European
Commission

JRC REFERENCE REPORTS

A Conceptual Model for Developing Interoperability Specifications in Spatial Data Infrastructures

Katalin Tóth, Clemens Portele, Andreas Illert,
Michael Lutz, Maria Nunes de Lima

2012

Report EUR 25280 EN

Joint
Research
Centre

European Commission
Joint Research Centre
Institute for Environment and Sustainability

Contact information
Katalin Tóth
Address: Joint Research Centre, Via Enrico Fermi 2749, TP 262, 21027 Ispra (VA), Italy
E-mail: katalin.toth@jrc.ec.europa.eu
Tel.: +39 0332 78 6491
Fax: +39 0332 78 6325

http://ies.jrc.ec.europa.eu
http://www.jrc.ec.europa.eu

This publication is a Reference Report by the Joint Research Centre
of the European Commission.

JRC69484

EUR 25280 EN

ISBN 978-92-79-22552-9 (pdf)
ISBN 978-92-79-22551-2 (print)

ISSN 1018-5593 (print)
ISSN 1831-9424 (online)

doi:10.2788/21003

Luxembourg: Publications Office of the European Union, 2012

Printed in Italy

A Conceptual Model for Developing Interoperability Specifications in Spatial Data Infrastructures

Tóth, Katalin [1]
Portele, Clemens [2]
Illert, Andreas [3]
Lutz, Michael [1]
Nunes de Lima, Vanda [1]

[1]European Commission, Joint Research Centre, [2]interactive instruments Gesellschaft für Software-Entwicklung mbH, [3] Bundesamt für Kartographie und Geodäsie

Keywords: spatial data infrastructure, interoperability, generic conceptual model, data specification development

Executive summary

Today, geographic information is being collected, processed, and used in domains as diverse as hydrology, disaster mitigation, statistics, public health, geology, civil protection, agriculture, nature conservation, and many others. The challenges regarding the lack of availability, quality, organisation, accessibility, and sharing of spatial information are common to a large number of policies and activities, and are experienced across the various levels of public authority in Europe.

Directive 2007/2/EC of the European Parliament and of the Council, adopted on 14 March 2007, takes measures to address these challenges by establishing an Infrastructure for Spatial Information in the European Community (**INSPIRE**) for environmental policies, or policies and activities that have an impact on the environment. Moreover, Spatial Data Infrastructures (SDIs) are becoming more and more linked to and integrated with systems developed in the context of e-Government. An important driver of this evolution is the Digital Agenda for Europe, which recommends "establishing a list of common cross-border services that allow businesses and citizens to operate independently or live anywhere in the EU" and "setting up systems of mutual recognition of electronic identities"[1].

This report addresses the question of how the reuse of geographic and environmental information created and maintained by different organisations in Europe can be enabled and facilitated. The main challenge related to this task is how to deal with the heterogeneity of data and how to establish information flow between communities that use geographic information in various environmental fields.

This report presents an integrated view of the data component of SDIs, highlighting the main features of the conceptual framework. We expect this document to be useful to

[1] http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/10/200&format=HTML&aged=0&language=EN&guiLanguage=en

- Decision makers responsible for the strategic development of SDIs who need to understand the benefits of using a conceptual framework and need to assess the complexity and the resources associated with this work,
- Leading civil servants from the Member State organisations that are legally mandated to implement INSPIRE,
- Scientists looking for a quick and comprehensive overview of the key elements of the data component in SDIs.

Section 1 introduces Spatial Data Infrastructures (SDIs) and how they have developed as a logical consequence of technological advances and the associated societal and technological challenges. With the development of information and communications technology, traditional paper maps have been replaced by digital geographic information and location-based services. This new digital technology could facilitate the reuse of geographic information, but is hampered by incomplete documentation, lack of compatibility among the spatial datasets, inconsistencies of data collection, and cultural, linguistic, financial and organisational barriers. SDIs propose organisational and technical measures to search, find, and reuse spatial data collected by other organisations.

One of the core concepts of SDIs is **interoperability**, which "means the possibility for spatial datasets 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 datasets and services is enhanced"[2]. INSPIRE, which is used as the main SDI initiative from which this report draws its examples and best practices, is built on the existing standards, information systems and infrastructures, professional and cultural practices of the 27 Member States of the European Union in all the 23 official and possibly also the minority languages of the EU.

Section 2 focuses on geographic information and details the **challenges and inconsistencies** that SDI users may face when trying to combine or reuse data retrieved from diverse sources. These challenges are ultimately rooted in the diversity of how geographic data is defined as a partial abstraction of reality. Geographic data, like any data, is always an abstraction, always partial, and always just one of many possible views. As a consequence, rivers may be represented as polygons in one dataset and as lines in another, the lines representing roads on both sides of a national border may not meet, and water may appear to flow uphill when combining a hydrological and an elevation dataset. These and further challenges of data reuse in SDIs are illustrated and explained in this section.

The main part of the report is found in **section 3, 4, and 5**, which describe the framework for the development of data specifications that address a number of the challenges described above. These specifications define the interoperability targets and how existing data should be transformed in order to meet these targets. Section 3 is split into two main parts, both of which largely build on INSPIRE experiences and best practices:

- The Generic Conceptual Model (GCM) defines 25 aspects or elements relevant to achieving data interoperability in an SDI, and proposes methods and tools to

[2] Art. 3(7) of Directive 2007/2/EC (INSPIRE)

address them. These include, for example, registries, coordinate reference systems, identifier management, metadata and maintenance, to name just a few.

- The description of the methodology for developing data specifications for interoperability includes a detailed discussion of the relevant actors, steps and the overall workflow – from collecting user requirements to documenting and testing the specifications that emerge from this process.

Together, both subsections explain the organisational and technical aspects of how the data component of an SDI can be established, and how interoperability arrangements, data standardisation and harmonisation contribute to this process. Since 2005, INSPIRE has been pioneering the introduction, development, and application of a conceptual framework for establishing the data component of an SDI. This experience shows that the conceptual framework described in this report is robust enough to reinforce interoperability across the 34 data specifications developed for the SDI. Moreover, because the framework is platform- and theme independent, can deal with cultural diversity, and is based on best practice examples from Europe and beyond, it may also provide solutions for SDI challenges in other environments.

Contents

Glossary

AFIS	Amtliches Festpunktinformationssystem (Official Fixed Point Information System)
ALKIS	Amtliches Liegenschaftskataster Informationssystem (Official Real Estate Cadastre Information System)
ATKIS	Amtliches Topographisch-Kartographisches Informationssystem (Official Topographic Cartographic Information System)
ATS	Abstract Test Suite
BAG	Bathymetry Attributed Grid
DNF	Digital National Framework
EC	European Commission
EU	European Union
GBIF	The Global Biodiversity Information Facility
GEOSS	Global Earth Observation System of Systems
GCM	Generic Conceptual Model
GIS	Geographic Information Systems
GML	Geography Markup Language
HY	Hydrography, hydrology
ICAO	International Civil Aviation Organisation
INSPIRE	Infrastructure for Spatial Information in Europe
ISO	International Standards Organisation
KML	Keyhole Markup Language
NUTS	Nomenclature des unités territoriales statistiques (Nomenclature of territorial units for statistics)
OGC	Open Geospatial Consortium
OWL	Ontology Web Language
SDI	Spatial Data Infrastructure
SI	Système international d'unités (International system of units)
SLD	Styled Layer Descriptor
SKOS	Simple Knowledge Organization System
SDI	Global Spatial Data Infrastructure
TAPIR	Taxonomic Databases Working Group Access Protocol for Information Retrieval
TC	Technical Committee
THREDDS	Thematic Real-time Environmental Distributed Data Services
TIFF	Tagged Image File Format
UK	United Kingdom
UML	Unified Modelling Language
UTC	Universal Time Coordinates
XML	Extensible Markup Language
WMS	Web Mapping Service

Foreword

Geographic information, spatial data infrastructures (SDIs), interoperability, and shared information systems are notions that developers in information and communications technology, decision makers responsible for public sector information, as well as scientists, engineers, and public servants may come across daily – whether they are working in domains such as hydrology, disaster mitigation, statistics, public health, geology, civil protection, agriculture, nature conservation, or one of many other disciplines.

Should they be concerned? Do they have an easy way to respond to the challenge of reading the ever growing, scattered, and sometimes highly technical documentation? Is it possible to understand the core ideas without an insight into policies, organisational aspects, workflows, and without prior knowledge of the subject matter and related technology?

While the answer to the first question is a definite 'yes', for most people it is probably 'no' for the other two. This report tries to address these questions by explaining the basic concepts and principles, summarising what interoperability means for the domain of geographic information and showing how SDIs can be key to solve the associated challenges. All of this will be explained from point of view of spatial data, touching upon the other components of SDIs[3] only to illustrate the connections.

Readers that are familiar with the concept of SDIs may ask why such special attention is being devoted to the data component, when this is, perhaps, the component for which achieving interoperability is the most difficult. We list only a few reasons:

- The data component is the best for setting the scene as to why interoperability is needed,
- Spatial data is an asset that has been accumulated over a long period of time by many different organisations. They are rightfully concerned by the impact of SDIs on their work. An understanding of the spirit of interoperability can help clarify potential misunderstandings,
- Current users of geographic information spend 80 percent of their time collating and managing the information and only 20 percent analysing it to solve problems and generate benefits (Geographic Information Panel (2008).
- Human psychology: the name "Spatial *Data* Infrastructure" implies the data subject.

There are many SDI initiatives across the world. The authors, all actively involved in INSPIRE, inevitably take most references from this initiative. Nevertheless, they try to emphasise those features of INSPIRE that are likely to be valid in other environments, complementing them with references to other initiatives.

The main objectives of this report is to explain the aspects of the framework necessary for development of information models and interoperability specifications in SDIs without going too deep into technicalities; allowing an "informed policy maker" possessing everyday IT literacy skills to understand them. In order to help the readers, basic definitions are given in green, while examples are given in light brown boxes.

[3] The definition and short description of SDIs will be given in section 1.1

1 Spatial Data Infrastructures – Setting the scene

1.1 From maps to Spatial Data Infrastructures

Facts, stand-alone bits of data and pieces of information, however accurate they may be, can never achieve the same effect as when they are put in a context of time and spaces, which are the most frequently used data references.

For thousands years of spatial observations[4], the final products of this effort were maps which graphically presented the spatial context (Klinghammer, I. (1995). Ancient maps were used to accomplish the most important missions of the state: navigation, discovery and colonisation of new territories, taxation, warfare, etc. Possession of maps brought with it the power to monopolise and gain luxuries. After the diffusion of modern typography, some popular products such as city, road, and tourist maps, and geographical atlases became more widely used.

However the majority of maps remained accessible to specialists only. Each type of map followed its own production line and thematic scope. The reuse of these maps was limited. Only topographic maps found wider diffusion as they gave general descriptions of the surface of the Earth and provided a geometrical basis for thematic mapping.

> Spatial analysis is the process of extracting or deriving new information by modelling, assessing, understanding and evaluating natural and social phenomena in the context of a geographic location.

With the development of information and communications technology, traditional paper maps have been gradually replaced by digital geographic information from map digitisation, Earth observation satellites, in-situ digital sensors and global positioning systems. Paper maps are still used for visualisation, but computers and other hardware[5] have become the main arena for spatial analysis, engineering design, and location-based services.

Geographic Information Systems (GIS) are integrated collections of computer software and data used to view and manage geographic information in order to analyse spatial relationships and to model spatial processes (Wade, T. and Sommer, S. (editors) 2006). The early implementations of GIS somewhat repeated the steps followed by analogue data processing, using data that was collected explicitly for the specific task to be solved and thereby missing out on benefiting from the potential reuse of digital data.

The diffusion of the Internet and widespread computer literacy have opened a genuinely new paradigm in spatial data handling, promoting data sharing across different communities and various applications. The frameworks for data sharing are the Spatial Data Infrastructures (SDIs)[6] that can be interpreted as extensions of a

---

[4]Cartographic science goes back as far Eratosthenes and Ptolemy.

[5] Personal and portable computers, mobile phones and specific devices such as those used for navigation offer applications based on spatial data.

[6] Sometimes, SDIs are also referred to as "spatial *information* infrastructures" to highlight the fact that they usually provide access to data through (value-added) services. However, we use the more widely established term "spatial data infrastructure" in this report.

desktop GIS (Craglia, M. (2010), where data collected by other organisations can be searched, retrieved and used according to well-defined access policies.

According to the Global Spatial Data Infrastructure (GSDI) Association's Cookbook (Nebert, D. D. (editor) 2004) "an SDI hosts geographic data and attributes, sufficient documentation (metadata), a means to discover, visualize, and evaluate the data (catalogues and Web mapping), and some method to provide access to the geographic data. Beyond this are additional services or software to support applications of the data. To make an SDI functional, it must also include the organisational agreements needed to coordinate and administer it on a local, regional, national, and or trans-national scale".

The description of GSDI classifies SDI components as data, metadata, services (technology), and organisational agreements. According to Craglia et al. (2003), "Spatial Data Infrastructures (SDIs) encapsulate policies, institutional and legal arrangements, technologies, and data that enable sharing and effective usage of geographic information". This definition adds an aspect of utmost importance – the effective usage of geographic data, which sets the requirement of interoperability.

The degree of SDI development strongly correlates with the development of the information society in general, use of information technology by the population, and the diffusion of the Internet. An SDI can be established at global, supranational, national, regional, cross-border, or local levels. In the ideal case, these levels are interconnected, accommodating each other's relevant components.

1.2 Examples of SDI initiatives

The establishment of an SDI requires the collaboration of many parties. This collaboration can be based on voluntary agreements between the interested parties, or it can be more formally regulated, or even legally enforced, mandating the targeted organisations to fulfil the provisions of legal acts. Voluntary initiatives, such as GSDI and some national SDIs, are often coordinated by international and national associations or umbrella organisations.

According to Longley at al. (2011) there are over 150 SDI initiatives described in the literature. The following examples mention only those that are referred to in context of this report. Two of these initiatives are established at the global level, one at the national level, and one at the supranational level in the European Union.

GSDI

The Global Spatial Data Infrastructure Association was founded in 1998 to "promote international cooperation and collaboration in support of local, national and international spatial data infrastructure developments that will allow nations to better address social, economic, and environmental issues of pressing importance"[7]. As an international voluntary organisation, the GSDI does not aim to establish a global spatial infrastructure, but rather focuses on raising awareness and exchanging best practice examples.

[7] http://www.gsdi.org/

GEOSS

The Global Earth Observation System of Systems aims to provide decision-support tools to a wide variety of users. As a "system of systems", GEOSS is based on existing observation, data processing, data exchange and dissemination systems, and includes in situ, airborne, and space-based observations. In order to reach interoperability, information and data providers are expected to adopt a necessary level of coordination and technical arrangements which include specifications for collecting, processing, storing, and disseminating shared data, metadata, and products.

Interoperability in GEOSS focuses on interfaces so as to minimise any impact on the component systems. As part of its 10-year implementation plan (2005), GEOSS draws on existing spatial data infrastructure components in areas such as geodetic reference frames, common geographic data, and standard protocols. The thematic scope of GEOSS covers the 'Societal Benefit Areas' related to Disasters, Health, Energy, Climate, Agriculture, Ecosystems, Biodiversity, Water and Weather.

UK Location Strategy

The UK Location Strategy was launched in 2008. It aims to "maximise exploitation and benefit to the public, the government and to UK Industry from geographic information and to provide a framework to assist European, national, regional and local initiatives. The Strategy will create an infrastructure for location information to assist policy, service delivery and operational decision making" (Geographic Information Panel 2008).

The strategy document provides for a gallery whereby local information is applied to public policy and strategic actions are proposed for better use of geographic information. It also defines a small number of key datasets (Core Reference Geographies), which will form common information frameworks that are defined, endorsed, and used by all data holders in both the public and private sectors. The Core Reference Geographies contain Geodetic frameworks (including ground height information), Geographic names, Addresses, Streets, Land and property ownership, Hydrology/Hydrography, Statistical boundaries, and Administrative boundaries. In frame of the Location Strategy, the Digital National Framework (DNF) has been defined as the mechanism for integrating and sharing location-based UK information from multiple sources.

INSPIRE

INSPIRE is a prominent example of a legally enforced infrastructure. The INSPIRE Directive of the European Parliament and the Council (2007/2/EC of 14 March 2007) sets up an infrastructure for spatial information in Europe to support environmental policies or activities that may have an impact on the environment.

According to Craglia (2011), INSPIRE has some characteristics that make it particularly challenging:

1. The infrastructure is built on those of 27 Member States of the European Union in more than 23 languages[8]. This requires the coexistence and collaboration of very different information systems, professional and cultural practices,

[8] 23 official languages of the EU as well as minority languages.

2. Given this complexity, it was necessary to adopt a consensus-building process, involving hundreds of national experts, to develop the technical specifications for INSPIRE,
3. Existing standards must be tested in real distributed and multilingual settings,
4. Standards that are not mature enough, or leave too much room for different interpretation (because of the legally mandated implementation) have to be refined,
5. Standards which do not yet exist must be developed,[9]
6. Inconsistency and incompatibility of data and metadata must be addressed for the 34 themes that fall within the scope of the Directive (see Table 1).

The data themes of INSPIRE are divided in modular blocks. “Annexes I and II focus on reference data, while Annex III focuses on data for environmental analysis and impact assessment.

Annex I	Annex III
1. Coordinate reference systems	14. Statistical units
2. Geographical grid systems	15. Buildings
3. Geographical names	16. Soil
4. Administrative units	17. Land use
5. Addresses	18. Human health and safety
6. Cadastral parcels	19. Utility and governmental services
7. Transport networks	20. Environmental monitoring facilities
8. Hydrography	21. Production and industrial facilities
9. Protected sites	22. Agricultural and aquaculture facilities
	23. Population distribution – demography
	24. Area management/restriction/regulation zones & reporting units
	25. Natural risk zones
Annex II	26. Atmospheric conditions
10. Elevation	27. Meteorological geographical features
11. Land cover	28. Oceanographic geographical features
12. Ortho-imagery	29. Sea regions
13. Geology	30. Bio-geographical regions
	31. Habitats and biotopes
	32. Species distribution
	33. Energy Resources
	34. Mineral resources

Table 1: Data themes of INSPIRE

The Directive does not require new data collection and does not set any obligation for data providers to change existing workflows. By enabling interoperability, data can be used coherently, independent of whether the existing dataset is actually transformed (harmonised) permanently or is only temporarily transformed by a network service in order to publish it in INSPIRE.

The SDI envisioned by INSPIRE is still under construction. The legislative process is continually evolving, complementing the Directive with ‘implementing rules’ that define the Member States’ obligations in concrete technical and legal terms. Each implementing rule is accompanied by technical guidelines which, in addition to providing general support for implementation, may give directions as to how to further improve interoperability.

[9] For example, standards are needed for the “invoke” services for service chaining, or interoperability target specifications for spatial data.

The experience of INSPIRE is notable given its size and results. Besides covering an unusually large number of data themes and involving participation from hundreds (if not thousands) of stakeholder organisation in the European Union and beyond, it has led to agreements that are legally binding in the Member States.

1.3 Interoperability and data harmonisation

The objective of effective use brings interoperability to the forefront. According to the 10-year Implementation Plan (GEOSS, 2005a), interoperability refers to the ability of applications to operate across otherwise incompatible systems.

There are three basic architectures for interoperable systems (Lasshuyt and van Hekken, 2001) as shown in Figure 1.

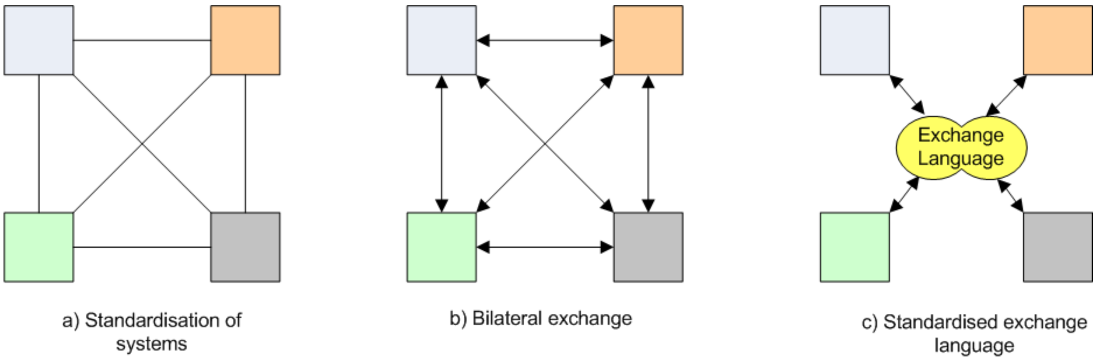


Figure 1: Basic architectures for interoperability (adopted from Lasshuyt and van Hekken 2001)

As shown in Figure 1a, when the systems are standardised they communicate with each other in a fully interoperable way. In most cases this approach does not work as each system has been developed according to the standards, conventions, or best practices of a particular organisation or user community.

In the case of bilateral exchanges (Figure 1b), dedicated interfaces are required between each pair of interconnected systems. The number of interfaces rapidly grows with the number of different systems. The third option (Figure 1c) is commonly considered the most practical solution for interoperability. This is a flexible system of systems, to which new systems can be added without having to adapt the existing ones or add new interfaces.

Even though there is no unique definition of the term system of systems, an SDI definitely fulfils its main requirements (management and operational independence, evolutionary development, emergent behaviours, large geographic extent). An SDI links the geographically dispersed system of various data providers at local, regional, national, transnational, and global levels. Each system works independently under local governance; they communicate with each other using agreed standards. According to best practises SDIs should be established and developed using a stepwise approach, with a continuously growing participants and widening scope. The emergent behaviour (the capacity to perform functions that do not reside in the components) can be detected through better decision making, when information is integrated in a trans-boundary or cross-theme context.

According to INSPIRE, interoperability is defined as "the possibility for spatial datasets 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 datasets and services is enhanced". This definition shifts the focus from how the systems interact[10] to how their users can benefit by removing the barriers commonly faced when trying to combine data from various sources.

In SDIs, interoperability bridges the heterogeneity between the communicating systems in two ways:

1. Transformation of spatial data (using information and communications technology); and
2. Harmonisation of the data the systems contain.

Data is transformed by specific software to produce a standardised presentation of the data. The transformation can be performed on- or offline. In the on-line process data is frequently transformed by web-based services. In the offline method an interoperable view (copy) is produced and stored to be accessed by a download service. In both cases the initial semantics and structure of data are preserved to fulfil the original user requirements for which they have been created.

Harmonisation is necessary when technical arrangements fail to bridge the interoperability gap and changes in the underlying data are needed. Harmonisation approximates the semantics and structure of the data and removes the remaining inconsistencies that cannot be solved by available technology. Both interoperability arrangements and harmonisation lead to standardisation of the output information.

> **Data harmonisation** is the process of modifying / fine-tuning semantics and data structure to facilitate compliance with agreements (specifications, standards, or legal acts) across borders and/or user communities.

Standards in the geospatial domain are mainly introduced at national or international levels. The Technical Committee (TC) 211 of the International Organization for Standardization (ISO) and the Open Geospatial Consortium (OGC) define the basis for the creation of geospatial information that has to be made coherent across domains. ISO standards are formulated in collaboration with national standardisation bodies, while OGC standards are created with the support of technology users and providers. Both organisations accumulate knowledge about best international expertise, which facilitates the worldwide diffusion of these standards.

Theme-oriented standardisation takes place in various international organisations such as the International Hydrographic Organisation (IHO), the North Atlantic Treaty Organisation (NATO), the World Meteorological Organisation (WMO), etc. In topics of common interest these organisations collaborate both in formal standardisation processes[11] and in SDI initiatives[12] leading to the further convergence of geographic information.

---

[10] This does not mean that INSPIRE ignores the interoperability of systems. The Network services component also covers IT technology.

[11] http://www.dgiwg.org/dgiwg/htm/activities/external_c_c.htm

[12] http://www.iho.int/iho_pubs/CB/C-17_e1.1.0_2011_EN.pdf and http://www.ungiwg.org/contact.htm

In addition to the abovementioned de-jure standards, best community practices (de-facto standards), such as GeoTIFF for geo-referenced imagery, GBIF and TAPIR for biodiversity, THREDDS for real time environmental data, or BAG for bathymetric data may be considered to achieve interoperability.

It is evident that interoperability arrangements and data harmonisation go hand in hand in SDI; the interoperability gap can be only breached by balancing both. Interoperable systems, in spite of their increased potential for effective reuse, must remain perfectly fit for the purpose for which they have been created.

2 Spatial data

2.1 From real world to spatial data

Spatial data is any data with a direct or indirect reference to a specific location or geographical area[13]. Spatial information contains spatial data that is structured for a specific purpose. In addition to describing the location and distribution of different phenomena in our terrestrial environment, spatial information explores context and relationships between spatial and non-spatial data.

> Geographic or spatial information?
> Geographic information is linked to a specific location on the Earth's surface. Spatial information points to a location on (topography), beneath (geology), or above (meteorology) the surface of the Earth. In addition, spatial data may relate to local, sometimes micro-systems (e.g. data from close-range photogrammetry).

It is important to note that "any description of reality is always an abstraction, always partial, and always just one of many possible views" (ISO TC 211 2005a). Diverse descriptions (abstractions) lead to multiplication of information related to the same geographic/spatial location. The abstraction process can involve various points of view, may be related to different moments of time, and may yield varying levels of detail in the information about the described area[14]. The three approaches that lead to a multiplication of geographic data are:

1. multiple views (multi-thematic views),
2. multi-temporal representations,
3. and multi-scale (resolution) representations.

1. Multiple views

Depending on the context and the point of view, the same phenomenon can be represented in various ways. Each community emphasises those properties of the phenomenon that are of interest to a specific field or task. A river, for example, can be regarded as a part of a hydrological network, a means of transport, part of a state's boundary, or a habitat of protected species. Each description is valid; the river section is the same, but the data collected and the information derived from this data is different for each

> A spatial data theme comprises all spatial objects that are relevant when describing the real world from a specific viewpoint. Spatial objects (features) are abstract representation of selected entities of the real world.

[13] Art. 3(2) of INSPIRE Directive

[14] ISO TC 211 standards use the term "Universe of discourse" to emphasise the fact that only some selected entities of the real world are targeted in a modelling process.

scenario. Each viewpoint outlines a specific thematic field. The term ‘spatial data theme’ is often used to refer to the collection and classification of spatial objects which is carried out from the same viewpoint.

Two potential views of hydrological data are shown in Figure 2. A ‘network’ view of hydrography is very useful for flood modelling, while the ‘mapping’ view is necessary for planning engineering facilities.

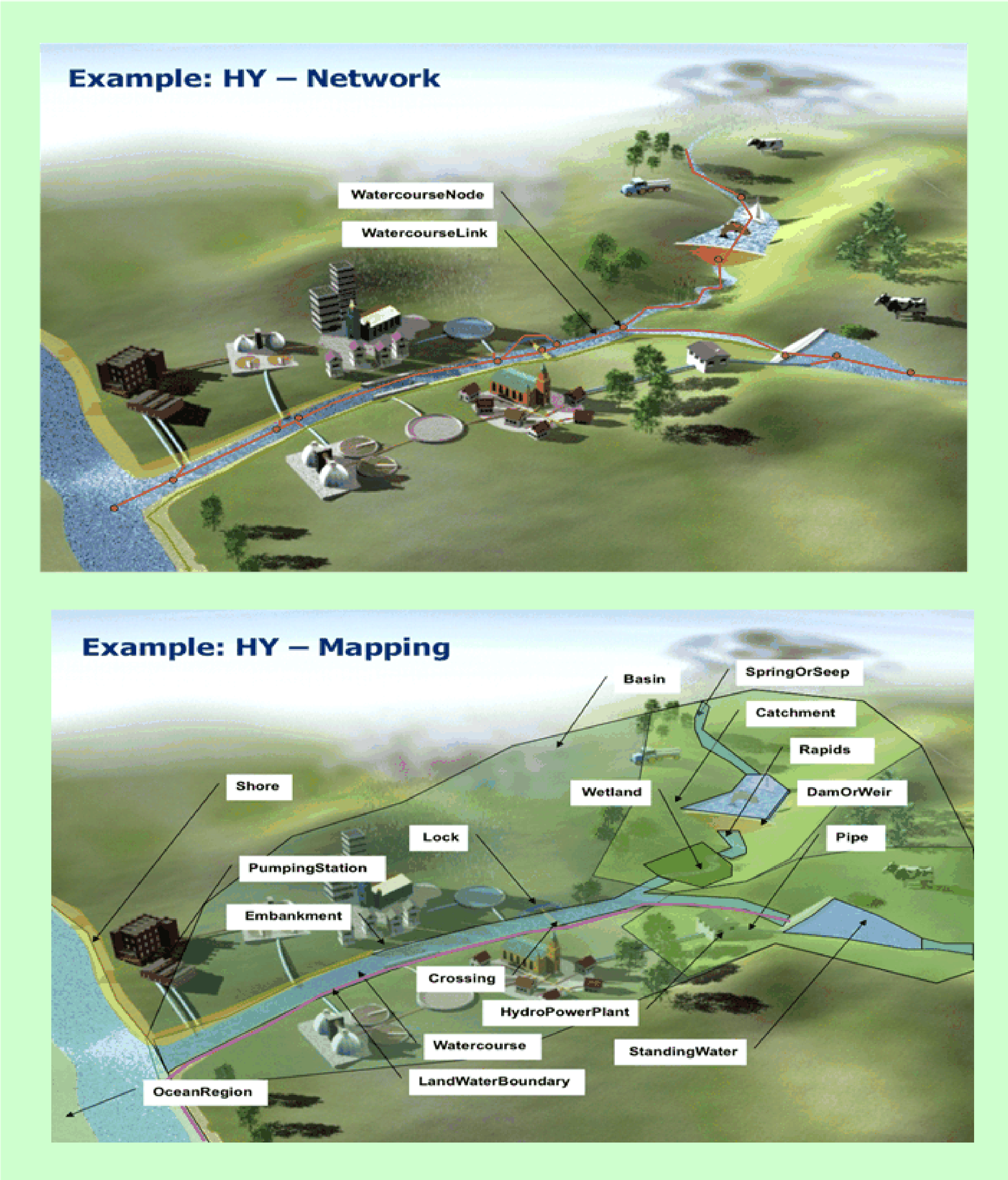


Figure 2: Multiple views on hydrology

2. Multi-temporal representations

Our world changes over time and this should be reflected in empirical data descriptions. Multi-temporal representation is a multiplicity principle which links a spatial object that is valid in a specific moment of time with its predecessor(s) and/or successor(s).

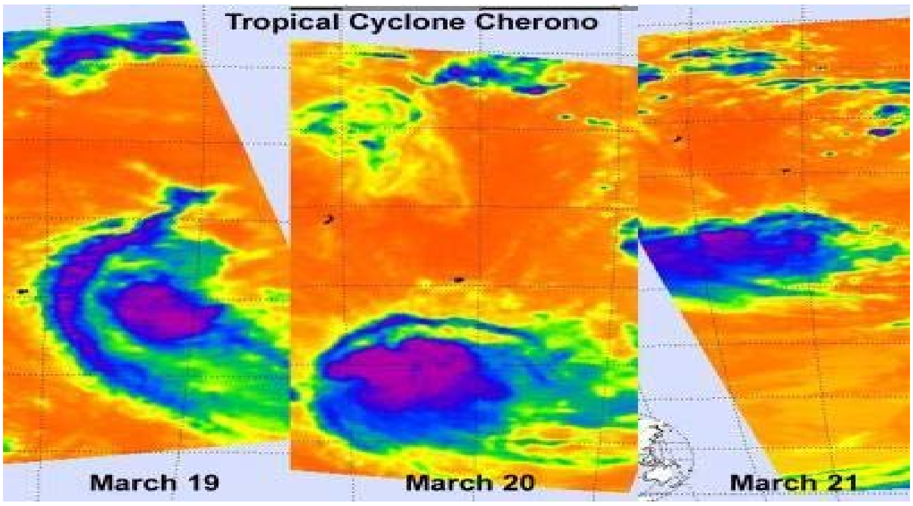


Figure 3: Multi-temporal representation

Rapidly changing natural phenomena, such as meteorological cyclones, are tracked using time series of satellite images. Here the identity of the cyclone remains the same, but its position, extent, and physical properties change over time.

The frequency of data capture can be very high, especially when automatic sensors are used. This information can be aggregated over time to represent the status and/or the values of a phenomenon at selected moments in time or by average values for a defined period. Climatic data is derived by aggregating meteorological observations from various periods of time.

3. Multi-scale (resolution) representations

Within a data theme, the entities of the real world can be described with varying levels of detail. The process of generalisation involves reducing the amount of detail in the representation of information. In the case of describing a settlement, as seen in Figure 4, a very detailed description could include single buildings and all the streets in the area, a less detailed one provides only blocks of buildings and main roads, while in small scales all the blocks of buildings are represented as one built-up area. The less detailed representations will include only a small number of the most important thematic properties (e.g. a point representing the whole settlement and its geographic name).

As a rule, detailed representations depict objects with the best approximation of their shape and true position, while less detailed representations allow simplification, which is important for preserving clarity and legibility of spatial information on maps or screens. The approach that associates different levels of detail is called a multi-scale or multi-resolution representation, but they are often referred to simply as multiple-representations.

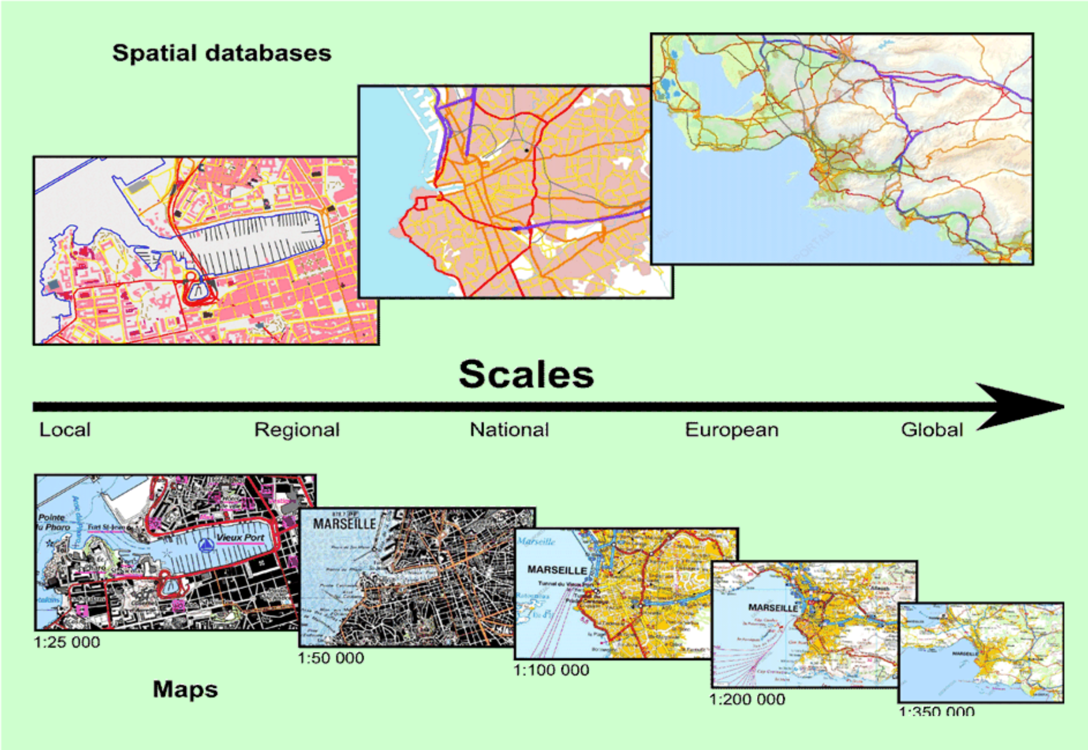


Figure 4: Multi-scale representation and generalisation (source of images: www.geoportail.fr)

Multiplicity of information that relates to the same place or to the same phenomenon in different moments of time offers enormous potential for gaining a better understanding of our world, because simultaneous or comparative analyses may explore new, otherwise hidden aspects.

Multiple geospatial information may be more demanding in terms of data processing and maintenance because of the potential inconsistencies of representations involved. The following section describes the challenges of integrating information from various sources.

2.2 Issues of incompatibility and inconsistency of spatial data

Users trying to integrate spatial data from disparate sources or to reuse information developed in other systems frequently face the problem of data incompatibility and inconsistency. The root of the problem lies in the different political, economic, cultural, and technical drivers of data production, which are expressed in in differences of syntax, semantics, spatial and temporal representations, as well as a lack of consideration for the co-dependencies between the themes.

Syntax is the internal structural pattern of natural or machine-readable language. The simplest examples of syntactic differences are the file storage formats used by different software and the grammatical rules of human languages. Without agreed syntax or a thorough knowledge of encoding languages, the communication between the systems cannot take place.

Syntactic differences can be bridged by technology and organisational solutions. Technology provides, for example, software tools to convert the formats of storage files. Harmonised presentation of the data can be achieved by agreements on the use of specific, preferably open source encodings.

Semantics is the study of meaning. It focuses on the relation between signifiers, such as words, signs, and symbols, and what they represent. Semantic consistency means that any two persons or any two systems will derive the same inferences from the same information. Semantic variability of geographic information and data results from abstraction processes whereby different communities in multinational or multidisciplinary environments describe the real world in different ways.

The concepts used for describing real world entities may not match in terms of their content (definition), degree of aggregation (semantic resolution) and the richness of description (number of properties or attributes), leading to differences in classification and/or in aggregation level, as illustrated in Table 2.

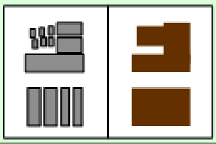
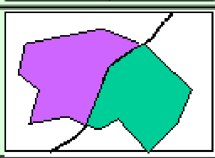
Examples of semantic differences		
Different aggregation level		The same real world entity is represented at different aggregation levels (houses vs. blocks)
Different classifications		The same entity differently classified at the two sides of a boundary (industrial zone vs. built-up area)

Table 2: Examples of semantic differences of spatial data

Semantic differences can be bridged by harmonising the concepts or by using technologies developed within the context of the semantic web[15]. Concept dictionaries, taxonomies, classification schemes, code lists, etc. are some of the vehicles use to publish agreed and harmonised concepts of spatial data.

Spatial representation may cause a further challenge to the integration of geographical data. Inconsistencies frequently occur at the graphical representation and may also lead to problems in data processing. Some typical examples are shown in Table 3.

[15] See section 4.1.6

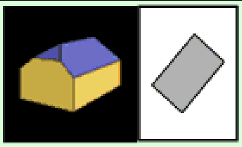
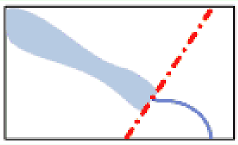
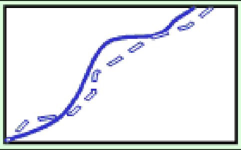
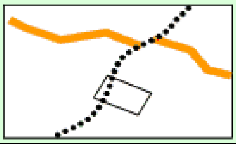
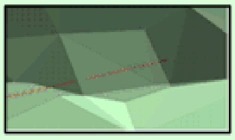
Different Spatial representations		Limited capabilities - overlay of raster (orthoimage) and vector (roads) representations
Different representation geometries (3D vs.2D)		The same building represented in 3 and 2 dimensional geometries
Different planar representation geometries		The river is represented by a polygon on one side of a boundary, while on the other by the center line
Different boundaries		Possible causes: absence of agreement between authorities, measurement/transformation errors, different generalisation
Overlapping spatial objects and geometrical shift		Errors along a boundary presumabely because of the different original projection systems
Inconsistency between data themes (Digital Elevation Model and Roads)		Violation of natural co-dependencies (the road crosses the land surface without a tunnel)

Table 3: Interoperability problems connected to spatial representation

Interoperability arrangements and data harmonisation in SDIs aim to eliminate incompatibility and inconsistency of data, thereby exempting the users from having to undertake onerous data manipulations before they start using data in their applications. The following paragraphs give some examples of interoperability problems related to differences in spatial representation, as illustrated in Table 3.

The first example in Table 3 shows spatial incompatibility arising from different spatial representations. Integrating coverage (raster) and vector data[16] rarely goes beyond overlaying and visual analysis because of the incompatibility of the processing algorithms. While converting vector data into simple coverage data (e.g. rasters) is relatively easy and can be carried out automatically, converting coverage data into vector data may require map digitisation.

Depending on the intended use of the data, the spatial characteristics of real-world phenomena may be represented using different geometric models. These include volumes in three-dimensional (3D) models, or surfaces in 2D models. The data about same or similar entities which are modelled using different geometry types need to be modified in order to be integrated. It should be noted that, without additional information, different representation forms can generally be transformed only by decreasing the dimension. For example, a river can be represented by a surface area or a centre line, as shown in Table 3. In order to arrive to a common and interoperable

[16] The main types of spatial representation are described in section 4.1.7

representation, the surface has to be collapsed into a centre line, which can be implemented by various algorithms.

The real world position of entities of social and political character (such as administrative boundaries, management units, etc.) has to be agreed by the competent authorities before they are delivered as geographic data. The absence of such agreements may lead to inconsistent representations of the adjacent and intersecting spatial objects along the boundaries of such entities. Differences in the position of boundaries, especially state boundaries, may be caused by using different reference and projection systems[17], which may manifest in unjustified overlays or discontinuities, as shown in the fifth example of Table 3.

Describing the real world using abstract representations from a specific viewpoint may ignore the natural dependencies of real world phenomena. This becomes evident when data from various sources is integrated. As shown in the last example of Table 3, the representation of the road that intersects the surface of the digital elevation, without a tunnel, provides an inconsistent model of the reality.

2.3 The subject of SDIs

As stated in section 2.1, describing our environment from different points of view, at different moments of time, and with different levels of detail leads to multiplication of spatial data, where each description serves a well-defined purpose. The descriptions, however, may contain common elements. The deeper we go into any specific aspect the less common elements we find. Vice versa: some aspects, like methods of describing spatial position, are shared across all applications.

> The purpose of **Reference data** is to establish a generic location context that can be reused (i.e. referred to) for other information.

Where does an SDI find its place among the countless number of applications that use spatial data? SDIs should encompass the common spatial aspects constituting a generic location context for a wide variety of applications. For example, demographic data can be linked to addresses, or can reuse the geometric position of administrative units. The use of reference data as an anchor to link other geographic or business information is one of the core concepts of the United Kingdom's Digital National Framework.

The means of defining the scope of an SDI is illustrated in Figure 5. A thematic SDI, like INSPIRE, may include generic concepts related to the target thematic field, for example, spatial data related to hydrology. Following the principle described above, only those spatial objects that have a strong potential for reuse should be included in the infrastructure. Specific applications, such as those that deliver business information, are out of the scope of the infrastructure.

---

[17] Reference systems define the frame for describing the position of spatial objects using coordinates. Projections are needed to represent the curved surface of the Earth on planar (paper or screen) media.

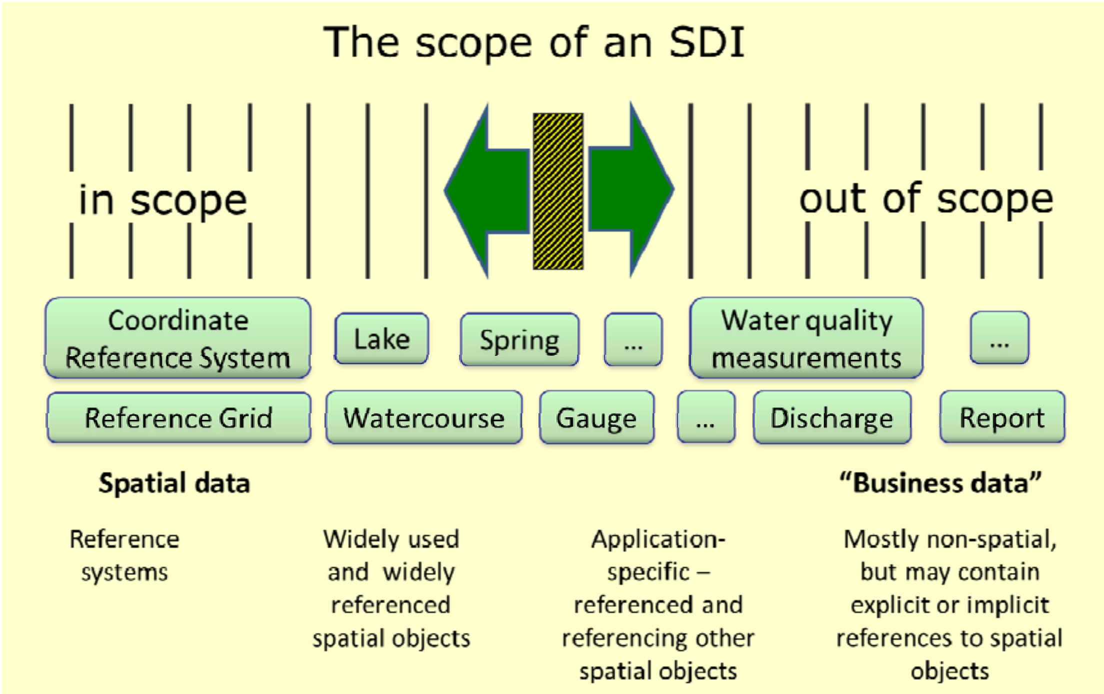


Figure 5: Scope of an SDI

Instead of including very specific details in the SDI, potential users should be informed, how the spatial framework provided by the infrastructure can be used and/or extended for their purposes.

3 The Conceptual Framework for Data Modelling in SDIs

A **spatial data model** is a mathematical construct to formalise the perception of space. A **conceptual model** encapsulates semantics (concepts) to categorise spatial objects within the scope of the description (universe of discourse). An **application schema** adds logical structure to the semantics defined in the conceptual model.

Spatial data represents real world phenomena in abstracted form, which can be structured in data models. Within a stakeholder community, the concepts of the data models in use are well known, and are sometimes even formally agreed on. People in the land registry domain have a common understanding of cadastral parcels, nature protection specialists know what a designated area is, and topographers don't need explanations about contour lines. In summary, each community abides by some fundamental agreements related to the data models they use. These agreements are often published as regulations, standards, or are shared as conventions and good practice examples.

Data modelling and data specifications are linked, in the first place, to data collection and data product delivery. But what role do they have in SDIs?

The interoperability in an SDI means that users are able to integrate spatial data from disparate sources "without repetitive manual intervention", i.e. the datasets they retrieve from the infrastructure follow a common structure and shared semantics. One way of achieving such interoperability would be to select one of the datasets and make the others comply with it. However, there is an infinite number of ways in which datasets can be combined; therefore each time a dataset is selected as a target model, all others would have to be transformed to comply with its specifications. This would also require publishing the data models for each source dataset. This is not a cost effective solution and does not add much value above the solutions already available in desktop GIS.

Instead of defining targets for interoperability on an ad-hoc basis, it is generally preferable to agree on common interoperability targets that are formalised and documented for each data theme so that they can read and used both by humans and machines.

> Data specification in the broader sense refers to both the data product specification, which is used for creating a specific dataset or product, and the interoperability target specification in SDIs, which is used for transforming existing data so that they share common characteristics. In this report, the term data specification refers to the interoperability target specification.

A data specification contains the data model and other relevant provisions concerning the data, such as rules for data capture, encoding, and delivery, as well as data quality requirements, metadata for evaluation and use, data consistency, etc.

A critical success factor for any SDI is its acceptance by the stakeholders. A bottom-up approach that creates a participatory environment in the specification development process foresees various interactions and feedback to the stakeholders' communities. Therefore, a collaborative model is needed that incorporates the safeguards necessary for consensus building processes.

Since an SDI is usually composed of many data themes where cross-theme interoperability may be required, a robust framework should be established that drives the development process of the data component in a coherent way. This idea was proposed in Germany as early as 1997 in the form of a harmonised conceptual base model ("AAA-Basisschema") for three national databases: the Official Fixed Point Information System (AFIS), the Official Real Estate Cadastre Information System (ALKIS), and the Official Topographic Cartographic Information System (ATKIS)[18]. The Geospatial Blue Book initiative in the USA (2005), which aimed to create "GIS for the Nation Data Model"[19], suggested keeping the application schemas of the data themes in a common information system that reinforced the consistent treatment of common concepts.

In the European Union, INSPIRE has adopted a conceptual framework that consists of two main sections as shown in Figure 6:

- The Generic Conceptual Model and
- The methodology for data specification development.

[18] http://web.archive.org/web/19981206200623/http:/www.adv-online.de/neues/oinhalt.htm
[19] http://support.esri.com/en/downloads/datamodel/detail/42

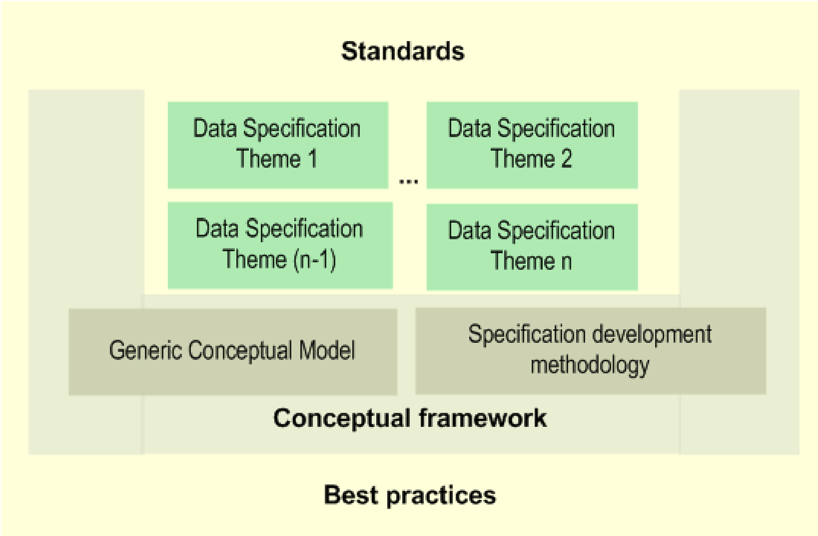


Figure 6: Relations of a conceptual framework

The main role of the conceptual framework is to provide a repeatable data specification development methodology and general provisions for the data specification process, which is valid for all spatial data themes. The conceptual framework outlines a step-wise and iterative process for establishing the data component: work should start by defining the common parts that must be followed by theme-specific tasks. In other words, the specification process of the data themes can only begin when the conceptual framework is sufficiently developed.[20]

The introduction of the conceptual framework is in line with the principle of reuse. In the context of SDIs, reuse relates not only to sharing data in different applications, but also to sharing knowledge, technical solutions, tools and components. Standards and examples of good practices of spatial data providers and user communities represent the basis for defining the conceptual framework and the data specification process.

The complexity involved in arriving at agreements on interoperability grows with the number of data themes and with the number of participating stakeholders. INSPIRE, with its 34 data themes, hundreds of participating experts, and rigorous documentation, is a good example for illustrating the role of the conceptual framework. Therefore, **chapter 4 and 5 are mainly based on the experience of INSPIRE**, and are complemented by inputs from other initiatives where appropriate.

One of the main tasks of the INSPIRE initiative is to enable the interoperability and, where practicable, the harmonisation of spatial datasets and data services in Europe. It is important to note that interoperability must go beyond particular communities and take the various cross-community information needs into account (Portele C. (editor) 2010a).

The generic conceptual model (GCM) makes the concepts of interoperability and data harmonisation more tangible by using a set of **interoperability elements**. These elements are derived from the requirements and the objectives of the infrastructure,

[20] The conceptual framework can be developed by reviews of the stakeholder communities, testing, and maintenance. This latter is connected to the modifications that stem from the application of the conceptual framework in the data specification development process.

matching them with the corresponding technical terms of geospatial technology and information modelling.

A valid question is whether the data component of an SDI can be established without a generic conceptual model. No generic conceptual model is needed for reaching interoperability within a single data theme, where a single interoperability specification would resolve the lack of interoperability. An SDI, however, consists of many data themes that do not form isolated flows of information. Interoperability and harmonisation is necessary if the infrastructure aims to share semantics, spatial representation, and syntax across themes.

In Figure 7, each box represents a well defined element of the application schema that can be a semantic spatial object, a geometric representation, an imported schema, a code list, etc. Because of the overlap between the data themes and the limited number of applicable standards, some of these elements have to be treated in a similar way.

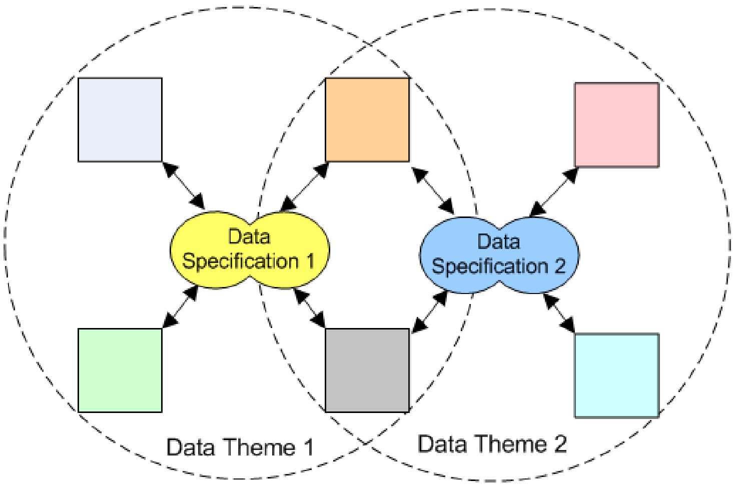


Figure 7: Cross-theme interoperability in SDIs (adapted from Lasschuyt, E. & van Hekken, M., 2001)

The GCM incorporates the shared concepts of data modelling and data specification development. The elements included in the GCM should not be specified in the data specifications of the individual themes. Vice versa: when common elements are discovered in the data specifications of two or more themes, these elements must be removed from the data specifications and included in the GCM.

Even though it is not called as a Generic Conceptual Model, the United Kingdom's Digital National Framework (DNF) sets principles, concepts and methods to establish better integrity of spatial information. It targets cross-cutting issues, such as:

- Linking information from multiple sources to a definitive location reference using unique identifiers,
- Structured presentation and formalisation to support data sharing and reuse,
- Reliability and data integrity,
- Flexibility enabling information exchange and cross-business applications.

A synthetic presentation of the interoperability elements contained in a GCM is given in Table 4. The majority of the presented elements were defined at the very beginning of the technical work on INSPIRE. They were later complemented by the outcome of research initiatives (e.g. the use of ontologies) and the practical experience of the INSPIRE development process (adding the consolidated model repository and migrating data specification maintenance from the GCM to the specification development methodology).

Fundamentals	Data Modelling	Data Management
– Requirements – Reference model – Architectural support for interoperability – Terminology – Multi-lingual text and cultural adaptability – Use of ontologies – Coordinate referencing and units of measurements – Registers and Registires	– Object referencing – Spatial and temporal aspects – Rules for application schemas and feature catalogues – Shared application schemas – Consolidated model repository – Multiple representation – Extension points	– Identifier management – Consistency between data – Data and information quality – Metadata – Conformance – Data capturing rules – Data transformation guidelines – Rules for data maintenance – Portrayal – Data delivery

Table 4: Interoperability elements for the data component of an SDI

The first group of the interoperability elements defines a starting point for the data specification process both in theory and practical tools. The second group supports the data specification process, while the third underpins interoperability from the view of data management.

Some elements, such as the reference model, shared application schemas, coordinate referencing, etc., have to be modelled, agreed and published. Others have to be managed and published in registries to support information sharing during the specification development phase and the operational phase of the infrastructure (i.e. when users can retrieve data according to the interoperability specifications). There are also elements that provide guidelines and best practice examples to support consistent implementation. Each element applies to all spatial data themes, but the degree of significance varies from theme to theme.

The INSPIRE GCM is being developed in an iterative fashion. The first version was derived by the Data Specifications Drafting Team according to the requirements of the INSPIRE Directive, matching these with technical provisions found in international standards and other reference materials describing good practice examples. Having improved the draft GCM on the basis of consultation with the stakeholders, the baseline version was delivered to the Thematic Working Groups responsible for developing the data specifications for Annex I themes.

The GCM has been updated over the course of the development of the Annex I data specifications. The main change was the introduction of the generic network model, because it was found that the network representation form was used in two themes. The Thematic Working Groups responsible for the

> The **Generic Conceptual Model of INSPIRE** contains a generic network model, which has been introduced when Hydrography and Transport Networks data theme started to model the spatial data as networks. The generic network model ensures that the same geometric principles are used. Later on, it has been reused in the Utilities data theme.

development of data specifications for the Annex II and III themes started to work with the updated GCM, and introduced other shared elements during their activities, such as the coverage schema and the observation and measurement model. Since the development of data specifications for Annexes II and III is still ongoing, other updates may still be made to the GCM. Further modifications may arise during the maintenance process of the specifications as presented in Figure 8.

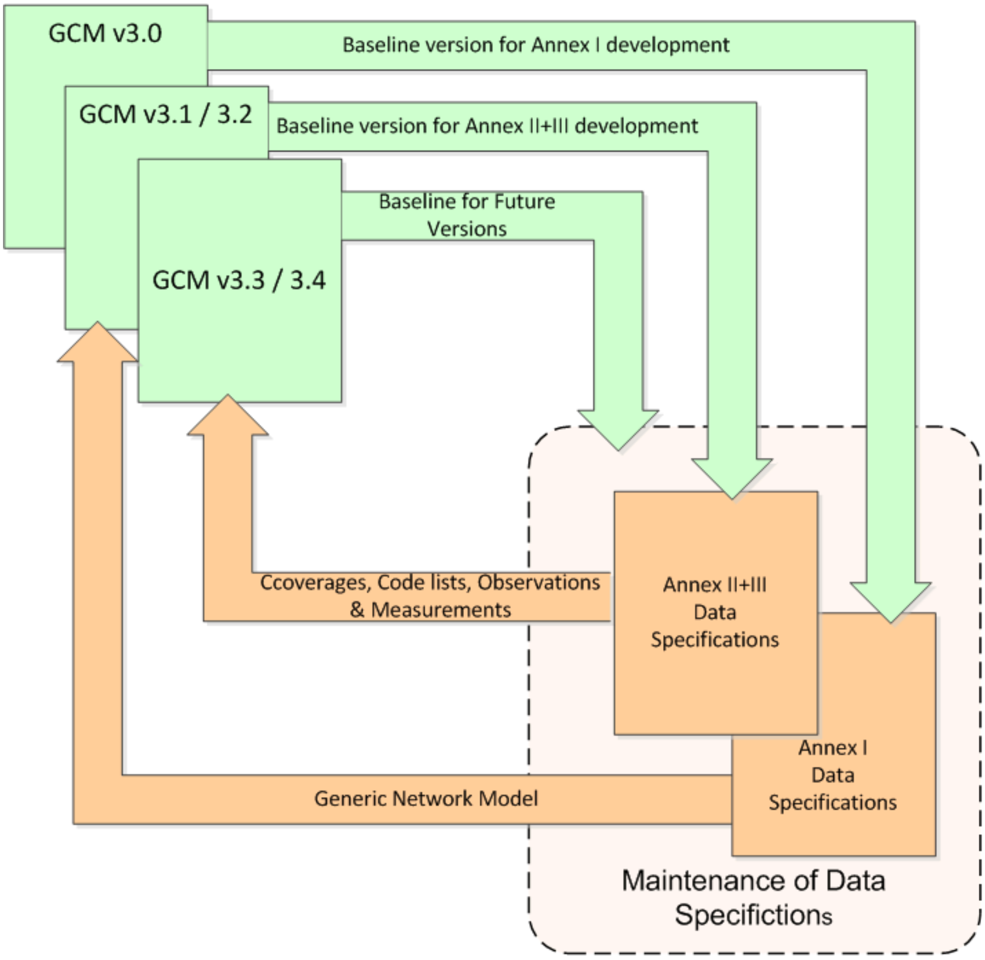


Figure 8: Iterative development of the Generic Conceptual Model

The following sections give further details of each interoperability element included in Table 4. Because of the nature of the topic, these sections are inevitably more technical. Readers more interested in the process may wish to skip section 4 and go straight to section 5, the Methodology for Data Specification Development.

4 Generic Conceptual Model

4.1 Fundamentals

4.1.1 Requirements

Experience shows that the requirements and implementation principles of SDIs might be dispersed over various policy papers, legal acts, technical studies, and other documents. In order to outline the extent of the required technical activities, these requirements and principles must be collected and systemised. Without being exhaustive, such principles may include:

- No obligation for new data collection: the arrangements target existing data and future data collections initiated by the competent organisations of the stakeholders,
- Inclusiveness: any data is better than no data,
- User driven approach: to delineate what should be included (re-usable geographical information) and what level of description is appropriate,
- No obligation for changing existing workflows: only publishing data according to the agreed interoperability target via network services,
- Instead of re-engineering, priority is given to transforming existing data,
- Reuse of existing standards, conventions and initiatives,
- Technical feasibility and proportionality (even though limitations of software components are not the main focus) to ensure that the specifications can be aligned with the ICT infrastructure of the data providers,
- Step-wise approach for implementation,
- Financial proportionality and cost-benefit considerations to ensure an optimal solution,
- Consistency of data/information referring to the same spatial location, presented in different scales and resolutions, and along boundaries (state and regional boundaries, etc.).

Clarification of such high level requirements is the first step in defining the GCM because these requirements are then translated into modelling constructs and specification elements.

4.1.2 Reference model

The reference model states where standards are applicable and how they should be used for developing the data component of the SDI. Since standards, as a rule, have broader scope, it is necessary to agree on the principles for adapting them to a specific purpose. This process of adaptation is referred to as profiling. The reference model also lists the types of information technology services that might be used for accessing, processing, and sharing geographic data and related information in the infrastructure. An example of a reference model is ISO 19101 – Geographic information, Reference Model, which provides a high level description how geographic information is created and how the standards relevant to this field fit together.

The INSPIRE generic conceptual model can be regarded as a specific reference model that serves as the basis for the data specification development. The GCM may also be used for developing other infrastructures in other geographic or thematic contexts.

4.1.3 Architectural support for data interoperability

Embedding spatial data in infrastructure means that access to the data is supported by the other building blocks of the SDI. These building blocks include data, metadata, network services, as well as arrangements for data sharing.

For efficient performance, the building blocks of the SDI have to be interlinked, which requires their coordination and fine tuning with respect of each other's functionalities and technical characteristics. This interoperability component also summarises the rules and technologies applied to publish information items necessary for understanding and interpretation of geographic information.

In SDIs, spatial data is accessed via the Internet through services providing specific functions such as discovery, access, mapping, transformation or other processing operations. The data component of the SDI has to take into account the technical characteristics of the **Network services**. For example, the View service may require that data be provided in a defined coordinate reference system or pre-defined styles for display/visualisation. This should be reflected in the data component.

Metadata provides information about the datasets and services integrated in the infrastructure. The primary function of metadata is to help discover existing data and services, and to help evaluate their fitness for purpose. Metadata for evaluation and use is tightly coupled with the data models and other specification elements. Data structures, semantics, encodings, eventual quality requirements, and other technical characteristics are fixed in the data specifications that are reported to the users as metadata. Ideally, data and metadata production go hand in hand.

The purpose of **data and service sharing** is to establish harmonised conditions of access to different groups of users. In an ideal SDI, all conditions of use are clear, complete, available to the public, and published online in various languages in a global context. The rights assigned to different groups of users in SDIs are managed through an access control function.

Registry services provide access to registers[21]. Since they play and important role in the data specification development process, they are included as an interoperability element in the generic conceptual model.

4.1.4 Terminology

Consistency of language is vitally important to semantic interoperability. The SDI needs a reference tool for sharing terms and their definitions. Glossaries, together with Feature Concept Dictionaries, support the coherent development of technical documents (specifications, web pages), improve their consistency, and allow stakeholders to better understand the data and the services present in the infrastructure. For better accessibility they must be implemented as registries.

4.1.5 Multi-lingual text and cultural adaptability

SDIs can span linguistic and cultural frontiers as well as competence areas of communities. It is therefore necessary to establish mechanisms to bridge any difficulties in reaching a common understanding of terms.

“The solution to multi-lingual issues is not the translation of everything into a common language (e.g. English). Often, it is sufficient to obtain resources in their original production language, rather than in its translated version” (European Committee for Standardization, (2011). This statement raises two issues:

- What should be translated; and
- When and how translation should take place.

[21] More details are included in section 4.1.8.

To allow machine-readability, the use of linguistic text in SDIs should be kept to a minimum, especially in the technical specifications. Ideally, the terms are kept in central (multi-lingual) dictionaries where they are translated into all of the languages of the addressed users[22]. Such centrally managed vocabularies can be used by humans or machine translation tools, thereby helping to eliminate the need for ad-hoc translation by the users who are not necessarily familiar with technical terms. For data access and to facilitate understanding it is useful to develop cross-language information retrieval strategies. That is why code lists, feature concept dictionaries, and feature catalogues compliant with ISO standards should be multilingual.

The rules for geographic names are different from those for linguistic text. Since geographic names are indirect spatial references that are widely used in querying other spatial information, it is essential that the names and their corresponding exonymes[23] be provided in majority and minority languages; none of these geographic names can be replaced by translations.

4.1.6 Use of ontologies

Ontologies are formal representations of semantics that can promote cultural adaptability and the dialogue between different groups of stakeholders. The Simple Knowledge Organization System (SKOS) Reference provides a standard, low-cost migration for porting existing knowledge from different systems - such as thesauri, taxonomies, classification schemes, etc., to the Semantic Web based on the similarities in their structures. It may be used on its own or in combination with formal knowledge representation languages such as the Ontology Web Language (OWL).

Ontologies are helpful in capturing multi-cultural aspects only if they are rich enough to include the contextual information necessary for different communities to reach a shared understanding. This interoperability component provides guidance for ontology development in SDIs.

It should be noted that although the operational use of ontologies in SDIs, including INSPIRE, is limited, research projects and emerging Semantic Web technologies are opening up new perspectives for their application.

4.1.7 Coordinate referencing and units of measurement

> In INSPIRE, the International Terrestrial Reference System (ITRS) with its European version (ETRS) are used for horizontal coordinates, while for the vertical component the European Vertical Reference System (EVRS) is used. The recommended projections are the Lambert Azimuthal Equal Area (ETRS89-LAEA), the Lambert Conformal Conic (ETRS89-LCC), and the Transverse Mercator (ETRS89-TMzn) projections.

Spatial position can be defined by the coordinate values of geometric points that represent the spatial object. A reference system is needed to define coordinates. Furthermore, for representing the curved surface of the Earth on planar media (paper maps, screens, etc.), a projection system is required. The selection of coordinate reference systems and projections varies from

[22] In INSPIRE all the official languages of the European Union are used.

[23] A geographical name used in a specific language for a spatial object situated outside the area in which that language is spoken; for example the English name “Brussels” is an exonyme of Bruxelles and Brussel.

country to country (to minimise the associated errors) and from community to community (to optimise spatial analysis and representations according to the use). In order to integrate data originally defined in different reference systems and/or projections, it is necessary to transform the data into a common system.

The common reference and projection systems selected for enabling interoperability should be precisely described. The coexistence of different reference systems requires their registration together with the specific transformation parameters needed to get from one system to another.

The GCM should also regulate the units of measurement. Based on international standardisation initiatives, preference is given to the International system of units (SI) except for the angles, which are usually reported in degrees. Parametric, or on non-length-based systems[24] may be used in addition to linear systems.

4.1.8 Registers and registries

An SDI involves a number of items that require clear descriptions and the possibility to be referenced. Registers assign identifiers to items and their definition and/or description. They are frequently implemented as registries, i.e. information systems for the maintenance of registers Registries are tools for information and knowledge sharing. In order to facilitate the reuse of concepts and components in the development phase of the infrastructure they are included in the GCM. For operational SDIs they help users to better understand the semantics and structure of the data.

Without being exhaustive, here are some examples of registers that are relevant for SDIs:

- **Glossary**: documentation of the terminology used in the infrastructure,
- **Feature Concept Dictionary**: This establishes a set of feature-related concepts (name, definition, description) that may be used to describe geographic information,
- **Feature Catalogue Register**: This register, based on ISO 19110 feature catalogues, contains definitions and descriptions of the spatial object types, their properties and associated components occurring in one or more datasets, together with any operations that may be applied,
- **Consolidated Model Repository**: A collection of all data models in a selected conceptual schema language, which permits the interdependencies between the models to be managed,
- **Code List Register**: An extendable controlled vocabulary describing the value domains of selected properties in an application schema, which is managed separately in its own dictionary,
- **Coordinate Reference System Register**: A register of coordinate reference systems, data, projection systems and coordinate operations which are used in the infrastructure,
- **Units of Measurements Register**: A register of units of measurements which may be used in spatial datasets,
- **Namespaces Register**: This manages the uniqueness of namespaces that can be reused, for example, for external object identifiers within the infrastructure,

[24] Such as barometric, or other length systems (e.g. miles).

- **Portrayal Register**: A register supporting the configuration of view services and the sharing of user-defined styles,
- **Encoding Schema Register**: This collects the specifications of data encoding used in the infrastructure.

4.2 Data modelling

4.2.1 Object referencing

Instead of assigning coordinates directly, the location of a phenomenon can be defined in relation to an existing spatial object. Such indirect referencing is possible by

- specifying references to other spatial objects,
- using a geographic identifier from a gazetteer.

Object referencing reuses the geometric coordinates of the referenced spatial object, specifying how the new information can be linked to existing coordinates. For example, in the case of linear referencing, an existing linear object (e.g. a road section) can be used to locate another spatial object (e.g. a bus stop) by indicating the distance from the beginning of the section.

A gazetteer allows a client to search and retrieve elements of a geo-referenced vocabulary. This alternative referencing method is especially useful in the case of geographical names and addresses.

4.2.2 Spatial and temporal aspects

There are two ways to describe the spatial extent or distribution of spatial objects: representing data as vector or 'coverage' datasets.

Traditionally, the geographic approach regards the world as being composed of identifiable structures with objective properties. This approach leads to vector data, where each phenomenon is conceived of as a separate spatial object with a separate identity. These objects are represented as points, surfaces, or volumes (in true 3D representations). The properties of such objects are described as attributes. Vector data addresses the question "*Where are the spatial objects belonging to a specific type and what are their properties?*

Another way of describing the world is the continuous field view, where a phenomenon is represented by a number of variables, each measurable at any point on the Earth's surface. These values change across the space and/or time (Longley, P. A. et al., 2011). This representation method, which is frequently referred as 'coverage', is very common in observations and measurements, including Earth observation. From a mathematical point of view, a coverage is a function that answers the question: *What is the value (of a specific property) at a specific location?* The assigned values often represent distributions such as temperature, elevation, or human population. The most frequently used coverages are grids that contain a set of values, each associated with one of the elements in a regular array of points or cells.

Both spatial representation forms are required since they "express […] the world: as a space populated by things, or as a space within which properties vary" (Woolf et al., 2010). It should be noted that the spatial representation form is not pre-defined by the data content. Within the same application they may be transformed into each other. For example, a stereoscopic pair of digital areal or satellite images (a coverage) can be used for extracting elevation data which can then be represented either as vector data (a collection of contour lines, elevation points, breaklines, etc.) or as an elevation grid (coverage data).

For temporal references it is necessary to state the time zone and the calendar used. The general usage of the Gregorian calendar together with a selected and agreed time zone facilitates data handling. For international and global SDIs it is reasonable to use the Coordinated Universal Time (UTC) standard. Interoperability is further supported by unambiguous and well-defined methods of representing dates and times according to ISO 8601 – Data elements and interchange formats – Information interchange – Representation of dates and times.

4.2.3 Rules for application schemas and feature catalogues

As already outlined, an application schema is a conceptual data model that is developed for a specific application (in data production), or for setting the interoperability target for a data theme in SDIs. It contains the spatial object types, their relationships and attributes, as well as eventual constraints applicable to the elements of the model. In SDIs each data theme contains at least one application schema. More application schemas can be introduced when

1. The data theme is too “big” and logical division according to different viewpoints is possible. This situation has arisen in the INSPIRE “Transport networks” data theme, where separate application schemas were developed for road, rail, water, air transport and cableways,
2. The data theme contains a core data model that is legally binding for implementation and one or more extended data models that are recommended, but not mandatory,
3. Different aggregation levels (different scales or resolutions) have to be modelled explicitly.

The rules for conceptual modelling regulate how the real world should be represented as application schema. A common Feature Concept Dictionary maintained for all data themes contributes to data consistency and eliminates redundancies.

The rules for application schemas contain the modelling constructs that are used in constructing the application schemas. Simpler homogeneous models facilitate both the specification process and the implementation of the specifications by the data providers.

The use of a common conceptual schema language[25] for formal documentation of the data models allows automated processing of application schemas. Nowadays the most frequently used conceptual schema language is the Unified Modelling Language

[25] A conceptual schema language is a formal language based on conceptual formalism for the purpose of representing conceptual schemas (ISO 19101:2005). It is usually machine readable to support the transition to the encoding schemas.

(UML). The SDI stakeholders may agree on a UML profile, i.e. on the eventual restrictions on the UML elements used.

A feature catalogue is an equivalent representation of the information in the application schema. The feature catalogues play an important role as:
- They support the conversion of the application schema information into text that is readable by humans,
- They support multilingualism as they are translated into the languages of the stakeholders (the application schema should be managed in one common language only)
- They facilitate searches and access to individual elements in the application schema, by human users and by software, as they are published via a registry service.

4.2.4 Shared application schemas

This data interoperability element collects reusable component models that are applicable in multiple schemas. In Figure 7 in page 24 the reusable components can be found at the intersection of the two data themes. Such schemas can be either defined for the infrastructure or can be imported from other initiatives.

A small but widely used model is the schema for unique identifiers. The structure of unique identifiers is described in section 4.3.1. Another example is the already mentioned generic network model. The “Observation and Measurement” application schema is shared by a number of INSPIRE data themes, such as Environmental Monitoring Facilities, Oceanographic geographical features, Atmospheric conditions, Meteorological geographical features, Soil, and Geology.

Shared application schemas are important tools for reinforcing cross-theme consistency and interoperability. It makes sense, therefore, to check existing application schemas before developing a new data theme in an SDI. The consolidated model repository described in the following section provides a straightforward access to all application schemas developed in context of a given SDI. When such a repository does not exist, developers have to check standards and other reference materials as described in the ‘as-is’ analysis paragraph of section 5.3.

4.2.5 Consolidated model repository

In an SDI context, where different theme-specific groups may be developing and maintaining data models, it is crucial to have a comprehensive yet concise overview of all agreements and results of the data modelling process. A specific tool is needed to provide this overview and thus allow the consistent (re)use of models developed by other groups.

The data specification process in INSPIRE adopted a **consolidated model repository** containing the agreed foundation models (such as ISO and other standards), the generic conceptual model, and the application schemas of the data themes. The introduction of the consolidated model repository was the only feasible way to jointly develop consistent data models and application schemas for 34 spatial data themes, because it allowed the expert groups working on the theme-specific data models to follow each other’s work and to detect similar modelling approaches, overlaps and gaps. The INSPIRE experience has shown the considerable value of this approach, which is summarised as follows.

First, the foundation models are scattered over various standards and are usually presented as static graphics or diagrams. The consolidated model repository makes them available in one place in a reusable form. Using specific information modelling software, it is possible to directly work with the data models included in these standards, importing their relevant components (profiles) into theme-specific data models. Consequently, standards are implemented in each theme in similar way.

Secondly, any spatial object, regardless of the application schema or theme in which it is created, can be referenced from other application schemas (in other themes). This is a crucial step for reinforcing consistency between the data models in different themes and thus for interoperability.

Thirdly, presenting data models in conceptual schema language (e.g. UML using ISO/TS 19103:2005) and in a graphical way (e.g. as UML diagrams) provides a quick and easily understood presentation of the data, which is also readable by machines. The narrative presentations of the schemas (feature catalogues) as well as the elements of the Feature Concept Dictionary can be derived automatically from the documentation of the data models in the consolidated repository. This feature helps to avoid inconsistencies in the narrative documentation of the specifications.

Finally, the repository makes it possible to generate the models automatically using GML/XML encoding schema[26]. It is recommended to make both the UML models and the GML/XML encoding schemas available as registries within the infrastructure in order to support the uptake and implementation of the models. For example, stakeholders may use the UML models as a basis for creating extensions that cover domain- or country/region-specific requirements. They can also be used by stakeholders to automatically generate other encodings.

For implementation, it is crucial to have access to the encoding schemas related to a specific data specification, e.g. in order to allow automatic validation. When models and schemas are updated as part of the maintenance procedure, it is vitally important that the different versions of the data model and the encodings can be accessed in order to be able to find out their status (valid, deprecated, etc.).

4.2.6 Multiple representations

As mentioned in section 2.1, real world phenomena can be described at different levels of detail. These are expressed in the aggregation levels of the concepts used for the abstraction (single houses vs. a built-up area) and/or in the spatial representation (river represented by a surface or a centre line). Scale/resolution is always selected as a function of concrete user requirements.

Should the need for different scales/resolution arise for a specific theme in an SDI, the different levels of detail can be modelled explicitly using separate application schemas that provide multiple representations of the real world. In order to keep the representations coherent, the application schemas have to be interlinked. The spatial

[26] Encoding is addressed in more details in section 4.3.10.

aggregation process should be supported by generalisation-specialisation hierarchies of the model. For example, a spatial object defined as block of houses in a small scale representation should be linked with the houses in a large scale representation through aggregation relationship. This practice has a positive effect on the maintenance of data, supporting the automatic propagation of updates from larger scales to small scales. Using the previous example, the area of the block will change automatically with the number of houses linked to that block.

Multiple-representation increases the complexity of the application schemas. Therefore this approach should be justified by strong user requirements. Generally it is advised to model as few levels of detail as possible. The experience of INSPIRE shows that it was possible to stay with one generic application schema in the vast majority of data themes.

4.2.7 Extension points

The interoperability specifications are developed taking account of requirements that are shared by many users. In order to underpin concrete applications or link business information users may wish to extend the data specifications provided in the infrastructure. Such extensions may be valuable contributions to the further development of the infrastructure provided that the extension does not

- change anything in the interoperability target specification, but normatively references it with all its requirements, or
- add a requirement that breaks any requirement of the interoperability target specification or of the generic conceptual model.

Extensions may add new application schemas, new spatial object and data types, new constraints to the application schemas, and define additional portrayal rules, etc. The code list may also be enlarged, as long as the infrastructure does not identify it as a centrally managed code list.

4.3 Data management

4.3.1 Identifier management

Unique identifiers (UID) are necessary for referencing new spatial objects to existing ones, and for retrieving geographic data. Two types of identifiers can be distinguished: external object identifiers, which uniquely identify the abstracted spatial object, and thematic identifiers, which are used to uniquely identify real-world phenomena.

External identifiers should satisfy the following conditions:

- Uniqueness: no two spatial objects may have the same identifier,
- Persistency: it does not change during the lifetime of the spatial object and is never re-assigned,
- Traceability: a mechanism exists to find a spatial object in the infrastructure based on its identifier,
- Feasibility: the UID can be created in the infrastructure based on the UID maintained by different organisations.

The identifiers assigned within a GIS application do not fulfil the criterion of uniqueness, because there is no guarantee that the same sequence of alpha-numeric digits is not used in another place or application. Therefore unique identifiers must be external and consist of two parts:

– A namespace to identify the data source. The namespace is owned by the data provider and should be registered in the Namespaces Register,
– A local identifier, assigned by the data provider. The local identifier is unique to the namespace, i.e. no other spatial object carries the same unique identifier.

Thematic object identifiers (for example ICAO location identifiers for airports or NUTS codes for statistical units) carry encoded knowledge that is relevant for the SDI. However, in most cases they cannot be considered as external identifiers mainly because not all four conditions described above are met. They should therefore be provided as thematic attributes of spatial objects.

Thematic identifiers may be used to establish relationships between spatial objects in different datasets that refer to the same real-world object. For example, objects from a dataset containing information about the geometry of a river network could be integrated with objects from another dataset with information on water quality if both use the same thematic identifier, e.g. the identifier of the river (segment) according to some environmental legislation or register. For this reason thematic identifiers for real world objects are also maintained, for example, in the United Kingdom's open data activities (Chief Technology Officer Council 2011).

4.3.2 Consistency between data

Having transformed[27] the data according to the interoperability specifications, some residual differences may still persist[28] when data is integrated from different sources. For the sake of consistency, data providers must match their data based on mutual agreements on the classification and/or the position of the corresponding spatial objects.

This interoperability element provides guidelines as to when the matching of data is applicable and how the process can be organised. Some themes in the infrastructure, such as atmospheric conditions, meteorological geographic features, oceanographic geographic features or sea regions, etc., are less concerned by this component because of their cross-border, transitory or fuzzy nature. Positional data matching does not apply to non-contemporaneous datasets. "Inconsistencies" related to temporal differences are not classified as inconsistencies in the strict sense.

When data matching is justified, for example along boundaries, the data providers should agree either on the 'true' position of the spatial objects to be matched or on the principles of the matching process. Consistency between different themes should be required only within the same or closely similar levels of detail.

When different pieces of geographic information relate to the same location, natural dependencies must be reflected. For example, a road and a river cannot cross each

[27] Data transformation is addressed in section 4.3.7
[28] See examples in section 2.2

other in the absence of a bridge, tunnel, or ferry connection. An initial list of co-dependencies between the themes comes from the scoping process[29].

4.3.3 Data and information quality

Data quality is an important aspect when users need to decide on the data's fitness for use. For the convenience of the users, the presentation of data quality should be similar across the themes whenever possible.

From point of view of an SDI, poor data quality may compromise interoperability. However, no data should be excluded from the infrastructure because of low quality. 'Poor' data is better than no data. Consequently, it should be carefully assessed as to which requirements are indispensable for the proper functioning of the infrastructure. For example, from the point of view of interoperability, requirements of logical consistency (which defined the semantics and data structures) are more 'important' than those of positional accuracy.

In the context of SDIs, rather than setting a priori requirements on data quality, it is more appropriate to recommend targeted results. The targeted results also depend on the nature of the data – more stringent values apply to reference data, which is used for object referencing.

The objective of this interoperability element is to fix a conceptual model for the applicable data quality elements as defined in the relevant standards[30], as well as threshold target results for conformance testing[31]. The final aim is to give the end-user some assurance about the reliability of the information using traceable indicators[32] or data quality measures on selected data quality elements (such as completeness, consistency, currency, accuracy, etc.) or on the conformity of a dataset as a whole.

4.3.4 Metadata

Metadata provides "information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data" (ISO TC 211, 2003a). Metadata describing geospatial resources is closely linked to the data that they represent. Therefore, the ideal development cycle streamlines the two.

For organisational reasons, metadata and data specification developments are sometimes separated by drawing a line between metadata for discovery and metadata for evaluation and use. The rationale behind this is to anticipate data sharing within the infrastructure even when the data is not in conformity with the interoperability target specifications. Therefore, metadata advocating discovery and first level evaluation (i.e. describing basic technical characteristics such as scale/resolution, geographic extent, spatial representation form, etc.) are published to be complemented or refined by metadata coming from data specification processes.

---

[29] See section 5.1
[30] ISO 19113, ISO/TS 19138, which will be replaced by ISO 19158
[31] See in details in chapter 4.3.5
[32] See QA4EO of GEOSS

Metadata is the main resource that gives information about the actual quality of the data. Contrary to a priori data, metadata on data quality gives an ex-post evaluation, which - in context of interoperable data usage in SDIs - depends on two main factors:

- the quality of the input data, and
- the success of the transformation process necessary to achieve interoperability.

After transforming the data for the infrastructure, the metadata related to the original data may no longer be valid. Strictly speaking, they should be re-evaluated or transformed if data transformations bring systematic changes in data quality. This is an extra burden for the data providers, and may not be the first priority in the course of establishing the infrastructure. As a temporary solution, the original metadata could be published with a description of the transformation process steps in order to provide sufficient information to the users about data quality.

Users also may judge the relevance and usability of data based on the metadata. The uncertainty associated with the "objective" data quality measures and the potentially subjective usability descriptions sometimes create more barriers than support for the users. Product certification and labelling may offer a user friendly solution. The "GEO Label" initiative will mark the quality of Earth observation products based on a range of well defined measures assessing the quality of the data or information provided by a system (GEO Task ST-09-02 Committee 2010).

4.3.5 Conformance

Conformance is defined by ISO 19105 as the fulfilment of specified requirements. Obviously, conformance of data in SDIs has to be evaluated against the interoperability target specifications. The scope of conformance evaluation may relate to a single specification element (e.g. the application schema, data capture rules, or selected data quality elements, etc.) or aggregated to the level of the specification as a whole.

Any product claiming conformance to the specifications as a whole has to pass all the tests described in the abstract test suites (ATS), which refer to the requirements to be tested and list the applicable tests, the quality measures and the corresponding threshold values.

A dataset can conform to one or more specifications at any one time. In order to fully inform users about the conformity of data, it is advisable to declare conformance with all the specifications against which the data has been tested.

4.3.6 Data capturing rules

Data capturing rules provide guidelines as to which real-world phenomena should be included in a data theme. They are also the main elements used to specify a targeted level of detail. The typical selection criteria are minimum area, length, or functional characteristics.

Since SDIs typically target existing data, the determination of data capture methods (such as surveying and measurement methods, applicable sensor types, etc.) are not relevant for this component.

4.3.7 Data transformation model/guidelines

In a successful SDI, all the data providers publish their data according to the agreed interoperability specifications. This can be achieved by maintaining the data in conformity with the interoperability specifications for direct access through a download service. The viability of this solution is limited. On one hand, stakeholders communities have well-established requirements to stick to their own specifications. On the other hand, especially in the case of transnational infrastructures, the transformation of projected coordinates is almost always necessary.

The best theoretical solution is, therefore, to keep original data structures and publish data in the SDI through transformation. Transformation between source and target application schemas is a key transformation type, but other transformations (e.g. coordinate transformation, edge-matching, language translation, format transformation, etc.) might also be required.

To make data available through a download service, data is typically transformed offline to create a static view that is compliant with the interoperability target specification. Alternatively, data can be transformed inside the download service 'on-the-fly', according to previously defined mapping rules. A third option is to use a separate transformation service that executes predefined or user-defined mapping. It should be the responsibility of each data provider to choose the method and enable the necessary data transformation according to this choice.

4.3.8 Rules for data maintenance

As the infrastructure is based on existing data, the maintenance of datasets occurs at the source, i.e. with the data providers, following their own business processes. There are two issues to be resolved from the point of view of the infrastructure:

- Ensure that the updates are transmitted in a timely manner to the data publishers according to their interoperability specifications;
- Provide a mechanism to distinguish between current and historical data.

The first issue is automatically resolved when the data is maintained by the data providers in conformity with the interoperability specifications, or when the transformations to reach interoperability are automated. In this case the data in the SDI is kept up-to-date with minimal human efforts.

When data is transformed offline, specific attention should be paid to the issue of propagating the updates to the data presented according to the interoperability specifications. Therefore, the maximum delay for introducing the changes should be agreed or regulated.

In general, the capacity to provide data updates will depend on the availability of life-cycle information in the application schema, which documents the time at which new spatial objects were inserted, or existing spatial objects were updated or retired. Life-cycle information can be used in search queries to select only those spatial objects that were affected by changes since a point in time specified by the user.

4.3.9 Portrayal

The graphic presentation of geographic information depends on many factors such as the information content, the medium of representation[33], the eventual portrayal conventions within the stakeholder communities, etc. In SDIs, the main emphasis is on reusing and combining data from different sources, which creates an infinite variety of data that must coexist in the course of spatial analysis. The harmonisation of portrayal rules is therefore a complex task.

Following the principle of step-wise implementation, the first step may aim to support the view service only, which is used in the discovery stage. This approach has been adopted in INSPIRE, where portrayal is addressed from the perspective of the single themes. The schema for portrayal rules and symbology for geographic features specify basic rules (layer structure) and a standardised set of default styles.

The most frequently used visualisation methods are based on OGC Styled Layer Descriptor (SLD), which allows user-defined symbolisation and colouring when data is displayed in a Web Mapping Service (WMS).

The Keyhole Markup Language (KML)[34] is an XML language that focuses on geographic display/visualisation, including annotation of maps and images. Geographic visualisation represents graphical data on the globe and guides the user's navigation in the sense of where to go and where to look.

In order to avoid clashes of styles used in different themes, some basic harmonisation is necessary. Where there is no harmonisation, for example, the same blue line could be used to represent bathymetry, waterways, and boundaries of sea regions. Sharing SLDs in a registry can help this harmonisation process, e.g. by enabling queries for styles defined for different data themes. A registry may also be used to share user-defined styles (e.g. for specific purposes, such as coastal zone mapping).

4.3.10 Data delivery

For exchanging spatial data, efficient methods for encoding and data delivery are required. The encoding rule specifies the data types to be converted, as well as the syntax, structure and coding schemes. It presents data in a format suitable for transport and storage. Clear definition of data formats helps to ensure syntactic interoperability.

Because of the diversity of data present in the infrastructure (vector, raster, etc.) a unique encoding rule and output data structure cannot be mandated. Thus, every data specification should specify at least one encoding rule that is mandatory for that specific theme.

While flexibility to support additional encoding rules is a valid approach, harmonisation and reduction of the spread of encoding rules is also important. It is reasonable to maintain the list of recognised encoding rules and output data structure schemas in a registry. Encoding rules should be based on international, preferably

[33] Paper map, computer screen, mobile devices, mobile phones, etc.
[34] SLD is recommended by INSPIRE, KLM is supported by GEOSS.

open, standards and should be compliant with ISO 19118 Geographic Information – Encoding.

In INSPIRE, unless otherwise specified for a specific data theme, the recommended encoding is the OGC's Geography Markup Language (GML) as defined in ISO 19136. For large volume coverage data such as orthoimagery or computer simulations (e.g. weather forecasts), other, more efficient, file-based encodings (e.g. geoTIFF) may be defined as the default encoding language. These encoding schemas are widely supported and can be inserted in the majority of GIS.

In an SDI, spatial data is accessible via download and view services. This interoperability component also includes the services used to deliver data and a reference to the encoding formats applied for exchanging data between systems.

5 Methodology for Data Specification Development

5.1 *Definition of the scope of the data themes*

The definition of the INSPIRE themes started with analysis of the requirements of European environmental legislation. The preliminary list drawn in the position paper of the Environmental Thematic Coordination Group was discussed in wide before being defined in the annexes of the Directive. Because of the changes introduced in the consultation process, it was necessary to revisit the theme descriptions before the data specification process started. This has been carried out by the Data Specification Drafting Team in the "Definition of Annex Themes and Scope" document. Defining interdependencies between the themes, this document represented an important input for the data specification process.

Defining the scope of the data themes and the infrastructure requires careful considerations and consensus building among the stakeholders, including the data users, producers, technology providers, and politicians responsible for the strategic development of the relevant field. Surveys, state-of-play studies, formal written opinions, web consultations, and public hearings are some examples of instruments that can be employed in this process.

For historical and organisational reasons, spatial data is collected and maintained by many different organisations. Since their activities are not necessarily coordinated, there can be overlaps or gaps in the data content. As redundancies are important sources of data inconsistency, it is necessary to outline the borders between the data themes. A clear definition of the scope of the data themes will help stakeholders to judge how their interests might be influenced by the emerging infrastructure, and where they may need to interact.

When overlaps between two or more themes are discovered, the following decisions have to be taken:

- Are the apparently overlapping parts justified from a conceptual point of view? Do the spatial objects describe **different** abstractions of the same real world entity (e.g. a river section as part of hydrography vs. a river section as part of waterway navigation)? If yes, the spatial objects should be modelled in both themes. If not, it should be decided which theme is the most appropriate to deal with the spatial object in question.
- If the separation is conceptually justified, how can the difference be made visible (choice of terminology), what are the critical points that make the difference, and

should a relationship between the two concepts be established (e.g. identifying the hydrological river section(s) to which a water transport river section corresponds)?

It should be noted that the conceptual framework does not consider or resolve organisational constraints (i.e. in case of unjustified overlaps, which organisation is duplicating the information?); it only flags where efforts for coordination are needed. Coordination is equally necessary when interdependencies between two or more data themes are discovered.

Based on prominent use cases and reference materials, the scoping process also outlines the possible content of the data themes in terms of key spatial object types and their attributes. This non-exhaustive list should not be an attempt to define the full content, it is rather an illustration for better understanding. The proper analysis of references and the definition of data requirements should occur during the course of the data specification development. The main outcome of the scoping process is well-defined starting point for the data specification process.

5.2 Principles of data specification development

As part of the conceptual framework, the specification development methodology guides the process so that the general principles of the SDI such as reuse, feasibility, and proportionality are followed. The methodology gives instructions as to which actions need to be taken at the different steps of the process.

The specification development process can be driven by data providers and/or data users. In a provider-driven approach, the main principle is to find a common denominator between the existing datasets belonging to a specific theme. Without external benchmarks, however, interoperability requirements may remain unclear in this approach, which could lead to the following problems:

- The data delivered according to the interoperability arrangement does not meet the requirements of the users;
- Rather than seeking an optimum level of interoperability, the strongest stakeholders may promote their solutions in order to minimise the potential transformations/changes to the datasets they produce.

In the user-driven approach, the external benchmarks stem from the requirements of the users, which are carefully analysed and formalised at the beginning of the specification development process. This approach may be associated with the following risks:

- It is difficult to capture detailed user requirements up front,
- The expressed requirements might be too ambitious, leading to excessive costs or the impossibility of implementation based on the existing data,
- Instead focusing on reuse, the specification process may yield a product specification fulfilling the needs of a “strong” user.

Experience shows that, in practice, a combination of these two approaches tends to be used, balancing aspirations with technical and financial viability.

The methodology described in this chapter provides the details of the data specification development process used for INSPIRE. This methodology has incorporated the results and experience of scientific research projects[35] as well as best practices of SDI development. Furthermore, this methodology has been formally described and tested, delivering tangible results for each of the 34 data themes included in INSPIRE. INSPIRE takes an iterative approach to an incrementally growing SDI, which is based on stakeholders' commitment. This predictable and repeatable development process model allows feasible and mutually satisfactory system solutions to be reached. The main steps of the process are shown in Figure 9.

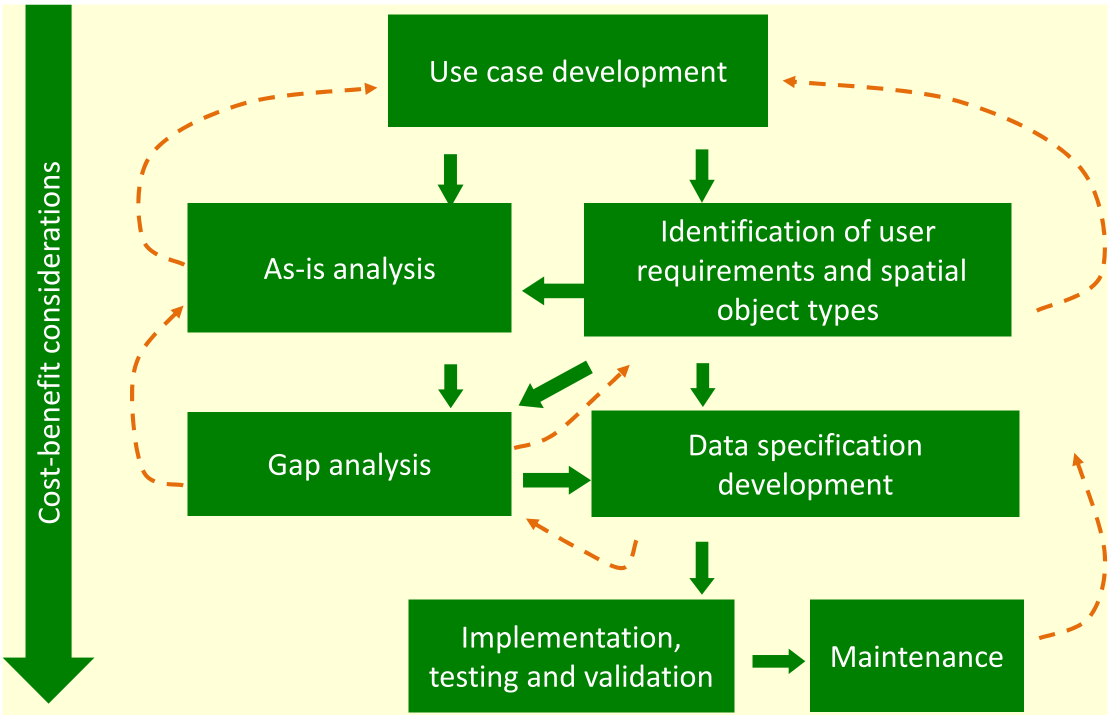


Figure 9: Steps in the data specification cycle

This approach helps to balance ambitions and feasibility. If ambitions are too high, this may lead to complex specifications, which will be difficult and expensive to implement. Furthermore, if specifications are too complex, there is a risk that they will not be supported by the data provider communities and that they will not be adopted by the users. However, overly simple data specifications may lead to insufficient interoperability, and the critical mass that makes the related efforts worthwhile may not be achieved, rendering the benefits of the infrastructure intangible. The main points of the challenge to be solved are illustrated in Figure 10.

---

[35] RISE ftp://ftp.cordis.europa.eu/pub/ist/docs/environment/rise_en.pdf
MOTIIVE https://www.seegrid.csiro.au/wiki/Marineweb/MOTIIVE

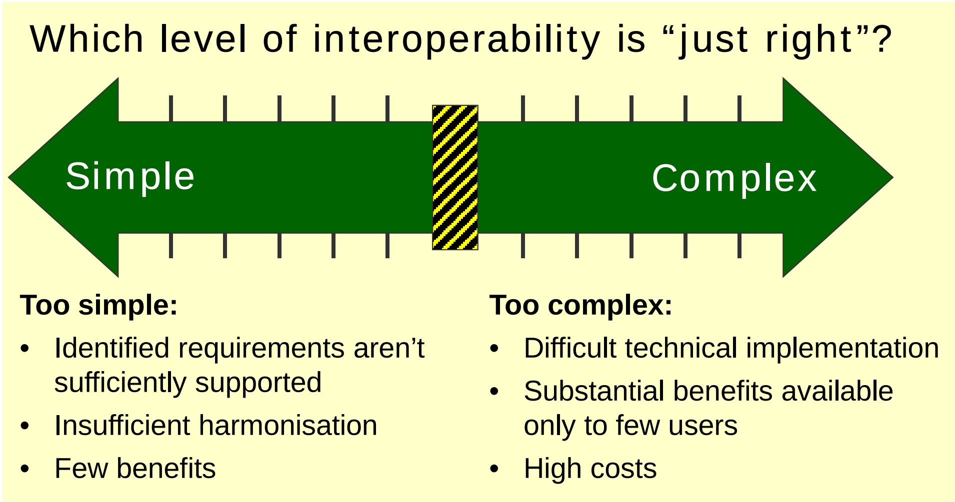


Figure 10: The challenge of finding a balance in the data specification process

A good approach to finding a balance is to apply two principles:

1. The focus of activities should be on generating consistent spatial (and temporal) information for wider use, leaving out information regarding the execution of business processes, scientific simulations, or specific reporting requirements.
2. Extension mechanisms should be provided for the models and it should be shown how other spatial and non-spatial aspects can be linked to the models.

The following sections describe in greater detail the steps to be taken in the data specification process.

5.3 The data specification development cycle

'Use case' collection and development

> A use case defines a goal-oriented set of interactions between actors and the system under consideration. Use cases help to understand the requirements of the users and define the data that is necessary to fulfill them.

The scoping phase of the infrastructure outlines what user needs are to be supported. These are further refined and documented during the initial phase of the data specification process. Use cases are widely used in information technology to formalise the descriptions of how users interact with the system to be developed. In SDI development, they illustrate the possible uses of data.

A use case may cover several data themes. For example, a use case describing flood risk analysis in a particular area may require data from hydrography, elevation, meteorology, etc., and may result in input for the "Natural risk zones" data theme. Common use cases also help to clarify eventual cross-theme dependencies. Therefore, use cases considered in the infrastructure should reflect the multiplicity of data usage.

For proper weighting of requirements, the use cases have to be ranked according to priority. High priority should be assigned to those use cases that are part of many user scenarios or are time-critical (disaster management, flooding, etc.). These are "quick win" areas where the benefits of SDI yield immediate and tangible results.

In practice, however, it can be difficult to collect use cases from the stakeholders. Data users are less aware of the benefits of SDIs or of SDI development initiatives. This should not jeopardise the specification development process. Specification development can start with preliminary use cases provided by the data providers, since they are usually aware of the tasks for which their clients use the data. Data users can be activated in parallel. The consultations included in the later phases of the data specification development cycle may provide the necessary feedback for improvements and convergence with users' needs.

Identification of user requirements and spatial object types

Use cases are used to identify the spatial data requirements in the 'first cut' data model. This model contains the candidate list of spatial object types, draft definitions and descriptions, and an initial set of other data specification elements. Each of these elements is defined according to the level of detail, which is determined based on user requirements. The concepts of spatial object types should be shared and harmonised across the different themes. A useful tool in this context is the Feature Concept Dictionary[36].

'As-is' analysis

Pursuing the principle that the SDI should bring existing data together, the data requirements from the use cases should be compared with the existing 'as-is' situation. This analysis reveals whether the requested data can be supplied by the data providers. If so, it also shows the complexity of the related transformation work. If there is no one-to-one relationship between the proposed harmonised schema and the theme-related datasets, data integration might be still required at the level of the data sources or by the users. The 'as-is' analysis is frequently performed in parallel with the gap analysis.

Gap analysis

Gap analysis identifies user requirements that cannot be met by the current available data. There are two kinds of gaps. Technical gaps can be filled by integrating data from any relevant dataset or data transformation, while content gaps can be addressed only by data collection. Existing state-of-the-art studies may provide a baseline for comparison.

Filling technical gaps provides undisputable value for the users, but may involve substantial costs to data producers. Technically sound and cost effective approaches may help, such as automatic tools for data integration and transformation. However, such transformation tools are not always available at the current technology level. Therefore a prudent approach that compares the benefits with the possible costs must be taken.

Data specification development

"First cut" data models and the other initial data specification elements outlined in the requirement analysis have to be adjusted according to the result of the 'as-is' and gap analyses. In order to respect technical and financial feasibility, the content of data specification can be earmarked for mandatory or optional implementation.

---

[36] See section 4.1.8

In INSPIRE, the mandatory elements are defined as "requirements" while the optional elements are defined as "recommendations".
Profiles have been applied, for example, in the Protected sites and the Buildings data themes.

According to the practice in INSPIRE, the data models should be implemented in their entirety; no spatial object type can be omitted. If there is a need to distinguish between more and less "important" spatial object types, the two groups must be packaged in separate data models that are also referred to as "profiles". Spatial objects that are indispensable to supporting the key requirements are placed in the core model. The extended models may guide voluntary implementation and the stepwise and coherent development of the infrastructure by setting targets for consecutive data collections and maintenance.

In addition to the technical elements, the data specifications may also contain explanations and examples to support better understanding and implementation.

Implementation, testing, and validation

Specifications must be reviewed and tested by a wider stakeholders group in order to verify whether data specifications are fit for the purposes of the infrastructure and contain enough information to support implementation.

In INSPIRE, three iterations were carried out. After the first iteration, the data specifications were reviewed by the Thematic Working Groups. The main purpose of this phase was to eliminate inconsistencies between the specifications between the various data themes. The second iteration comprised a review and a testing phase in which all the stakeholder communities could participate. In order to accelerate the process, consultative meetings – the 'comment resolution workshops' – were convened to resolve divergent opinions of the stakeholders. Based on the outcomes the specifications were again revised and published as implementation guidelines in the third iteration. Selected parts of the guidelines have been included in the legislative acts mandating the implementation of INSPIRE by the Member States of the European Union.

Specification testing can be carried out to deliver feedback on feasibility or fitness for use. Feasibility testing assesses the efforts of data providers required to transform their data to be compliant with the interoperability target specification. This results in feedback on technical feasibility and the associated costs of implementation.

Application testing assesses how much interoperability has facilitated the work of users. This test is performed by the data users to assess whether the data provided in conformity with the interoperability target specifications facilitates their performance. The results of testing and stakeholder consultations can be used for reiterating the data specification process from any step, most probably from the 'as-is' and the gap analyses. The iterations can be repeated until consensus is reached. After this validation process, the specifications are published so that they can be used by the general public.

Commission Regulation 1089/2010 implementing interoperability of spatial data-sets and services contains a subset of the INSPIRE Generic Conceptual Model and the data specifications. While there is one data specification for each theme, the legally mandated sections are collected in a singe "implementing" rule.

For legally reinforced SDIs, an additional step is necessary. The technical drafts should be made into legal acts fulfilling the legislative requirements while maintaining the technical content. One way of ensuring legal reinforcement is to mandate only the parameters for the services through which the data is made available in the infrastructure, leaving the semantic models in the guidelines. Another option is to select a subset of the data specifications, comprising the semantic model, based on technical feasibility and cost-benefit issues. In this case the data specifications with the full technical content serve as guidelines for the stakeholders, enabling further coherent development of the infrastructure.

5.4 Maintenance of specifications

Changes in requirements or in an 'as-is'[37] situation may trigger a revision of the data specification, and the associated registers, documents and tools necessary for supporting technical and documentation activities. The request for changes in data specification may be triggered by the following:

- Issues detected at a later stage in the course of the step-wise data specification process and in the implementation phase,
- Changes in the legislative frame with an impact on the requirements for spatial data,
- New initiatives and programmes influencing the development of SDIs (emerging SDI initiatives at higher level, eGovernment, etc.),
- Need for harmonisation with international standards and other initiatives,
- New relevant user requirements and use cases,
- Changes in the 'as-is' situation of the stakeholders and progress in technology,
- Errors or ambiguities within the documents,
- Inconsistencies with other building blocks of the infrastructure,
- Cost-benefit considerations.

From an organisational point of view, the maintenance procedure should be as open and participatory as the specification development process, which guarantees coherence between implementation, development and maintenance. Therefore the persons and organisations that have to be involved in the process, as well as the methods and workflows have to be defined.

The maintenance process basically follows three methods of change. The "fix and align" method serves to correct errors and (re)establish consistency with other components or building blocks of the infrastructure. The "depreciate" method is used to discard elements[38] that are no longer used or that are replaced with new items, while the "add" method allows new items to be introduced.

Minor corrections allow for a backwards compatible revision, i.e. all datasets that conform to the previous version are still conformant with the revision. Major revisions introduce significant changes. Where feasible and appropriate, a major revision should remain backwards compatible. This type of revision is allowed when absolutely necessary for the domain, e.g. to introduce a significant number of additional spatial object types to a theme, or to upgrade the Generic Conceptual Model or a data specification in a fundamental way.

In order to support the maintenance process, it is recommended that version control systems of repositories be used both for the consolidated data model and the technical documents.

---

[37] See section 5.3.
[38] In the interests of traceability, no item should be simply deleted.

5.5 Cost-benefit considerations

Besides being based on technical feasibility, the interoperability arrangements should be based on careful analysis of the related costs and the benefits, as shown in Figure 10 in page 43. Cost-benefit analysis in the data specification development process must be carried out throughout the specification process.

In cost-benefit analysis, the expected costs and benefits are converted into comparable units, usually monetary values. Carrying out a strict cost-benefit analysis is rather difficult for SDIs, especially in terms of benefits. Benefits are generally incurred by the users and society in general. Furthermore, before being visible, the benefits of SDIs may need time to mature, i.e. a transition period during which a critical mass of available datasets is transformed to reach interoperability.

Cost-benefit considerations give an overall presentation of quantitative and qualitative assessment criteria for SDIs. Instead of trying to convert each cost-benefit aspect into comparable (monetary) units, they contain statements as to

- Where and how costs and benefits are likely to occur,
- How to avoid or reduce costs by undertaking appropriate decisions and technical measures,
- How to highlight the possible benefits and make them visible to stakeholders.

The main means of detecting the possible costs related to the implementation of the interoperability specifications is the testing process, where data providers can record the investments necessary to reach interoperability in terms of expertise, time, new software and hardware, and educational needs. In INSPIRE, this type of testing is called ‘transformation testing’.

The other type of testing - application testing - helps to quantify the benefits to the users by comparing the time necessary for performing a specific task using data that is compliant with the interoperability specifications and the data supplied in its original form. If data in conformity with interoperability specification facilitates the performance of users’ tasks, the benefits of the infrastructure are visible. The benefits can be quantified in terms of time reduction, performing the tasks with less qualified personnel, etc.

In order to get a broader picture of the costs and benefits of the infrastructure, an extended impact assessment and a direct survey among the stakeholders have been carried out for INSPIRE. Table 5 summarises the main points relevant to SDI cost-benefit considerations.

COSTS	BENEFITS	
— Costs related to the development of the specifications — Costs of reengineering the databases — As an alternative, costs in developing schema mapping from old to new specifications — Hardware and software costs if new systems were required — Costs in running/ checking/validating the transformation	Direct User Value/Benefit — Increased data availability — Increased ease of use — Better data sharing ability — Reduced cost of integrating data Social Value — Enables better decision making — Reduces barriers between organisations — Increases institutional effectiveness — Promotes more efficient use of (taxpayer) funds	Operational benefits for institutions — Promotes intra-institutional collaboration — Promotes inter-institutional collaboration — Reduces data integration cost across institutions — Promotes reuse of existing datasets — Decreases costs of IT/ information management — Overall cost savings for info management — Achieves cost avoidance (as opposed to savings) — Fosters closer working relationships — Supports improved decision making — Supports other information infrastructure

Table 5: Aspects involved in cost-benefit analyses of SDIs

5.6 Actors in the data specification process

The organisational structure of establishing the data component of the SDI is defined by the following conditions:

1. The process should be based on consensus building;
2. Establishing and running an SDI aiming at cross-theme interoperability needs the involvement of numerous organisations;
3. Cross-theme interoperability requires tools and organisational measures for continuous flow of information between the stakeholders.

These conditions imply the need for coordination in order to ensure communication, planning, providing and maintaining the tools during the specification process.

The more data themes are included in the infrastructure, the bigger is the demand for a well-structured process. A modular approach allows more freedom from an organisational point of view. It might be difficult to engage the necessary resources to develop the interoperability specification for many data themes in parallel. When the modules are scheduled in the right order, the knowledge accumulated at the beginning can be used for the later stages. It is worthwhile to start the process with reference data, where stakeholders are “spatially aware”.

Meaningful discussions with stakeholder communities can only take place based on good proposals. The technical drafts for the interoperability target specifications have to be proposed by a competent body. Following the participatory principle in SDIs, the best organisational forms are the technical expert groups, composed of representatives of the stakeholders. The expertise of these groups includes:

- Expertise in geographic information modelling and the relevant standards,

- Thematic (domain) expertise (knowledge of the data to be used in the representative use cases),
- SDI expertise: knowledge about the underpinning policies and the standard SDI architecture,
- Network services expertise (knowledge about data access),
- Software expertise: expertise about the implementation and deployment of the relevant specifications.

In INSPIRE, the coordination body is called "Consolidation Team", which is composed of employees of the European Commission. For the data component two types of expert groups are distinguished: the Data Specification Drafting Team, which is responsible for the development and maintenance of the conceptual framework, and the Thematic Working Groups that are responsible for developing the interoperability target specifications for each data theme. The members of these expert groups are delegated by the communities of stakeholders. Stakeholders also participate in reviews and testing. The legally mandatory part of the specifications is adopted by the INSPIRE Committee, which is composed of official representative of the Member States of the European Union.

For effective work organisation, specific roles are foreseen in the expert groups. The **group leader** schedules the work, distributes the tasks among the members, and mediates the discussions with the experts in the group and the external partners. In the conceptual framework development phase the group should have a good overview of SDI developments and demonstrate a strong background in information modelling and standardisation. In the data specification phase the emphasis is on domain expertise and the knowledge of the conceptual framework.

The results of the specification work are documented by the **editor**, according to pre-defined templates. The editor must be a good technical writer, who prepares the narrative documentation and masters the selected conceptual schema language to present the data models in machine-readable format.

5.7 Supporting tools

Many different stakeholders are involved in the data specification development phase. The outcome of their work must be comparable. Each data specification should follow the same structure in the documentation, which facilitates communication between the expert groups and the uptake by the user communities. The expert groups responsible for technical drafting should be helped by tools and templates that guide the work, keep the results coherent, and help to share knowledge from the very beginning of the process.

The tools can be classified as shared document templates, document repositories, internet-based discussion fora, and registers. Shared document templates reinforce harmonised documentation and ensure that all the aspects that have to be considered are covered in the same way. In INSPIRE the most prominent example of templates is the data specification template, which is based on ISO 19131. In order to facilitate the work, other templates and checklists (e.g. for use case description and analysing the reference materials) can also be provided.

Document repositories help to share reference materials and working drafts primarily amongst the members of the expert groups. Making the drafts visible to all groups helps to foster coherence between the data themes. Version control systems of document repositories give the opportunity to return to a previous proposal in any time. In addition, keeping records of changes makes the process traceable and transparent.

6 Conclusions

The wealth of digital spatial data accumulated over the past 30-40 years and the advances of information and communications technology have opened new perspectives for analysing our physical and societal environment. Spatial analysis, decision support, and location-based services frequently reuse data that has been originally created for other purposes, achieving considerable economies in system development.

Integrating spatial data from disparate sources is often jeopardised by limited data sharing and the lack of interoperability. Spatial data infrastructures provide a means for overcoming these obstacles by offering online services for discovering, evaluating, retrieving and transforming data. One of the causes of limited interoperability is inconsistency and incompatibility. In most cases, data has to be transformed to share common characteristics and thereby achieve interoperability.

Without an SDI, these transformations are performed by the users on an ad-hoc basis. In SDIs, interoperability is enabled at the source; data providers should supply the data according to pre-defined and agreed norms. The technical presentations of these norms are the interoperability target specifications, frequently referred as data specifications.

The interoperability gap in the context of spatial data can be bridged in two ways: by using interoperability arrangements, which comprise technological and organisational solutions, and data harmonisation. In an SDI the preferred solution is the first, because data providers do not need to change their original data structures. They may deploy technology (e.g. batch or on-the-fly data transformation) to meet interoperability requirements. However, current technology does not always fully cover the interoperability gap. Data harmonisation brings the data structures of the different providers closer in line with each other. Experience shows that the combination of these two approaches provides the best solution.

An SDI is a collection of several data themes. The interoperability target has to be defined for each of them in the form of interoperability (or data) specifications. In order to achieve cross-theme interoperability, a robust framework is needed that reinforces common technical measures, efficient information exchange, and standardised methodology for data specification development across the infrastructure. This is the conceptual framework. Based on the experience of INSPIRE, this framework has two components: the generic conceptual model and the specification development methodology.

The generic conceptual model (GCM) turns interoperability arrangements and data harmonisation into a set of interoperability elements, matching them with the corresponding elements of information modelling and geospatial technology. Containing the shared concepts, the GCM is the principal tool for reinforcing interoperability across all the data themes included in the infrastructure.

The GCM approach has been rigorously implemented in INSPIRE, paying special attention to continuous sharing of the results of technical work. The publicly available registries and the use of the consolidated model repository mark an innovative approach to establishing the data component of an SDI. “In the future, this conceptual model is expected to influence, in many cases, modelling activities for spatial data at national level, because it adds value to the national spatial data infrastructure and simplifies transformation to the INSPIRE data specifications” (Portele C. (editor), 2010a). The technological convergence of the data providers is a key element of SDI initiatives.

As part of the conceptual framework, the specification development methodology reinforces the requirement that the general principles of the infrastructure such as reuse, feasibility, and proportionality be followed. The safeguards built into the process ensure that all the necessary steps and actions are completed in each of the themes included in the infrastructure. The methodology has to provide a predictable and repeatable development process, which leads to feasible and mutually satisfactory solutions. The methodology should also describe the roles that the stakeholders play during the different stages of the process.

The legislative framework of INSPIRE has established a strong precedent for an incrementally growing SDI based on stakeholders’ commitment. This experience shows that such methodology can deliver tangible results even when the scope of the SDI is broad, hundreds of stakeholders from more than 30 countries[39] are involved, and the technical work has to be prepared in a relatively short time[40]. That is why the data specification methodology proposed by INSPIRE has been adopted by the United Nations Spatial Data Infrastructure (Atkinson, R. and Box, P., 2008).

The particular value of the conceptual framework described in this report is that it collects the best practices of ongoing initiatives. Both the methodology for specification development and the generic conceptual model have been tested in real-life conditions in the course of the development of the data specifications. Even though this development process resulted in the 9 finalised and the 25 draft interoperability specifications it should be noted that their implementation is still underway and users’ benefits can be properly assessed only in the future.

The data specifications that have been carefully reviewed, tested, and endorsed by the stakeholders’ communities, prove the viability of the approach, crystallising collective knowledge from Europe and beyond. The ever growing participation in the process, the advances in the legal reinforcement, and the broad feedback received as a result of the testing and implementation process signify that a similar conceptual framework might also be a success factor in other initiatives.

---

[39] Besides the Member States of the European Union, stakeholders from the European Economic Area, Switzerland, USA, and EU candidate countries also joined the process.

[40] The technical work on the INSPIRE data component started in 2005 and is expected to be finished in April 2012.

Acknowledgements

The authors would like to acknowledge the work of numerous experts and stakeholders that contributed to the development of the technical guidelines and the implementing rules of the INSPIRE Directive. We especially appreciate the work of the INSPIRE Data Specification Drafting Team, which has systematically and meticulously put together and documented the conceptual framework of INSPIRE. We also thank INSPIRE stakeholders who, with their comments, testing and queries, tirelessly contributed to improving the work of the experts.

We would equally thank our reviewers from the JRC, Max Craglia, Michel Millot, and Katalin Bódis, who helped the authors not to get lost in the technical details. Their comments were very useful in removing assumptions and ambiguities, and in filling the gaps with information that hopefully made the heavy technical subject digestible. They played the role of "informed policy maker" excellently!

Our external reviewers, Siri Jodha Khalsa, Zdisław Kurczyński, Stefano Nativi, and Daniele Rizzi, helped us to take a step back from our data- and INSPIRE oriented view and to put the report in a broader perspective. They had a decisive role in shaping the report in its final, hopefully better structured and homogenous form.

Bibliography

Atkinson R. and Box, P. (2008): United Nations Spatial Data Infrastructure (UNSDI) Proposed Technical Governance Framework v1.1 (pp. 4-52). Retrieved from http://www.ungiwg.org/docs/unsdi/TechnicalGov/Proposed_UNSDI_Tech_Gov_Framework_v1.1.doc

Chief Technology Officer Council (2011): Designing URI Sets for Location. A report from the Public Sector Information Domain of the CTO Council's cross Government Enterprise Architecture, and the UK Location Council.Version 1.0 (pp. 1-18). Retrieved from http://location.defra.gov.uk/wp-content/uploads/2011/09/Designing_URI_Sets_for_Location-V1.0.pdf

Craglia M. et al. (2003): Spatial data infrastructures. GI in the Wider Europe (pp. 19-20). European Commission. Retrieved from http://www.ec-gis.org/ginie/doc/ginie_book.pdf

Craglia M. et al. (2008): Next-Generation Digital Earth. A position paper from the Vespucci Initiative for the Advancement of Geographic Information Science. *International Journal of Spatial Data Infrastructures Research*, *Vol.3*, 146-167.

Craglia, M. (2010): Building INSPIRE: The Spatial Data Infrastructure for Europe. *ARC News*, 5-7. Redlands, California. Retrieved from http://www.esri.com/news/arcnews/spring10articles/building-inspire.html

Craglia, M. and Nowak J., editors (2006): Report of International Workshop on Spatial Data Infrastructures "Cost-Benefit / Return on Investment" (pp. 3-61). Luxembourg. Retrieved from http://www.ec-gis.org/sdi/ws/costbenefit2006/reports/report_sdi_crossbenefit .pdf

European Commission (2008a): Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata. *Official Journal of the European Union*, *L 326*, 12–30. European Commission. Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32008R1205:EN:NOT

European Commission (2008b): European Interoperability Framework v 2.0 (pp. 1-79). Retrieved from http://ec.europa.eu/idabc/servlets/Docb0db.pdf?id=31597

European Commission (2009a): COMMISSION DECISION of 5 June 2009 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards monitoring and reporting. *Official Journal of the European Union*, *148*, 18-26. Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:148:0018:0026:EN:PDF

European Commission (2009b): Commission Regulation (EC) No 976/2009 of 19 October 2009 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards the Network Services. *Official Journal of the European Union*, *L 148*, 18-26. Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:148:0018:0026:EN:PDF

European Commission (2010a): Commission Regulation (EU) No 1088/2010 of 23 November 2010 amending Regulation (EC) No 976/2009 as regards download services and transformation services. *Official Journal of the European Union*, *L 323*, 1-10. Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:323:0001:0010:EN:PDF

European Commission (2010b): Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A Digital Agenda for Europe. *Official Journal of the European Union*. Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0245:FIN:EN:PDF

European Commission (2010c): Commission Regulation (EU) No 1089/2010 of 23 November 2010 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial data sets and services. *Official Journal of the European Union*, *L 323*, 11-102. Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:323:0011:0102:EN:PDF

European Commission (2010d): Commission Regulation (EU) No 268/2010 of 29 March 2010 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards the access to spatial data sets and services of the Member States by Community institutions and bodies. *Official Journal of the European Union*, *L 83*, 8-9. Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:083:0008:0009:EN:PDF

European Commission (2011): Commission Regulation (EU) No 102/2011 of 4 February 2011 amending Regulation (EU) No 1089/2010 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial data sets and services. *Official Journal of the European Union*, *L 31*, 13-34. Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:031:0013:0034:EN:PDF

European Committee for Standardisation (2011): CEN/TR 15449:2011 Geographic information - Standards, specifications, technical reports and guidelines, required to implement Spatial Data Infrastructures

European Parliament and European Council (2003): Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the reuse of public sector information. *Official Journal of the European Union*, *L 345*, 90-96. Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:345:0090:0096:EN:PDF

European Parliament and European Council (2007) Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) . *Official Journal of the European Union*, *L* (108), 1-14. Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:EN:PDF

GEO Task ST-09-02 Committee (2010): A GEO Label: Informing Users About the Quality, Relevance and Acceptance of Services, Data Sets and Products Provided by GEOSS (pp. 1-12). Retrieved from http://www.iiasa.ac.at/Research/FOR/downloads/ian/Egida/geo_label_concept_v01.pdf

GEOSS (2005a): The Global Earth Observation System of Systems (GEOSS) 10-Year Implementation Plan. Retrieved from http://www.earthobservations.org/documents/10-Year Implementation Plan.pdf

GEOSS (2005b): GEOSS 10-Year Implementation Plan Reference Document (pp. 3-209). Noordwijk. Retrieved from http://www.earthobservations.org/documents/10-Year Plan Reference Document.pdf

Geographic Information Panel (2008): Place Matters: the Location Strategy for the United Kingdom (pp. 8-39). London. Retrieved from http://www.communities.gov.uk/documents/communities/pdf/locationstrategy.pdf

Illert, A. (editor) (2008a): INSPIRE Definition of Annex Themes and Scope (pp. 1-132). Retrieved from http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/DataSpecifications/D2.3_Definition_of_Annex_Themes_and_scope_v3.2.pdf

Illert, A. (editor) (2008b): INSPIRE Methodology for the development of data specifications (pp. 1-123). Retrieved from http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/DataSpecifications/D2.6_v3.0.pdf

INSPIRE Data and Service Sharing Drafting Team (2010): INSPIRE Good practice in data and service sharing (pp. 1-66). Retrieved from http://inspire.jrc.ec.europa.eu/documents/Data_and_Service_Sharing/INSPIRE_GoodPractice_ DataService Sharing_v1.pdf

INSPIRE Metadata Drafting Team (2009): INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119 (pp. 1-74). Retrieved from http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/metadata/MD_IR_and_ISO_20090218.pdf

INSPIRE Network Services Drafting Team (2008): INSPIRE Network Services Architecture (pp. 1-30). Retrieved from http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/network/D3_5_INSPIRE_NS_Architecture_v3-0.pdf

ISO TC 211 (2000): ISO 19105 Geographic information - Conformance and testing (pp. 1-21)

ISO TC 211 (2002: ISO 19108 Geographic information - Temporal schema (pp. 1-48)

ISO TC 211 (2003a): ISO 19107 Geographic information - Spatial schema (p. 166)

ISO TC 211 (2003b): ISO 19115 Geographic information - Metadata (pp. 1-140)

ISO TC 211 (2005a): ISO/TS 19103 Geographic information - Conceptual schema language (pp. 1-67)

ISO TC 211 (2005b): ISO 19109 Geographic information - Rules for application schema (pp. 1-81)

ISO TC 211 (2005c): ISO 19110 Geographic information - Methodology for feature cataloguing (pp 1-55)

ISO TC 211 (2005d): ISO 19123 Geographic information - Schema for coverage geometry and functions (pp. 1-65).

ISO TC 211 (2005e): ISO 19128 Geographic information - Web map server interface (pp. 1-76)

ISO TC 211 (2005f): ISO/NP 19135-1 Geographic information - Procedures for item registration

ISO TC 211 (2007a): ISO 19131Geographic information - Data product specifications (pp. 1-40)

ISO TC 211 (2007b): ISO 19136 Geographic information - Geography Markup Language (GML) (pp. 1-394)

ISO TC 211 (2007c): ISO 19131 Geographic information - Data product specifications (pp. 1-40)

ISO TC 211 (2007d): ISO 19111 Geographic information - Spatial referencing by coordinates (pp. 1-78)

ISO TC 211 (2008): ISO TS 19101 Geographic information - Reference model- Part 1: Fundamentals (pp. 1-40)

ISO TC 211 (2010): ISO 19142 Geographic information - Web Feature Service (pp. 1-238)

ISO TC 211 (2011): ISO 19118 Geographic information - Encoding (pp. 1-69)

Klinghammer I. (1995): A térképészet tudománya. Membership inauguration lecture at the Hungarian Academy of Science. Retrieved from http://lazarus.elte.hu/hun/tantort/2005/szekfoglalo/klinghammer-istvan.pdf

Lasschuyt, E. and van Hekken, M. (2001): Information Interoperability and Information Standardisation for NATO C2 – A Practical Approach (pp. 1-20). The Hague. Retrieved from http://ftp.rta.nato.int/PubFullText/RTO/MP/RTO-MP-064/MP-064-05.pdf

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2011): Geographic Information. Systems & Science (Third edition, pp. 1-525). Hoboken: John Wiley & Sons, Inc.

Nebert, D. D. (editor) (2004): GSDI Cookbook (pp. 1-250). Retrieved from http://memberservices.gsdi.org/files/?artifact_id=655

Open Geospaial Consortium (2005): OpenGIS Implementation Specification for Geographic information - Simple feature access - Part 1:Common architecture (pp. 1-51). Retrieved from http://portal.opengeospatial.org/files/?artifact_id=13227&passcode=pcq7e0gzzeat5n7erwhr

Open Geospatial Consortium (2007a): Geospatial Digital Rights Management Reference Model (GeoDRM RM) (pp. 1-130). Retrieved from http://portal.opengeospatial.org/files/?artifact_id=14085&passcode=pcq7e0gzzeat5n7erwhr

Open Geospatial Consortium (2007b): Styled Layer Descriptor profile of the Web Map Service Implementation Specification (pp. 1-53). Retrieved from http://portal.opengeospatial.org/files/?artifact_id=22364&passcode=pcq7e0gzzeat5n7erwhr

Open Geospatial Consortium (2007c): Sensor Observation Service (pp. 1-104). Retrieved from http://portal.opengeospatial.org/files/?artifact_id=26667&passcode=pcq7e0gzzeat5n7erwhr

Open Geospatial Consortium (2008): Web Coverage Service (WCS) Implementation Standard (pp. 1-133). Retrieved from http://portal.opengeospatial.org/files/?artifact_id=27297&passcode=pcq7e0gzzeat5n7erwhr

Portele, C. (editor) (2010a): INSPIRE Generic Conceptual Model (pp. 1-137). Retrieved from http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.5_v3_3.pdf

Portele, C. (editor) (2010b): INSPIRE Guidelines for the encoding of spatial data (pp. 1-38). Retrieved from http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.7_v3.2.pdf

Rackham, L. (2010): Digital National Framework (DNF) – Overview v3.0 (pp. 5-47). Retrieved from http://www.dnf.org/images/uploads/guides/DNF0001_3_00_Overview_1.pdf

Tóth, K. (2010): Tér-tudatos információs társadalom. *Információs társadalom*, *X*(2), 7-16.

Tóth, K. and Tomas, R. (2011): Quality in Geographic Information – Simple Concept with Complex Details. Proceedings of the 25th International Cartographic Conference (pp. 1-11). Paris: International Cartographic Association.

Tóth, K. and Smits, P. (2009): Cost-Benefit Considerations in Establishing Interoperability of the Data Component of Spatial Data Infrastructures. Proceeding of the 24th International Cartographic Conference - The World`s Geo-Spatial Solutions, Vol. XXIV. (pp. 1-10). Santiago de Chile: International Cartographic Association. Retrieved from http://icaci.org/files/documents/ICC_proceedings/ICC2009/

Wade, T. and Sommer, S. (editors) (2006): A to Z GIS. An illustrated dictionary of geographic information science (2nd edition pp. 1-265). Redlands: ESRI Press.

Woolf A. et al. (2010): GEOSS AIP-3 Contribution - Data Harmonization (pp. 1-40). Retrieved from http://www.thegigasforum.eu/cgi-bin/download.pl?f=545.pdf

European Commission
EUR 25280 – Joint Research Centre – Institute for Environment and Sustainability

Title: **A Conceptual Model for Developing Interoperability Specifications in Spatial Data Infrastructures**

Authors: Katalin Tóth, Clemens Portele, Andreas Illert, Michael Lutz, Maria Nunes de Lima

Luxembourg: Publications Office of the European Union

2012 – 57 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1018-5593 (print), ISSN 1831-9424 (online)

ISBN 978-92-79-22552-9 (PDF)
ISBN 978-92-79-22551-2 (print)

doi:10.2788/21003

Abstract

This report addresses the question of how geographic and environmental information created and maintained by different organisations in Europe can be embedded in Spatial Data Infrastructures (SDIs) and reused in various applications by different people. The main challenge related to this task is to deal with the heterogeneity of data managed by others.

The core concept of SDIs is interoperability, which "means the 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, which is used as the main SDI initiative from which this report draws its examples and best practices, is built on the existing standards, information systems and infrastructures, professional and cultural practices of 27 Member States of the European Union in more than 23 languages.

The main part of this report describes the conceptual framework for the development of interoperability specifications that define the targets to which existing data should be transformed. The conceptual framework is composed of two fundamental parts: the Generic Conceptual Model (GCM) and the methodology for data specification development.

The GCM defines 26 aspects or elements for achieving data interoperability in an SDI. These include registers and registries, coordinate reference systems, identifier management, metadata, maintenance, to name just a few.

The description of the methodology for developing data specifications for interoperability includes a detailed discussion of the relevant actors, steps and the overall workflow – from capturing user requirements to documenting and testing the specifications that emerge from this process.

The GCM and the methodology together help to understand the organisational and technical aspects how the data component of an SDI can be established, how interoperability arrangements, data standardisation and harmonisation contribute to this process.

Since 2005 INSPIRE has been pioneering the introduction, development, and application of a conceptual framework for establishing the data component in an SDI. This experience shows that the conceptual framework described in this report is robust enough to reinforce interoperability across the 34 data specifications developed for the infrastructure. Moreover, because the framework is platform and theme independent, able to deal with the cultural diversity, and based on best practice examples from Europe and beyond, it may provide solutions for SDI challenges in other environments too.

LB-NA-25280-EN-N

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new standards, methods and tools, and sharing and transferring its know-how to the Member States and international community.

Key policy areas include: environment and climate change; energy and transport; agriculture and food security; health and consumer protection; information society and digital agenda; safety and security including nuclear; all supported through a cross-cutting and multi-disciplinary approach.

Publications Office

ISBN 978-92-79-22552-9

9 789279 225529