



Harmonização de dados: o exemplo da COS

Teresa Zuna

tzuna@dgterritorio.pt



Identificação do conjunto de dados geográficos
(*source schema*)

Identificação do tema INSPIRE
(*target schema*)

Identificação de atributos, CRS, qualidade dos dados, conversão CAD/SIG

Análise

Criação
matching table
.xls

Preenchimento da *Matching table*

Mapeamento

Escolha da ferramenta ETL

Transformação dos dados

Funções de transformação

Transformação

Exportação
GML

Validação

Validação

Erros?

sim

não

Publicação dos dados harmonizados
WFS, WMS

Análise - Source Schema: Carta de Ocupação e Uso do Solo 2010 N5

- Especificações técnicas

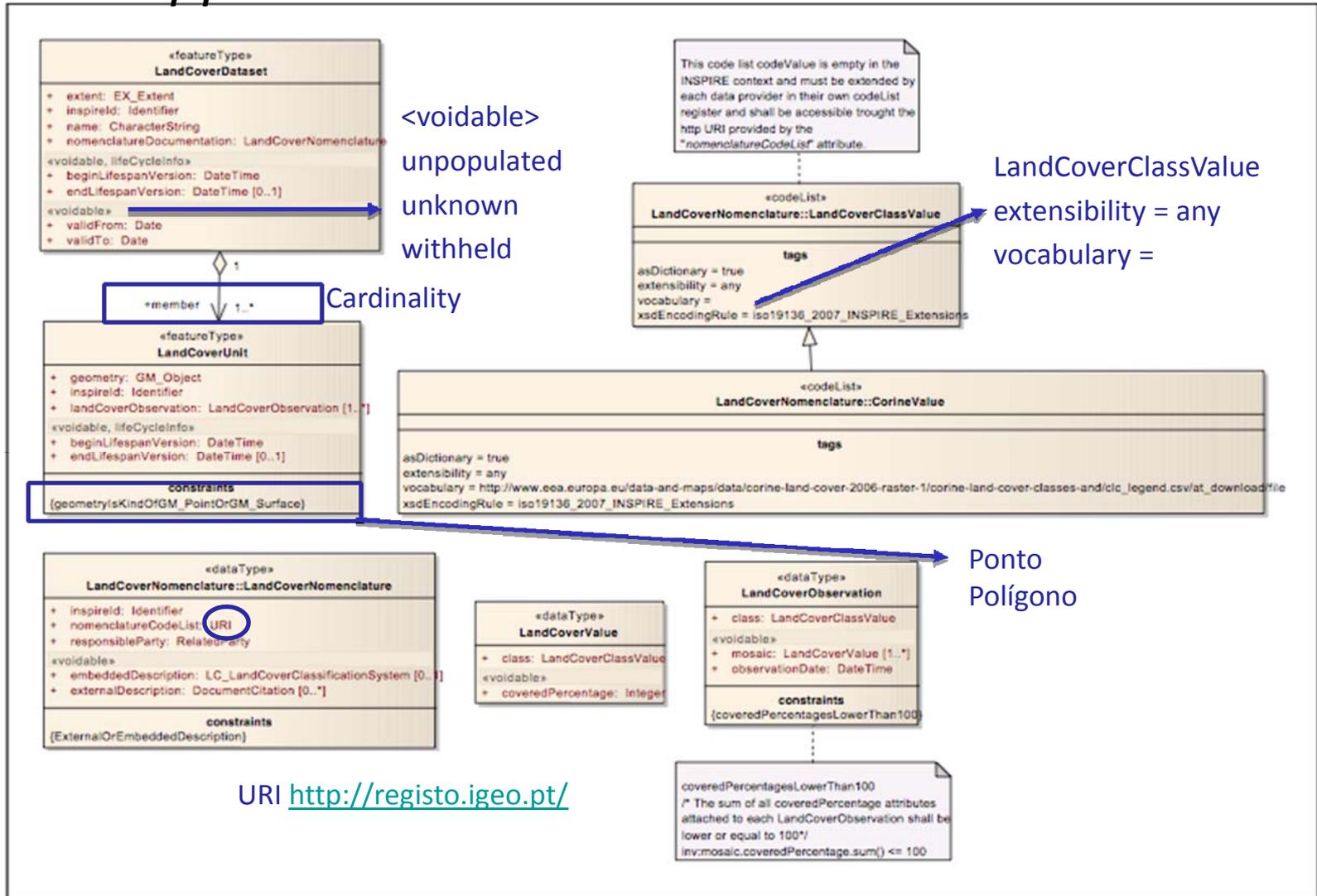
Modelo de dados	Vectorial
Estrutura de dados	Polígonos
Sistema de referência	ETRS89 (European Terrestrial Reference System 1989) PT-TM06
Unidade Mínima Cartográfica (UMC)	1 ha
Distância mínima entre linhas	20 metros
Nomenclatura	Nomenclatura hierárquica com 5 níveis de detalhe e 226 classes

Análise - Source schema: Carta de Ocupação e Uso do Solo 2010 N5

- Atributos da tabela alfanumérica

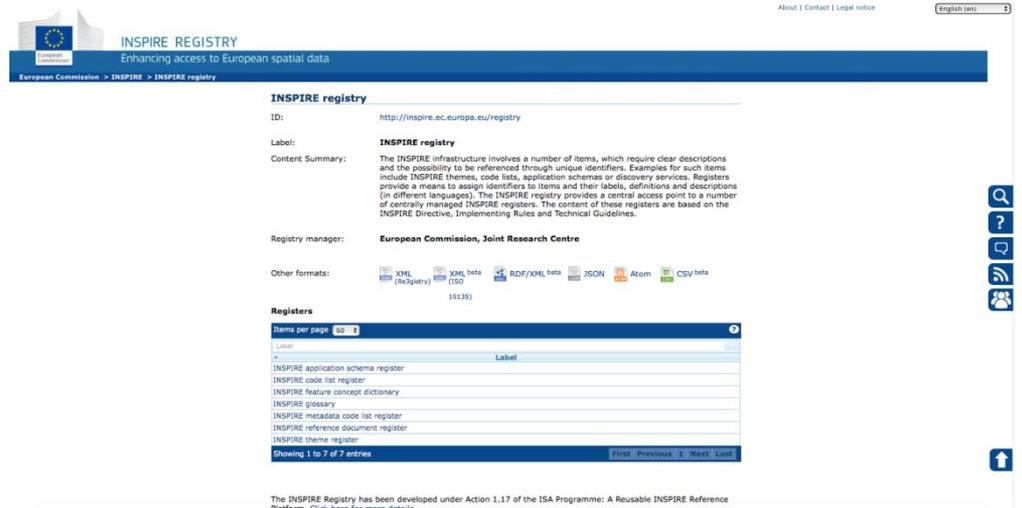
Designação dos atributos	Terminologia/ Formato
Área (ha)	AREA/ Float (19 algarismos/10 casas decimais)
Identificador único	FID/ OID (4 caracteres)
Classe de Ocupação do Solo	USO/ String (10 caracteres)
Geometria de cada polígono	THE_GEOM (Shapefile, geometry)

Análise - Target schema: Application schema LandCoverVector.xsd



Code list (Lista de códigos)

- É permitida apenas uma nomenclatura para a COS 2010
- Os valores são geridos fora da application schema *LandCoverVector.xsd*
- O produtor de dados é o responsável pela nomenclatura
- Lista hierárquica (parent value)
- INSPIRE register
 - http://INSPIRE.ec.europa.eu/codelist_register/codelist
- <http://registo.igeo.pt/listadecodigo/CartaOcupacaoSoloValue>
- URI, legenda



INSPIRE REGISTRY
Enhancing access to European spatial data

European Commission > INSPIRE > INSPIRE registry

INSPIRE registry

ID: <http://inspire.ec.europa.eu/registry>

Label: **INSPIRE registry**

Content Summary: The INSPIRE infrastructure involves a number of items, which require clear descriptions and the possibility to be referenced through unique identifiers. Examples for such items include INSPIRE themes, code lists, application schemas or discovery services. Registers provide a means to assign identifiers to items and their labels, definitions and descriptions (in different languages). The INSPIRE registry provides a central access point to a number of centrally managed INSPIRE registers. The content of these registers are based on the INSPIRE Directive, Implementing Rules and Technical Guidelines.

Registry manager: **European Commission, Joint Research Centre**

Other formats: [XML \(Registry\)](#) [XML beta \(ISO\)](#) [RDF/XML beta](#) [JSON](#) [Atom](#) [CSY beta](#)

19135

Registers

Items per page: 10

Label	Label
INSPIRE application schema register	
INSPIRE code list register	
INSPIRE feature concept dictionary	
INSPIRE glossary	
INSPIRE metadata code list register	
INSPIRE reference document register	
INSPIRE theme register	

Showing 1 to 7 of 7 entries

The INSPIRE Registry has been developed under Action 1.17 of the ISA Programme: A Reusable INSPIRE Reference Platform. Click here for more details.

Code list: COS 2010_N5

```

<value id="http://registo.igeo.pt/listadecodigo/CartaOcupacaoSoloValue/1.1.1.01.1">
  <thisversion>http://registo.igeo.pt/listadecodigo/CartaOcupacaoSoloValue/1.1.1.01.1</thisversion>
  <latestversion>http://registo.igeo.pt/listadecodigo/CartaOcupacaoSoloValue/1.1.1.01.1</latestversion>
  <label xml:lang="pt">Tecido urbano contínuo predominantemente vertical</label>
  <governance-level id="http://INSPIRE.ec.europa.eu/registry/governance-level/eu-legal">
    <label>eu-legal</label>
  </governance-level>
  <codelist id="http://registo.igeo.pt/listadecodigos/CartaOcupacaoSoloValue">
    <label xml:lang="pt">Lista de códigos da Carta de Ocupação do Solo</label>
  </codelist>
  <theme id="http://registo.igeo.pt/tema/cos">
    <label xml:lang="pt">Carta de Ocupação do Solo</label>
  </theme>
  <applicationschema id="http://INSPIRE.ec.europa.eu/applicationschema/lcn">
    <label xml:lang="pt">Nomenclatura da Carta de Ocupação do Solo</label>
  </applicationschema>
  <itemclass uriname="listadecodigos">
    <label xml:lang="pt">Lista de códigos</label>
  </itemclass>
  <status id="http://INSPIRE.ec.europa.eu/registry/status/valid">
    <label xml:lang="en">Valid</label>
  </status>
  <register id="http://registo.igeo.pt/listadecodigos">
    <label xml:lang="pt">Registo da lista de códigos INSPIRE-PT</label>
    <registry id="http://registo.igeo.pt">
      <label xml:lang="pt">Registo INSPIRE-PT</label>
    </registry>
  </register>
</value>

```

Mapeamento

Matching table (Tabela de correspondências)

- Disponível no site INSPIRE/DataModels/Mapping table
LandCoverVectorMappingTable.xml

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	Re
gml/Base	The attribute gml:id supports namespaces of a base 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 Namespace uniquely identifying the data source of	localid			id		external feature nr id		1..1505			Not available	
		beginLifespanVersion	The identifier of the particular version of the spatial object	version	1	voidable					PT.GEO.LC.COS2010_PTCOS_N5		unpopulated	Not available	
		endLifespanVersion	Date and time at which this version of the spatial object	DateTime	1	voidable								Not available	
		geom	Date and time at which this version of the spatial object	DateTime	0..1	voidable	the_geom							1:1	
LandCoverDataset	A vector representation for Land Cover data. This representation allows Land Cover data being supported by a vector geometry.						id							Not available	
		beginLifespanVersion	The identifier of the particular version of the spatial object, with a maximum length of 255	version	1	voidable								Not available	
		endLifespanVersion	Date and time at which this version of the spatial object	DateTime	1	voidable								Not available	
		extent	Date and time at which this version of the spatial object	DateTime	0..1	voidable	the_geom		polygon					1:1	
		name	Contains the extent of the data set.	EX_Extent	1		file_name							1:1	
		nomenclatureDocumentation	Name of the Land Cover data set.	CharacterString	1									Not available	
		validFrom	Information about the nomenclature used in this data set.	LandCoverNomenclature	1									Not available	
		validTo	The time when the phenomenon started to exist in the real world.	Date	1	voidable								Not available	
		member	The time from which the phenomenon no longer exists in the real world.	Date	1	voidable								Not available	
			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		LandCoverClassValue	1									1:1	
		mosaic	The assignment of a land cover class to a land cover unit through a classification.	LandCoverValue	1..*	voidable								1:1	
		observationDate	List of classification values describing into details a land cover unit, associated with percentages.	DateTime	1	voidable								Not available	
LandCoverValue	Generic class supporting Land Cover value and percentage.	class		LandCoverClassValue	1									Not available	
		coveredPercentage	Assignment of a land cover spatial object to a land.	Integer	1	voidable								Not available	

Modelo INSPIRE
LandCoverVector.xsd

Carta de Ocupação do Solo 2010 N5

Mapeamento

LandCoverVectorMappingTable.xml

gml: identificador único

Application schema	Documentation	Attribute/ Association role/ Constraint	Attribute / Association role / Constraint documentation	Values / Enumeration	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 Namespace uniquely identifying the data source of the spatial object.	localid namespace	1	
		beginLifespanVersion	The identifier of the particular version of the spatial object.	version	1	
		endLifespanVersion	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 Namespace uniquely identifying the data source of the spatial object.	localid namespace	1	
		beginLifespanVersion	The identifier of the particular version of the spatial object, with a maximum length of 25	version	1	
		endLifespanVersion	Date and time at which this version of the spatial object was inserted or changed in the	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.					

Application schemas

Descrição application schema

Member - LandCoverUnit

Multiplicity

Voidable

Atributos

Descrição dos atributos

Mapeamento COS 2010 N5

Campos
adicionados

Attribute name	Documentation	Attribute Association role Constraint	Attribute / Association role / Constraint documentation	Values / Enumerations	Multiplicity	Voidable / Non-Voidable	Status	Remarks
gml id				PT_COS2010N5_1..*			Not available	
				PT_COS2010N5_LCU_1..*			Not available	
Id		internal feature nr id		1..1505			Not available	
				PT.IGEO.LC.COS2010_PTCO_N5			Not available	
						unpopulated	Not available	
							Not available	
							Not available	
the_geom		polygon					1:1	

Status

1:1	Match
Easy	Necessita de algum processamento
Difficult	Necessita de processamento mais complexo
Not available	Quando não consta na COS 2010 nível 5

							Easy	
							1:1	
							1:1	
							Not available	
							Not available	

Transformação dos dados: HALE (Humboldt Alignment Editor)

Codelist - COS2010N5.xsd

Exportar .GML

The screenshot displays the HALE (Humboldt Alignment Editor) interface. The main window is titled 'HUMBOLDT Alignment Editor 2.9.3 - COS2010_10DEZ - C:\01\Docs\Dados\COS2010_10DEZ.hale'. The interface is divided into several panes:

- Source Data:** Shows a tree view of the source data 'COS2010_n5'. Elements include 'AREA', 'CLC1', 'filename', 'ID', 'ORIG_FID', 'the_geom', and 'USO'. A blue box highlights this pane with the text 'COS2010_N5'.
- Target Data:** Shows a tree view of the target data 'LandCoverDataset'. Elements include 'location', 'beginLifespanVersion', 'boundedBy', 'description', 'descriptionReference', 'endLifespanVersion', 'extent', 'id', 'identifier', 'inspireId', 'member', 'metaDataProperty', 'name', and 'nomenclatureDocumentation'. A blue box highlights this pane with the text 'LandCoverVector.xsd'.
- Transformed Data:** A table showing the mapping of source data to target data. The title 'Dados transformados' is overlaid on this pane. The table has columns for source and target identifiers and their corresponding values.
- Functions:** A pane at the bottom showing a flow of functions used in the transformation. It includes 'Retype', 'Classification', 'Rename', 'Generate Unique...', 'Assign', and 'id' functions. A blue box highlights this pane with the text 'funções'.
- Map:** A map view on the right side showing the spatial data being transformed, with labels like 'La Codosera' and 'Sagovia'.

Transformação dos dados: HALE

The screenshot shows an XML document editor with the following code:

```

1 <?xml version="1.0" ?>
2 <gml:FeatureCollection xmlns:ns2="urn:x-inspire:specification:gmlas:BaseTypes:3.2" xmlns:tn-a="urn:x-inspire:specification:gmlas:AirTransportNetwork:3.0"
3   xmlns:tn="urn:x-inspire:specification:gmlas:CommonTransportElements:3.0" xmlns:gts="http://www.isotc211.org/2005/gts" xmlns:base="http://inspire.ec.europa.eu/schemas/base/3.3rc3/"
4   xmlns:gco="http://www.isotc211.org/2005/gco" xmlns:net="urn:x-inspire:specification:gmlas:Network:3.2" xmlns:hfp="http://www.w3.org/2001/XMLSchema-hasFacetAndProperty"
5   xmlns:lcn="http://inspire.ec.europa.eu/schemas/lcn/3.0rc3" xmlns:base2="http://inspire.ec.europa.eu/schemas/base2/1.0rc3" xmlns:xlink="http://www.w3.org/1999/xlink"
6   xmlns:ad="urn:x-inspire:specification:gmlas:Addresses:3.0" xmlns:stat="urn:x-inspire:specification:gmlas:StatisticalUnits:0.0" xmlns:cp="urn:x-inspire:specification:gmlas:CadastralParcels:3.0"
7   xmlns:gn="urn:x-inspire:specification:gmlas:GeographicalNames:3.0" xmlns:ns1="http://www.w3.org/1999/xhtml" xmlns:gml="http://www.opengis.net/gml/3.2" xmlns:bui="urn:x-inspire:specification:gmlas:Buildings:0.0"
8   xmlns:sc="http://www.interactive-instruments.de/ShapeChange/AppInfo" xmlns:lcv="http://inspire.ec.europa.eu/schemas/lcv/3.0rc3" xmlns:gss="http://www.isotc211.org/2005/gss"
9   xmlns:au="urn:x-inspire:specification:gmlas:AdministrativeUnits:3.0" xmlns:gsr="http://www.isotc211.org/2005/gsr" xmlns:gmd="http://www.isotc211.org/2005/gmd" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
10  gml:id="_78b5aef6-9a8f-4d5e-a930-7cacff4b4e63" xsi:schemaLocation="http://inspire.ec.europa.eu/schemas/lcv/3.0rc3 file:/C:/01.Docs/Dados/LandCoverVector.xsd.xml"/>
11 <gml:featureMember>
12 <lcv:LandCoverDataset gml:id="lcv1">
13 <lcv:inspireId>
14 <base:Identificor>
15 <base:localId>2</base:localId>
16 <base:namespace>PTCON.INSPIRE.LC.LCLU2010 </base:namespace>
17 <base:versionId nilReason="unpopulated" xsi:nil="true"></base:versionId>
18 </base:Identificor>
19 </lcv:inspireId>
20 <lcv:beginLifespanVersion xsi:nil="true"/>
21 <lcv:extent>
22 <gmd:EX_Extent>
23 <gmd:geographicElement>
24 <gmd:EX_BoundingPolygon>
25 <gmd:polygon>
26 <gml:Polygon gml:id="_556f29e9-934b-4651-9da6-d480cd8664e3" srsDimension="2">
27 <gml:exterior>
28 <gml:LinearRing>
29 <gml:posList>87697.66995392863 -66691.42999999964 87774.31995392937 -66751.90000000101 87856.609953929 -66828.1099999997 87784.2299539285 -66915.76000000039 87742.41995393013
30 -66970.68999999906 87770.31995392937 -66990.16000000057 87745.36995392869 -67033.14000000073 87714.89995392945 -67046.84000000009 87660.03995392968 -67070.03000000009 87655.20995392927 -67072.57000000057
31 87639.3999539286 -67061.27000000083 87616.51995392903 -67067.48999999961 87566.24995392896 -67086.70000000072 87542.80995392922 -67094.32999999978 87524.44995392821 -67097.15000000034 87482.3699539296
32 -67109.02000000001 87464.6599539296 -67129.04999999957 87460.14995392866 -67138.08999999892 87441.69995392906 -67142.25000000006 87414.56995392896 -67153.35999999911 87386.37995392979 -67158.70000000046
33 87359.24995392968 -67172.58999999991 87347.02995392871 -67178.03999999956 87306.56995392934 -67190.86000000009 87284.98995392954 -67196.70000000033 87260.92995392834 -67209.94000000009 87231.70995392965
34 -67228.38000000076 87211.02995392919 -67236.46999999933 87185.90995392892 -67193.74999999956 87184.05995392881 -67190.59999999969 87214.17995392907 -67169.90999999938 87219.56995392965 -67164.06999999991
35 87198.44995392936 -67143.38000000105 87251.93995392822 -67115.48999999944 87325.21995392806 -67072.76999999968 87363.43995392894 -67045.33999999908 87376.9199539281 -67017.47000000009 87421.42995392889
36 -66946.71000000105 87446.59995392838 -66917.48000000029 87462.23995392949 -66889.54999999926 87444.25995392923 -66885.16000000093 87446.78995392987 -66882.14999999989 87475.27995392881 -66848.27999999992
37 87526.97995392885 -66789.81999999989 87566.84995392851 -66744.05000000092 87583.0299539301 -66736.84999999961 87600.55995392846 -66717.52000000037 87623.48995392947 -66692.78000000086 87633.8299539286
38 -66695.93000000074 87648.21995392913 -66684.24000000004 87654.95995392872 -66676.13999999977 87670.69995392827 -66675.69000000009 87671.92995392912 -66676.63000000102 87678.33995392939 -66681.54000000018
39 87697.66995392863 </gml:LinearRing>
40 </gml:exterior>
41 </gml:Polygon>
42 </gmd:polygon>
43 </gmd:EX_BoundingPolygon>
44 </gmd:geographicElement>
45 </gmd:EX_Extent>
46 </lcv:extent>
47 <lcv:name>C052010_n5</lcv:name>
48 <lcv:nomenclatureDocumentation>
49 <lcv:LandCoverNomenclature>
50 <lcv:inspireId>
51 <base:Identificor>

```

Annotations in the image:

- A blue arrow points to the `xmlns:gml="http://www.opengis.net/gml/3.2"` declaration, with the text "Apontadores para os esquemas" next to it.
- Another blue arrow points to the `gml:posList` element, with the text "Parte geométrica" next to it.

Validação dos resultados

1. Land Cover Vector.xsd (v.4)
2. GML Schematron
3. LandCoverVector Schematron (v.3.3)

Categorias de conformidade:

- Esquema de aplicação
- Geometria/Topologia
- Sistema de Referência
- Qualidade dos dados
- Metadados
- Representação gráfica

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

Validação dos resultados: COS 2010 N5 GML

The screenshot shows the XPath 2.0 application interface. The main window displays the XML schema for Schematron constraints for GML / ISO 19136. The schema includes various rules and assertions, such as those for ValueArray, PolygonPatch, and srsName attributes.

The bottom panel shows the validation results, listing several errors:

Info	Description - 37 items	Resource
❌	E [Xerces] s4s-elt-character: Non-whitespace characters are not allowed in schema elements other than 'xs:appinfo' and 'xs:documentation'. Saw 'All values in the domain set shall be gml:MultiS...	SchematronConstri
❌	E [Xerces] s4s-elt-character: Non-whitespace characters are not allowed in schema elements other than 'xs:appinfo' and 'xs:documentation'. Saw 'All values in the domain set shall be gml:MultiS...	SchematronConstri
❌	E [Xerces] s4s-elt-character: Non-whitespace characters are not allowed in schema elements other than 'xs:appinfo' and 'xs:documentation'. Saw 'All values in the domain set shall be gml:Grid ...	SchematronConstri
❌	E [Xerces] s4s-elt-character: Non-whitespace characters are not allowed in schema elements other than 'xs:appinfo' and 'xs:documentation'. Saw 'All values in the domain set shall be gml:Rectifi...	SchematronConstri
❌	E [Xerces] s4s-elt-schema-ns: The namespace of element 'schema' must be from the schema namespace, 'http://www.w3.org/2001/XMLSchema'.	SchematronConstri
❌	E [Xerces] s4s-elt-invalid-content.1: The content of 'schema' is invalid. Element 'title' is invalid, misplaced, or occurs too often.	SchematronConstri
❌	E [Xerces] s4s-elt-invalid-content.1: The content of 'schema' is invalid. Element 'ns' is invalid, misplaced, or occurs too often.	SchematronConstri
❌	E [Xerces] s4s-elt-invalid-content.1: The content of 'schema' is invalid. Element 'ns' is invalid, misplaced, or occurs too often.	SchematronConstri

At the bottom of the application, the URL <http://schemas.opengis.net/gml/3.2.1/SchematronConstraints.xml> is visible, along with a status bar showing "Learn completed", "U+0073", and "1 : 2".

Notas finais

- Conhecimentos sobre XML/GML, UML, XSD
- Compreensão dos princípios de interoperabilidade e harmonização de conjuntos de dados geográficos
 - Implementing Rules e Technical Guidelines
- Consulta dos Grupos de Trabalho por tema
- Partilha de informação e conhecimento