

PD CEN/TR 15449-3:2012



BSI Standards Publication

Geographic information — Spatial data infrastructures

Part 3: Data centric view

bsi.

...making excellence a habit.™

National foreword

This Published Document is the UK implementation of CEN/TR 15449-3:2012. Together with PD CEN/TR 15449-1:2012, PD CEN/TR 15449-2:2012 and PD CEN/TR 15449-4, it supersedes PD CEN/TR 15449:2011, which will be withdrawn upon publication of all parts of the series.

The UK participation in its preparation was entrusted to Technical Committee IST/36, Geographic information.

A list of organizations represented on this committee can be obtained on request to its secretary.

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

© The British Standards Institution 2013. Published by BSI Standards Limited 2013

ISBN 978 0 580 75643 6

ICS 07.040; 35.240.70

Compliance with a British Standard cannot confer immunity from legal obligations.

This Published Document was published under the authority of the Standards Policy and Strategy Committee on 31 July 2013.

Amendments issued since publication

Date	Text affected
------	---------------

TECHNICAL REPORT
RAPPORT TECHNIQUE
TECHNISCHER BERICHT

CEN/TR 15449-3

October 2012

ICS 35.240.70; 07.040

Supersedes CEN/TR 15449:2011

English Version

**Geographic information - Spatial data infrastructures - Part 3:
Data centric view**

Information géographique - Infrastructures de données spatiales - Partie 3: vue centrée sur les données d'une infrastructure de données spatiales (IDS)

Geoinformation - Geodateninfrastrukturen - Teil 3: Datenzentrierte Sicht

This Technical Report was approved by CEN on 27 May 2012. It has been drawn up by the Technical Committee CEN/TC 287.

CEN members are the national standards bodies of Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, Former Yugoslav Republic of Macedonia, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey and United Kingdom.



EUROPEAN COMMITTEE FOR STANDARDIZATION
COMITÉ EUROPÉEN DE NORMALISATION
EUROPÄISCHES KOMITEE FÜR NORMUNG

Management Centre: Avenue Marnix 17, B-1000 Brussels

Contents

Page

Foreword.....	4
Introduction	5
1 Scope	7
2 Normative references	7
3 Terms and definitions	7
4 Abbreviated terms	8
5 Data-centric view on SDI.....	10
5.1 Introduction	10
5.2 The model-driven approach.....	11
6 Aspects of data specifications	12
6.1 General.....	12
6.2 Semantics and semantic interoperability.....	12
6.3 Conceptual schema language	13
6.3.1 Overview	13
6.3.2 Relevant standards.....	14
6.3.3 Examples and tools	14
6.4 Application schema	14
6.4.1 Overview	14
6.4.2 Relevant standards.....	16
6.4.3 Examples and tools	16
6.5 Features and feature catalogues.....	17
6.5.1 Overview	17
6.5.2 Relevant standards.....	18
6.5.3 Examples and tools	18
6.6 Portrayal	18
6.6.1 Overview	18
6.6.2 Relevant standards.....	18
6.6.3 Examples and tools	19
6.7 Encoding.....	19
6.7.1 Overview	19
6.7.2 Relevant standards.....	20
7 Data management.....	20
7.1 Accessing data.....	20
7.2 Quality and conformity of spatial datasets	20
7.2.1 Overview	20
7.2.2 Relevant standards.....	21
7.3 Spatial referencing.....	22
7.3.1 Overview	22
7.3.2 Relevant standards.....	23
7.3.3 Examples and tools	23
7.4 Identifier management	23
7.4.1 Overview	23
7.4.2 Relevant standards.....	24
8 Metadata	24
8.1 Metadata types	24
8.1.1 Introduction	24
8.1.2 Discovery metadata	24

8.1.3	Feature level metadata.....	24
8.1.4	Dataset metadata.....	24
8.3	Examples and tools.....	25
9	Data Product Specification.....	25
9.1	Role of a Data Product Specification	25
9.2	Stepwise approach.....	25
9.2.1	General	25
9.2.2	Step 1 – Use case development.....	26
9.2.3	Step 2 – Identification of the user requirements and spatial object types.....	27
9.2.4	Step 3 – As-is analysis.....	27
9.2.5	Step 4 – Gap analysis.....	28
9.2.6	Step 5 – Data Specification Development.....	28
9.2.7	Step 6 – Implementation, test and validation	28
9.2.8	Step 7 – Cost-benefit analysis	28
9.3	Content of a Data Product Specification.....	29
9.4	Relevant standards	29
9.5	Examples and tools.....	29
	Bibliography.....	30

Foreword

This document (CEN/TR 15449-3:2012) has been prepared by Technical Committee CEN/TC 287 “Geographic information”, the secretariat of which is held by BSI.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. CEN [and/or CENELEC] shall not be held responsible for identifying any or all such patent rights.

This document supersedes CEN/TR 15449:2011.

The present standard comprises the following parts:

- CEN/TR 15449-1, *Geographic information — Spatial data infrastructures — Part 1: Reference model*
- CEN/TR 15449-2, *Geographic information — Spatial data infrastructures — Part 2: Best practices*
- CEN/TR 15449-3, *Geographic information — Spatial data infrastructures — Part 3: Data centric view (the present part);*
- CEN/TR 15449-4, *Geographic information — Spatial Data Infrastructure — Part 4: Service centric view*

Introduction

Spatial data infrastructure (SDI) is a general term for the computerised environment for handling data that relates to a position on or near the surface of the earth. It may be defined in a range of ways, in different circumstances, from the local up to the global level.

This Technical Report focuses on the technical aspects of SDIs, thereby limiting the term SDI to mean an implementation neutral technological infrastructure for geospatial data and services, based upon standards and specifications. It does not consider an SDI as a carefully designed and dedicated information system; rather, it is viewed as a collaborative framework of disparate information systems that contain resources that stakeholders desire to share. The common denominator of SDI resources, which can be data or services, is their spatial nature. It is understood that the framework is in constant evolution, and that therefore the requirements for standards and specifications supporting SDI implementations evolve continuously.

SDIs are becoming more and more linked and integrated with systems developed in the context of e-Government. Important drivers for this evolution are the Digital Agenda for Europe, and related policies (see Part 1). By sharing emerging requirements at an early stage with the standardization bodies, users of SDIs can help influence the revision of existing or the conception of new standards.

The users of an SDI are considered to be those individuals or organisations that, in the context of their business processes, need to share and access geo-resources in a meaningful and sustainable way. Based on platform- and vendor-neutral standards and specifications, an SDI aims at assisting organisations and individuals in publishing, finding, delivering, and eventually, using geographic information and services over the internet across borders of information communities in a more cost-effective manner.

Existing material about SDIs abounds. The criteria used for determining if a given standard or specification is referred to in this report are that the publication addresses an aspect of SDI, and that it is non-proprietary in nature.

Based on these considerations, the following reports have been taken into account:

- legal texts and guidelines produced in the context of INSPIRE;
- documents produced by ISO/TC 211 (and co-published by CEN);
- documents produced by the Open Geospatial Consortium (OGC), including the OpenGIS Reference Model (ORM);
- the European Interoperability Framework and related documents;
- deliverables from the European Union-funded projects (e.g. GIGAS, SANY).

Considering the complexity of the subject and the need to capture and formalise different conceptual and modelling views, CEN/TR 15449 is comprised of multiple parts:

- Part 1: Reference model: this provides a general context model for the other Parts, applying general IT architecture standards;
- Part 2: Best Practice: this provides best practices guidance for implementing SDI, through the evaluation of the projects in the frame of the European Union funding programmes;
- Part 3: Data centric view: this addresses concerns related to the data, which includes application schemas and metadata;

- Part 4: Service centric view (in preparation): this includes the taxonomy of services, concepts of interoperability, service architecture, service catalogue, and the underlying IT standards.

Further parts may be added in the future.

1 Scope

Part 3 of the Technical Report describes a data-centric view of a Spatial Data Infrastructure (SDI). The Data Centric view addresses the concepts of semantic interoperability, the methodology for developing data specifications through the application of the relevant International Standards, and the content of such specifications including Application Schemas, Feature Catalogues, General Feature Model, Data Lifecycle Management and Data Quality, Data Access and Data Transformation.

The intended readership of this Technical Report are those people who are responsible for creating frameworks for SDI, experts contributing to INSPIRE, experts in information and communication technologies and e-government that need to familiarise themselves with geographic information and SDI concepts, and standards developers and writers.

2 Normative references

Not applicable.

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

3.1

conceptual formalism

set of modelling concepts used to describe a conceptual model

EXAMPLE UML meta model, EXPRESS meta model.

Note 1 to entry: One conceptual formalism can be expressed in several conceptual schema languages.

[SOURCE: EN ISO 19101:2005]

3.2

conceptual model

model that defines concepts of a universe of discourse

[SOURCE: EN ISO 19101:2005]

3.3

conceptual schema

formal description of a conceptual model

[SOURCE: EN ISO 19101:2005]

3.4

conceptual schema language

formal language based on a conceptual formalism for the purpose of representing conceptual schemas

EXAMPLE UML, EXPRESS, IDEF1X.

Note 1 to entry: A conceptual schema language may be lexical or graphical. Several conceptual schema languages can be based on the same conceptual formalism.

[SOURCE: EN ISO 19101:2005]

3.5
conformance

fulfilment of specified requirements

[SOURCE: EN ISO 19113:2005]

3.6
component

physical, replaceable part of a system that packages implementation and provides the realisation of a set of interfaces

[SOURCE: ISO/TS 19103:2005]

3.7
identifier

linguistically independent sequence of characters capable of uniquely and permanently identifying that with which it is associated

[SOURCE: ISO/IEC 11179-3:2003]

3.8
interoperability

capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units

[SOURCE: ISO/IEC 2382-1:1993]

3.9
reference frame

aggregation of the data needed by different components of an information system

3.10
resource

asset or means that fulfils a requirement

[SOURCE: EN ISO 19115:2005]

3.11
spatial data infrastructure

SDI
policies, standards and procedures under which organisations and technologies interact to foster more efficient use, management and production of geo-spatial data

[SOURCE: United Nations SDI initiative (UNSDI)]

3.12
Use Case

specification of a sequence of actions, including variants, that a system (or other entity) can perform, interacting with actors of the system

[SOURCE: OMG UML Specification]

4 Abbreviated terms

API Application Programming Interface

ATS Abstract Test Suite

CEN	European Committee for Standardization / Comité Européen de Normalisation
CRS	Coordinate Reference System
DCE	Distributed Computing Environment
DPS	Data Product Specification
ebXML	Electronic Business using eXtensible Markup Language
EDR	Entity Relationship Diagrams
EN	European Standard (CEN deliverable)
EPSG	European Petroleum Survey Group
ESDIN	European Spatial Data Infrastructure Best Practice Network
INSPIRE	Infrastructure for Spatial Information in Europe
IT	Information Technology
GEOSS	Global Earth Observation System of Systems
GIGAS	GEOSS, INSPIRE and GMES an Action in Support
gmd	Geographic MetaData
GMES	Global Monitoring for Environment and Security
GML	Geography Markup Language
GSDI	Global Spatial Data Infrastructure Association
IEC	International Electrotechnical Commission
ISO	International Organisation for Standardization
NMA	National Mapping Agency
OCL	Object Constraint Language
ODP	Open Distributed Processing
OGC	Open Geospatial Consortium
OMG	Object Management Group
OSI	Open System Interconnection
RM-ODP	Reference Model of Open Distributed Processing
REST	Representational State Transfer
SDI	Spatial Data Infrastructure
SOA	Service Oriented Architecture

SOAP	Simple Object Access Protocol
SQL	Standard Query Language
TC	Technical Committee
TR	Technical Report
TS	Technical Specification
UML	Unified Modelling Language
UNSDI	United Nations SDI
URI	Uniform Resource Identifier
UUID	Universally Unique Identifier
WSDL	Web Service Description Language
XMI	eXtensible Markup Interface
XML	eXtensible Markup Language

5 Data-centric view on SDI

5.1 Introduction

Exchange of and access to spatial data is the principal objective of an SDI. The data are at the heart of an SDI. The spatial data in an SDI are a model of the real world. This model is developed according to well defined methodologies described in different standards. The model is made explicit through a concise description of data specifications in data specification documents. These specifications can then be used to develop new datasets or to transform existing datasets to the specifications by mapping the existing model to the model described in the specifications. In this way, semantic interoperability can be achieved: i.e. different datasets can be used together and be understood by different users in the same way. Metadata are part of the datasets and should get proper attention during the data modelling. Metadata will play a crucial role in documenting and understanding the content of the data model and data product specification, in achieving technical interoperability.

On top of the data, and by making use of the metadata, services can be built to make the data accessible through the web and to use them in any information system by viewing, downloading or processing them. This is often referred to as a Service Oriented Architecture (SOA). A SOA enables new and existing enterprise systems to share services, information and data across technical platforms, departments and ultimately across organisational, regional and national boundaries. The benefit is that this leads from a stand-alone system-centric view to an enterprise data-centric view of IT. The transition to a data-centric SOA allows an SDI to better leverage new and existing IT investments to support such an infrastructure. The data-centric transition builds a strategy around the organisations and their geospatial data infrastructure both to preserve the IT investment and to provide better access to authoritative data sources.

In the next clauses and sub-clauses the data modelling approach, the different aspects and role of data specifications, as well as the data management aspects are elaborated. The service centric view and SOA will be further developed in a separate part of this Technical Report. Where appropriate, the relevant international standards will be summarised and examples of implementations are given, as well as existing tools for implementation.

5.2 The model-driven approach

The model-driven approach follows the concepts developed in the model-driven architecture defined by OMG¹). The lifetime of a technical implementation is shorter than the lifetime of the information it handles. This makes it necessary to describe the information in a way that allows for new techniques and implementation environments to be applied.

The starting point of information modelling is the universe of discourse. This is a specific part of the real world that we want to describe in a model. The universe of discourse may include not only features such as roads, watercourses, lakes, property boundaries, but also their attributes, their functions and the relationships that exist among those features. A universe of discourse is described in a conceptual model. This model is formally represented in one or more conceptual schemas using a conceptual schema language. A conceptual schema that defines how a universe of discourse is described as data and can be used by one or more applications, is called an application schema. It provides a description of the semantic structure of the spatial dataset. The application schema also identifies the spatial object types and reference systems required to provide a complete description of geographic information in the dataset. Figure 1 describes the relationship between modelling the real world, the conceptual model and the resulting conceptual schema that represents it.

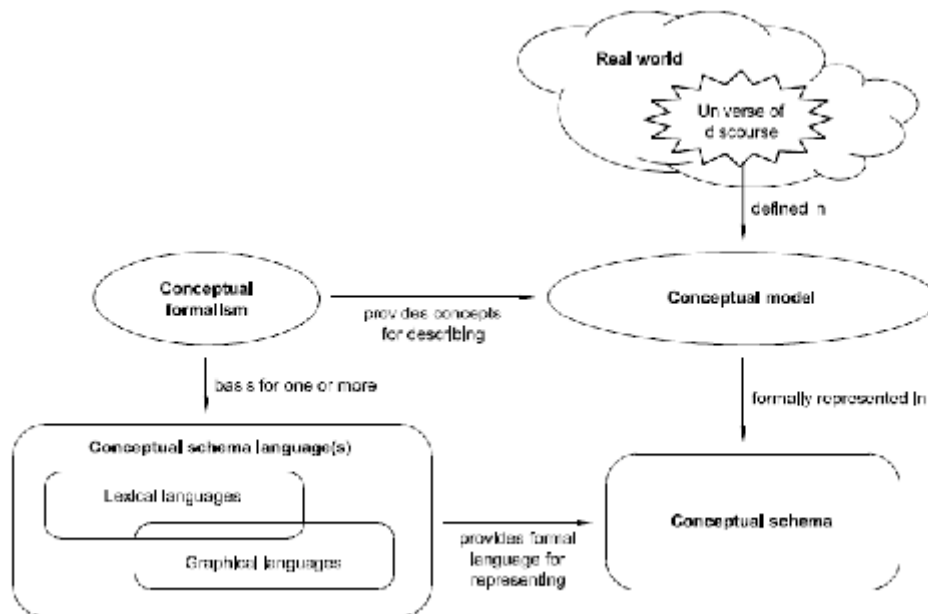


Figure 1 — The Universe of Discourse as the starting point for conceptual modelling and application schema (EN ISO 19101)

The EN ISO 19100 series of standards provide the mechanism for such a model-driven approach: the information is described by a formal, implementation-independent schema. Implementations for various techniques (e.g. XML file transfer, different types of web services²), relational database) and implementation environments (e.g. J2EE, .Net) can be derived from the schema in a more or less automatic way. Changes in information requirements are applied to the schema; never directly to the implementation. Figure 2 depicts these principles.

1) OMG, 2003. Object Management Group, Model Driven Architecture Guide Version 1.0.1. [Online] <http://www.omg.org/mda/> (last visited 2011-11-09).

2) Emphasis has been moving away from SOAP based services towards RESTful services. The latter do not require XML, SOAP, or WSDL service-API definitions.

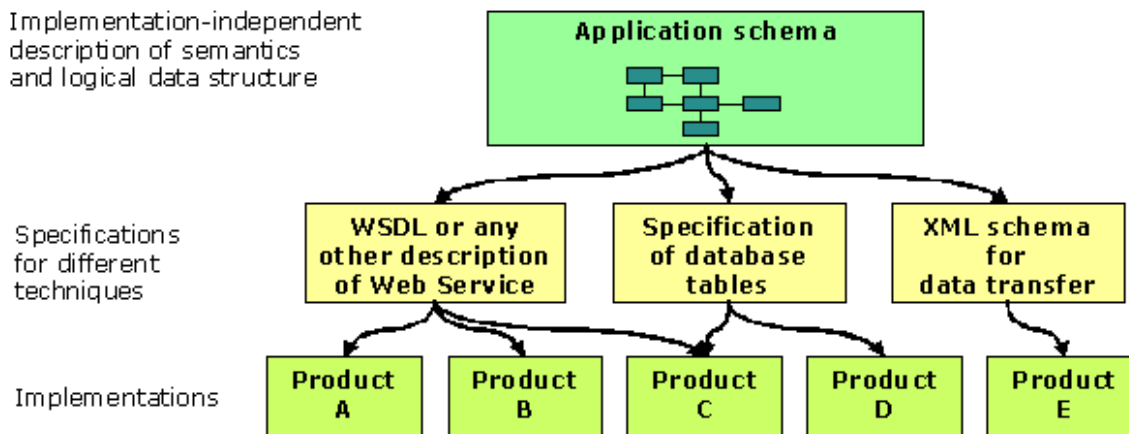


Figure 2 — The model-driven approach is promoted for SDI development

6 Aspects of data specifications

6.1 General

In this clause several aspects of data specifications are described. First, we explain the basis for data specifications, the concept of semantics and semantic interoperability. Next, we describe the different aspects which play an important role in the data specifications development:

- the conceptual schema language used to describe the application schema;
- the application schema itself;
- the features which make up these application schemas and the organisation of features in feature catalogues;
- portrayal aspects to define the way features are visualised;
- and the implementation through encoding rules.

For each of the aspects we give an overview, the relevant standards and examples and possible tools where relevant.

6.2 Semantics and semantic interoperability

In general terms, semantics relates to the meaning of words. In the context of spatial data and data specifications it relates to the meaning of spatial objects and their attributes. While the semantics refer to the content, the syntax refers to the structuring or ordering of things. In order to reach interoperability, both elements should be considered.

The information viewpoint from the ISO Reference Model for Open Distributed Processing (RM ODP³) is focusing on the semantics of information and information processing. A specification developed from this viewpoint provides a model of the information that could be used in a GIS or similar system. The information viewpoint is the most important viewpoint for the ISO 19100 series of standards (EN ISO 10101).

3) ISO/IEC – RM ODP: Reference Model for Open Distributed Processing, www.rm-odp.net/.

Semantic interoperability refers to applications and people interpreting data consistently in the same way in order to ensure they are understood as it was intended by the creator of the data. Semantic interoperability may be achieved using translators to convert data from a database to an application. The schemas and implementations described in the ISO 19100 series of standards support this level of interoperability. To achieve interoperability between heterogeneous systems two issues need to be addressed. First to define the semantics of the content and logical structure of geographic data. This should be done in an application schema. Secondly to define a system and platform independent data structure that can represent data corresponding to the application schema.

Semantics and syntactic issues become very important when spatial data are interchanged between systems. This is illustrated in Figure 3. System B has to be able to use data from system A. Data in the data transfer is structured according to the application schema C and encoded/decoded according to a standard (EN ISO 19118). M_{AC} and M_{CB} are data mappings defining how an existing schema A can be transformed to the application schema C, or vice versa how the data according to the application schema C can be transformed to an existing schema B. If the data structure of system A or B is different from that of C, such mappings may be difficult to achieve. However, if the semantics of system A or B are different from that of C, such mappings may be even impossible to achieve. This makes semantics a very important issue. It is the application schema C that defines the semantics of the content and the logical structures of the spatial data and that might make the preservation of the semantics between A and B eventually possible.

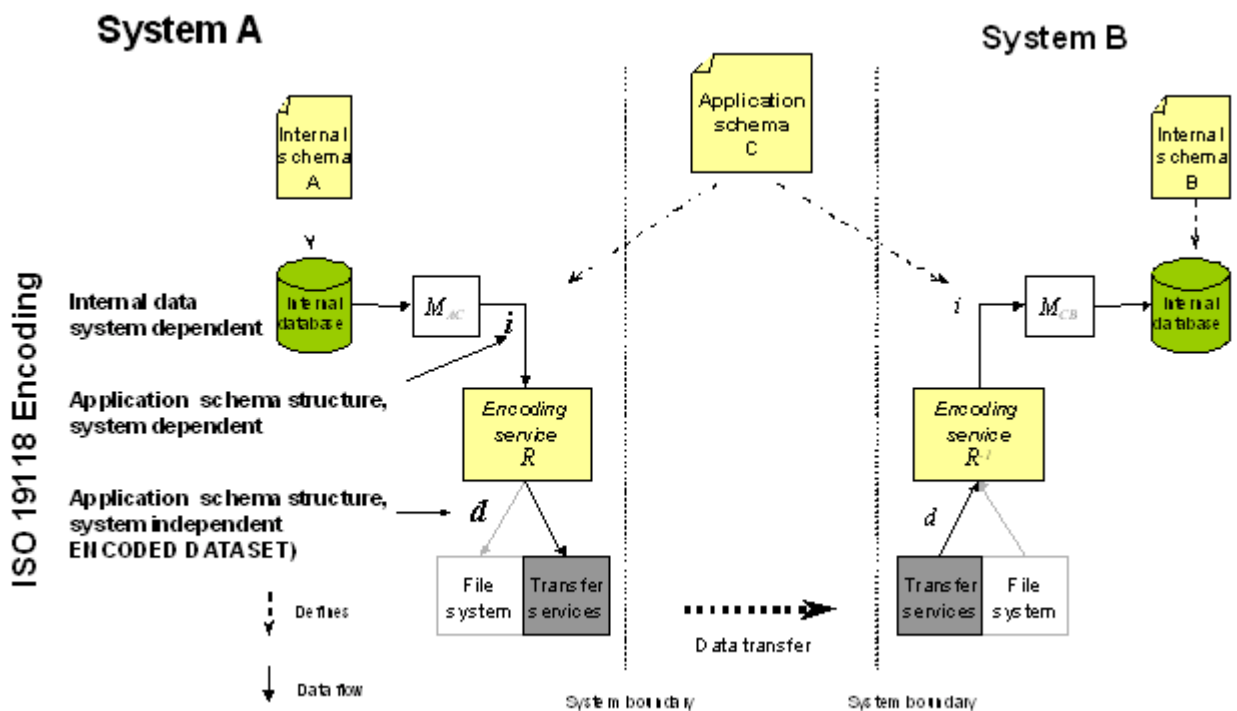


Figure 3 — Preserving semantics between systems

6.3 Conceptual schema language

6.3.1 Overview

A conceptual schema language is a formal language based on a conceptual formalism (see Figure 1). It provides the semantic and syntactic elements used to describe the conceptual model rigorously in order to convey meanings consistently. A conceptual model described using a conceptual schema language is called a conceptual schema. In order to support the goal of interoperability in the ISO 19100 series of geographic information standards, conceptual schema language should be used to develop an application schema, data interchange mechanisms and service implementations (EN ISO 19101). In this case, several platforms can be

supported. Because a conceptual schema language provides a uniform method and format for describing information, it is possible to read and update the resulting conceptual schema by computer systems as well as human beings. A conceptual schema language is based upon a conceptual formalism. The conceptual formalism provides the rules, constraints, inheritance mechanisms, events, functions, processes and other elements that make up a conceptual schema language. ISO/TS 19103 and ISO 19109 describe how conceptual schema languages are applied to create application schemas for geographic applications.

A conceptual schema language may have a clearly defined graphical notation (such as UML for class diagrams) but also a machine-readable format (such as XMI). ISO/TS 19103 specifies the requirements for conceptual schema languages. ISO/TC 211 has adopted the Unified Modelling Language (UML) as the conceptual schema language to be used. In order to define requirements and apply constraints to a UML model, the Object Constraint Language (OCL) has been developed in addition to the UML concepts.

Examples of OCL constraints are:

- the area of a building is at least 25 square metres (context Building; geometry()->area() > 25);
- the startTime of a feature should be before the endTime (context Feature; FeatureStartTime < FeatureEndTime).

A conceptual modelling language can also be used to model software and hardware. The benefit from this within the ISO geographic information standards is that the same conceptual schema language (e.g. UML) can be used for both domain models (information modelling) and for system models (service modelling). Although the ISO 19000 series has adopted UML as the conceptual schema language to be used, other languages exist.

6.3.2 Relevant standards

Relevant standards for conceptual schema language are:

- ISO/TS 19103 provides rules and guidelines for the use of a conceptual schema language. It identifies the combination of the Unified Modelling Language (UML) static structure diagram with its associated Object Constraint Language (OCL) and a set of basic type definitions as the conceptual schema language for specification of geographic information.

6.3.3 Examples and tools

Examples of conceptual schema languages are: UML; ArchiMate; Entity Relationship Diagrams (ERD); EXPRESS; INTERLIS.

6.4 Application schema

6.4.1 Overview

An application schema is a conceptual schema that defines how a universe of discourse is described as data. Application schemas contain semantics for data interpretation as well as data structures for generation of, for example, XML-schemas. The application schema provides a description of the semantic structure of the dataset. The application schema also identifies the spatial feature types, feature attribute types, feature relationship types and feature operation types. These might originate from a feature catalogue. It also defines the reference system and quality elements required to provide a complete description of geographic information in the dataset (see Figure 4). An application schema addresses the logical organisation, rather than the physical. The purpose of an application schema is:

- to provide a computer-readable data description defining the data structure, which makes it possible to apply automated mechanisms for data management; and

— to achieve a common and correct understanding of the data, by documenting the data content of the particular application field, thereby making it possible to unambiguously retrieve information from the data.

In the ISO 19100 series of standards, non-geographic features are also valid. An application schema might be described explicitly in a conceptual schema language, or might be derived from the internal structure of a software program.

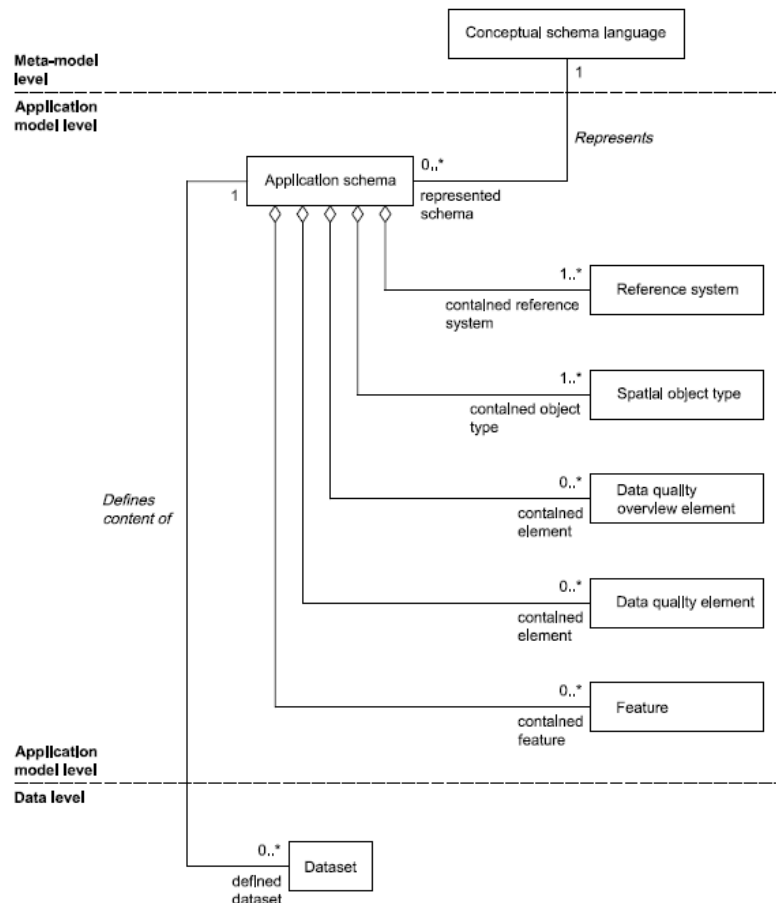


Figure 4 — Application schema and the elements it contains, and the dataset it defines

In order to facilitate automated processing of geographic datasets the elaboration of application schema is rigorously defined in the ISO 19100 series of standards. According to EN ISO 19109, the development of an application schema for a specific domain has to follow certain rules. For example, spatial aspects should be expressed according to EN ISO 19107 while temporal aspects should be expressed according to EN ISO 19108. On the other hand, there are several manners to ensure flexibility in the application schema: by using constraints on features and properties (attributes and association roles); by allowing flexible lists of possible values for an attribute; or by offering options for representing the same real-world phenomena. Figure 5 shows the main steps for implementing a schema derived from a set of specifications.

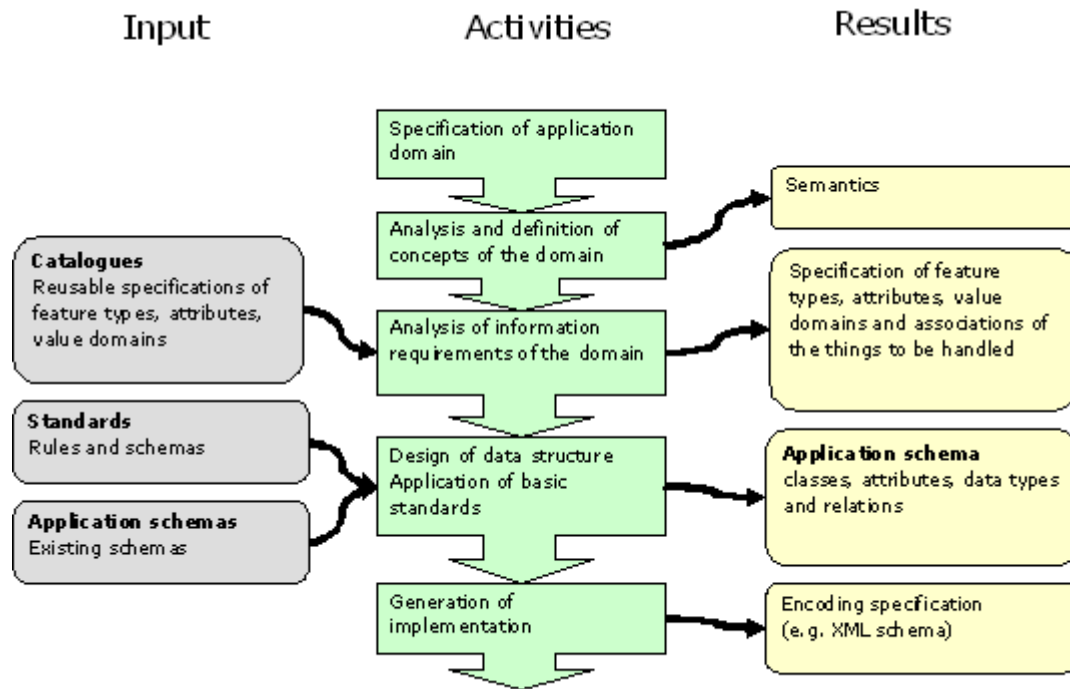


Figure 5 — From specification to application

6.4.2 Relevant standards

Relevant standards for application schema are:

- EN ISO 19107 provides conceptual schemas for describing and manipulating the spatial characteristics of geographic features. A feature is an abstraction of a real world phenomenon; it is a geographic feature if it is associated with a location relative to the Earth.
- EN ISO 19108 defines the standard concepts needed to describe the temporal characteristics of geographic information as they are abstracted from the real world. Temporal characteristics of geographic information include feature attributes, feature operations, feature associations, and metadata elements that take a value in the temporal domain.
- EN ISO 19109 defines rules for creating and documenting application schemas, including principles for the definition of features. An application schema provides the formal description of the data structure and content required by one or more applications. An application schema contains the descriptions of both geographic data and related data.
- EN ISO 19137 defines rules for a core profile of the spatial schema specified in ISO 19107 that specifies, in accordance with ISO 19106, a minimal set of geometric elements necessary for the efficient creation of application schemata.

6.4.3 Examples and tools

Example of a simple application schema – Energy resources for INSPIRE.

Example of extended application schemas – ESDIN.

6.5 Features and feature catalogues

6.5.1 Overview

A feature is an abstraction of a real world phenomena (EN ISO 19101). A feature concept is a concept that may be specified as one or more spatial object types or feature types each with a different set of properties appropriate for a particular application. Examples of feature concepts are 'road', 'river' or 'building'.

Registers are used to collect and store all the possible feature concepts, their definitions and all the information that is necessary to understand the concepts in an unambiguous way. These registers are called common feature concept dictionaries and are described in EN ISO 19126. They can serve several application schemas. A common feature concept dictionary is key to reach harmonisation across thematic fields and their corresponding application schema.

A feature catalogue contains more detailed definitions and descriptions of the spatial object types, their attributes and associated components occurring in one or more spatial datasets, together with any operations that may be applied to them. A feature catalogue is linked to an application schema. The way a feature catalogue is developed is described in EN ISO 19110. A feature catalogue might be automatically derived from an application schema and might be published via a registry service. Feature catalogues are published for the purposes of styling into a human readable presentation and access to the individual elements in the application schema.

Common feature concept dictionaries and feature catalogues are usually multi-lingual. It is believed that application schemas, feature catalogues and feature concept dictionaries promote the dissemination, sharing, and use of geographic data through providing a better understanding of the content and meaning of the data (INSPIRE, 2010). Figure 6 shows the role of the common feature concept dictionary, the feature catalogue and other data-related components of the SDI.

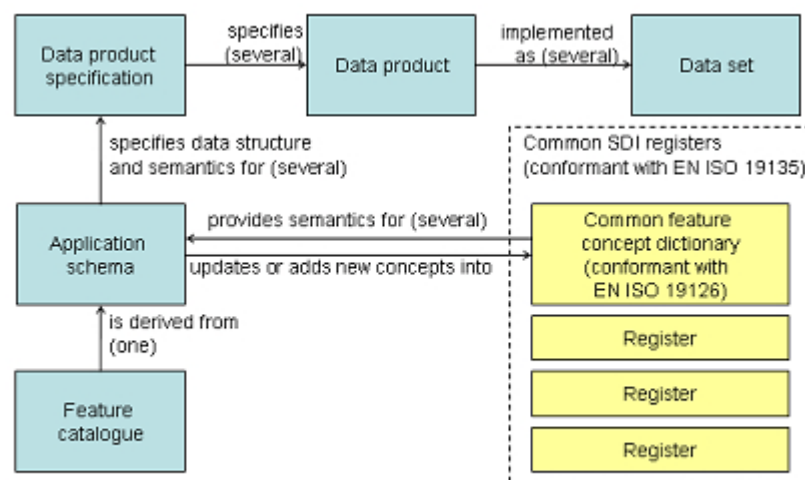


Figure 6 — A common feature concept dictionary provides semantics for application schemas of different domains

EN ISO 19125 deals with the definition and implementation of simple feature geometries which are supported by many implementations and used in many operational systems. In practice, many applications go beyond what Simple Features provides and supports. The standard consists of two parts, a common architecture part (EN ISO 19125-1), and a part on an SQL option for simple feature access (EN ISO 19125-2) which may be less relevant for SDI implementation. The GML Simple Features Profile from OGC (OGC, Geography Markup Language (GML) simple features profile) provides a widely implemented exchange standard for simple features.

6.5.2 Relevant standards

Relevant standards for features and feature catalogues are:

- EN ISO 19110 defines the methodology for cataloguing feature types. It specifies how a classification of feature types is organised into a feature catalogue and presented to the users of a set of geographic data. It applies specifically to the cataloguing of feature types that are represented in digital form but its principles can be extended to the cataloguing of other forms of geographic data. There exists an amended version: EN ISO 19110:2005/Amd-1:2011.
- EN ISO 19125-1 defines the simple feature model. It is an abstract model independent from any computer platform. Simple features are geometries restricted to two dimensions with linear interpolations between the vertices and containing spatial and non-spatial attributes.
- EN ISO 19125-2 defines a database access of simple features through an SQL interface.
- EN ISO 19126 specifies a schema for feature concept dictionaries to be established and managed as registers. It does not specify schemas for feature catalogues or for the management of feature catalogues as registers. However, because feature catalogue are often derived from feature concept dictionaries, EN ISO 19126 specifies a schema for a hierarchical register of feature concept dictionaries and feature catalogues. These registers are in accordance with EN ISO 19135.
- OGC Geography Markup Language (GML) simple features profile defines a simplified profile of GML 3.2 that supports GML features and a limited set of geometric types. A set of application schema encoding rules is defined that allow features to be encoded using GML application schemas.

6.5.3 Examples and tools

Feature catalogue of protected sites.

6.6 Portrayal

6.6.1 Overview

In general terms, portrayal refers to a representation by picture or symbols. Cartographic portrayal relates to the representation of spatial datasets. It is important as an aid to immediate user understanding. In the context of SDI, it defines the way graphic output is created for datasets and metadata of the ISO 19100 family of standards. However, the EN ISO 19117 standard does not include standardisation of cartographic symbols. The cartographic symbolisation itself is kept separate from the feature types of the dataset. The definition of the cartographic representation of a feature is stored in a portrayal catalogue⁴).

The portrayal within the context of data specifications will define all the portrayal rules for data. These rules might be defined by the thematic communities. They will clarify how standardised portrayal catalogues can be used to harmonise the portrayal of data across the different view services that will be developed on top of the datasets. In practice, the portrayal rules define how each spatial feature of a dataset will be depicted. Rules can vary from simple to more complex. For example a black solid line, 1 pixel, for a simple road; or a double red line, 3 pixels each, for a highway with two tracks in each direction with a capacity of 2 500 cars/hour per track. For rendering layers of features of a dataset, standard styles can be used. For each feature type, a default style should be available to allow rendering if the rule based search for a given feature type fails.

6.6.2 Relevant standards

Relevant standards for portrayal are:

4) Kresse, W. and Fadaie, K. (2004). ISO Standards for Geographic Information. Berlin, Germany: Springer-Verlag.

- EN ISO 19117 defines a schema describing the portrayal of geographic information in a form understandable by humans. It includes the methodology for describing symbols and mapping of the schema to an application schema. It does not include standardization of cartographic symbols, nor their geometric and functional description.

6.6.3 Examples and tools

Standardised portrayal of land cover data.

6.7 Encoding

6.7.1 Overview

In the context of an SDI, geospatial datasets are often exchanged from one organisation to another and from one system to another. The primary goal of the ISO 19100 family of standards is to reach full interoperability between different systems. Firstly, this is guaranteed by defining the semantics of the content and the logical structures of the geospatial data (see 6.2). This is achieved by implementing a common application schema (see 6.4) and defining a platform-independent data structure to represent data corresponding to the application schema. Encoding rules allow to encode geographic information defined in application schemas into a system independent data structure suitable for data transfer and storage. In practice, data transfer between systems will occur in different steps:

- the source system maps the internal data structure to the common application schema;
- an encoding service applies the encoding rules to the data (creating a new file or transfer the data to a transfer service);
- the source system invokes a transfer service;
- the destination system receives the dataset;
- the destination system applies the inverse encoding rule to interpret the data received; and
- the destination system translates the application schema to the internal data structure.

Extensible Markup Language (XML) is selected as the basis for an implementation-neutral exchange format. The transition from UML to a common XML format should be well defined, and should preferably be done by automatic code generation. EN ISO 19118 describes the requirements for creating XML encoding rules based on UML schemas, the requirements for creating encoding services, and the requirements for XML-based encoding rules. In addition to general XML, the Geography Markup Language (GML) was developed to provide a common XML encoding for spatial data (EN ISO 19136). It allows profiles that support proper subsets of GML framework descriptive capabilities and supports the description of geospatial application schemas for specialised domains and information communities. Implementers may decide to store geographic application schemas and information in GML, or they may decide to convert from some other storage format and use GML only for schema and data transport. INSPIRE has adopted GML for encoding geospatial data. The core XML schemas of GML have not been established using a model-driven approach, but it is fully consistent to establish the GML encoding of application schemas applying a model driven approach. GML 3.2 is jointly developed by ISO/TC 211 and OGC and published as EN ISO 19136. It defines a set of encoding rules between UML and XML. Those encoding rules are applicable to classes having the <<featureType>> stereotype, their properties and types (recursively).

CEN ISO/TS 19139 provides a common XML encoding for EN ISO 19115 conceptual schemas on metadata. It was designed for the model driven approach providing encoding rules for a clearly defined set of stereotypes excluding the <<feature type>> stereotype. CEN ISO/TS 19139 encoding rules can more generally be applied to any conceptual schemas related to geographic data. Even based on a model driven approach, the CEN ISO/TS 19139 method for encoding geographic data allows manual encoding designs addressing the following requirements:

- part or the whole conceptual schema may include parts of the existing XML implementation. CEN ISO/TS 19139 and the encoding of EN ISO 19107 and EN ISO 19108 defined by EN ISO 19136;
- some classes may require a custom implementation designed for efficiency or simplicity.

The specification of these manual encoding designs needs to be formalised in a (standard) specification. The XML Schema implementation of feature catalogues and data dictionaries will be addressed respectively in an amendment of EN ISO 19110 and in EN ISO 19126.

6.7.2 Relevant standards

Relevant standards for encoding are:

- EN ISO 19118 specifies the requirements for defining encoding rules based on UML schemas to be used for interchange of geographic data within the ISO 19100 series of International Standards. The standard also specifies requirements for creating encoding services and an informative XML based encoding rule for neutral interchange of geographic data.
- EN ISO 19136 defines an XML encoding schema known as Geographic Markup Language (GML). GML is in compliance with EN ISO 19118 and is used for the transport and storage of geographic information modelled according to the conceptual modelling framework used in the ISO 19100 series of International Standards and including both the spatial and non-spatial properties of geographic features. It defines the XML Schema syntax, mechanisms, and conventions. GML specifies XML encodings of a number of the conceptual classes defined in the ISO 19100 series of International Standards and the OpenGIS Abstract Specification in conformance with these standards and specifications.
- CEN ISO/TS 19139 defines Geographic MetaData XML (gmd) encoding, an XML Schema implementation derived from EN ISO 19115.

7 Data management

7.1 Accessing data

A Spatial Data Infrastructure is populated with data from different data custodians. These can be traditional data providers like National Mapping Agencies (NMA) or Cadastre Agencies. But more and more other data custodians contribute to the infrastructure: Environmental Agencies, various Ministries, but also Research Institutes or the private sector. This means that data repositories are distributed over several places and that users need to access the data remotely. In many cases, accessing data should not be based on copying the data and on local data storage, but on access through web based services. In some cases data will be discovered, viewed and even processed through these services. In other cases, the data are downloaded to have a copy of the full datasets (or parts thereof) in view of efficient processing. But also in the latter case, it is up to the services to provide data transfer, i.e. to download (parts of) the datasets and to transform them in order to make them usable within a specific environment.

Direct access to relational databases is not regarded as a part of an SDI. A relation database is regarded as a system's internal schema, and is an implementation based upon an implementation- and vendor- independent application schemas, and therefore outside the scope of an SDI.

7.2 Quality and conformity of spatial datasets

7.2.1 Overview

Quality is not to be considered in absolute terms, but in terms of user requirements. Users require information on the quality of datasets to assess whether the datasets are useful for them or not ('fitness for purpose'). Therefore, the first question is: what are the data quality user requirements? Quality levels of each spatial

dataset are defined using the criteria defined in the ISO 19100 series of standards, including completeness, consistency, currency and accuracy. This will include methods of best practice in publishing:

- 1) acceptable quality levels of each spatial dataset; and
- 2) attainment against those levels for each spatial dataset.

Data quality management requires interaction with the user on quality: what are the requirements, are there any published quality levels, how will quality be evaluated and do we need conformance testing?

Quality information associated with individual spatial objects is part of the metadata associated with the respective spatial objects (see Clause 8) and will in general be described as part of the application schema. A data specification should specify all data quality elements and sub-elements that are to be provided with the metadata of the dataset in accordance with EN ISO 19113. This should include a statement on applicable data quality measures as defined in ISO/TS 19138. EN 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. ISO/TS 19138 defines a set of data quality measures that can be used when reporting data quality for the sub-elements in EN ISO 19113.

The quality elements as defined in EN ISO 19113 are:

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

EN ISO 19113 is also applicable to data producers which should provide quality information to describe and assess how well a dataset meets the mapping of the universe of discourse as specified in the product specification.

Conformance is the fulfilment of specified requirements as described in the ISO 19100 international family of standards, or in the context of the spatial data of an SDI, the conformance with the data specifications. The conformance assessment process is the process for assessing the conformance of an implementation to an International Standard. Conformance testing is the testing of a product to determine the extent to which the product is a conforming implementation.

EN ISO 19105 provides a framework, concepts, and methodology for testing and the criteria to be achieved to claim conformance to the ISO 19100 family of International Standards. This International Standard is based in part on ISO 9646-1 which describes conformance and testing in Open Systems Interconnection (OSI), ISO 10303-31 which describes conformance and testing in industrial automation systems and integration, and ISO 10641 which describes conformance and testing for computer graphics and image processing.

Legacy datasets are datasets that are not conformant with the standards and specifications identified for an SDI. However, they are still representing great value. Metadata catalogues have to be able to store metadata of legacy datasets.

7.2.2 Relevant standards

Relevant standards for quality and conformity issues are:

- EN ISO 19113 establishes the principles for describing the quality of geographic data, or parts thereof, as well of datasets series or parts and specifies components for reporting quality information. It also provides an approach to organizing information about data quality.

- EN ISO 19114 provides a framework of procedures for determining and evaluating quality that is applicable to digital geographic datasets, consistent with the data quality principles defined in EN ISO 19113. It also establishes a framework for evaluating and reporting data quality results, either as part of data quality metadata or also as a quality evaluation report.
- ISO/TS 19138 defines a set of data quality measures. These can be used when reporting data quality for the data quality sub-elements identified in EN ISO 19113. Multiple measures are defined for each data quality sub-element, and the choice of which to use will depend on the type of data and its intended purpose.

The standards EN ISO 19113, EN ISO 19114 and ISO/TS 19138 will be replaced by ISO 19157.

- EN ISO 19105 defines two classes of conformance: class A and class B. Class A concerns conformance of specifications, including any profile or functional standard, with the series of ISO geographic information standards as a whole. Class B concerns conformance of conformance clauses as defined by this International Standard. It also provides a framework for Abstract Test Suites (ATS) to test conformity.

7.3 Spatial referencing

7.3.1 Overview

Geographic information contains spatial references which relate the features represented in the data to positions in the real world. There are two major ways of spatial referencing: by coordinates or by geographic identifiers.

A coordinate reference system (CRS) combines a coordinate system with a datum, which gives the relationship of the coordinate system to the surface and shape of the Earth. A CRS defines the coordinate space such that the coordinate values are unambiguous. The coordinate system defines a set of mathematical rules for specifying how coordinates are to be assigned to points. The position of each point in space is usually defined by coordinate tuples, for example the X, Y and Z coordinates. The datum consists of a series of parameters that define the position of the origin, the scale, and the orientation of the coordinate system.

In information systems, and especially when exchanging spatial datasets between such systems, the transformation of one coordinate reference system to another is often needed (e.g. in cross border applications). Therefore operations may be needed on spatial datasets (which contain a collection of coordinate tuples referenced to the same coordinate reference system). Coordinates can be transformed (in case of a different datum) or converted (in case of the same datum).

A map projection is coordinate conversion from an ellipsoidal coordinate system to a plane. This is applied when e.g. representing spatial data on a map or on the screen. Latitude/longitude coordinate systems reference locations directly to the surface of an ellipsoid using angles as coordinate units. In this case, there is no map projection associated. Latitude/longitude coordinates are widely used for publicly-available spatial datasets. However, the use of latitude/longitude coordinate can complicate display (distortions) and especially spatial analysis.

EN ISO 19111 defines and models coordinate reference systems and coordinate transformations and conversions. ISO/TS 19127 *Geodetic codes and parameters* provides the definitions of different Coordinate Reference Systems. In practice, the lists of geodetic codes and parameters are stored in specific registers.

The second approach is based on geographic identifiers. The position of spatial features is described through their spatial relation to other spatial features. Typical identifiers are: names, or addresses; or local measurement systems, e.g. along routes or rivers, in which the position is defined relative to a fixed point or points along the geographic feature or features; identifiers through containment, e.g. a city within an administrative unit; or even loose relationships, e.g. between landmark x and landmark y. The localisation of such positions are addressed in EN ISO 19112, *Spatial referencing by geographic identifiers*. The referencing often assumes a hierarchical relationship, e.g. country – province – municipality – postal code – address. A geographic feature in this context is called a location instance.

7.3.2 Relevant standards

Relevant standards for spatial referencing are:

- EN ISO 19111 defines the conceptual schema for the description of spatial referencing by coordinates, optionally extended to spatio-temporal referencing. It describes the minimum requirements to define one-, two- and three-dimensional spatial coordinate reference systems. It also describes the information required to change coordinates from one coordinate reference system to another. ISO 19111-2 extends the first part to include the use of parametric values or functions.
- EN ISO 19112 defines the conceptual schema for spatial references based on geographic identifiers. It establishes a general model for spatial referencing using those geographic identifiers. It defines the components of a spatial reference system, as well as the essential components of a gazetteer.
- ISO/TS 19127 defines the rules for the population and maintenance of dedicated registers of geodetic codes and parameters and identifies the data elements, in compliance with EN ISO 19135 and EN ISO 19111.

7.3.3 Examples and tools

European Petroleum Survey Group (EPSG) Geodetic Parameter Dataset (www.epsg.org/geodetic.html)

7.4 Identifier management

7.4.1 Overview

The consistent identification of geographical items is paramount in SDI implementation. It allows efficient discovery, access, management, and supply of data; it provides a mechanism for referencing thematic information to reference data, etc.

Different requirements exist for identifiers depending on the context of use. It has to be guaranteed that identifiers are distinct when processed or used within the same application. Application may use on-the-fly identification whenever necessary. Volatile identifiers can be built through chained services having a life-time equal to the request/response time. Some identifiers have a local application but in many cases a Universally Unique Identifier (UUID) is required. A UUID is an identifier that is not specific for an environment, feature type or application domain. There are various mechanisms used to guarantee that UUIDs are unique: time and node based, name based, random number based etc. Common practice is that the creation of a UUID is done through combinations of hardware addresses, time stamps and random seeds.

In general, reference to the responsible party of the data (or feature) should be part of the metadata. In a community where geographic information is distributed in the same way as music files, metadata is not always supplied. One possibility to trace an instance of a feature is a UUID within a namespace registered in a registry. If tracing is not a concern, a general UUID without any namespace should be sufficient. In other cases, e.g. in the framework of INSPIRE, where several infrastructures interact, a more pragmatic approach can be applied in which a balance of guaranteed uniqueness and traceability is envisaged. In fact, INSPIRE adopts a hierarchical approach in which registered namespaces in combination with locally unique identifiers ensure global uniqueness, rather than requiring each participant to adopt the same solution.

Many thematic datasets (e.g. cadastral parcel numbers, protected sites) have been in existence for several decades and have their own identification method which is not likely to be changed now. For example, the protected sites within the context of the European Natura 2000 program have unique identifier based on a combination of an ISO country code and unique number. Such cases need to be supported and included in any cross referencing. There should be some sort of structure in the identifiers that gives the possibility to find the object/feature in a net of servers.

7.4.2 Relevant standards

Relevant standards for identifier management are:

- EN ISO 19135 specifies procedures to be followed in establishing, maintaining and publishing registers of unique, unambiguous and permanent identifiers and meanings that are assigned to items of geographic information. It specifies elements of information that are necessary to provide identification and meaning to the registered items and to manage the registration of these items.

8 Metadata

8.1 Metadata types

8.1.1 Introduction

Metadata provides information about data on different levels: from complete datasets to information related to single features. Metadata may be stored in a catalogue referring to a dataset (e.g. for discovery), as a separate entity together with the dataset (e.g. for quality evaluation) or integrated in the dataset (e.g. for positional accuracy for a certain measurement). EN ISO 19115 specifies a schema to be used in all cases. However, in practice there are different requirements in the different use cases.

8.1.2 Discovery metadata

In an SDI, users might search for a particular dataset according to certain criteria. A user wants to know which datasets exist that match the search criteria (e.g. all the data related to soil for a cross border region between Belgium, The Netherlands and Germany), and what are the major characteristics and the conditions for use. Data discovery will, in addition to a common metadata set, require services to register and search for the datasets.

8.1.3 Feature level metadata

EN ISO 19109 defines the rules for application schema and is comprised of two categories of models that are related to metadata:

- a General Feature Model that determines the particular way metadata and quality elements relate to geographic features;
- two interchange models: the traditional data transfer model and the interoperability model; each of them implying an interrelation of the metadata with its resources.

8.1.4 Dataset metadata

With regards to the ISO 19109 interchange models, the *interoperability* model is based on data interchange by transactions and is designed for a large number of transactions involving simple interchanges. In contrast, the *transfer model* is designed for a lesser number of transactions with large amounts of well-organised data. In both cases, the interchanged data are organised in datasets having possibly their own metadata. The organisation of the datasets and their metadata is application dependent and is called the metadata application information (see 6.2 of EN ISO 19115:2005).

The transfer model may imply quite complex metadata application information and the variety of metadata application information model used by the past has been until now a curb on interoperable interchange of geospatial information. Interoperable interchanges by transfer need to go further in terms of standardization. This is the purpose of the “Extensions for metadata based transfers of geospatial information” presented in CEN ISO/TS 19139.

8.2 Relevant standards

- EN ISO 19115 defines metadata elements, provides a schema and establishes a common set of metadata terminology, definitions, and extension procedures. It defines the schema required for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data. A corrigendum exists for this European Standard: EN ISO 19115:2005/AC:2008.
- EN ISO 19115-2 extends the existing geographic metadata standard by defining the schema required for describing imagery and gridded data. It provides information about the properties of the measuring equipment used to acquire the data, the geometry of the measuring process employed by the equipment, and the production process used to digitise the raw data. This extension deals with metadata needed to describe the derivation of geographic information from raw data, including the properties of the measuring system, and the numerical methods and computational procedures used in the derivation.
- CEN ISO/TS 19139 defines Geographic metadata XML (gmd) encoding, an XML Schema implementation derived from EN ISO 19115.

8.3 Examples and tools

Metadata editor of the INSPIRE geoportal.

9 Data Product Specification

9.1 Role of a Data Product Specification

A Data Product Specification (DPS) is a detailed description of a dataset or dataset series to enable the dataset or dataset series to be created, supplied to and used by another party. It is a precise technical description of the data product in terms of the requirements that it will or may fulfil. However, the data product specification only defines how the dataset should be. For various reasons, compromises may need to be made in the implementation. The metadata associated with the product dataset should reflect how the dataset actually is. A data product specification may be created and used on different occasions, by different parties and for different reasons. It may, for example, be used for the original process of collecting data as well as for products derived from already existing data. It may be created by producers to specify their product or by users to state their requirements (EN ISO 19131).

In the case of INSPIRE, DPS (termed data specification within that context) have been created for the 34 themes of the 3 annexes of the Directive. Countries should transform or harmonise their existing datasets to match them with the specifications as described and required by the INSPIRE Directive and its implementing rules. The mapping of existing datasets to the common application schema might be done by using transformation services (on the fly or not), by using interactive tools available on the market or by re-engineering the existing datasets. The DPS will also be used in practice to create new datasets or datasets series that match those requirements.

9.2 Stepwise approach

9.2.1 General

In order to define the data product specification for a specific domain a stepwise approach should be followed. The following steps are defined based on the experience of INSPIRE. Figure 7 gives a schematic view of these steps.

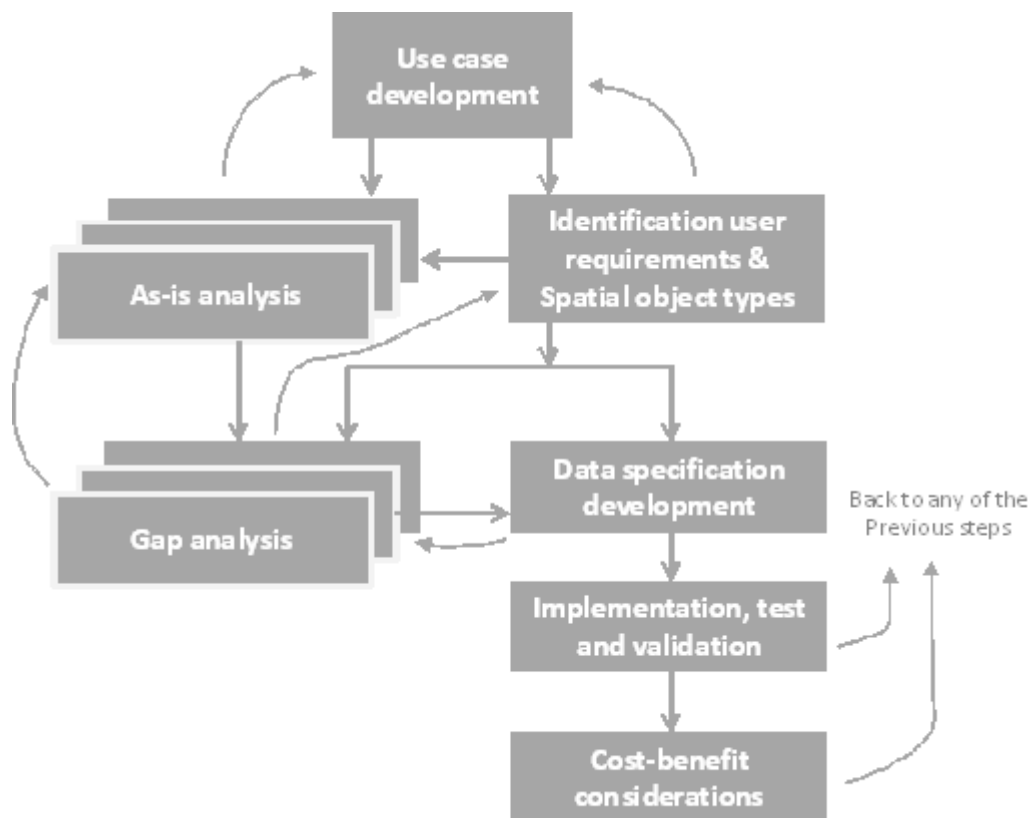


Figure 7 — Possible steps for data specifications development

It should be emphasised that these steps are based on the experience of INSPIRE⁵⁾. Depending on specific situations not all steps mentioned are necessarily to be followed, and not all the steps are of equal importance. First, some steps might run in parallel: e.g. use case and user requirements identification. Second, the development of use cases, the identification of user requirements and the development and testing of the specifications are certainly key to the whole process. Thirdly, some steps might be omitted because they were already performed before and are thus described somewhere else (e.g. ‘as-is’ analysis), or are less relevant all together (e.g. cost benefit considerations).

9.2.2 Step 1 – Use case development

Use cases and application scenarios can be described for any user domain and user community. The use cases are described in sufficient detail to clarify the data requirements regarding the data for a particular thematic field. These use cases may be described according to a template for use case descriptions and might include following elements: the name & description of the use case, the priority for the community, the pre- and post-conditions, the flow of events, and the data sources needed including their description, provider, geographic & thematic scope, scale or resolution, delivery and any other documentation.

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 language using the vocabulary of the domain.

5) Based on INSPIRE - D2.6: Methodology for the development of data specifications. [Online] http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/DataSpecifications/D2.6_v3.0.pdf

9.2.3 Step 2 – Identification of the user requirements and spatial object types

The requirements regarding the data of a particular domain are extracted from the use cases and application scenarios. This includes the identification of the required levels of detail. The result is a description of the relevant universe of discourse for that domain. In this step, a candidate list of spatial object types with definitions and descriptions is established as well. It is recommended to capture the understanding of the scope of the domain 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.

In this step it is important to take into consideration all the data interoperability aspects:

- the reference model (EN ISO 19101);
- terminology;
- application schema & feature catalogues;
- spatial & temporal aspects;
- multi-lingual aspects;
- coordinate referencing;
- units of measurement;
- object referencing;
- metadata;
- portrayal;
- ID-management;
- registers;
- quality issues;
- maintenance;
- data capturing;
- data transfer;
- multiple representations;
- consistency and conformity checks.

It might be necessary to add use cases or to modify existing ones.

9.2.4 Step 3 – As-is analysis

The as-is analysis of the current situation focus on existing spatial datasets for a specific domain. It can be carried out in parallel to step 2. A checklist with the different topics as described in step 2 can assist in identifying the relevant data interoperability aspects. The result provides the basis for the next step, the gap analysis.

The as-is analysis can be performed on existing spatial datasets, catalogues or thesauri, but also on descriptions of models on paper or in UML format. This step should be done in cooperation with experts within the thematic domain(s) which is (are) under investigation.

9.2.5 Step 4 – Gap analysis

The gap analysis identifies user requirements that cannot be met by the current data offerings: e.g. missing feature types or attributes, but also differences in reference systems, quality, etc. 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. This analysis compares the results of each as-is analysis with the first-cut data specification, and evaluates if the identified datasets are close enough to the data specification. A similar checklist as in step 3 can be used to do so. It also identifies how to extract information from these data sources into the application schema.

The resulting document contains information on the identified gaps. Most of the gaps might be solved by extracting information from the data sources by transformation tools. Some gaps will need greater efforts than transformation. If the identified source dataset is far from the first-cut data specification a new as-is analysis or a change of the user requirements may be required. Alternatively new datasets might be developed, or, if feasible existing datasets will be re-engineered (e.g. by adding new codes or classifications).

9.2.6 Step 5 – Data Specification Development

For each thematic domain, one or more application schemas can be developed with the spatial object (feature) types with their properties, range of valid property values and constraints. The data specification itself will be documented according to EN ISO 19131, the International Standard specifying the contents of data product specifications in the field of geographic information. The application schema, accompanied by a corresponding feature catalogue derived from the application schema, constitutes the core component of the data specification. It includes a schema in UML as well as the corresponding feature catalogue and GML application schema. The specification of the valid encoding(s) of spatial data for a domain is an integral part of each data specification. A data product specification is the basis for the development or transformation of data products that might be implemented as one or more datasets which are accompanied by metadata (see Figure 8).

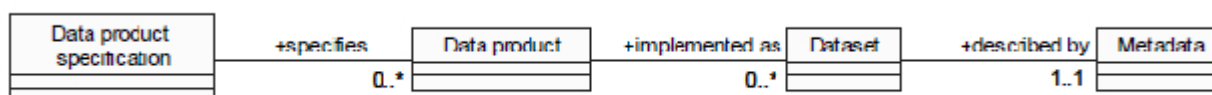


Figure 8 — From data product specification to dataset including metadata

9.2.7 Step 6 – Implementation, test and validation

The resulting data specification should be evaluated by the stakeholders of the specific domain and tested within one or more pilots under real world conditions using the use cases developed in step 1 to test the proposed specification for consistency, completeness and feasibility. Testing can be done in different ways, but should always happen on real datasets and in real applications. The result of this step is a documented report with detailed results of the testing.

9.2.8 Step 7 – Cost-benefit analysis

During the implementation, testing and validation the estimation of costs and benefits is an important step. It entails identifying cost and benefits related to the feasibility of transforming existing datasets to the defined data specifications. It includes aspects such as understanding of the specifications, training needed, data collection, attribute mapping, setting up of transformation infrastructure, execution of transformation, documentation and management & coordination. Also the fitness for purpose should be assessed. This

includes the costs and benefits for the user, the social value, the institutional operational benefits, the institutional financial value and the strategic and political value.

9.3 Content of a Data Product Specification

A Data Product Specification (DPS) defines the requirements for a data product. In clause 6 we described several aspects that DPS are dealing with. It forms the basis for producing or acquiring data. DPS will help potential users to evaluate the data product to determine its fitness for use by them. The information contained in a DPS is different from the information contained in metadata, which provides information about a particular physical dataset. Thus, metadata describes how a dataset actually is, whilst a data product specification describes how it should be.

A DPS document should cover at least following aspects of data products:

- Overview;
- Specification scopes;
- Data product identification;
- Data content and structure;
- Reference systems;
- Data quality;
- Data product delivery;
- Metadata.

In addition, the DPS could also cover following aspects:

- Data capture;
- Data maintenance;
- Portrayal;
- Additional information.

9.4 Relevant standards

- EN ISO 19131 specifies requirements for the specification of geographic data products, based upon the concepts of other ISO 19100 International Standards. It also provides help in the creation of data product specifications, so that they are easily understood and fit for their intended purpose.

9.5 Examples and tools

EuroRegionalMap

Bibliography

The following is a list of international and European standards that are relevant for this Part of CEN/TR 15449:

- [1] EN ISO 19101:2005, *Geographic information — Reference model (ISO 19101:2002)*
- [2] ISO/TS 19103:2005, *Geographic information — Conceptual schema language*
- [3] EN ISO 19105:2005, *Geographic information — Conformance and testing (ISO 19105:2000)*
- [4] EN ISO 19107:2003, *Geographic information — Spatial schema (ISO 19107:2003)*
- [5] EN ISO 19108:2005, *Geographic information — Temporal schema (ISO 19108:2002)*
- [6] EN ISO 19109:2006, *Geographic information — Rules for application schema (ISO 19109:2005)*
- [7] EN ISO 19110:2006, *Geographic information — Methodology for feature cataloguing (ISO 19110:2005)*
- [8] EN ISO 19111:2007, *Geographic information — Spatial referencing by coordinates (ISO 19111:2007)*
- [9] EN ISO 19111-2:2012, *Geographic information — Spatial referencing by coordinates — Part 2: Extensions for parametric values (ISO 19111-2:2009)*
- [10] EN ISO 19112:2005, *Geographic information — Spatial referencing by geographic identifiers (ISO 19112:2003)*
- [11] EN ISO 19113:2005, *Geographic information — Quality principles (ISO 19113:2002)*
- [12] EN ISO 19114:2005, *Geographic information — Quality evaluation procedures (ISO 19114:2003)*
- [13] EN ISO 19115:2005, *Geographic information — Metadata (ISO 19115:2003)*
- [14] EN ISO 19115-2:2010, *Geographic information — Metadata — Part 2: Extensions for imagery and gridded data (ISO 19115-2:2009)*
- [15] EN ISO 19117:2006, *Geographic information — Portrayal (ISO 19117:2005)*
- [16] EN ISO 19118:2011, *Geographic information — Encoding (ISO 19118:2011)*
- [17] EN ISO 19125-1:2006, *Geographic information — Simple feature access — Part 1: Common architecture (ISO 19125-1:2004)*
- [18] EN ISO 19125-2:2006, *Geographic information — Simple feature access — Part 2: SQL option (ISO 19125-2:2004)*
- [19] EN ISO 19126:2009, *Geographic information — Feature concept dictionaries and registers (ISO 19126:2009)*
- [20] ISO/TS 19127:2005, *Geographic information — Geodetic codes and parameters*
- [21] EN ISO 19131:2008, *Geographic information — Data product specifications (ISO 19131:2007)*
- [22] EN ISO 19135:2007, *Geographic information — Procedures for item registration (ISO 19135:2005)*

[23] EN ISO 19136:2009, *Geographic information — Geography Markup Language (GML) (ISO 19136:2007)*

[24] ISO/TS 19138:2006, *Geographic information — Data quality measures*

[25] CEN ISO/TS 19139:2009, *Geographic information — Metadata — XML schema implementation (ISO/TS 19139:2007)*

British Standards Institution (BSI)

BSI is the national body responsible for preparing British Standards and other standards-related publications, information and services.

BSI is incorporated by Royal Charter. British Standards and other standardization products are published by BSI Standards Limited.

About us

We bring together business, industry, government, consumers, innovators and others to shape their combined experience and expertise into standards-based solutions.

The knowledge embodied in our standards has been carefully assembled in a dependable format and refined through our open consultation process. Organizations of all sizes and across all sectors choose standards to help them achieve their goals.

Information on standards

We can provide you with the knowledge that your organization needs to succeed. Find out more about British Standards by visiting our website at bsigroup.com/standards or contacting our Customer Services team or Knowledge Centre.

Buying standards

You can buy and download PDF versions of BSI publications, including British and adopted European and international standards, through our website at bsigroup.com/shop, where hard copies can also be purchased.

If you need international and foreign standards from other Standards Development Organizations, hard copies can be ordered from our Customer Services team.

Subscriptions

Our range of subscription services are designed to make using standards easier for you. For further information on our subscription products go to bsigroup.com/subscriptions.

With **British Standards Online (BSOL)** you'll have instant access to over 55,000 British and adopted European and international standards from your desktop. It's available 24/7 and is refreshed daily so you'll always be up to date.

You can keep in touch with standards developments and receive substantial discounts on the purchase price of standards, both in single copy and subscription format, by becoming a **BSI Subscribing Member**.

PLUS is an updating service exclusive to BSI Subscribing Members. You will automatically receive the latest hard copy of your standards when they're revised or replaced.

To find out more about becoming a BSI Subscribing Member and the benefits of membership, please visit bsigroup.com/shop.

With a **Multi-User Network Licence (MUNL)** you are able to host standards publications on your intranet. Licences can cover as few or as many users as you wish. With updates supplied as soon as they're available, you can be sure your documentation is current. For further information, email bsmusales@bsigroup.com.

BSI Group Headquarters

389 Chiswick High Road London W4 4AL UK

Revisions

Our British Standards and other publications are updated by amendment or revision.

We continually improve the quality of our products and services to benefit your business. If you find an inaccuracy or ambiguity within a British Standard or other BSI publication please inform the Knowledge Centre.

Copyright

All the data, software and documentation set out in all British Standards and other BSI publications are the property of and copyrighted by BSI, or some person or entity that owns copyright in the information used (such as the international standardization bodies) and has formally licensed such information to BSI for commercial publication and use. Except as permitted under the Copyright, Designs and Patents Act 1988 no extract may be reproduced, stored in a retrieval system or transmitted in any form or by any means – electronic, photocopying, recording or otherwise – without prior written permission from BSI. Details and advice can be obtained from the Copyright & Licensing Department.

Useful Contacts:

Customer Services

Tel: +44 845 086 9001

Email (orders): orders@bsigroup.com

Email (enquiries): cservices@bsigroup.com

Subscriptions

Tel: +44 845 086 9001

Email: subscriptions@bsigroup.com

Knowledge Centre

Tel: +44 20 8996 7004

Email: knowledgecentre@bsigroup.com

Copyright & Licensing

Tel: +44 20 8996 7070

Email: copyright@bsigroup.com



...making excellence a habit.™