Code lists for interoperability – Principles and best practices in INSPIRE

Michael Lutz
EGU, ESSI2.5, 24 April 2012

www.jrc.ec.europa.eu

Serving society
Stimulating innovation
Supporting legislation
Content

• INSPIRE context
• Why code lists?
• Starting point
• Code lists in INSPIRE data specifications
• Implementation
“Infrastructure for Spatial Information in the European Community”

Distributed infrastructure
27 countries
23 languages

European legislation

General rules for establishment

Environment
34 spatial data themes
## INSPIRE Thematic Scope

### Annex I
- 1. Coordinate reference systems
- 2. Geographical grid systems
- 3. Geographical names
- 4. Administrative units
- 5. Addresses
- 6. Cadastral parcels
- 7. Transport networks
- 8. Hydrography
- 9. Protected sites

### Annex II
- 1. Elevation
- 2. Land cover
- 3. Ortho-imagery
- 4. Geology

### Annex III
- 1. Statistical units
- 2. Buildings
- 3. Soil
- 4. Land use
- 5. Human health and safety
- 6. Utility and governmental services
- 7. Environmental monitoring facilities
- 8. Production and industrial facilities
- 9. Agricultural and aquaculture facilities
- 11. Area management/restriction/regulation zones & reporting units
- 12. Natural risk zones
- 13. Atmospheric conditions
- 14. Meteorological geographical features
- 15. Oceanographic geographical features
- 16. Sea regions
- 17. Bio-geographical regions
- 18. Habitats and biotopes
- 19. Species distribution
- 20. Energy Resources
- 21. Mineral resources
Data interoperability

The starting point ...

- Access to spatial data in various ways
- User has to deal with interpreting heterogeneous data in different formats, identify, extract and post-process the data he needs → lack of interoperability
Data interoperability

... and what INSPIRE is aiming at

- Provide access to spatial data 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 data model and the harmonised data specification
Some examples

<table>
<thead>
<tr>
<th>Code</th>
<th>Country/territory</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>Andorra</td>
</tr>
<tr>
<td>AE</td>
<td>United Arab Emirates</td>
</tr>
<tr>
<td>AF</td>
<td>Afghanistan</td>
</tr>
<tr>
<td>AG</td>
<td>Antigua and Barbuda</td>
</tr>
<tr>
<td>AI</td>
<td>Anguilla</td>
</tr>
<tr>
<td>AL</td>
<td>Albania</td>
</tr>
<tr>
<td>AM</td>
<td>Armenia</td>
</tr>
<tr>
<td>AO</td>
<td>Angola</td>
</tr>
<tr>
<td>AQ</td>
<td>Antarctica</td>
</tr>
<tr>
<td>AR</td>
<td>Argentina</td>
</tr>
<tr>
<td>AS</td>
<td>American Samoa</td>
</tr>
<tr>
<td>AT</td>
<td>Austria</td>
</tr>
<tr>
<td>AU</td>
<td>Australia</td>
</tr>
<tr>
<td>AW</td>
<td>Aruba</td>
</tr>
<tr>
<td>AX</td>
<td>Åland Islands</td>
</tr>
<tr>
<td>AZ</td>
<td>Azerbaijan</td>
</tr>
<tr>
<td>BA</td>
<td>Bosnia and Herzegovina</td>
</tr>
<tr>
<td>BB</td>
<td>Barbados</td>
</tr>
<tr>
<td>BD</td>
<td>Bangladesh</td>
</tr>
<tr>
<td>BE</td>
<td>Belgium</td>
</tr>
<tr>
<td>BF</td>
<td>Burkina Faso</td>
</tr>
<tr>
<td>BG</td>
<td>Bulgaria</td>
</tr>
<tr>
<td>BH</td>
<td>Bahrain</td>
</tr>
</tbody>
</table>

--- Further files and information ---

**Top of classification**

<table>
<thead>
<tr>
<th>Detail</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Infectious and parasitic diseases</td>
</tr>
<tr>
<td>06</td>
<td>Neoplasms</td>
</tr>
<tr>
<td>25</td>
<td>Diseases of the blood-(forming organs), immunol. disorders</td>
</tr>
<tr>
<td>26</td>
<td>Endocrine, nutritional and metabolic diseases</td>
</tr>
<tr>
<td>28</td>
<td>Mental and behavioural disorders</td>
</tr>
<tr>
<td>31</td>
<td>Diseases of the nervous system and the sense organs</td>
</tr>
<tr>
<td>33</td>
<td>Diseases of the circulatory system</td>
</tr>
<tr>
<td>37</td>
<td>Diseases of the respiratory system</td>
</tr>
<tr>
<td>42</td>
<td>Diseases of the digestive system</td>
</tr>
<tr>
<td>45</td>
<td>Diseases of the skin and subcutaneous tissue</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Coastal and halophytic habitats</td>
<td></td>
</tr>
<tr>
<td>11. Open sea and tidal areas</td>
<td></td>
</tr>
<tr>
<td>Sandbanks which are slightly covered by sea water all the time</td>
<td>1110</td>
</tr>
<tr>
<td>Posidonia beds (Posidonia oceanicae)</td>
<td>1120</td>
</tr>
<tr>
<td>Estuaries</td>
<td>1130</td>
</tr>
<tr>
<td>Mudflats and sandflats not covered by seawater at low tide</td>
<td>1140</td>
</tr>
<tr>
<td>Coastal lagoons</td>
<td>1150</td>
</tr>
<tr>
<td>Large shallow inlets and bays</td>
<td>1160</td>
</tr>
<tr>
<td>Reefs</td>
<td>1170</td>
</tr>
<tr>
<td>Submarine structures made by leaking gases</td>
<td>1180</td>
</tr>
<tr>
<td>12. Sea cliffs and shingle or stony beaches</td>
<td></td>
</tr>
<tr>
<td>Annual vegetation of drift lines</td>
<td>1210</td>
</tr>
<tr>
<td>Perennial vegetation of stony banks</td>
<td>1220</td>
</tr>
<tr>
<td>Vegetated sea cliffs of the Atlantic and Baltic Coasts</td>
<td>1230</td>
</tr>
<tr>
<td>Vegetated sea cliffs of the Mediterranean coasts with endemic Limonium spp</td>
<td>1240</td>
</tr>
<tr>
<td>Vegetated sea cliffs with endemic flora of the Macaronesian coasts</td>
<td>1250</td>
</tr>
</tbody>
</table>

---
Why use code lists?

• Provide agreed set of values with multi-lingual names, definitions and descriptions to be (re-)used as values of properties
  ⇒ “controlled vocabularies” for the values of properties
  • Example: observedProperty of an observation
  ⇒ Less variation in coding, minimising the duplication of datasets (compared with free text)
  ⇒ Data consumers (client applications) know and understand the values used by data providers
  ⇒ More reliable searching & recovery of data items
  ⇒ Enhanced interoperability
The starting point

- Use & adoption of code lists is growing and evolving in many domains
  - Code lists for spatial data already exist and should be reused in INSPIRE (wherever possible)
- Code lists in INSPIRE may be centrally governed by the EC, within thematic communities or international organisations (e.g. WMO)
  - They may have Member State-specific extensions
  - Maintenance rules should be (but are not always) defined
- Existing code lists vary widely in terms of
  - structure
  - maintenance
  - availability (formats, services, ...)


CLs in Annex II+III data specifications

- **Obligation:**
  - requirement → only specified values SHALL be used
  - recommendation → only specified values SHOULD be used

- **Governance:**
  - INSPIRE-governed → include values in data specs
  - externally governed → include reference in specs, plus:
    - external organisation responsible for maintaining the CL
    - version to be used
    - availability: URL/citation
    - formats, e.g. SKOS/RDF, XML, HTML, PDF, ...
    - specific subsets to be used
    - rules for encoding values (URIs, labels)

- **Vocabulary:** unique id (URI) of the code list values

- **Extensibility**
CL & interoperability – 3 patterns

1) One code list is mandated
   • Example: use only NASA SWEET ontology on physical properties to describe observedProperty
   → CL is well-known by client; data integration/queries possible
   → used CL need not be declared by provider

2) One of several code lists is mandated
   • Example: use ICAO or IATA code list to describe airportCode
   → CLs are well-known by client; data integration/queries possible if mappings exist
   → used CL needs to be declared by provider

3) Any appropriate code list may be used
   • Example: any soil classification scheme can be used
   → CLs are not well-known by client; data integration/queries difficult
   → used CL needs to be declared by provider

→ Combine approaches 1/2 and 3
## Updates vs. Extensions

<table>
<thead>
<tr>
<th>FormOfRoadNodeValue</th>
<th>to be used in INSPIRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>enclosedTrafficArea</td>
<td></td>
</tr>
<tr>
<td>junction</td>
<td></td>
</tr>
<tr>
<td>levelCrossing</td>
<td>superseded</td>
</tr>
<tr>
<td>roundabout</td>
<td>deprecated</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FormOfRoadNodeValue</th>
<th>to be used in INSPIRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>enclosedTrafficArea</td>
<td></td>
</tr>
<tr>
<td>junction</td>
<td></td>
</tr>
<tr>
<td>levelCrossingWithSignal</td>
<td></td>
</tr>
<tr>
<td>levelCrossingWithoutSignal</td>
<td></td>
</tr>
<tr>
<td>(roundabout)</td>
<td></td>
</tr>
<tr>
<td>trafficSquare</td>
<td></td>
</tr>
<tr>
<td>roadServiceArea</td>
<td></td>
</tr>
</tbody>
</table>
## Updates vs. Extensions

<table>
<thead>
<tr>
<th>FormOfRoadNode Value</th>
<th>Extension to be used in INSPIRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>enclosedTrafficArea</td>
<td></td>
</tr>
<tr>
<td>junction</td>
<td></td>
</tr>
<tr>
<td>levelCrossing</td>
<td></td>
</tr>
<tr>
<td>roundabout</td>
<td></td>
</tr>
</tbody>
</table>

### Additional values
- trafficSquare
- roadServiceArea
Updates

- Addition and supersession will lead to different values in both versions of CL
- Deprecation will (ultimately) lead to a code no longer being used in (new) data
- This may lead to problems
  - in data integration if different providers use different versions
  - in queries if client and provider use different versions
- Supersessions can be handled as mappings
  - avoid simple addition and deprecation where possible?
Extensibility in INSPIRE

- Are values other than those included in the CL allowed for a CL-valued attribute? → 3 types
  - None: only the values included in the CL may be used → high interoperability, low flexibility
  - narrower: only the values included in the CL or narrower values may be used → high interoperability & flexibility, but additional reasoning may be needed for answering queries
  - any: any values may be used in addition to those specified in the CL → low interoperability & high flexibility

- If data providers use additional codes, data may contain values that are not well-known to client
  - This may lead to problems in data integration and queries
INSPIRE CL registry requirements

- manage CL values from INSPIRE DSs & IRs
  - multi-lingual
  - partly hierarchical
  - also allow references to external code lists
  - following ISO 19135 procedures, roles & metadata model
  - consistent with INSPIRE maintenance procedures and roles

- make CL values available to INSPIRE users
  - human-readable (HTML) → look up and understand values, find translations
  - machine-readable (XML, RDF, web service?) → populate user interfaces, check data compliance, data transformation, URIs for encoding

- Register may be referred to from Annex II+III data specifications and possibly legal act
Encoding

• Use URI of the value for interoperability plus label for display in UIs.

• Example: In GML, the value identifier and label for the CountryCode for Germany (value id: DE) could be encoded using the xlink:href and xlink:title properties.

```xml
<SomeFeature gml:id="abc123">
  (...)
</SomeFeature>
```

• Unclear how to include multi-lingual labels.
More information

• INSPIRE
  • http://inspire.jrc.ec.europa.eu/

• INSPIRE data specifications
  • Overview
  • Data models
    ▪ http://inspire.jrc.ec.europa.eu/index.cfm/pageid/2/list/datamodels
  • Schemas
    ▪ http://inspire.ec.europa.eu/schemas/
1. Coastal and halophytic habitats
  11. Open sea and tidal areas
     Sandbanks which are slightly covered by sea water all the time
     Posidonia beds (Posidonion oceanicae)
     Estuaries
     Mudflats and sandflats not covered by seawater at low tide
     Coastal lagoons
     Large shallow inlets and bays
     Reefs
     Submarine structures made by leaking gases

12. Sea cliffs and shingle or stony beaches
    Annual vegetation of drift lines
    Perennial vegetation of stony banks
    Vegetated sea cliffs of the Atlantic and Baltic Coasts
    Vegetated sea cliffs of the Mediterranean coasts with endemic Limonium spp
    Vegetated sea cliffs with endemic flora of the Macaronesian coasts
    ...