	Traffic importance of the road	
Value domain	0	The most important roads in a given network [GDF]
	1	The second most important road in a given network
	…	…
	n	The least important roads in a given network. [GDF] 10 classes in GDF, RADEF
	Unknown	A not classified road e.g. a road measured by photogrammetry
Metadata	Number of classes in delivered data set	

The functional importance of a road segment has 10 possible values in GDF (standard for ITS) but it may have 5 possible values in one country, 8 possible values in another one, ...

The reasons for allowing different number of classes according to countries are the following:
- In existing data, the number of values that are used from this list depend on the road classification system, which in turn typically varies from one country to another.
- The user requirements are not known; some users may want to analyse or display data using 4 classes, whereas other users would prefer 5, 6 or more.
- moreover, as EuroRoadS is a project aimed to business users, some harmonisation work may be let to users, as they have the technical tools and skills to do so (for instance,

mapping the heterogeneous national classifications in the one required by their application; to do so, they need to know the number of classes used in each classification, it is why this information is required as metadata by the EuroRoadS data product specification).

This example illustrates that different member states or organisations may have different classification systems for the same spatial object type that can be supported by such a model.

Example 2: A reason for update of cadastral parcels in the German AAA model

In the German AAA model, every object version can carry information about the cause for the change that created this version.

The AAA model itself specifies a base list of causes in a codelist includes values such as "parcels merged", "parcel split", "parcels split and merged", "change in ownership", etc. In addition, every German state may create additional codelist values for such causes depending on their legal or organisational framework.

Example 3: Sea information in Eurospec

The Eurospec Feature Catalogue on hydrographic theme offers three ways to model sea information:

- o Sea water
- o Foreshore
- o Land Water Boundary

These 3 feature types are conditional: at least, one of them has to be present.

Example 4: Bridges and tunnels in EuroRoadS (grade separated crossings)

Grade separated crossings can be described in different ways. As road elements without node (omission of a node implies no crossing in the same level or two links cross each other without being cut in their intersection implies no crossing at the same level), see figure16. It can be complemented with different point features e.g. separated crossing with the complementary attribute upper/lower. The separated crossing can also been described with the coordinate points height value at each link.

Figure 8 – Example A is a network only with nodes at level crossings and example B a planar network with a node in the crossing. This node will be complemented with the information separated crossing

The following table defines the three possible levels for the road link representing the grade separated crossing:

Attribute	**Definition**
startLevel	Used when the road link is connected to a start node which in the planar topology case represents a grade separated crossing. The level shall be an integer number between -9 and +9 where the lowest number indicates lowest level.
endLevel	Used when the road link is connected to a end node which in the planar topology case represents a grade separated crossing. The level shall be an integer number between -9 and +9 where the lowest number indicates lowest level.
intermediateLevel	Used in the non-planar topology case where z values either are missing or are unreliable to indicate the level for the link. The level shall be an integer number between -9 and +9 where the lowest number indicates lowest level.
Constraint	**Definition**
correctIntermediateLevel	If intermediateLevel is set, neither startLevel nor endLevel shall be set. This implies that startLevel or endLevel can not be set at the same time that intermediateLevel is set. OCL: intermediateLevel->size() = 1 implies startLevel->size() + endLevel->size() = 0
correctStartLevelNode	…
correctEndLevelNode	…

Of course, it is no use to have a perfect conceptual model with no data in it. So, when INSPIRE semantic application schemas are developed it should be taken into account that the application schema needs to facilitate (easy) transformation rules.

Recommendation 12
Avoid enumeration with predefined quantitative intervals.

Example : Width of watercourses

- Watercourse in ROUTE 500 (IGN France) :
  Geometry type: Line
  Width :
  - less than 15 m
  - between 15 and 50 m
  - more than 50 m.

This application schema is convenient for a national database but not for data interoperability.

- ***Watercourse in ERM***
  Geometry type: Line

  WD7 Width Lower range
  Data Type : short integer
  Measurement unit : 1 meter
  Domain : Range value, >
  WD8 Width Upper range
  Data Type : short integer
  Measurement unit : 1 meter
  Domain : Range value, ≤

This application schema is convenient for data interoperability.

Recommendation 13
Use pyramidal classifications, when appropriate

Pyramidal classifications have the advantage of allowing several levels of detail for semantic information. Data providers with little existing information will fill the first level (the most generic) as a priority; later, they may fill the other level(s) perhaps after upgrading their data. Data providers with rich information will fill all the levels (even the most specific).

Example 1: Corine Land Cover classification

- 4 Wetlands
  - 4.1 Inland wetlands
    - 4.1.1 Inland marshes
    - 4.1.2 Peat bogs
  - 4.2 Maritime wetlands
    - 4.2.1 Salt marshes
    - 4.2.2 Salines
    - 4.2.3 Intertidal flats
- 5 Water bodies
  - 5.1 Inland waters
    - 5.1.1 Water courses
    - 5.1.2 Water bodies
  - 5.2 Marine waters
    - 5.2.1 Coastal lagoons
    - 5.2.2 Estuaries
    - 5.2.3 Sea and ocean

NOTE 1: This example is given to illustrate how a pyramidal classification looks like. It is not a recommendation to use Corine Land Cover classification for the INSPIRE theme “land cover”. Other ways to model land cover exist and it will be to the Thematic Working Group in charge of this theme to choose the best way to model land cover data.

NOTE 2: Classifications should not be used mixing different attributes, e.g. values as “local road with 2 lanes” must be avoided.

Example 2: Eurospec

Choice of a very generic feature type e.g. Watercourse
More detailed information is given by an attribute (Watercourse category) with possible values:

- o Ditch
- o Aqueduct
- o Canal – channel
- o Culvert
- o River

A.6 Spatial and temporal aspects

Recommendation 14
For geometry and topology, choose the minimum profile of ISO 19107 and ISO 19108 which may fulfil the user requirements.

NOTE 1: Examples on temporal aspects are missing at present.

NOTE 2: a minimum profile of ISO 19107 is recommended by the Generic Conceptual Model (recommendation 11) : use of the Simple Feature spatial schema as defined by OGC document 06-103r.

Example1: curve interpolations

ATKIS data in Germany may use three different curve interpolation types for topographic data like buildings:

- o linear (line strings, GM_LineString in ISO 19107)
- o circularArc3Points (arcs with 3 control points on the arc, GM_Arc in ISO 19107)
- o cubicSpline (a cubic spline, GM_CubicSpline in ISO 19107).

This is expressed in the model using OCL constraints. For example, the constraint on spatial object types with a surface geometry is:

```
context AU_Flaechenobjekt inv:
self.position.oclIsTypeOf(GM_PolyhedralSurface) implies
 self.position.boundary().exterior->forAll( r : GM_Ring |
  r->forAll( oc : GM_OrientableCurve |
   oc.primitive.segment->forAll( s : GM_CurveSegment |
    s.interpolation = 'linear' or
    s.interpolation = 'circularArc3Points' or
    s.interpolation = 'cubicSpline')))
  and
 self.position.boundary().interior->forAll( r : GM_Ring |
  r->forAll( oc : GM_OrientableCurve |
   oc.primitive.segment->forAll( s : GM_CurveSegment |
    s.interpolation = 'linear' or
    s.interpolation = 'circularArc3Points' or
    s.interpolation = 'cubicSpline')))
```

An additional constraint on a specific spatial object type, in this case a building ("Gebäude"), which may not use cubic splines in the surface boundaries is expressed as:

```
context AX_Gebaeude inv:
self.position.oclIsTypeOf(GM_PolyhedralSurface) implies
 self.position.boundary().exterior->forAll( r : GM_Ring |
  r->forAll( oc : GM_OrientableCurve |
   oc.primitive.segment->forAll( s : GM_CurveSegment |
    s.interpolation <> 'cubicSpline')))
  and
 self.position.boundary().interior->forAll( r : GM_Ring |
  r->forAll( oc : GM_OrientableCurve |
   oc.primitive.segment->forAll( s : GM_CurveSegment |
    s.interpolation <> 'cubicSpline')))
```

Other member states use different sets of interpolation types. In a harmonised model it may be appropriate to use linear interpolation only unless there is a user requirement for specific other interpolation types by the applications using the data.

Recommendation 15
Describe topology rules based on user requirements.

Example 1: ERM

A general and important point in data interoperability is that the data model characteristics of many databases maintained by NMAs are still very much determined by their cartographic origin: topological consistency between all area and line features as well as linear connectivity of networked line features are not always strict requirements for producing graphically consistent maps, but they are needed for almost every useful implementation of GIS based applications. Therefore the pragmatic approach adopted by the ERM project did not accept any compromise on data model aspects, and the consortium partners did upgrade their existing databases wherever necessary to obtain full topological consistency and connectivity.

Example 2: EuroRoadS

The EuroRoads data model offers 2 options:

- the topological one : there are nodes where a road link starts and ends or in crossings at the same level (fig 17)
- the geometric one : there are nodes only if they carry a specific information, e.g. a named junction (fig 18).

The decision to allow these 2 options has been based on the following reasons:

- some data providers have data in the topological option, some other have data in the geometric one
- no explicit user requirement about topology is known; moreover, as EuroRoadS is a project aimed to business users, some harmonisation work may be let to users, as they have the technical tools and skills to do so (for instance, transforming the geometrical option in the topological one, if they realise later that they need it).

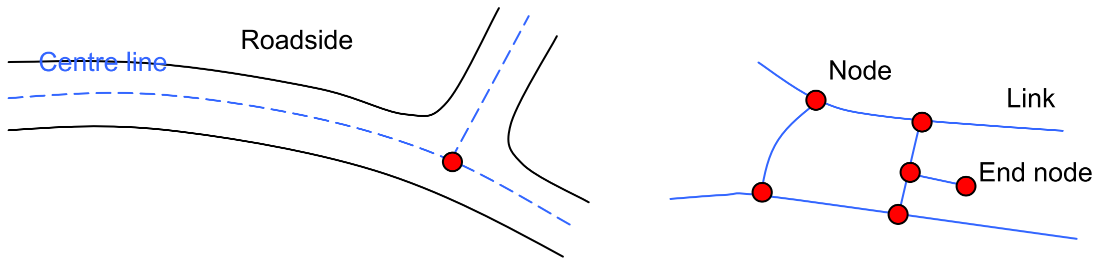


Figure 9 – The road represented as a centre line. A level crossing and a road end are represented with a node

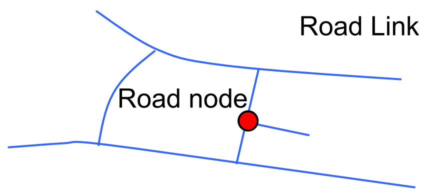


Figure 10 – Example of a road network spatially represented by geometry

The topological option is modelled as shown in figure Figure 11 – Topological representation option.

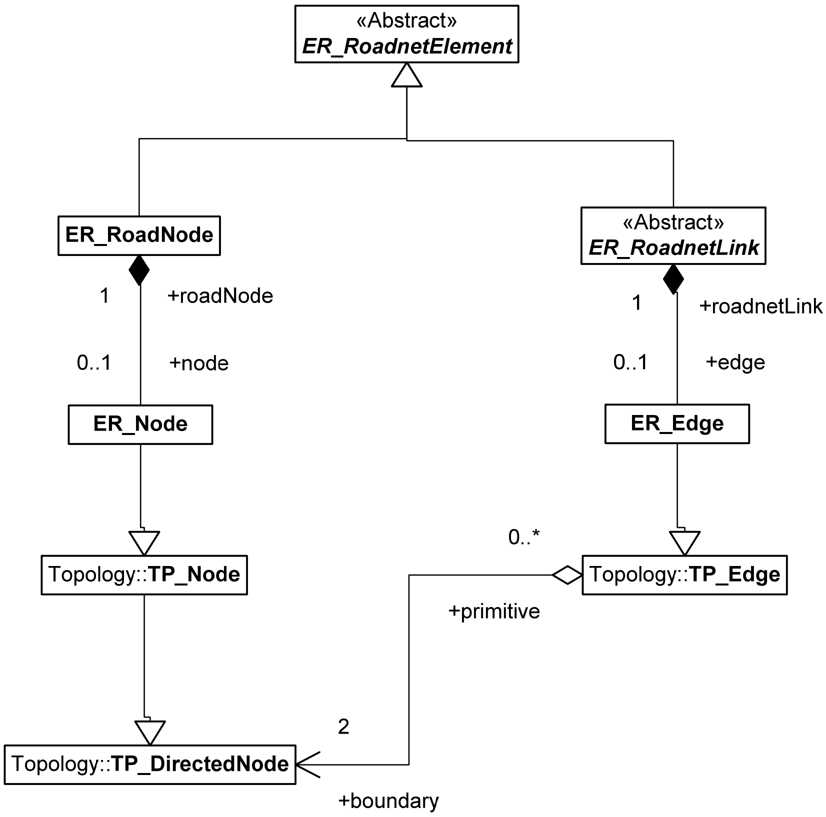


Figure 11 – Topological representation option

The geometric option is modelled as shown in figure Figure 12 – Geometric representation option.

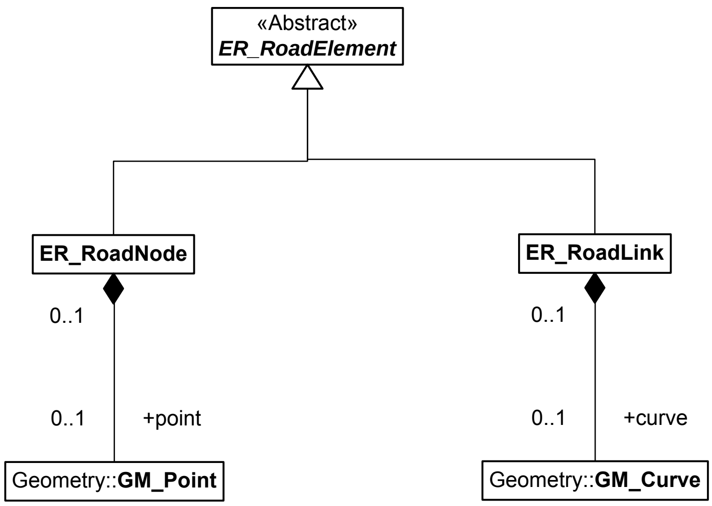


Figure 12 – Geometric representation option

A.7 Multilingual text

See requirements and recommendations in D2.5.

A.8 Coordinate referencing – units of measurements

D2.5 provides recommendations about coordinate reference systems to be used in INSPIRE. Local or national data will often be in different CRS. The transformation will be achieved by a coordinate transformation service (a service offering coordinate operations between two coordinate reference systems according to ISO 19111).

The next question is whether the quality of the transformation service will have any significant impact on the data quality as such.

Example :

It is possible to transform coordinates from the local CRS to ETRS89 (CRS recommended in D2.5):
- by a transformation with 7 parameters
- by using a grid

The first transformation is "less accurate" (e.g. for France, it implies errors of some meters) but requires less effort. The use of this transformation will imply some loss of accuracy. This transformation is adapted to middle or small scale data. For large scale data, it will deteriorate the data accuracy.

The second transformation is "accurate" but requires more effort: the grid is not always available, the computation is longer.

Recommendation 16
Influence of conversions between different CRS has to be considered when defining requirements about positional accuracy.

A.9 Object referencing modelling

Recommendations about this component are in D2.5 clause 13.

A.10 Data transformation model/guidelines

Recommendation 17
The development of application schemas should take into account that both the schema and its documentation need to facilitate transformation processes, as far as possible.

Clause 5.3 describes the two ways to implement the common application schema, i.e. how to transform existing data in the required common application schema:
- on the fly-conversion
- conversion to a derived data set

The transformations are under the responsibility of the Member States (and of data producers, data providers) which are responsible and free to choose the transformation method. The following examples are just given to provide TWG and data producers with a better understanding of this component.

Example 1: RISE for DEM

Making a common DEM in a cross-border region from different national grids ideally implies the following steps:

1- transform the national grids from national horizontal CRS to European common horizontal CRS

2- transform the altitudes from national vertical CRS to European common vertical CRS
3- re-interpolate the grids (resample) to agreed spatial resolution
4- combine harmonised national grids to single dataset for cross-border region
5- ensure consistency at national boundaries (e.g. smoothing to avoid artificial “cliffs”).

Steps 1-4 were demonstrated in RISE with steps 1-3 being done in the WCS.

With regard to the two CRS transforms – whether a WCS can do it depends on the transformation library that it is accessing. And this library tends to be more generic than the WCS. For example, RISE used the Minnesota WCS which made use of GDAL, the PROJ4 libraries and EPSG projection codes. It appears that the horizontal CRS and the vertical CRS are handled at the same time. One specifies both the horizontal and vertical CRS of the input file and the horizontal and vertical CRS for the output file.
When calling a WCS one can specify the horizontal CRS (with an EPSG code), but the vertical datum has to be set up in the WCS configuration. So one would need one WCS for projections based on WGS84; and a different WCS for output projections based, for example, on the Swedish datum.
One doesn’t need to modify the input data in either case.

With regard to the re-sampling, the WCS can resample either up or down. In the RISE test we only showed it going down, i.e., going from 25m to 50m. But it is possible to choose any sampling interval. It is also possible to select the interpolation function as a WCS configuration item. In RISE we just used nearest neighbour, but one can select linear or bicubic spline.

WCS can only output data for one layer of information at a time. So, if it is loaded with say a Swedish and a Norwegian DEM and one wanted a DEM for a cross-border region, then one must make two calls to the WCS (one for the Swedish data, and one for the Norwegian data) and then combine them in a separate software package. The combination process is very simple, since both datasets can have the same projection and sampling interval, so a simple merging of the grids is all that is required. This is what we did in RISE to get the DEM of the cross-border region.

The reprojection on the fly is relatively quick and is not an issue, assuming one is not trying to handle huge areas at very high resolution

Example 2 : SDIGER experience

The rules of translation between national and common model are simple, only based on the schemas (on features and attributes), e.g. spatial requests are not required.

Three main categories of approaches have been identified to harmonise or restructure the local repositories to be compliant with the agreed common application schemas for the SDIGER use-case scenarios:

- Data reorganisation. This is the off-line database conversions from the local models into the common application schemas defined.

- Using views. If the geographic data are managed in a relational database management system, they can be re-organised through views. That is to say, the approach consists in the creation of database views of the local schemas, but being these views compliant with the common application schema.

- Reorganise the data in real-time. This third approach consists of providing on-the-fly transformations. One of the possible tools in this category is the analysis of the on-the-fly schema translation tool from the GIMODIG project (data harmonisation project for cartographic purposes in Germany, Finland, Norway, Sweden), which is based on the use of XSLT documents. Figure 21 shows how the GIMODIG allows the transformation of GML data obtained from local Web Feature Services into GML data compliant to the GML schema derived from the common models.

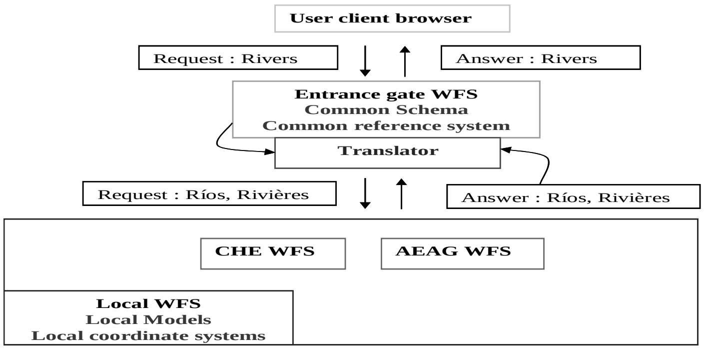


Figure 21: coordinate transformation and schema translation in the SDIGER project

The SDIGER project has been investigating all these alternatives. Apart from studying the feasibility of the GIMODIG approach, the more pragmatic approach of “using views” is also taken into account for the development of the Web Application. The Web Application explores both the direct access to the database to retrieve views data; and the access to Web Feature Services with database connection to these views and that provide this data in GML format.

Conversion of data on the fly is not very efficient. The GIMODIG approach has been tested with different configurations of software and hardware. These series have revealed that the bottleneck in the process seems to be the XSLT transformation in itself.

<u>Example 3:</u> schema translation tools

Some SDIC/LMO have already investigated this area and launched research projects in this domain.
For instance, a PhD made at IGN France has developed a method (implemented in a prototype) to enable a user to choose its application schema at conceptual level by modifying the application schema of the producer data, the transformation being checked (consistency controls) and "metadated". Then, this schema translation is propagated to the other levels (for implementation).
Swisstopo and BKG have launched a project to develop mdWFS (model driven WFS); the tool enables the user to define matching rules between two conceptual schemas and then propagate this schema translation to the other levels (for implementation). A prototype of this tool should be available since end of 2007.

<u>Example 4:</u> from geometry oriented modelling to object oriented modelling (ISO 19109 clause 8.7.6)

The traditional way of structuring geographic data does not distinguish between features and geometric primitives, but includes geometric information in the definition of a feature type. Thus, features are classified as point features, line features and area features because of the nature of the geometry. Large amounts of existing geographic data and functional standards are based on this way of structuring the geographic data.

This International Standard uses the geographic feature as the fundamental unit of geographic information. The geometry is one of several ways of describing the feature. Since a feature type is not defined on the basis of its geometry, several geometric descriptions may be associated to the same feature. It is recommended that point features, line features and area features are redefined in a generalized form as geographic features.

Rules:
1. A point feature shall take a GM_Point as the value of its spatial attribute.
2. A line feature shall take a GM_Curve or a GM_CompositeCurve as the value of its spatial attribute.
3. An area feature shall take a GM_Surface or a GM_CompositeSurface as the value of its spatial attribute.

NOTE A study about schema transformation has been proposed to the Consolidation Team. This study should provide more input for this component. Other guidelines will come from the Implementing rule about Transformation Services.

A.11 Portrayal

Recommendation 18
The symbology should be chosen in order to have cartographically correct data representation.

Data representation is considered correct if the relationships between features are respected:
- differentiation (e.g. between different feature types)
- association (e.g. within a same feature type)
- order (e.g. according to different classifications within a feature type)

Figure 13

For instance, these three legends are very different but are all considered as cartographically correct as they respect the principles of differentiation, association and order.

Furthermore, most of the usual cartographic rules should have to be respected (e.g. the symbol size must be big enough in order to be perceptible by a human eye but not too big in order not to hide the surroundings).

Recommendation 19
Portrayal rules in INSPIRE should build upon the existing cartographic experience and upon common traditions, if they exist.

For many years, people have read maps; of course, the legends are different across Europe. However, there are some common conventions (e.g. hydrographic features are generally represented with the blue colour, forests with the green one). When such conventions exist, it is better to keep them, as the user may recognise a feature type or at least a theme without looking at the legend.

Example 1:

In the previous illustrations, the legend used in the middle example, which is more traditional and where differentiation is more obvious would probably be the preferred option in the INSPIRE context.

Recommendation 20
If, for a given theme, harmonised rules about portrayal already exist, they should be adopted by INSPIRE data specifications.

Example 2:

One of the major objectives of the International Commission of Stratigraphy is the establishment of a standard, globally applicable stratigraphic scale. The International Commission of Stratigraphy has elaborated a common chart (http://www.stratigraphy.org/cheu.pdf) for similar display of the age of sedimentary rocks on geological paper maps, digital maps and web mapping applications.

For the colours, the ICS adapted the colours traditional in use for small-scale geological maps in Europe, promoted by the Commission of the Geological Map of the World.

If the age definitions and boundaries are adopted by INSPIRE as standard for the description of the age of rocks, the colours defined by this chart could be used by WMS Services

A.12 Identifier management

D2.5 states that unique identifiers should be persistent; however the life-cycle rules do vary from Member State to Member State.

Recommendation 21
Spatial object life-cycle rules will likely vary from member state to member state, so the requirements on this topic stated in the INSPIRE data product specifications should be as flexible as possible because feasibility is important and needs to be taken into account.

For instance, TWG should only specified that a change in a particular property shall lead to a change in the identity of a spatial object (a new spatial object with a new spatial object identifier) where this is a known user requirement and where this can be accommodated by maintenance regimes for the existing data sets.

But, as stated in requirement 68 in D2.5, when required by users, the life-cycle rules for spatial object types in a spatial data set shall be documented, with enough details, in the data set metadata. The first example below shows why life-cycle rules may be required by users. The following examples provide best practice about how to document these life-cycle rules.

Example 1: management of a drilling rig by several organisms

Imagine a drilling rig – for this story, a complex steel structure, owned and operated by a commercial organisation, and used for several years in a variety of places. A number of organisations will have an interest in this rig, at least the owning / operating company; the ‘department of industry’ in the operating country, responsible for licensing the rig to drill; the hydrographic department (or other maritime safety organisation) responsible for the waters where the rig operates.
Assume that each of these organisations maintains a geographic database, or at least a database with positional information alongside their other records, and needs to have a record of this rig.

The Company

The company will probably create its information record when it first considers acquiring the rig (ordering it to be built, or buying it from another company). This record will (or at least, ideally should) keep its 'identity' so long as the company is responsible for the rig – which may be many years after they have sold it on or scrapped it. This identity will ideally be captured in a single identifier, which can be the target for links from e.g. staff records, production records, maintenance records.

When the rig is first deployed, when it is relocated, even if all its components are replaced over time, it will keep the same identity.

The Department of Industry

Consider their role in licensing the rig to drill. The department will create an information record when the company first applies to drill in their waters. They may have a policy of tracking individual rigs, or they may be more interested in drilling sites. But even if they track the rig, they are starting after the company, and will 'cancel' the record if the rig transfers to drill in a different countries waters. If, some years later, the company moves the rig back into the first countries waters, they may not need to associate this new operation with the previous one.

The Hydrographic Office (HO)

The focus of interest here is with the rig as a hazard to navigation, and also as a destination for shipping. If it is a large rig, Hydrographic Office may notify shipping as it is moved from its build location out to see, but would not generally create a geographical record for this temporary and mobile hazard. HO would create a record as soon as HO is made aware that the rig will be operating at a particular location; ideally this would be in advance of the work starting.

HO would associate other records with it – lights, AIS or radar signature, safe clearance areas.

If the rig moves a few hundred metres, HO may keep the same hazard record and change its position. Any move more than that, and HO would most likely cancel the hazard and create a new one in the new location.

The Moral

A same entity will be considered by different organisms as different spatial objects with different life-cycle rules. To the company, it is a rig and changes location, colour or whatever over time; to the licensing agency, it is a facility which is licensed to do certain things in a certain place for a certain period; to the maritime agency, it is a hazard which exists in a place for a period.

Each of these agencies would do well to keep a record of the other identifiers, but will need to understand the other data managers' life cycle rules.

<u>Example 2:</u> TOP 10 in Netherlands

What happens with a modified spatial object depends on the size of a modification. At 'small' (or minor) modifications a spatial object gets a new version date. At 'large' (or very significant) modifications the original spatial object terminates and a new spatial object is erected with a new TOP10_ID and a starting date. Then the original spatial object is considered to have disappeared, and is assigned an end date. If a modification is 'small' or 'large', is subjective. For that reason rules have been defined, to judge to what extent spatial object modifications leads up to a new version date or new TOP10_ID. When there is a modification, it can be a modification of the spatial object type, a change of attributes, or the geometry has changed. Depending on the kind of modification certain rules are applied.

A change of spatial object type is always considered a large change, and therefore the old spatial object gets an end (spatial object + version) date and the new spatial object gets a new TOP10_ID and a start (spatial object + version) date equal to end date of previous spatial object. A change of the geometry type (e.g. from line to point) is also always considered a large change. When an attribute has changed, in most cases, only the version date is changed.

In order to decide if a geometry change is large or small the *'overlap rule'* is applied (see below for more details). This rule is applied to the main geometry of the spatial object (as a spatial object can have multiple geometry attributes). Changes in the coordinates of the non-main geometry are considered a small change. In case of point or line main geometry attribute, a buffer operation is applied first (with buffer size is 3 meters at both sides, so total width is then 6 meters; see figure 22). When the changes are within the buffer (based on the different rules below), the spatial object keeps the same TOP10_ID and alone the version date is changed.

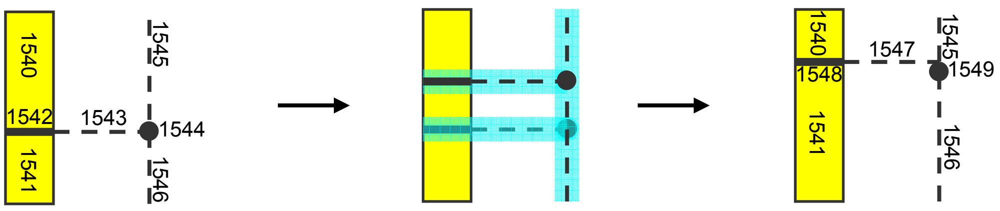


Figure 14 – Application of buffers (light blue) around point and line spatial objects.

A watercourse gets another course.
The new watercourse overlaps enough with the original watercourse; so, according to the overlap rule, it takes over the TOP10_ID. See figure 23.

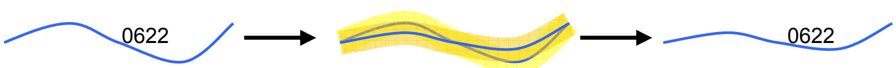


Figure 23 – New version of watercourse is within buffer around old version, so the TOP10_ID is kept.

- *Overlap rule*

In case a new spatial object overlaps with an old spatial object of the same spatial object type (and at same height level) and the area of the overlap is more than half of the area of the old spatial object then the new spatial object gets the TOP10_ID of the old spatial object (see figure 24). That is, this is considered a small change. Note that an old TOP10_ID can at most be assigned to one new spatial object (as only one new spatial object can have more than half of the area of the old spatial object). However, in case a new spatial object would get more TOP10_ID's then the id of the largest overlapping spatial object is used.

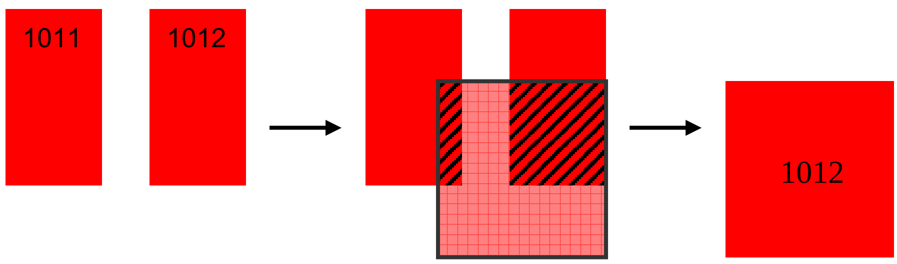


Figure 24 – Application of the overlap rule: where the new spatial object has an overlap with more than half of the area of the old spatial object, then the new spatial object gets the old TOP10_ID. Here this is the case with spatial object 1012 (but not with spatial object 1011).

- *Aggregation of spatial objects*

For this case a rule is defined to prevent that there can be more TOP10_ID's allocated to one spatial object and is applied when spatial objects are joined together. When the size of the new spatial object amounts less than 200% of the size of one of the original spatial objects, then it gets the existing TOP10_ID with a new version date. When the size of a new spatial object amounts 200% or more, compared to the original spatial object, then it does not get the original TOP10_ID (see figure 25).

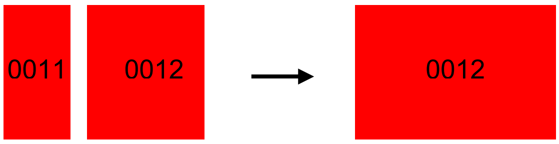


Figure 25 – Application of the aggregation rule. The size of the new spatial object is more than 200% of the size of the original spatial object with TOP10_ID 0011, so it does not use this ID. However, the size of the new spatial object is less than 200% of the size of the original spatial object with TOP10_ID 0012, the new spatial object keeps this ID.

- *Splitting rule*

This rule prevents several new spatial objects from getting the same TOP10_ID after splitting up a spatial object. When the size of a new spatial object is more than 50% of the size of the original spatial object, then it obtains the same TOP10_ID with a new version date. The other new erected spatial objects obtain a new TOP10_ID (see figure 26).

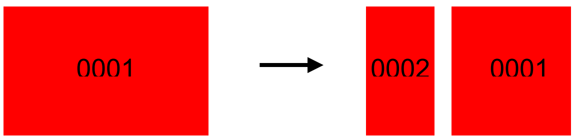


Figure 26 – Application of the splitting rule. The area of the small new spatial object is less than 50% of the area of the original spatial object. It doesn't take over the TOP10_ID but gets a new TOP10_ID. The area of the large new spatial object is more than 50% of the surface of original spatial object and takes over the TOP10_ID 0001.

Other examples

The spatial object characteristic metalling of a road changed from ' unpaved ' to ' paved or metalled '.
In this situation the spatial object only gets a new version date. See figure 27.

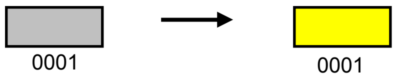


Figure 27 – Attribute change

The boundaries of two areas are changed: a terrain area becomes larger, at the cost of another terrain area. The light coloured area takes, according to the aggregation rule, the TOP10_ID of the original area (0491). The other area takes the TOP10_ID of the original dark coloured area (0492). See figure 28.

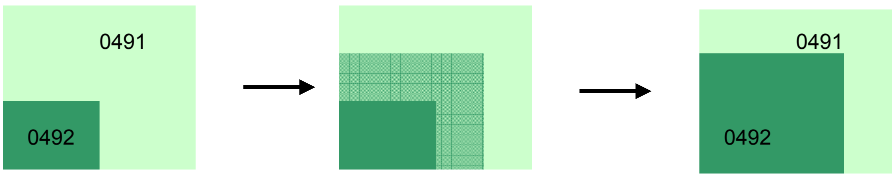


Figure 28 – Changing area

Example 3: (MasterMap – OS in UK)

The features within OS MasterMap vector layers are viewed as having a life cycle. The life cycle of each feature is matched, where practically possible, to that of the real-world object it represents. For example, a new building will become a new object in the Ordnance Survey main holding of the data and will be treated as the same feature – even if it undergoes change – until the building is demolished. By adopting this approach Ordnance Survey is emulating real-world behaviour within a digital model and therefore creating a more realistic version of the real world in a computer.

Feature life cycles are established and maintained in accordance with a set of published rules. Essentially, these rules indicate when an OS MasterMap feature will be retained and when it will be replaced, for different types of feature and different change scenarios. These rules are not only there to guide surveyors (from Ordnance Survey) collecting and attributing the features but also to provide customers with a consistent definition of how real-world change is handled by Ordnance Survey.

However, not all changes to the real-world object will be reflected in changes to the feature. For example, the addition of a new porch to a house would usually be considered too minor a change for Ordnance Survey data capture. Different customers with different applications think of feature life cycles in different ways. Understanding change is important to understanding the OS MasterMap product and to deriving the optimum value from it.

This information is given by the attribute “Change History” which can have the following possible values:

- New - new feature in the database.
- Position - feature changed geometry through accuracy improvement.
- Modified - feature has been edited by an operator e.g.
    - Geometry of a topographic feature changed after real-world change.
- Software - feature has been adjusted by an automatic software process.
- Reclassified - descriptive attributes of a feature have changed.
- TextChange - text string of text feature has changed.
- Restructured - new line feature(s) created from existing feature(s) where:
    - The feature is split into two or more features.
    - Two or more features are joined together.
- Attributes - applied to features with non-geometric attribute change
- Incomplete – area or line feature is incomplete e.g. during revision process

Life cycle rules adopt the approach of allowing features to persist through changes so far as is reasonable. There is inevitably some degree of subjectivity involved in judging that a real-world object has changed so much it can no longer be considered the same object, and therefore the OS MasterMap feature(s) representing it should be deleted and replaced.
The general way to deal with life-cycle rules and versioning is shown in the following figure.

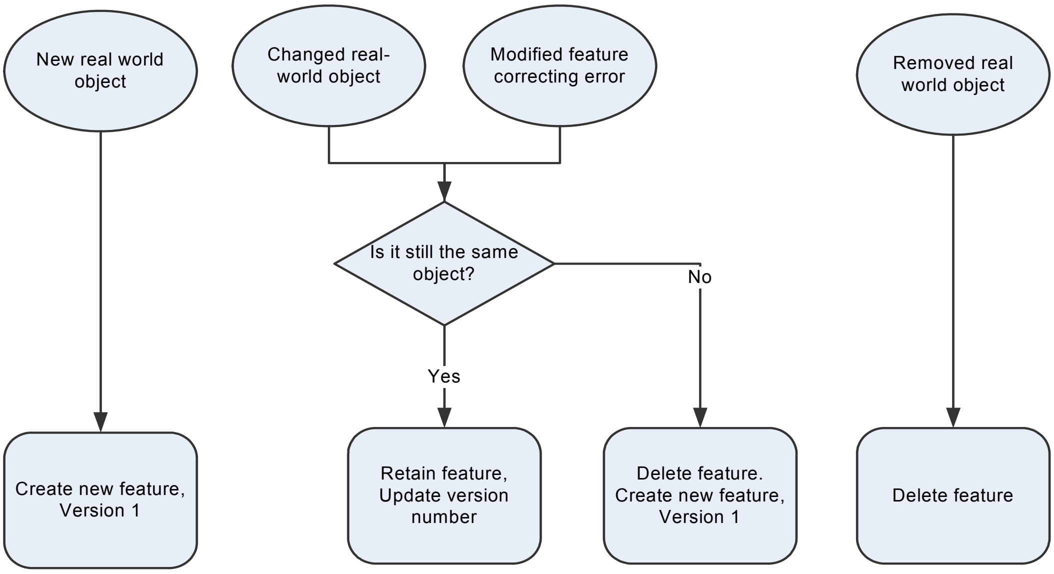


Figure 15

Each feature in OS MasterMap Topography Layer has a unique reference called a TOID. The vast majority of database systems used within organisations rely on the use of unique referencing for the efficient management of the data stored within them. Each feature also has a version number and a version date as well as the TOID. As the real-world feature that it represents changes during its life, so the feature within OS MasterMap will also change in terms of its shape or its attribution. Whilst it remains the same feature – essentially occupying the same space and having the same function – the TOID will not change, but the version number will increment and the version date will change. This allows an instance of a feature to be identified in both space and time.

NOTE: The previous examples are about vector data that are, more or less regularly, updated. The way to document life-cycles rules of spatial objects may be quite simpler for data which are not updated but re-observed or re-produced, generally coverage data.

Example 4: BD ORTHO, IGN France

BD ORTHO is made from photos taken every 5 years
.

- The rhythm of 5 years is given in the data product specification, in the “maintenance” component.
- The year when the photos have been taken is part of the identifier of each tile of BD ORTHO and gives the temporal validity of the data.

D2.5 considers only external identifiers. However, it will often be required to consider also thematic identifiers in application schemas.

Recommendation 22
If thematic identifiers need to be harmonised for some themes, it should be done by the concatenation of a two letters prefix for the member state with the national thematic identifier.

Example 1: Hydrological code

The RISE use case has shown the importance of the hydrological code. The method recommended by WFD to provide European hydrological codes is the following:

- use national hydrological code (free but maximum : 22 characters)
- add a two letters code to identify the member state.

So, it is compatible with the D2.5 recommendations about "unique identifiers".

Remark: in some cases, it may be required to harmonise the whole structure of the thematic identifier.

Example2: Code SHN and SHI (EuroBoundaryMap)

The SHN code represents the unique identifier of an administrative unit at any level within a country.
The SHI code is the key for deriving higher level units from polygons of the lowest level. It represents the number of characters that need to be replaced by zero from the right of the lowest level unit SHN code.

Example
The following French example illustrates the implicit relation between administrative units of different levels of hierarchy. SHN code of lower level unit contains SHN codes of higher level units. The key for deriving SHN codes of higher level units is SHI code in ISN table.

Name *(**GEN** attribute in NAM table)*	***SHI*** *(ISN table)*	***example of SHN*** *(PAT and NAM table)*
Commune	*0*	***4157705356***
Canton de rattachement	*3*	***4157705**000*
Arrondissement	*5*	***41577**00000*
Département	*6*	***4157**000000*
Région	*8*	***41**00000000*
République	*10*	*0000000000*

A.13 Registers and Registries

The examples in this chapter refer to feature catalogues and feature concept dictionaries.

Feature catalogue is a key component for feature-based data. During the harmonisation process, most of the discussions will be about the definition of features and attributes; generally, it will be relevant to use spreadsheets displaying, either the proposal for the feature catalogue or the matching tables with existing data.

These spreadsheets are very useful at different stages of the methodology:

- proposal (e.g. after the study of user requirements)
- census of existing data (may be used for questionnaires)

Example (Eurospec): matching table for hydrographic theme (for census of existing data)

Eurospec Feature Catalogue					Large scale National Database			Medium Scale National Database		
Feature name	Geometry	Feature code	Definition	Requirement	Presence	Geometry	Comment	Presence	Geometry	Comment
Watercourse	L, S	BH502	A natural or man-made flowing watercourse or stream	M						
Waterfall	P, L, S	BH180	A vertically decending part of a watercourse where it falls from a height (for example: over a rock or a precipice).	O						
Rapids	L, S	BH120	Portions of a stream with accelerated current where it descends rapidly but without a break in the slope of the bed sufficient to form a waterfall.	O						
River Bank	S	BH141	The region along the edge of a river that is between the water and the first break in slope.	O						
Dam	P, L, S	BI020	A barrier constructed to hold back water and raise its level to form a reservoir or to prevent flooding.	O						
Lock	P, L, S	BI030	An enclosure with a pair or series of gates used for raising or lowering vessels as they pass from one water level to another.	O						
Lake/Pond	S	BH080	A body of water entirely surrounded by land.	M						
Island	P, S	BA030	A land mass, other than a continent, surrounded by water.	O						

Recommendation 23
Choose terms and definitions as generic as possible.

Example 3: Eurospec

The Eurospec Feature Concept Dictionary will be based mostly on the DFDD (DGIWG Feature Data Dictionary).

DFDD may offer several terms for similar notions, e.g:

- Bog: A poorly drained tract of wet spongy ground consisting of decaying vegetation, that retains stagnant water, and is too soft to bear the weight of any heavy body.
- Marsh: A soft, poorly drained wetland that is characterised by the growth of only non-woody plants (for example: grasses) and often forms a transition region between a waterbody and land.
- Swamp: A seasonally flooded, poorly drained wetland with more woody plants than a marsh and better drainage than a bog.
- Wetland: A marsh, swamp, or other stretch of land that is usually saturated with water.

Wetland has been preferred for Eurospec because it is the most generic term (so matching from existing data to common model will be easier).

NOTE EuroSpec is an example of a feature catalogue. DFDD is an example of a feature concept dictionary.

A.14 Metadata

Metadata have a key role in a harmonised product. Metadata will enable data providers to supply information about:

- the options chosen in the common specifications (e.g. optional features and attributes)
- the recommendations from the common specifications that have not been fulfilled or only partly.

Example: EuroRoadS (examples of metadata)

Feature and/or attribute	Metadata information	Free text and examples
Form of way	Possible values: A list of possible values	This is the value domain used in NVDB: Motorway Dual Carriageway Single Carriageway Roundabout Pedestrian zone
National road class	Number of classes: byte	10 values in GDF and NVDB
Road node type	Possible values: A list of possible values	- Nodes are used only in junctions - Pseudo nodes are used to split links where an attribute change its value
Roundabout, minimum diameter	Value: byte Unit: meter feet	- Roundabouts with a centreline diameter less than 20 meter are stored as a node with the attribute value roundabout. Larger roundabout is stored with its real geometry

NVDB: National road Database in Sweden
GDF: Standard for road data and navigation applications.

A.15 Maintenance

This is an optional component in ISO 19131 - Data Product Specification. However, as it is more or less related with identifiers and versioned objects, see the corresponding chapters in D2.5 and in this document.

Users often want to know the data currency:

TABLE 3.2 POSSIBLE MEASURE OF CURRENCY

Measure	ISO 19100 quality element/sub-element	Comments
Units of change over period of time	Lineage Temporal accuracy/ new element: currency	There are two options a producer can select. It can be reported as lineage information or producer can make a new sub-element under temporal accuracy
Date of last update	Lineage	This can be reported as lineage for a certain area if updating is based on a certain area. Feature instance in the database should contain date
Rate of change	Lineage	An estimate of change of feature types over period of time
Temporal validity	Temporal accuracy/ Temporal validity	Validity of data with respect of time. e.g. valid/non valid/not_yet_valid
Update frequency	Lineage or as metadata MD_ MaintenanceInformation	Planned update based on quality model

From the Guidelines for Implementing the ISO 19100 Geographic Information Quality Standards in National Mapping and Cadastral Agencies

Some recommendation about data currency should be given in the INSPIRE product specification, if there are user requirements on this topic.

However, it will be impossible to enhance data currency without great effort from data producers. Generally, the best solution will be just to give t his information for each data set.

Recommendation 24
For each data set, register data currency in metadata

A.16 Quality

ISO/TC 211 has standardised a quality model, modelled as part of the metadata standard ISO 19115. The model consists of data quality information according to 5 quality elements, and lineage information. The 5 quality elements are:

- Positional accuracy
- Completeness
- Thematic accuracy
- Logical consistency
- Temporal accuracy

- Positional accuracy is generally needed by users. It should be considered by TWG, specially when several levels of detail are required for a theme, as it will contribute to define these levels of detail.

However, as INSPIRE is based on existing data, it won't be possible that all data sets will be compliant with the quality required (or at least desirable).

Recommendation 25
Specify the positional accuracy desirable in the common data specification; accept deviations and ask data providers to register them in metadata, at data set level.

Example: ERM

Features in EuroRegionalMap should have 125m of accuracy or better. However, a less accurate level is accepted depending on the data sources accuracy. The absolute horizontal accuracy and data sources have to be indicated in the metadata.

- Completeness has meaning only if rules of selection have been clearly defined.

Example 1: ERM

Presence or absence: availability of the information in the data (Yes/No)
For example: AP030 ROAD : Yes

Completeness: % of the selection of the features according to the specifications:
- if selection is fulfilling the selection requirements then 100 %
- % of “empty“ value for the attributes

For example: AP030 ROAD: Yes 100%
RTN National Route Number Yes 60%
Remarks: Comments if any
For example:
RTN National Route Number Yes 60% local roads not populated.

Regarding the % value, there is no need to proceed to an accurate calculation and this value can be roughly estimated. In this case, you can add a ± 60% or give a range of value 50-65%.

Example 2:
for an orthoimage, completeness may be given as the maximum percentage of cloudy areas.

- Thematic accuracy deals with the following questions:
  - Are there any requirement related to correctness of thematic classification?
  - Are there any requirements related to the accuracy of non-quantitative attributes?
  - Are there any requirements related to the accuracy of quantitative attributes?

Thematic accuracy may be given by error percentage.

- Temporal accuracy: generally, users rather want to know data currency.

Data currency is considered in this document under component (N) Maintenance, as data currency is not considered as a quality element by ISO.

- Logical consistency has meaning only if consistency rules have been specified. Main recommendations should be to have:
  - clear consistency rules
  - consistency rules that can be checked automatically (but it is generally the case).

  Example (consistency rules): WFD

  Consistency refers to the absence of apparent contradictions in the data set, database or transfer file. Consistency is a measure of the internal validity of a database, and is assessed using information that is contained in the database.

  Due to the lack of reference data, the most important part of the quality assurance process will be the assurance of the logical consistency of the data. The consistency applies to the features, the attribute-tables as well as to the attributes, and to the relationships. The relationships comprise the defined relationships between feature classes and attribute classes as well as to geometric relationships, e.g. sub-basins are covered by river basins.

  Conceptual Consistency

  The checks for conceptual consistency should include checking for the existence of the feature classes, the attribute classes, and the relationships that are defined in the model. The next step is to verify the existence and the correct definition of the features, attributes, domains, and relations in the database. Then it should be verified that attribute values exist, where these are defined, and that the relations are valid. The cardinality of the relations should conform to their definition. These quality checks will be applied by the EC when integrating the national GIS layers into the EU geographical database.

  In the data model it is expressed that simple features are stored in the feature classes. Consequently it should be verified that the features in the database are consistent with the definition of simple features. This includes, for example, that polygons are closed, that boundaries of the polygons must not intersect, and that holes and exclaves are considered correctly. Quality assurance on the validity of simple features are vital for the consistency of the database and should be applied by the Member States and reported by the EC.

  Table 3.5.4: Conceptual Consistency Elements

Element	**Obligation**	**reported by**
Existence of GIS layers, attribute tables, relationships, domains	mandatory	EC
Definition of attribute	mandatory	EC
Existence of attribute values, where mandatory	mandatory	EC
Verification of cardinality of relationships	mandatory	EC
Simple features definition	mandatory	EC

  Domain Consistency

  In the data model, a number of domains are defined. It should be verified that the definition of the domains is correct. Then it should be checked that the attribute values in the feature and attribute classes are consistent with the domain values. In addition to the existing domains, so-called value range domains should be set up, as soon as the dimensions for the items concerned are defined. The checks on domain consistency should be applied by the Member States and will be verified during the integration process that generates the European database.

Topological Consistency
There are a number of GIS layers and attributes that can be tested for topological consistency. Some of the GIS layers have a country indication. The Member States should ensure that the appropriate country code is used.

The water bodies have an attribute indicating the relation to the EcoRegion GIS layer. The relation between water bodies and its parent river basin district can be verified by overlaying the water bodies with the river basins. The EC will test the correctness of the assignment by overlaying the respective layers.

The Appendix V contains a set of topological rules applicable to the GIS layers. The rules will be tested by the EC when merging the national GIS layers. The correctness should be reported as part of the data quality element topological consistency.

The WFD database will be set up as a collection of data sets provided by the EU countries. It is recommended that the features crossing national boundaries should be coherent. This principle should apply to the geometry as well as to the attributes, e.g. the boundaries of river basins should meet at the border. The coding of the basin should be the same. The feature classes which could cover more than one country are in principle all polygon and line features, i.e. water bodies and river basins, sub-basins. This situation will be analysed by the EC when integrating the national GIS layers into a European database.

It is recommended that the hydrographical GIS layers should constitute a network. The directions of the lines should indicate flow directions. Flow lines should connect the incoming and outgoing river lines through a standing water body (e.g., lake). These connecting flow lines are termed *continua* in the data model. The data will be analysed by the EC when integrating the national GIS layers.

A.17 Data transfer

See requirements and recommendation in D2.7.

A.18 Consistency between data

Consistency includes logical consistency (coherence of data with the model) and the inter-representation consistency (absence of contradictions between the different representations), which itself includes

a) coherence between spatial objects of the same theme at different levels of detail
b) coherence between different spatial objects within a same area
c) along state boundaries.

The logical consistency is considered in component (O) Data & Information Quality.

Other examples about consistency at different levels of detail are given in component (R) Multiple representations.

Consistency between themes has to be modelled in the application schema by constraints (an example is given in annex C), as recommended in the Generic Conceptual Model.

However, consistency, as other components, has to be considered in the INSPIRE data product specification, only when appropriate, i.e. when required by users and when feasible.

<u>Example 1:</u> meteorological data
For meteorological ensembles, there may be more than 50 forecasts from the initial dateTime. The ensembles can be viewed as a Monte-Carlo simulation and exploration of the atmospheric chaotic attractor. In which case the lack of consistency between forecasts may be the important property to recognise.

Recommendation 26
The main point to ensure consistency across national boundaries is to combine data sets at similar levels of detail.

It would be meaningless to combine data sets with very different densities of information; so, this criteria has to be taken into account to decide which existing data have to be considered in the as-is analysis.

Recommendation 27
The possible levels of detail to be considered for INSPIRE data specifications are:
- European
- National
- Regional
- Local

Thematic Working Groups will have to define which scale range is involved in each level of detail.

Example 1 (geology):
- european level (continent) : from 1 : 5 000 000 to 1 : 1 500 000.
- national level (nation) : 1 : 250 000 to 1 : 1 000 000
- regional level : from 1 : 250 000 to 1 : 25 0000
- local level : scale > 1 : 25 000.

Example 2 (cadastral parcels) :
only local level : scale ≥ 1 : 10 000

Example 3 (land use):
- local high level : around 1 : 1000
- local low level : around 1 : 5 000
- regional level : around 1 : 50 000

Example 4 (topographic maps – hydrography, transport and land cover):

Robinson (Robinson et al., 1995) states that there is no consensus on the quantitative limits of the terms small, medium, and large scale in topographic maps, since the terms are relative. Most cartographers would agree, however, that a classification may be defined as follows:
- large scale: ≥ 1 : 50.000
- medium scale: < 1 : 50.000 and > 1: 500.000
- small scale: ≤ 1: 500.000

These threshold values constitute a compromise. Classification may deviate according to the density of population in particular countries, because this parameter is closely related with the density of information in topographic maps. National and local scales are also dictated by the size of the country. For instance Germany (Neudeck, 2001):
- large scale: ≥ 1: 10.000
- medium scale: < 1: 10.000 and > 1: 250.000
- small scale: ≤ 1: 250.000

or Russia (Berljant, 2002):
- large scale: ≥ 1: 200.000
- medium scale: < 1: 200.000 and ≥ 1: 1.000.000
- small scale: < 1: 1.000.000

Consistency across national boundaries also involves edge-matching issues. This question is considered in Annex B of this document.

A.19 Derived reporting/multiple representation

The Network Services Drafting Team investigated generalisation as a possible part of the INSPIRE data transformation services at the Vienna meeting on 19-20 April 2006. Because of its immaturity the DT did not recommend the use of generalisation services, but supported multiple-representation.

A workshop about data consistency and multiple representation was held in Ispra on 7 and 8 November 2006. The report of this workshop provides state-of-the-art about these topics and gives useful recommendations, such as:

- Automatic generalisation methods are not mature enough to be considered as a service in the ESDI. For practical reasons multiple scale representation is generally needed, which can be completed by generalisation

- Establishing and/or preserving links between different representations contributes to update propagation, thus to data consistency.

- There are different modelling approaches for MRDB (Multiple Representation Databases) that are usually used for linking different Levels of Detail.

- If required, correspondences between the databases can be established by various tools of data matching (data mining, ontologies and formal specifications) and transformations (conflation, generalisation, matching geometries)

It will be up to each TWG to decide if links between spatial objects at large scale and the spatial objects at smaller scales are required or not, of course, in the case of a theme including several levels of detail.

Example 1:

> In Germany in the AAA model, links between the representations on different scales/LoDs are not stored. Metadata with the feature about the LoDs to which it applies is stored, so any application can pick the right features for whatever it wants to do.

Example 2 : Ordnance Survey (UK)

> Generally and when practicable it is more efficient to capture spatial objects at the highest resolution. This improves reliability in comparison and analysis when combined with other data sets and strengthens data integrity. Such an approach also allows the creator to then publish the information at various resolutions.
> Moreover, the method of high resolution capture multiple resolution publication has been successfully employed by geography statisticians for a number of years now. With such an approach it is possible to aggregate statistical data into an area unit for reporting and this may use object referencing to reference the aggregated level unit to existing spatial objects at each level of aggregation.

Statistical Example:

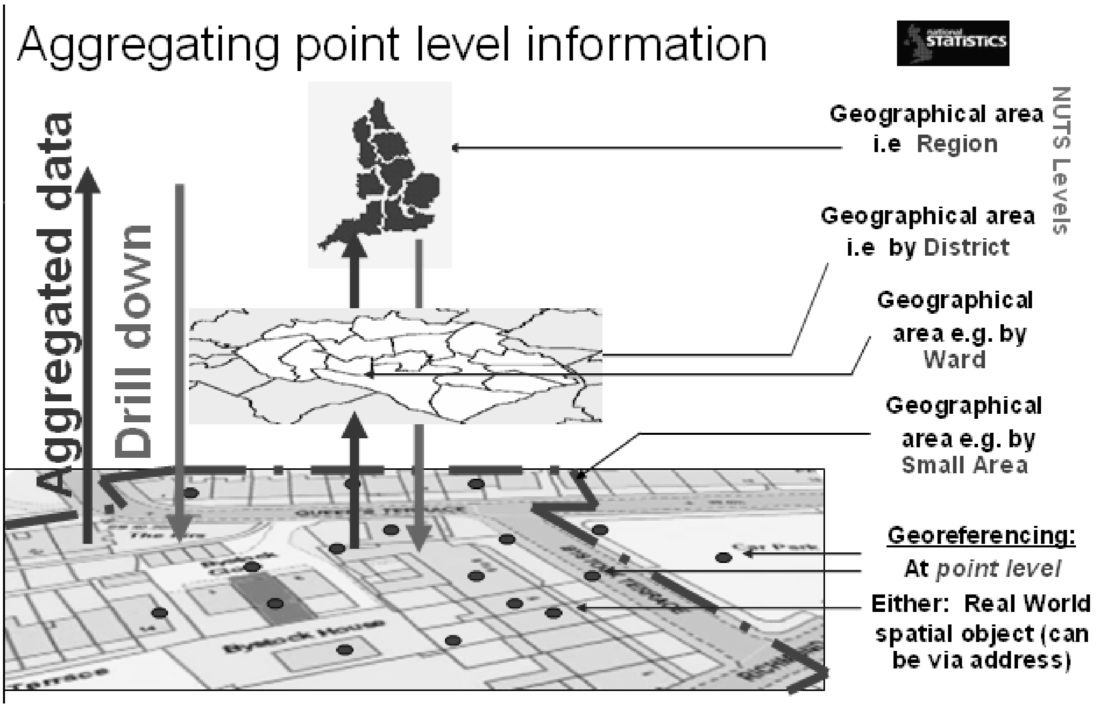


Figure 16 - Aggregated Spatial Objects for Statistical Reporting.

As is shown above the first level units can then be aggregated up to achieve this. The associated statistical (application) data is also aggregated up. This process can continue from local to regional to national and European levels and this is already reflected in the NUTS [Nomenclature of Units for Territorial Statistics] levels for statistics.

To ensure consistency the spatial objects and the statistical data are aggregated in harmony, to derive lower resolution reporting units by reusing existing spatial objects. Therefore the entire publication process becomes more efficient to manage in terms of aggregation and disaggregation as demanded by the user irrespective of the starting point being at the European level or the local level - as the diagram below demonstrates.

Such an approach can also be applied to other themes and spatial object types. For each theme there must be a defined use case to ensure that the approach is relevant to a particular theme. For example, the approach is well suited to a river or road network but there may be little immediate benefit in aggregating cadastral parcels.

A River Example: In the case below the user may wish to refer to an entire spatial object such as the River Seine (or equally the N121 motorway). The River Seine spatial object may be referenced by many sets of application data such as water quality. At the lowest resolution the report may simply be an average of water quality for the entire river. However a user may then progressively wish to enlarge the resolution and reveal more information – *consistently,* until the most detailed level is reached and all the application information is disaggregated. The unique identifiers of one complete spatial object (eg the entire River Seine) can reference several sub-objects (as in the statistical units example above, note that this is a case of object referencing) that make up a more detailed components of the spatial object of interest and so on.

Where real-time generalisation is not possible then the geometry is likely to be cached otherwise it may be derived as required. Linear networks lend themselves to this approach compared with surface objects. In all cases the application information is held separately and only once. It is cross referenced to the intermediate spatial object by unique identifier.

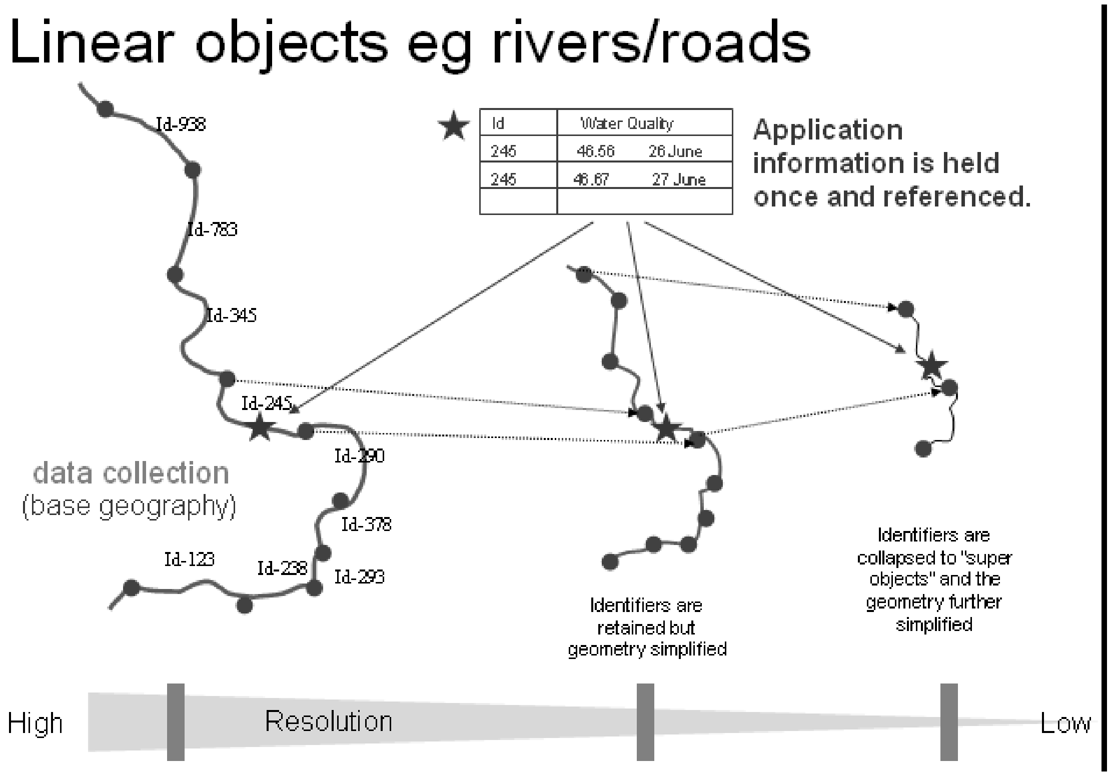


Figure 17

Constraints
Ordnance Survey applies the following constraints:

1. The aggregated units retain logical consistency throughout
2. A minimum number of levels of detail are employed.
3. Dynamic generalisation is preferred over cached derived geometry (so long as constraint 1 is adhered to).
4. The unique identifier references the aggregation of higher resolution spatial objects that it represents.
5. The aggregated units collectively maintain internal positional consistency and with spatial objects in other themes.
6. Cartographic portrayal is important as an aid to immediate user understanding

NOTE 1 Application data is not embedded in the spatial object but is referenced to each level.

A.20 Data capture

This component deals with:

- which spatial objects/attributes have to be captured (e.g. all the roads/only main roads)
- how spatial objects have to be represented (e.g. by points, lines, polygons).

How the spatial objects are captured (e.g. by stereo-plotting, by field survey, …) is out of the INSPIRE scope: this point does not have to be harmonised. Nevertheless, TWGs may ask data producers to give information on that topic in metadata, as lineage information, if they consider this information is necessary for users.

Example1:

> Different producers may have the same definition of "Forest" (groups of trees...), but they may capture the forest with very different selection rules (height of trees, minimum % of area covered by trees, minimum size of forests...)

It is a key component as selection rules will define density of information and so, they are the main way to define level of detail i.e. the scale.

For instance, a road database will mainly contain roads with their attributes; its feature catalogue will be more or less the same at scale 1: 10 000 and at scale 1: 5 000 000. The different levels of details will be defined mostly by the rules of selection and by the positional accuracy.

However, it will be difficult to harmonise the selection rules:
- if existing data has more data than what is needed, harmonisation will require generalisation : it may be done more or less automatically ; nevertheless, it is not costless
- if existing data has less data than what is needed, harmonisation will require new data capture (very costly).

Example (ERM) :

> “A very specific point of harmonisation has to do with the 'graphic' resolution of the vector data corresponding to the scale or resolution of EuroRegionalMap.
>
> Although this problem is not pragmatic by nature (there simply is no point in presenting vectors that are of a higher resolution than what the accuracy of the data supports), the 'graphic' resolution with which the various national databases were initially digitised (and for which there often were reasons that have to do with cartographic purposes) were not always consistent with the objective accuracy of the data, and these variations of 'graphic' resolution between the different national databases are indeed transferred to EuroRegionalMap. It would of course have been possible to use processing power to 'downgrade' exaggerated 'graphic' resolution of vector data wherever necessary (although 'downgrading' could be trickier than it appears on first sight due to topological constraints), but in many instances this would have conflicted with another guiding principle of EuroRegionalMap, namely that EuroRegionalMap data should never become 'disconnected' from the national databases: if we want EuroRegionalMap to be the result of nationally maintained databases converging towards a common European model, but NMCAs have strong reasons to keep the 'graphic' resolution of their national databases for other reasons than EuroRegionalMap and therefore refuse to converge towards a common 'graphic' resolution, than the point of convergence has been reached and some heterogeneity must be tolerated as a price to be paid for ease of maintenance.”

Recommendation 28
- if user requirements exist, put them in the specifications.
- if no user requirement exist (or if they are weak, not well known), give selection rules in a very general way (e.g. by the scale of representation).
- anyway, there will be deviations from the common model ; they will have to be registered in metadata.

Example 1: ERM

> EuroRegionalMap data are collected at a density of detail that approximates the medium scale product range (from 1:200 000 to 1:300 000). Portrayal criteria mentioned in the data dictionary are general guidelines. It is up to producers to settle in detail its own portrayal criteria.

Selection rules have to be given in a harmonised data product specification to give a general idea of the density of information expected but often, in existing data, these rules may be different because adapted to the size of the country. It is one of the reasons to accept deviations from the general recommendation.

For instance, the minimum size of a polygon in Corine specification is 25 ha, which is not adapted to small countries, such as Malta (whose total area is 31 500 ha). An opposite case is given in the following example.

Example 2: RISE

Requirements from WFD: The main factor determining the necessary spatial detail of data gathering under the WFD is the size of the smallest spatial object to be shown on the maps. In the WFD the only direct indication in this context are the size thresholds given for the typology according to system A (WFD Annex II). These thresholds are set to a 0.5 km^2 surface for lakes and to a 10 km^2 catchment area for rivers. Although these thresholds do not imply that all water bodies larger than these numbers need to be reported, these figures can be used to estimate the required detail of data gathering or the input scale.

Existing data: However this requirement is not fulfilled by Sweden which applies the following rule: “Catchment data size present in Swedish coverage is > 40 km^2, lakes > 1 km^2 “ because of the great number of lakes in the territory.

Recommendation 29
The way spatial objects are going to be represented in the database has also to be specified (e.g. by points, lines, polygons).

Example 3: roundabout in EuroRoads (road network)

A roundabout should be represented by its true geometry if the diameter is more than X meter. The attribute value *Roundabout* should be stored in the included links.

If the diameter is less than X meter, the roundabout should be generalised and represented as a node. The attribute value *Roundabout* should be stored on the node.

The diameter X shall be stored in metadata. EuroRoadS specification recommends a diameter of 20 metres for the centre line.

If a roundabout is stored at an aggregated level (always represented as a node), this information shall be stored in metadata.

The feature Roundabout should be stored as the attribute value Roundabout in the attribute Form of node or Form of way depending on the generalisation rules.

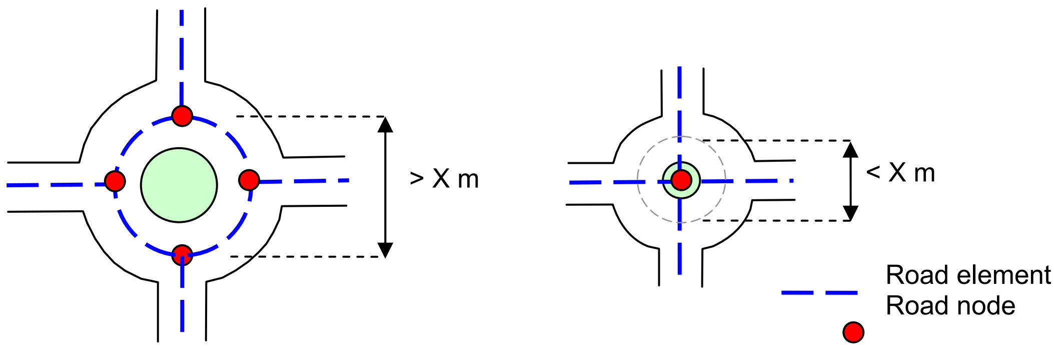


Figure 18 – Example of a roundabout with the limit of 25 meters for the outer diameter, as a level for generalisation.

Example 4: ERM

Watercourse BH502
Definition: A natural or man-made flowing watercourse or stream.
Feature class: WATRCRSA
Geometry type: Area
Primitive type: Face
Portrayal criteria: Watercourses that form up a logical water network at a map scale 1:250 000.
Watercourse with width ≥ 125 m.

Watercourse BH502
Feature class: WATRCRSL
Geometry type: Line
Portrayal criteria: Watercourses that form up a logical water network at a map scale 1:250 000.
Watercourse with width < 125 m.

A.21 Conformance

Recommendation 30
Each theme specification should define clearly what is mandatory or conditional and what is only recommended or even just optional.

NOTE Conformance classes and conformance checks are described in clause 25 of the Generic Conceptual Schema and clause 7 of this document.

Annex B
(informative)

Management of connections at international boundaries

B.1 Introduction

Requirements about connections between spatial objects at international boundaries (i.e. edge-matching) come from the Directive, Article 10 (2): in order to ensure that spatial data relating to a spatial object the location of which spans the frontier between two Member States are coherent, Member States shall, where appropriate, decide by mutual consent on the depiction and position of such common spatial objects.

The challenge is to match two data sets on the same theme; the two data sets have already been transformed in the common application schemas, in the same CRS (generally ETRS89 for planimetry and EVRF2000 for altimetry, if any), their levels of detail are the same (or almost the same) and they come from two neighbour countries.

Two issues can be distinguished: a) the issue of representing national boundaries and b) the issue of edge-matching national data. It would be possible to edge-match national data (except administrative units, cadastral parcels and statistical units) even though there is not always a precise and agreed representation of the boundary. However, matching precisely at the national boundary is the ideal solution.

NOTE 1: There might be also edge-matching issues within countries, e.g. in federally organised countries. This point is not specifically considered in this chapter.
Generally, it is assumed that the country has already some organisation to solve the issue (and so, may provide interesting feed-back to INSPIRE from their experience on this topic). If not, most of the recommendations given in this chapter may also apply for edge-matching within a Member State.

NOTE 2: This chapter provides some recommendations for edge-matching and gives examples, generally from EuroGeographics experience. The numeric values given as thresholds in the example apply only for the given example. Furthermore, these numeric values are often used only for automatic processing of edge-matching.

It will be to each “Thematic Working Group” to define the appropriate thresholds, if required, in a given data product specification, for each case of edge-matching.

B.2 Which spatial objects have to be matched?

The issue is not exactly the same for vector data and for coverages.

B.2.1 Vector data

Recommendation 31
Spatial objects have to be matched if they represent the same real-world phenomena along or across the national boundary.

- they must have the same spatial object type and the same geometric primitive (e.g. lines, polygons, …).
- they must be close to each other, geometrically and/or semantically.
    - o for geometry, the condition may be given by a threshold which will be function of the database accuracy
    - o for semantics, spatial objects having the same code or same name (if the 2 neighbour countries have the same language) or corresponding geographical names (i.e the endonym in one database being an exonym in the neighbour database, if the 2

neighbour countries do not have the same language) : they may be considered as representing the same real-world phenomena

NOTE 1 the use of distance threshold is convenient to enable some automation of the process. Nevertheless, the main point is that in the real world the two spatial object lines (or areas) are really connecting (or they are not) or that the two point features are really the same

Each "Thematic Working Group" will have to specify more in detail what is meant by "being close to each other".
The main points to be decided:
- the threshold to be applied, at least to enable some automatic procedure
- the attributes to be taken into account.

NOTE 2 as general rule, the threshold has to be defined by the "Thematic Working Group" taking in account the accuracy of the required database and the source databases. Nevertheless, some adaptation may be required, when matching two neighbour countries, taking in account the accuracy of existing data on each side and (at least for large scale data), the alterations implied by the CRS transformation.

<u>Example 1 for point features (EGM, accuracy ≈ 1000 m):</u>

If the same point or node feature type is portrayed in adjacent layers more than once, only one of these features is left to the final, combined database. For automatic edge-matching, points and nodes are considered to be the same feature if they have the same feature type and if the distance between features is less than (about) 250 meters. However, in some special case it's obvious that the two features are the same even if the distance is more than 250 meters.

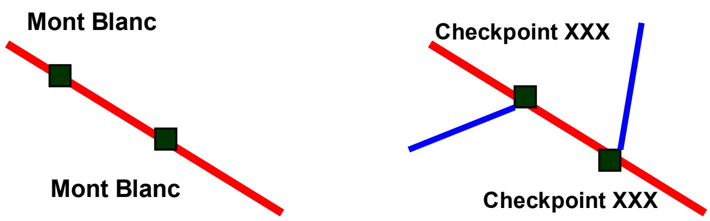


Figure 19 – Same point/node feature duplicated in layers ELEVATION POINT and TRANSPORT

<u>Example 2 for line features (EGM, accuracy ≈ 1000 m):</u>

Two line features will only be moved automatically to match each other if they are of the same feature type and if the gap is no greater than approximately 1000 meters. The gap between two line features along the edge matching boundary is the shortest distance between the edges (not necessarily between nodes) composing the line features.

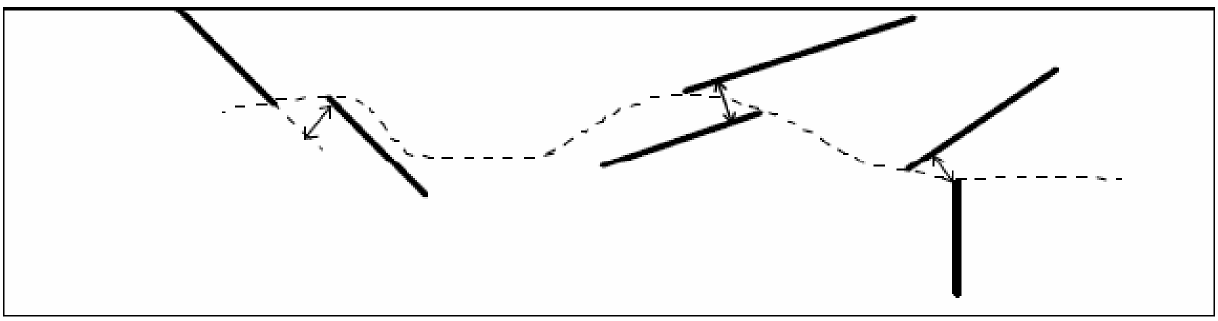


Figure 20 – Measuring gap between lines

Rules for edge-matching are less obvious for the area features, because quite depending on the selection criteria and generalisation criteria used to portray the area features.

<u>Example 3 (ERM, accuracy ≈ 125 m, scale range of 1: 250 000) for area features:</u>

Edge-matching can only be performed at a distance not exceeding the geometrical resolution accuracy of the data set, and the minimum area size resolution.

The matching distance between borderlines of area feature is a maximum of 125m (accuracy resolution). If bigger distance is noticed, it is not needed to do the edge-matching. An exception is done for the water bodies and islands because it is a tremendous condition to get a continuous water network.

When the extended part to be added to the spatial object is less than 20 ha (for area feature: minimum area size resolution) this can be ignored and not added.

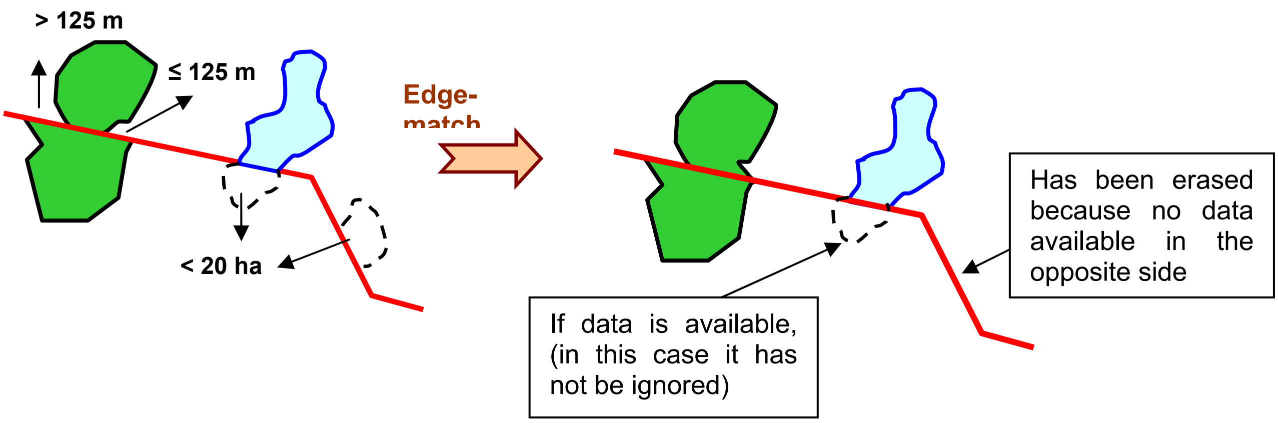


Figure 21

Of course, this geometrical resolution accuracy and minimum area size values will vary according to the scale range of Level of Detail.

When possible, trans-border spatial objects should be treated as one spatial object at least when it comes to quantitative measurements.

<u>Example:</u> There might be rules (minimum surface) for including a lake in a map/database. Applying these rules only on national parts might lead to erroneous results.

Recommendation 32
When a line segment (for example a watercourse) is consistent with the national boundary, the segment should have exactly the same geometry in both national components.

In other words, the geometry of the linear feature should fit the geometry of the national boundary, knowing that this national boundary has been preliminary portrayed in a unique manner in both countries**.**

The fact that the watercourse is consistent with the national boundary should be proofed by the boundary treaty.

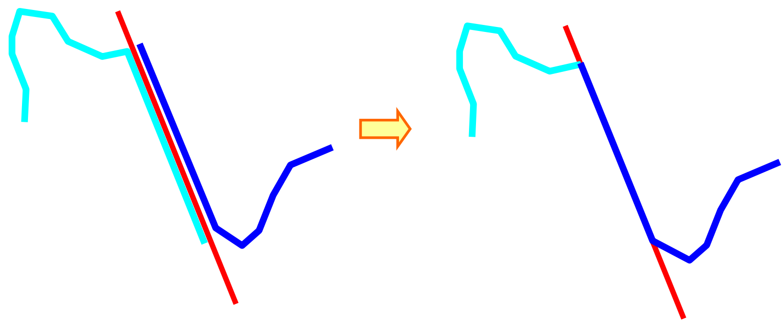


Figure 22

When real world objects are effectively continuous in cross-border areas, selection and representation criteria should be consistent in the different data sets of similar LoDs (same spatial object type and geometric primitive).

This is mainly important for linear features in order to assure a continuous network all over national boundaries.

When discrepancy appears in the selection criteria of linear features, it may be also necessary to do some edge-matching to ensure network continuity. It may imply to add or to delete some small size spatial objects.

Example ERM and EGM:

> When a spatial object is obviously stopped at the national boundary with no counterpart at the other side, the decision whether to erase it or to extend it should be decided in common agreement between neighbour countries.

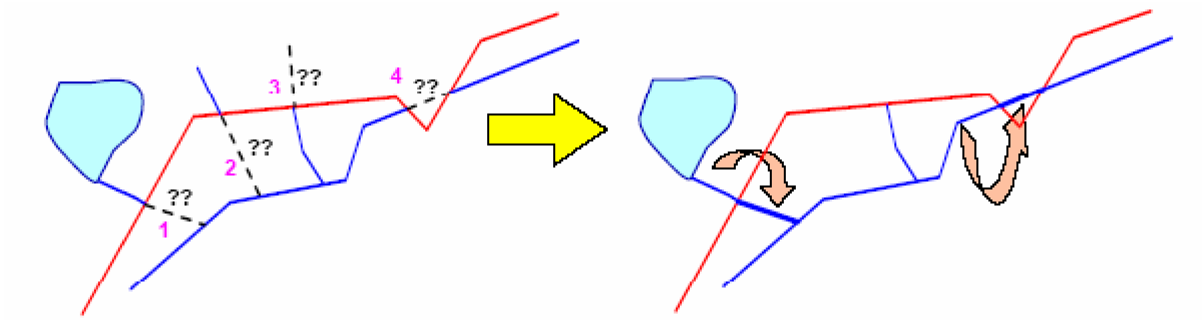


Figure 23 – Two new short lines (1, 4) added. Lines 2 and 3 are too long or too uncertain. Existing line above the missing line number 2 removed. New spatial objects get attributes from those lines they are joined to (existing dangling node).

The threshold to authorise adding or deleting spatial objects will have to be defined for each product (at least, for each theme and for each scale range) by each “Thematic Working Group”.

B.2.2 Coverage data

At least, elevation and orthoimages (and may be, other INSPIRE themes) are concerned by this type of data.

Let us assume that the coverages in two neighbour Member States are already harmonised (same grid type, same grid size, same CRS, same function).

The first point to be decided is how to make the mosaic between the two countries:

- should we cut the data set exactly at the national boundary?
- is it enough to adopt some rectangular “cutting” along the national boundary (e.g. keep tiles of 1 km^2 or 100 km^2).

The second issue is the edge-matching of the function values along the boundary: probably, some inconsistencies will appear when gathering data. For instance, for DEM, there may be artificial 'cliffs' and 'steps' at the edges of the national contributions.

Each “Thematic Working Group” will have to decide if it is necessary or not to do some smoothing at national boundaries.

Example 1: about DEM (BKG):

> Experience at BKG has shown that the most effective method for edge-matching between DEM is to re-calculate the border area from the original measurements, putting together the information from both sides of the national border. Only in cases where the original measurements are not available, a blending algorithm may be applied on the edge between the interpolated DEM files.
>
> In case two DEMs overlap, a blending method has been proposed by Norbert Pfeifer (TU Vienna) and provides good results for DEM with different accuracy (see his powerpoint slides from the joint EuroGeographics / EuroSDR workshop on "NMCAs and GMES", 5 - 6 September 2006, at the BKG in Frankfurt am Main: http://www.eurogeographics.org/eng/documents/08_Pfeifer_DTMquality.pdf)
>
> Generally, edge-matching has to be done by agreement between two neighbour countries but, in some cases, it may be necessary to have it done at centralised level.

Example 2: radar rainfall

> Radar edge matching occurs on the overlap between circular radar images. The edge matching is dependent on different radar properties and even on the specific weather. Here data exchange mechanisms happen as frequently as every 2 minutes, while the European Composite Radar is delivered every 30 minutes. This has to be a centralised service due to the data exchange overload: the merged data is centrally provided, even if it is supplied through different routes by individual EUMETNET members.
> In fact this is a specialised procedure and so has to be done once, and by a central system, not only because of the time constraints. In a way, this is a generalisation of the borders issue. Here it isn't one data provider having to match between another, but all data providers submitting to standards and having the data centrally adapted.

B.3 How to match spatial objects?

It is still an open issue whether if it is necessary to undertake geometric edge-matching for each pair of corresponding spatial objects or if it is enough to establish the link between the two spatial objects (e.g. by using their identifiers).

- The first solution will fulfil more requirements, as it ensures continuity across the boundary, both for computers and for human beings (visualisation). However, it may lead to significant deformations, especially if the two neighbour data sets are not at the same level of detail (and so, not at the same accuracy). This can normally be accommodated where both parties agree the need to resolve both the match and the geometric fidelity, there are examples where topographic objects are matched seamlessly. The advantage is that once this is achieved it is resolved for the remainder of time and is reusable.

- The second solution may be enough e.g. for itinerary computation (it is more or less the solution chosen by EuroRoadS, with its border nodes, each node "knowing" the identifier of the border node on the other side). It ensures continuity only for computers, not for the human beings.

Connecting (or border) nodes or edges are required with the second solution, they may be useful also even with the first solution.

B.3.1 Geometry edge-matching:

Recommendation 33
To achieve geometric edge-matching, spatial objects will be moved according to the following rules:
- Both spatial objects will be moved to reach the middle position if the two sources used for capturing the data are of the same quality level.
- If one of the two sources used is of higher quality, the positions of spatial objects captured from this source will be preserved and the other spatial objects will be moved toward the best source.

NOTE : "higher quality source" has to be understood as the most reliable data source ; it generally and mainly involves geometric accuracy and data currency.

When displacing the spatial objects, we must take care that the relative position of the other spatial objects will be preserved (e.g. a river located on one side of a road should remain on the same side of the road after edge matching).

<u>Rules for points:</u>
The only points to be matched are the points located on the boundary. A common position has to be agreed and this common position must be on the agreed national boundary.

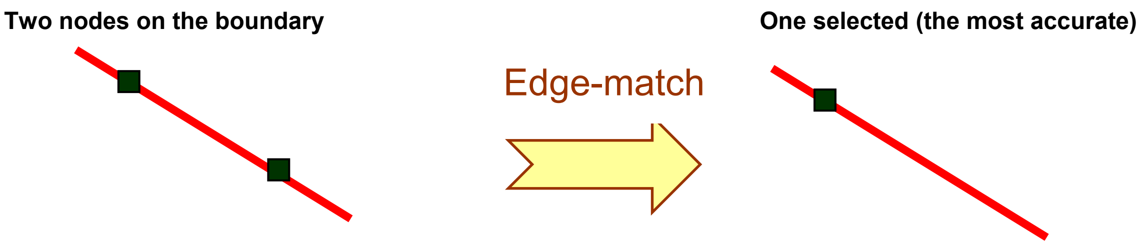


Figure 24 – The matching location may be a choice between the two locations.

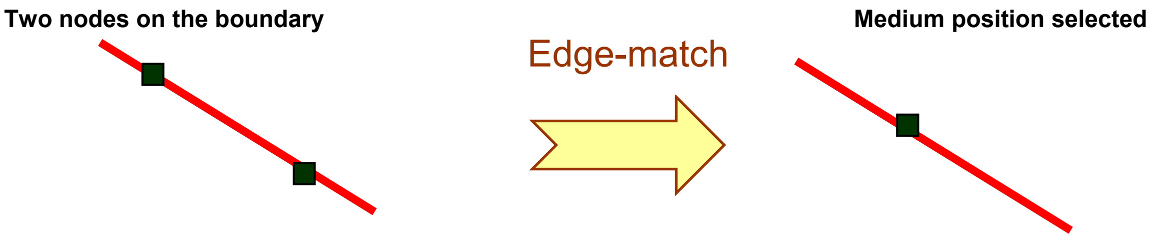


Figure 25 – The matching location may be a medium position between the two locations (the middle between the two locations).

<u>Rules for lines crossing the national boundary:</u>

The edge-matching must be done:
- respecting the characteristics of the spatial object.
- as near as possible to the national boundary.

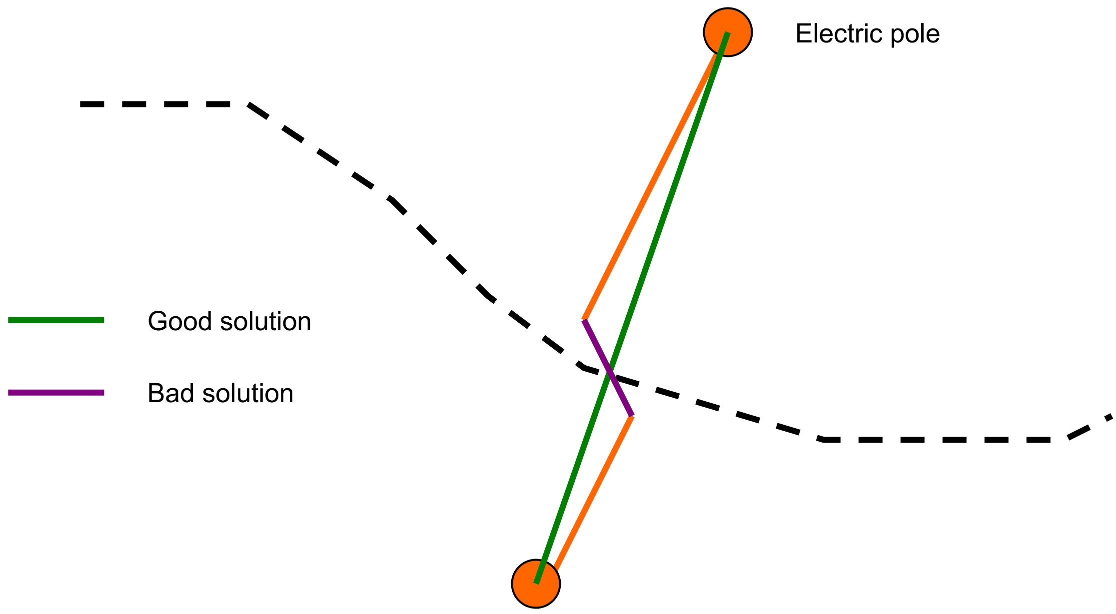


Figure 26 – Electric lines are straight lines between two poles.

In other cases, it may be quite better to have some elastic deformation. Specific geometric methods can be used to recognise homologous points and to stretch the spatial objects geometry one to another. Rubber sheeting is a famous method to stretch data one to another (see for example Laurini 1996; Haunert 2005).

The stretching also raises the problem of geometrical deformations. Theoretically the geometrical anamorphous can be applied on both side or on one side only (see following figure). Rules will be required, e.g. applying the anamorphous on less accurate data.

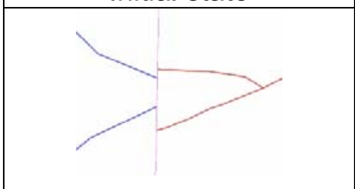
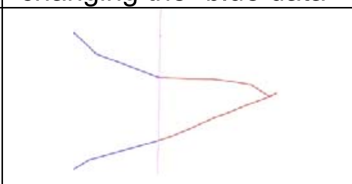
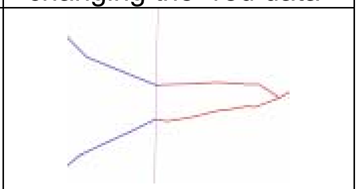
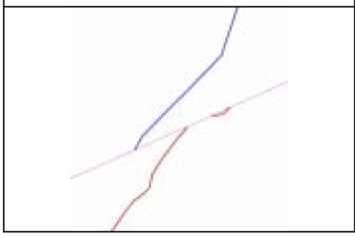
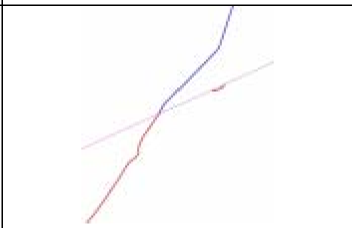
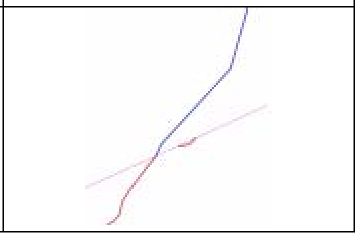
Initial state	changing the “blue data”	changing the “red data”
		
		

Figure 27 – Stretching the data along a boundary (from Lamine and Mustière 2005)

Example EGM (accuracy around 1000 m):

Modifying lines to facilitate a smooth edge match line: Lines composing the line feature will be modified to the next node, or (maximum) up to 5 kilometres away from the boundary.

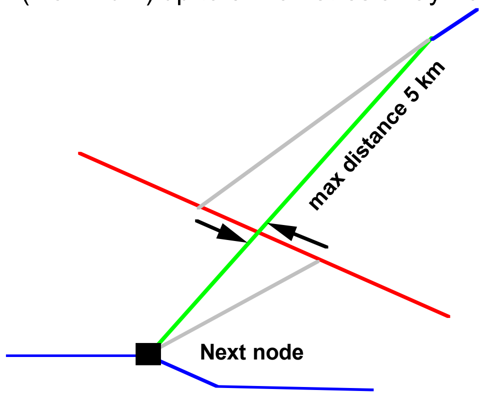


Figure 28

For some themes, it may be useful to consider more complex cases:

Example ERM (accuracy around 125 m):

When one spatial object on one side of the boundary corresponds to two spatial objects of the same type on the other side of the boundary, and if the gap is no greater than approximately 125 meters, the spatial objects will be moved equal distance to form a fork.

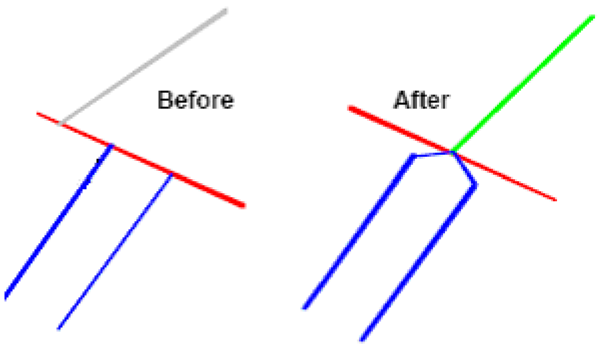


Figure 29

In ERM, this situation may happen for parallel powerlines. The situation might also happen if two watercourses or roads meet near the national border and continue as one beyond the national border. At large scale, it may happen for dual carriageway roads.

NOTE: for some INSPIRE themes, perfect edge-matching at international boundaries won't be always possible (and even may be, won't be required)

Example (Sea Regions)

Another example which may be informative is the rather esoteric task of defining Sea Regions – which INSPIRE has chosen as a distinct theme. The responsibility belongs to IHO in Monaco, but the documentation is only in paper form. The difficulties arise because in the normal case there is little requirement for high precision. Sea regions are often locally agreed between nations rather than internationally, but imprecisions arise because the points are defined usually between land promontories and islands. Promontories erode, and the sea region border may go through islands or skirt the coasts without any consistency.

B.3.2 Creation of connecting spatial objects:

The use of connecting points (or lines) will have to be decided by each “Thematic Working Group”, according to the characteristics of the spatial objects (e.g. it is of no use for administrative units or parcels (as they do not cross the boundary) and according to the users requirements.

The main issue will be to decide how to manage these connecting nodes:

- should they be in the initial databases? or only in the INSPIRE data set?
- when matching two spatial objects, should we have:
    - 2 connecting points or only one connecting point?
    - if 2 connecting points, should they have the same location or are different locations allowed?

If it is decided that connecting nodes should be on the same location, this location should be on the agreed national boundary.

However, the technical implementation of this principle seems to have raised problems, e.g. the management of coordinates and topology is different depending on different GIS.

Connecting nodes must ensure the link between the two spatial objects to be connected, either by the same location, or by a relationship between the two connecting nodes.

<u>Example 1: EuroRoadS</u>

> If a road continues in reality after the end of the data set border, it has to be registered as an attribute to the node. If the corresponding node identification number is known it shall be stored.

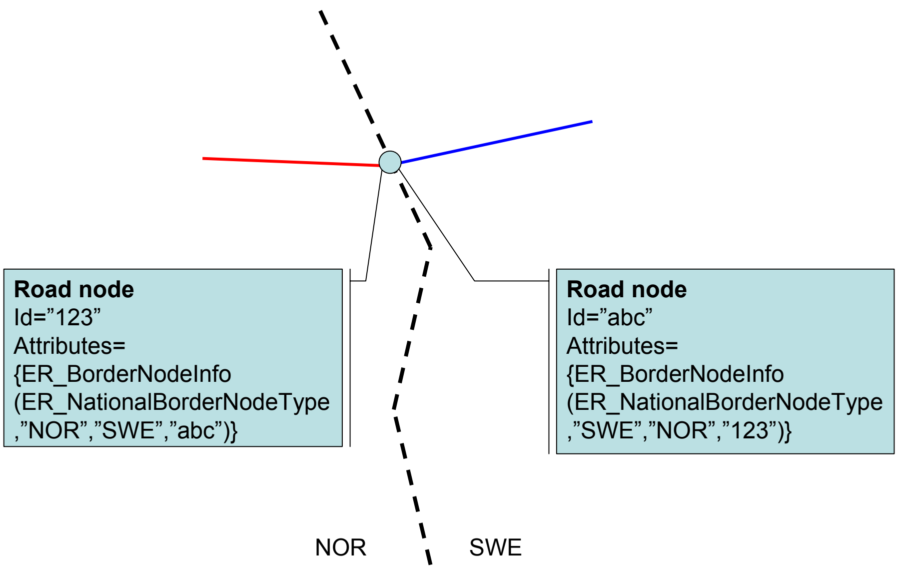


Figure 30

B.4 How to match attributes?

There are two cases:

- When a spatial object lies on the national boundary and is part of the two countries. This occurs mainly with line and point features
- When a spatial object over crosses the national boundary. This occurs mainly with line and area features

B.4.1 Matching rule for duplicated spatial objects located on the national boundary

The general rule is that all the matched attributes must have the same value.

- The unknown and unpopulated values must be replaced by a populated one if available.
- When no real information is provided at both sides, then we can give a priority in the null values

However, there may be exceptions for geographical names (according to the model chosen for mutilingual texts) and for some national codes:

- in France, “Rhin” will be the national name and “Rhein” may be included as “exonym”
- in Germany, “Rhein” will be the national name and “Rhin” may be included as “exonym”

B.4.2 Matching rule for spatial objects crossing the national boundary

Some attribute values need to be consistent at both side of the boundary. This is mainly applied to line and area features.

For each theme, define which attributes must have the same value (or corresponding values) and which attributes may have independent values.

<u>Example (ERM):</u>

The attributes, which need to be logically matched, are highlighted in bold. Matched value:

- Y means that the spatial object should be logically matched at both side of the boundary
- N means that inconsistency in attribute values is allowed.

Table 1: Extract of matching attribution for cross-border spatial objects

HYDRO	Area/ Line	BH502	Watercourse	
	Area/ Line	**EXS**	**Existence Category**	**Y**
	Area/ Line	**HOC**	**Hydrographical Origin Category**	**Y**
	Area/ Line	**HYC**	**Hydrological Category**	**Y**
	Line	LOC	Location Category	N
	Area/ Line	NHI	National Hydrological Identification Code	N
	Area/ Line	NAMN1	Name in first national language	N
	Area/ Line	NAMN2	Name in second national language	N

B.5 Which national boundaries?

Another issue of edge-matching at national boundaries is to know where the national boundary is, i.e. to agree on a common representation of the boundary to be used.

The national boundaries are generally defined in bilateral treaties between the neighbouring countries, but it does not mean that an agreed geographical representation of the boundary is available.

Recommendation 34
In order to avoid manifold work, it is recommended to use already agreed national boundaries, when and where they exist. Generally, for one level of detail, there should be only one representation of the national boundary; this common representation for a special scale can be created by the mapping authorities of two neighbour countries.

Some work on this topic has been already done by EuroGeographics and a set of agreed boundaries is more or less available:

- the EuroBoundaries project aims to provide legally agreed national boundaries at the highest available accuracy; however, as it involves legal procedure, the collection and preparation of these boundaries of the whole Europe will be quite long (about 10 years) and until its full achievement, the EuroBoundaries data set is planned to contain :
  - one common representation based on the legally agreed national boundaries at high accuracy for some parts of the boundaries
  - two different national representations of some other parts if there are such
  - one provisional line in cases the two others are not available on the basis of best possible information as temporary limited representation.
- EuroRegionalMap is a topographic pan-European database at scale 1 : 250 000 ; it includes 30 countries (including all countries of the European Community, except Bulgaria). The administrative theme contains agreed representations of national boundaries; however, the agreement is not a legal one but just an agreement between NMCAs. In disputed areas, the two national representations are represented by ERM. Consistency between national boundary and topographic objects included in ERM (roads, rivers, …) is ensured.
- EuroGlobalMap is a topographic pan-European database at scale 1:1 000 000 ; it includes 36 countries (incl. all countries in the European Community). As for ERM, the administrative theme contains agreed representations of national boundaries; however, the agreement is not a legal one but just an agreement between NMCAs. In disputed areas, the two national representations are represented by EGM. Consistency between national boundary and topographic objects (roads, rivers, …) included in EGM is ensured.
- EuroBoundaryMap is an administrative pan-european database; data is represented at two resolutions 30 m and 200 m.; it includes 36 countries (incl. all countries of the European Community). The representations of national boundaries in EuroBoundaryMap are not legally agreed but proposed by the data set coordinator and approved by NMCAs.

Recommendation 35
Boundaries in dispute for geopolitical reason can be maintained separately. Boundaries in dispute owing to technical or administrative reasons cannot be considered as such and should be solved.

In case, there is not yet an agreed boundary, at the required Level of Detail [LoD] between two neighbouring countries, criteria for a suitable fixed boundary may be the following:

- The first criterion to resolve a boundary is to agree on the geometric resolution accuracy and a degree of generalisation suitable for the different scales or LoDs involved.

The degree of generalisation focuses on the density of vertices. Geometric data resolution in the density of vertices on an edge should be as low as possible keeping a realistic size and shape of the spatial object.

When two boundaries are presented having similar resolution, the preference should be given to a given boundary with the best positional accuracy and degree of generalisation.

- The second criterion is to keep the full consistency of the national boundary with the topography or at least the relative topological relation of the national boundary with the topographic objects, even to the detriment of its absolute positional accuracy. For medium and small scale data, this means that the relative position of the boundary with the topographic situation should prevail on its absolute position .

- When the national boundary is determined by real-world topographical objects (like a mountain range or a river), the geometry of the boundary should exactly fit the geometry of the topographical object. When practicable, neighbouring countries should agree on common representation of those topographical objects coincident or coterminous with the national border line. Consistency of the national boundary with the water network has the highest priority.

- The referring coordinate system of the agreed geometrical location of the boundary vertices should be the European reference system (ETRS 89)

- All decisions concerning the representation of national boundaries should be based on traceable arguments (like boundary treaties and supporting documents as orthoimages, maps).

Furthermore, the EuroBoundaries project may offer the contact person responsible for the country's national boundary.

Recommendation 36
As soon as fixed, those agreed international boundaries should be stored and structured in a common data schema. This international boundaries data set would serve as reference data for sustainable maintenance at national level and could be stored at European level.

B.6 Organisational point of view

Member States must agree on shared edge-matching responsibilities. Generally, it does not mean that they must agree individually for each pair of corresponding spatial objects; they may agree on some automatic method (based on the above recommendations) at least to detect which spatial objects may be matched and to match spatial objects when the distance between these spatial objects is considered as "small". Nevertheless, some discussion may be required in more complicated cases. Edge-matching in INSPIRE should be done, using as much as possible automatic procedures but there might be cases that require manual interventions. In which case the economics of "resolve once – use many times" will be more important from a cost benefit perspective.

NOTE: the mutual agreement for each pair of corresponding spatial objects has to be done for EuroBoundaries, as this project is expected to supply reference international boundaries for INSPIRE.

It is likely that during the matching phase, some inconsistencies will be shown. Data providers should be encouraged to correct these inconsistencies in the source/original data sets as soon as possible.

The edge-matching has to be considered not only once but also at each update of one of the neighbour databases. If updates are frequent, use of automatic methods to detect and correct inconsistencies seems the only solution.

Annex C
(informative)

Data specification document template and example

The Data Specification template and the example will be be moved to a separate document, published on the INSPIRE web site (http://www.ec-gis.org/inspire/ds/).

Annex D
(informative)

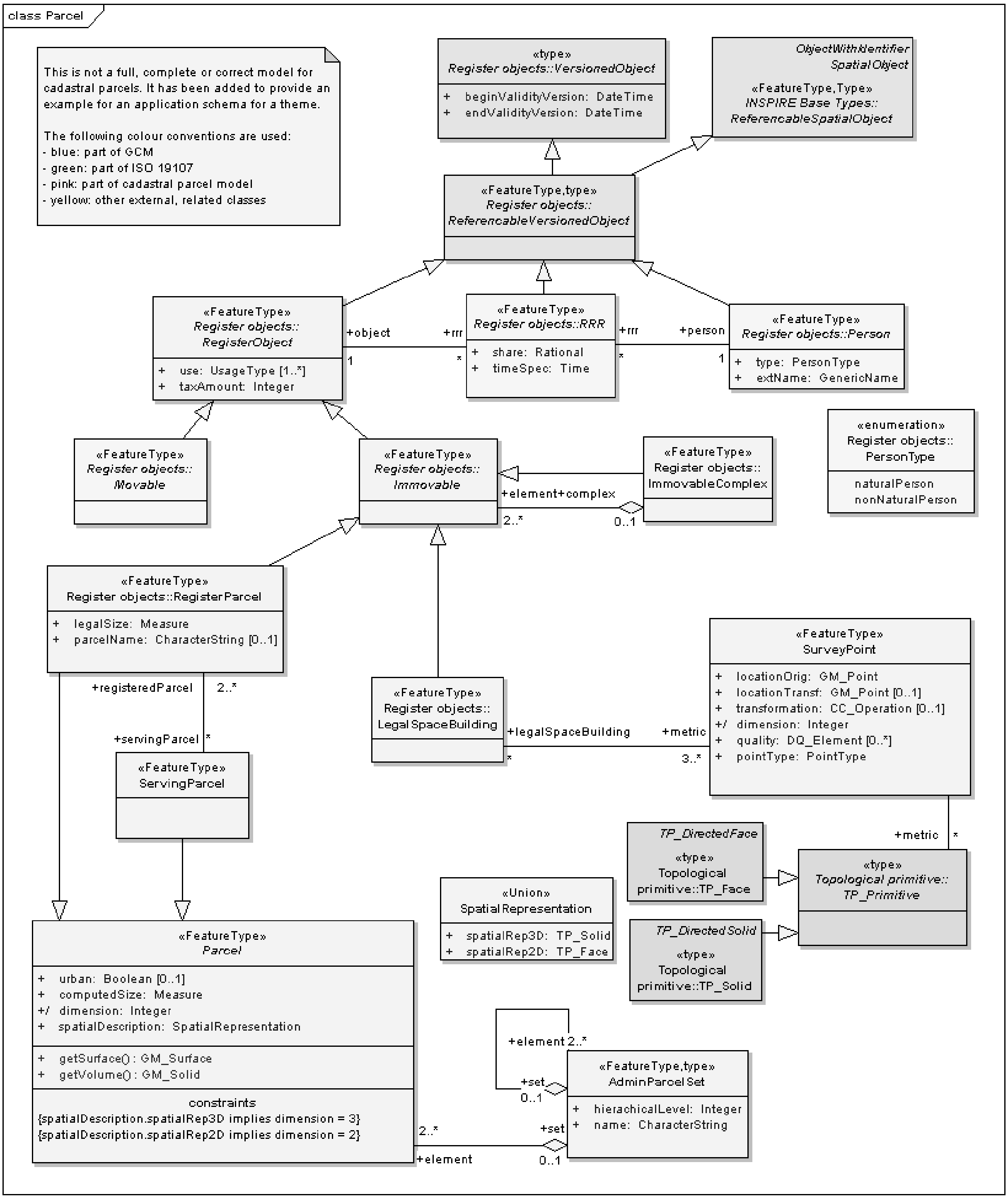
Example application schemas in UML

D.1 General remarks

For illustration purposes, this annex contains draft application schemas for aspects of the themes "cadastral parcels" (D.2), "elevation" (D.3), "meteorology" (D.4), and "geology" (D.5). These build upon the base schemas of the ISO 19100 series and the proposed Generic Conceptual Model. **These application schemas are in no way intended as a proposal for the future application schema for these themes nor does the Drafting Team endorse the content of the schemas.** The schemas have been included to provide an example for application schemas and relevant modelling patterns.

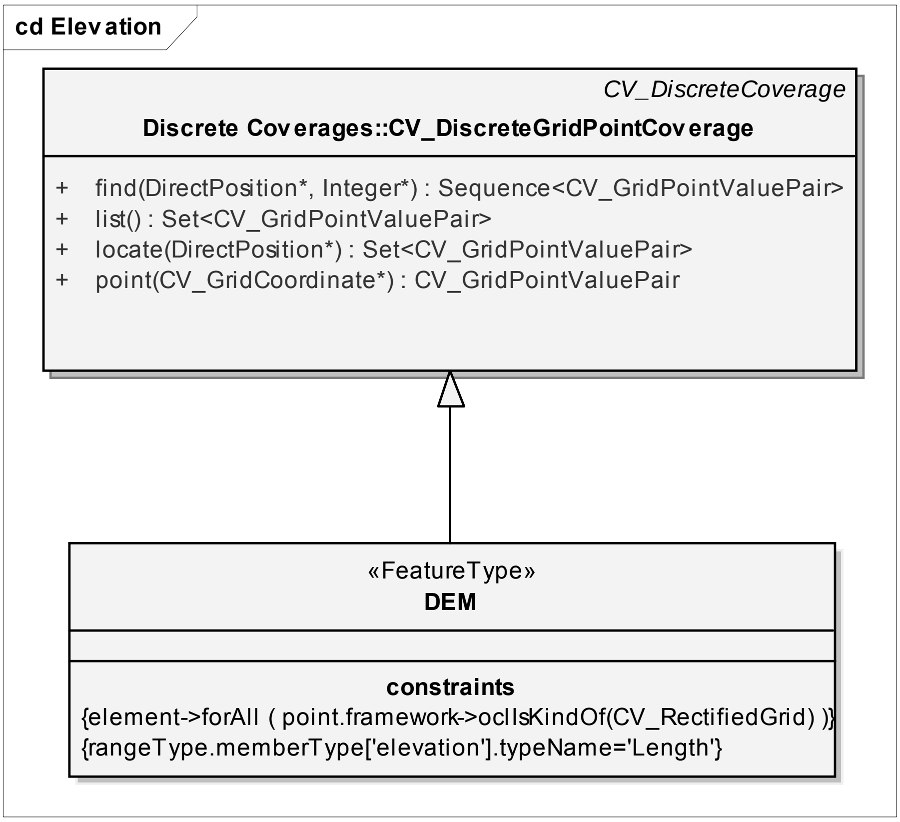
At this point, the examples are incomplete. See in particular the note in the package diagrams. Furthermore, no documentation for the schemas is included at this stage.

D.2 Application schema example: "Cadastral parcels"



.

D.3 Application schema example: "Elevation"



D.4 Application schema example: "Meteorology"

The UML diagrams in this subclause represent simple models of various meteorological data classes: surface and vertical sounding observations, and gridded analysis fields. These modelled information classes are consistent with the scoping of the Meteorology theme in document D2.3.

The models are based on information elements contained in the WMO exchange formats BUFR and GRIB. Furthermore, the models are factored to conform to the OGC Observations and Measurements model (OGC 07-022r1 “Observations and Measurements – Part 1 – Observation schema”, and OGC 07-002r3 “Observations and Measurements – Part 2 – Sampling Features”).

D.4.1 Synoptic Observations

Meteorological synoptic observations are real-time meteorological and atmospheric observations made from a variety of platforms – fixed stations on the earth’s surface, moving ships, ascending radiosondes, etc. Modelled here are the two broad classes of fixed surface and radiosonde observations.

Observable and utility types

A number of utility types are defined to characterise the meteorological properties observed at a fixed weather station – pressure, clouds, and current weather.

As well, various codelists and enumerations are defined for identifying dominant cloud and weather states, and the type of observing station and radiosonde. These correspond more or less directly to WMO code tables associated to the GRIB and BUFR specifications.

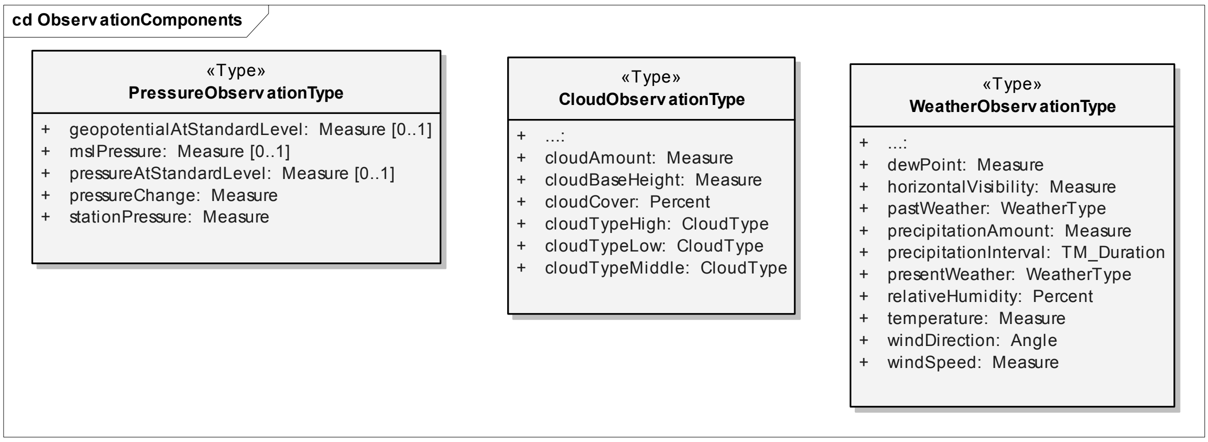


Figure 31

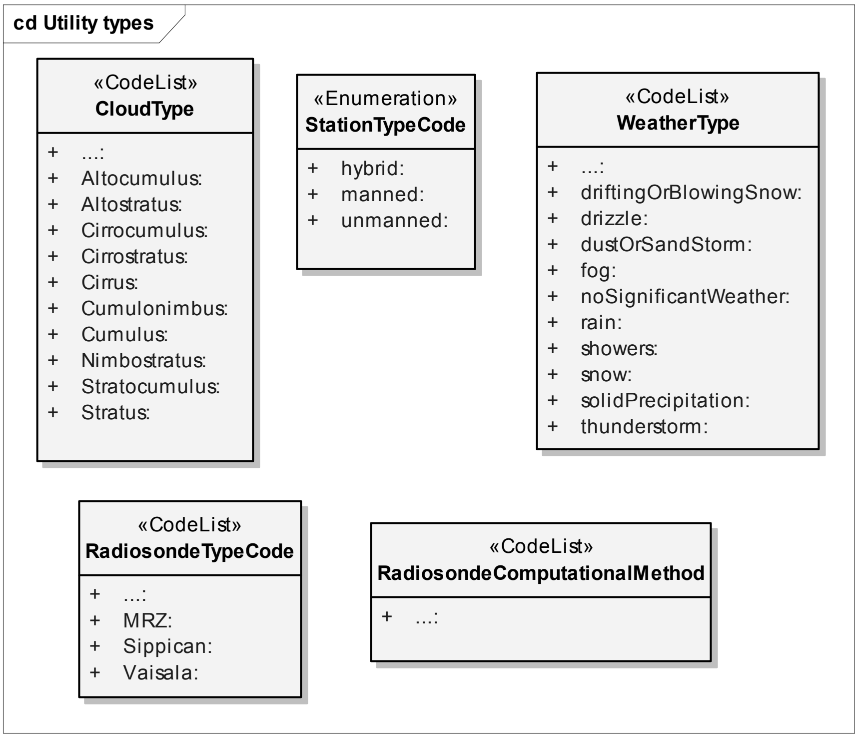


Figure 32

SoundingObservations

Some types of sounding observations are made through the release and ascent of a weather balloon carrying a radiosonde. These devices are one of the major sources of in-situ synoptic observation made globally for meteorological analysis and forecast. They measure primarily pressure, temperature, wind and humidity vertically through the atmosphere. Other types of soundings are made through satellites (e.g. ATOVS), lidars, sodars, drop sondes, etc.

The OGC Observations and Measurements pattern has been applied here: an *observation* (SondeObservation) is made on a *feature-of-interest* (MeteorologicalSounding) using a specified *procedure* (Radiosonde). The *observed property* is a composite (MetSoundingParameters) that defines the type of the observation *result*. The result itself is a discrete coverage over a vertical domain. The ‘feature-of-interest’ (MeteorologicalSounding) is a so-called ‘Sampling Feature’ (OGC 07-002r3) that samples along a particular path the meteorological state of the atmosphere (here represented through the 'sampledFeature' association to a canonical spatial object, AtmosphereMeteorologicalConditions, representing the complete state).

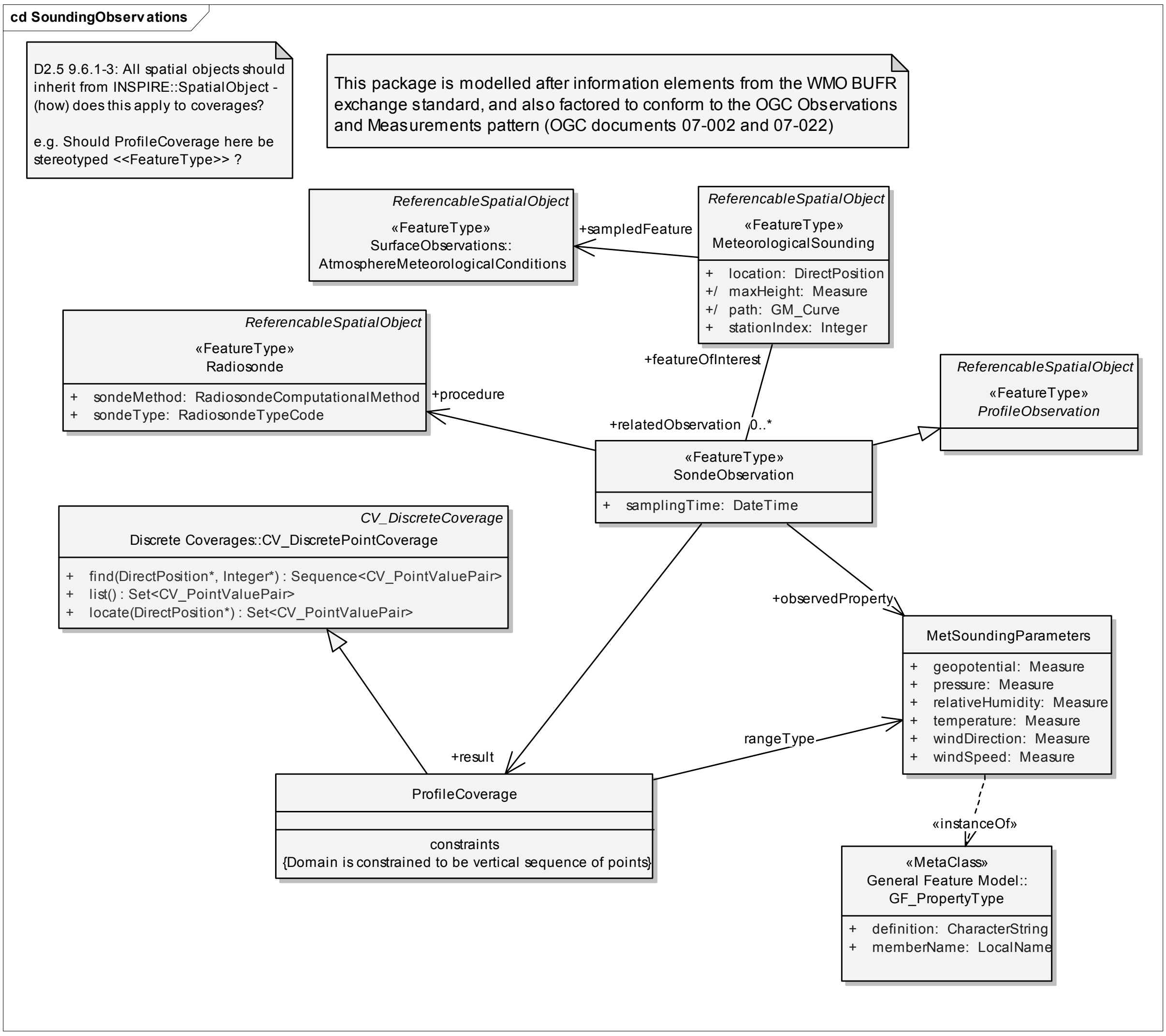


Figure 33

SurfaceObservations

Surface observations of current weather are made globally either by human or automatically at fixed or ship-borne weather stations. The model here represents observations at fixed stations, omitting sea-ice and ocean-state observations etc. that would be included for ship-based stations. The information elements (as with those in the SoundingObservation model) are modelled after the WMO BUFR format and associated code tables.

As above, the OGC Observations and Measurements patterns have also been used. In this case, the *observation* (SurfaceObservation) is made on a fixed point 'Sampling Feature' (StationFacility). The observation *procedure* is a (manned or automatic) MeterologicalStation, while the result (Weather) aggregates observations of cloud, pressure, and current meteorological conditions (wind, temperature, humidity, visibility, weather). As previously, the model identifies a canonical spatial object class (AtmosphereMeteorologicalConditions) as proxy for the real, complete atmospheric state that is being observed.

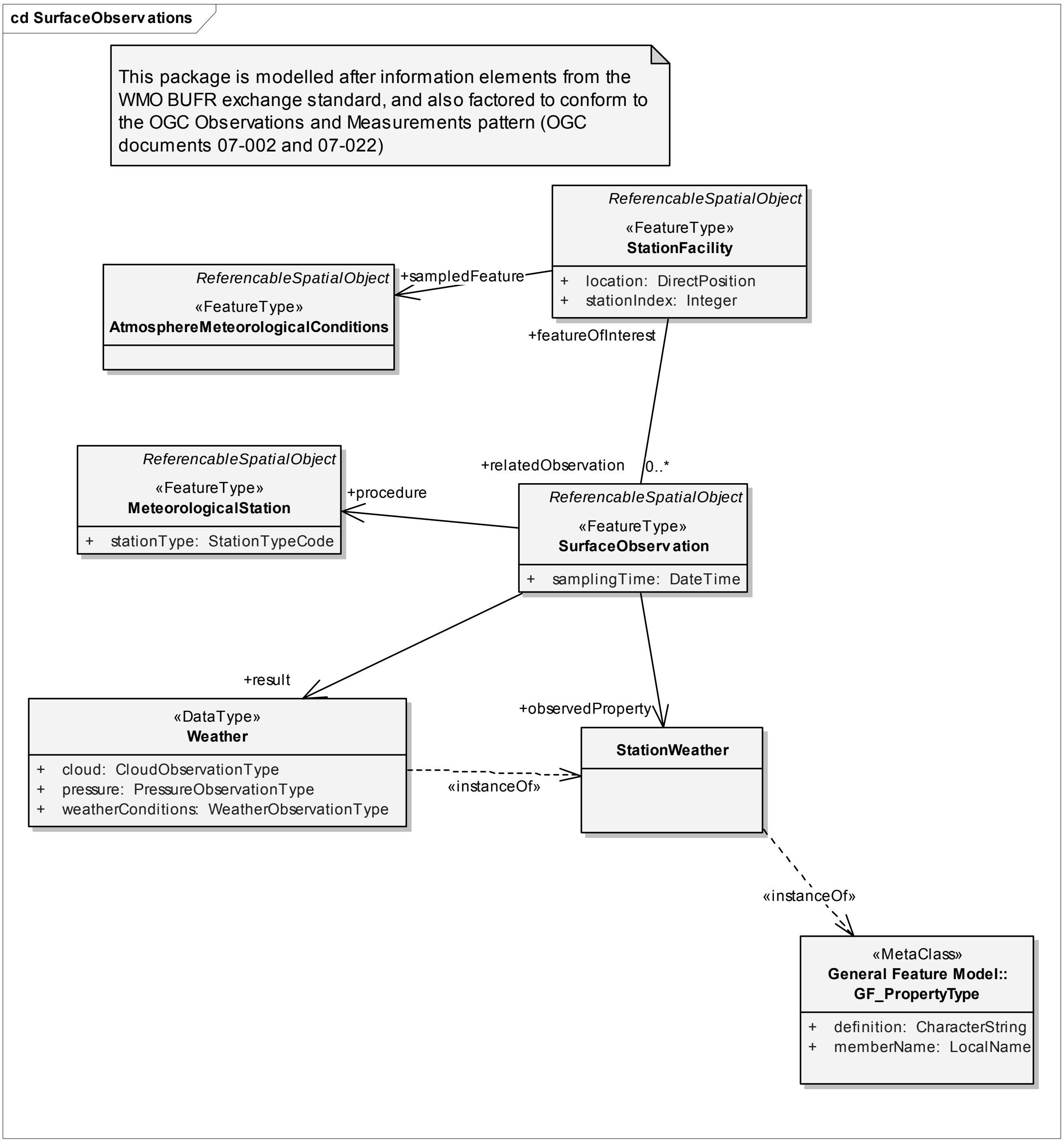


Figure 34

D.4.2 Analyses

Uniform analyses of weather conditions based on discrete observations are undertaken at meteorological centres on an operational (e.g. six-hourly) basis. They are one type of gridded product. Satellite and radar data are also gridded, but generally not stored in GRIB format.

Utility types

Some utility types are needed to characterise the generating meteorological centre and the type of product. While these are here modelled after WMO code tables, they could also be represented through a richer hierarchy of gridded product types (e.g. distinguishing analyses and forecasts through different spatial object types).

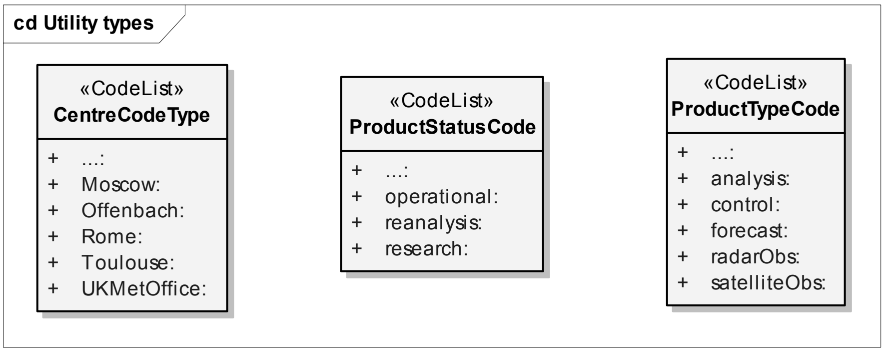


Figure 35

GriddedProducts

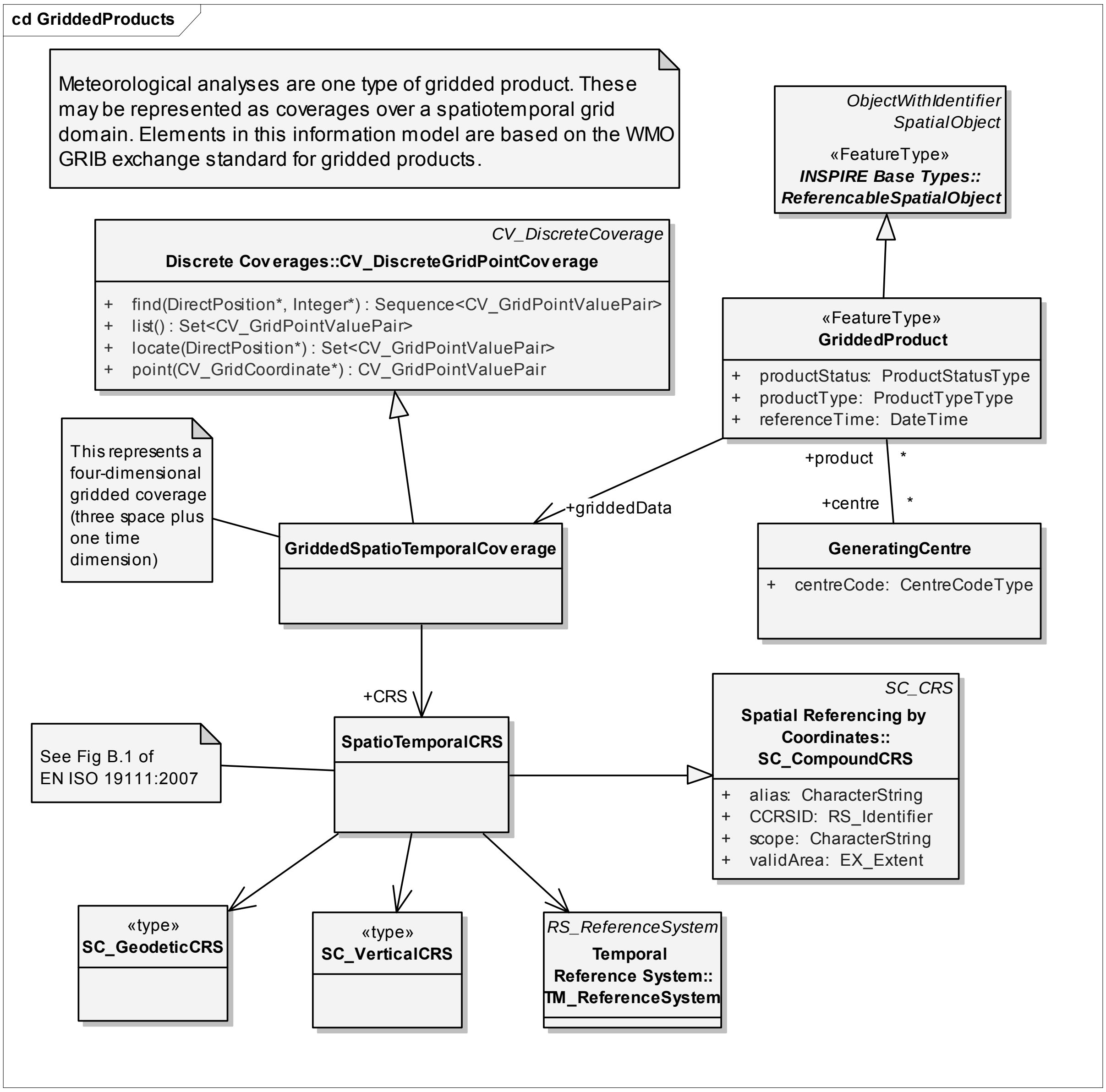


Figure 36

Conceptually, analysis products may be modelled as coverages over a gridded spatiotemporal domain. Depending on the analysed parameter, the spatial axes of the grid may be two-dimensional (e.g, surface pressure) or three-dimensional (e.g. temperature). The combined spatiotemporal nature of the grid domain is modelled by associating the coverage to a compound CRS (SC_CompoundCRS) that combines a horizontal (SC_GeodeticCRS), a vertical (SC_VerticalCRS) and a temporal (TM_ReferenceSystem) coordinate reference system as provided by ISO 19111:2007.

The model here is motivated by elements in the WMO GRIB information model. While it is less explicitly conformant to the Observations and Measurements model, it may nevertheless be factored in a similar manner: the GriddedProduct may play the role of an *observation*, with the GeneratingCentre providing the *procedure*, and the gridded coverage (GriddedSpatioTemporalCoverage) providing the *result*.

D.5 Application schema example: "Geology"

The following two UML diagrams cover selected aspects of a geological application schema. This is based on a draft of GeoSciML version 2.

The types depicted in the diagrams use some types specified elsewhere in the model. This includes the candidate OGC standard Observations & Measurements schema (part 1: observations, part 2: sampling features) as well as a range of base value types from properties (using a prefix "CGI_").

D.5.1 Mapped Feature

A Mapped Feature provides a link between a notional feature (description package) and one spatial representation of it, or part of it (exposures, surface traces and intercepts, etc).

Its properties are:

- the specific bounded occurrence, such as an outcrop or map polygon
- the Mapped Feature carries a geometry (shape)
- the association with a Geologic Feature (legend item) provides specification of all the other descriptors
- the association with a Sampling Feature provides the context and dimensionality

A Mapped Feature is always associated with some sampling feature - e.g. a mapping surface, a section, a Borehole, etc. As noted on the diagram, if the associated sampling feature is a Borehole, then the shape associated with the Mapped Feature will usually be either a point or an interval. This reconciles the 2-D ("map", section) and 1-D (borehole, traverse) viewpoints in a common abstraction.

NOTE While geology is an Annex II theme and thus unique identifiers should be assigned to spatial objects in accordance with the Generic Conceptual Model, it currently is not common practice in geology to assign identifiers to spatial objects. Therefore, the features derive from SpatialObject instead of ReferencableSpatialObject.

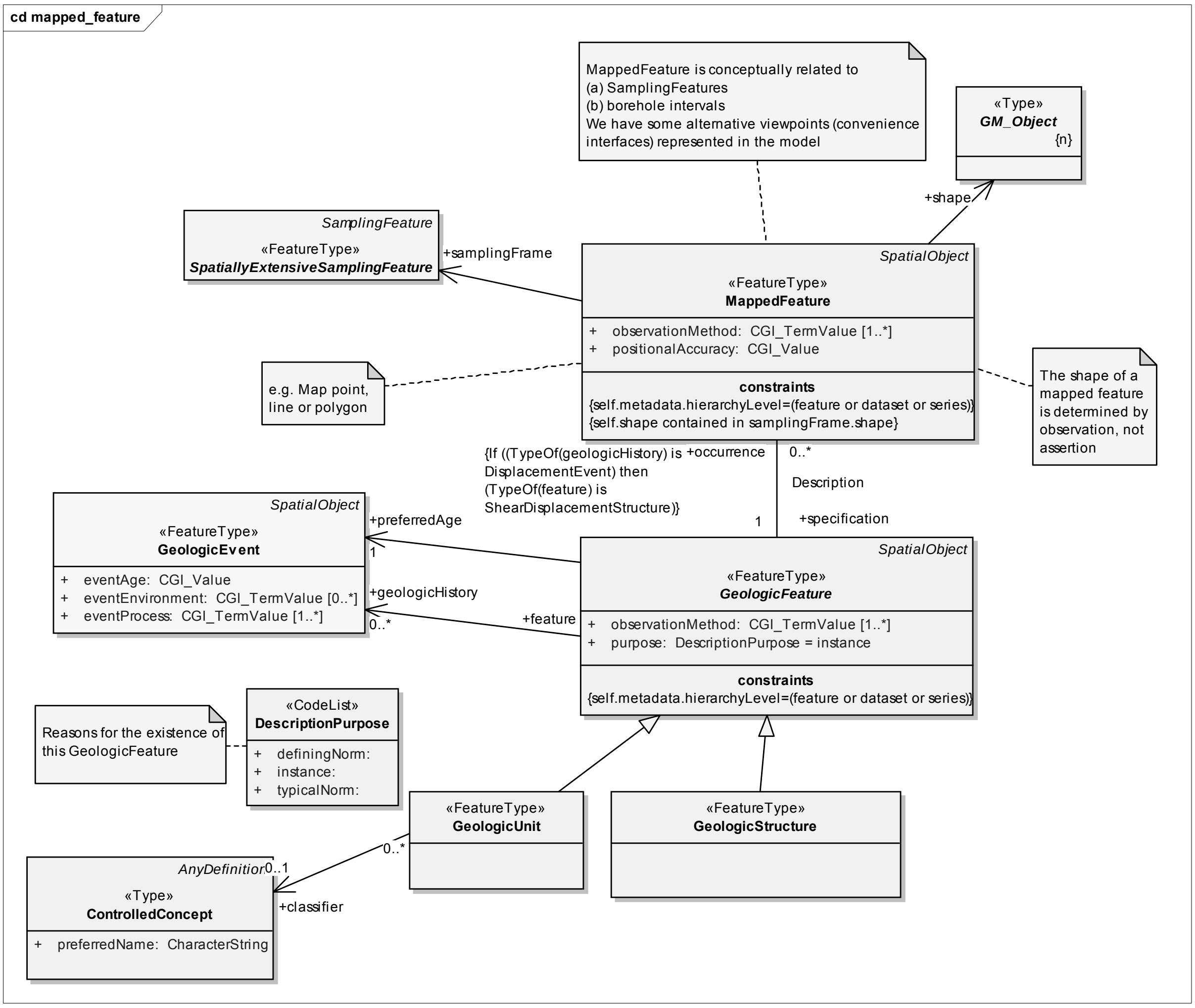


Figure 37

D.5.2 Geologic Unit

Operationally, the Geologic Unit is a container used to associate geologic properties with some mapped occurrence (through GeologicFeature.occurrence → MappedFeature link), or with a geologic unit controlled concept in a vocabulary (through the GeologicUnit.classifier → ControlledConcept link).

Conceptually, the geologic unit may represent a body of material in the Earth whose complete and precise extent is inferred to exist (NADM geologic unit, stratigraphic unit in sense of NACSN or Intnl Stratigraphic Code), or a classifier used to characterise parts of the Earth (e.g. lithologic map unit like 'granitic rock' or 'alluvial deposit', surficial units like 'till' or 'old alluvium').

Spatial properties are only available through association with a Mapped Feature.

Geologic unit includes both formal units (i.e. formally adopted and named in the official lexicon) and informal units (i.e. named but not promoted to the lexicon) and unnamed units (i.e. recognisable and described and delineable in the field but not otherwise formalised).

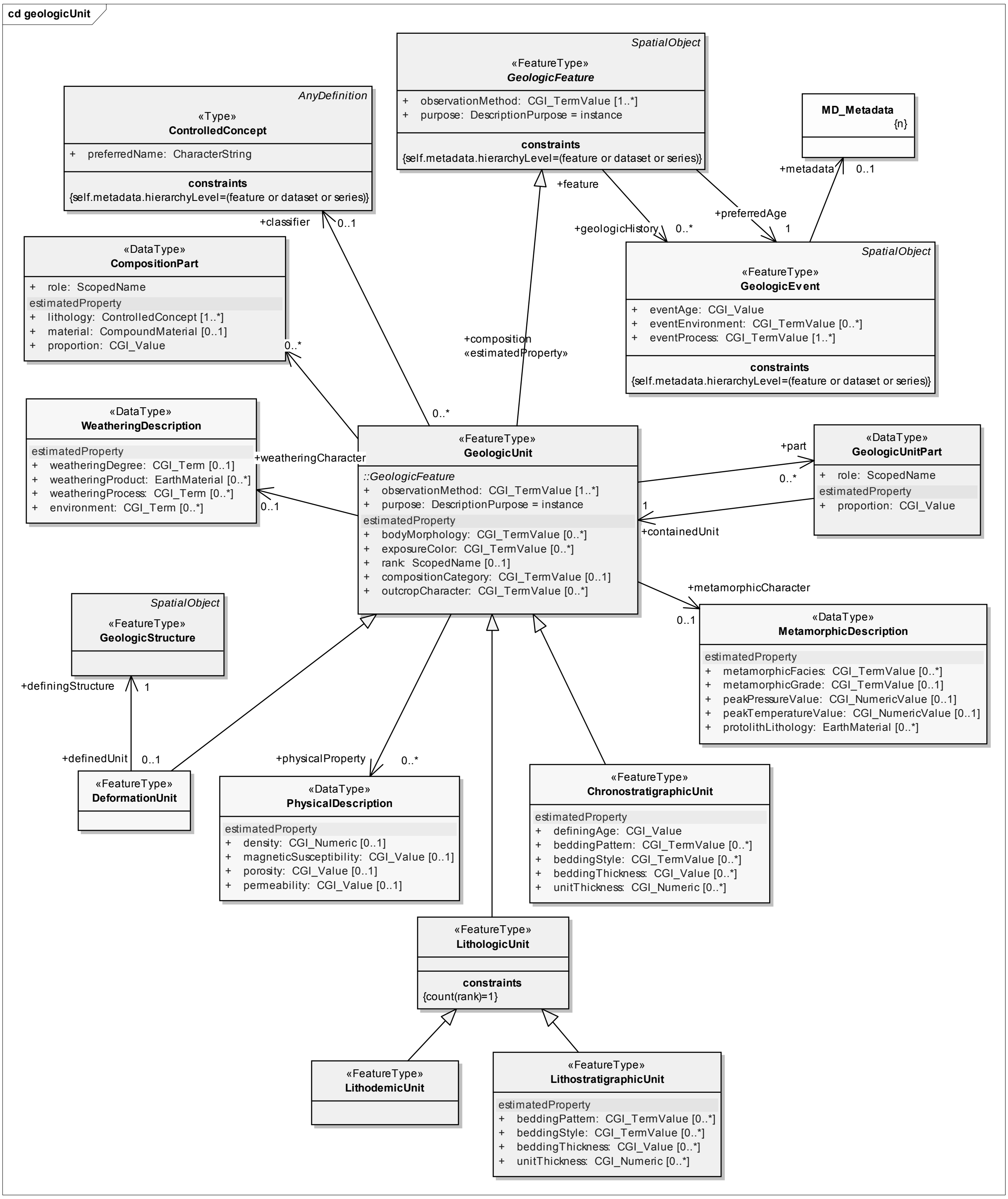


Figure 38

NOTE The ControlledConcept.preferredName property currently does not support multilingual names. This would need to be addressed in the further work on the application schema.

Annex E
(informative)

Use case template

Document source: RISE (2006) Methodology and Guidelines on Use Case and Schema Development, version 1.1, deliverable number 15

A use case is initiated by a user with a particular goal in mind, and completes successfully when that goal is satisfied. It describes the sequence of interactions between actors and the system necessary to deliver the service that satisfies the goal. Actors are parties outside the system that interact with the system. An actor may be a class of users, roles users can play, or other systems." This should clarify the question related to whose "goal" we are talking about.

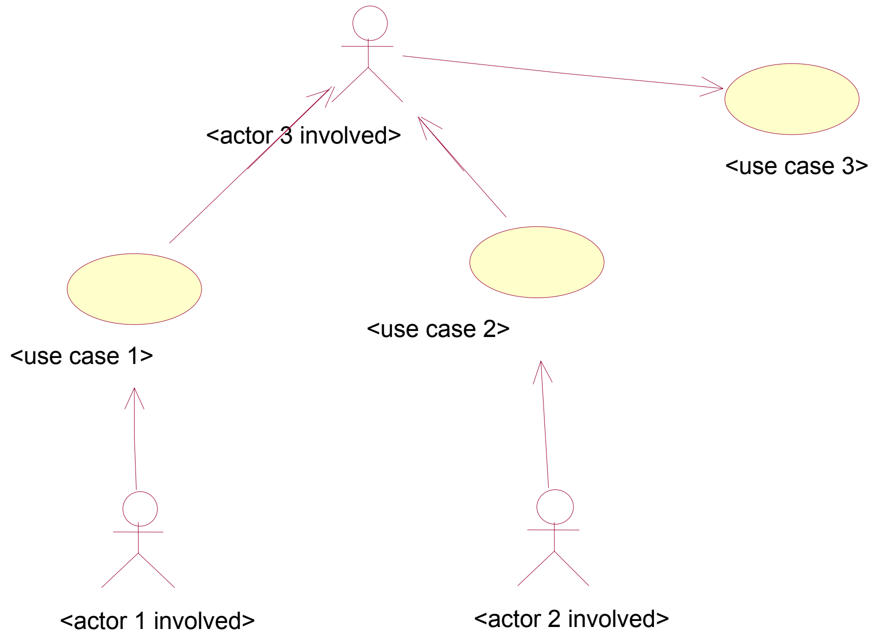
Generally, use case steps are written in an easy-to-understand structured narrative using the vocabulary of the domain. This is engaging for users who can easily follow and validate the use cases, and the accessibility encourages users to be actively involved in defining the requirements.

The following is a use case template that may be used to describe use cases.

Describe the use case(s) with as much detail as possible. This information will be the input in order to (i) detect possible new requirements, and (ii) assess the services that need to be implemented according to the needs of the end users and stakeholders.

The description consists of three parts:

Part 1: UML use case diagram to provide an overview of the use case(s) and the involved actors (example below).



Part 2: Narrative explanation of the use case(s)

Part 3: Detailed, structured description of the use case (template below)

Use Case Description	
Name	<name of use case>
Priority	<high/medium/low>
Description	<short description>
Pre-condition	<What are the pre-requisites from other use cases? What input is required?>
Flow of Events – Basic Path	
Step 1.	
…	
Step m.	
Step m+1.	
…	
Step n.	
Flow of Events – Alternative Paths	
Step m.	
Step m+1.	
Post-condition	<What is the output for other use cases? What are the anticipated next steps?>
Data source: <Name> ***[repeat per data source]***	
Description	
Data provider	
Geographic scope	
Thematic scope	
Scale, resolution	
Delivery	
Documentation	

Repeat for additional use cases

Annex F
(informative)

Checklist for data interoperability

F.1 General remarks

A use case methodology has to be applied. The use case template and the checklist along with guidelines on their use in this document may help in this step, but their application by TWG is optional as long as a result of the step the user requirements are identified and the data interoperability components have been considered.

The goal of the checklist provided here is to assist the TWG for the exploitation of Reference Material and/or of the user requirement survey.

The checklist was originally developed as part of the RISE methodology (*Document source: RISE (2006) Methodology and Guidelines on Use Case and Schema Development, version 1.1, deliverable number 15*). Its purpose was and is to assist the facilitator of the data specification development process. The checklist is intended to aide in addressing all relevant aspects of data interoperability.

F.2 Overview

The harmonisation process consists of several steps, each addressing the data interoperability components documented in the RISE methodology document. This spreadsheet shall help to identify the questions to be addressed in each step for each component.

The important steps supported by the checklist are:

- Requirements: the expected results as derived from the use case description
- As-is analysis: situation with respect to the existing data sources
- Gap analysis: identification of issues (gap between requirements and current situation)
- Harmonisation approach: proposed approach for the harmonised data specification

Note that not all components are relevant in every step.

It is proposed to document at least summary information of the analysis in a matrix representation so that it is easy to get an overview per step (e.g. an overview of all requirements) or per data interoperability component (e.g. spatial profile required and used in all existing data sources). additional information may be provided in separate documents that are linked to the spreadsheet.

Replace the checklist information in the individual cells by the use case specific information addressing the topics of the checklist and any other relevant information identified in the use case analysis.

Note that it may be required/helpful to add additional columns; for example, if the analysis of multiple data sources in the as-is analysis cannot be displayed in a single column in a clear way.

In the general harmonisation process, the use case description comes first and feeds into the column ‘Use case requirements’ and partially 'As-is analysis'. As the first-cut application schema is being developed, the 'As-Is analysis' will be enhanced and the entries in 'Gap Analysis' and, as a result, 'Harmonisation approach' will be added. In practice the steps do overlap and are performed in iteration. For instance, known information on known gaps will and should already be added in the phase of the collection of the use case requirements. However, it is important to note that the checklist will not be available in full before all steps in the harmonisation process have been completed.

F.3 The checklist

Data interoperability component	**Use case requirements**	**As-is analysis**	**Gap analysis**	**Harmonisation approach**
	Describe the use case requirements in particular for (harmonised) input data sources in view of a harmonised product as described in the use case. The checklist may refer to a use case description by the domain experts (users), if one exists	Identify and describe the available data sources as well as possible for the use case. The checklist may refer to a use case description by the domain experts (users), if one exists	Check the data interoperability components for differences and/or contradictions between the available data sources (As-is analysis) and the target specification (Use case requirements)	If differences or contradictions have been identified in the Gap analysis, investigate suitable harmonisation methods
0 Context	What are the main characteristics of the use case (general purpose, required level of detail, geographic extent, ...).	What are the existing data sets that have been identified for this use case?	If already identified, what are the main issue(s) in this use case?	Is the list of existing data sets relevant? - is it exhaustive? - if not, what are the reasons to reject some of them ?
(A) INSPIRE Principles Answers for this component will probably be "not applicable".	Nothing to be added here as it is assumed that all GMES/INSPIRE data specifications will adhere to the INSPIRE principles. However, if a deviation from these principles has been identified as required, it shall be documented here.	How does the context in which the data source has been designed and created align with the INSPIRE principles?		Potential conflicts with INSPIRE principles? For instance: is it required to improve the data specification of the input data source, or to create new data to fill gaps?
(B) Terminology	Which concepts are required? If there is a document about user requirements, does it include a glossary?	Which concepts are used in existing data? Do the documents about existing data include a glossary?	Which concepts do not match between the use case and the Inspire Glossary?	Which concepts from the use case may be matched with concepts in the INSPIRE Glossary? Which concepts have to be added and defined in the INSPIRE Glossary?

(C) Reference Model In INSPIRE, answer for this component will probably be "not applicable"	Has a reference model been defined for the application ? If yes: - Is it based on ISO 19101 ? On the OpenGIS Reference Model (ORM) ? Anything else ? - What methodology has been used in the reference model (e.g. RM-ODP) ?	Is a reference model available for the application providing source data ? If yes: - Is it based on ISO 19101 ? On the OpenGIS Reference Model (ORM) ? Anything else ? - What methodology has been used in the reference model (e.g. RM-ODP) ?		

(D) Application schemas and feature catalogues	The point is to identify the requirements about application schema and specially the required features and attributes. Is there a documentation about these requirements? Is there already an application schema ? If yes, which Conceptual Language Schema does it use? Which elements of General Feature Model are required : - features - attributes - association - inheritance relation - properties - constraints - operations - others. Then, identify the list of required features, attributes, associations, ...	The point is to identify the existing application schemas and specially the existing features and attributes. Is there a documentation about existing data? Are there already application schema(s) ? If yes, which Conceptual Schema Language(s) do it use? Which elements of General Feature Model are used in existing data : - features - attributes - association - inheritance relation - properties - constraints - operations - others. Then, identify the list of available features, attributes, associations, ...	Which features/attributes/… are missing in existing data? (may be displayed in auxilliary matching tables).	Which features/attributes should be kept as mandatory elements? Which features/attributes may be kept only as optional elements?
(E) Spatial and temporal aspects (Vector geometry)	Geometries of these feature types: - Dimensionality of the geometries (0D, 1D, 2D, 3D)? - Interpolation types for curves and surfaces ? - Sharing of geometry objects required ? For which features and in which cases ?	Geometries used: - Dimensionality of the geometries (0D, 1D, 2D, 3D)? - Interpolation types for curves and surfaces ? - Sharing of geometry objects required ? For which features and in which cases ?	Does the data source differ from the use case requirements with respect to geometry (e.g. curve interpolation) ?	Processing of Geometry?

(E) spatial and temporal aspects (Topology)	Topology required? If yes: - Which requirements (e.g need for continuous network ? Initial and final nodes ?) ? - Is is only required for internal processing or data consistency? Or is it of interest to the user ? - How is topology related to geometry?	Topology used? If yes: - Which rules (e.g need for continuous network ? Initial and final nodes ?) ? - How is topology related to geometry?	Does the data source differ from the use case requirements with respect to Topology (e.g. implicit/explicit topology) ?	Processing of Topology ?
(E) Spatial and temporal aspects (Coverages)	Are Coverages required? If yes: - Which types (raster, triangulated irregular networks, point coverages, polygon coverages, etc.) ? For grids : - type of grid (DTM, TIN, ...) ? - Resolution of grids, etc. ? For images : - source (aerial photo, satellite image, ...) ? - type of geometric rectification, correction level (e.g. orthophoto)? - pixel size ? - radiometry (black and white, colour, infrared, ...) ?	Are Coverages used? If yes: - Which types (raster, triangulated irregular networks, point coverages, polygon coverages, etc.) ? For grids : - type of grid (DTM, TIN, ...) ? - Resolution of grids, etc. ? - which applications ? - [consistency and differences between the data sources] For images : - Which source (aerial photo, satellite image, ...) ? Which date ? - type of geometric rectification, correction level (e.g. orthophoto) ? - pixel size ? - radiometry (black and white, colour, infrared, ...) ? - which applications ?	If coverages are used, does the data source differ from the use case requirements with respect to resolution, type of grid etc. ? Which problems do you have when working with heterogeneous coverage data ?	Which processings do you carry out to work with different grids ? Which processings do you carry out to work with different types of images ? (e.g. resampling, radiometric equalization, …) ?

(E) Spatial and temporal aspects) (Temporal profile)	Temporal model - Support for features that move or change geometry with time? - Support for multiple versions of a feature (historic data)? Or even versioning of properties?	Temporal model - Support for features that move or change geometry with time? - Support for multiple versions of a feature (historic data)? Or even versioning of properties?		
(F) Multi-lingual text and cultural adaptibility (specification)	Is the data specification (or schema/feature catalogue) required in several languages ? Which languages?	In which languages is documentation (specification or feature catalogue) about existing data ?	Are some language(s) missing? Which?	What solution have you adopted or are you envisaging (e.g. multilingual thesauri, automatic translation) ?
(F) Multi-lingual text and cultural adaptibility (content)	Are geographical names required in several languages ? Which languages.	Are geographical names available in several languages ? Which languages?		What solution have you adopted or are you envisaging (e.g. multilingual text) ?
(G) Coordinate referencing and units model	Reference systems required - Coordinate Reference Systems (horizontal, vertical) - Temporal Reference Systems - Units of measurement	Reference systems available - Coordinate Reference Systems (horizontal, vertical) - Temporal Reference Systems - Units of measurement	Does the data source differ from the use case requirements with respect to Coordinate Reference Systems and/or Units (e.g. different geodetic datum, map projection, units of measurement) ? Does the data source differ from the use case requirements with respect to temporal reference (data sets refer to different situations in time, different time intervals etc.) ?	Coordinate transformation (Datum transformation, map projection) ? For Coordinate reference system transformation : - method used for horizontal datum transformation? - if image data, is there resampling ? Which method (nearest neighbour, bilinear, bicubic, ...) ? Is there "retiling" ? - method used for vertical datum? - Assessment of the results ? Transformation of source data forth/back in time?

(H) Object referencing model	Object referencing methods required? If yes, for which applications (e.g. gazetteer service, to enhance consistency between data,...)? Which methods are required/desirable?: - by name ? - by code (e.g. administrative unit code) ? - by identifier ? - by geometry ? - other (e.g. linear referencing, by address) ? - requirements for identifiers and object referencing in general	Are object referencing methods applied? If yes, which methods: - by name? - by code (e.g. administrative unit code)? - by identifier? - by geometry? - other (e.g. linear referencing, by address)?	If the use case requires object referencing methods: does the data source meet the requirements of the object referencing methods?	
(I) Data transformation / guidelines	n/a			To be investigated for all transformation methods applied in the process of harmonisation: - What efforts need to be taken ? (Feasibility!) - Is the transformation performed on-the-fly ?) - Is the data transformation model driven? - Are the results put into a redundant secondary data repository ? - For methods performed on-the-fly: how are errors and conflicts treated ?

(J) Portrayal	Which data do you need to display? How? - map service and/or feature portrayal service or other? - which scales? which symbolisation? which portrayal rules? … Is a portrayal catalogue required?	Is existing data supplied with symbolisation? If yes, how ? (e.g. by view services, within GIS formats, …). For which scales? Are there existing portrayal catalogues?	If the use case requires portrayal: does the data source provide the necessary input for portrayal?	Is portrayal catalogue to be created new? Is existing portrayal catalogue adopted ? If yes, does it need to be modified?
(K) Identifier Management	Are identifiers required ? For which features ? Which roles do identifiers for entities play? Which are the required characteristics of identiiers (e.g. unique, stable)? Is there a management for such identifier specified?	Are there identifiers in existing data sets ? What do you know about them ? (Which roles do identifiers for entities play? Is there a management for such identifier specified? relevant for existing data ?)	Is identifier definition and management consistent?	Which solution have you adopted (or are you envisaging) if existing data does not have the appropriate identifiers ?
(L) Registers and registries	Which registers are required (if any) : - reference system - units of measurement - feature concept dictionary - features catalogues -codelists - thesauri - portrayal catalogues - other; Do these registers require to be conform to ISO standards ? Other standards ?	Which registers are available (if any) : - reference system - units of measurement - feature concept dictionary - features catalogues -codelists - thesauri - portrayal catalogues - other. Are these registers conform to ISO standards ? To other standards ?		Are registers to be created new? Are existing registers adopted ? If yes, do they need to be modified?

(M) Metadata	Metadata required : - Discovery level / Exploration level /Exploitation level ? - ISO 19115 compliant? - Language?	Metadata available : - Discovery level / Exploration level /Exploitation level ? - ISO 19115 compliant? - Language?	Does the data source differ from the use case requirements with respect to metadata ? (level of detail, language etc.)	Is the metadata transformed in the process of data interoperability? (at data set level, at feature level ?)
(N) Maintenance	Requirements to maintenance of data (update cycle? Incremental update?)	Update procedures of data sources ?	Does the data source meet the use case requirements with respect to update procedures ?	How will harmonisation work with future updates?
(O) Data quality	Data quality requirements ? If yes: - Positional accuracy ? - Other aspects - completeness, thematic accuracy, logical consistency, temporal accuracy? Data quality management: - Interaction with user on quality requirements, e.g. published quality levels ? - Quality evaluation / conformance testing ?	Data quality requirements in existing data? If yes: - Positional accuracy ? - Other aspects - completeness, thematic accuracy, logical consistency, temporal accuracy?	Does the data source meet the use case requirements with respect to data quality ?	Does the harmonisation method have an impact on data quality, e.g. positional accuracy ?

(P) Data transfer	Data transfer: - methods for encoding application and reference data ? - support for access to and update of data ? - change-only updates ? If yes – why and for which purpose ? - model-driven encoding, i.e. fully determined by the application schema in UML ? Have delivery methods (formats, etc) been specified ? If yes, same questions as for existing data ? If yes – has the service aspect been considered ? Can the required services be determined from the use case descriptions? Are existing services / service types taken into account?	How is the source data available: - formats, versions - service interfaces - usage constraints How is vector data supplied ? Which media ? Which format ? How is image data supplied ? Which image format ? If compressed format, which method ? Which factor ? How is image georeferencing provided ? Volume of image data to handle? (e.g. how many tiles ?) Is there any tool (e.g. image index) to help the handling of data ? How is other coverage data supplied ?	Does the data source differ from the use case requirements with respect to the data format (required - provided) ?	Is format conversion required ? Do service interfaces need to modified?
(Q) Consistency between data (between themes)	Are there consistency requirements between data from different sectors or themes ? If yes, between which features? Which consistency rules are required? Which application(s) require these consistency rules?	Which consistency rules are there already in existing data sets?	Which consistency rules are missing in existing data?	Is there some cooperation between the data providers of the different themes/sectors?

(Q) Consistency between data (between levels of detail)	Are data at different levels of detail required? If yes, for which features? For which levels of detail? Are there consistency rules required between different levels of detail? Which rules? For which applications?	Are data at different levels of detail available? If yes, which levels of detail? Are there links between levels of detail (e.g. same application schema, same currency, …) ?	Which consistency rules are missing in existing data? Which inconsistencies are an issue for the application(s)?	
(Q)Consistency between data (accross boundaries)	Do you need edge-matching between different areas ? Which areas? which boundaries (e.g. between regions, between countries, between sea and land) ? Which areas ? which boundaries (e.g. between regions, between countries, between sea and land) ? For which applications (e.g. cartography, navigation, …) ? For which features ? Which kind of edge-matching do you need ? -Is geometric edge-matching required ? If yes, is an accurate knowledge of the boundary required ? -Is semantic edge-matching required? for which attributes?	Are the boundaries well defined ?	For edge-matching , which problems have you met and what are their reasons ? - related to different positional accuracy. - related to different currency : some features may be just missing on one side - related to different interpolation type. For example: feature type with linear interpolation does not intersect with a feature type using arc by bulge interpolation. - related to different datums and/or projections. - related to accuracy of the transformation applied. For example using theoretic transformation (7 parameters transformation) instead of empiric transformation taking into account errors in the original reference frame.	Is there some cooperation between the neighbour countries ? Which solutions have you implemented or are you envisaging, for edge-matching ?

(R)Multiple representations	For products/services that deal with data at multiple scales / resolutions - Which applications require data at multiple scales / resolutions ? Which features are involved ? Which scales /resolutions are required ? - What are the requirements for consistency between the representation of the same entity in different scales / resolutions ? - Are requirements known how data is supposed to be aggregated/linked across different resolutions (“generalisation” of data) ?	Are data at different levels of detail available? If yes, which levels of detail? Are there links between the different levels of detail (e.g. generalisation)?		How is Data aggregated/linked across different resolutions (“generalisation” of data) ?
(S) Data capturing	What is the level of detail required? Which selection criteria are required (e.g. for a specific feature type : all features, features whose length/area is bigger than a given threshold)? Where are they defined ?	What are the levels of detail available? Which selection criteria are there in existing data? Are they documented?	For which features are there more data available than required ? For which features are there less data than required?	If more data is an issue, is some generalisation of data possible? If less data is an issue, is new data capture possible (at least, in future)?

(T)Conformance	Do conformance rules exist or can they be derived from the descriptions? Is it required to specify conformance rules for the data specification? Is conformance with the INSPIRE implementing rules required (which are not yet known/defined)? With any other conformance rules ?	Does the data source explicitly conform to any specifications?		

Annex G
(informative)

Tools

G.1 UML model

It is planned that the consolidated UML model will be maintained using the UML tool Enterprise Architect from Sparx Systems (http://www.sparxsystems.com/products/ea.html). The details, how the concurrent work on the individual packages will be managed are still to be defined. The model will be maintained with version control to allow tracking of the different versions of the model.

NOTE 1 More requirements on tooling need to be identified and described, e.g. for registers and the registry.

NOTE 2 This Annex may be removed from this document in the next revision and continued as part of a separate 'Tools' document, under the responsibility of the CT.

Bibliography

This clause lists documents that were used in the drafting of this document, but which are not normative.

- Bakker, Nico J., 2005: Developing a new geographical object database. Experiences from idea to delivering datasets TOP10NL. Paper presented at the International Cartographic Conference.
- Berljant, 2002: Kartographie. Aspekt Press. Moskau
- Cockburn A.: 2000, 'Writing effective Use Cases', Addison-Wesley Professional.
- DGIWG: DFDD
- Domingues C. & Bucher B., 2006 : Application d'aide à la conception de légendes. Actes du colloque SAGEO, Strasbourg, 2006
- EuroBoundaryMap, 2007: User Guide. v1.1, EuroGeographics 2007
- European Commission, 1994: CORINE Land cover technical guide. Office for the Official Publications of the European Communities, Luxembourg.
- European Commission, 2000: CORINE Land Cover Technical guide, Addendum 2000. Technical Report N)40, Copenhagen, EEA, May 2000
- European Commission, 2003: Implementing the Geographical Information System Elements (GIS) of the Water Framework Directive, 2003.
- European Commission, 2006. INSPIRE Multiple-Representation and Data Consistency Workshop. Ispra November 2006.
- EuroGlobalMap, 2005: Technical Guide for Data Collecting and Finalizing. v1.4, National Land Survey of Finland, 1.8.2005
- EuroRegionalMap, 2004: Specification and Data Catalogue. IGN Belgium, v4.0, 13/01/2004
- EuroRegionalMap, 2005: Edge Matching Directives On cross border area. v1.0, 22 Dec 05, IGN Belgium
- EuroRoadS Project, 2006: Final draft specification of core European road data. Deliverable D6.5, v2.0, National Land Survey of Sweden, 18/01/2006.
- EuroRoadS Project, 2006: Final specification of Road Network Information Model. Deliverable D6.3, v2.0, National Land Survey of Sweden, 23/01/2006
- Fowler, M (2003): UML Distilled: A Brief Guide to the Standard Object Modeling Language (3rd Edition) (The Addison-Wesley Object Technology Series), Addison-Wesley Professional
- Furthermore: The page of the Object Management Group with the UML specification (http://www.uml.org/)
- GeoScience Markup Language (GeoSciML), https://www.seegrid.csiro.au/twiki/bin/view/CGIModel/GeoSciML
- IGN France, 2006 : Route 5000®, descriptif de contenu. Edition 5, 2006
- INSPIRE RDM Working group, 2002: Reference Data and Metadata Position Paper. V4.2, 02/10/2002.
- Institut Géographique National : BD Ortho version 2 descriptif de livraison
- International Commission of Stratigraphy: http://www.stratigraphy.org
- JRC, SDI unit (Ed.), 2006: presentations from the INSPIRE Workshop on Multiple-Representation and Data Consistency. Ispra, November 7-8 2006. http://sdi.jrc.it/ws/multiple_rep/

- Lamine K. and S. Mustière, 2005: Intégration de données transfrontalières relatives à la randonnée pédestre. Laboratoire COGIT, Paris 6. European Project WoW: Walk on Web; IST-2-004688-STP.
- Larman C. (2004): 'Applying UML and patterns: An introduction to object-oriented programming and design and iterative development'. Pearson Education Inc
- Lenkungsgremium GDI-DE (Steering Committee GDI-DE), BKG: Documentation on the Modelling of Geoinformation of Official Surveying and Mapping GeoInfoDok
- Nebert, Douglas D. (editor), The SDI Cookbook, January 2004, GSDI
- Neudeck, Stefan, 2001, Stefan: Zur Gestaltung topographischer Karten für die Bildschirmvisualisierung. Universität der Bundeswehr. Studiengang Geodäsie und Geoinformation. München
- RISE, 2006: Methodology & Guidelines on Use case & Schema Development. Deliverable 19, version 1.1, September 2006.
- Robinson et. al., 1995: Elements of Cartography. John Wiley & Sons, Inc.
- OGC, 2007: Observations and Measurements – Part 1 – Observation schema (OGC document 07-022r1)
- OGC, 2007: Observations and Measurements – Part 2 – Sampling Features (OGC document 07-002r3)
- Webster, Third New International Dictionary of the English Language (Unabridged), Merriam-Webster Publisher. June 2002