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)
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 geoweb services.
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.

<table>
<thead>
<tr>
<th>Data model</th>
<th>Vectorial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data structure</td>
<td>Polygons</td>
</tr>
<tr>
<td>CRS</td>
<td>ETRS89 (European Terrestrial Reference System 1989) PT-TM06</td>
</tr>
<tr>
<td>Minimum Mapping Unit (UMC)</td>
<td>1 hectare</td>
</tr>
<tr>
<td>Minimum distance between lines</td>
<td>20 meters</td>
</tr>
<tr>
<td>Nomenclature</td>
<td>Hierarchical classification with five levels and 225 classes</td>
</tr>
</tbody>
</table>
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
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)
- URI, legend: [http://registo.igeo.pt/listadecodigo/CartaOcupacaoSoloValue](http://registo.igeo.pt/listadecodigo/CartaOcupacaoSoloValue)
1. Data analysis

2. Mapping

3. Data transformation

4. Validation
2. Mapping

Correspondence 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.

<table>
<thead>
<tr>
<th>Feature type</th>
<th>Feature type description</th>
<th>Feature type definition</th>
<th>Stdtype</th>
<th>Inspire classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>LandCoverVector.xsd</td>
<td>Portuguese Land Cover Map</td>
<td>LandCoverVector.xsd</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.
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 automatic 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

<table>
<thead>
<tr>
<th>Part 1 (normative)</th>
<th>Conformance classes</th>
<th>Abstract Tests</th>
<th>Related FT</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.1 Application Schema Conformance Class</td>
<td>A.1.1 Schema element denomination test</td>
<td>E.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.1.2 Value type test</td>
<td>E.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.1.3 Value test *</td>
<td>E.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.1.4 Attributes/Associations completeness test</td>
<td>E.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.1.5 Abstract spatial object test</td>
<td>E.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.1.6 Coordinates test *</td>
<td>E.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.1.7 Geometry representation test*</td>
<td>E.1</td>
<td></td>
</tr>
<tr>
<td>A.2 Reference Systems Conformance Class</td>
<td>A.2.1 Dataset test *</td>
<td>E.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.2.2 Coordinate reference system test *</td>
<td>E.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.2.3 Grid test</td>
<td>E.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.2.4 View service CRS test</td>
<td>E.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.2.5 Temporal reference system test</td>
<td>E.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.2.6 Units of measurements test</td>
<td>E.2</td>
<td></td>
</tr>
<tr>
<td>A.3 Data Consistency Conformance Class</td>
<td>A.3.1 Unique identifier persistence test</td>
<td>E.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.3.2 Version consistency test</td>
<td>E.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.3.3 Life cycle time sequence test*</td>
<td>E.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.3.4 Validity time sequence test *</td>
<td>E.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.3.5 Update frequency test</td>
<td>E.3</td>
<td></td>
</tr>
<tr>
<td>A.4 Metadata IR Conformance Class</td>
<td>A.4.1 Metadata for interoperability test</td>
<td>E.4</td>
<td></td>
</tr>
<tr>
<td>A.5 Information Accessibility Conformance Class</td>
<td>A.5.1 Code list publication test</td>
<td>E.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.5.2 CRS publication test *</td>
<td>E.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.5.3 CRS identification test *</td>
<td>E.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.5.4 Grid identification test</td>
<td>E.5</td>
<td></td>
</tr>
<tr>
<td>A.6 Data Delivery Conformance Class</td>
<td>A.6.1 Encoding compliance test</td>
<td>E.1</td>
<td></td>
</tr>
<tr>
<td>A.7 Portugal Conformance Class</td>
<td>A.7.1 Layer designation test</td>
<td>E.6</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Part 2 (informative)</th>
<th>Conformance classes</th>
<th>Abstract Tests</th>
<th>Related FT</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.8 Technical Guideline Conformance Class</td>
<td>A.8.1 Multiplicity test</td>
<td>E.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.8.2 CRS http URL test</td>
<td>E.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.8.3 Metadata encoding schema validation test</td>
<td>E.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.8.4 Metadata occurrence test</td>
<td>E.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.8.5 Metadata consistency test</td>
<td>E.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.8.6 Encoding schema validation test</td>
<td>E.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.8.7 Coverage multipart representation test</td>
<td>E.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.8.8 Coverage domain consistency test</td>
<td>E.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A.8.9 Style test</td>
<td>E.10</td>
<td></td>
</tr>
</tbody>
</table>
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:

**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