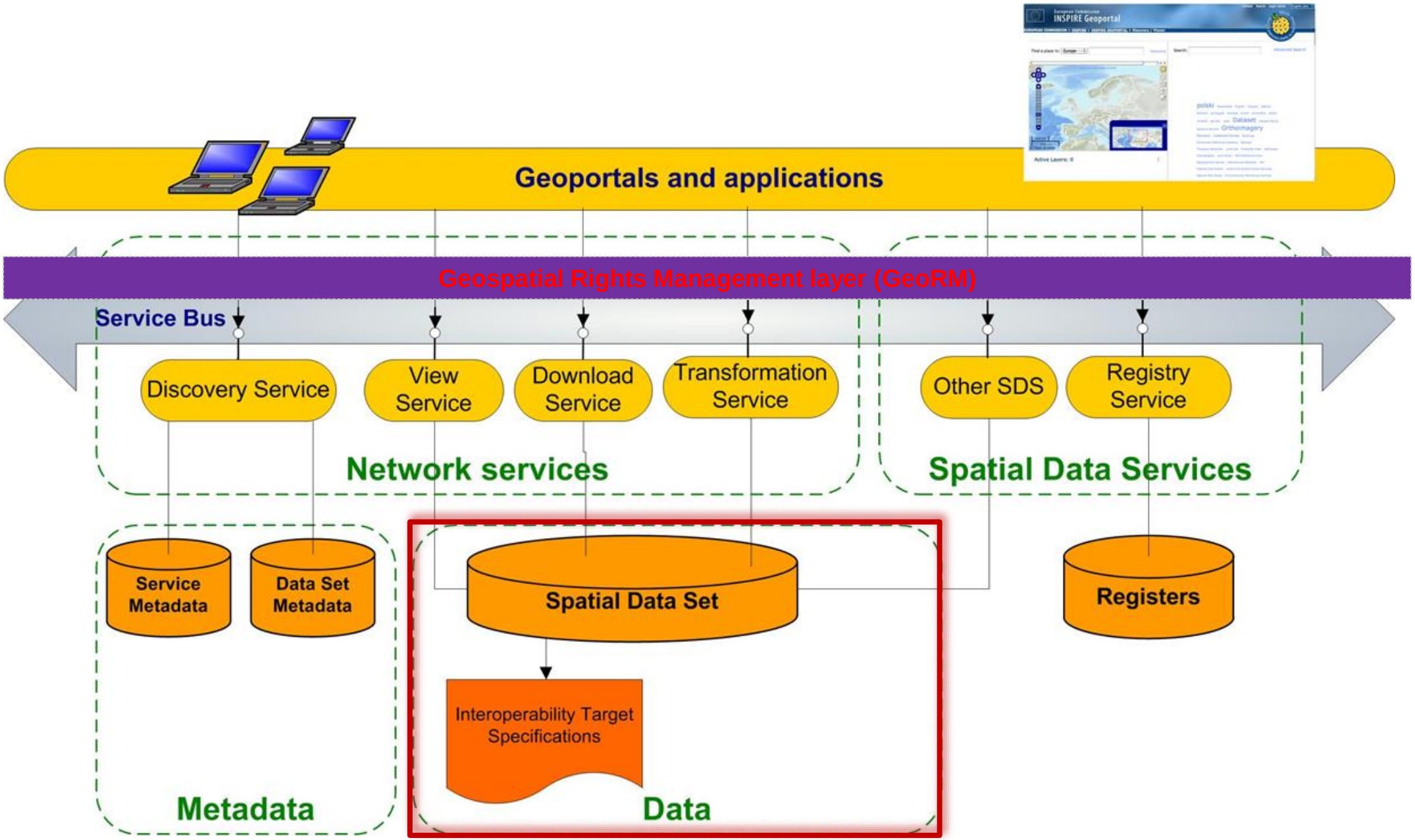
TECHNICAL INTRODUCTION TO SDI

Data models and data specifications

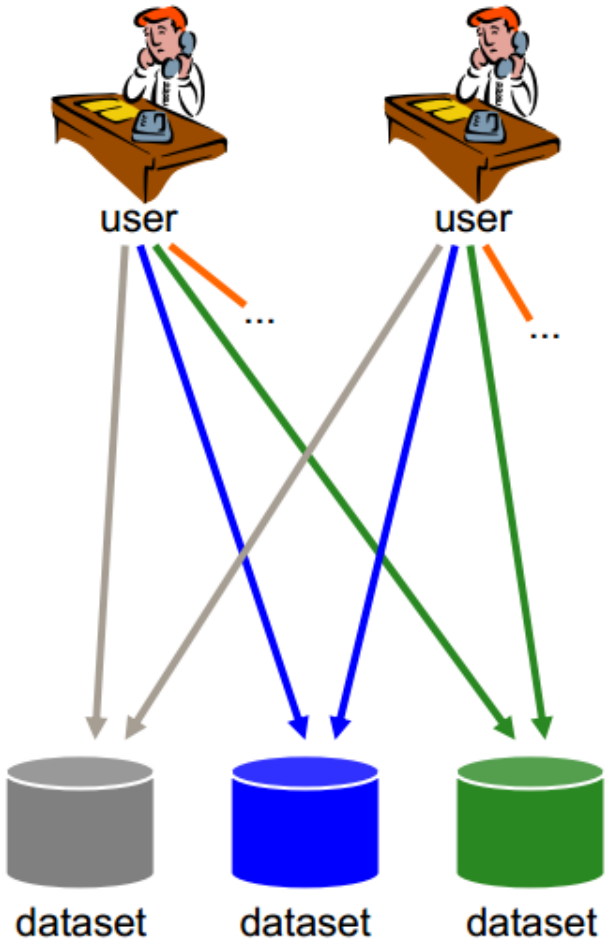
Danny Vandenbroucke

Structure of the Module

1. Motivation and Background
2. Scope & objectives of data specifications for SDI’s
3. The modelling framework for data specifications (ISO 19131)
4. Development of data specifications
5. Examples of data models: ISO 19152
6. Exercise


Geoportals and applications
Geospatial Rights Management layer (GeoRM)
Service Bus
Discovery Service
View Service
Download Service
Transformation Service
Other SDS
Registry Service
Network services
Spatial Data Services
Service Metadata
Data Set Metadata
Spatial Data Set
Registers
Interoperability Target Specifications
Metadata
Data

The starting point



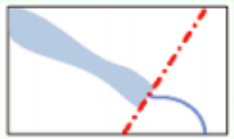
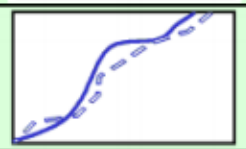
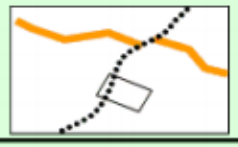
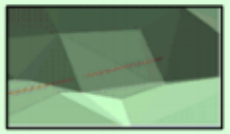
- Access to spatial data in various ways: copies via CD
- User has to deal with interpreting **heterogeneous data** in different formats, identify, extract and post-process the data needed

→ **Lack of interoperability**

Semantic and schematic differences

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

Different spatial representations

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

- **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.

- **Semantic** heterogeneity

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

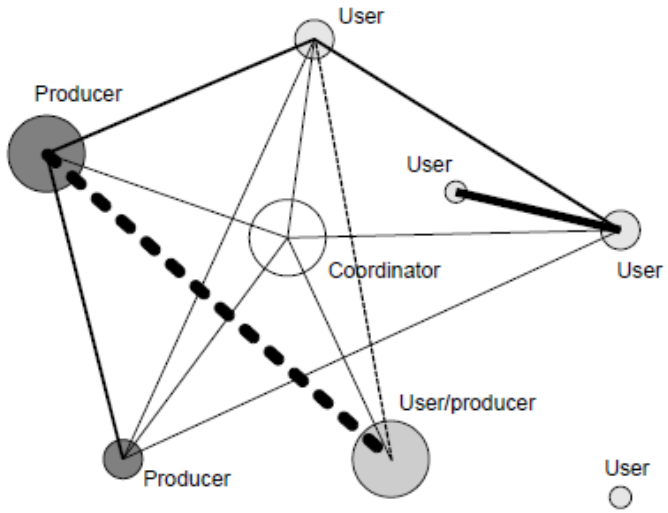
  e.g. Different classifications/definitions of roads when viewed from different perspectives: traffic network route directions, spatial planning… <> 1 on 1 match

- **Technical** interoperability

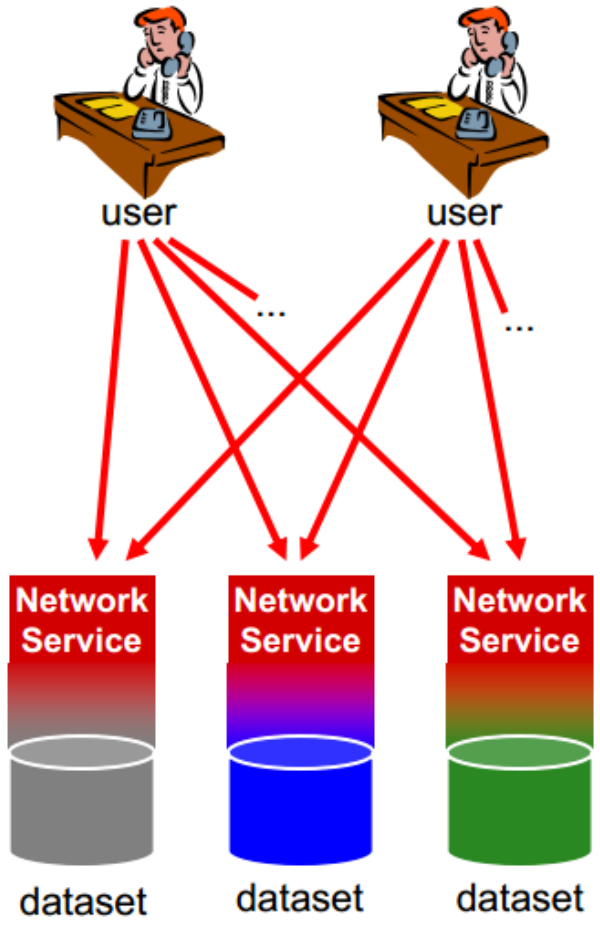
  should guarantee that system components can interoperate

- **Semantic** interoperability

  should guarantee that data content is understood by all in the same way

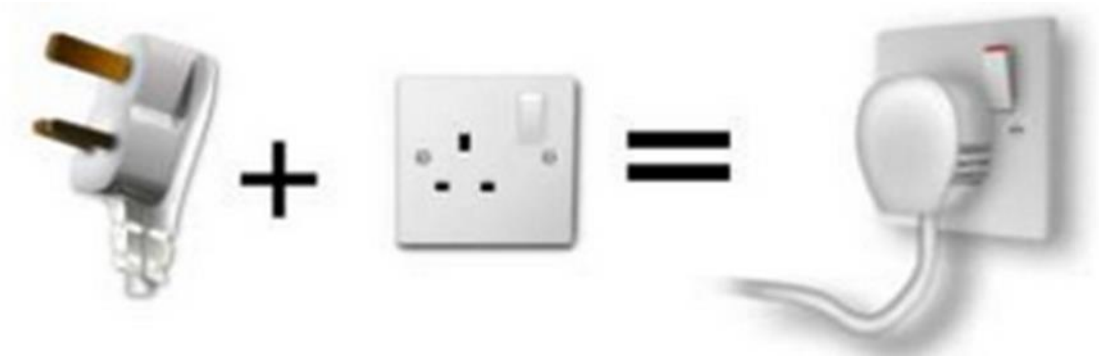


…what a SDI aiming at

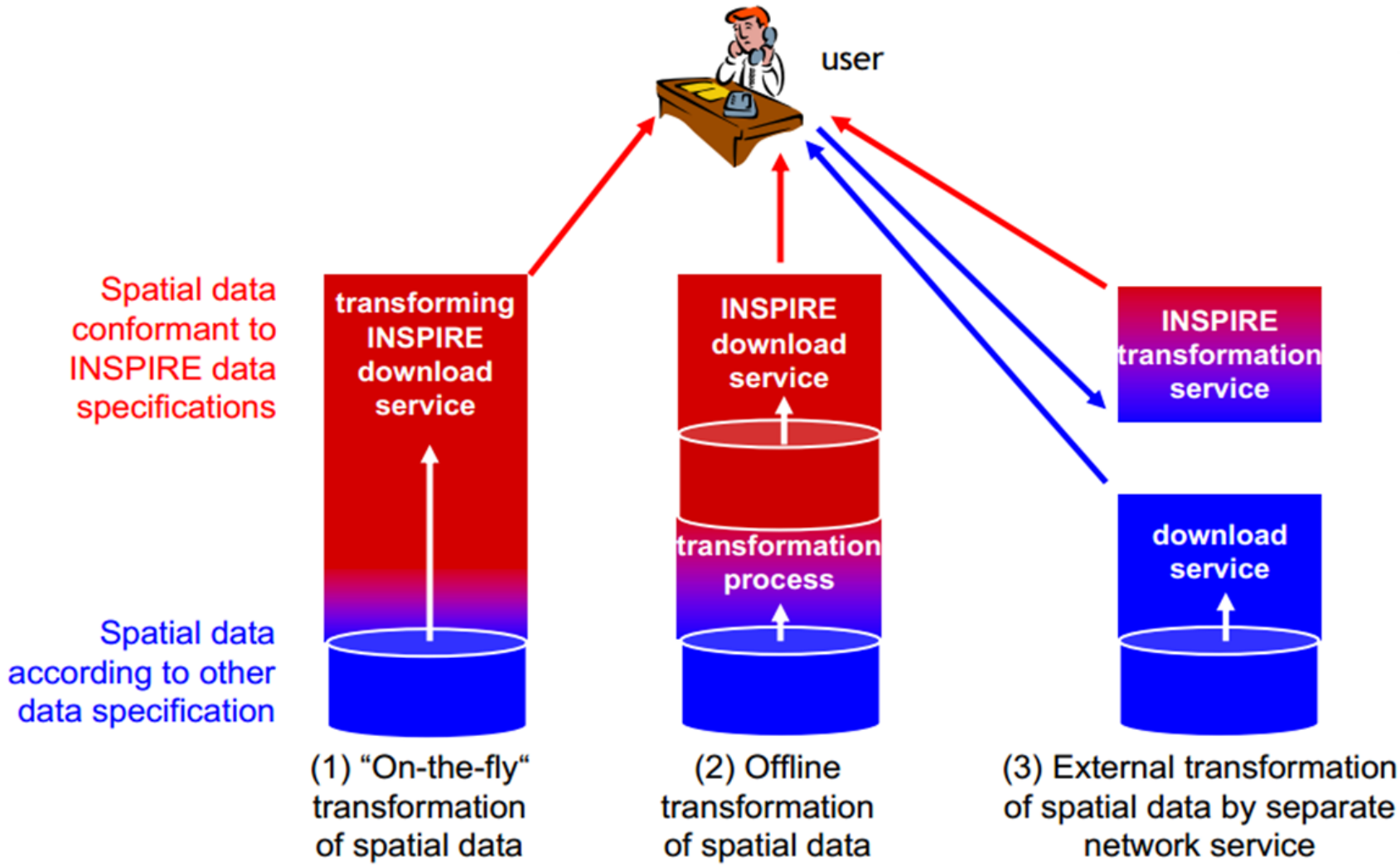


- Provide access to spatial data via **network services** and according to a **harmonised data specification** to achieve **interoperability** of data
- Datasets used within organizations may remain unchanged
- Data or service providers have to provide a transformation between their **internal model** and the **harmonised data specification**

- Facilitate data use and interoperability by adopting **common cross-domain models** to exchange data

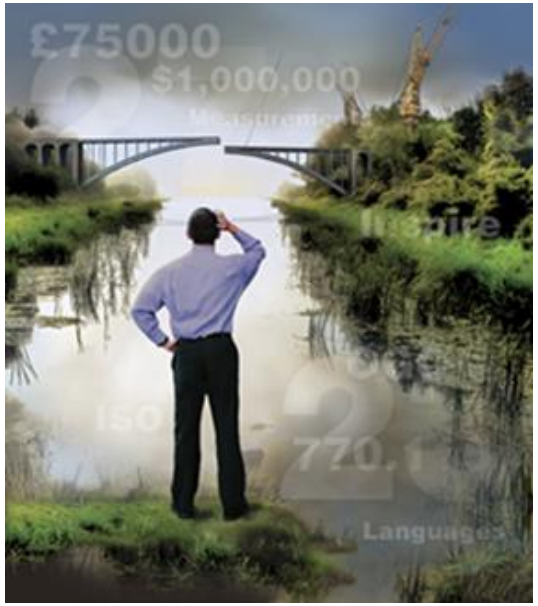


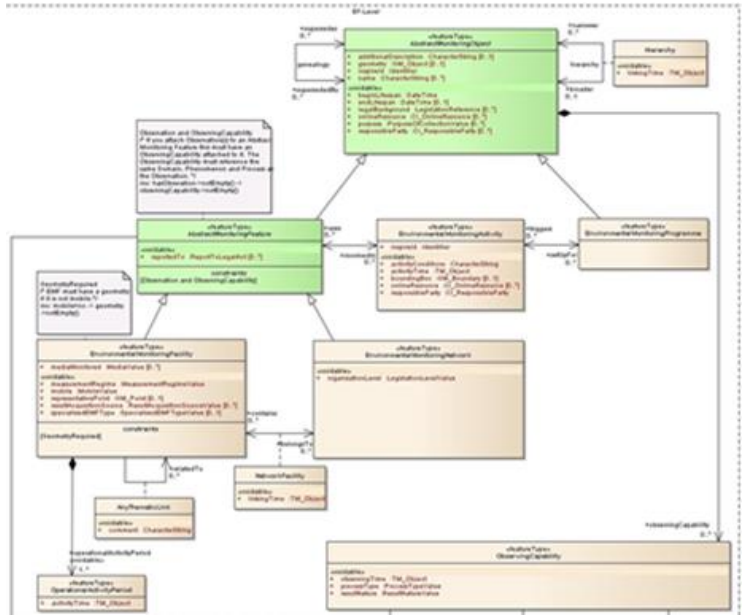
DATA INTEROPERABILITY


user
Spatial data conformant to INSPIRE data specifications
Spatial data according to other data specification
transforming INSPIRE download service
INSPIRE download service
transformation process
INSPIRE transformation service
download service
(1) "On-the-fly" transformation of spatial data
(2) Offline transformation of spatial data
(3) External transformation of spatial data by separate network service

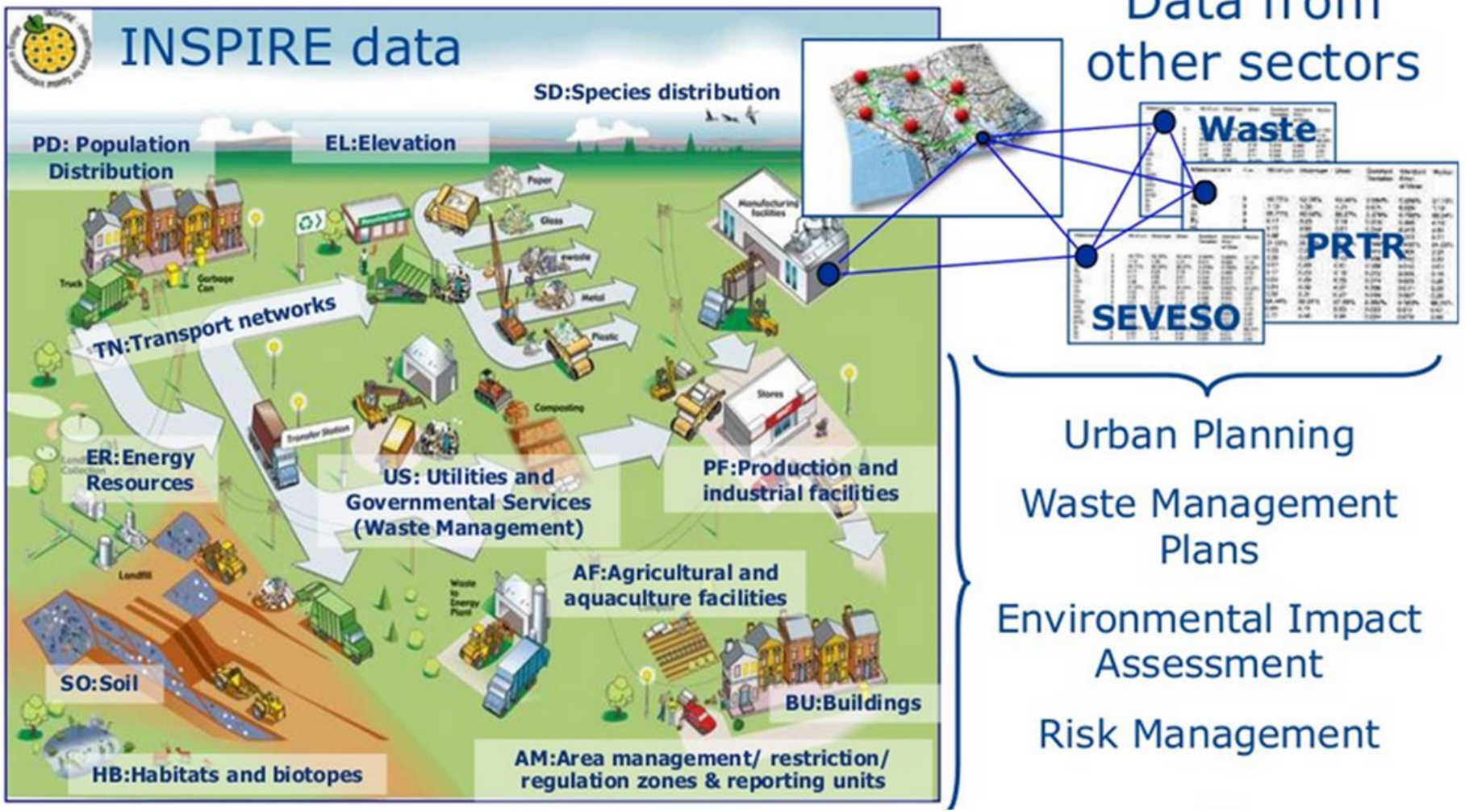
- Member States should make data available within the scope of INSPIRE using
  - the same spatial **object types** (and definitions)
  - the same **attributes** (and definitions, types, code lists) and relationships to other types, e.g. BuildingHeight, BuildingSize
  - a common **encoding** (GML application schemas)
  - common **portrayal** rules

- This facilitates interoperability and pan-European/cross-border applications (e.g. information systems, reporting systems, forecasting models)





Cross-sector data interoperability



Source: EC Joint Research Centre

- 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."*

- Art 8(2): The implementing rules shall address the following aspects of spatial data:

  (a) a common framework for the unique identification of spatial objects, to which identifiers under national systems can be mapped in order to ensure interoperability between them;

  (b) the relationship between spatial objects;

  (c) the key attributes and the corresponding multilingual thesauri commonly required for policies which may have an impact on the environment;

  (d) information on the temporal dimension of the data;

  (e) updates of the data.

- …

Data Specifications

Scope

Annex I

Coordinate reference systems


Geographical grid systems

Geographical names


Administrative units


Addresses


Cadastral parcels


Transport networks


Hydrography


Protected sites

Annex II

Elevation


Land cover


Ortho-imagery


Geology

Annex III


Statistical units


Buildings


Soil


Land use


Human health and safety


Utility and governmental services


Environmental monitoring facilities


Production and industrial facilities


Agricultural and aquaculture facilities


Population distribution – demography


Area management/restriction/regulation zones & reporting units


Natural risk zones


Atmospheric conditions


Meteorological geographical features


Oceanographic geographical features


Sea regions


Bio-geographical regions

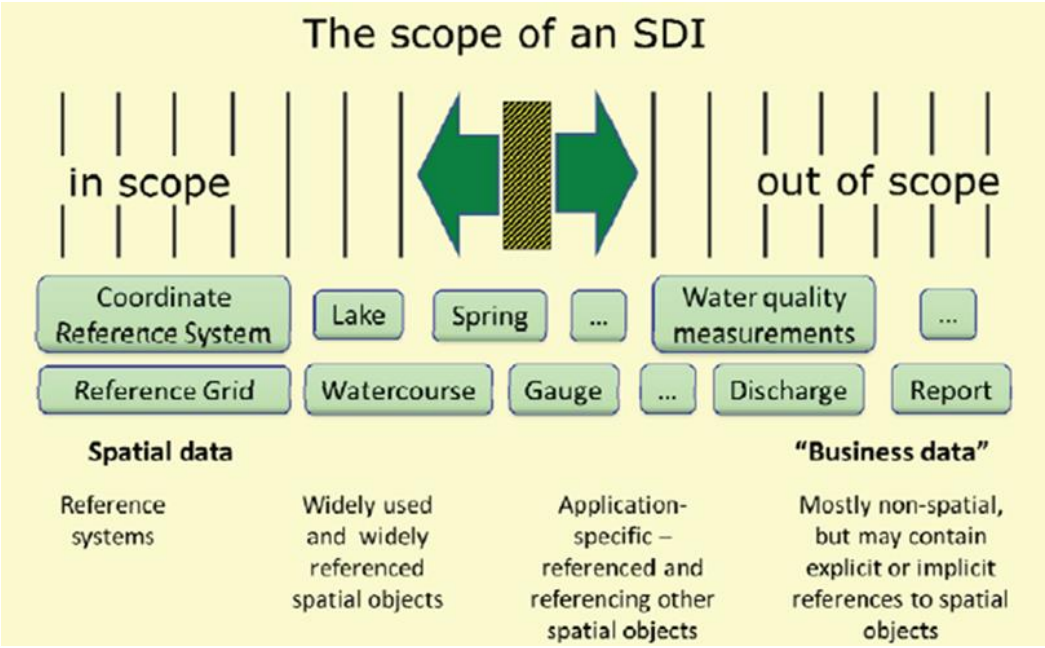

Habitats and biotopes


Species distribution


Energy Resources


Mineral resources

- Scope is **spatial** data – not all kinds of thematic/descriptive data



- Re-use the INSPIRE data specs for own usage
  - Extensions
  - Additional constraints
  - Re-use of common objects

Data specifications

Exercise 1: Find your scope

Exercise 1: Find your scope

Data Specification Scope

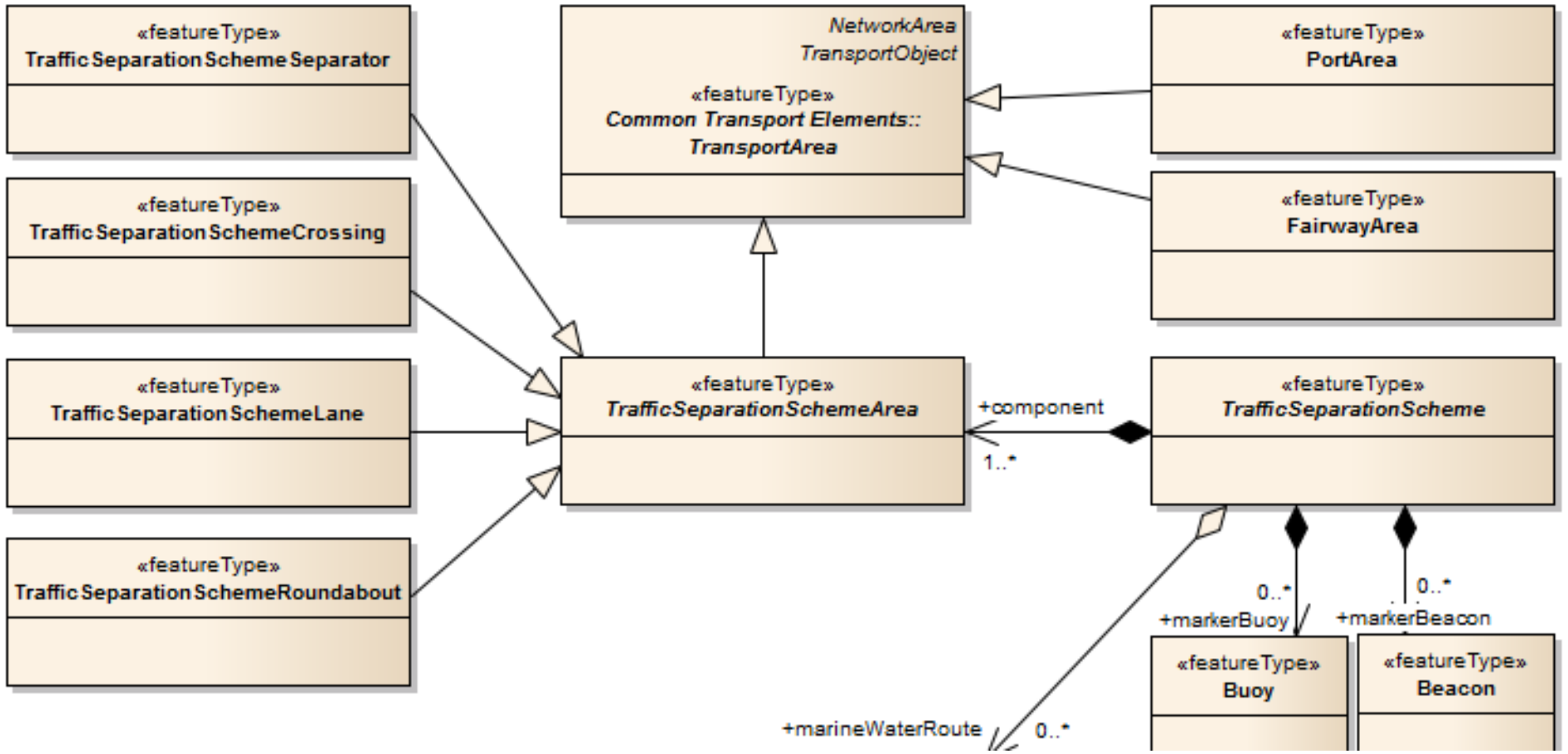
Example of INSPIRE

- Go to INSPIRE website (https://inspire.ec.europa.eu/inspire-tools)
- Use the tool “Find your scope” (toolkit):
  1. In catalogue of INSPIRE objects:
     - find “zone” --> limit to only “Spatial object type” --> narrow search “terrestrial zone”
     - Which Object, INSPIRE Data Theme, Application Schema
     - What are the other possible specialisations of ***TransportArea***?
  2. To find the Spatial object that should be used for a dataset that stores the locations of stations where magnetic measurements are performed. (use “Direct Search”)
  3. Find your own scope…

Exercise 1: Result (1)

Inspire Data Specification Scope

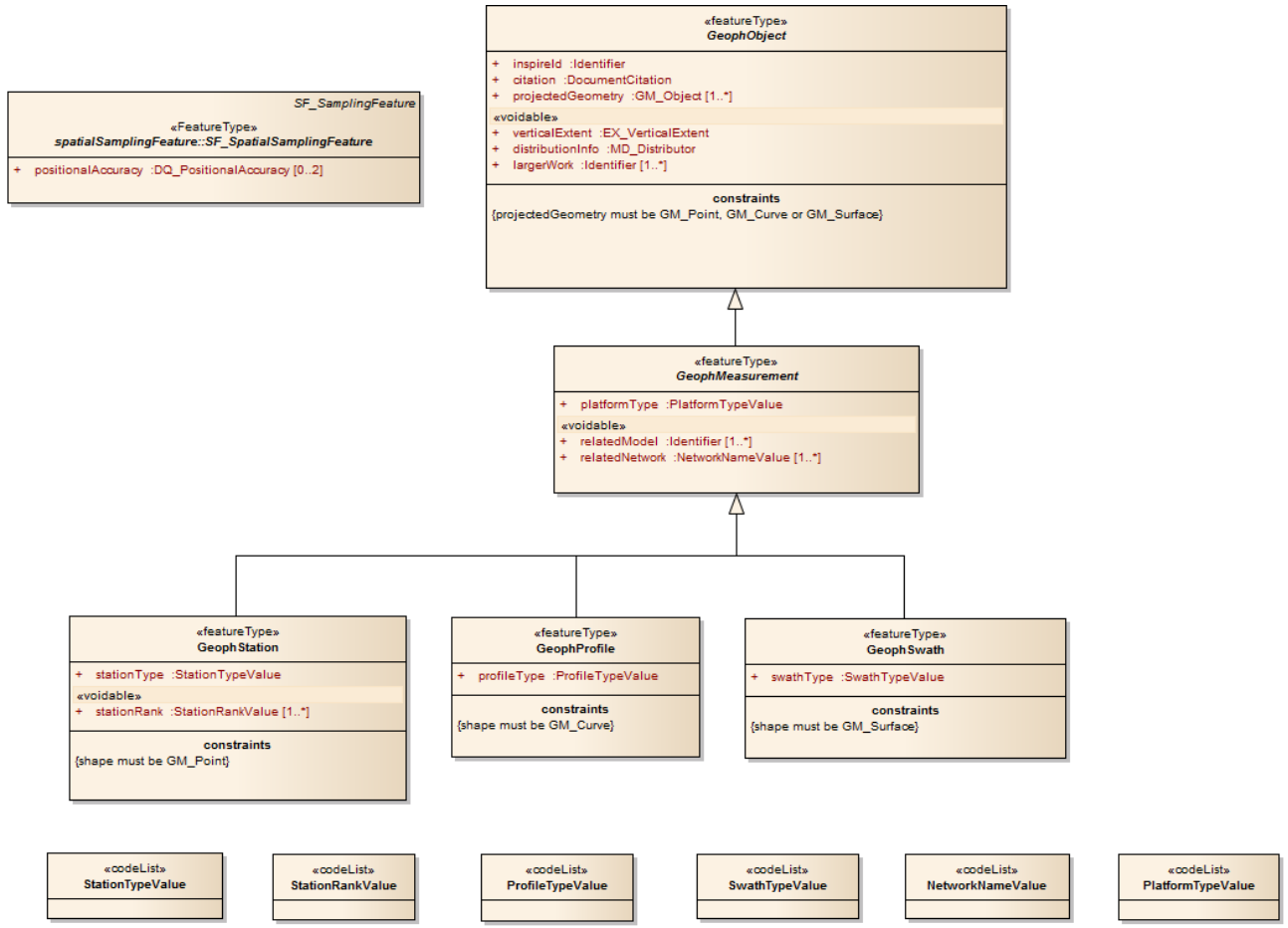
1. PortArea – Transport Networks – Water TN



Exercise 1: Result (2)

Inspire Data Specification Scope

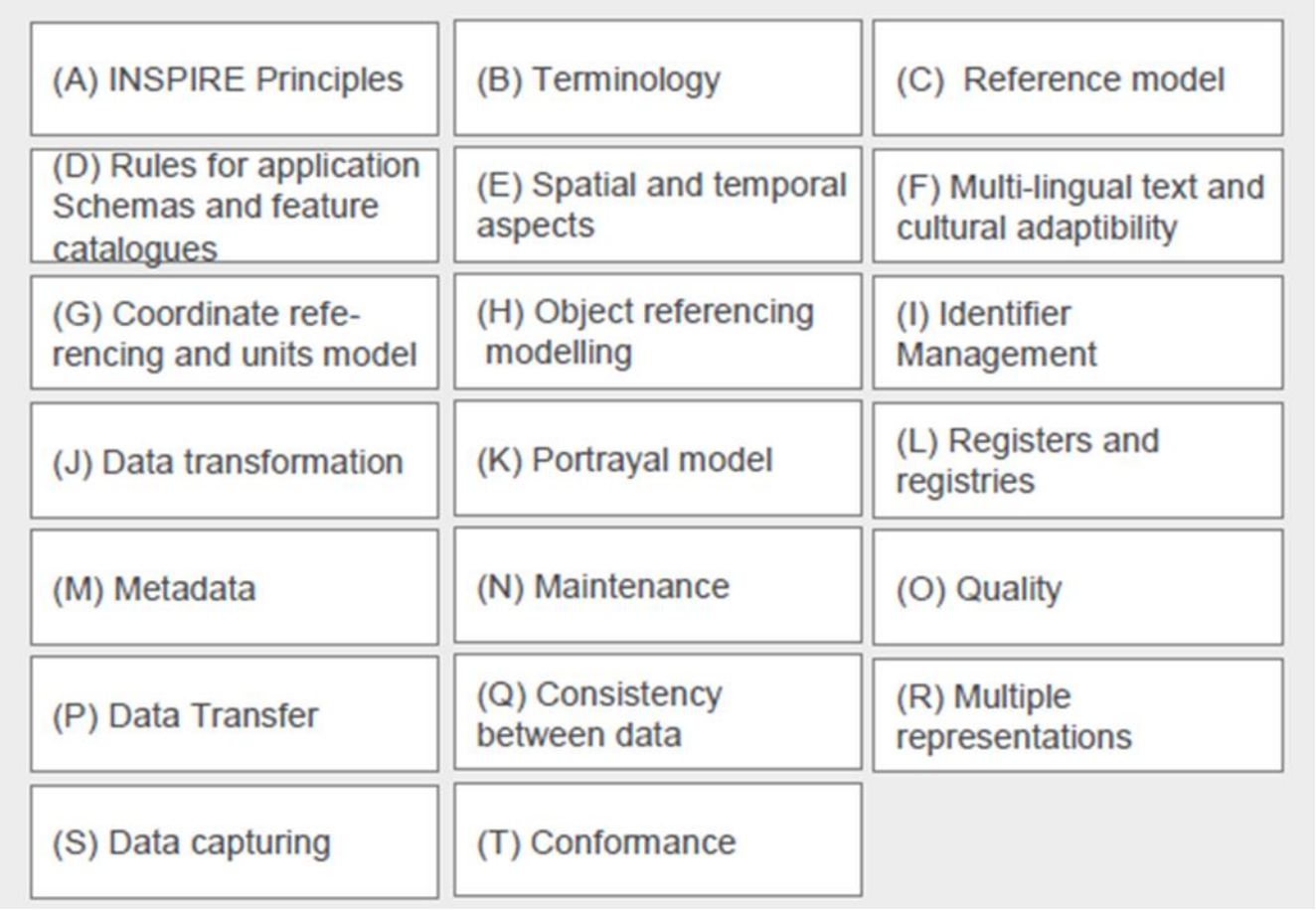
- Direct search: “magnetic field”
  - Relevant objects? Observed Event (NZ) vs **Geoph Station (GE - Geophysics)**



Data Specifications

A modelling framework supported by the ISO 19100 series of standards

Example from INSPIRE



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.

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.

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.

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**

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)

P. Data Transfer	•This component will describe methods for encoding application and reference data as well as information products. •The encoding of spatial objects will in general be model-driven, i.e. fully determined by the application schema in UML. Where appropriate, existing encodings will continue to be used.(D2.7) •To support network services that are implemented as web services, spatial objects are expected to be primarily encoded in GML. Coverage data is expected to use existing encodings for the range part, e.g. for the pixels of an orthophoto.
S. Data Capturing	This component will describe the data specification-specific criteria regarding which spatial objects are to be captured and which locations/points will captured to represent the given spatial object (e.g. all lakes larger than 2 ha, all roads of the Trans European Road Network, etc.).
T. Conformance	how conformance of data to a data specification is tested, i.e. it will be necessary to apply conformance tests as specified in the individual data specification.

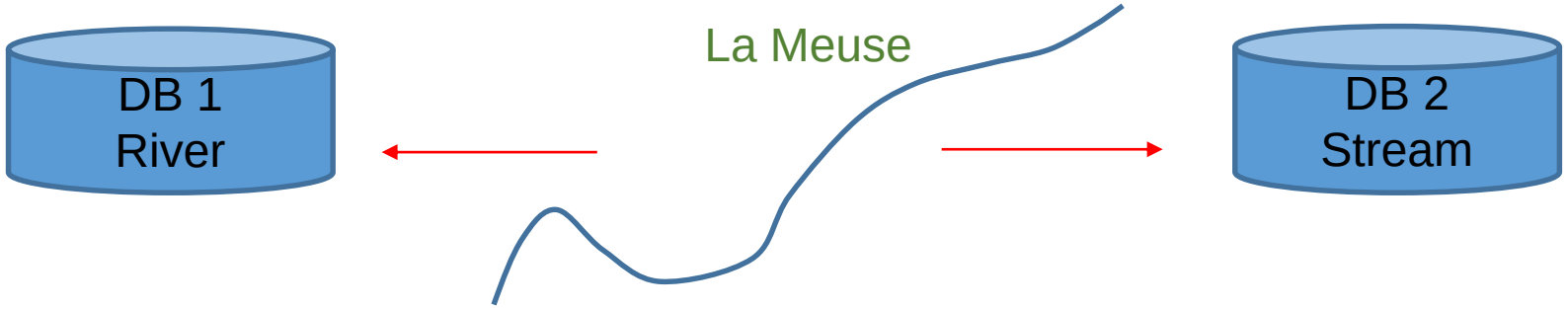
Generic Conceptual Model

- Where does the abbreviation GCM stands for (used in the INSPIRE context)?
- How many thematic domains, divided in how many annexes, are addressed by the INSPIRE directive?

34 themes, 3 annexes

- Which kind of heterogeneity?

Semantic Heterogeneity



- What are next from “Encoding” and “Harmonised vocabularies” the other two major cornerstones of data interoperability?

“Conceptual data models” & “Registers”

Data Specifications

Development

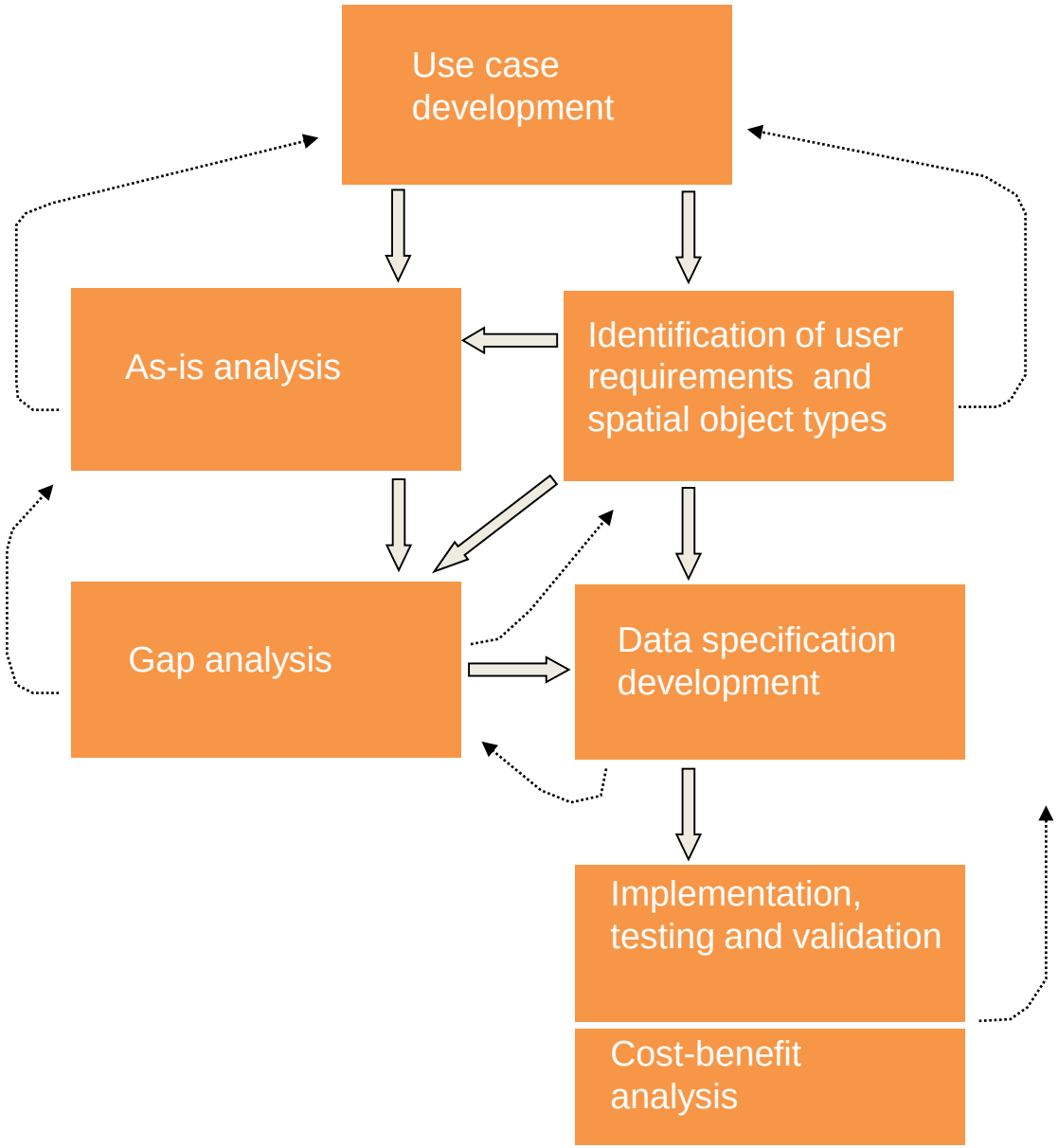
DATA SPECIFICATION

||

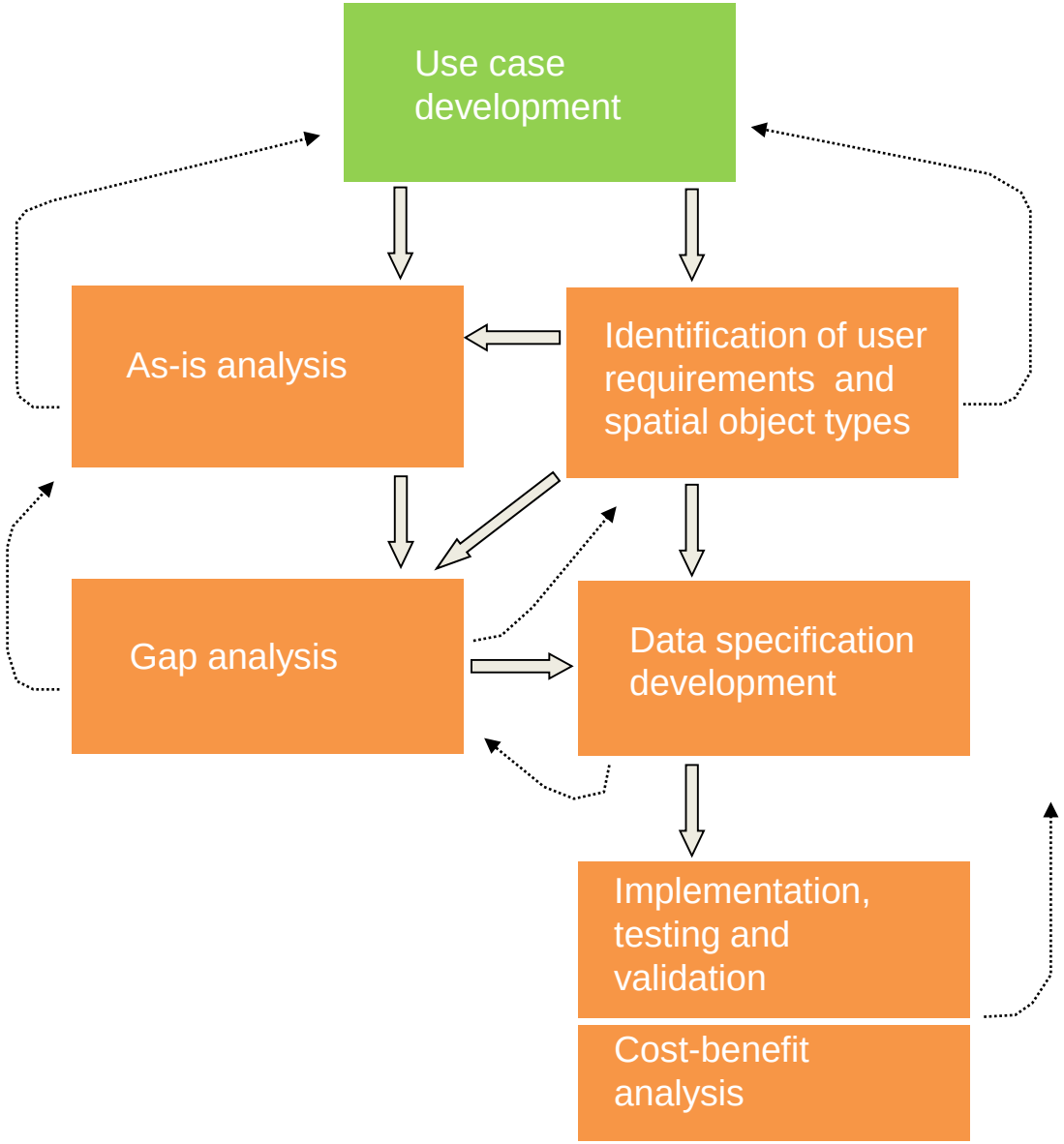
Synonym to data product specification

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]



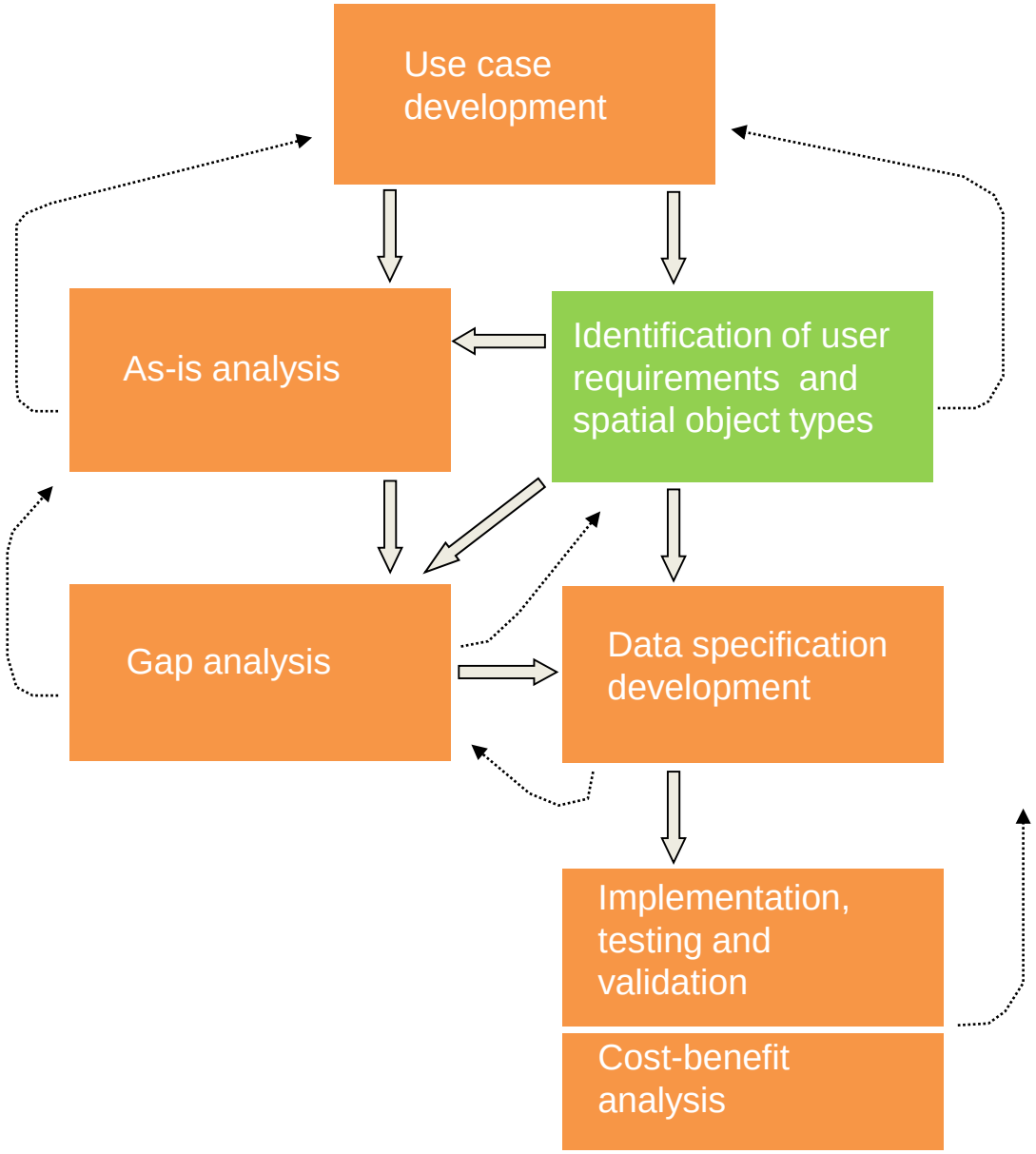
Step-wise methodology



Step 1

Major sources are:

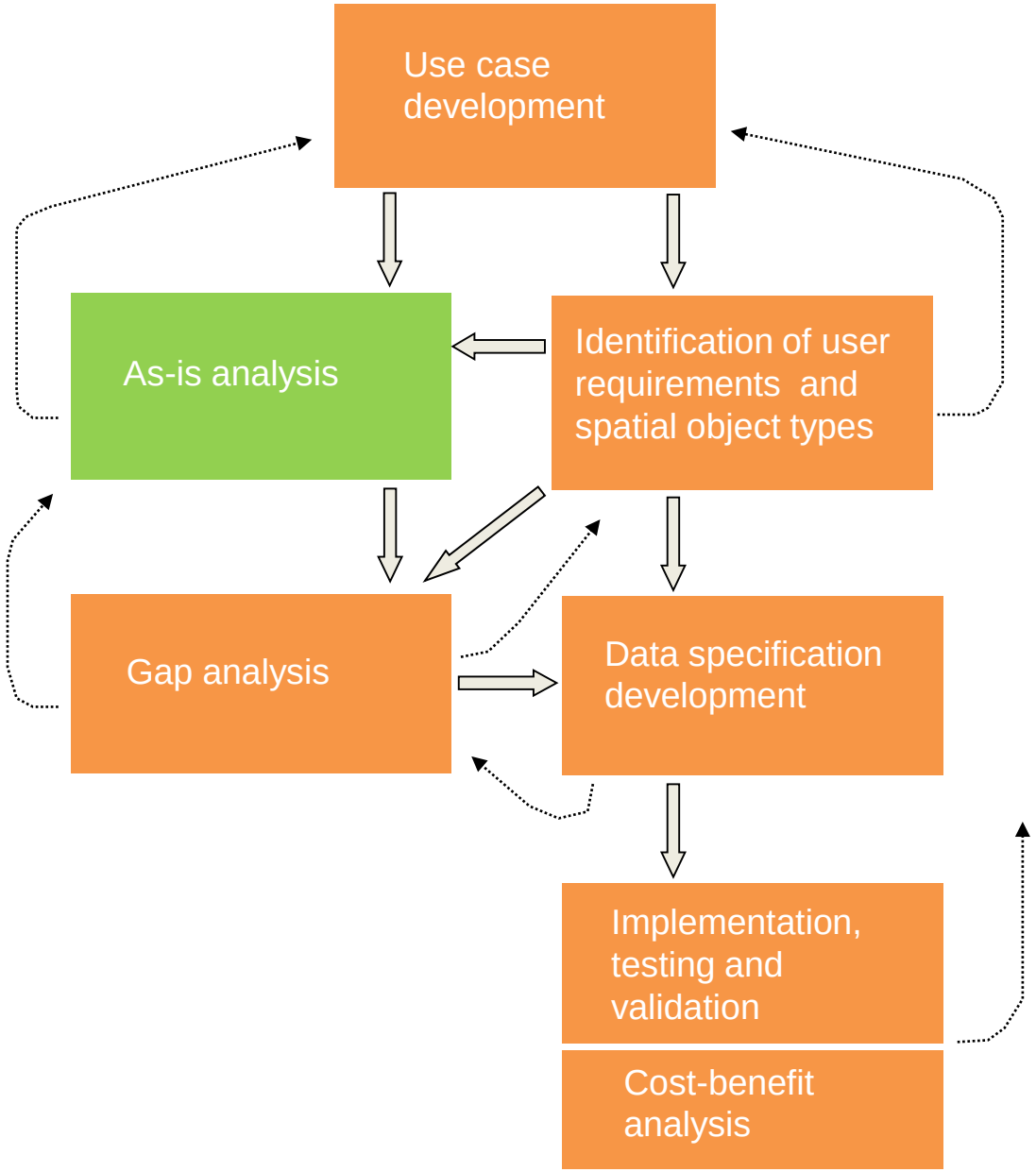
- European environmental policies
- User requirements survey
- SDIC/LMO reference material
- EU-funded initiatives and projects



Step 2

Identify requirements on:

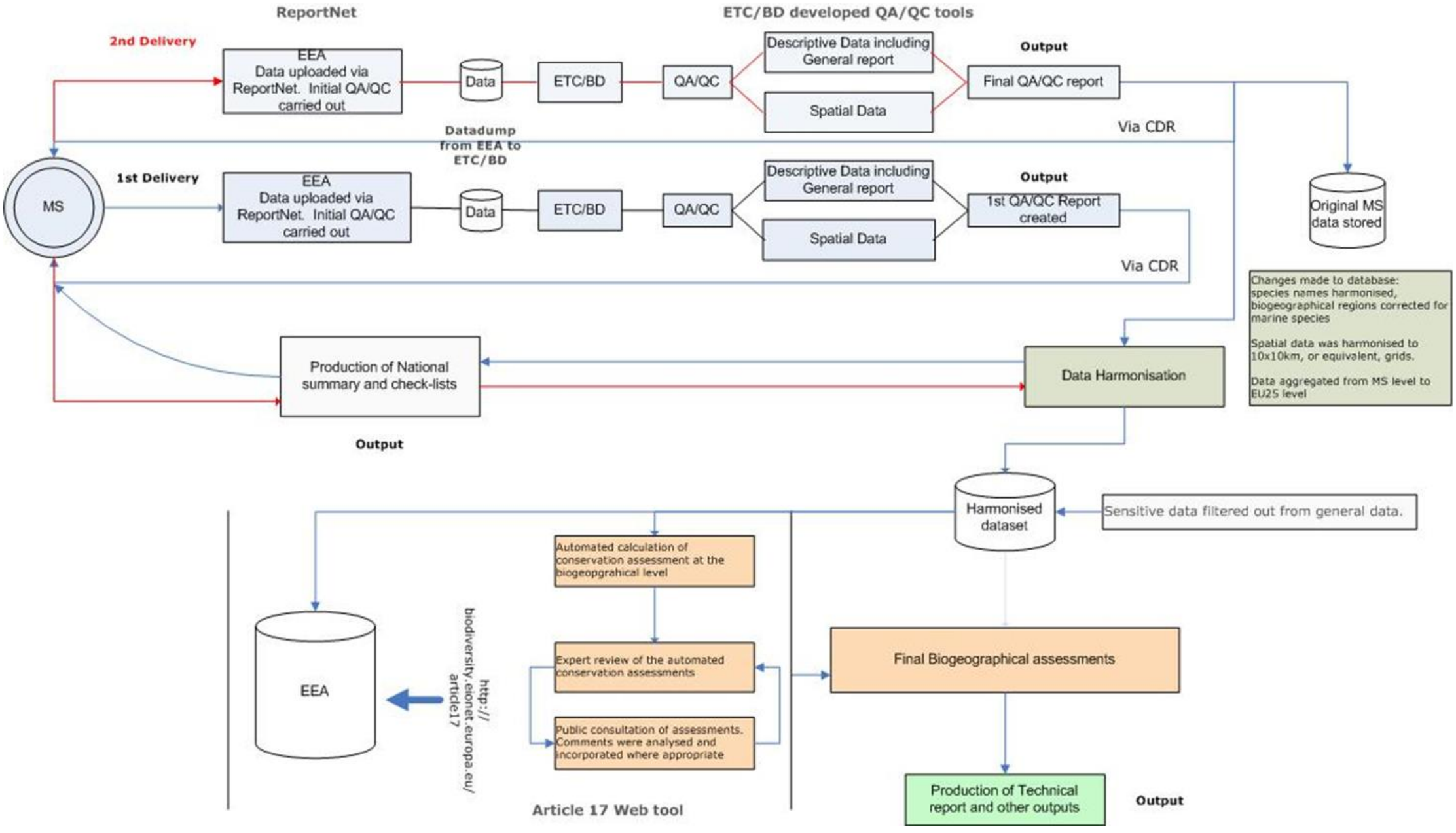
- the data content
- metadata, data quality, portrayal and other elements of the data specification

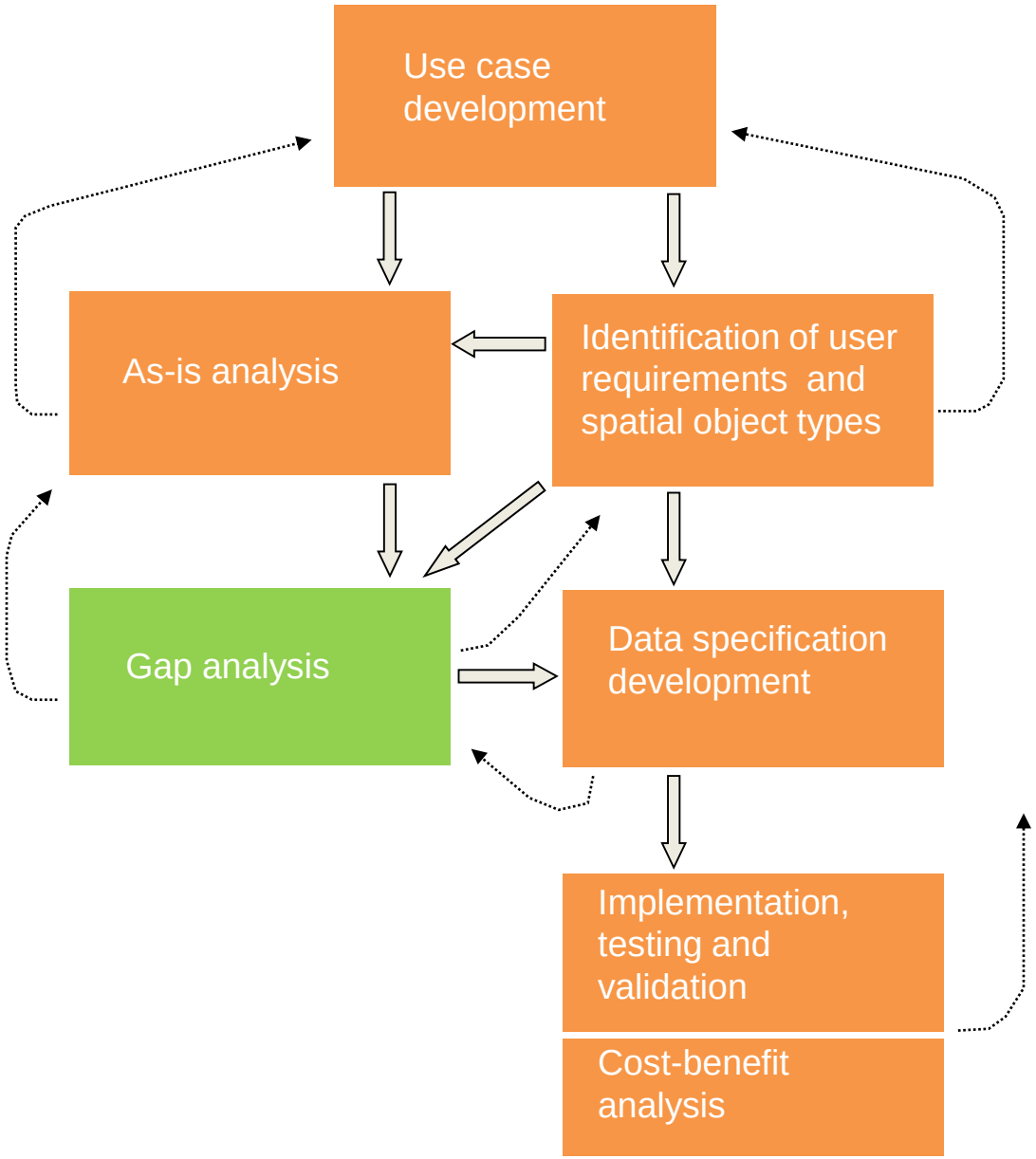


Step 3

Analyse the current situation regarding spatial data sets for the theme, based on:

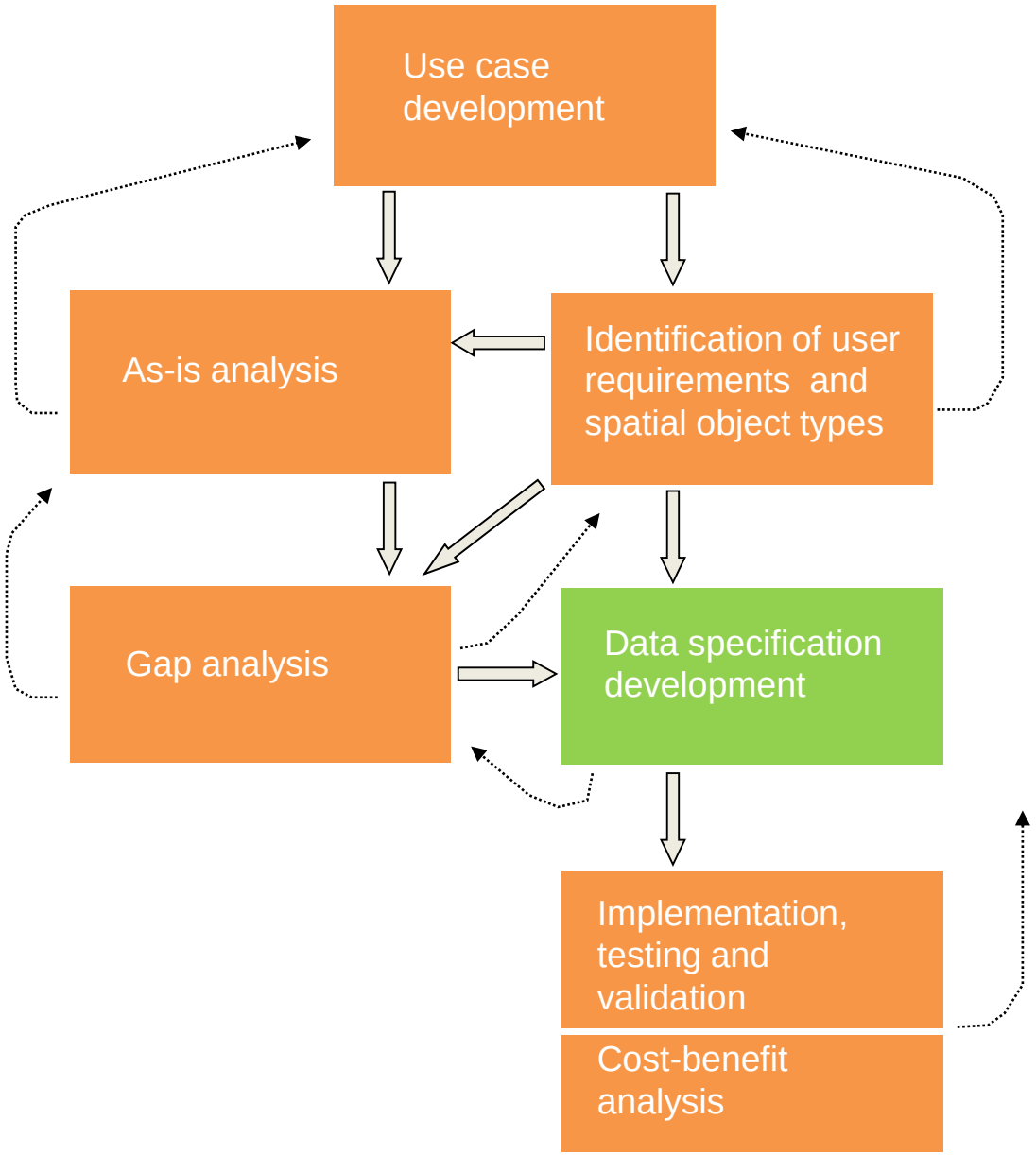
- the reference material
- existing internationally standardised data specifications
- expertise/field experts


ReportNet
ETC/BD developed QA/QC tools
2nd Delivery
EEA
Data uploaded via ReportNet. Initial QA/QC carried out
Data
ETC/BD
QA/QC
Descriptive Data including General report
Spatial Data
Output
Final QA/QC report
Via CDR
Datadump from EEA to ETC/BD
MS
1st Delivery
EEA
Data uploaded via ReportNet. Initial QA/QC carried out
Data
ETC/BD
QA/QC
Descriptive Data including General report
Spatial Data
Output
1st QA/QC Report created
Via CDR
Original MS data stored
Changes made to database: species names harmonised, biogeographical regions corrected for marine species
Spatial data was harmonised to 10x10km, or equivalent, grids.
Data aggregated from MS level to EU25 level
Production of National summary and check-lists
Output
Data Harmonisation
Harmonised dataset
Sensitive data filtered out from general data.
Automated calculation of conservation assessment at the biogeopgrahical level
Expert review of the automated conservation assessments
Public consultation of assessments. Comments were analysed and incorporated where appropriate
Article 17 Web tool
http://biodiversity.eionet.europa.eu/article17
EEA
Final Biogeographical assessments
Production of Technical report and other outputs
Output



Step 4

Compare identified data sources with identified user requirements

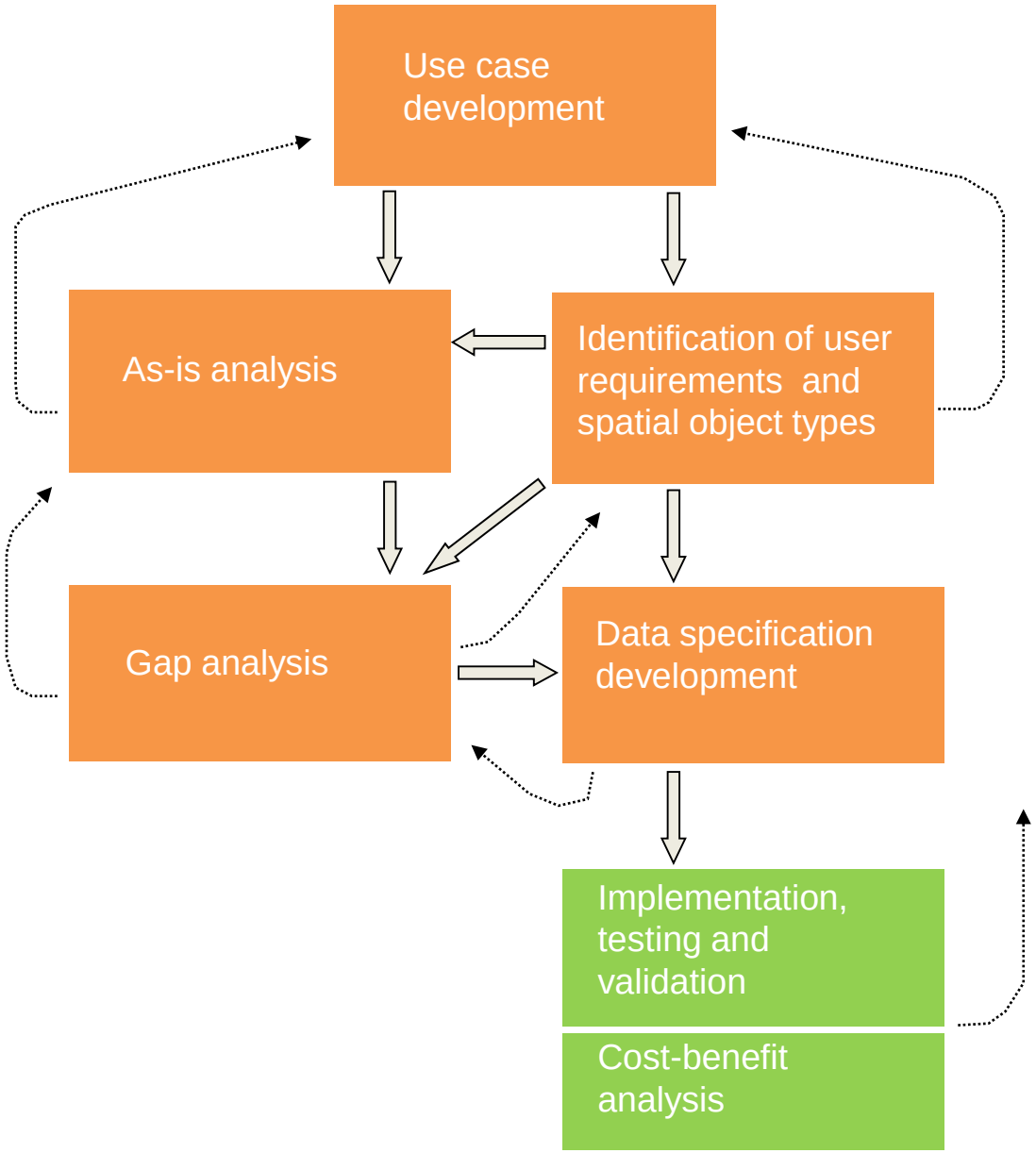


Step 5

The data specifications must be designed to ensure easy mapping between existing data and the harmonised data specification.

Consider:

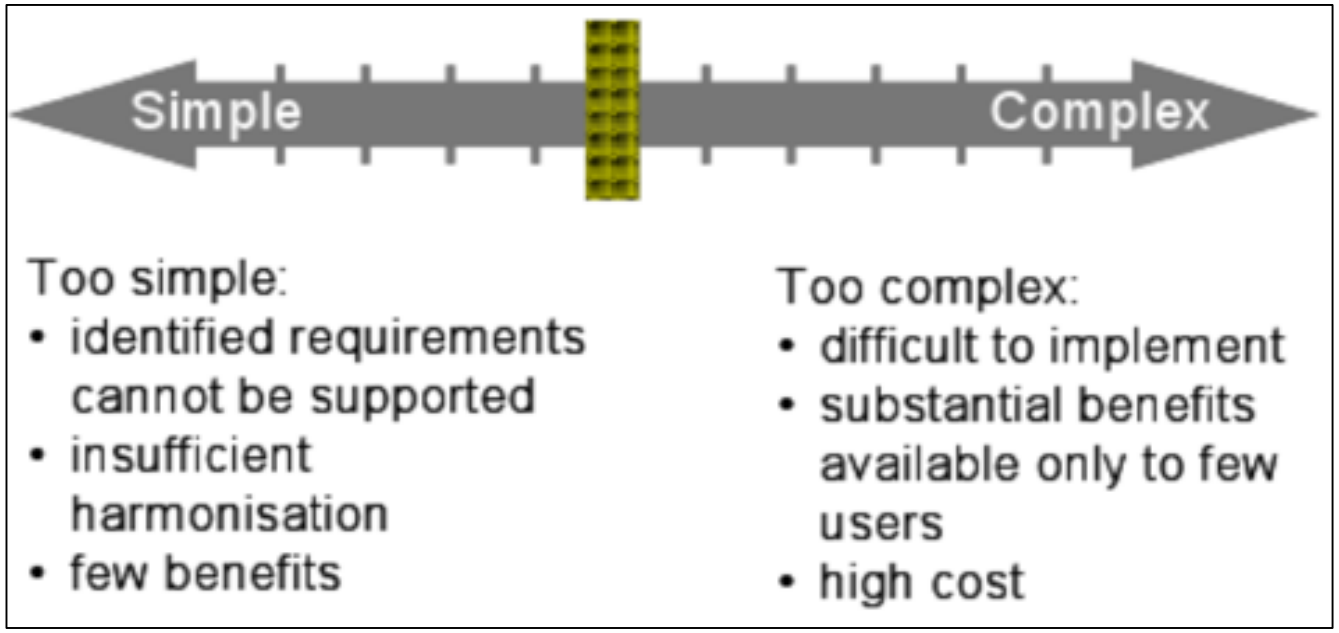
- No excessive costs
- No collection of new data!



Step 6-7

- review process
- test under real world conditions
- analyse costs and benefits

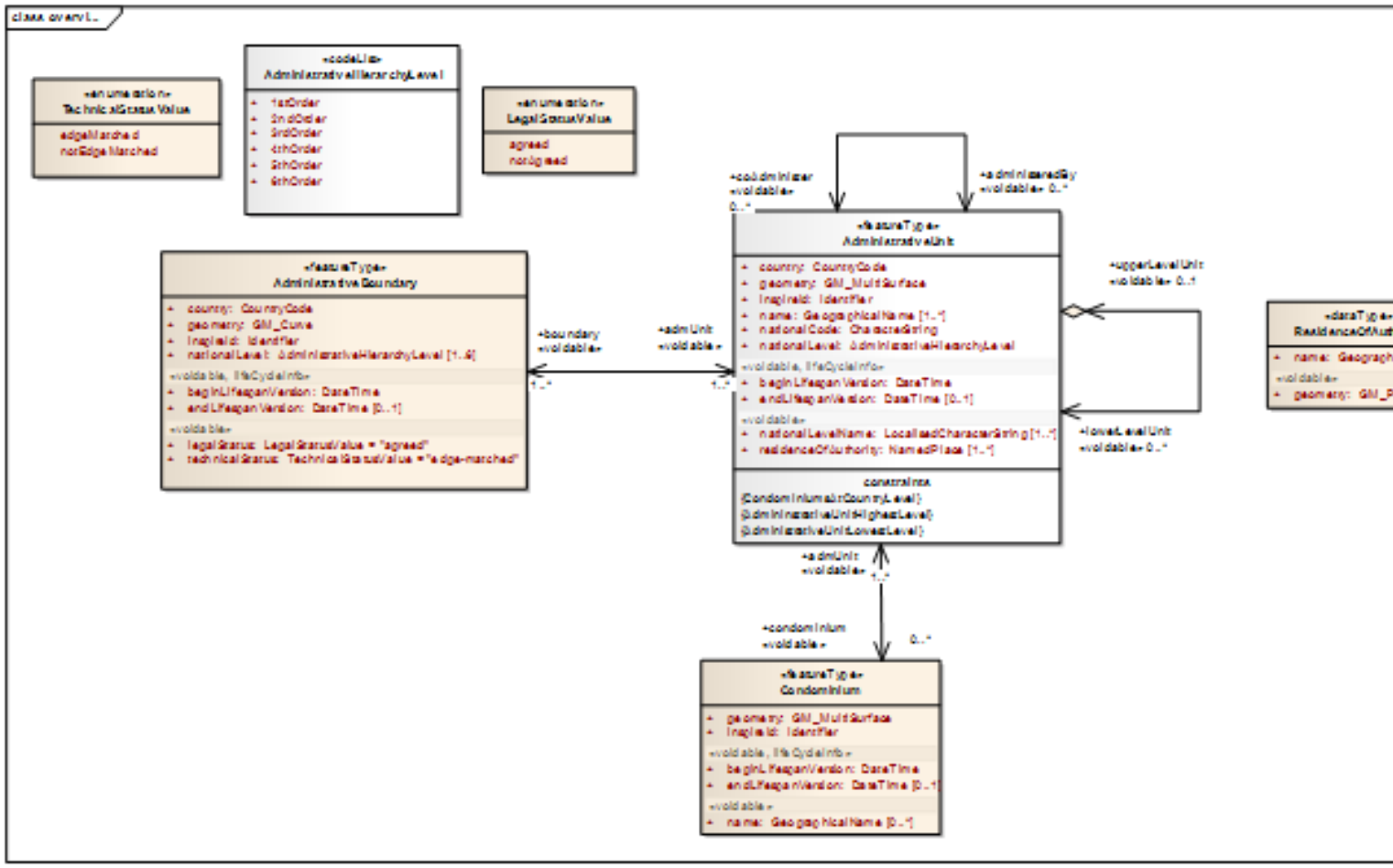
Final round of harmonisation



Requires:

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

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



Data Specification example

Cadastral Parcels (CP) – the INSPIRE approach

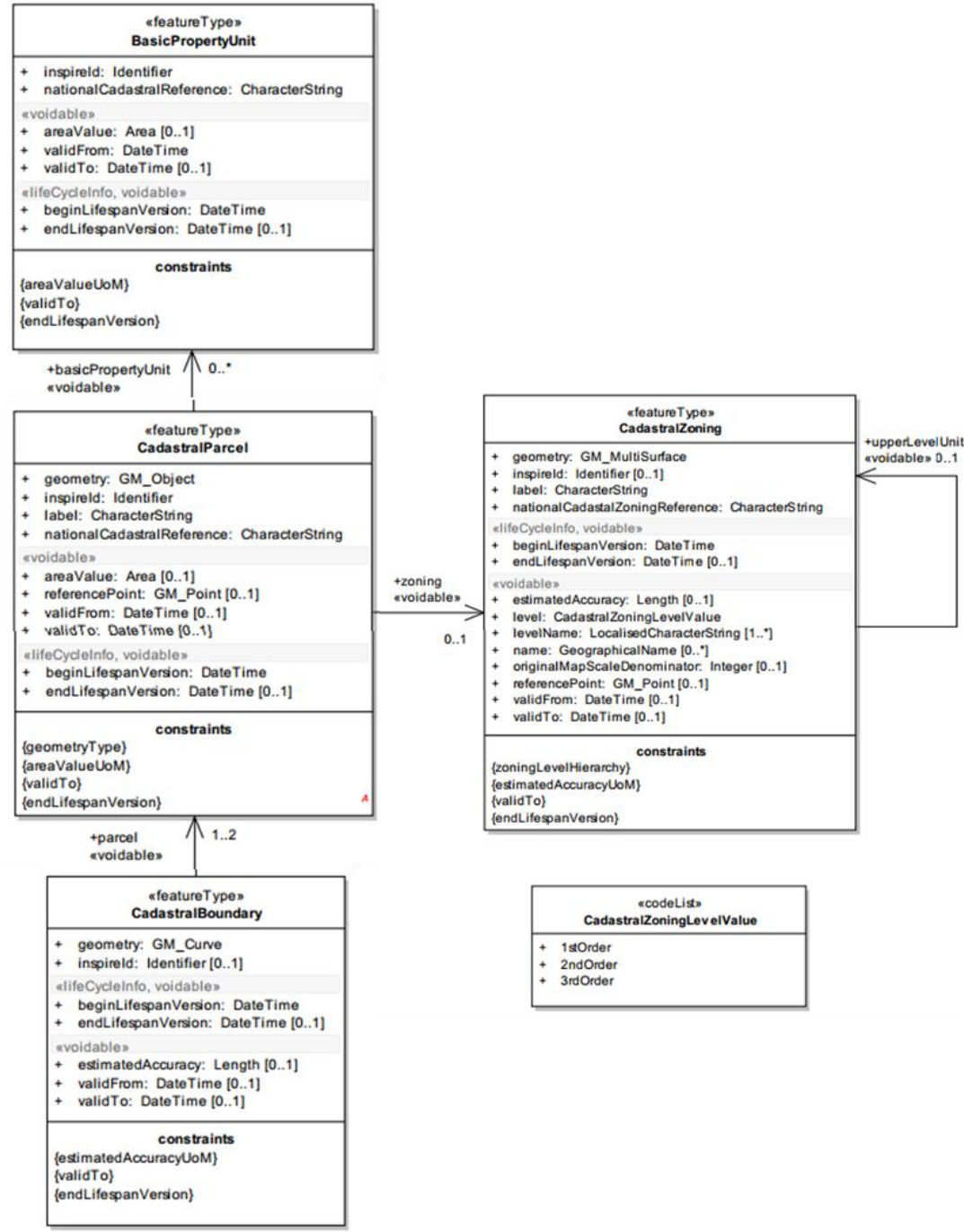
- The scope of the cadastral information in the INSPIRE context is limited to the **geographic side** of the cadastral information systems (land administration)

- 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.

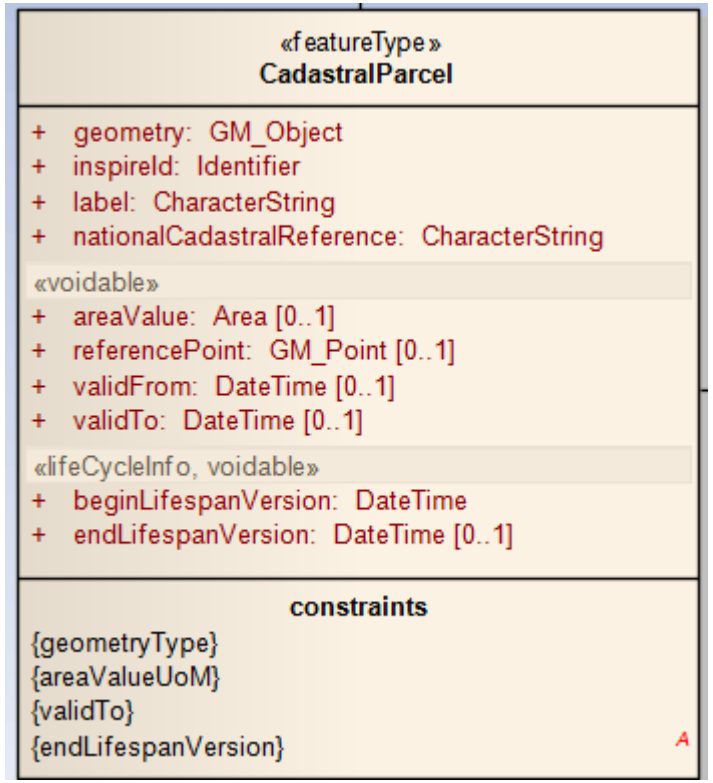
- All countries run a register
  - Usually a partition of the country with exceptions
- Basic unit of the system is the parcel
- The cadastral parcels should be, as much as possible, single areas of Earth surface (land and/or water) under homogenous real property rights and unique ownership, where real property rights and ownership are defined by national laws.

- 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


«featureType»
BasicPropertyUnit
+ inspireId: Identifier
+ nationalCadastralReference: CharacterString
«voidable»
+ areaValue: Area [0..1]
+ validFrom: DateTime
+ validTo: DateTime [0..1]
«lifeCycleInfo, voidable»
+ beginLifespanVersion: DateTime
+ endLifespanVersion: DateTime [0..1]
constraints
{areaValueUoM}
{validTo}
{endLifespanVersion}
+basicPropertyUnit
«voidable»
0..*
«featureType»
CadastralParcel
+ geometry: GM_Object
+ inspireId: Identifier
+ label: CharacterString
+ nationalCadastralReference: CharacterString
«voidable»
+ areaValue: Area [0..1]
+ referencePoint: GM_Point [0..1]
+ validFrom: DateTime [0..1]
+ validTo: DateTime [0..1]
«lifeCycleInfo, voidable»
+ beginLifespanVersion: DateTime
+ endLifespanVersion: DateTime [0..1]
constraints
{geometryType}
{areaValueUoM}
{validTo}
{endLifespanVersion}
+zoning
«voidable»
0..1
«featureType»
CadastralZoning
+ geometry: GM_MultiSurface
+ inspireId: Identifier [0..1]
+ label: CharacterString
+ nationalCadastalZoningReference: CharacterString
«lifeCycleInfo, voidable»
+ beginLifespanVersion: DateTime
+ endLifespanVersion: DateTime [0..1]
«voidable»
+ estimatedAccuracy: Length [0..1]
+ level: CadastralZoningLevelValue
+ levelName: LocalisedCharacterString [1..*]
+ name: GeographicalName [0..*]
+ originalMapScaleDenominator: Integer [0..1]
+ referencePoint: GM_Point [0..1]
+ validFrom: DateTime [0..1]
+ validTo: DateTime [0..1]
constraints
{zoningLevelHierarchy}
{estimatedAccuracyUoM}
{validTo}
{endLifespanVersion}
+upperLevelUnit
«voidable» 0..1
+parcel
«voidable»
1..2
«featureType»
CadastralBoundary
+ geometry: GM_Curve
+ inspireId: Identifier [0..1]
«lifeCycleInfo, voidable»
+ beginLifespanVersion: DateTime
+ endLifespanVersion: DateTime [0..1]
«voidable»
+ estimatedAccuracy: Length [0..1]
+ validFrom: DateTime [0..1]
+ validTo: DateTime [0..1]
constraints
{estimatedAccuracyUoM}
{validTo}
{endLifespanVersion}
«codeList»
CadastralZoningLevelValue
+ 1stOrder
+ 2ndOrder
+ 3rdOrder

- CadastralParcel (mandatory)
- CadastralZoning (auxiliary)
- CadastralBoundary (auxiliary)
- BasicPropertyUnit (auxiliary)

Core Profile



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".

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

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.

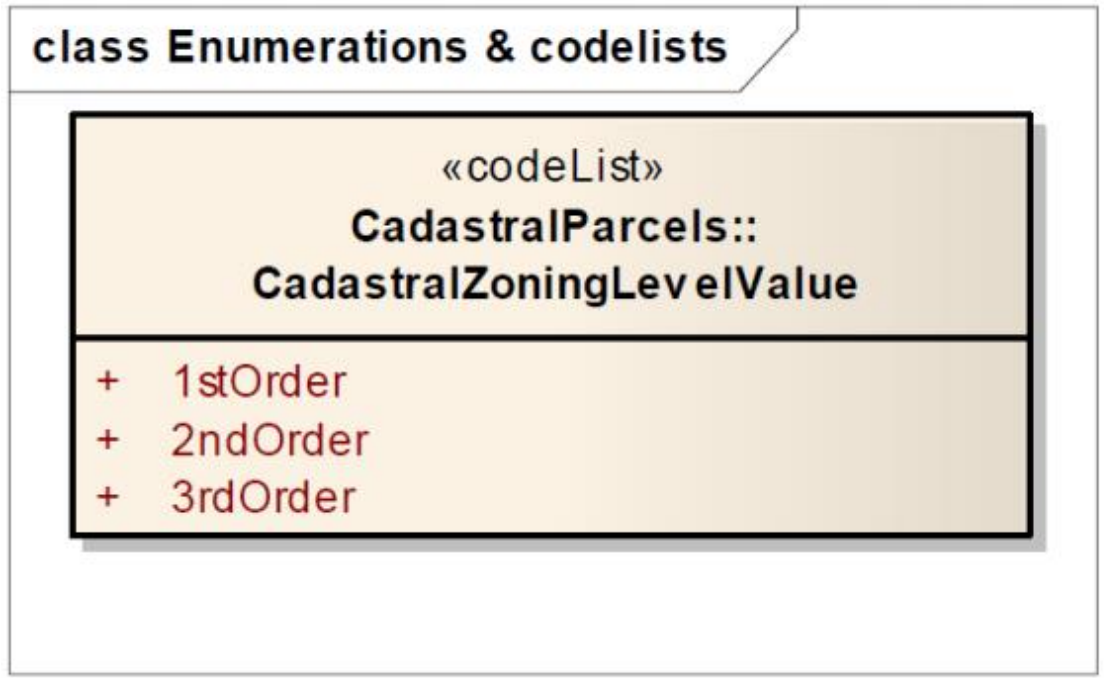
- 0-, 1-,2-,2,5 dimensional geometries

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.

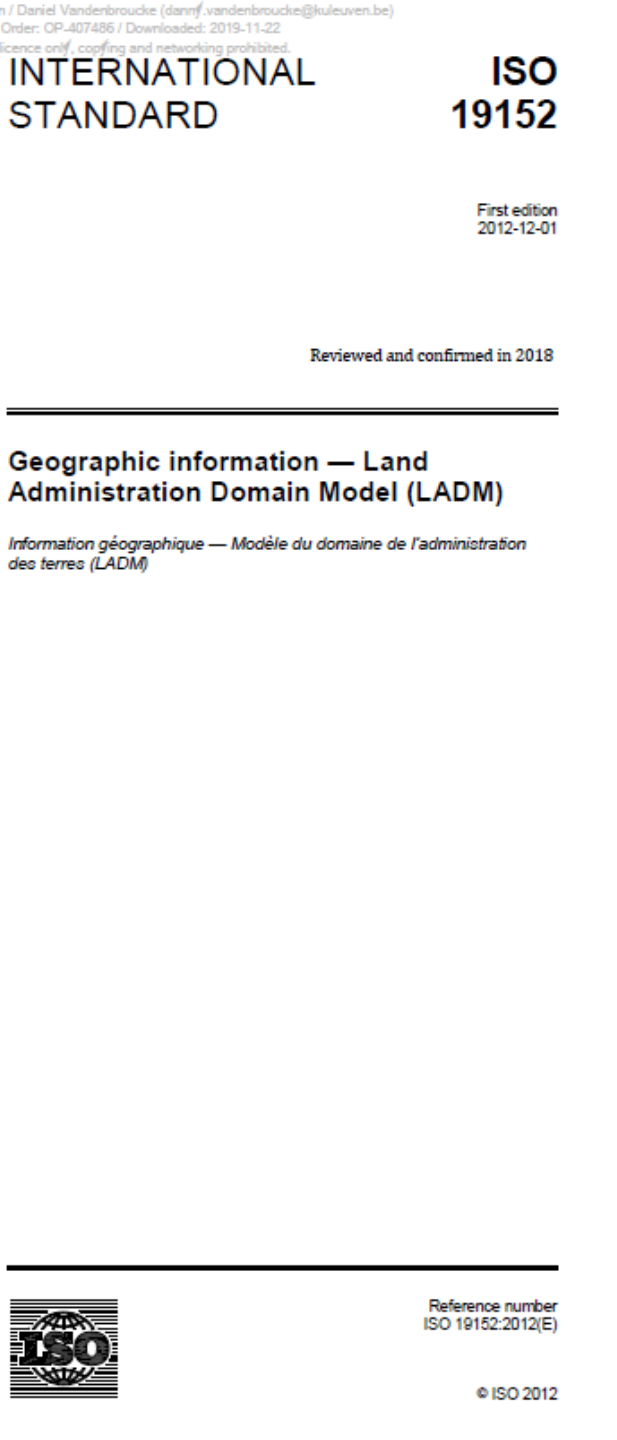
Recommendation 6 There should be no topological overlaps between cadastral parcels.

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



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

- Applied in Mozambique by DINAT
  - Based on the ISO 19100 series of standards and 19131 in particular
  - Follows a regular schema of an ISO standard
  - Relationship with standards explained in annex



INTERNATIONAL STANDARD

ISO 19152

First edition
2012-12-01

Reviewed and confirmed in 2018

Geographic information — Land Administration Domain Model (LADM)

Information géographique — Modèle du domaine de l'administration des terres (LADM)

Reference number
ISO 19152:2012(E)

- Major packages of the data model
  - Party
  - Administrative
  - Spatial unit
  - Surveying and representation

- Exercise: analyzing the standard (model)