INSPIRE
DATA SPECIFICATIONS

Author: Diederik Tirry, Stijn Keijers (KU Leuven - SADL)

Sources: smeSpire project, 2013.

V: 3.0 | 01/2018

Structure

2

1. Scope and objectives of INSPIRE data specifications
2. The modeling framework for INSPIRE data specifications
3. Development of INSPIRE data specifications
4. Understanding the Technical Guidelines: INSPIRE data specifications

After the training you will be able to:

3

- Explain why INSPIRE requires data specifications (interoperability);
- Understand the main elements of the INSPIRE modeling framework;
- Explain the development process of INSPIRE data specifications;
- Explain how data specifications are elaborated;
- Read & understand the Technical Guidance documents;
- Identify and describe the different levels of metadata (discovery, evaluation, use);

Target audience

4

This seminar aims at :

- GI professionals, ICT professionals and managers who need to understand the requirements of the INSPIRE rules and guidelines.

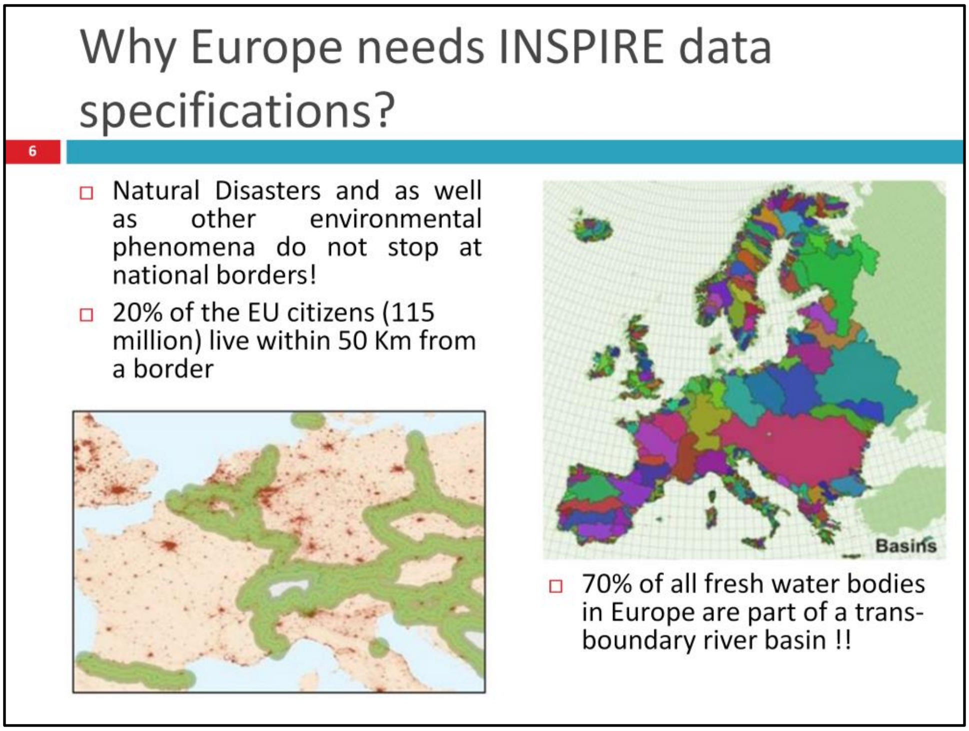
Prior knowledge:

- no explicit pre-requisites are required. However, participants should have a basic understanding of the INSPIRE directive or Spatial Data Infrastructures.

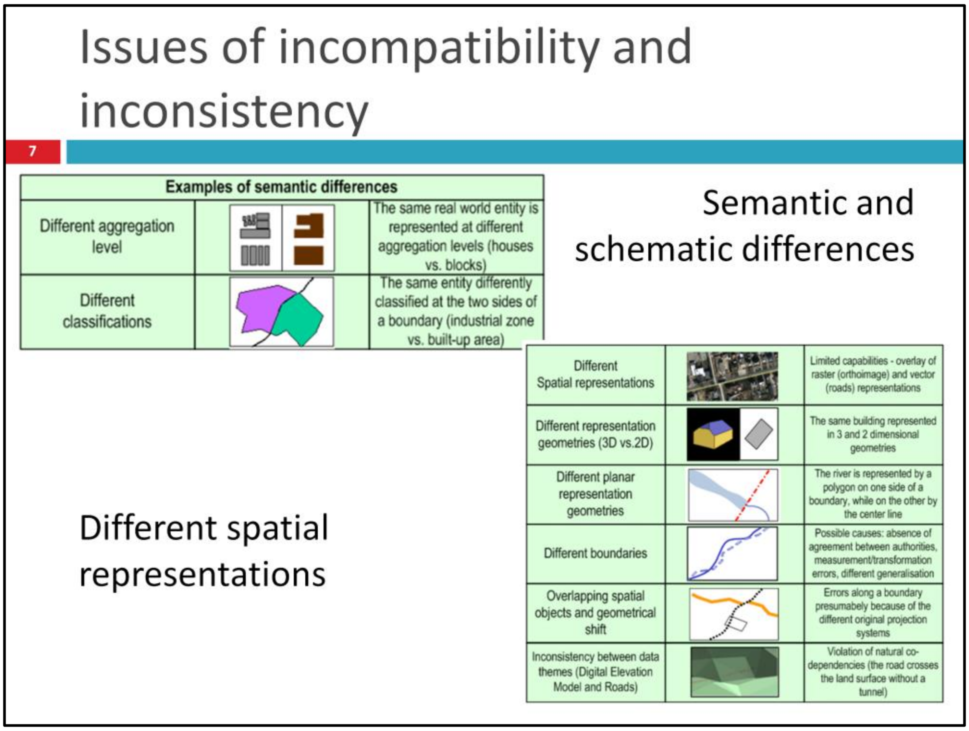
5

Part 1

Scope and objectives of INSPIRE data specifications



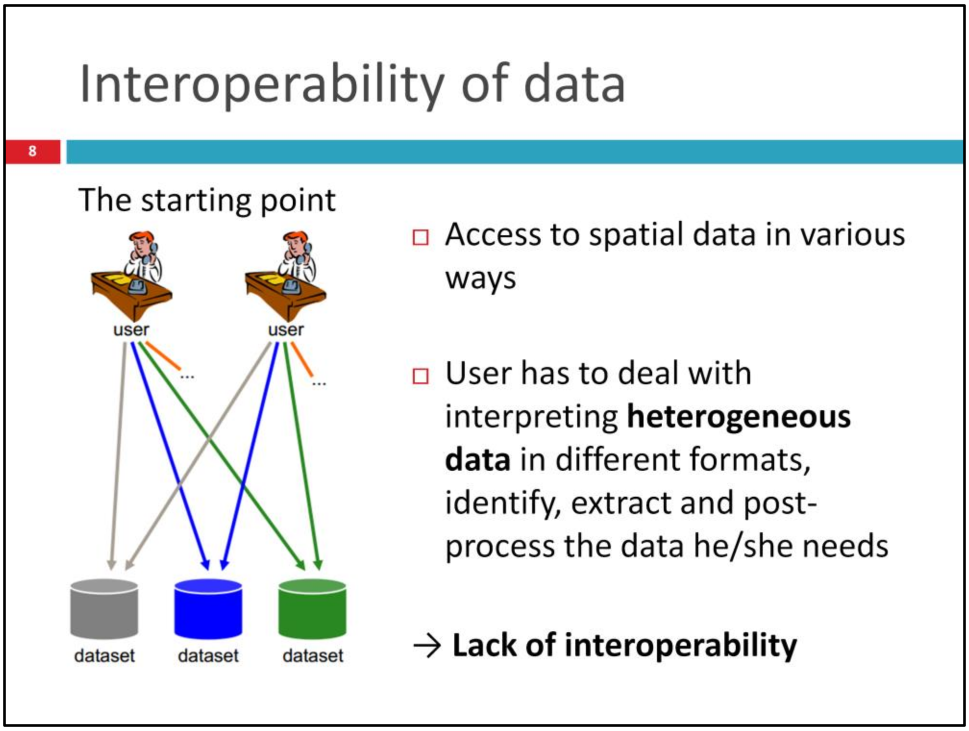
Why do we actually need this INSPIRE DS?
With this slide we would like to illustrate the need for cross border Environmental information.
The fact is that Natural Disasters as well as other environmental phenomena do not stop at national borders.
The picture on the left shows that a substantial part of the EU population lives within a distance of 50km from a EU border. Which also means that many people can be affected by cross border natural disasters. Therefor good EU wide knowledge/information of the environmental phenomena is needed. Until now each MS collected information on Env. Phenomena in its own specific way. In order assemble all this information on a EU scale, there is a need for exchanging this information between MS in a uniform way. This uniform way is described in the data specifications.
The picture on the right is a map of river basins of Europe, as you can see they do not stop a national borders. These basins are important units for reporting on and managing water quality.
These two examples on this slide show that there is a need for a comprehensive strategy to address specific environmental problems, like water pollution, and to come up with guidelines for exchanging data on these environmental topics.



This slide addresses some issues of incompatibility and inconsistency which can happen at the borders (in different MS).

We make a distinction between two types of inconsistencies.

1. On the one hand, the semantic and schematic differences, illustrated here with two examples
   a. different ways of modelling houses: as individual houses or as building blocks (an aggregation of individual houses)
      OR
   a. different classifications for the same entity (industrial zone at one side of the border and Built-up area at the other side of the border.
1. On the other hand, different spatial representations:
   a. Vector vs raster
   b. 2D vs 3D
   c. River as poly or as line (depending on scale but also on the content you want to model and on the use of the data).
   d. Different boundaries → agree to use same geometry
   e. Overlapping and shift
   f. Inconsistency between data themes (a road that does not follow the DEM)



In the coming slides were are going to talk about interoperability of data.
This slide shows the situation when there is no interoperability at all, this reflects how it used to be.

A user had to search for input data by contacting different data producers, all having their own rules for access and use of their data products. In many cases these datasets were retrieved in different formats (f.e. shapefile, ASCII-files, …) and it was up to the user to interpret this heterogeneous data and to identify, extract and post process the data in order to end up with a harmonised dataset to be used for further analysis.

So, in the future this is something we would like to solve. The whole process of searching, retrieving and combining different data sources should be made much easier and preferably in automatic way.

Levels of heterogeneity

9

- **Syntactic** heterogeneity

  Data may be implemented in a different syntax of **different paradigms**, such as relational or object- oriented models. Syntactic heterogeneity is also related to the **geometric representation** of geographic objects, e.g., raster and vector representations.

- **Structural** or **schematic** heterogeneity

  Objects in one database are considered as properties in another, or object classes can have **different aggregation** or **generalisation hierarchies**, although they might describe the same Real World concepts.

We just talked about the lack of interoperability in different data sources.
Let's have a closer look now to the different levels of heterogeneity.

The first level of heterogeneity is what I call the "Syntactic Heterogeneity". Data may be implemented in different syntax which can be translated as data that is stored in different formats: shapefile, ascii-file, relational versus object-oriented database models. But it is also related to the geometric representation of geographic objects f.e. some datasets may use the vector approach while other datasets are using the raster approach to depict the same type of objects.
With the existing technologies of today this is not an invincible problem. For most of the cases conversion algorithms are offered within the most commonly used software products. It would be very convenient though that everyone uses the same syntax so conversion will be unnecessary.

The second level of heterogeneity is the Structural or also called the schematic heterogeneity. This has to do with how you model the information in a certain structure inside the database. So even if you use the same syntax (or database format) and you deal with objects/concepts with a semantic meaning then you still can model them in a different way in different schemes. If you ask two people to model a certain piece of information in a same database model, you probably end up with two different database schemes but they cover the same information/semantic concepts. It all concerns the way the information is organised inside the database. You can have different levels of aggregation or detail. Objects in one schema can be considered as properties in another but in the end it concerns the same Real World concepts. You can see a schema as the personal perception of a somebody on a real world concept.
This second level of heterogeneity is also of a technical kind and can be solved by actual technologies that can align different schemas describing the same real world concepts. But again it would be very convenient that everybody applies the same schemas.

Levels of heterogeneity

10

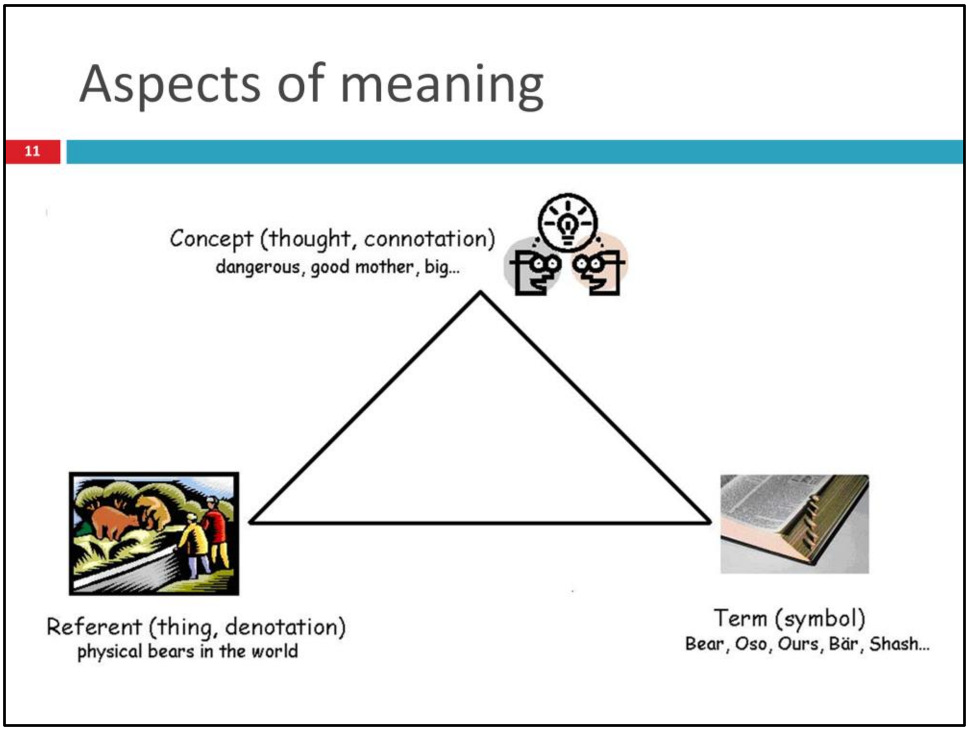
- **Semantic** heterogeneity

  A Real World concept may have more than one meaning to comply with various disciplines, giving as a consequence semantic heterogeneity.

  For example a road network for pavement management has different semantic descriptions from transportation data maintained in a GIS database designed for small scale topographic mapping applications.

The 3rd level of heterogeneity is the semantic heterogeneity.

This is the most complex level of the three to deal with. It means that people are giving different meanings to the same real world concept. This is not strange and it probably depends on the discipline from which point of view you are looking at the real world concept.
A typical example is the geographical object: road.
From the point of view of the road authority, responsible for the quality of the pavement of the road, the concept road will get a different meaning than looked at it through the eyes of the people responsible for the traffic flows, for them the type of pavement (concrete or asphalt) is not important they are more interested in the connectivity of the roads in order to manage traffic flows, traffic jams and so on.
The same concept can have different meanings and it is the challenge here to align these different meanings. So if in one MS they talk about a certain type of soil we would like to know that this type of soil has the same name, the same meaning in another MS.



This is another slide to explain how these different aspects of meaning occur.

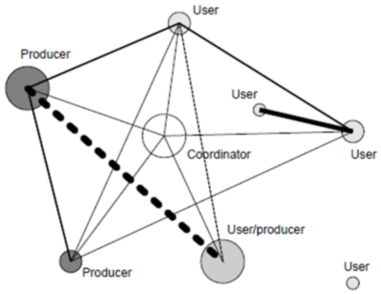
When we speak about f.e. a bear one of the aspects of meaning occurs in our mind, what are the first thoughts that come in our mind. This can differ from person to person.
But it also occurs when translating concepts into a language, which words do we use to describe, which terms are used in other languages and what is the definition given to these terms.
And the third aspect of meaning is of course the real thing, in the case of our example the physical animal: the bear.

All three aspects are connected but in order to be on the same terms of speaking we need to make good agreements on defining concepts that refer to the real thing and which terminology and symbols will be used to represent and to describe it.

Interoperability of data

12

- **Technical** interoperability
  - should guarantee that system components can interoperate
- **Semantic** interoperability
  - should guarantee that data content is understood by all in the same way



The scope now is to make data interoperable and to solve the different levels of heterogeneity that we talked about in the previous slides. The result would be that each producer and each user can easily exchange data amongst each other with our without the aid of a coordinator in between.
As I said before establishing technical interoperability in not that difficult because todays applications and software are able help in the harmonisation process. It only requires means and actions.
Implementing semantic interoperability however is something else. Here we have to make sure that the data content is understood by everybody in the same way. It means we all have to agree on using harmonised definitions and classifications for objects while at this moment many different definitions and classification systems exist for the same objects in different countries. The idea would be to map the existing heterogeneity to these universal definitions and classifications so data becomes exchangeable and understood by everybody in the same way.

What INSPIRE is aiming at

13

- Key requirements of the INSPIRE directive
  - Art 3(7): "**Interoperability** 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"
  - Art 7(1): "**Implementing rules** laying down technical arrangements for the **interoperability** and, where practicable, **harmonisation** of spatial data sets and services … shall be adopted…. Relevant user requirements, existing initiatives and international standards for the harmonisation of spatial data sets, as well as feasibility and cost-benefit considerations shall be taken into account in the development of the implementing rules."

The needs for interoperability are also translated in the INSPIRE directive as key requirements as you can read in the following articles:
Interoperability means the possibility for spatial data sets **to be combined**, **without repetitive manual intervention** (no need for preprocessing anymore), and in such a way that the result is **coherent** (it means that the result can be interpreted in its totality).
In another article we can read that INSPIRE will lay down implementing rules for reaching this interoperability and where practicable harmonization of spatial datasets. To develop these rules existing initiatives and especially international standards for the harmonization of spatial datasets will be taken into account.

What INSPIRE is aiming at

14

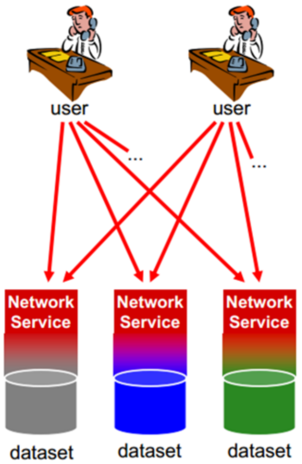
- Key requirements of the INSPIRE directive
  - Art 8(2): The implementing rules shall address the following aspects of spatial data:
    - a **common framework for the unique identification of spatial objects**, to which identifiers under national systems can be mapped in order to ensure interoperability between them;
    - the relationship between spatial objects;
    - the **key attributes** and the corresponding multilingual thesauri commonly required for policies which may have an impact on the environment;
    - information on the temporal dimension of the data;
    - updates of the data.
  - ...

In article 8 we see that these implementing rules will address a certain number of aspects of spatial data:
One of these aspects is the use of a common framework for the unique identification of spatial objects, which means that each spatial object will be identifiable through a persistent and unique key.
Another important aspect is that there will be some key attributes defined for every geographical object and that multilingual thesauri will be required for different policies which may have an impact on the environment.
Based on this articles we can say that INSPIRE is about making existing heterogeneous data in different MS, exchangeable and interoperable within the whole European Union.

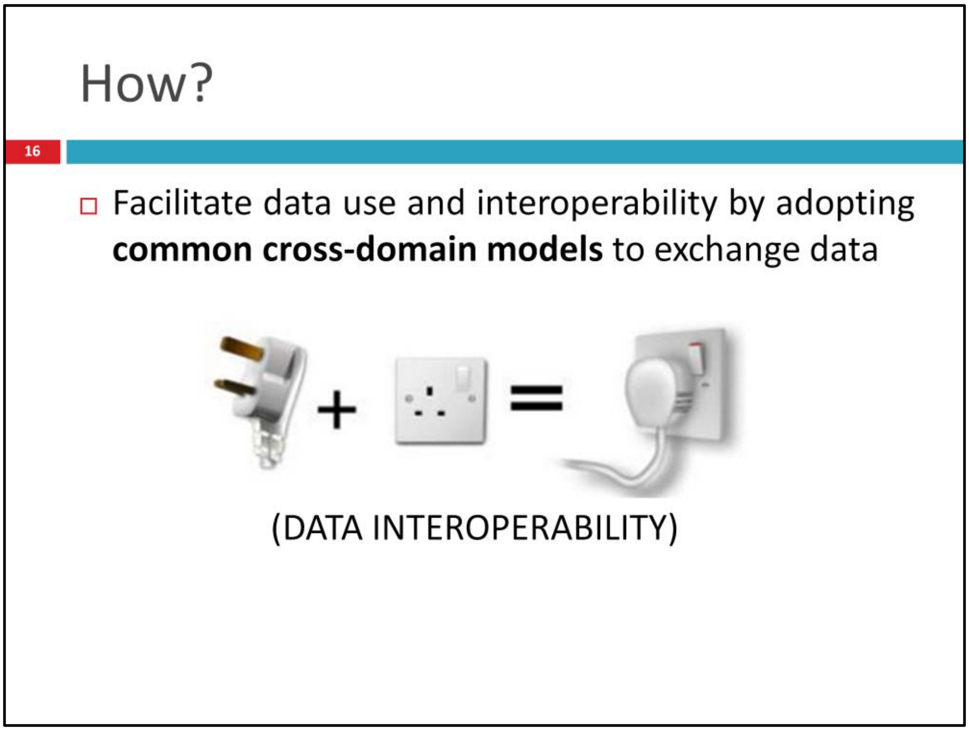
What INSPIRE is aiming at

15

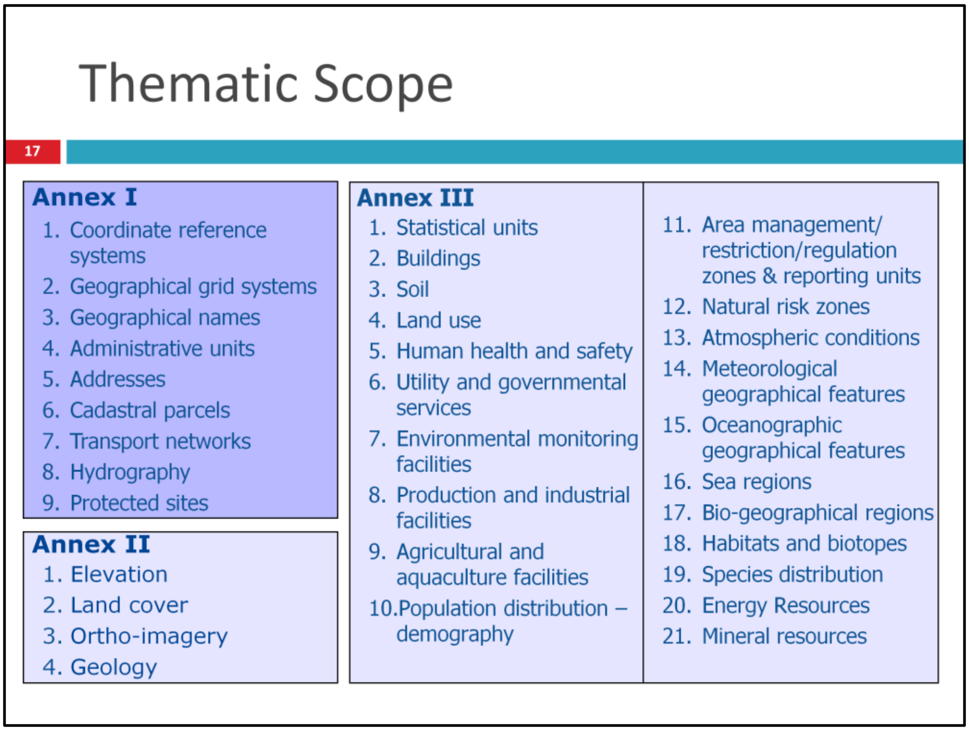
- Provide access to spatial data via **network services** and according to a **harmonised data specification** to achieve **interoperability** of data
- Datasets used in Member States may stay as they are
- Data or service providers have to provide a **transformation** between their **internal model** and the **harmonised data specification**



This slide shows in a graphical way what INSPIRE is aiming at:
Instead that every user is searching with his own means for a dataset of his needs, spending a lot of time in pre-processing it to the desired format in order to combine it with other datasets, INSPIRE would offer access to spatial data via network services and the data that would be made available follows the structure of a harmonised data specification to achieve interoperability.
There is an important remark to make about this provision of interoperability. The MS are not obliged to modify the datasets that are internally used. But at least the data providers should offer a transformation between their internal model and the harmonised data specification which can be exchanged.



Right plugHow can we make this facilitation of data access, data use and interoperability work. It is through defining common cross-domain models that will be used to exchange the data. You can compare it with selecting the right plug that fits the socket.
s and sockets.



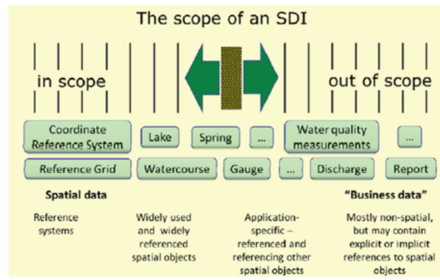
Until now we talked about the need for data specifications in order to achieve interoperability. But what is now the thematic scope of INSPIRE? Which thematic domains are addressed by INSPIRE?
In general we can say that INSPIRE is aiming at thematic domains that concern environmental policies. This slide shows the complete list of thematic domains (34 in total) which are grouped in 3 annexes because of different priorities, different timeline and deadlines to implement data components of the INSPIRE directive.

Annex 1 themes are the basic reference data and they were the first themes to be modelled because themes from the other annexes could depend on them. At this stage in the INSPIRE process the DS for all themes are ready and available through the INSPIRE website.

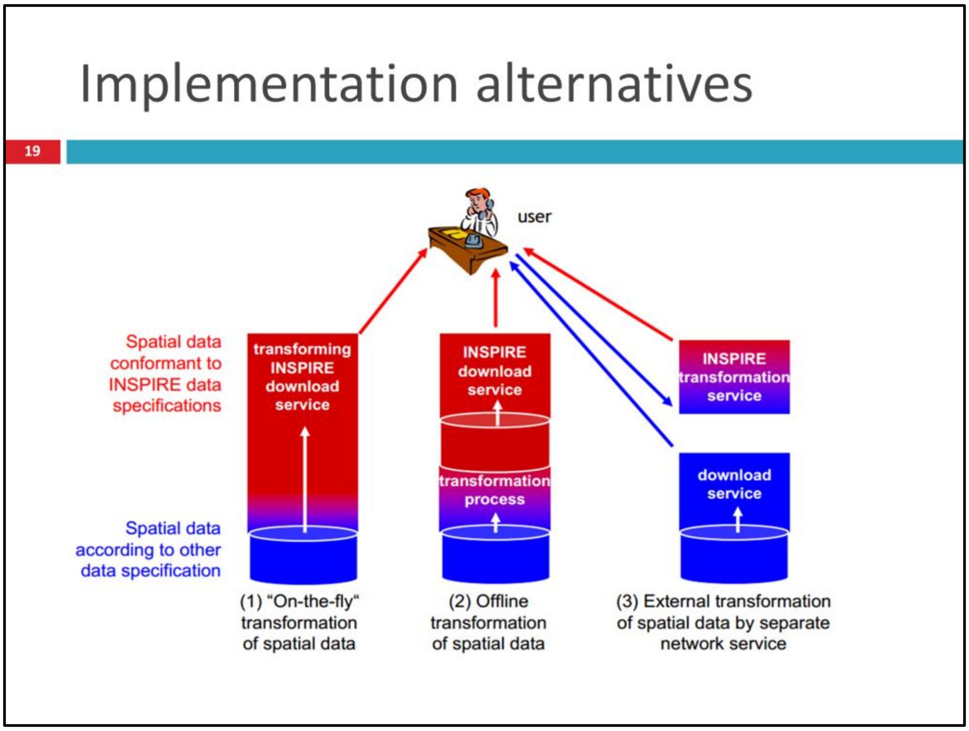
INSPIRE data scope

18

- Scope is **spatial** data – not all kinds of thematic data
- MS are encouraged to re-use the INSPIRE data specs for their own usage.
  - Extensions
  - Additional constraints
  - Re-use of INSPIRE objects



We have seen now the thematic domains that INSPIRE focusses on. But the next question is: What kind of thematic data falls within the scope? Here we often can notice misinterpretations of the directive The data scope of INSPIRE is strictly "Spatial data" and not all kind of thematic data. It really concerns the spatial objects that can be identified within a certain thematic domain or within multiple domains, including some crucial key attributes necessary to describe these objects. Business data, which mostly is non-spatial but which can be spatially referenced, is out of scope. As an example: Information from water quality measurements in surface water is out of scope but the spatial objects delineating these surface waters are definitely within the scope.
MS are encouraged to re-use the INSPIRE DS for their own usage and there is a possibility to extend them for example to include business data, but it is not an obligation.



What are now the different implementation alternatives to reach this harmonisation towards the INSPIRE DS? At the start of the INSPIRE story, three different implementation strategies were proposed:

The first one is the on-the-fly transformation of spatial data through download services that automatically transform the source data to the schema of the INSPIRE DS. The data is on the fly processed after the request for the data has been send by the user;

The second proposed implementation strategy, is the offline transformation of spatial data. It means that each data provider transforms his own data into the appropriate INSPIRE DS and offers the result to the user by means of an INSPIRE download service.

The third possibility is the external transformation of spatial data by a separate network service. Here the idea is that the user first downloads the source dataset and makes use of a separate transformation service to transform the data into the INPSIRE DS.

At this point in time, the implementation of INSPIRE is in full process and it became clear that the second option is the most preferred one and also the most implemented strategy.

Why are the other options not so popular?

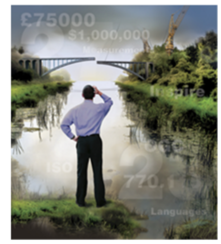
The first alternative is technically possible but the disadvantage is that the processing happens on-the-fly and this requires time to retrieve the final result, certainly with large datasets.

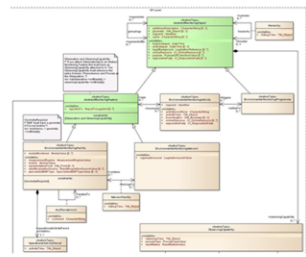
The last option has the same problem as the first one, but here is also another reason why it is not commonly used in practise: each dataset requires a specific transformation to arrive to the DS, so it means that a specific transformation service must be created for each existing data source, which makes it difficult to keep the overview of all services.

Common Data Specifications goal

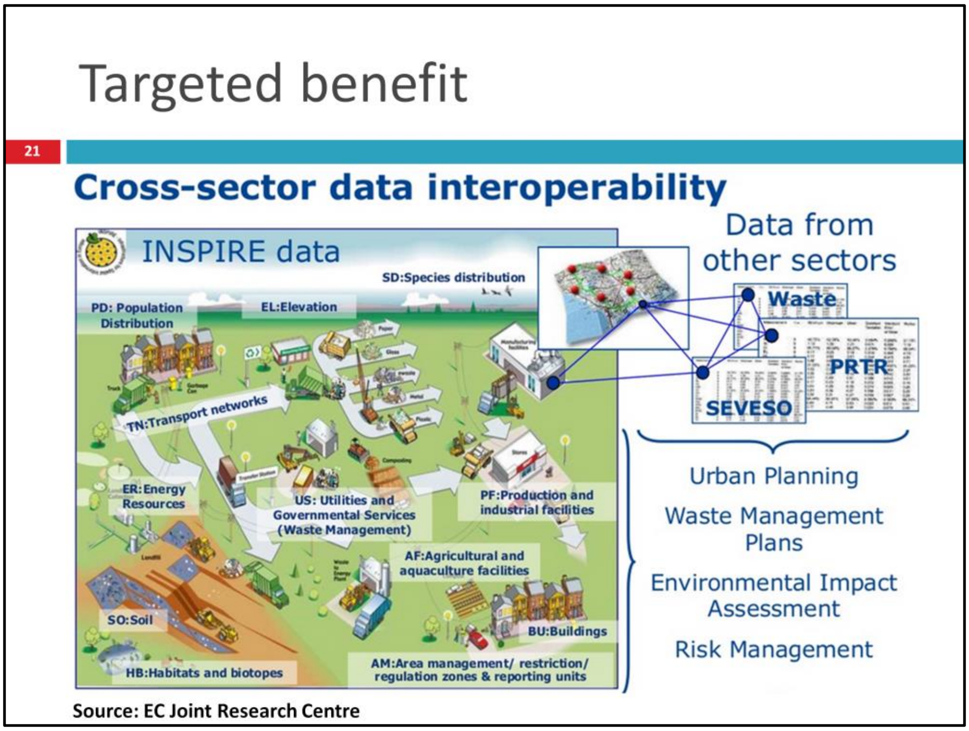
20

- Member States should make available data within the scope of INSPIRE using
  - the same spatial **object types** (and definitions)
  - the same **attributes** (and definitions, types, code lists) and relationships to other types, e.g. BuildingHeight, BuildingSize
  - a common **encoding** (GML application schemas)
  - common **portrayal** rules
- This facilitates interoperability and pan-European/cross-border applications (e.g. information systems, reporting systems, forecasting models)





The common data specification goal can be resumed as follows:
When MSs make their spatial data available within the scope of INSPIRE it means by using the same spatial object types, same attributes and relationships with other types, using a common GML encoding and common portrayal rules as specified in the DS technical guidelines, then interoperability will be facilitated. This will result in valuable Pan-European/cross-border applications which can benefit from the availability of harmonised datasets. Applications that can be useful to policy makers or applications useful to risk management and so on…



So when the whole spatial data infrastructure of INSPIRE will be established by the year 2020 it would look like on the picture: we see a situation where there is cross-domain/ cross-sector interoperability of spatial data. This makes it easy to exchange and combine different datasets from different sectors in order to perform all kinds of analysis.

At this moment we are not there yet, the implementation process is still in progress.

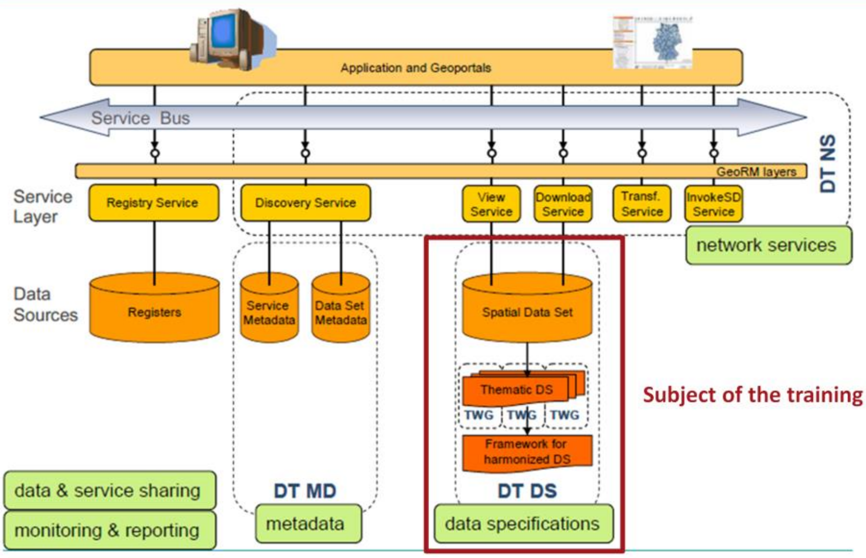
22 Part 2

The modeling framework for INSPIRE data specifications

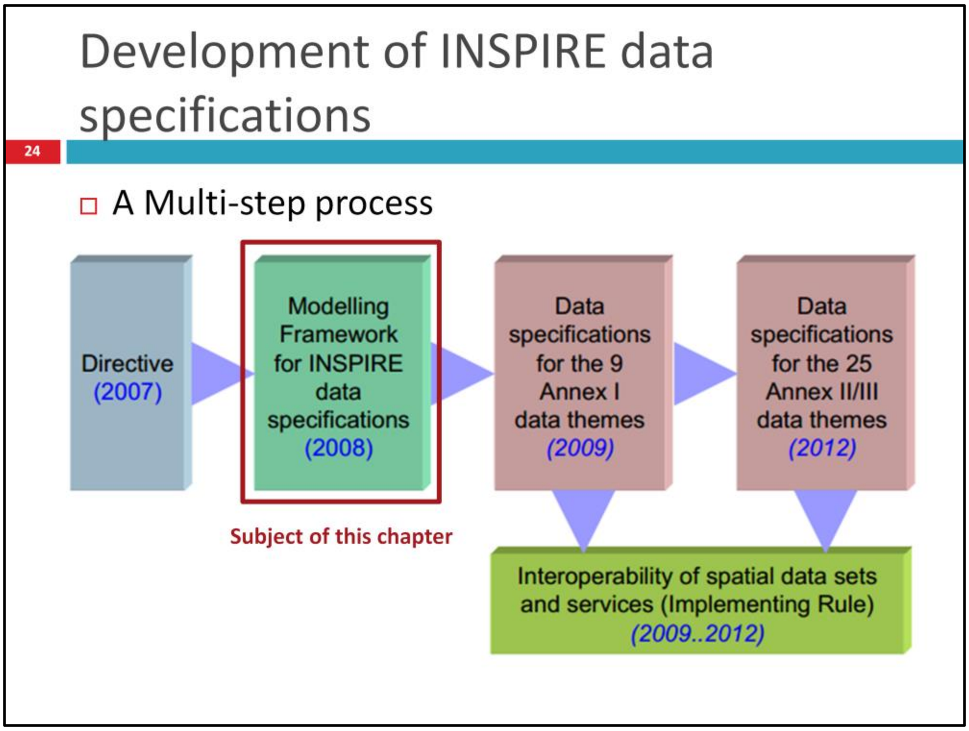
In part one we talked about the scope and the objectives of the Inspire DS.
Part two will focus on the modelling framework for the INSPIRE DS. What is behind the scenes or what are the cornerstones or building blocks for these DS?

INSPIRE Architecture

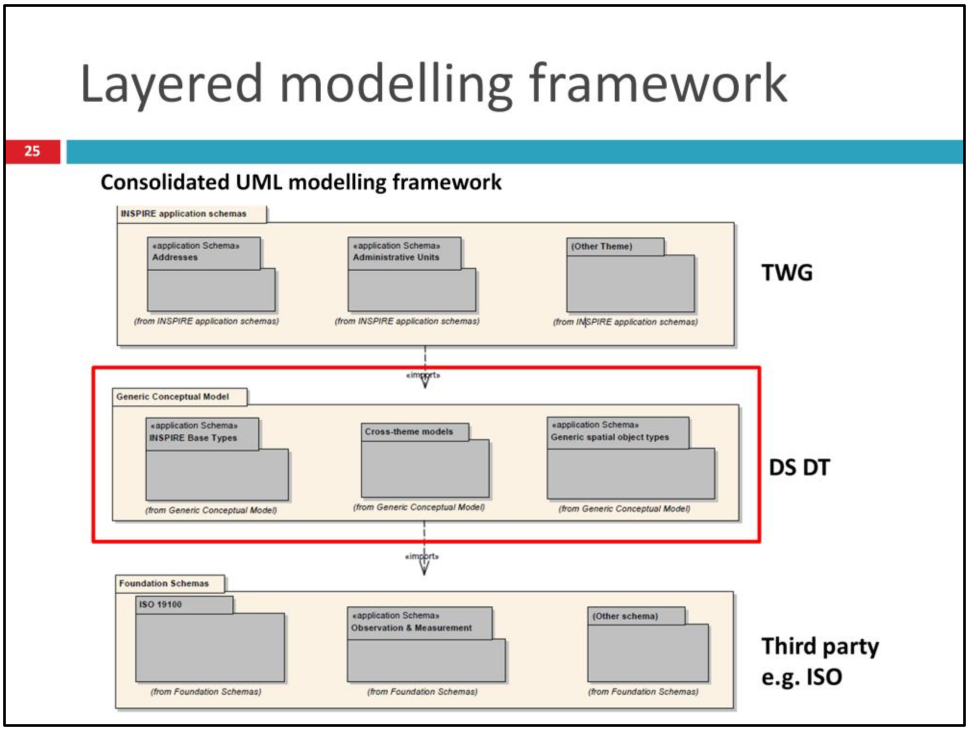
23



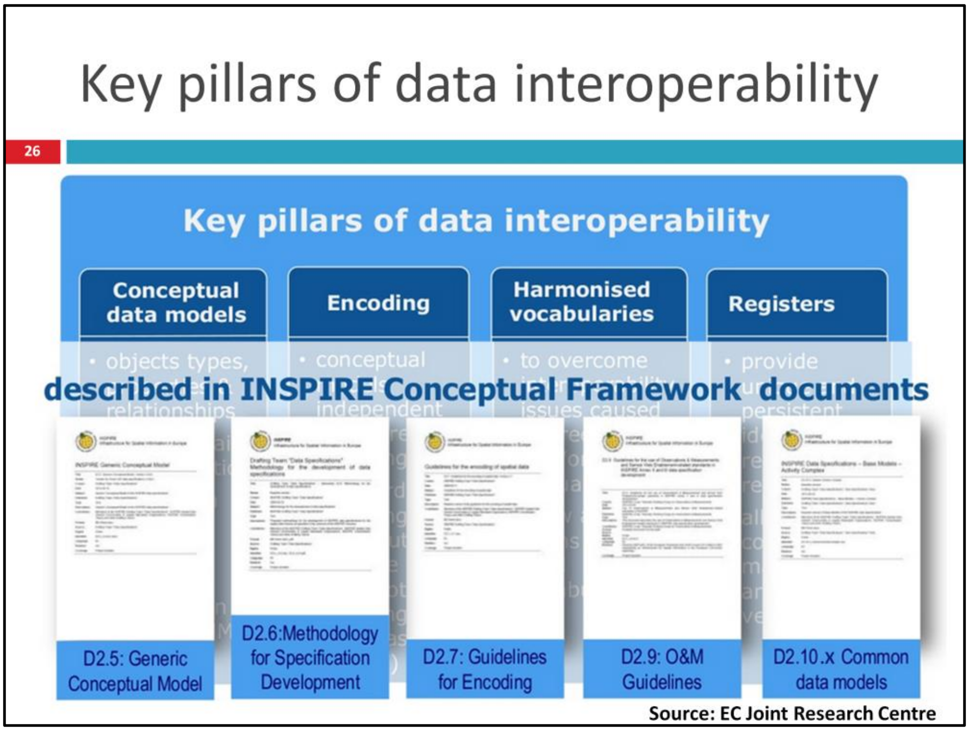
Let's first have a look at the overall INSPIRE architecture to position the DS in the whole context. This picture you probably have seen before.
At the bottom you have the data components of the infrastructure containing the spatial data itself but also the metadata that describe the data and that make it possible for user to find and to understand the offered data sets. The data component side also holds different registers to store information commonly used by different datasets: code lists, thesauri, and so on...
On top of the data layer we find the service layer containing services that can be called by users and applications to retrieve information on the data (like metadata or in the form of webmaps) or to retrieve the data itself. These services are grouped/collected by the "Service Bus". In between the services and the service bus there might be a layer managing the access to the different services the Geospatial Rights Management layer (GeoRM).
And completely at the top you'll have the application and geoportal layer that will be used by the users of data.
The subject of this training module is the part indicated by the red rectangle it concerns the DS and the framework used to build these common DS.



The development process of the INSPIRE DS was a complex multi-step process.
It all started with the approval of the INSPIRE directive in 2007, one year later the modelling framework was created by a pool of experts and the following years the DS were elaborated, starting with the annex I themes, followed a few years later by the Annex II and III themes. Those DS were then integrated in the Implementing Rule on interoperability of spatial data sets and services.
In this chapter we will focus on the modelling framework for the DS



Here you get an overview of the modelling framework as a layered UML modelling framework.
The basis is formed by international standards like the ISO191xx series, O&M schemas,… On top of that we have the Generic Conceptual Model as a middle layer which provides some useful common application schemas that define some essential INSPIRE base types (f.e. the INSPIRE Id) and also generic spatial objects that can be used by multiple themes. It was created by the DS Drafting team. And then we have the third layer which is established by the thematic working groups, which contains the developed application schemas of the data models.
This whole framework was modelled in UML and it can also be downloaded from the INSPIRE website.

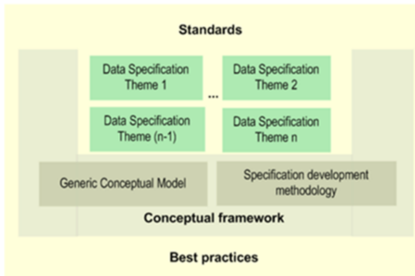


The INSPIRE modelling framework is based on 4 cornerstones of assuring the data interoperability.
The most important one, is the pillar of the conceptual data models, needed for cross-domain harmonisation. These models define for each theme the essential common object types, their properties and the relationship to other object types and are managed in a common UML repository.
The second pillar is the “encoding”. We need to use a common way to encode the conceptual data models. The standard approach in INSPIRE is to encode all data sources as GML which is the most common file format for exchanging geographical data. But also other encodings might be necessary f.e; for rasterdata, where GML is not so suited, we can use GeoTiff or other raster formats.
The third pillar is the use of harmonised vocabularies which is important to overcome different semantic meanings. So it concerns common classifications and code lists that list the definitions and values, translated in the different official languages, to be used by all data providers. Some vocabularies allow additional terms from local vocabularies.
And that brings us to the 4th cornerstone: the registers:
The vocabularies will be managed in an online register which allows a consistent management. The register will provide unique and persistent identifiers for reference to resources. Registers will also be used to maintain the versioning of concepts, concepts can evolve overtime, some become deprecated while new concepts will be defined.
All these cornerstones are described in INSPIRE Conceptual Framework documents (the D2.x range of documents).

INSPIRE conceptual framework

27

- To provide a repeatable data specification development methodology and general provisions for the data specification process, which is valid for all spatial data themes



We have talked now about the Modelling framework, how it fits in the development process of the INSPIRE DS. That it is structured in layers. And on which key concepts of interoperability it relies. But you still might wonder: What is the use of it?
Since we would like to harmonise data models for different themes it is important to have a common basis to start from. And this common basis is provided by the Conceptual Model. You can see it as a number of best practises in the form of common/general defined concepts and a repeatable DS development methodology which is applicable for all data themes. So it forms the basis for all DS and this is also reflected in the DS itself as common chapters and paragraphs as we will see later when looking in to detail to the DS.

INSPIRE conceptual framework documents

28

Framework documents D2.X

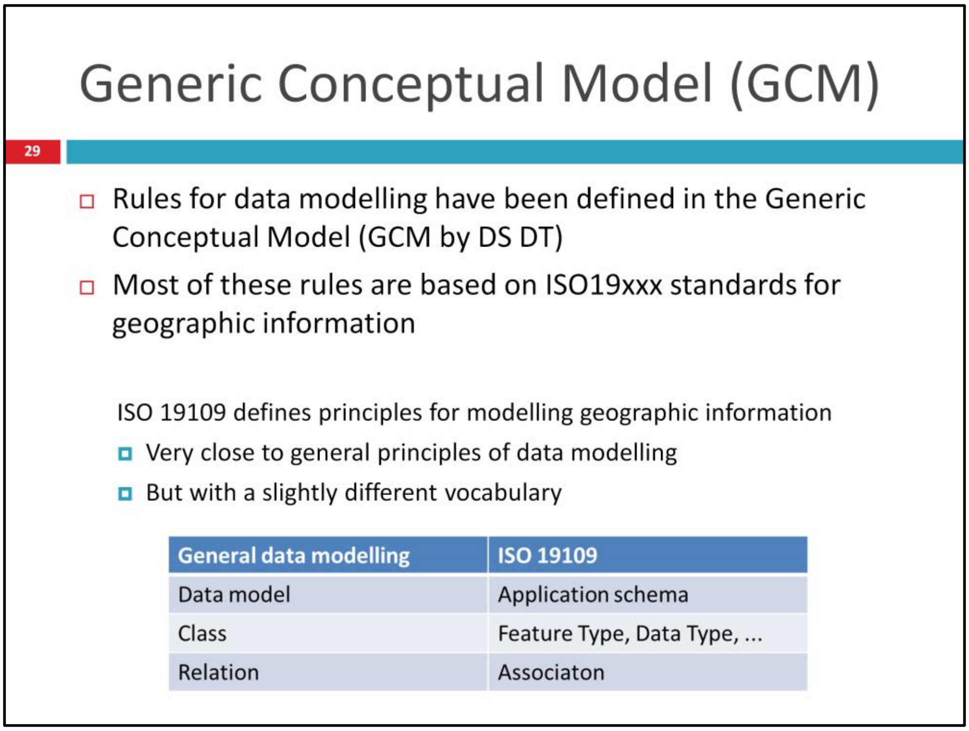
- Generic Conceptual Model **(D2.5)**
- Methodology for specification development **(D2.6)**
- Guidelines for the encoding of Spatial Data **(D2.7)**

  See “Harmonisation aspects” => P. Data Transfer
- Common data models **(D2.10.x)**

  The need arose during the modelling of data specifications to model concepts that are common to different INSPIRE themes
- O&M guidelines **(D2.9)**

  The need arose during the modelling of data specifications to provide **common guidelines** for modelling “Observations and Measurements”

As I mentioned before, when talking about the major cornerstones of data interoperability, the INSPIRE Conceptual Framework is described in several documents which are listed here.
In the following slides I would like to go in more detail to the most important of these framework documents which is the generic conceptual model.



The Generic Conceptual Model (abbreviated as the GCM) is laying down the rules for data modelling and was created by a group of experts (DS DT).

Most of the rules in this document are based on existing international standards from the ISO19xxx series, concerning geographic information. So the idea was not to reinvent the wheel but to reuse as much as possible from these well-known standards.

In the ISO19109 defines the principles for modelling geographic information which is not much different from the general data modelling apart from the vocabulary used:

Some examples that demonstrate this difference:

- Data model vs application schema
- Class vs Feature type (specific class for identifiable spatial objects) /Data type
- Relation vs association

Generic Conceptual Model (GCM)

30

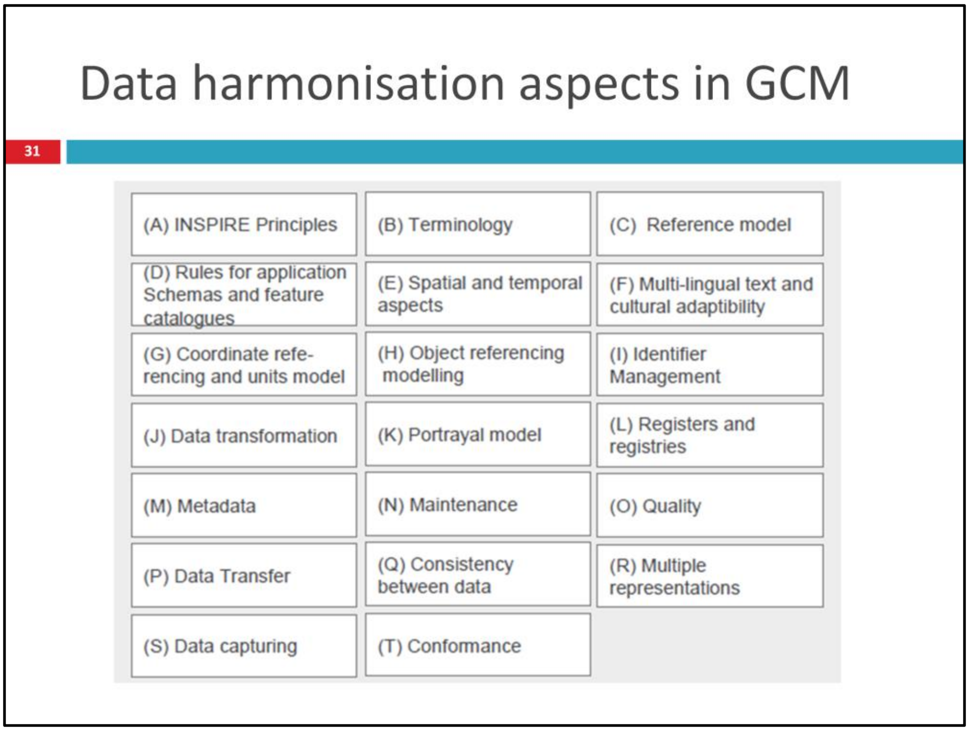
- The INSPIRE Data Specifications drafting team (DS DT) has defined the "Generic Conceptual Model" (GCM) as a foundation for the work on the Annex Theme data specifications in the respective Theme Working Groups.

- The GCM is not a "ready-to-use" common data model for INSPIRE, it involves the following aspects:
  - Common terminology and basic concepts
  - requirements and recommendations for the creation of the Theme data specifications;
  - a set of base types to use in the Theme data models (called application schemas in INSPIRE)
  - Specification of cross-theme concepts including a INSPIRE identifier, Generic Network model and Gazetteers

GCM also known in the INSPIRE community as the D2.5 document is meant to be the foundation for the modelling work performed by the TWG for all the Annex theme DS.

It is not a “ready-to-use” common data model by it highlights certain common aspects, like:

- The use of a common terminology
- It lists requirements and recommendations for DS creation applicable to all themes
- It defines and describes a set of base types that can/should be used by all themes in their application schemas
- It includes some cross-theme concepts that were identified from the start, like the INSPIRE identifier (what is it? And how should it be constructed), it contains a Generic Network model for themes that deal with networks (fe the theme transport networks), the concept of a gazetteer (which is a kind of geographical dictionary to find specific geographic entities).



We already saw that the GCM is the guidance for good modelling practises for all DS. The result of the modelling should be a harmonised set of DS for all themes.

This harmonisation can be reached by following the different harmonisation aspects covered in the GCM and they are listed on this slide (from the INSPIRE principles over identifier management, data capturing requirements to conformance issues). In the next slides we will highlight some of the most important harmonisation aspects.

Harmonisation General Principles

32

A. INSPIRE principles	•that spatial data are stored, made available and maintained at the most appropriate level; •that it is possible to combine spatial data from different sources across the Community in a consistent way and share them between several users and applications; •that it is possible for spatial data collected at one level of public authority to be shared between other public authorities.
B. Terminology	•use of a **consistent language** when referring to terms via a **glossary** •The ESDI needs to establish a **common terminology**
C. Reference Model	the **framework** of the **technical parts** •information modelling and data administration •the components to be described in a consistent manner.

We can distinguish some categories in the list of harmonisation aspects. First of all there are some “General principles”:

- Of course these include the INSPIRE principles saying that:
    - spatial data must be stored and made available on the most appropriate level;
    - That it must be possible to share and combine spatial data from different sources in a consistent way;
- another principle is the use of consistent language when referring to terms to overcome semantic problems. This can be reached by making use of a glossary with common definitions.
- but also the possibility to fall back on a reference model is one of the base principles. It is the framework of all the technical parts that support us in modelling information and administration of data;

INSPIRE principles: Data harmonisation is a methodology to reach these goals.
Terminolgy: This component will support the use of a ***consistent language*** when referring to terms via a ***glossary***. This needs to be registered and managed through change control with ***multi-lingual support***. The ESDI needs to select a common terminology from all of the existing terminologies and/or their translations.
Ref Model: This component will define the framework of the technical parts including topics like information modelling (i.e. conceptual modelling framework with rules for application schemas) and data administration (i.e. reference systems). It will provide a structure which allows the components of INSPIRE which are related to data specifications to be described in a consistent manner.

Harmonisation Schemas

33

D. Rules for application schemas and feature catalogues	•**Feature catalogues** define the types of spatial objects and their properties •The **full description of contents and structure** of a spatial dataset is given by the application schema which is expressed in a **formal conceptual schema language**
E. Spatial and temporal aspects	**Conceptual schema** for describing the spatial and temporal characteristics of spatial objects: - **Spatial** geometry and topology - **Temporal** geometry and topology - Coverage functions
G. Coordinate referencing	spatial and temporal **reference systems, units of measurements**, the parameters of transformations and conversions. **European geographical grids**.
H. Object Reference Modelling	This component will describe how information is **referenced to existing spatial objects**, typically base topographic spatial objects, rather than directly via coordinates.

Beside the general principles we also have some principles related to schemas:

One aspect will deal with the question: How to describe these schemas? If everybody is allowed to use his own way of documenting these schemas, we will end up again with an undesired heterogeneity. So there must be some general rules that define a common way of documenting. The rule is to establish feature catalogues which define the types of spatial objects and their properties. While the application schema gives a full description of the content and structure expressed in a conceptual schema language like UML.

Harmonising data also means dealing with spatial and temporal aspects. There is a need for common information on spatial geometries and topology (what are the geometry types to be used and what are their characteristics) but also on the way temporal characteristics of data will be managed.

When talking about exchanging spatial data we cannot go around the aspect of coordinate referencing. So instructions/guidelines/recommendations towards common European reference systems, including the definition of European geographical grids, are needed for creating a harmonised data space across Europe.

To avoid storing geometries more than once, it is necessary to have a mechanism to reference information to existing base-topographic/spatial objects, f.e. you can use the geometry of the theme "buildings" to add other thematic information like Industrial facilities by referencing the relevant geometry instead of redefining it. How this should be done, is tackled in the GCM by the aspect of Object Reference Modelling.

A feature catalogue define the types of spatial objects and their properties (attributes, association roles, operations) as well as constraints and are required when turning the data into usable information.
An application schema is the full description of the contents and structure of a spatial dataset is given by the which is expressed in a formal conceptual schema language.
The feature catalogue defines the ***meaning*** of the spatial object types and their properties while the application schema describes the ***formal structure***. Text elements in the feature catalogues should be maintained at least in the official European languages.

In order to make it easier in making application schemas, three other general schemas have been developed

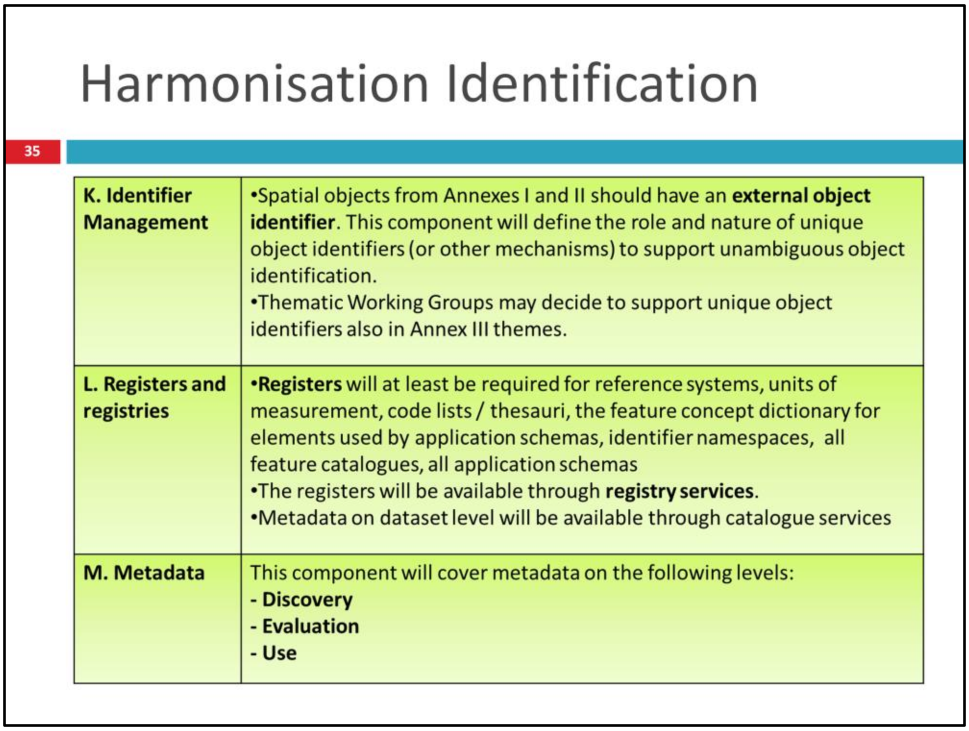
E <read>: Examples of coverage are raster's, triangulated irregular networks, point coverage's, polygon coverage's etc. While the component "reference model" specifies an overall framework, this component deals with the spatial and temporal aspects in more detail, for example, the types of spatial or temporal geometry that may be used to describe the spatial and temporal characteristics of a spatial object.

Harmonisation Translations

34

F: Multi-lingual text and cultural adaptability	•Feature catalogues, feature concept dictionaries, definition and geographical names, attributes / associations and enumerations / codelists •Application schemas not planned to be multi-lingual
I: Data translation model	This component is about translating from a national/local application schema to the INSPIRE application schema and vice versa. Translations are required for data and for queries.
J: Portrayal Model	This component will define a model for portrayal rules for data according to a data specification. It will clarify how standardised portrayal catalogues can be used to harmonise the portrayal of data.

The following slide brings us to the category of translation aspects.
As the INSPIRE directive will be applicable in all European MS. It is logical, in order to make the DS understandable by all MS, that translations are made in all official European languages. This at least for the most crucial information.
Data translation from a local application schema to the INSPIRE application schema of the relevant theme is also addressed in the GCM.
Another kind of translation is the “portrayal model”. This will clarify how standardised portrayal catalogues can be used to harmonised the portrayal of data, i.e. object of the same type will be displayed with the same symbol even if they come from different data sources.



Harmonisation Identification

35

K. Identifier Management	•Spatial objects from Annexes I and II should have an **external object identifier**. This component will define the role and nature of unique object identifiers (or other mechanisms) to support unambiguous object identification. •Thematic Working Groups may decide to support unique object identifiers also in Annex III themes.
L. Registers and registries	•**Registers** will at least be required for reference systems, units of measurement, code lists / thesauri, the feature concept dictionary for elements used by application schemas, identifier namespaces, all feature catalogues, all application schemas •The registers will be available through **registry services**. •Metadata on dataset level will be available through catalogue services
M. Metadata	This component will cover metadata on the following levels: **- Discovery** **- Evaluation** **- Use**

Next category of harmonisation aspects is concerning the identification of objects.

The identifier management is one of the key issues in INSPIRE. It states that every spatial object (at least for annex I, II) should get a unique and persistent identifier (i.e. INPSIRE identifier). All annex III themes followed this good practise although it was initially not required. The "INSPIRE id" makes it possible to reference each object by means of its identifier.

Registers are very important in the harmonisation process. They are functioning as the libraries/dictionaries that hold the information commonly agreed on. There will be registers for listing possible reference systems, listing possible "UOM"s, code lists (i.e. values of classifications) used in the different themes, thesauri, …

Those registers will become available through registry services so they can be used by other models or applications.

Metadata play also an important part in the identification of datasets and services. Information stored in the MD gives the user an idea of what he may expect from the described dataset of service. There are different levels of MD as we will see later on in this module.

Harmonisation Data Quality

36

N. Maintenance	•**change only updates** •**Versioning** of objects •spatial object life-cycle rules
O. Quality	This component will advise the need to publish quality levels of each spatial data set This will include methods of best practice in publishing: - Acceptable quality levels of each spatial data set - Attainment against those levels for each spatial data set
Q. Consistency between data	**Consistency** along or across borders, themes, sectors, or at different resolutions
R. Multiple representation	This component will describe best practices how data can be **aggregated** - across time and space - across different resolutions (“generalisation” of data)

A last category of harmonisation aspects concerns “data quality” elements.

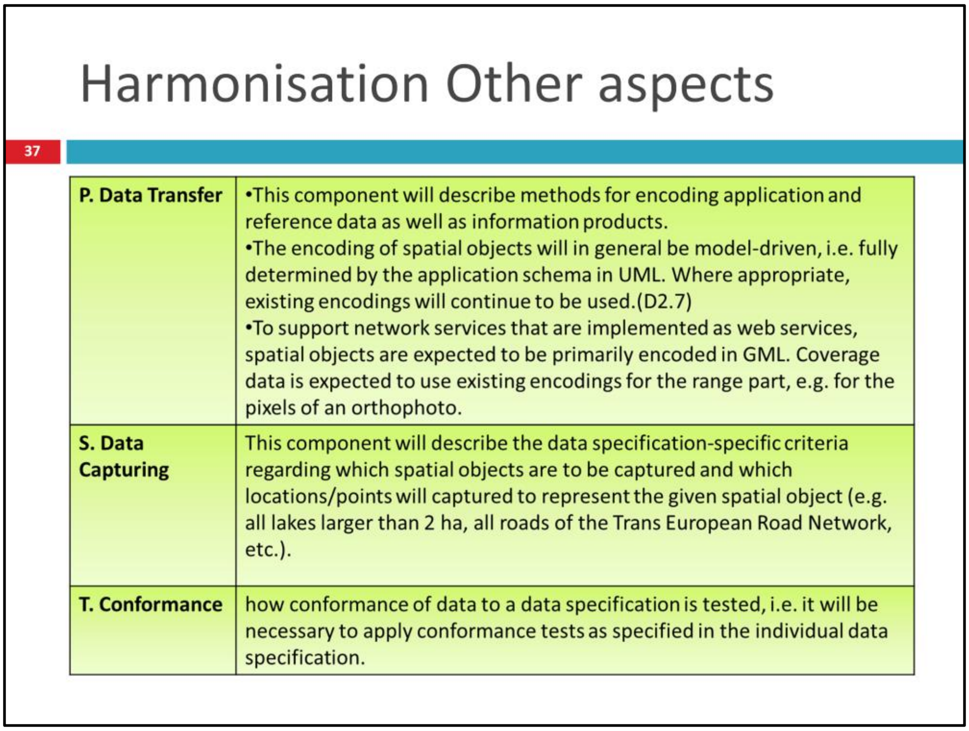
- One of these elements focusses on the maintenance of data products and spatial objects within those products. How to deal with updates? What are the best practises for versioning of objects (introducing a new object and deprecating the old one).
- the quality component will advise the need to publish quality levels of each spatial dataset using the criteria defined in the ISO 19100 series of standards, including completeness, consistency, currency and accuracy. This will include methods of best practice in publishing Acceptable Quality Levels…
- in this category also consistency between data is addressed and this in terms of format, logical and topological accordance, and so on.
- DQ concerns also multiple representations or the best practises of how data can be aggregated:
  - over time and space;
  - but also across different resolutions also called generalisation of data.

Maintenance: This component will define best practice in ensuring that application data can be managed against updates of reference information without interruption of services. This will require the definition of mechanisms by different stakeholder areas to manage where this is required and it is feasible. ***RSS feeds for change information?***

Quality: This component will advise the need to publish quality levels of each spatial dataset using the criteria defined in the ISO 19100 series of standards, including completeness, consistency, currency and accuracy. This will include methods of best practice in publishing AQL etc

Consistency: Format, logical, topological etc

- Multiple representations
- Derived reporting (example: typically water samples at 1 km intervals are reported to the European level)



All aspects on this slide do not fit in one of the previous categories however they cannot be disregarded when it concerns data harmonisation.

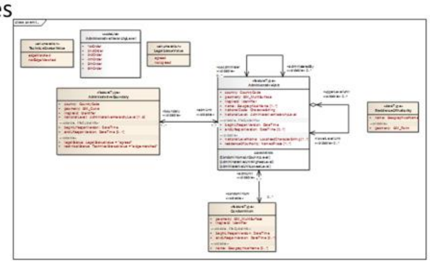
- The data transfer aspect focusses mainly on the encoding of data. As said before, within the INSPIRE framework, GML is seen as the standard for encoding data. But for coverage data this might not be the proper format to exchange data. Therefor the GCM will give some guidance in alternative encoding mechanisms.
- The data capturing aspect covers the DS-specific criteria regarding, which spatial objects are to be taken on board (in scope or out of scope) or which coordinates will represent certain spatial objects. Also a certain accuracy of data capture can be required.
- And last but not least the conformance aspect. For a dataset to be declared as INSPIRE conformant it needs to pass conformance tests as specified in the individual DS of the theme it belongs to.

This overview of harmonisation aspects is not to bore you but they all come back in the final DS documents throughout the different chapters. Some of these aspects are a repetition from one of the general framework documents (like GCM, guidelines for O&M) but others are described by the Data Specification document itself.

Results of modelling process

38

- Framework documents
  - Definition of Annex Themes and Scope
  - Generic Conceptual Model
  - Guidelines for the encoding of Spatial Data
  - Methodology for DS development
  - Common data models
  - O&M guidelines
- Data specification for all Annex Themes
  - Textual description of the data model
  - UML model
  - GML application schema
- Available here:
  http://inspire.jrc.ec.europa.eu



Let’s have a look now to the result of the modelling process.

First of all we have all the supporting frame work documents as listed here. These form the background information for the DSs.

But most important are of course the data specification documents for all of the annex themes and they consist of:

- A textual description of the data model;
- The model itself in UML format;
- GML application schema

All these documents are made available in a repository and can be consulted at the website of inspire.

It is quite a challenge to read through all of them but a least you know now how they were established and how they are organised.

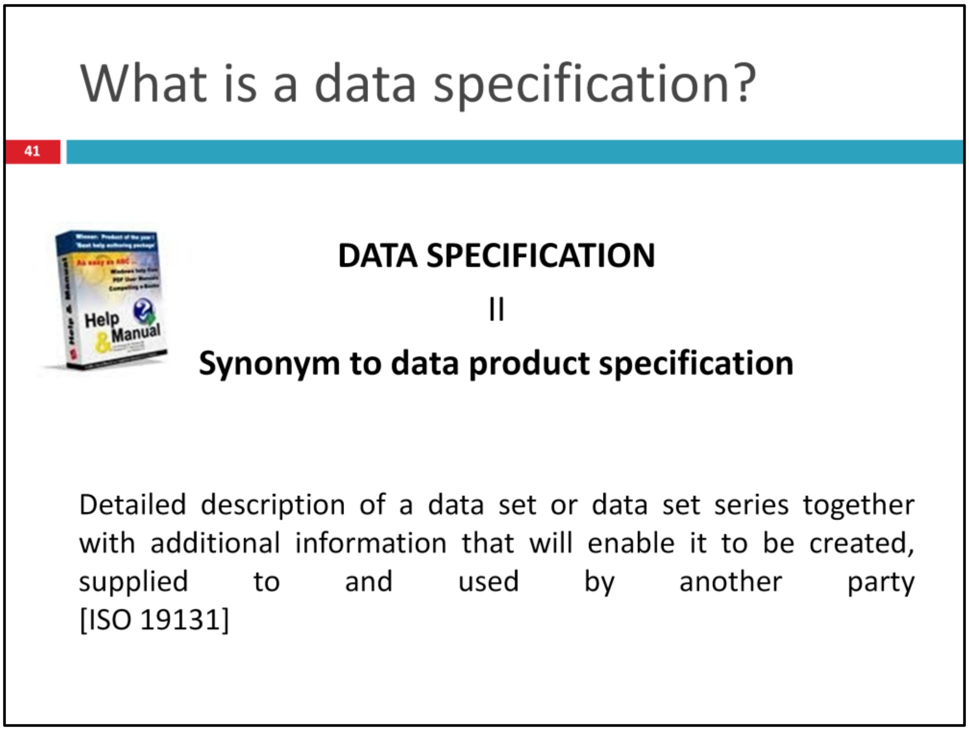
40

Part 3

Development of INSPIRE data specifications

Let’s have a quick look now to the development process of the DS in part 3 of this module.

The knowledge of the process is not crucial to understand the DS but it gives you some useful background on the different steps followed and the involvement of different people and communities.



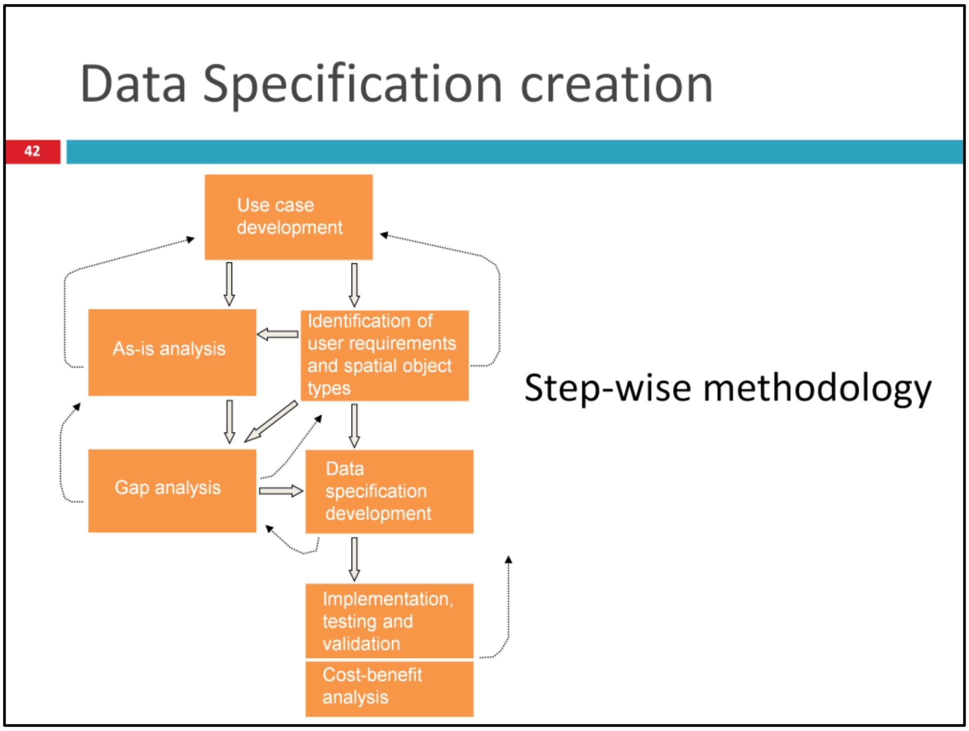
We already talked a lot about DS:

- What are their scope and objectives?
- What are the building blocks and the principles behind them?

But what is now a proper definition of a DS, also called a data product specification?

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

Or in human language and applied to the INSPIRE framework: It is the user manual for anyone that has to create, modify a dataset within the scope of INSPIRE.

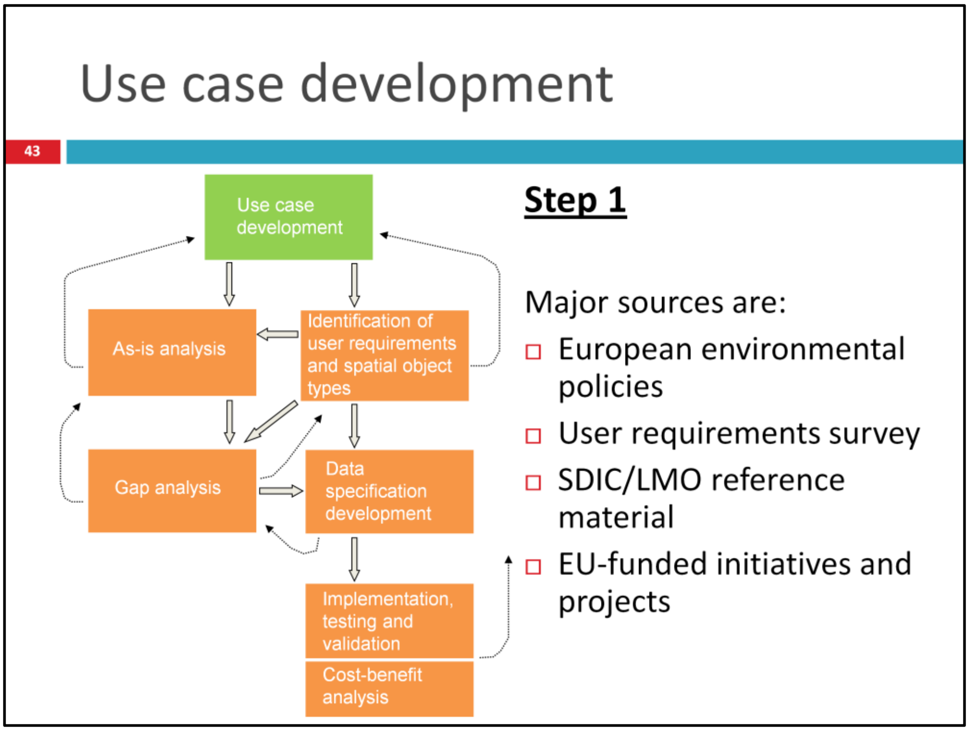


The creation process of such INSPIRE DS followed a step-wise methodology with moments of feedback and with the involvement of the relevant stakeholders. Here on this slide you can see the workflow of the process indicating as well the feedback possibilities (iterations) between the different steps.

It started with the development of relevant “use cases” from these use cases, user requirements were drafted and the spatial objects were identified. In parallel an as-is analysis was made of the existing datasets. Based on those requirements and the description of the real situation, gaps were identified and the results of all three processes were used to draft the first version of DS. Then followed a implementation, testing and validation phase after which a new round of feedback was possible and the comments were used to fine tune the DS.

To give you an idea of the duration of this process: For the annex II and III themes, that were developed in parallel with each other, it took approximately 2 years until the final DS were ready for adoption.

In the coming slides I will explain the separate steps in more detail.



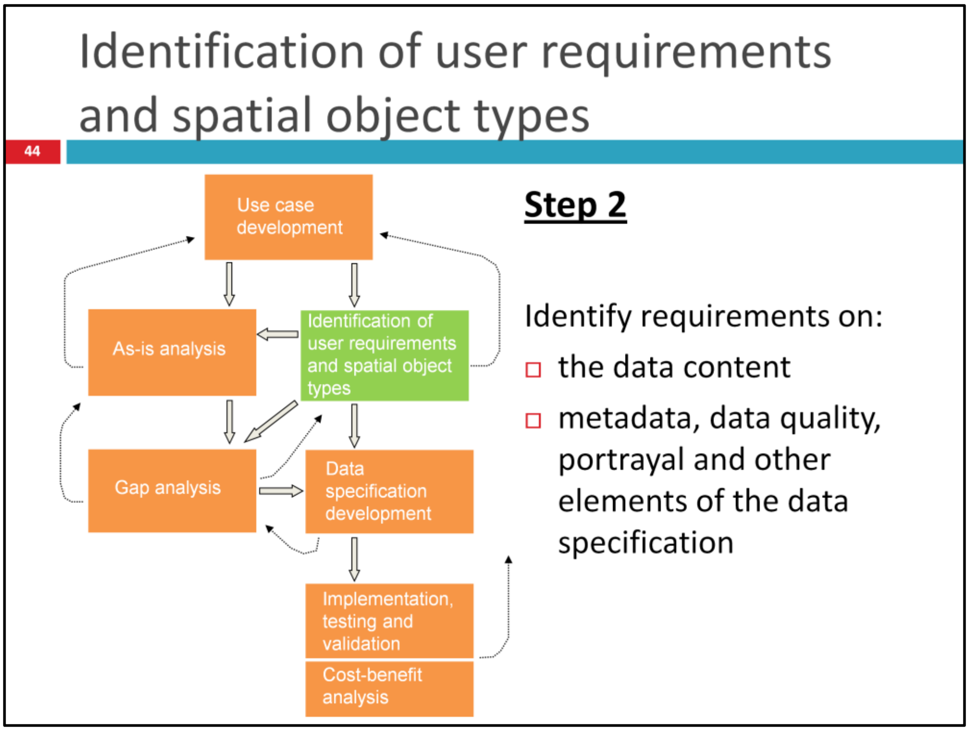
The first step in the development process was the defining of relevant use cases.

Many stakeholders could propose Use Cases that should be taken into account for the further development process. The major sources on which the UC were inspired, are listed here.

Relevant UC could be derived from or influenced by:

- existing European environmental policies, fe MS have already certain reporting obligations concerning environmental information;
- a survey on user requirements;
- reference material provided by Spatial Data Interest Communities (SDIC) and Legally Mandated Organisations (LMO). Both groups were already involved in other aspects of the INSPIRE framework. This material was based on real-life implementations of certain themes, which gave a good insight on the how and why datasets were modelled in a specific way.
- similar EU-funded initiatives and projects.

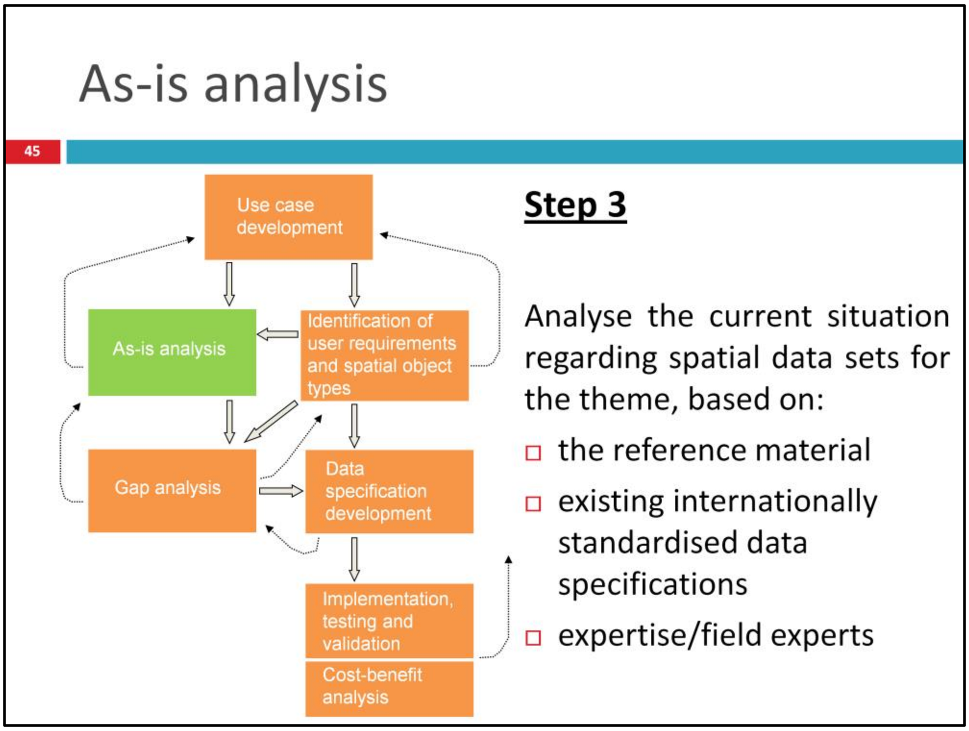
This all illustrates that the DS did not come out of the bleu, a lot of material came from existing situations and had to be considered as UC to serve as a base of the DS development.



The second step was, to identify the User Requirements and the essential spatial objects from all the material gathered in the first step.

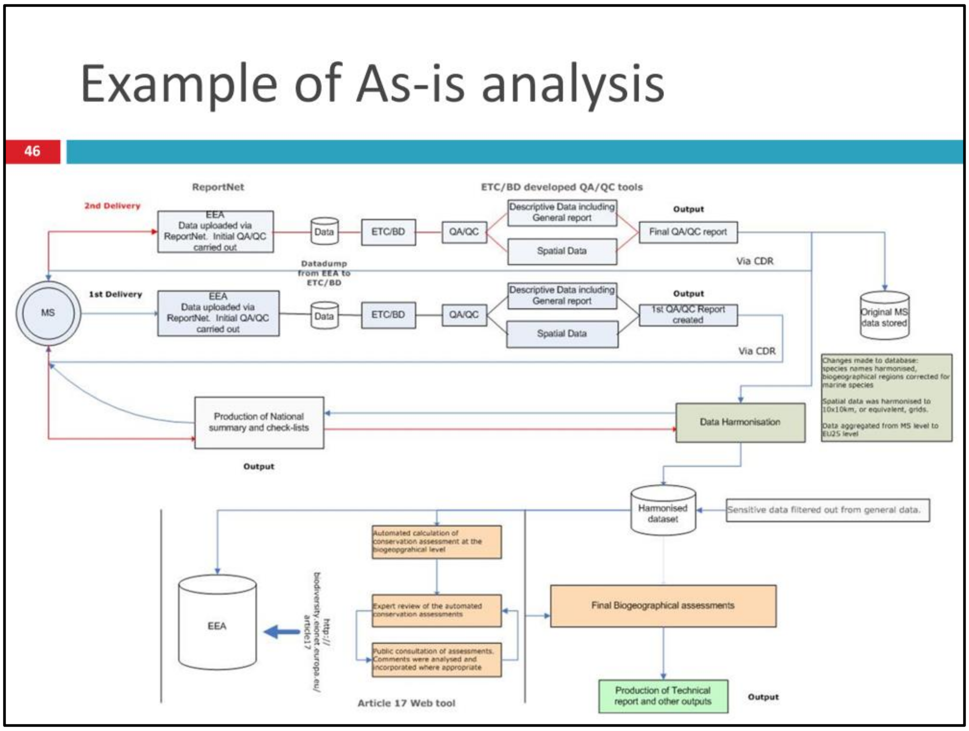
The requirements could regard:

- data content
- MD
- DQ
- Portrayal
- Other elements of the DS



In the third step is based on the reference material provided in the first step. It concerns an As-Is analysis of the current situation regarding current procedures and workflows of (theme specific) spatial data sets.

Apart from the reference material, there was also looked to which extent existing international standards were already in use and of course the knowledge of field experts could not be disregarded in this phase.

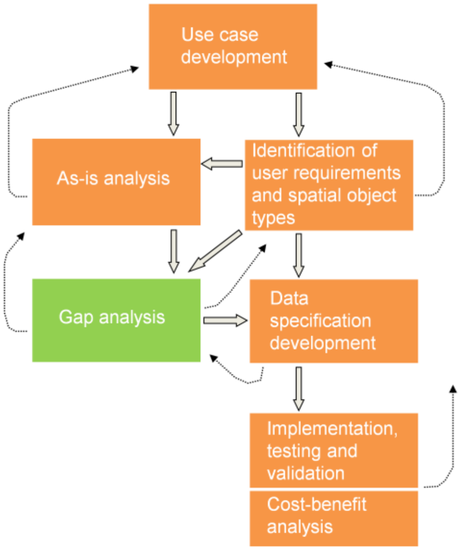


This slides gives an example of the workflow that MSs should follow, to report on Habitats according a certain article of the Habitat directive. The MSs need to report on and upload data to the EEA and ETC/BD, they perform a QC, based on the result a second delivery with corrections can be asked. The final result is then stored in a common repository from which a harmonised database on the state of habitats throughout Europe is created. This resulting database can then be used by the EEA to make new assessments according the Biogeographical Regions of Europe.

So this is an example of an established workflow which cannot be neglected in the development process of the DS.

Gap analysis

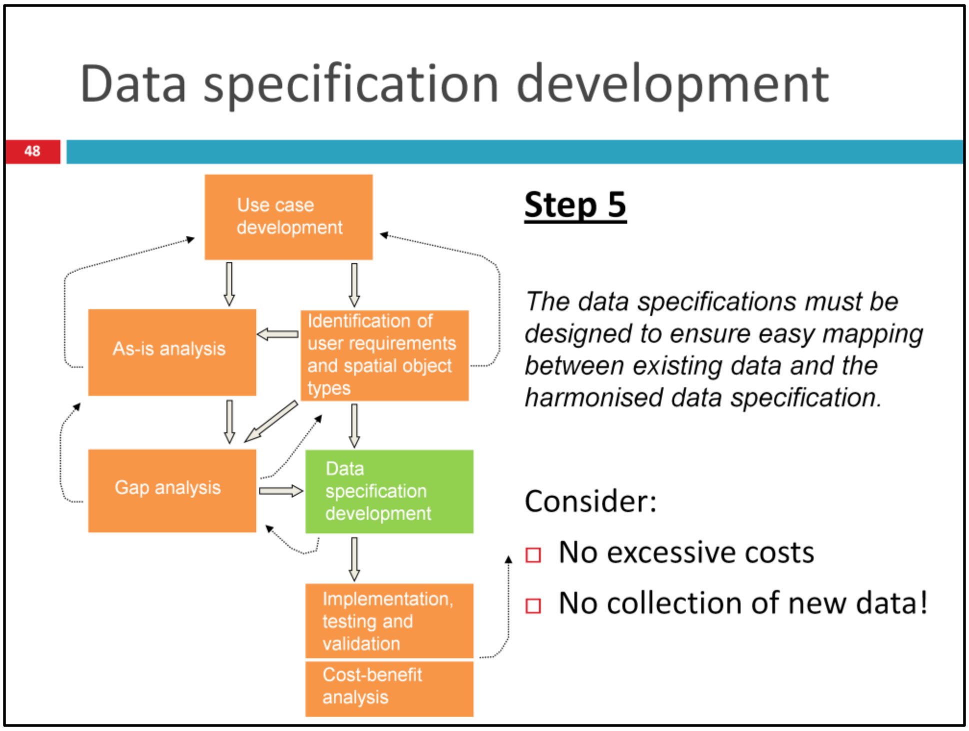
47



Step 4

Compare identified data sources with identified user requirements

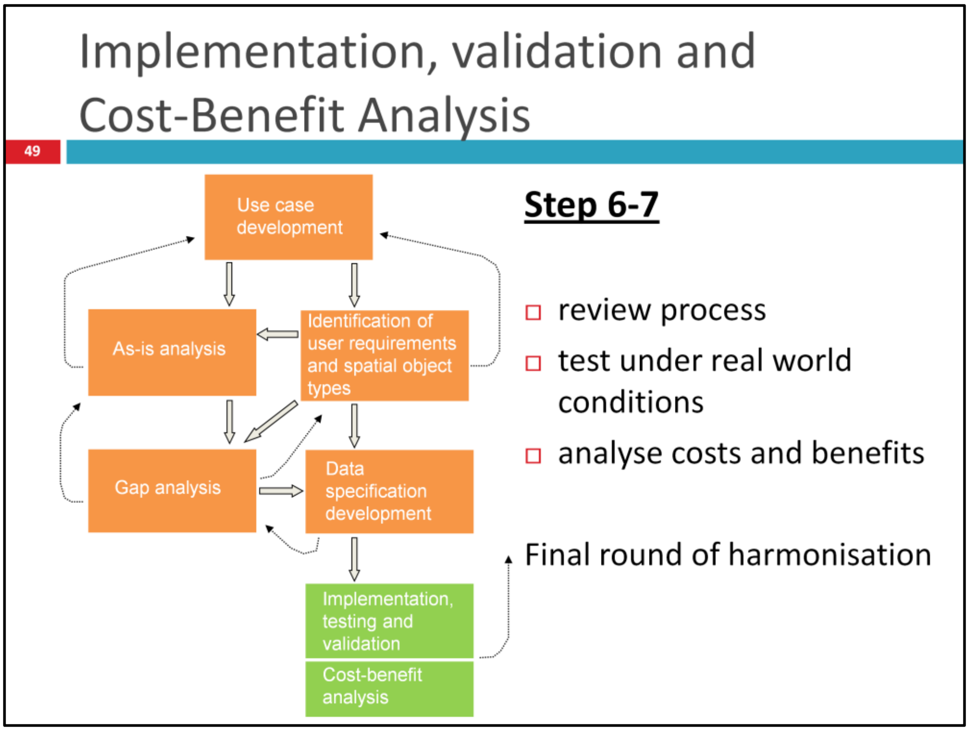
In the fourth step the user requirements are compared to what is already present in existing data sources. User requirements that are not met yet are identified as “Gaps”. This is called the gap analysis phase. DS must be designed to close these gaps.



All previous steps belong to the preparation phase for the real development of the DS. The development of the DS was done by a group of domain experts (TWG). That group had to design the DS in such way that the user requirements were met and that the mapping between the existing data and the final harmonised model was quite straight forward.

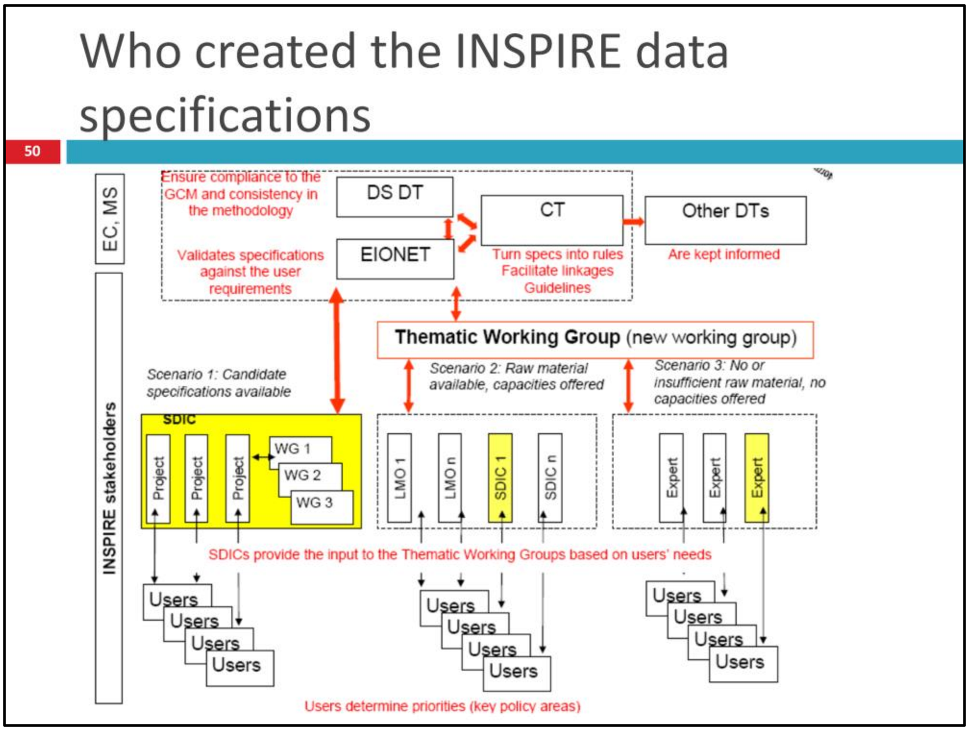
The group had to consider two major aspects that are related to each other :

1. The data harmonisation process should not lead to excessive costs for the data providers because then they will not do the effort to transform their data;
2. the data specification should not require collection of new data



The resulting draft versions of the DS were then implemented and send back to the data communities for testing and validation in real world conditions.
Their remarks were evaluated and processed in a new iteration round in order to come up with the final DS.

(Cost-benefit analysis is difficult to make in a quantitative way. Economic value is difficult to estimate, but several studies exist and already many papers on this topic are published in the IJSDIR.)



To get an idea of the people and communities that were involved in the DS creation process, I show you this slide. The actual work of designing the DS was done be a group of domain experts per theme, called the TWG. Data users could provide input through the SDICs and LMOs.

And the whole process was controlled by the DS DT to ensure compliance to the GCM and consistency in the methodology.

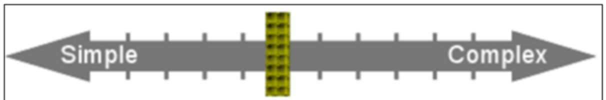
The validation of the DS against the user requirements was done by the partnership network EIONET of the EEA (involving experts and national institutions).

And finally the DS were translated by the INSPIRE Coordination Team (CT) into IR and technical guidelines while other DTs of INSPIRE were kept informed

So we can conclude that the development of the DS was a complex process to which a lot of people contributed.

Which level of harmonisation is “just right”?

51



Too simple:
- identified requirements cannot be supported
- insufficient harmonisation
- few benefits

Too complex:
- difficult to implement
- substantial benefits available only to few users
- high cost

Requires:
- an iterative process
- well-established requirements
- good understanding of the existing geographic information
- testing and validation

When developing DS it is important to find the appropriate level of harmonisation.

If the model is too simple, there will be no added value and nobody will use it because it cannot fulfil the user requirements.

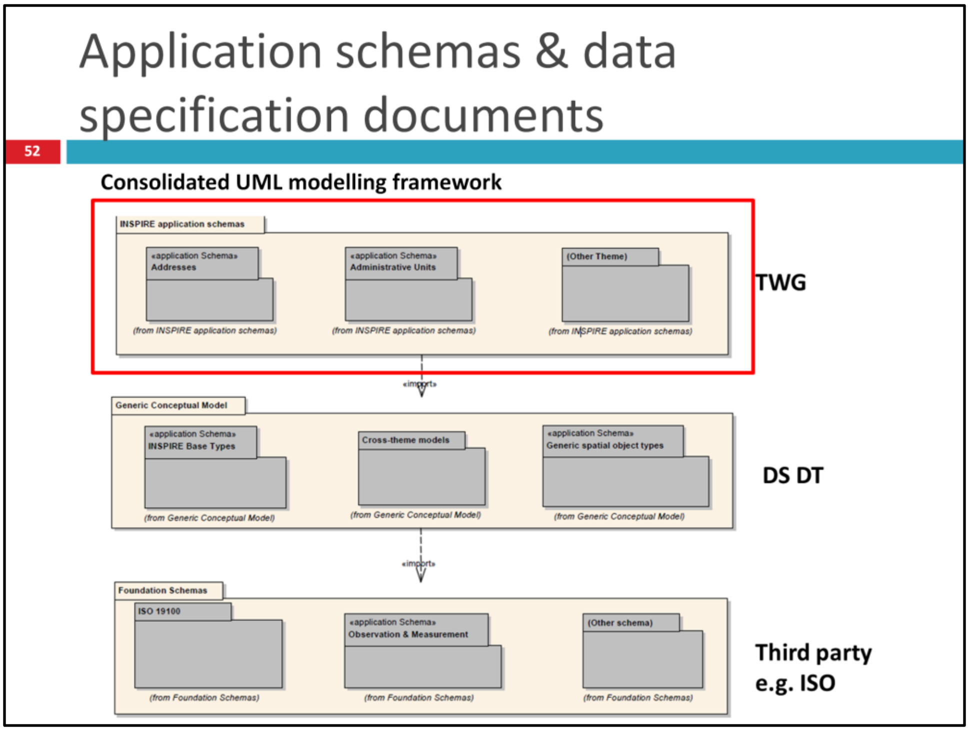
If the model is too complex on the other hand it might be difficult to implement. A number of benefits will only be available for a limited number of users. And the cost of harmonisation will be too high.

The compromise is somewhere in the middle and to find this equilibrium it required:

- an iterative process with testing, validation and adaption phases;
- user requirements that were well-defined;

AND

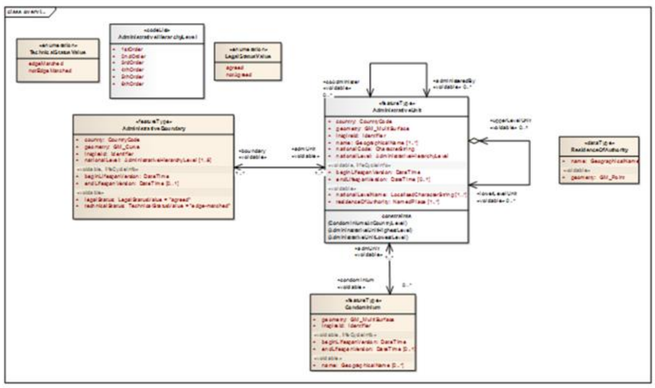
- a good understanding of the existing geographic information



If we go back to the representation of the layered modelling framework, we have arrived to the upper layer where the DS are created for each of the themes by the different TWG which also created the UML application schemas.

Result

- Data specification for all Annex Themes
  - Textual description of the data model
  - UML model
  - GML application schema



To resume and also to make the link to the next part of the module ‘understanding the technical guidelines’, I give you again the result of the DS process.

For each annex theme of inspire a DS document is created with a textual description of the model, the model as a UML diagram and the final GML application schema which will be used in the transformation process of data sources into the Data model of the DS.

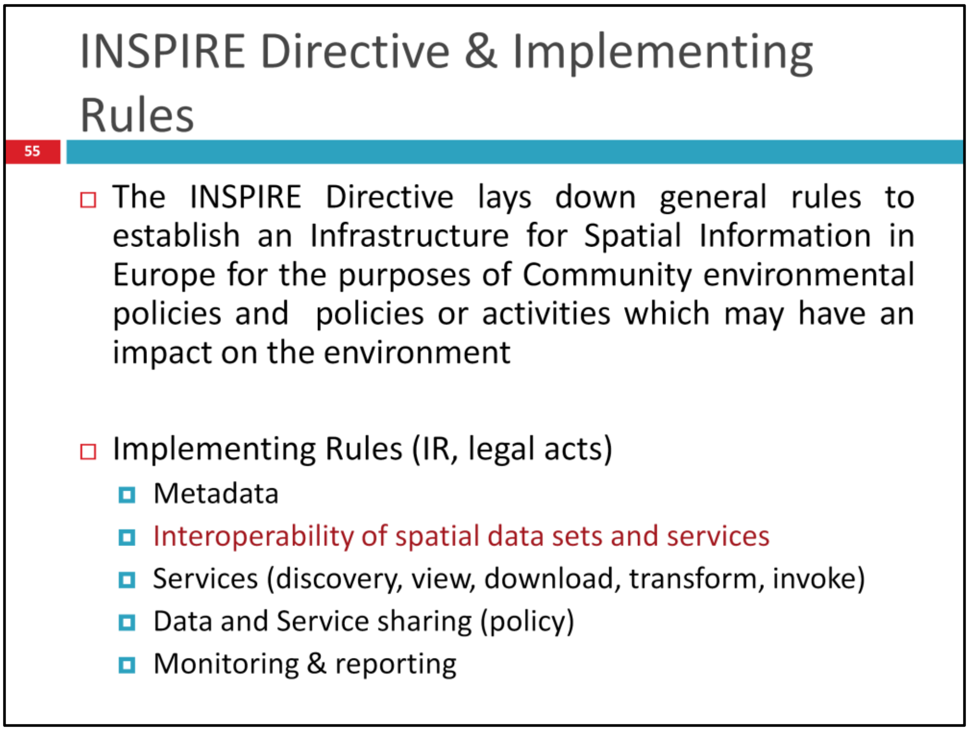
These documents will be the key components for data providers to deal with the transformation process. The other framework documents that we talked about can be considered as essential background material.

54

Part 4

Understanding the Technical Guidelines: INSPIRE Data specifications

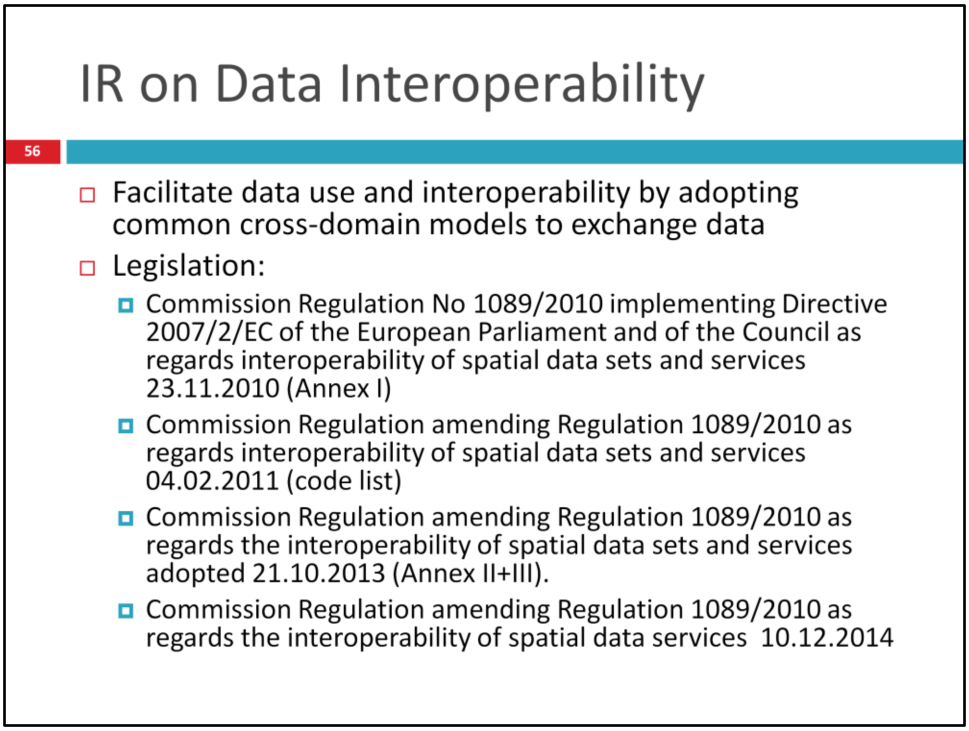
We know now what the scope and the objectives are for the DS. We have talked about the basis for these DS in the part on the modelling framework and we have shed a light on the development process of the DS. After this long introduction it is about time to have a closer look at the DS document itself in part 4 of this module. The objective of this part is to give an understanding of these DS, how they should be read by looking at the different chapters and highlighting the most important parts. We will talk about implementing rules (IR), and see how they differ from technical guidelines (TG). We will identify the parts in the documents that are common to all DS, it saves us time in reading through different DS. After this part you will be able to read through a DS document and interpret it in the correct way.



Before we go to the content of the DS, it is necessary to show you the position of the DS documents inside the whole INSPIRE story.
At first there was the INSPIRE directive which lays down “general rules” to establish an Infrastructure for Spatial Information in Europe related to environmental policies. These rules are translated in legal acts called the IR. There are common IR adopted in a number of specific areas:

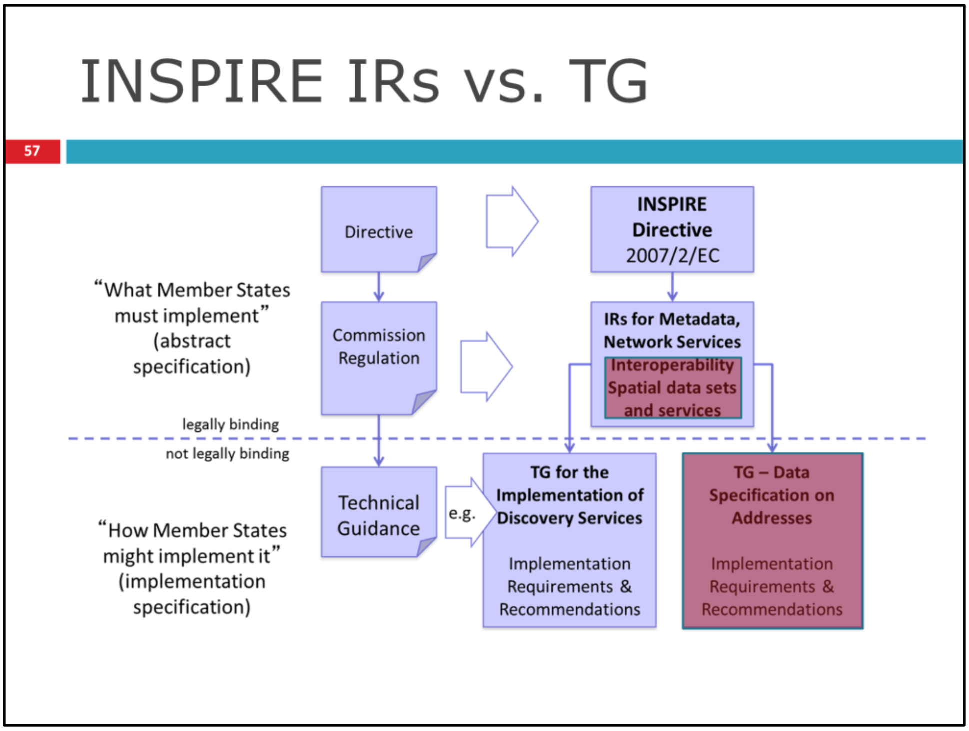
- Metadata: defining the mandatory elements to describe data sets and services
- Interoperability of spatial data sets and services: concerning adoption of cross domain data models for the exchange of data
- Network Services (discovery, view, download, transform, invoke)
- Data and Service sharing (policy): regulating the provision of access;
- Monitoring & reporting: continuous monitoring of the implementation of the Directive and regular reporting covering 4 main fields metadata, spatial data sets and services, network services, data sharing

These IRs are adopted as Commission Decisions or Regulations , and are binding in their entirety



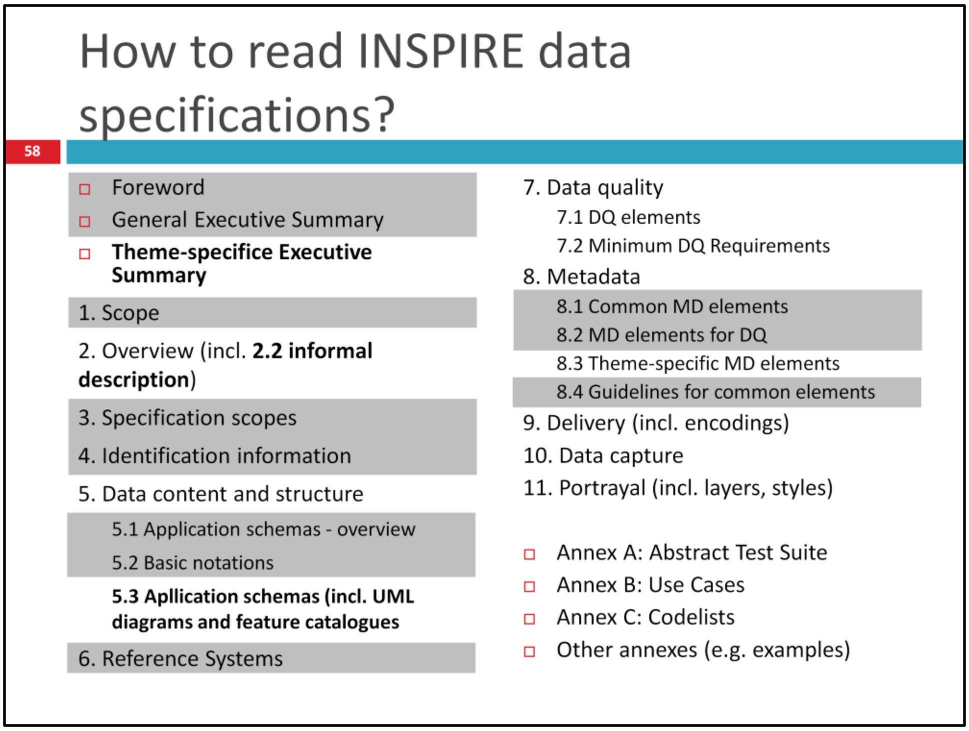
The IR's on data interoperability are designed to facilitate the data use by making the exchange of data easier through the use of harmonised data models. The legislation came into place in different stages resulting in a commission regulation with amendments:

- It started with IR concerning Annex I themes in November 2010;
- Amended with a regulation concerning code lists in February of 2011;
- In October 2013, the rules for Annex II and III were amended;
- And the last amendment was done in December 2014 giving IR for spatial data services apart from the standard network services (discovery, view and download).



How do the DS documents relate to these IR or commission regulations?

This slide shows that there must be made a distinction between the "legally binding" part, i.e. "What MS must implement" represented by the Directive and the commission regulations (IR or legal acts). And on the other side the "not legally binding" part, i.e. "How MS might implement it" represented by the Technical Guideline documents to which also the DS documents belong.
The Technical guidelines may contain specific technical requirements to meet the requirements of the IR. They also contain recommendations meant to facilitate the implementation but these are not mandatory.



Now we know that the DS documents will help you to implement the IR, let's have a look at their structure. All DS documents are based on a common template which means that they all have the same structure which is shown on this slide.

It starts with Introduction containing a summary with a general part and a theme specific part.

The first chapter "the scope" is only mentioning which theme is considered in this DS.

The second chapter contains an informal description of the theme to explain in human language the scope of the theme. Chapter 3 and 4 are very short and deal with the general scope and the identification information of the specific document.

The most important chapter is chapter 5 "data content and structure" which is actually the core of the DS document together with the chapters on data quality, metadata and encoding. It contains the Application schemas (data models) including UML diagrams and feature catalogue(s).

Chapter 6 concerns the rules for the use of different reference systems (spatial – temporal), units of measure and reference grids.

Chapter 7 and 8 are quite related and deal with Data Quality and Metadata, the general elements but also the theme specific elements. The relation between chapters can be explained by the fact that the data quality of a theme needs to be described by specific elements in the MD.

Chapter 9 gives information on the encodings to be used for data delivery. The default proposed encoding is GML. While for coverage data alternative encoding can be used.

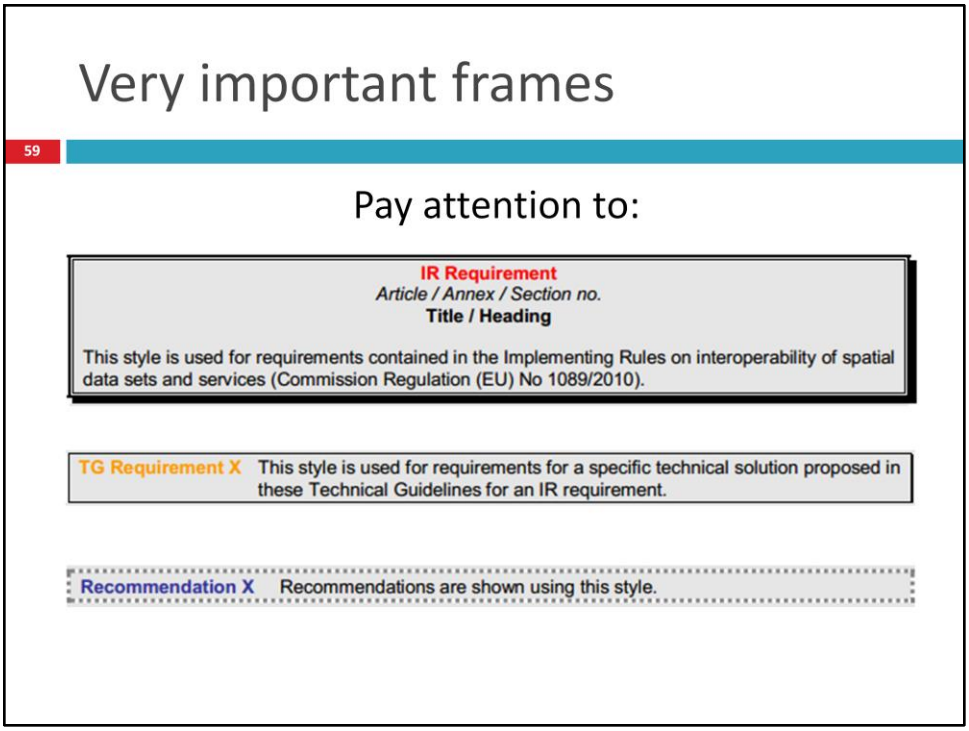
The chapter on Data Capture is meant to specify requirements or recommendations for the process of data capture for new datasets. Often this chapter is not elaborated.

Chapter 11 defines the rules for layers and styles to be used to display spatial object types in view services for the concerned theme.

Some parts in a DS document, as you also can see on the slide, are indicated with a shading. It concerns the parts that are common to all DS documents, so you only need to read them once because they are the same in all DS documents. It concerns generic parts that apply to all themes it is background information coming from conceptual framework documents like the GCM or some basic notions on INSPIRE, but also IR requirements coming from the IR on spatial interoperability, etc…

The DS document has also a series of annexes:

- Annex A: describes an abstract test suite which should guide you through the conformity testing of a transformed dataset against the DS.
- Annex B: gives you background information on the use cases that were considered in the creation process of the DS
- Annex C: contains the values of the theme specific code lists used in the models
- And other optional annexes f.e. on alternative encodings for coverages



In all DS you will find the same important types of styles using frames. They indicate the types of requirements and recommendations you will have to follow when implementing the DS:

- First of all there are the IR requirements indicated in red, these a rules that are legally binding, they must be respected to be in line with INSPIRE and they are coming from the IR documents. For each of IR requirements, the Technical Guidelines contain additional explanations and examples. Also called “**Legal Requirements”**
- Then you have in orange/yellow the TG requirements. TG may propose a specific technical implementation for satisfying an IR requirement. Therefore TG may contain additional technical requirements that need to be met in order to be conformant with the corresponding IR requirement when using this proposed technical implementation. In other words if you choose to be in line with the DS (which is not mandatory) these requirements should be met. Also called “**Implementation Requirements”**
- And a last you have the TG recommendations (in blue with dashed frame) for facilitating implementation or for further and coherent development of an interoperable infrastructure. The implementation of recommendations is not mandatory. Compliance with the Technical Guidelines or the legal obligation does not depend on the fulfilment of the recommendations. Also called “**Implementation Recommendations”**

It is important to understand the difference between these terms in order to know what is expected for the implementation.

CH5: data content and structure

60

Core chapter of each data specification document

- 5.2 Basic notions repeats relevant parts of the Generic Conceptual Model
  - Notation (UML)
  - Used Stereotypes
  - Voidable properties
  - Code Lists and Enumerations
  - Identifier management
  - Geometry representation
  - Temporality representation
  - Coverages

→ knowledge needed on Standards & GCM

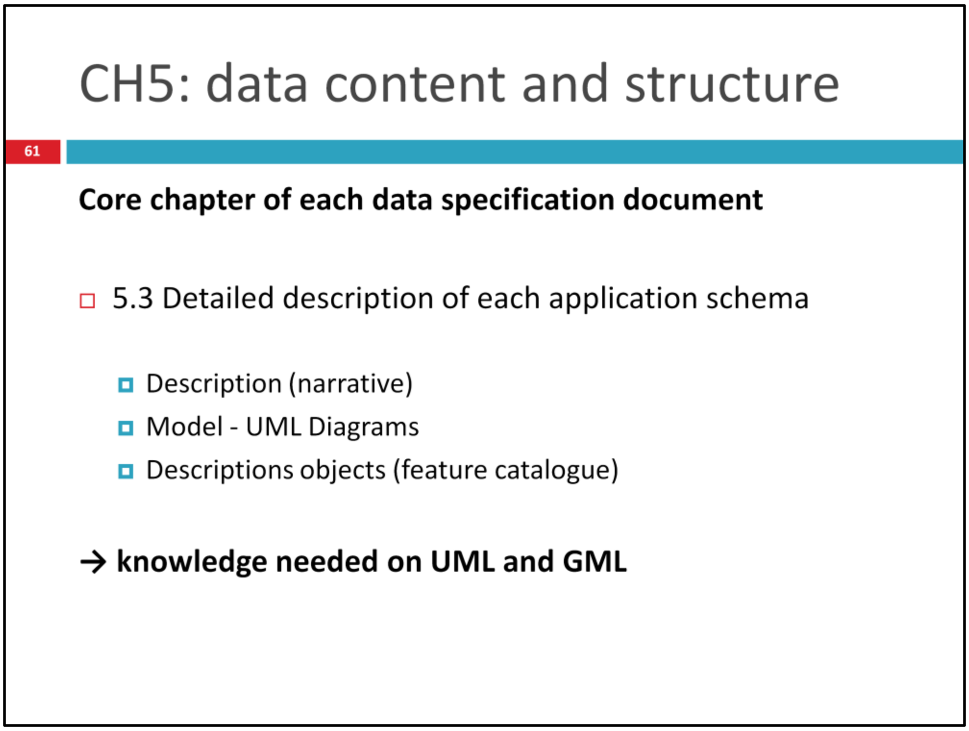
I gave you already an overview of the DS content in one of the previous slides. Let's no go in to more detail for the most important chapters of a DS document: 5, 7 and 8, Dealing respectively with "content and structure", "Data Quality" and "Metadata".

Chapter 5 is actually the heart of each DS. At first it mentions that the theme specific applications schemas defined by the TWG are included in the IR which means that the "Types for the Exchange and Classification of Spatial Objects" concerning that theme are the ones described in the specific application schemes dealing with that theme.

In a second subchapter some **general** basic notions are explained which are valid for each theme, it concerns relevant parts taken from the GCM, like:

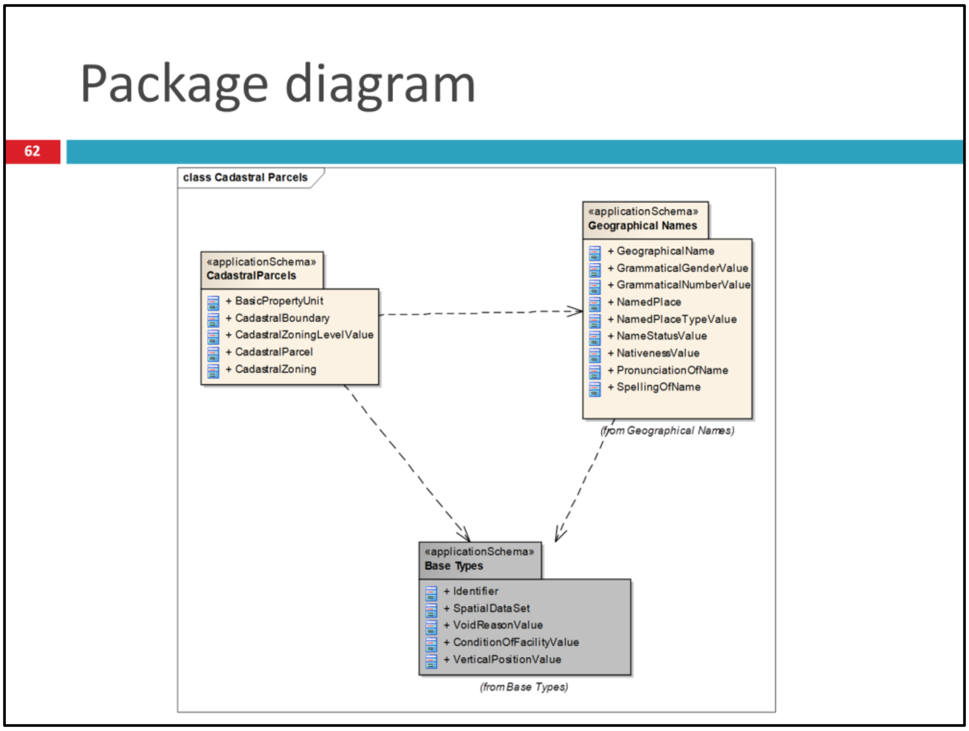
- UML Notation of the Application schemas as a common conceptual schema language;
- List of Stereotypes that can be used in the application schemas;
- Explanation of the concept of "voidable properties";
- Difference between enumerations and code lists;
- Concept of Identifier management;
- Geometry representation: Which geometry types are allowed?;
- Temporality representation: the concept of life-cycle of spatial objects;
- Information on how to deal with "coverages".

It gives the essential knowledge you need in order to understand the following paragraphs of the DS. More information on these topics can be found in the referenced standards and in the GCM.

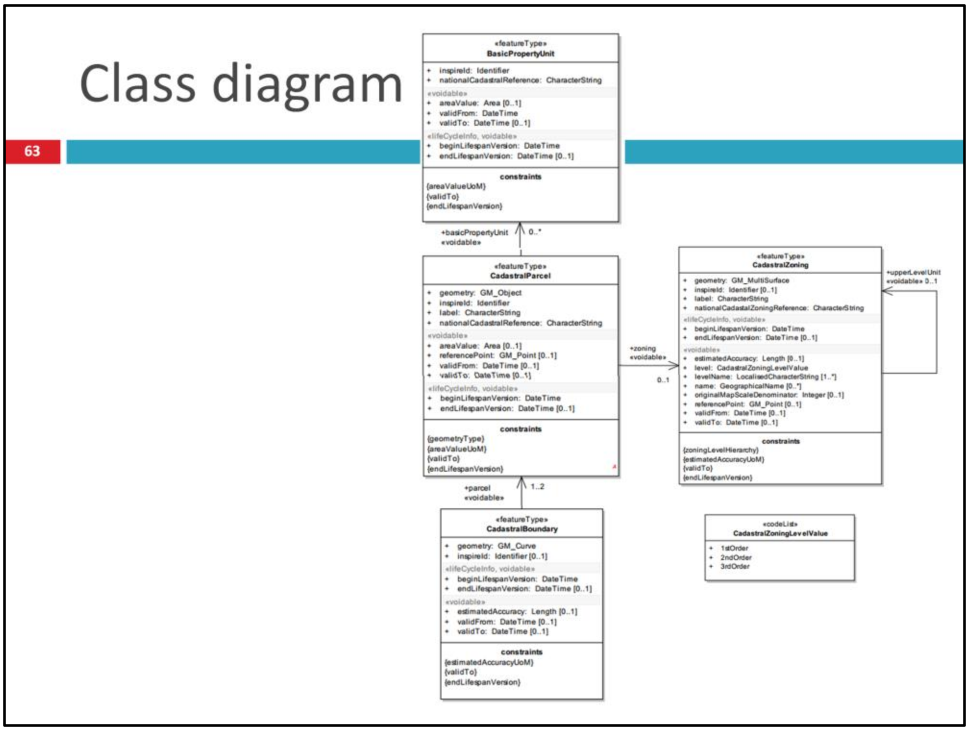


In Subchapter 5.3 you can find then a detailed description of each application schema. It contains a narrative description of the model next to the UML diagrams that demonstrate the different types/objects and their relationships. To read these diagrams, knowledge of UML (and GML) is necessary. The models themselves are then accompanied by the full description of all objects, attributes, the relationships and the used classifications also called the feature catalogue.

This is the “Core” of every DS document, people with experience in reading the DS can concentrate on this part of the document in order to understand what has been modelled and how it is structured. It provides the information needed to perform a mapping from existing resources to the related DS.



Through package diagrams relations between application schemas are displayed. In the example here on the slide you can see the AS of Cadastral parcels that makes use of the application schema of Geographical Names and the schema of Base Types defined in the GCM. These last two APs are probable most referred to by other APs because they contain essential types like Geographical name, NamedPlace and Identifier, VoidReasonValues and so on.

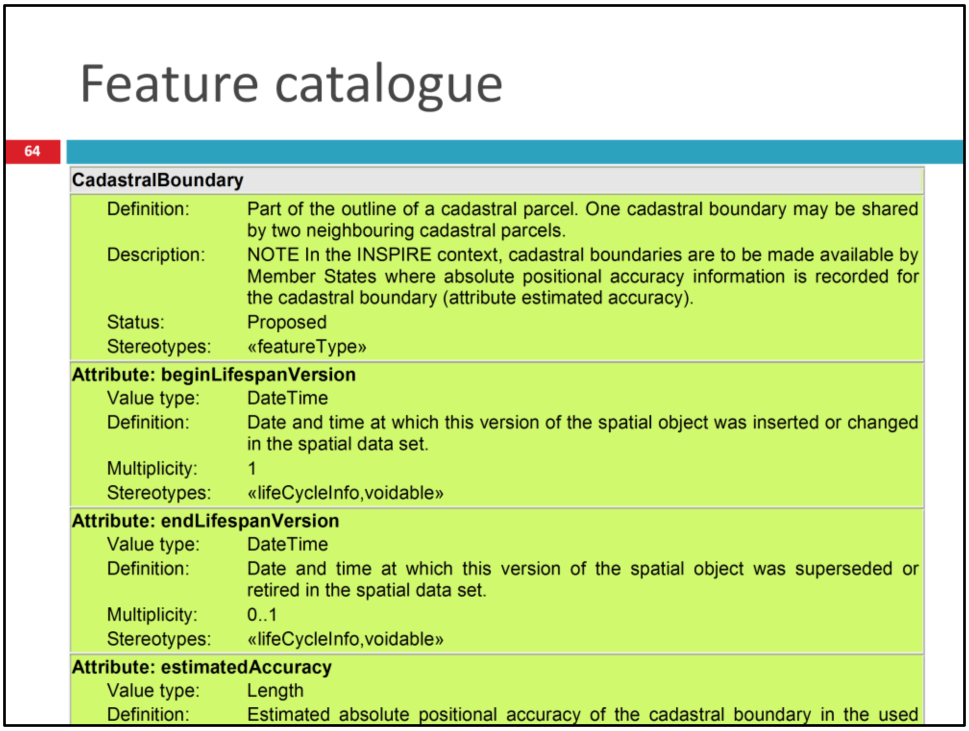


While the package diagrams show the dependencies between AP, the class diagrams depicts the different classes with their attributes and their relationships within 1 AP. Here in the example you see the AP of Cadastral parcels. Each box in the schema is a class represented by a class name and typed by a stereotype (above the name between brackets). Stereotypes can be defined on packages, ,on classes, on attributes, on associations/dependencies to attribute them specific properties. F.e. a stereotype “featureType” defined on top of a class makes this class special it will contain Spatial Objects meaning that each object inside that class will have a geometry an most likely also a InspireID.

Each class has some attributes attached, which are listed inside the box. Some are mandatory, others are not. Some are voidable others are not.

Constraints (see lowest compartment) can be defined on classes f.e. to restrict certain values of an attribute under certain circumstances. Classes can be related to each other which is displayed by the connection arrows in between them.

For full understanding of the diagrams a good knowledge of UML is required and also an understanding of the basic notions like described before.



Feature catalogue

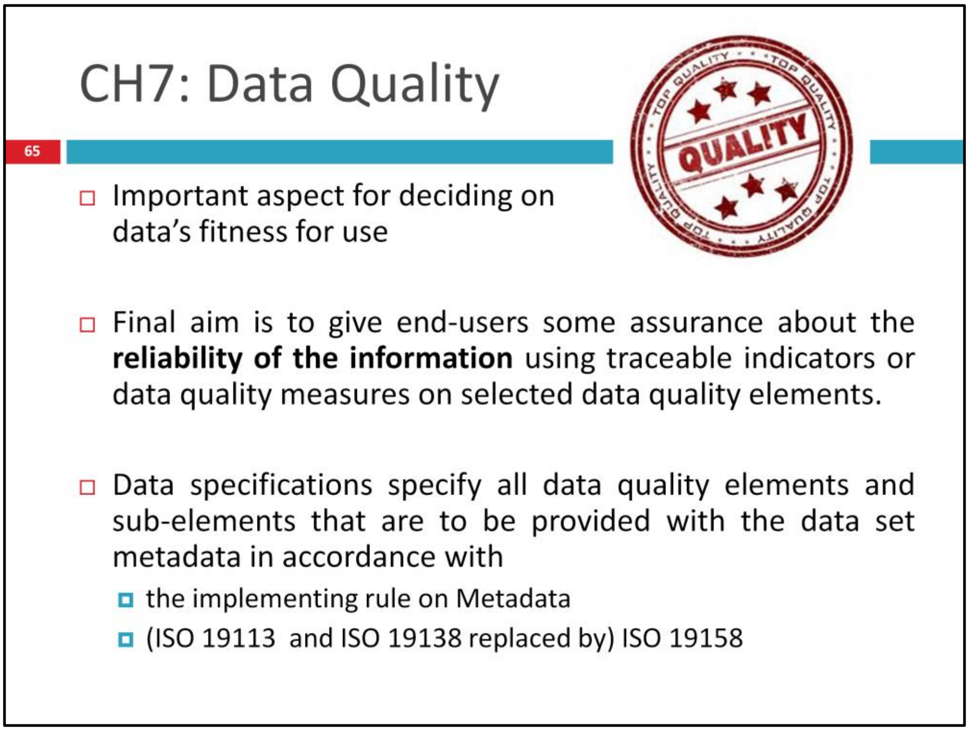
64

CadastralBoundary	
Definition:	Part of the outline of a cadastral parcel. One cadastral boundary may be shared by two neighbouring cadastral parcels.
Description:	NOTE In the INSPIRE context, cadastral boundaries are to be made available by Member States where absolute positional accuracy information is recorded for the cadastral boundary (attribute estimated accuracy).
Status:	Proposed
Stereotypes:	«featureType»
Attribute: beginLifespanVersion	
Value type:	DateTime
Definition:	Date and time at which this version of the spatial object was inserted or changed in the spatial data set.
Multiplicity:	1
Stereotypes:	«lifeCycleInfo,voidable»
Attribute: endLifespanVersion	
Value type:	DateTime
Definition:	Date and time at which this version of the spatial object was superseded or retired in the spatial data set.
Multiplicity:	0..1
Stereotypes:	«lifeCycleInfo,voidable»
Attribute: estimatedAccuracy	
Value type:	Length
Definition:	Estimated absolute positional accuracy of the cadastral boundary in the used

The class diagrams give an idea of the way the different objects with their properties are modelled but for a definition and a full description of the objects including their properties you should look to the feature catalogue which is part of the subchapter 5.3.

This slide gives you an extract from the feature catalogue of the CP application schema. Here you‘ll find the definition of the feature class “CadastralBoundary”, more information is given in the description, the stereotype is listed and you can see for all attributes information on their data type, their definition, description, multiplicity and the stereotype(s). Also the relationships and the constraints are explained in this Feature Catalogue

So the feature catalogue is essential to understand what is meant with all the modelled classes, their attributes and their relationships.



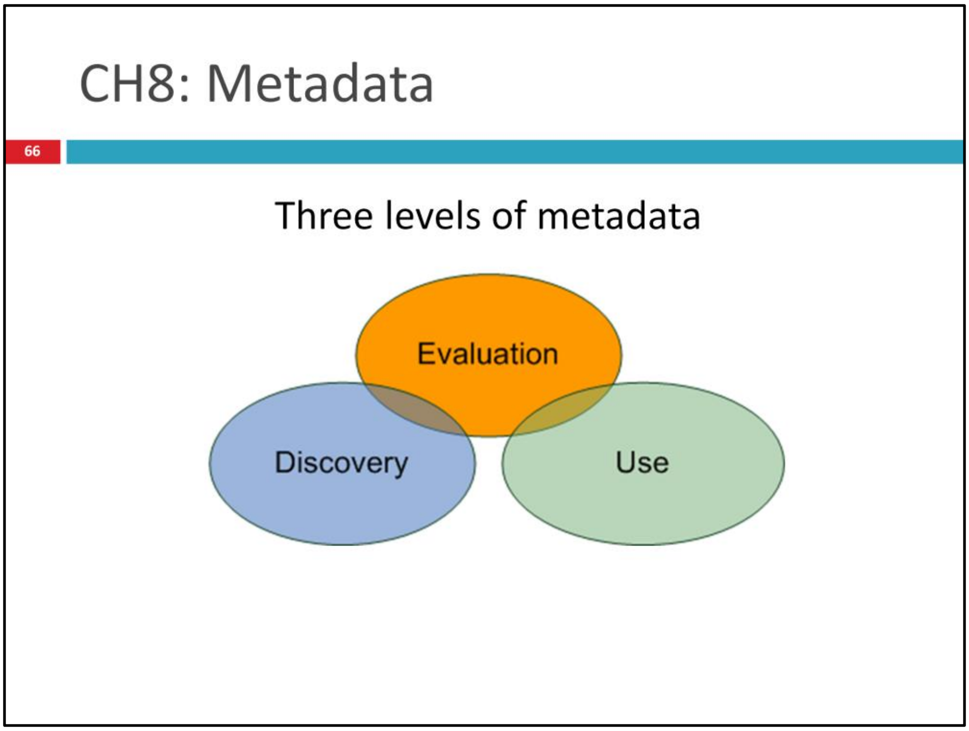
That was chapter 5 about the "**data content and the structure"**

Now we go over to another important chapter, chapter 7 on **"data quality"**. In this chapter some important aspects for deciding whether the data is fit for use can be collected.

The objective here is to give end-users some assurance about the reliability of the data. It can be that a data provider cannot deliver a good quality dataset but then it is really important to report this low quality so the user knows what he should expect and make up for himself is the data is fit for his use.

This chapter defines some standard indicators and data quality measures to be used to report the quality of a theme specific dataset. There are general data quality elements that apply to all themes but in some cases theme specific DQ elements were defined by the TWG.

For all the defined DQ elements there must be a corresponding MD element for documenting the quality, in accordance to the IR on Metadata. But in some cases additional MD elements should be taken from the ISO 19158 standard (which is replacing the former standards 19113 and 19138) and which is specifically designed for documenting DQ.



Chapter 8 specifies the dataset-level metadata elements, which should be used for documenting metadata for the complete dataset.

In general we can distinguish three levels in MD:

1. Discovery MD also called primary MD: typical elements used to find, search for data sets
2. The second level of MD is the evaluation MD:
3. The third level of MD are the MD concerning the use of data

In INSPIRE we speak of Discovery MD, MD for interoperability (corresponds to the MD for evaluation an use) and theme-specific MD elements. This differentiation corresponds with the INSPIRE documents in which the elements are defined/specified.

Metadata for discovery

67

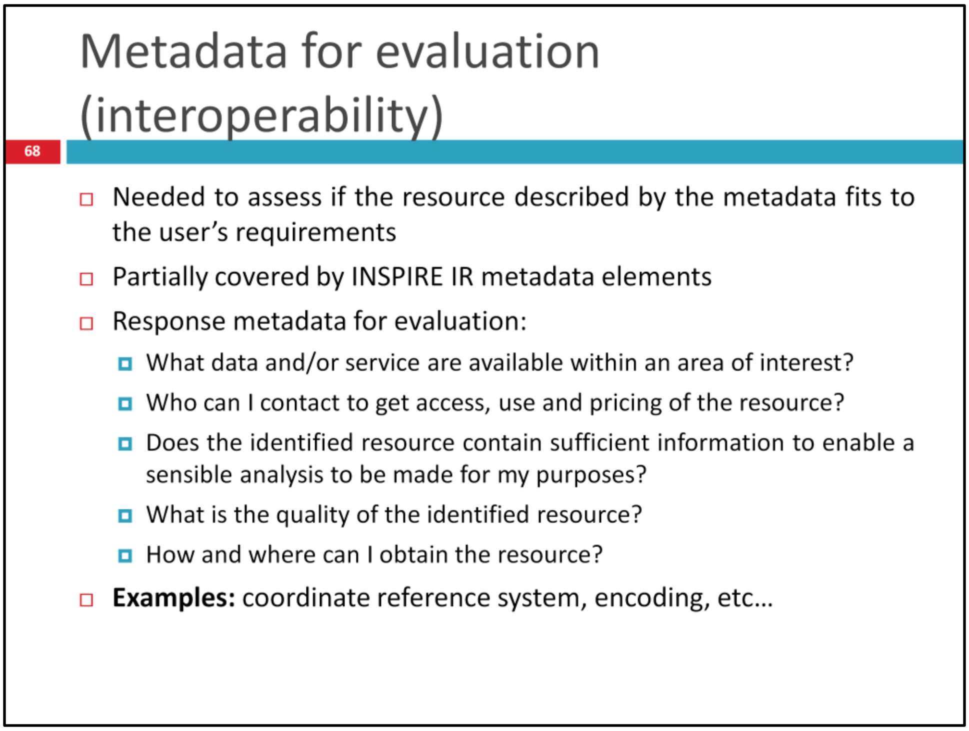
- Satisfy the need to search for resources
- Answers "what, why, when, who, where and how" questions about spatial resources.
- Plays a major role in the context on catalogues services where users formulate queries to search for metadata records
- These elements are often called queryables or search attributes
- **Examples:** resource title, resource abstract, lineage

Metadata elements	Multiplicity
Resource title	1
Resource abstract	1
Resource type	1
Resource locator	0..*
Unique resource identifier	1..*
Resource language	0..*
Topic category	1..*
Keyword	1..*
Geographic bounding box	1..*
Temporal reference	1..*
Lineage	1
Spatial resolution	0..*
Conformity	1..*
Conditions for access and use	1..*
Limitations on public access	1..*
Responsible organisation	1..*
Metadata point of contact	1..*
Metadata date	1
Metadata language	1

This slides list all elements mentioned in the IR for MD (i.e. **Commission Regulation (EC) No. 1205/2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata**)

These elements are also called Discovery MD by INPSIRE and they apply to datasets of all themes.

- These are typical elements that are used for searching and finding resources in MD catalogues through Geoportals or other MD search engines.
- They give answers to questions about spatial data, like: “what, why, when, who, where and how”;
- Therefore they are also called “queryables”
- Some examples are the Name or resource title, topic, abstract, lineage, keywords, …



The Metadata for evaluation is partially covered by elements listed in the INSPIRE IR for MD (see previous slide) but other required elements are listed in the IR for interoperability of spatial data sets and services (more specific in Art 13) (i.e. **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**).

It concerns the elements:

- Coordinate Reference System
- Temporal Reference System
- Encoding
- Topological Consistency
- Character Encoding

Basically these MD should provide the user with the information on access, pricing, fitness for certain purpose, data quality, on encoding (format) , coordinate reference used, …

Metadata for use: Theme specific

69

- Elements of this category come into play after a user has chosen a specific resource
- Specific metadata elements will support the user in fully understanding the data and using it properly resulting in more reliable analysis and more confidence in the results
- Elements of this category are out of scope of the INSPIRE Implementing Rules for Metadata. Instead, they are covered by the INSPIRE Data Specifications of the Annex themes I,II and II accordingly

Apart from the required MD elements for interoperability there are also recommended elements regarding interoperability and these are specified in the DS document in chapter 8. These elements are also called MD for use

Some of the recommended elements are common for all Spatial Data Themes from Annexes I, II and III, it concerns:

- Maintenance information
- DQ – logical consistency – conceptual consistency
- DQ – logical consistency – domain consistency

While each time can define additional recommended elements specific for that theme. Here also the specific data quality elements defined in chapter 7 will be listed.

These elements allow the user to fully understand the data and how he should use them, knowing the limitations concerning the quality.

So the different levels of MD might be confusing but you should keep in mind that MD elements are managed by the **IR for MD** and the **IR for interoperability**. And that the MD elements really applicable for using a certain dataset of a certain theme are listed in the INSPIRE DS document (chapter 8).

Annexes of Data specification document

70

- ATS – Abstract Test Suite
- Use cases
- Code list values
- (Encoding rules for TIFF and JPEG 2000)
- Examples
- Prototypes of Annex II & III models

We have focussed now on the most important chapters in the DS document with regard to understanding the themes application schemas. Keeping the transformation process from existing resources to the DS models in mind.

But the DS document has also some important annexes of which some are normative while others are only informative.

- For all themes there is Annex A the ATS with a normative part that describes some tests for testing the conformity of a transformed dataset with the **IR on interoperablility**, and an informative part describing tests for testing the conformity with the TG requirements.
- Annex B is informative and lists the all the Use Cases that were taken into account while developing the DS.
- Annex C is again normative since it includes the code list values defined in the Application schemas of the theme.
- Other annexes can be defined depending on the need of that specific theme, like specific encodings rules, examples of implementation, …

Conclusions

71

- Through INSPIRE public sector data will become available over the next years for 34 spatial data themes
- INSPIRE provides a comprehensive framework for interoperability of spatial data
- Understanding the principles of interoperability in INSPIRE is not a piece of cake, however it is a requirement to remove obstacles in data sharing
- Platforms to share methods and infrastructure components for re-use in other sectors
- INSPIRE data can be combined with other data to enable cross-sector & cross-border integration
- Hook to other interoperability frameworks => Linked Open Data

Before having a look at the DS document of CP as an example. I would like to draw some conclusions to resume this module.

- First of all as a result of the implementation of the INSPIRE directive by 2020, public sector data will be available for 34 spatial data themes in a harmonised way;
- What should have become clear to you through this seminar is that INSPIRE provides a comprehensive framework for interoperability of spatial data;
- But that understanding and applying the principles of interoperability is not evident although it is an obligation in order to make data sharing possible;
- At this moment there are also platforms that you can consult online, specially via the JRC, to share methods and components that can be re-used for establishing these data transformation operations in other sectors;
- And hopefully in the future these INSPIRE data can be combined with other data to reach a full cross-sector & cross-border integration.
- And also with regard to the future: apart from this INSPIRE framework there is also another big interoperability framework which is called **Linked Open Data (LOD)** and which tries to open up data to the public and put it online. Probably these frameworks can be connected together with other frameworks to benefit from one another.

72

Data Specification example

Introduction to Annex I data specifications for cadastral parcels

In the following slides I will try to guide you through the DS document of CP which is an annex I theme. I'll just highlight some elements that are really theme specific by following the DS document structure.

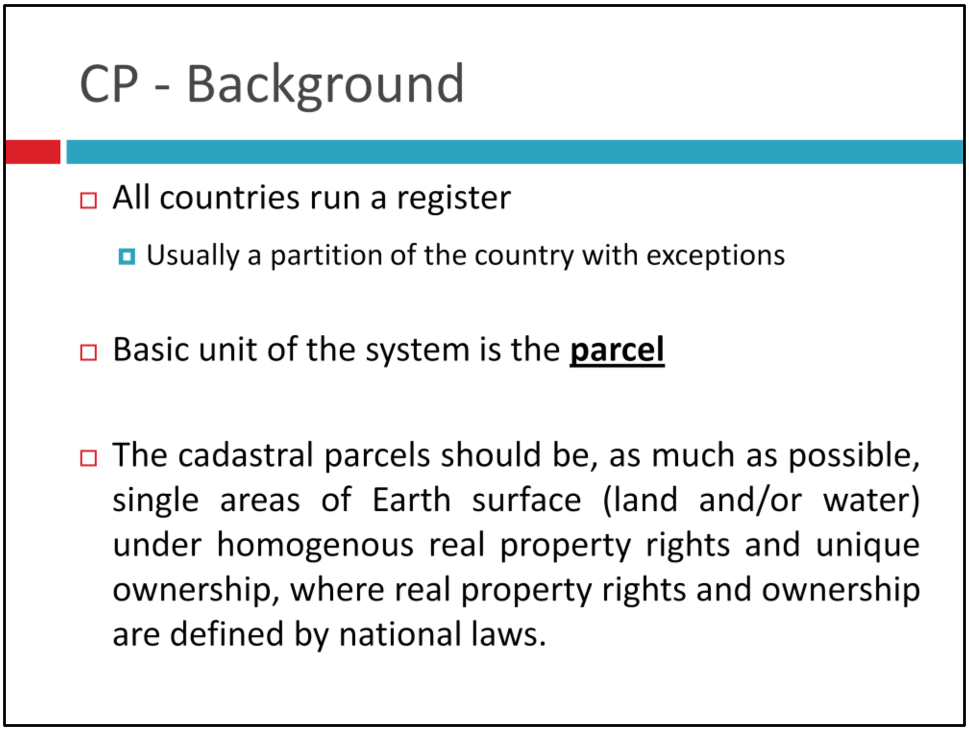
CP - Scope

- The scope of the cadastral information in the INSPIRE context is limited to the geographic side of the cadastral information systems (land administration)
- INSPIRE does not aim at harmonising the concepts of ownership and rights related to the parcels
- Cadastral parcels should serve the purpose of generic information locators. Having included the reference to the national registers as a property (attribute) of the INSPIRE parcels, national data sources can be reached.

If you look at the scope of CP as listed in the DS document you can see that in the INSPIRE context cadastral information is limited to the geographic side of the cadastral information systems (managed by a land administration). So all business information related to the geographic component is out of scope and should not be made available throughout INSPIRE.

Inspire does not aim at harmonising the concepts of ownership and rights related to the parcels. So this is also not modelled within INSPIRE.

Cadastral parcels should serve the purpose of generic information locators. This means that it is sufficient to include only a reference with a code or a key (as an attribute of a parcel object) to the national registers in order to reach the information stored in the national data sources.



By looking at the background documentation of the CP DS, you will see that:

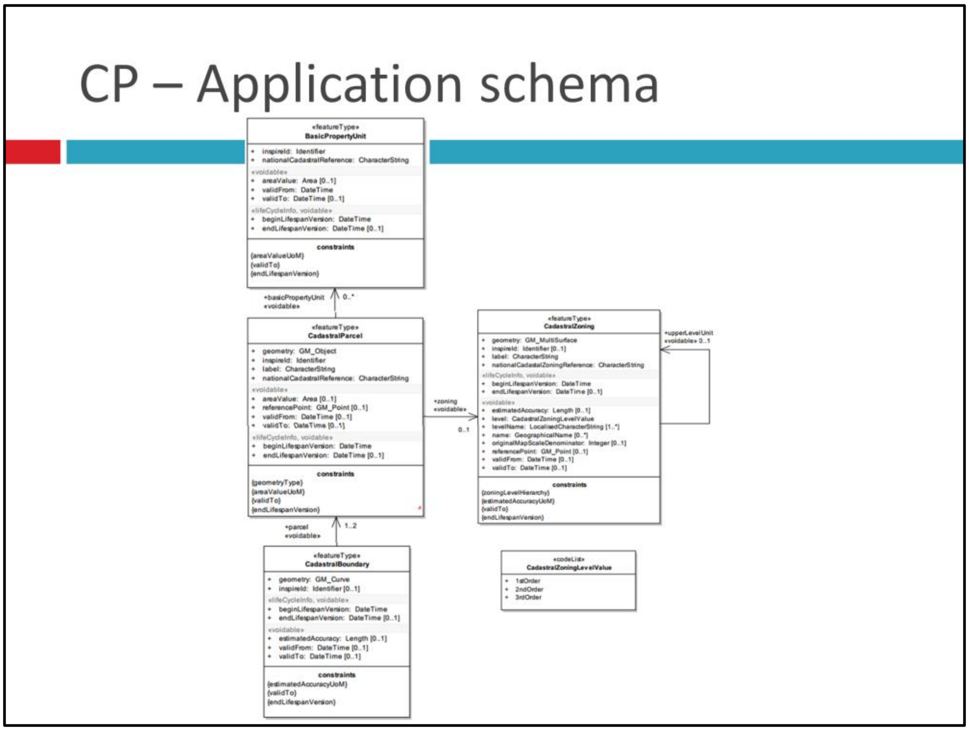
- All countries run a register (for storing cadastral information) – sometimes this is only covering a part of the country;
- The basic unit that is used in this registers is the parcel
- And one of the criteria is that a parcel should be a single area of Earth surface (if possible) under homogenous property rights and unique ownership.

CP – Basic components

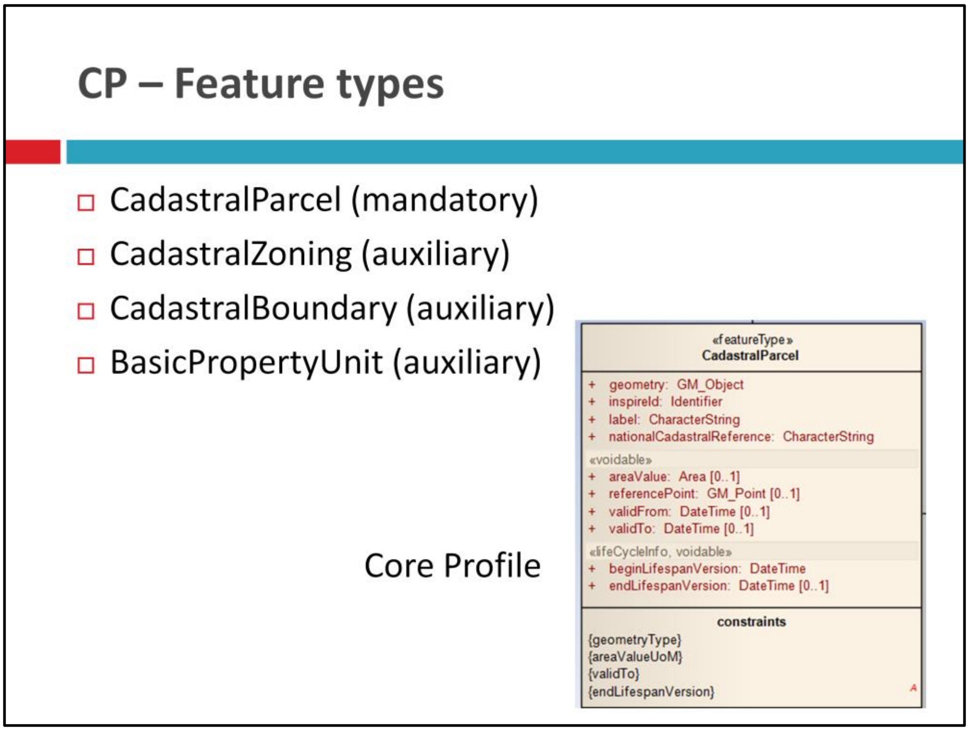
- Parcel (basic unit)
- Subdivision (municipalities, sections, districts, parishes, urban or rural blocks, etc)
  - Carry information for the parcels inside the subdivision: accuracy or scale
- Cadastral boundaries
  - Only neccessary if spatial accuracy is associated with them

By taking into account this background information we can better understand how the basic components of the CP data model were chosen.

- So first of all there is the Parcel defined as the basic unit
- Sometimes parcels are aggregated into higher level unit like municipalities, sections, districts and so on. These aggregations are modelled as “CadastralZoning” in the CP data model. And they can carry higher level information concerning that aggregation.
- In the case that cadastral information is attached to the boundary of cadastral parcels it is necessary to store the cadastral boundaries otherwise it is optional.



This slide shows you the UML diagram of the CP application schema which is not complicated. We can distinguish 4 classes that are related to each other. Centrally we have the cadastral parcel which is part of a cadastral zone (the class on the right). The cadastral parcel is defined by its boundaries (the class at the bottom) and on top of the slide we have the basic property unit which is optional for countries where national cadastral reference is given to one or a group of parcel(s) defined by unique ownership and homogeneous real property rights.



Here we see a list of the four classes as feature types (i.e. spatial identifiable objects). And we can see that there is only 1 feature type mandatory (cadastral parcel) all the rest is auxiliary (mandatory in specific conditions) but the core profile is actually limited to the Cadastral parcel feature type.

This is a good example that implementing INSPIRE DS is not so complex as often is argued. If we look to the attributes that belong to a CP, we see:

- geometry
- inspireID: External object identifier of the spatial object
- Label: Text commonly used to display the cadastral parcel identification
- nationalCadastralReference: Thematic identifier at national level of the cadastral parcel which ensures the link to the national cadastral register or equivalent.

Some voidable properties

- areaValue: Registered area value
- referencePoint: point within the parcel used for label placement
- validFrom and validTo: Official date and time when the cadastral parcel is legally established and deprecated;

two lifeCycleInfo properties:

- beginLifespanVersion and endLifespanVersion: Date and time at which this version of the spatial object was inserted or changed in the spatial data set and the date when it was superseded or retired

Apart from the attributes you can see in the last compartment of the class that there are also some constraints defined which put some extra limitations on the content of some attributes.

CP – Requirements & Recommendations

Recommendation 2 Edge-matching between cadastral parcels in adjacent data sets should be done. Ideally, there should be no topological gaps or topological overlaps between cadastral parcels in adjacent data sets. Status of edge-matching should be reported as metadata, under lineage element (see annex D).

Requirement 7 All instances of feature type CadastralParcel shall carry as a thematic identifier the attribute nationalCadastralReference. This attribute must enable users to make the link with rights, owners and other cadastral information in national cadastral registers or equivalent.

Recommendation 5 Cadastral parcels should be provided, as much as possible, as GM_Surface.

NOTE Some countries (e.g. Germany, Spain, France) have a few percentage of multi-surface parcels. These parcels may be provided as GM_MultiSurface.

Recommendation 6 There should be no topological overlaps between cadastral parcels.

Recommendation 7 There should be no topological gaps between cadastral parcels.

Recommendation 8 If life-cycle information is not maintained as part of the spatial data set, all spatial objects belonging to this data set should provide a void value with a reason of "Unpopulated".

On this slide you see an IR requirement (remember the style, red and the double border), and some recommendations that were defined for the CP theme.

The IR requirement makes the thematic identifier in the form of the nationalCadastralReference mandatory for a CP. It forms the link to the national cadastral registers with ownership and rights information.

There is a recommendation that advises to provide the geometry as a GM_Surface (ie a concept from ISO to model a polygon);

There are some topological recommendations that say to avoid topological overlaps and gaps between CPs. You can wonder why this is only a recommendation and not a requirement because for such a dataset as CP you would not expect such a weak demand on the topological quality.

But not all MS have National cadastral systems with that quality yet, this has to do with history with the inventory method of the parcels. In many countries there is already a digital layer but that does not have any legal value yet and the only legally binding document is an informative description in a kind of codex which does not avoid topological correctness. That is why this was only taken on board as a recommendation and not as a requirement.

CP – Requirements & Recommendations

NOTE 3 A spatial object may change in a way where it is still considered to be the same spatial object; in this case, there will be several versions of the same object.

EXAMPLE On 01/01/2008, there has been new delineation of a cadastral parcel (A) and a new value for attribute areaValue has been computed. Two cases may occur, depending on the life-cycle information management at national level:

- It is considered that it is a new cadastral parcel (B) with a new identifier
- It is considered that it is a new version of the same object (A) with unchanged identifier.

(For instance, in France, the first case occurs when there is a new survey of cadastral data on a whole area. The second case occurs when there is a new survey for an individual cadastral parcel).

In first case:

- the spatial object "parcel A" will get for attribute *endLifespanVersion the value* 01/01/2008.
- a new spatial object "parcel B" will be created; it will get a new identifier and this new spatial object "parcel B" will also get for attribute *beginLifespanVersion the value* 01/01/2008.

In second case,

- the spatial object "parcel A" will get for attribute *endLifespanVersion the value* 01/01/2008.
- a new version of the spatial object "parcel A" will be created and will get for attribute *beginLifespanVersion the value* 01/01/2008.

This new version of the spatial object "parcel A" will be identified by a new value for objectIdentifier.version.

Recommendation 9 Life-cycle rules are up to each data provider. They should be documented as metadata, under lineage element

This slide illustrates that the DS is a technical guidance for data providers. In this case it shows how to deal with lifecycle issues of CPs. It gives some examples from reality where CP can change in time and the way these types of changes should be treated: either as a new parcel with a new identifier or just as a new version of the parcel with the same identifier, which than implements that the old version becomes retired.

CP – Requirements & Recommendations

Recommendation 11 From temporal point of view, cadastral parcels should be published for INSPIRE if and only if they are published in national register. Cadastral parcels under internal updating process should not be published for INSPIRE.

EXAMPLE In most countries, there are parcels under dispute for which a provisory solution has been adopted in national register. These parcels should be published for INSPIRE. There are also parcels under splitting process; this splitting process is generally internally managed, new provisory parcels are created but these new parcels are published in national register only once all operations (survey, checking, validation, registration) have been achieved. In this case, the provisory parcels should not be published for INSPIRE but only the new definitive, validated ones.

Here we have an important recommendation that says that INSPIRE cadastral parcels should only be published when the parcels are officially published in the national register. So for work in progress it is not necessary to make it available through INSPIRE unless it is published in the National register i.e. when it is officially validated.

CP - Geometry

- 0-, 1-,2-,2,5 dimensional geometries

Recommendation 5 Cadastral parcels should be provided, as much as possible, as GM_Surface.

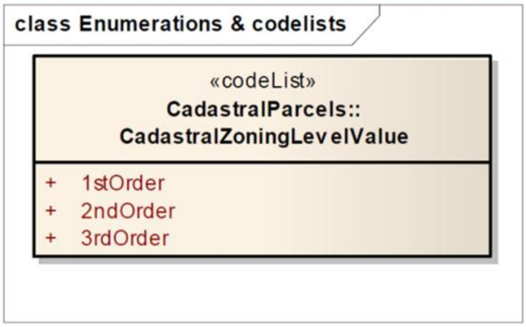
NOTE Some countries (e.g. Germany, Spain, France) have a few percentage of multi-surface parcels. These parcels may be provided as GM_MultiSurface.

Recommendation 6 There should be no topological overlaps between cadastral parcels.

Recommendation 7 There should be no topological gaps between cadastral parcels.

These recommendations on geometry we have seen already, so polygons are recommended but by looking at the model we have seen that 1-dimensional features (boundaries) are also acceptable under certain circumstances.

CP – Enumerations/codelists



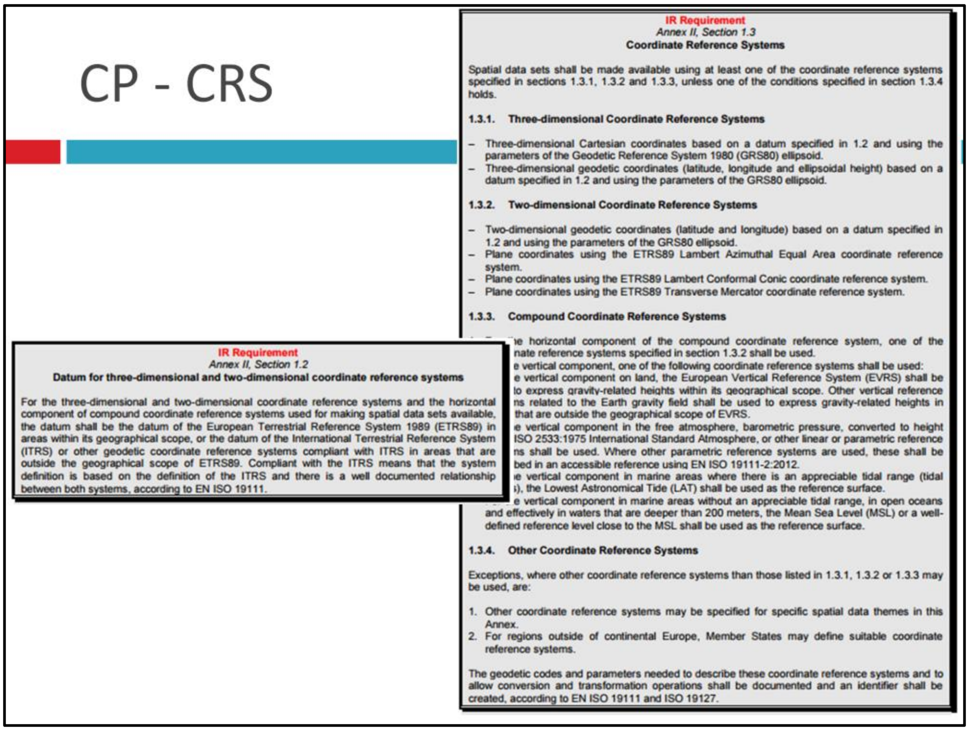
CadastralZoningLevelValue	
Definition:	Levels of hierarchy of the cadastral zonings.
Description:	NOTE The higher levels in the administrative units theme (province, state) are not repeated in this code list.
Status:	Proposed
Stereotypes:	«codeList»
Governance:	Centrally managed in INSPIRE code list register. URN: urn:x-inspire:def:codeList:INSPIRE:CadastralZoningLevelValue

Here you have an example of a codelist that provides the values that should be used to indicate the levels of aggregation (hierarchy) of cadastral zones.

CP - Codelist

```xml
  <description>-- Definition --&#13;
Levels of hierarchy of the cadastral zonings.&#13;
&#13;
-- Description --&#13;
NOTE The higher levels in the administrative units theme (province, state) are not repeated in this code list.</description>
  <identifier codeSpace="urn:x-inspire:specification:gmlas:CadastralParcels:3.0">CadastralZoningLevelValue</identifier>
  <dictionaryEntry>
    <Definition gml:id="_22835_40498">
      <description>-- Definition --&#13;
Uppermost level (largest areas) in the hierarchy of cadastral zonings, equal or equivalent to municipalities.</description>
      <identifier codeSpace="urn:x-inspire:specification:gmlas:CadastralParcels:3.0/CadastralZoningLevelValue">1stOrder</identifier>
      <name>1stOrder</name>
    </Definition>
  </dictionaryEntry>
  <dictionaryEntry>
    <Definition gml:id="_22835_40499">
      <description>-- Definition --&#13;
Second level in the hierarchy of cadastral zonings.</description>
      <identifier codeSpace="urn:x-inspire:specification:gmlas:CadastralParcels:3.0/CadastralZoningLevelValue">2ndOrder</identifier>
      <name>2ndOrder</name>
    </Definition>
  </dictionaryEntry>
  <dictionaryEntry>
    <Definition gml:id="_22835_40500">
      <description>-- Definition --&#13;
Third level in the hierarchy of cadastral zonings.</description>
      <identifier codeSpace="urn:x-inspire:specification:gmlas:CadastralParcels:3.0/CadastralZoningLevelValue">3rdOrder</identifier>
      <name>3rdOrder</name>
    </Definition>
  </dictionaryEntry>
</Dictionary>
```

This is the same codelist but provided in GML notation. You can see the gml-tags with the values of the codelist. All the INSPIRE managed codelists are made available online through the INSPIRE codelist register.



This is an extract from chapter 6 of the DS of CP. It shows some requirements concerning CRS. Both requirements apply to all data themes.

The first requirement defines the datum to be used as ETRS89 in areas within its geographical scope and ITRS or ITRS compliant should be used in areas outside the geographical scope of ETRS89.

The second requirement specifies the possible CRSs that can be used for spatial datasets. At least one of the possibilities should be chosen.

CP – Data Quality

- Rate of missing elements and accuracy
- No requirements on minimal data quality, only recommendations

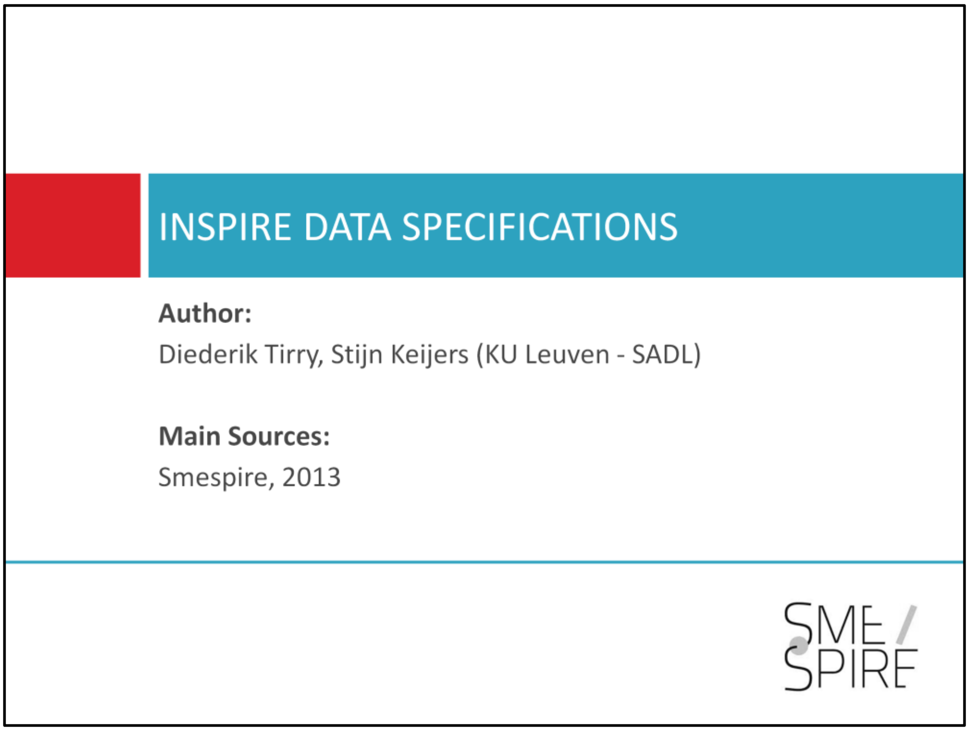
- Accuracy on three levels possible:
  - Statement for the whole dataset
  - Attribute on the cadastral zonings
  - Attribute on the cadastral boundaries

According the DS of CP some data quality elements will be used, apart from the general ones (logical consistency- conceptual and domain consistency), these are the completeness (rate of missing elements) and the positional accuracy (mean value of positional uncertainties, 1D or 2D).

For those elements no minimum requirements are defined but only some recommendations. F.e. the missing rate should target 0% missing elements) or the positional accuracy expressed as the Mean value of positional uncertainties must not be larger than 1m in urban areas and 2.5m in rural areas.

This information on accuracy can be stored on different levels:

1. as a general statement on the dataset level (MD element);
2. as an attribute on the object level of cadastral zoning;
3. or as an attribute on cadastral boundaries.



So far the illustration of the DS document of CP.
Here also ends this module of “basics of INSPIRE data specifications”. I thank you for your attention and I hope you will succeed in answering the questions of the self-test.