Land Cover spatial datasets harmonization in Portugal using HALE

Teresa Zuna, Alexandra Fonseca,
Danilo Furtado, Ana Luísa Gomes,
André Serronha, Paulo Patrício

Introduction

DGT is the entity responsible for the operational coordination of the National System for Geographic Information (SNIG) and the National Contact Point (NCP) for the INSPIRE Directive.

DGT has participated in different European projects associated to Spatial Data Infrastructures especially focused on the harmonization of spatial data according to INSPIRE Directive (e.g. HUMBOLDT, GIS4EU, NatureSDI*plus*, HELM, eENV*plus*, EAGLE 6).

Introduction

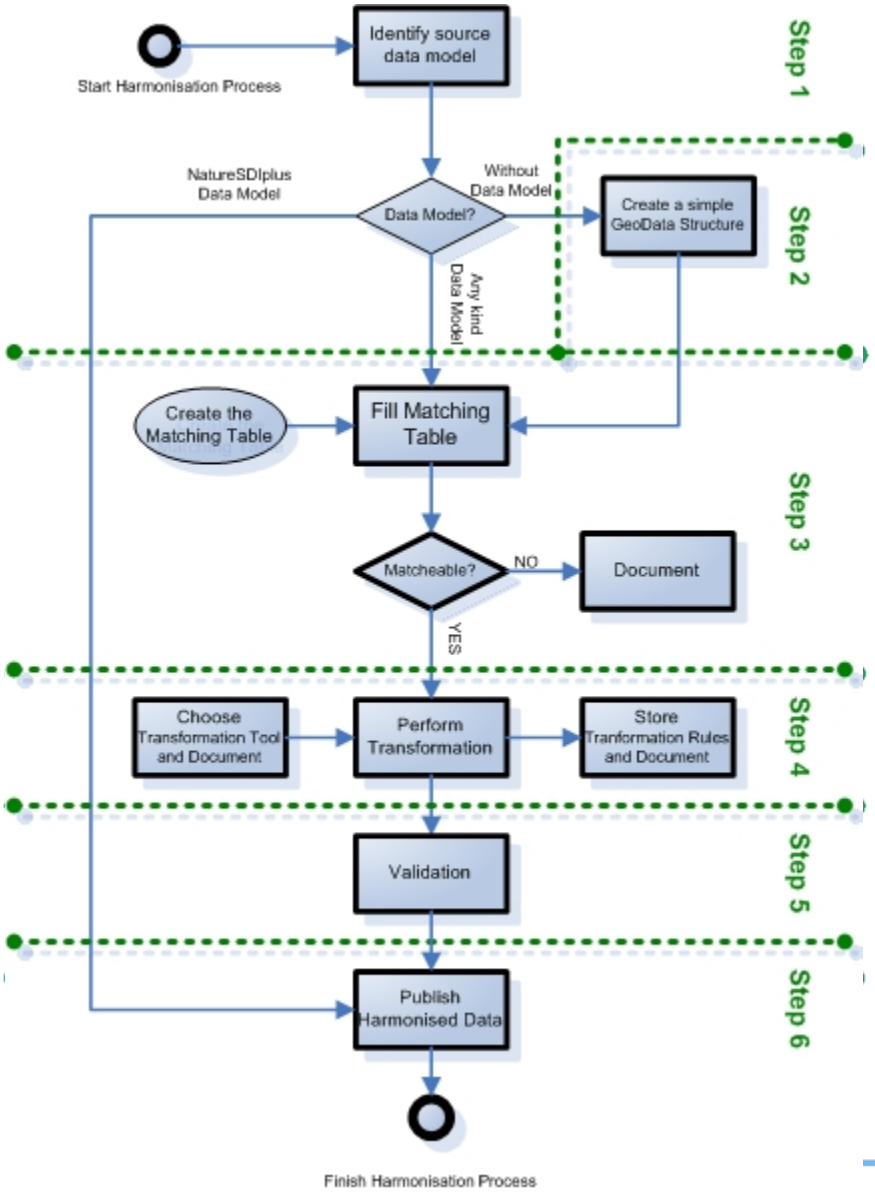
DGT is also responsible for the production of some thematic maps, namely the Land Cover Map (*Carta de Ocupação do Solo - COS*).

Being COS one of the Spatial Data Sets (SDS) produced by DGT with greater relevance to the development of environmental management and planning studies in Portugal and following the participation of DGT in the EAGLE 6 project, it was decided to proceed with the harmonization of COS according to the specifications of INSPIRE Directive.

Land Cover Map 2010 (COS2010)

Harmonization process

The harmonization process involves the analysis of the data models, the filling of the matching table, the transformation of the SDS into the target schema, the validation and the publishing of the SDS through geowebservices.



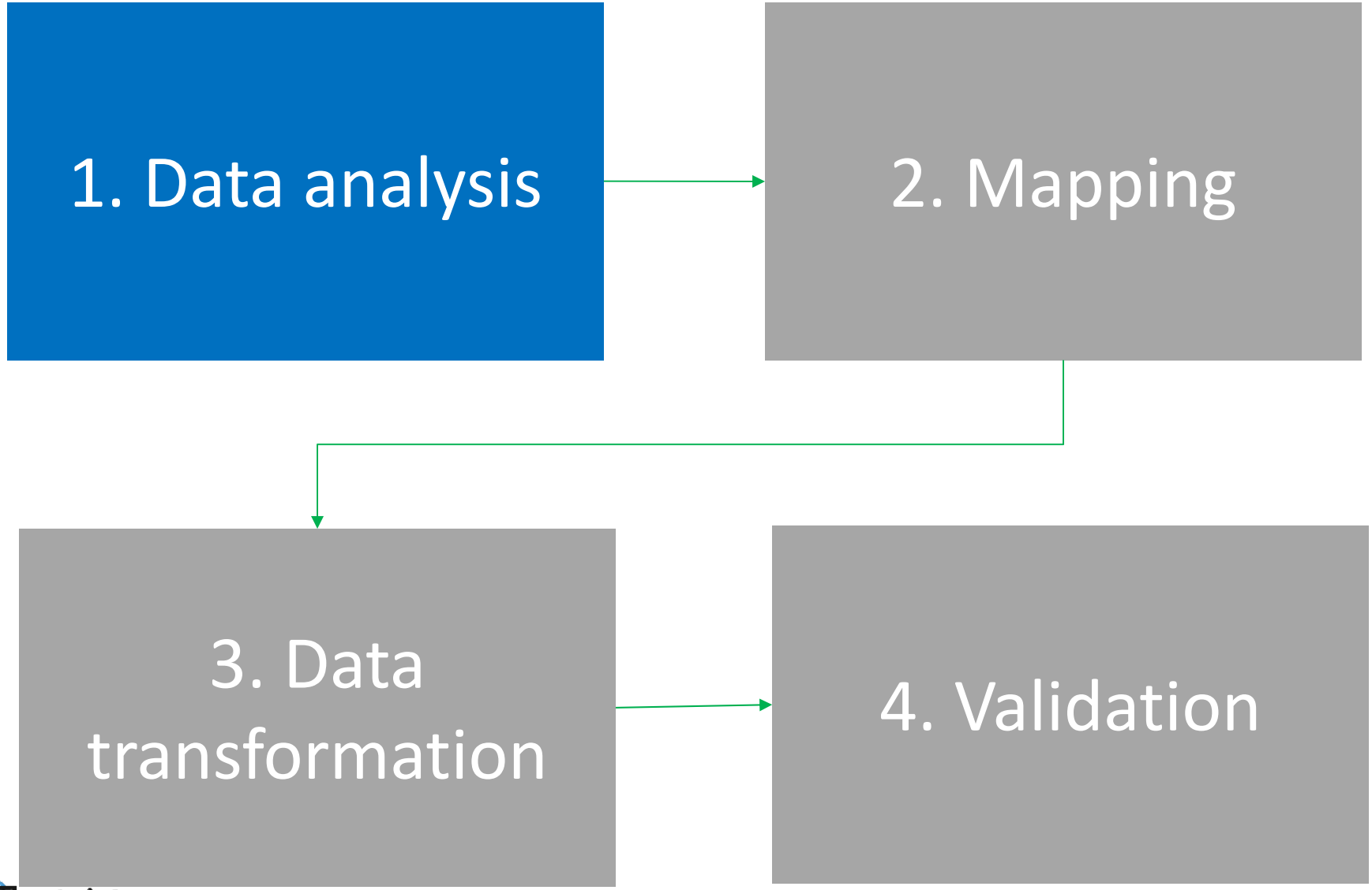
Data analysis

Mapping

Data transformation

Validation

Harmonized data publication


1. Data analysis
2. Mapping
3. Data transformation
4. Validation

1. Data analysis

Interpretation of source schema – COS2010

- Data format
- Spatial data representation
- Attributes
- Coordinate Reference System
- Metadata

1. Data analysis

The Land Cover Map 2010 (COS 2010) is a thematic map that aims to characterize in great detail the land cover and land use in mainland Portugal.

COS2010 presents an hierarchical structure with five levels, corresponding to 225 classes at the most detailed level (N5). This nomenclature is consistent with the nomenclature of the Corine Land Cover at the first three levels.

Data model	Vectorial
Data structure	Polygons
CRS	ETRS89 (European Terrestrial Reference System 1989) PT-TM06
Minimum Mapping Unit (UMC)	1 hectare
Minimum distance between lines	20 meters
Nomenclature	Hierarchical classification with five levels and 225 classes

1. Data analysis

Harmonization involves the transformation of the source data (source schema) in the data model described by the Directive (target schema) in an open format and oriented to services. This requires:

INSPIRE target schema

- INSPIRE theme
- INSPIRE documents
  - General Conceptual Model
  - Data Specifications


INSPIRE
Infrastructure for Spatial Information in Europe

D2.8.II.2 Data Specification on Land cover – Draft Technical Guidelines

Title	D2.8.II.2 INSPIRE Data Specification on *Land cover* – Draft Technical Guidelines
Creator	INSPIRE Thematic Working Group *Land cover*
Date	2013-02-04
Subject	INSPIRE Data Specification for the spatial data theme *Land cover*
Publisher	INSPIRE Thematic Working Group *Land cover*
Type	Text
Description	This document describes the INSPIRE Data Specification for the spatial data theme *Land cover*. This version (version 3, release candidate 3) reflects the content of the draft amendment to Commission Regulation (EU) No 1089/2010 for the Annex II+III spatial data themes as submitted to the INSPIRE Committee.
Contributor	Members of the INSPIRE Thematic Working Group *Land cover*
Format	Portable Document Format (pdf)
Source	
Rights	Public
Identifier	D2.8.II.2_v3.0rc3
Language	En
Relation	Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)
Coverage	Project duration

ANNEX 1

- Addresses
- Administrative units
- Cadastral parcels
- Coordinate reference systems
- Geographical grid systems
- Geographical names
- Hydrography
- Protected sites
- Transport networks

ANNEX 2

- Elevation
- Geology
- Land cover
- Orthoimagery

ANNEX 3

- Agricultural and aquaculture facilities
- Area management / restriction / regulation zones & reporting units
- Atmospheric conditions
- Bio-geographical regions
- Buildings
- Energy Resources
- Environmental monitoring Facilities
- Habitats and biotopes
- Human health and safety
- Land use
- Meteorological geographical features
- Mineral Resources
- Natural risk zones
- Oceanographic geographical features
- Population distribution and demography
- Production and industrial facilities
- Sea regions
- Soil
- Species distribution
- Statistical units
- Utility and governmental services

1. Data analysis

(Target data model)

- Theme INSPIRE – II.2 Land Cover

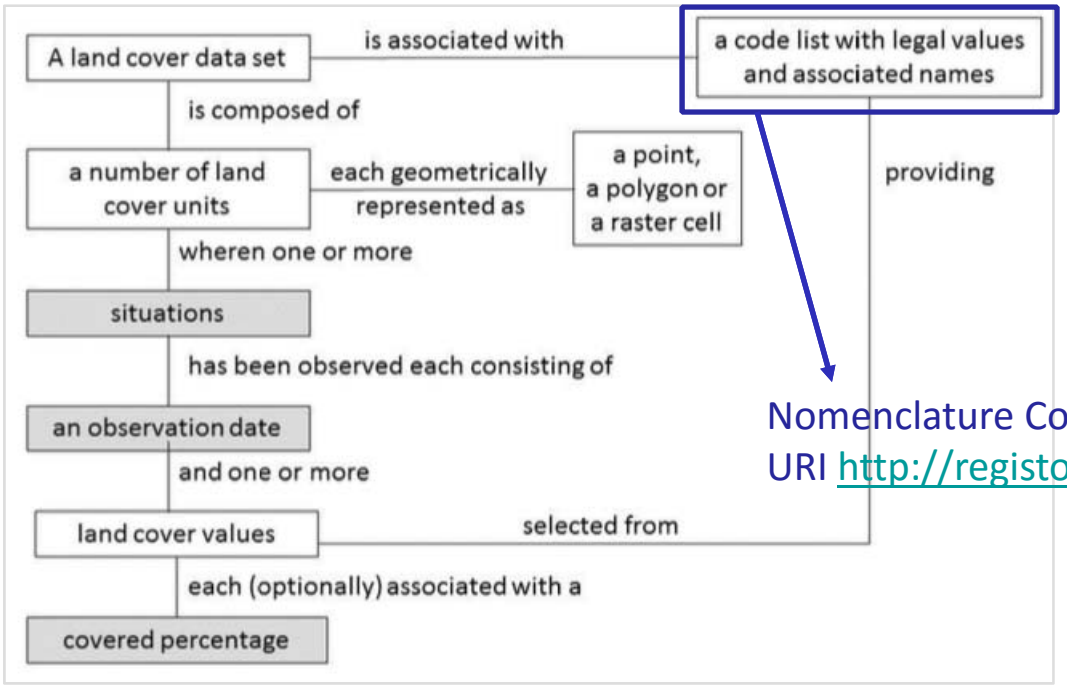
  – D2.5. INSPIRE Generic Conceptual Model versão 3.4rc3

  – D2.8.II.2 *Data Specification on Land Cover – Technical Guidelines*

  – *Application schema (LandCoverVector.XSD)*

  – UML diagram

  – Objects catalogue



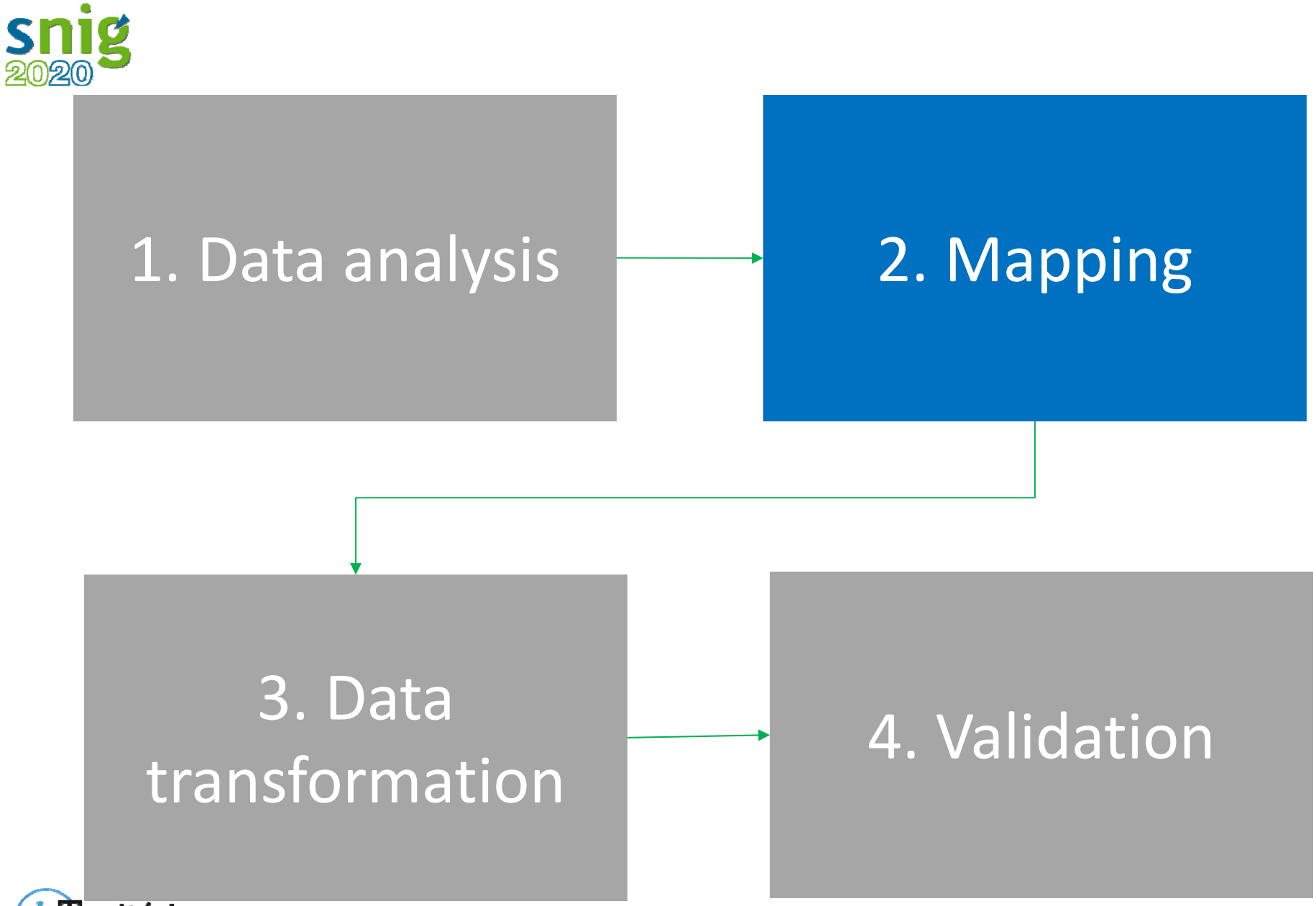
Nomenclature Code List
URI http://registo.igeo.pt/

1. Data analysis

LandCoverNomenclature

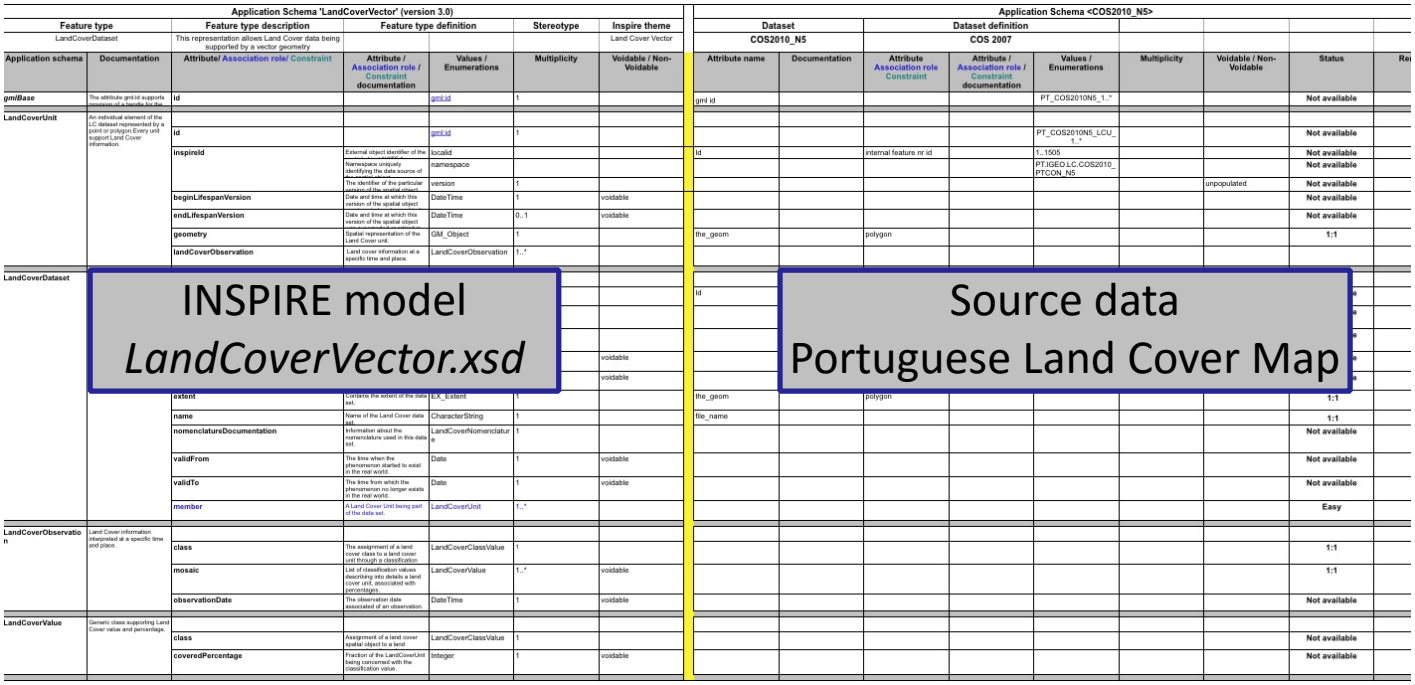
Code list - is a nomenclature of land cover classes, where each class is represented by a code and a description.

- Only one nomenclature for COS 2010
- The values are managed outside the application schema LandCoverVector.xsd
- Hierarchical list (parent value)
- INSPIRE register (http://inspire.ec.europa.eu/codelist_register/codelist)
- URI, legend: http://registo.igeo.pt/listadecodigo/CartaOcupacaoSoloValue


1. Data analysis
2. Mapping
3. Data transformation
4. Validation

2. Mapping

Correpondence tables, known as matching tables are used to establish correspondence between the attributes in the source data model (source schema) and the target data model structure (target schema). The matching table identifies and describes the classes, attributes, enumerations, code lists and associations between classes of both models.

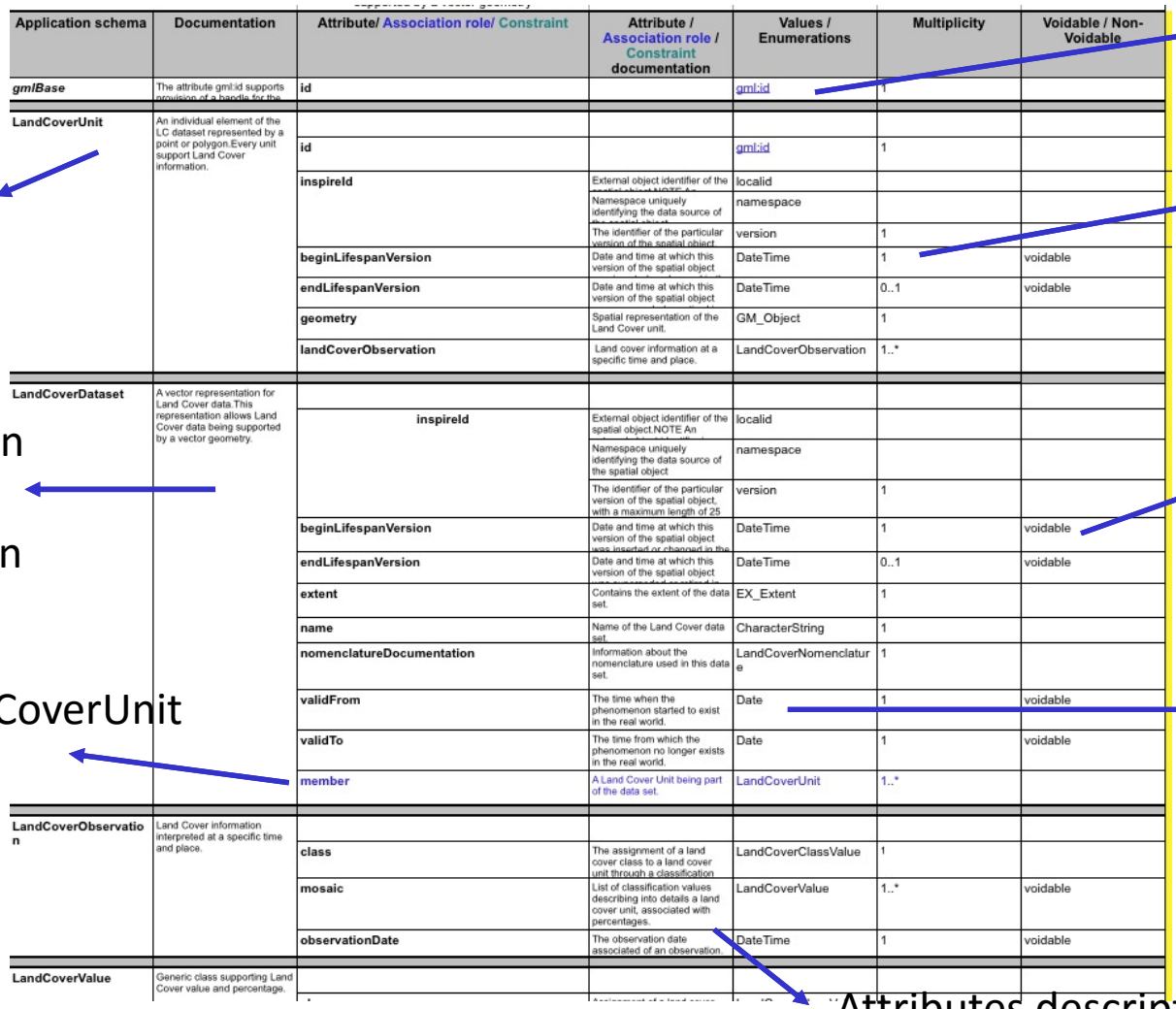


Application Schema 'LandCoverVector' (version 3.0)							Application Schema <COS2010_N5>							
Feature type		Feature type description	Feature type definition		Stereotype	Inspire theme	Dataset		Dataset definition					
LandCoverDataset		This representation allows Land Cover data being supported by a vector geometry				Land Cover Vector	COS2010_N5		COS 2007					
Application schema	Documentation	Attribute/ Association role/ Constraint	Attribute / Association role / Constraint documentation	Values / Enumerations	Multiplicity	Voidable / Non-Voidable	Attribute name	Documentation	Attribute Association role Constraint	Attribute / Association role / Constraint documentation	Values / Enumerations	Multiplicity	Voidable / Non-Voidable	Status
gmlBase	The attribute gml:id supports provision of a handle for the	id		gml:id	1		gml id				PT_COS2010N5_1..*			Not available
LandCoverUnit	An individual element of the LC dataset represented by a point or polygon.Every unit support Land Cover information.													
		id		gml:id	1						PT_COS2010N5_LCU_1..*			Not available
		inspireId	External object identifier of the	localId			Id		internal feature nr id		1..1505			Not available
			Namespace uniquely identifying the data source of	namespace							PT.IGEO.LC.COS2010_PTCON_N5			Not available
			The identifier of the particular	version	1								unpopulated	Not available
		beginLifespanVersion	Date and time at which this version of the spatial object	DateTime	1	voidable								Not available
		endLifespanVersion	Date and time at which this version of the spatial object	DateTime	0..1	voidable								Not available
		geometry	Spatial representation of the Land Cover unit.	GM_Object	1		the_geom		polygon					1:1
		landCoverObservation	Land cover information at a specific time and place.	LandCoverObservation	1..*									
LandCoverDataset														
							Id							*
														*
														*
						voidable								*
						voidable								*
		extent	Contains the extent of the data set.	EX_Extent	1		the_geom		polygon					1:1
		name	Name of the Land Cover data set.	CharacterString	1		file_name							1:1
		nomenclatureDocumentation	Information about the nomenclature used in this data set.	LandCoverNomenclature	1									Not available
		validFrom	The time when the phenomenon started to exist in the real world.	Date	1	voidable								Not available
		validTo	The time from which the phenomenon no longer exists in the real world.	Date	1	voidable								Not available
		member	A Land Cover Unit being part of the data set.	LandCoverUnit	1..*									Easy
LandCoverObservation	Land Cover information interpreted at a specific time and place.													
		class	The assignment of a land cover class to a land cover unit through a classification	LandCoverClassValue	1									1:1
		mosaic	List of classification values describing into details a land cover unit, associated with percentages.	LandCoverValue	1..*	voidable								1:1
		observationDate	The observation date associated of an observation.	DateTime	1	voidable								Not available
LandCoverValue	Generic class supporting Land Cover value and percentage.													
		class	Assignment of a land cover spatial object to a land	LandCoverClassValue	1									Not available
		coveredPercentage	Fraction of the LandCoverUnit being concerned with the classification value.	Integer	1	voidable								Not available

Filling in the matching table

2. Mapping

The matching tables are used to document the harmonization process by completing various fields where aspects related to the harmonization process are recorded.



Application schema	Documentation	Attribute/ Association role/ Constraint	Attribute / Association role / Constraint documentation	Values / Enumerations	Multiplicity	Voidable / Non-Voidable
gmlBase	The attribute gml:id supports provision of a handle for the	id		gml:id	1	
LandCoverUnit	An individual element of the LC dataset represented by a point or polygon.Every unit support Land Cover information.					
		id		gml:id	1	
		inspireId	External object identifier of the	localid		
			Namespace uniquely identifying the data source of	namespace		
			The identifier of the particular version of the spatial object	version	1	
		beginLifespanVersion	Date and time at which this version of the spatial object	DateTime	1	voidable
		endLifespanVersion	Date and time at which this version of the spatial object	DateTime	0..1	voidable
		geometry	Spatial representation of the Land Cover unit.	GM_Object	1	
		landCoverObservation	Land cover information at a specific time and place.	LandCoverObservation	1..*	
LandCoverDataset	A vector representation for Land Cover data.This representation allows Land Cover data being supported by a vector geometry.					
		inspireId	External object identifier of the spatial object.NOTE An	localid		
			Namespace uniquely identifying the data source of the spatial object	namespace		
			The identifier of the particular version of the spatial object, with a maximum length of 25	version	1	
		beginLifespanVersion	Date and time at which this version of the spatial object	DateTime	1	voidable
		endLifespanVersion	Date and time at which this version of the spatial object	DateTime	0..1	voidable
		extent	Contains the extent of the data set.	EX_Extent	1	
		name	Name of the Land Cover data set.	CharacterString	1	
		nomenclatureDocumentation	Information about the nomenclature used in this data set.	LandCoverNomenclature	1	
		validFrom	The time when the phenomenon started to exist in the real world.	Date	1	voidable
		validTo	The time from which the phenomenon no longer exists in the real world.	Date	1	voidable
		member	A Land Cover Unit being part of the data set.	LandCoverUnit	1..*	
LandCoverObservation	Land Cover information interpreted at a specific time and place.					
		class	The assignment of a land cover class to a land cover unit through a classification	LandCoverClassValue	1	
		mosaic	List of classification values describing into details a land cover unit, associated with percentages.	LandCoverValue	1..*	voidable
		observationDate	The observation date associated of an observation.	DateTime	1	voidable
LandCoverValue	Generic class supporting Land Cover value and percentage.					

gml: unique identifier

Application schemas

Multiplicity

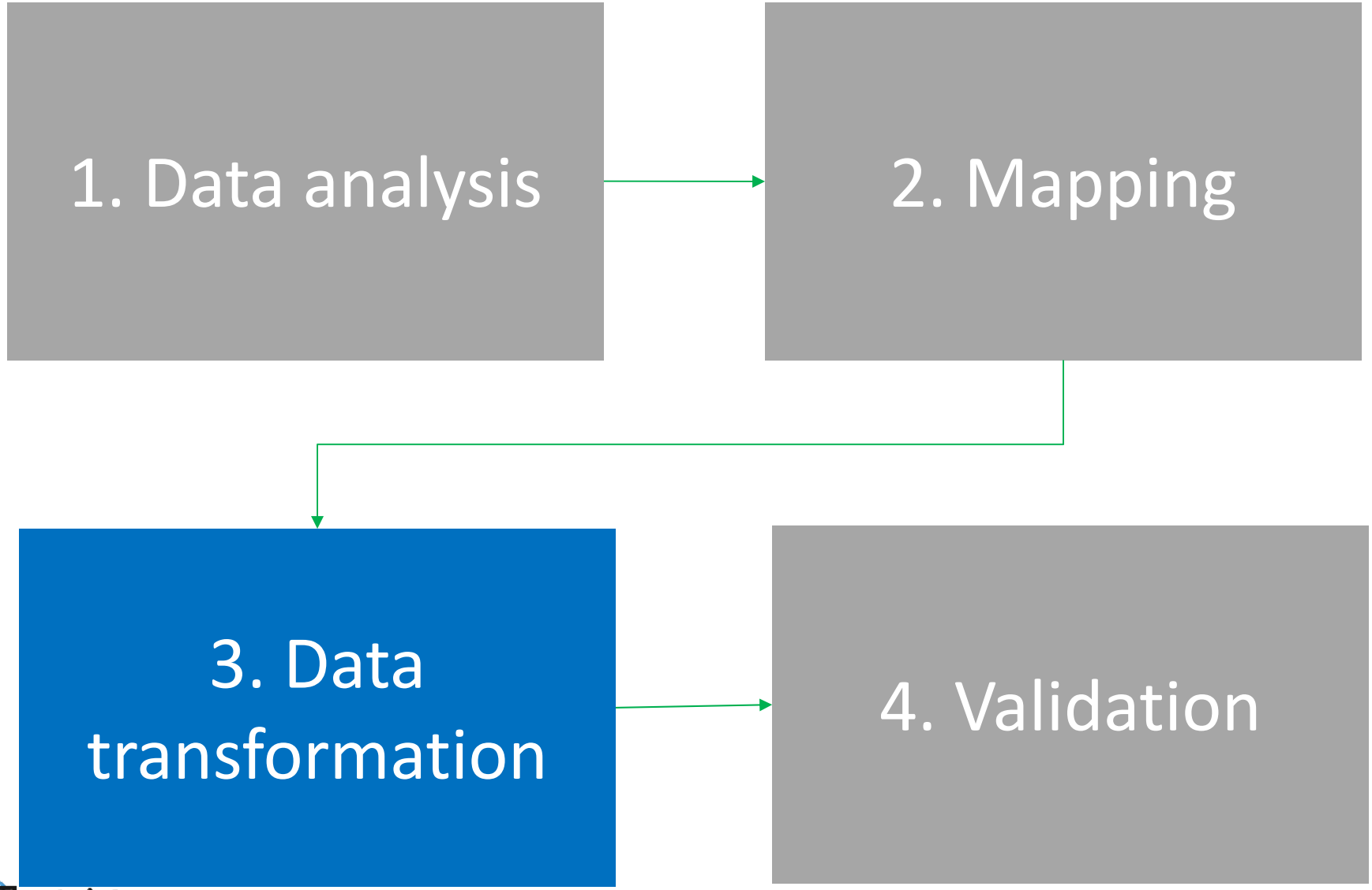
Application schema description

Voidable

Member - LandCoverUnit

Attributes

Attributes description

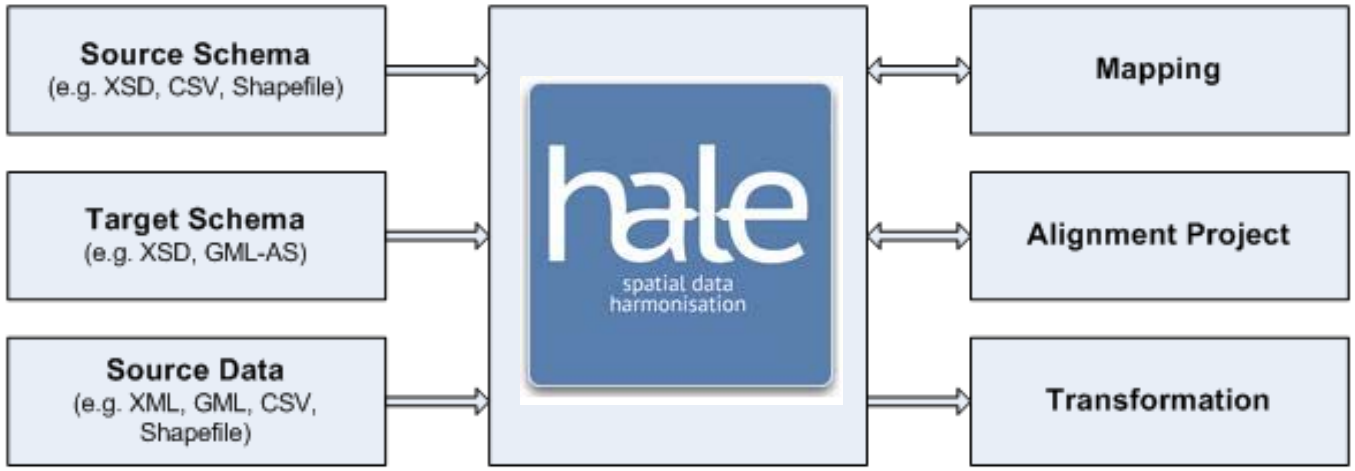

1. Data analysis
2. Mapping
3. Data transformation
4. Validation

3. Data transformation

HALE

The HUMBOLDT Alignment Editor (HALE) was developed by the European project HUMBOLDT (www.esdi-humboldt.org) aiming to contribute to the implementation of the INSPIRE Directive.

HALE is an open source tool, developed in order to support and facilitate SDS harmonization and transformation processes. It allows the user to establish relationships between schemas (source and target) and transform SDS automatically, based on the specifications defined in the application schemas.



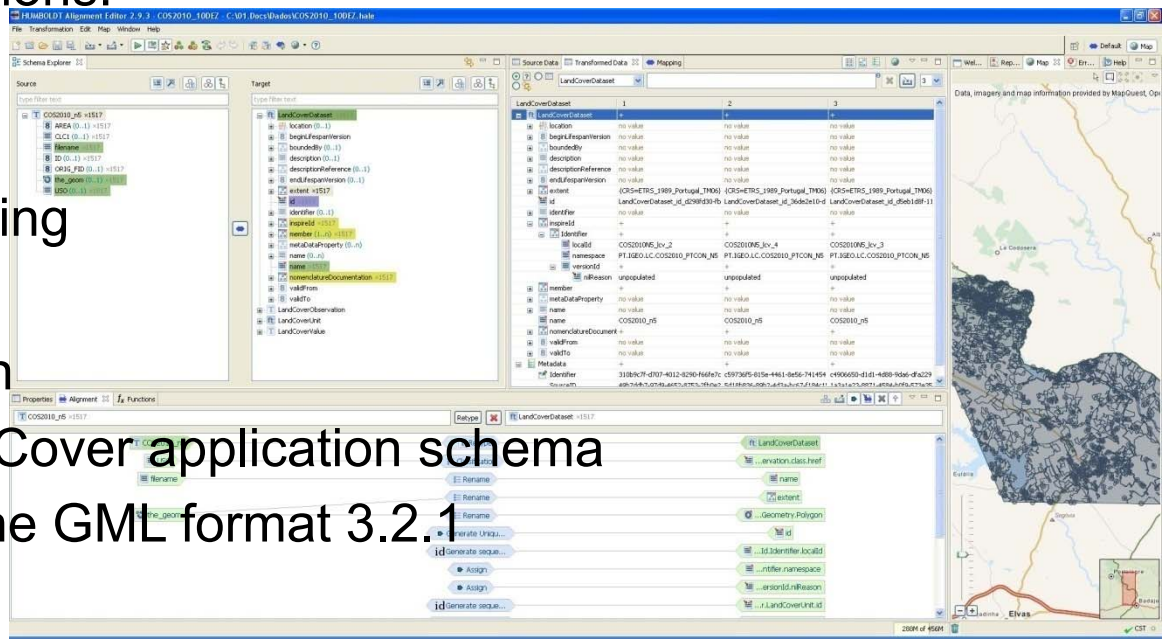
3. Data transformation

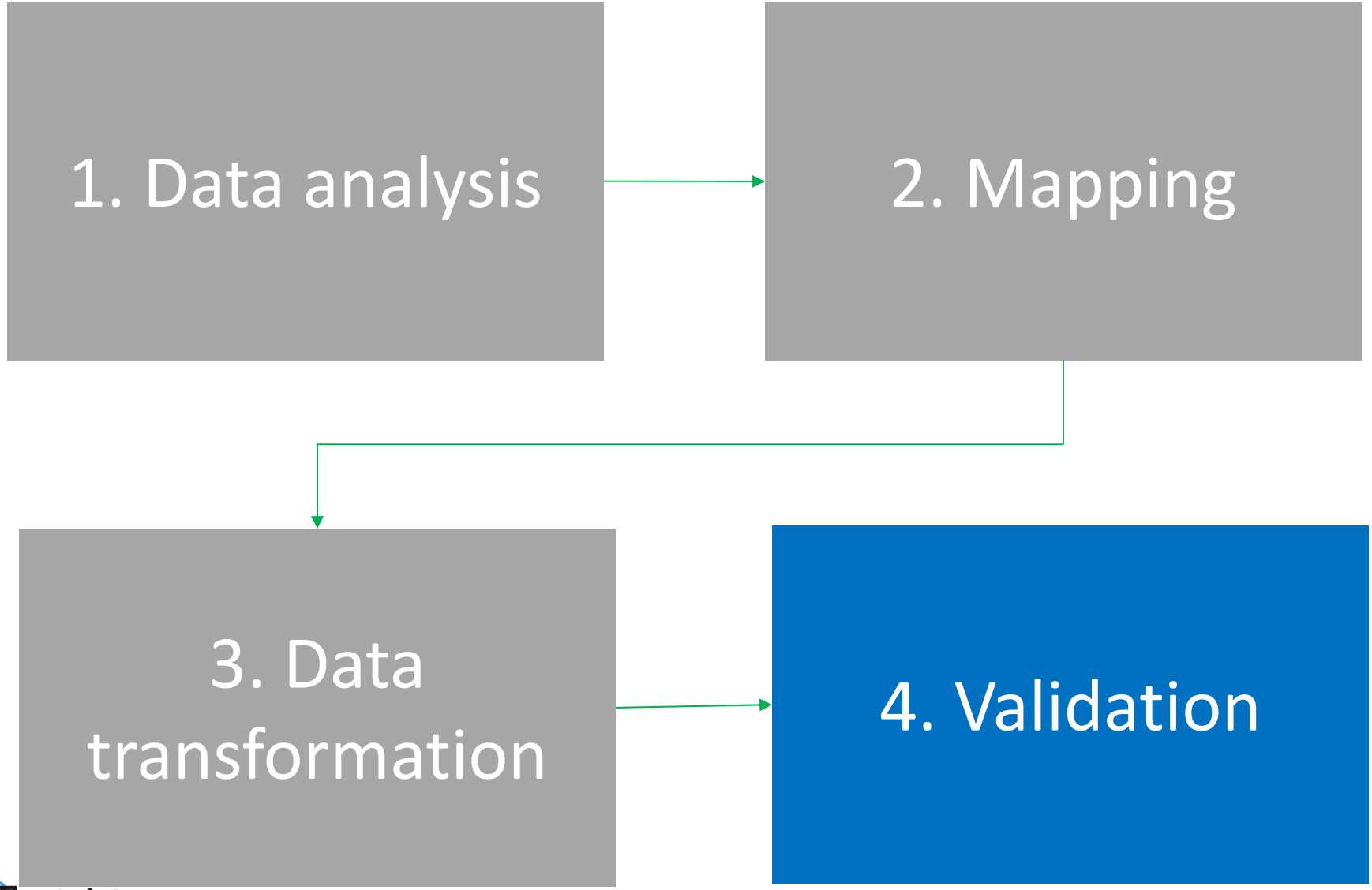
- **HALE**:
  - Exports to GML 3.2.1
  - Adapted to the INSPIRE Directive (Code Lists, application schemas ...)
  - Processing with real-time feedback
  - Online validation with the application schema
  - Allows scripting

3. Data transformation

HALE:

1. import source and target schemas
2. import the data source (having regard the ISO 8859-1 Latin alphabet)
3. import the code list in .XML or .CSV formats
4. mapping between the entities of the source and target schemas using the matching table
5. establish correspondence relationships between schemas using the following HALE functions:
    - retype
    - Assign
    - formatted string
    - Rename
    - Classification
6. validation with Land Cover application schema
7. export the result to the GML format 3.2.1




1. Data analysis
2. Mapping
3. Data transformation
4. Validation

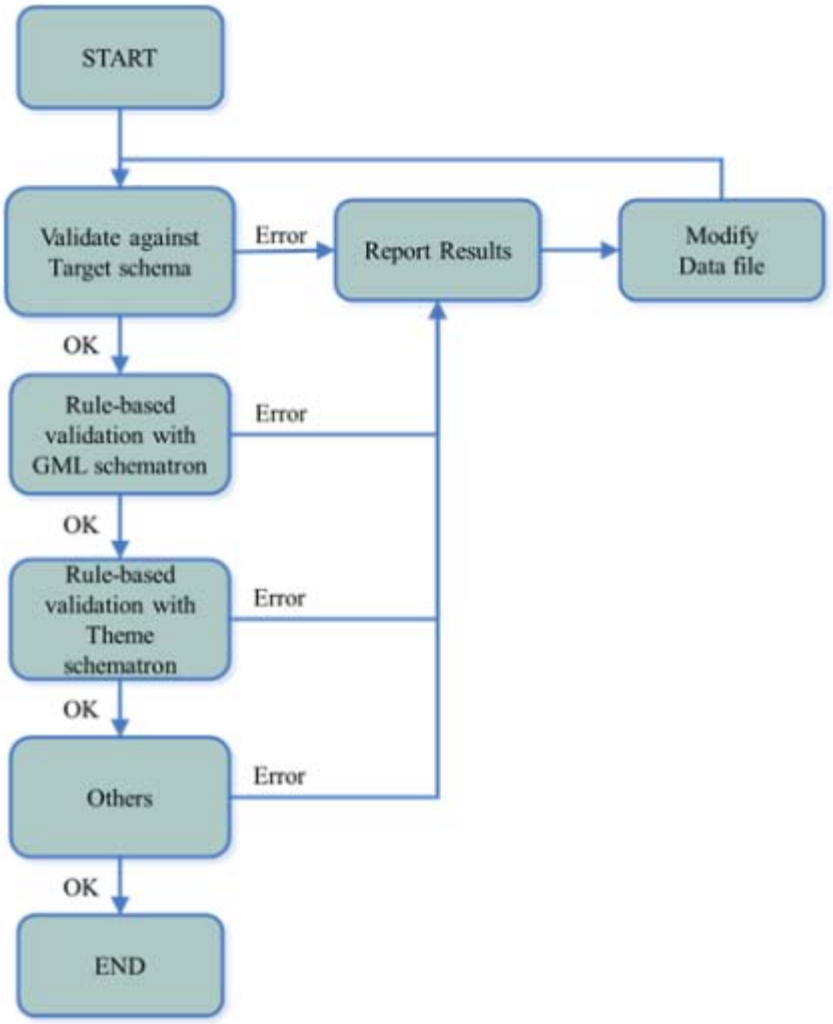
4. Validation

Abstract Test Suite (ATS): in the Annex A of the data specifications

- Abstract Test Suite (ATS)
  - Group 1 – normative
  - Group 2 – informative
- GML authomatic validation
  - LandCoverVector.xsd
  - GML Schematron 3.2.1
  - LandCover Schematron 4.0
- GML manual validation
  - Manual checking in the GML, of the characteristics specified by the ATS

ATS	Conformance classes	Abstract Tests	Related ET
Part 1 (normative)	A.1 Application Schema Conformance Class	A.1.1 Schema element denomination test	E.1
		A.1.2 Value type test	E.1
		A.1.3 Value test *	E.1
		A.1.4 Attributes/Associations completeness test	E.1
		A.1.5 Abstract spatial object test	E.1
		A.1.6 Constraints test *	E.1
		A.1.7 Geometry representation test*	E.1
	A.2 Reference Systems Conformance Class	A.2.1 Datum test *	E.1
		A.2.2 Coordinate reference system test *	E.1
		A.2.3 Grid test	E.2
		A.2.4 View service CRS test	E.2
		A.2.5 Temporal reference system test	E.2
		A.2.6 Units of measurements test	E.2
	A.3 Data Consistency Conformance Class	A.3.1 Unique identifier persistency test	E.3
		A.3.2 Version consistency test	E.3
		A.3.3 Life cycle time sequence test*	E.1
		A.3.4 Validity time sequence test *	E.1
		A.3.5 Update frequency test	E.3
	A.4 Metadata IR Conformance Class	A.4.1 Metadata for interoperability test	E.4
	A.5 Information Accessibility Conformance Class	A.5.1 Code list publication test	E.5
		A.5.2 CRS publication test *	E.1
		A.5.3 CRS identification test *	E.1
		A.5.4 Grid identification test	E.5
	A.6 Data Delivery Conformance Class	A.6.1 Encoding compliance test	E.1
	A.7 Portrayal Conformance Class	A.7.1 Layer designation test	E.6
Part 2 (informative)	A.8 Technical Guideline Conformance Class	A.8.1 Multiplicity test	E.1
		A.8.2 CRS http URI test	E.7
		A.8.3 Metadata encoding schema validation test	E.8
		A.8.4 Metadata occurrence test	E.8
		A.8.5 Metadata consistency test	E.8
		A.8.6 Encoding schema validation test	E.1
		A.8.7 Coverage multipart representation test	E.9
		A.8.8 Coverage domain consistency test	E.9
		A.8.9 Style test	E.10

4. Validation



In the COS 2010 INSPIRE Directive data specifications validation procedure, the following methodology was applied:

4. Validation

DGT was involved in 2015 in a project for the European Environmental Agency (EEA), EAGLE 6, which performed the harmonization of CORINE Land Cover and Urban Atlas data in accordance to the INSPIRE Directive.

In this project, a partnership with the **EPSILON Italia** company from the eENVplus project team was established, that resulted in the production of the **Land Cover v.4 Schematron file**, now also available in the eENVplus validator.

eENVplus provides a free online validation service (http://cloud.epsilon-italia.it/eenvplus_new/ATS.htm?), which allows the implementation of the ATS (Abstract Test Suite) included in the Annex A of the data specifications.

This Executable Test Suit (ETS) checks the conformity of the GML data sets in relation to the application schema, and also in relation to the ISO 19136: 2007 (schematron GML 3.2.1).

It also allows validation with schematron type files, for the themes already available.

4. Validation

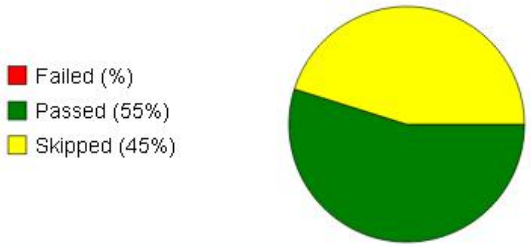
The **eENVplus** validator was used for the COS 2010 GML validation, with the Land Cover Vector application schema, the GML 3.2.1 Schematron and the Land Cover Schematron v.4 :

TestNG Results

Results overview
Reporter output

gml32-3.2.1-r18	55%
0 Groups	
0 / 17 / 14 / 31	
All GML application schemas	
GML application schemas defining features and feature collections	
GML application schemas defining spatial geometries	
GML application schemas defining time	
GML application schemas defining spatial topologies	
GML Documents	

Test suites overview



gml32-3.2.1-r18	0	17	14	31	55%
All GML application schemas	0	7	0	7	100%
GML application schemas defining features and feature collections	0	2	0	2	100%
GML application schemas defining spatial geometries	0	2	0	2	100%
GML application schemas defining time	0	0	2	2	%
GML application schemas defining spatial topologies	0	0	2	2	%
GML Documents	0	6	10	16	38%

Advantages of **eENVplus validator**:

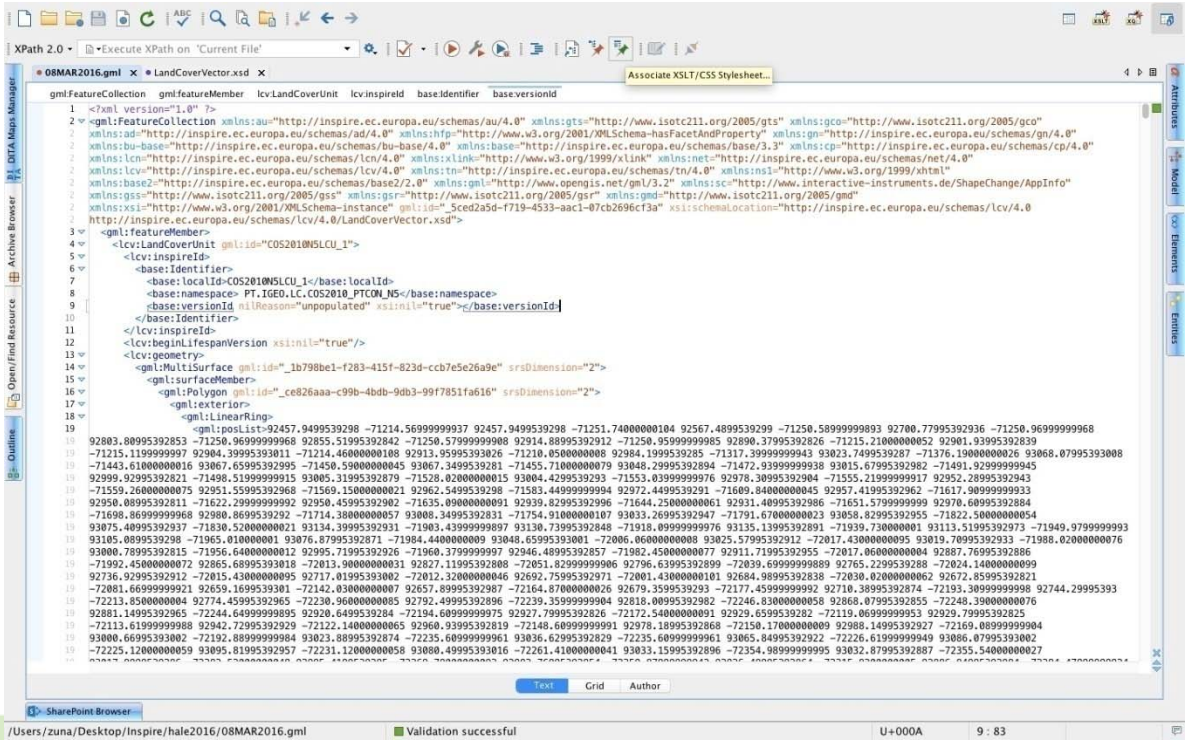
- Methodological guide on the validation process
- Online resource
- Graphical representation of the results

4. Validation

As an auxiliary for the validation of the transformed GML, it was also used **oXygen**, a software for the editing of XML, that allowed the visualization of the GML file and also the validation with the following schemes:

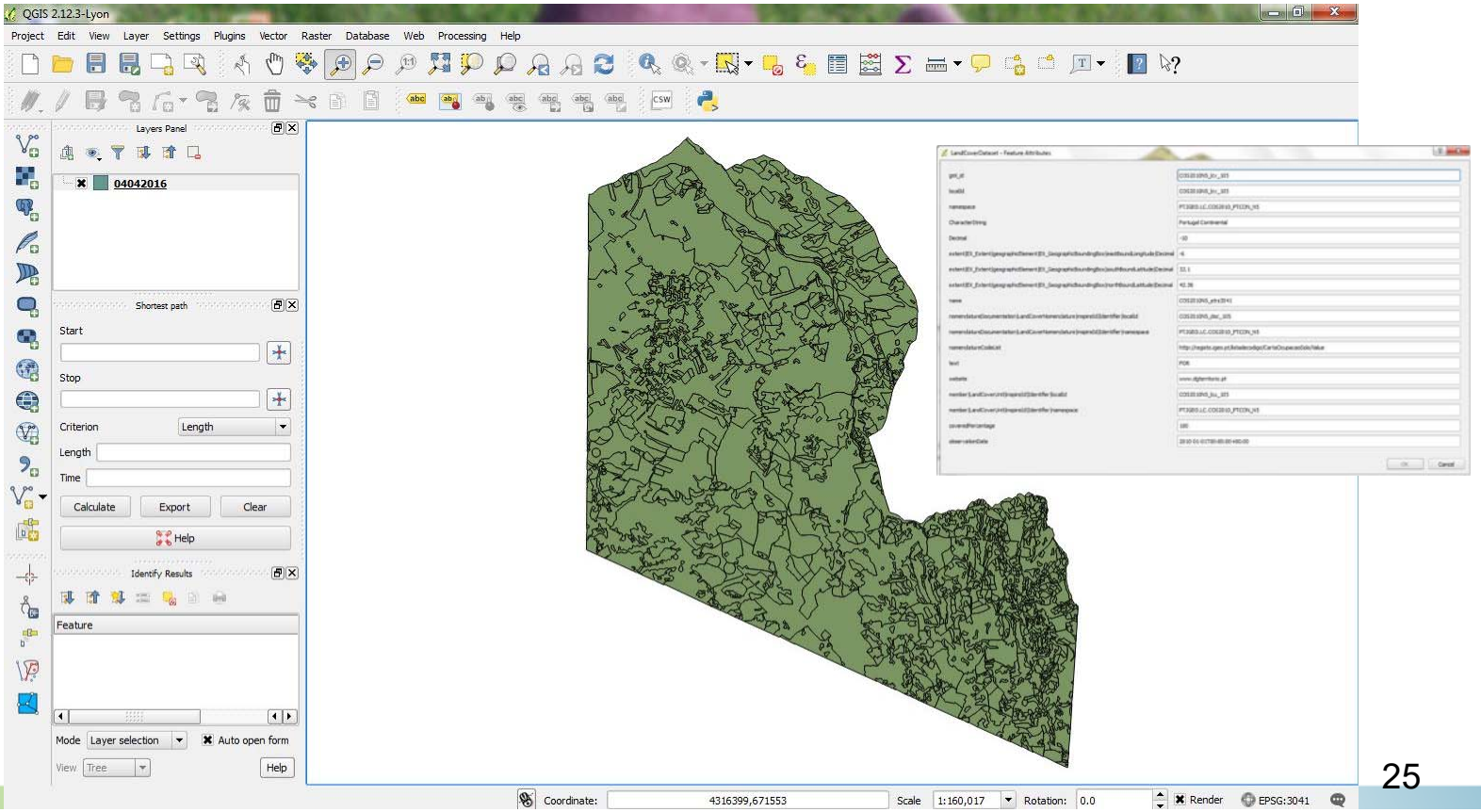
- LandCoverVector.xsd
- GML Schematron 3.2.1
- Land Cover Schematron 4.0

<oXygen/> *XML editor* helped to understand and correct the errors



Final considerations

The main objective of this work was the application of the Data Specifications for the II.2 Land Cover INSPIRE Directive Theme to the 2010 Land Cover Map of Portugal (COS 2010), in order to produce a harmonized file according to the rules established by INSPIRE Directive. A valid GML file was the final result of the harmonization process.



THANK YOU

Paulo Patrício

ppatricio@dgterritorio.pt