Contents

Annex D
(normative)
Example of data quality measure for CORINE land Cover Survey initiative

Developed in compliance with the guidelines and templates given in ISO 19114 ISO 19113 and ISO 19138.

Data quality components (Short name)	
DQ_Name	Duplicate coverage (overlaps)
DQ_Alias	
DQ_Scope	CLC coverage
DQ_Element	Completeness
DQ_Subelement	Commission
DQ_Measure	
DQ_MeasureDef	
DQ_MeasureDesc	Pass-fail
DQ_MeasureStructure	
DQ_MeasureRefSource	
DQ_MeasureID	
DQ_EvalMethod	
DQ_EvalMethodType	Internal
DQ_EvalMethodDesc	Check the complete CLC coverage for overlaps.
DQ_QualityResult	
DQ_ValueType	Boolean
DQ_Value (example)	False
DQ_ValueUnit	NA
DQ_Date	2011-04-15
DQ_ConformanceLevel	Zero violations in the coverage
Example dataset parameters	3 contiguous area covered by more than one CLC polygons found.
Example quality result meaning	The dataset fails. Commission exists. Overlapping CLC polygons found in the coverage.

Table 18 : Example for Completeness - commission: Duplicate coverage (overlaps)

Data quality components (Short name)	
DQ_Name	Gaps in the CLC coverage
DQ_Alias	
DQ_Scope	Entire area to be filled with CLC polygons
DQ_Element	Completeness
DQ_Subelement	Omission
DQ_Measure	
DQ_MeasureDef	
DQ_MeasureDesc	Pass-fail
DQ_MeasureStructure	
DQ_MeasureRefSource	
DQ_MeasureID	
DQ_EvalMethod	
DQ_EvalMethodType	Internal
DQ_EvalMethodDesc	Compare the target area of CLC mapping against the area covered by the CLC coverage.
DQ_QualityResult	
DQ_ValueType	Boolean
DQ_Value (example)	False
DQ_ValueUnit	NA
DQ_Date	2011-04-15
DQ_ConformanceLevel	Zero violations in the coverage
Example dataset parameters	2 small areas found internally, which are not covered by CLC polygons. Additionally, 76 ha of the country area along the state boundary is missing from the CLC interpretation.
Example quality result meaning	The coverage fails. Omissions exists. Country area is not entirely filled with CLC polygons.

Table 19 : Example for Completeness - omission: Gaps in the CLC coverage

Data quality components (Short name)	
DQ_Name	Existence of valid country-level metadata
DQ_Alias	
DQ_Scope	Contry-level metadata linked to CLC coverage
DQ_Element	Completeness
DQ_Subelement	Omission
DQ_Measure	
DQ_MeasureDef	
DQ_MeasureDesc	Pass-fail
DQ_MeasureStructure	
DQ_MeasureRefSource	
DQ_MeasureID	
DQ_EvalMethod	
DQ_EvalMethodType	Internal
DQ_EvalMethodDesc	Check metadata document for existence and validity.
DQ_QualityResult	
DQ_ValueType	Boolean
DQ_Value (example)	False
DQ_ValueUnit	NA
DQ_Date	2011-04-15
DQ_ConformanceLevel	Zero violations in the coverage
Example dataset parameters	Country-level metadata exist, but resource title and Metadata point of contact missing.
Example quality result meaning	The dataset fails. Country-level metadata is not complete.

Table 20 : Example for Completeness - omission: Existence of valid country-level metadata

Data quality components (Short name)	
DQ_Name	25 ha MMU conformance
DQ_Alias	
DQ_Scope	All polygons in the LC coverage, except polygons along the clipping boundary
DQ_Element	Logical consistency
DQ_Subelement	Conceptual consistency
DQ_Measure	
DQ_MeasureDef	
DQ_MeasureDesc	Pass-fail
DQ_MeasureStructure	
DQ_MeasureRefSource	
DQ_MeasureID	
DQ_EvalMethod	
DQ_EvalMethodType	Internal
DQ_EvalMethodDesc	Select polygons smaller than the pre-defined 25 ha MMU. Small polygons along clipping boundaries are excluded from the examination.
DQ_QualityResult	
DQ_ValueType	Boolean
DQ_Value (example)	False
DQ_ValueUnit	NA
DQ_Date	2011-04-15
DQ_ConformanceLevel	Zero violations in the coverage
Example dataset parameters	125 429 polygons are within data quality scope. 8 polygons with an area smaller than 25 ha MMU exists.
Example quality result meaning	The coverage fails. Violation of the conceptual schema exists.

Table 21 : Example for Logical consistency – conceptual consistency: 25 ha MMu conformance

Data quality components (Short name)	
DQ_Name	Valid codes
DQ_Alias	
DQ_Scope	All polygons in the CLC coverage
DQ_Element	Logical consistency
DQ_Subelement	Domain consistency
DQ_Measure	
DQ_MeasureDef	
DQ_MeasureDesc	Pass-fail
DQ_MeasureStructure	
DQ_MeasureRefSource	
DQ_MeasureID	
DQ_EvalMethod	
DQ_EvalMethodType	Internal
DQ_EvalMethodDesc	Compare the CLC codes linked as attribute to the polygons in the CLC coverage against the list of valid CLC codes.
DQ_QualityResult	
DQ_ValueType	Boolean
DQ_Value (example)	False
DQ_ValueUnit	NA
DQ_Date	2011-04-15
DQ_ConformanceLevel	Zero violations in the coverage.
Example dataset parameters	125 429 polygons are within data quality scope. 5 polygons with non-valid CLC code exists.
Example quality result meaning	The coverage fails. The attributes of at least one polygons violated the attribute domain.

Table 22 : Example for Logical consistency – domain consistency: Valid codes

Data quality components (Short name)	
DQ_Name	Self crossing polygons
DQ_Alias	
DQ_Scope	All polygons in the CLC coverage
DQ_Element	Logical consistency
DQ_Subelement	Topological consistency
DQ_Measure	
DQ_MeasureDef	
DQ_MeasureDesc	Pass-fail
DQ_MeasureStructure	
DQ_MeasureRefSource	
DQ_MeasureID	
DQ_EvalMethod	
DQ_EvalMethodType	Internal
DQ_EvalMethodDesc	Check the coverage against self crossing polygons („shape 8” polygons”).
DQ_QualityResult	
DQ_ValueType	Boolean
DQ_Value (example)	False
DQ_ValueUnit	NA
DQ_Date	2011-04-15
DQ_ConformanceLevel	Zero violations in the coverage.
Example dataset parameters	10 self crossing polygons have been found within 125 429 polygons (data quality scope).
Example quality result meaning	The coverage fails. Violation of the topologycal schema exists.

Table 23 : Example for Logical consistency – topological consistency: Self-crossing polygons

Data quality components (Short name)	
DQ_Name	Attributes name convention
DQ_Alias	
DQ_Scope	CLC coverage
DQ_Element	Logical consistency
DQ_Subelement	Format consistency
DQ_Measure	
DQ_MeasureDef	
DQ_MeasureDesc	Pass-fail
DQ_MeasureStructure	
DQ_MeasureRefSource	
DQ_MeasureID	
DQ_EvalMethod	
DQ_EvalMethodType	Internal
DQ_EvalMethodDesc	Check names of attribute fields against attribute name convention described int he CLC Technical Guidelines.
DQ_QualityResult	
DQ_ValueType	Boolean
DQ_Value (example)	False
DQ_ValueUnit	NA
DQ_Date	2011-04-15
DQ_ConformanceLevel	Zero violations in the coverage.
Example dataset parameters	CLC code field name in the CLC2006 coverage not correct: „Code" instead of the convention „Code_06".
Example quality result meaning	The coverage fails. The name of at least one attribute field do not correspond to the convention.

Table 24 : Example for Logical consistency – format consistency: Attributes name convention

Data quality components (Short name)	
DQ_Name	Overall accuracy
DQ_Alias	
DQ_Scope	Entire area covered by CLC polygons
DQ_Element	Thematic accuracy
DQ_Subelement	Classification correctness
DQ_Measure	
DQ_MeasureDef	
DQ_MeasureDesc	Percent correctly classified (PCC)
DQ_MeasureStructure	
DQ_MeasureRefSource	
DQ_MeasureID	
DQ_EvalMethod	
DQ_EvalMethodType	External
DQ_EvalMethodDesc	Compare CLC codes of coverage polygons against CLC codes of virtual polygons, the validation expert would draw around random sample points. Calculate the rate of sample points where the two CLC code agrees.
DQ_QualityResult	
DQ_ValueType	Percentage
DQ_Value (example)	86,1%
DQ_ValueUnit	Percent
DQ_Date	2011-04-15
DQ_ConformanceLevel	Greater than 85% of the area shall be correctly classified.
Example dataset parameters	An agreement between CLC codes has been found for 861 of 1000 samples. The agreement rate within the samples is 86,1%.
Example quality result meaning	The coverage passes. The estimated overall accuracy value is appr. 86,1% ± 1,1% on the P = 68.3% significance level.

Table 25 : Example for Thematic accuracy – classification correctness: Overall accuracy

Annex E
(informative)
Examples of Land Cover Parameters

As explained before in the narrative description, the earth's bio-physical surface is populated with landscape elements which combine to form the land cover, and these elements very frequently are collected by other mapping initiatives than the land cover surveys. Parameterization is essential not only for describing the way these landscape elements compound the different land cover situations, but also for adding land cover attributes to these situations (irrigation in crops, cutting of trees in forest areas, urban areas under construction or abandoned, etc.).

The next set of examples intends to show how a data model with land cover parameters allows presenting both the characteristics of land cover, and the relationship between each occurrence of land cover with the different landscape elements on it (trees, buildings, etc.). These examples are based both on the experience of some Members States who are already using object-oriented data models (according to ISO 19107), such as the German DLM (Digital Landscape Model), or Spanish SIOSE (Spanish Land Cover Information System), and the various exercises made by European Working Groups of land cover experts, with the aim to improve current European land cover data models, such us EEA's CORINE Land Cover (CLC), or Eurostat's LUCAS, by adding detailed information that already exist in many national inventories, in a harmonized and interoperable way.

The 'ParameterType' classes are designed to respond these needs for adding and exchange parameterized information within Inspire Land Cover Data Specifications. The 'ParameterType' abstract class can take shape in three specific classes, whom inherit attributes from it, that describe the different typology of considered parameters, according criteria of accountability and measurability of the phenomenon taken in account. These classes are:

- □ CountableParameter
- □ PresenceParameter
- □ PercentageParameter

So each example has a description of the thematic requirement which is addressing, with some visual examples, and a brief explanation of how the 'ParameterType' classes could be used to solve each case.

Example 1: Percentage Parameter

One of the key factors when producing CLC polygons in natural areas is precisely the determination of the density of tree crown cover in forests. In the CLC nomenclature, the class definition for '31 Forests' indicates that '31X' must be used ' with a canopy closure of 30 % at least '. So determining if there is a mass of forest trees with crown cover density of above or less than 30% is fundamental to assign 31X classes. Moreover, tree crown cover threshold is one of the main points in any forest nomenclature definition.

However, it is precisely this 30% minimum threshold one of the problems encountered when trying to compare CLC databases with other land cover inventories with different tree crown cover threshold. So, adding this parameter or descriptor to CLC polygons is a major improvement to move on comparability and thematic interoperability. In many cases, this parameter can be added to the CLC database from more detailed forest inventories (national or regional) or may be obtained by incorporating new technologies for detecting tree cover (e.g. detection and classification of LIDAR data or very high resolution remote or aerial imagery).

The current LC data model allows providing the CLC labels for a polygon dataset, but also, when existing, provides parametric information related to the percentage of tree crown cover in a forest area. In this context the data type ParameterType/Percentage Parameter is used, with the name 'TreeCrownCover', as shown in the following example, with labelled CLC polygons.

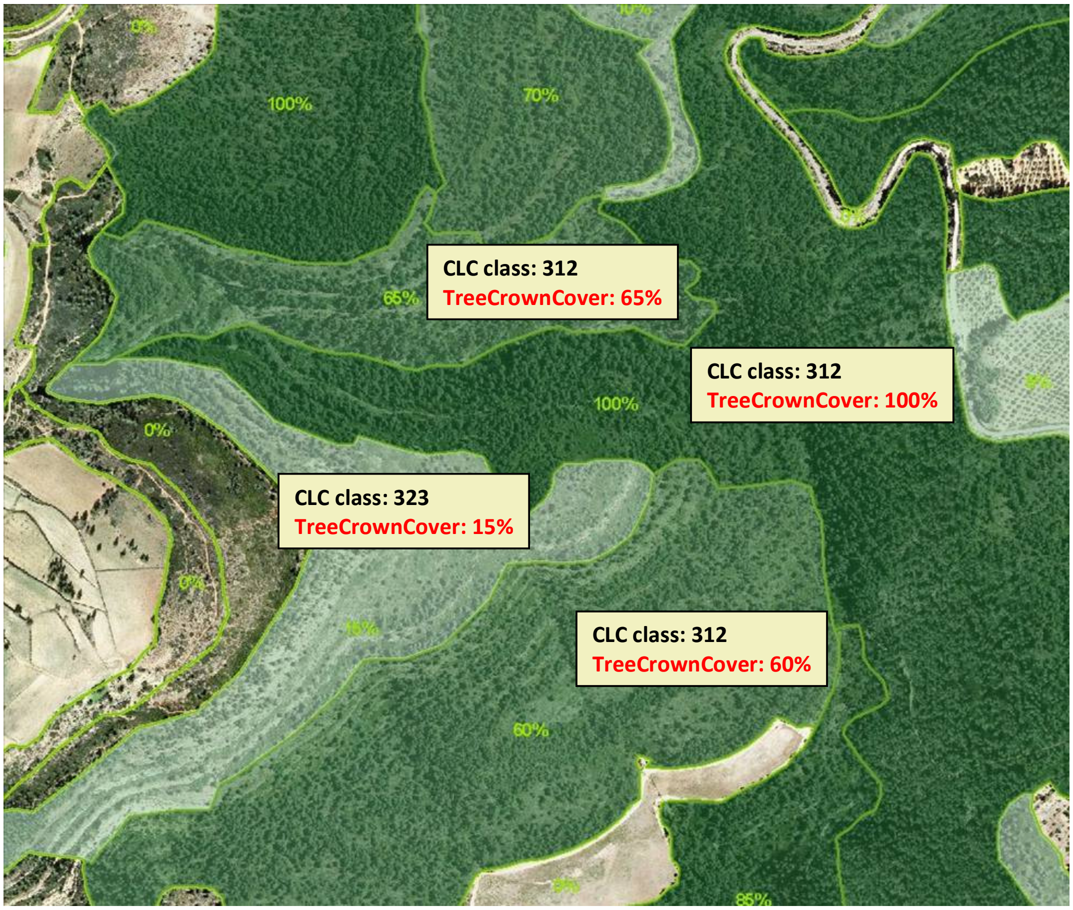


Figure 1: CLC polygons, adding tree crown cover information

Using this capability in the data model, it is possible to maintain the backward comparability between two CLC databases (e.g. CLC 1990 or 2000 with the future CLC databases), but also compare a CLC inventory with another land cover database, which is using a different definition for 'forest'.

Example 2: Presence Parameter

In this example, it can be seen how several land cover polygons with crops, from SIOSE, are being characterized by adding parametric information directly related with the land cover class in the polygon.

The data type ParameterType/PresenceParameter (name:'Irrigation'), is used here to characterize if the crop is being irrigated or not, at the observation time. SIOSE class 'crops' is a land cover parent class, grouping information related to general agricultural terms, such as 'Irrigation'. There are several land cover classes which inherit properties and attributes from the parent land cover class 'crops' (in the following example, 'herbaceous crops', 'non citrus fruit trees', and 'pastures'). It is possible to use the parameter 'Irrigation' both for analyze general crop data, and also when more specific information is needed, for example, when asking about the area covered by irrigated herbaceous crops in the Spanish territory.

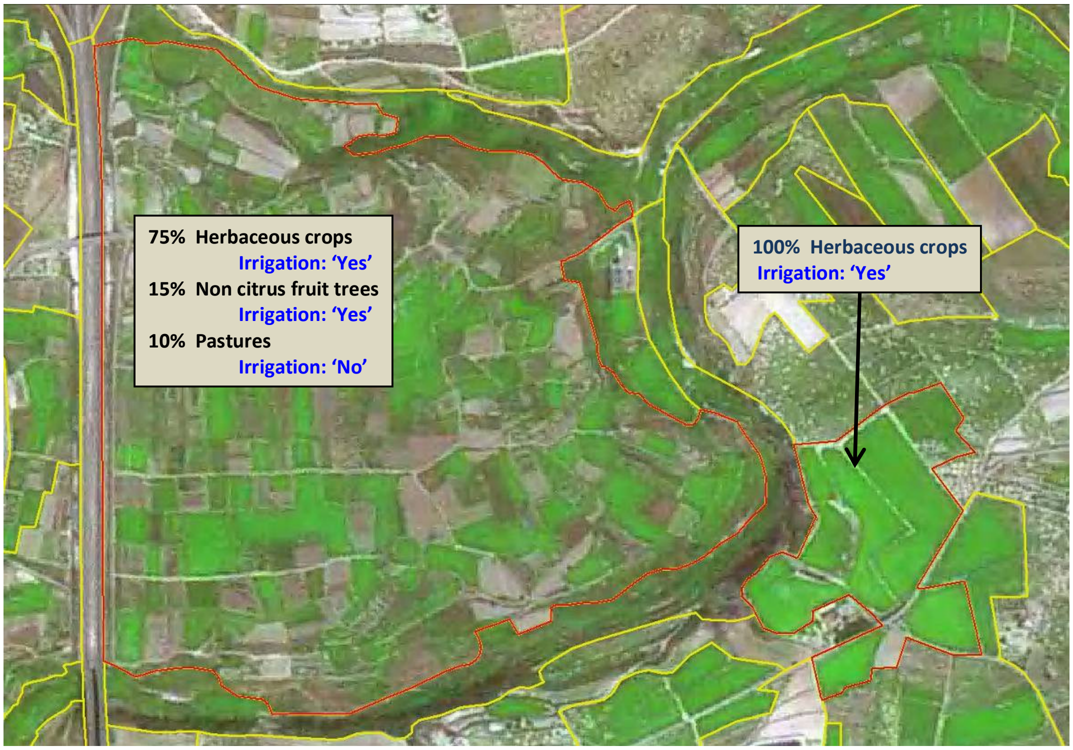


Figure 2: SIOSE polygons of an agricultural area, parameter 'Irrigation'

Example 3: Percentage Parameter

As explained in the informal description, there is a strong correlation between the existence of a certain class of land cover in the land and landscape elements that exist in the field. Some of these elements are also key elements to discriminate one category from another. For example, in the CLC database, according to a greater or lesser percentage of buildings that cover a given urban area (plus other artificial elements), it is possible to determine the corresponding CLC class (111 or 112 for example). Also the type of buildings (individual buildings, industrial buildings, etc.) determines the resulting CLC class in some cases.

Several National Land Cover databases are produced using more detailed mapped information from other topographic or thematic inventories (agricultural, forest, urban). The following figure shows how the German ATKIS Basis -DLM is used for deriving the Land Cover component in DLM (with CORINE Land Cover nomenclature) with a scale of 1:25.000, which will be used later for producing the German CLC for Europe (scale 1:100.000).

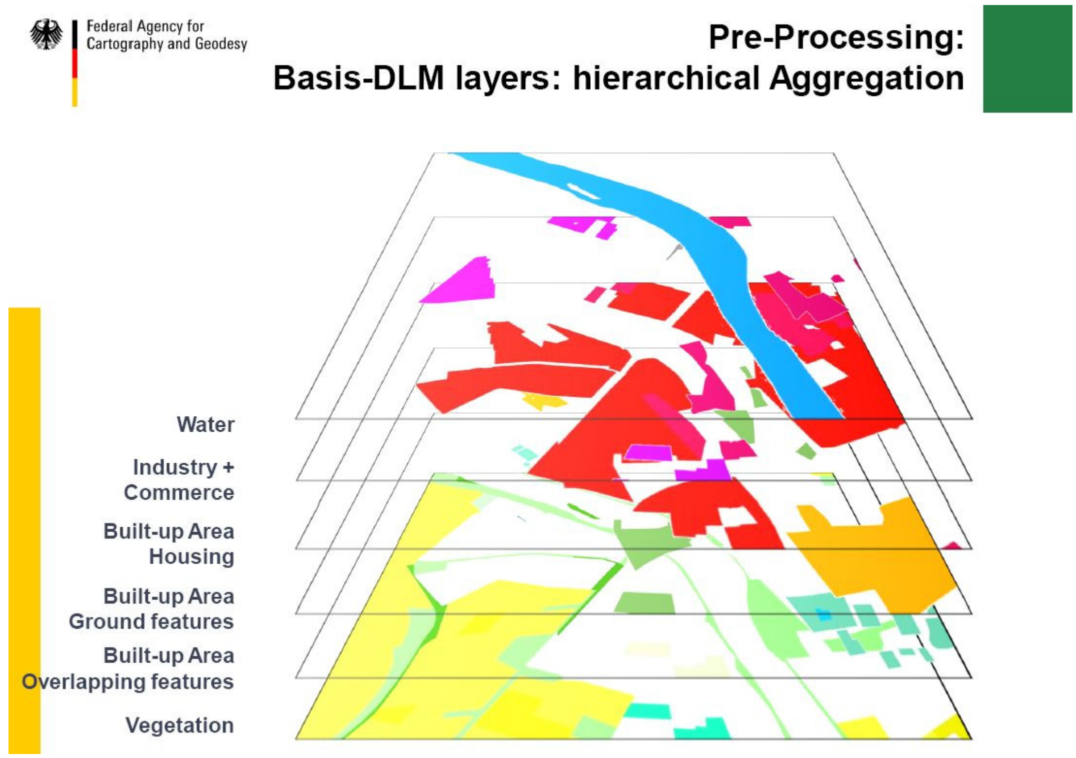


Figure 35: Hierarchical aggregation of DLM layer for producing Land Cover, courtesy of Stephan Arnold, BKG.

In the Spanish SIOSE database, buildings are also considered when describing the information in each SIOSE polygon, therefore the % of surface covered by buildings is stored in the database (with other parameters also related with buildings). So it is possible not only to use this parameter for deriving CLC labels for the SIOSE polygons (using the thresholds considered in the artificial CLC classes, for example), but also for more advanced queries to the database, such as, for example, generating thematic cloropleth maps of settlement in urban areas, using the percentage of area covered by buildings as the discrimination parameter.

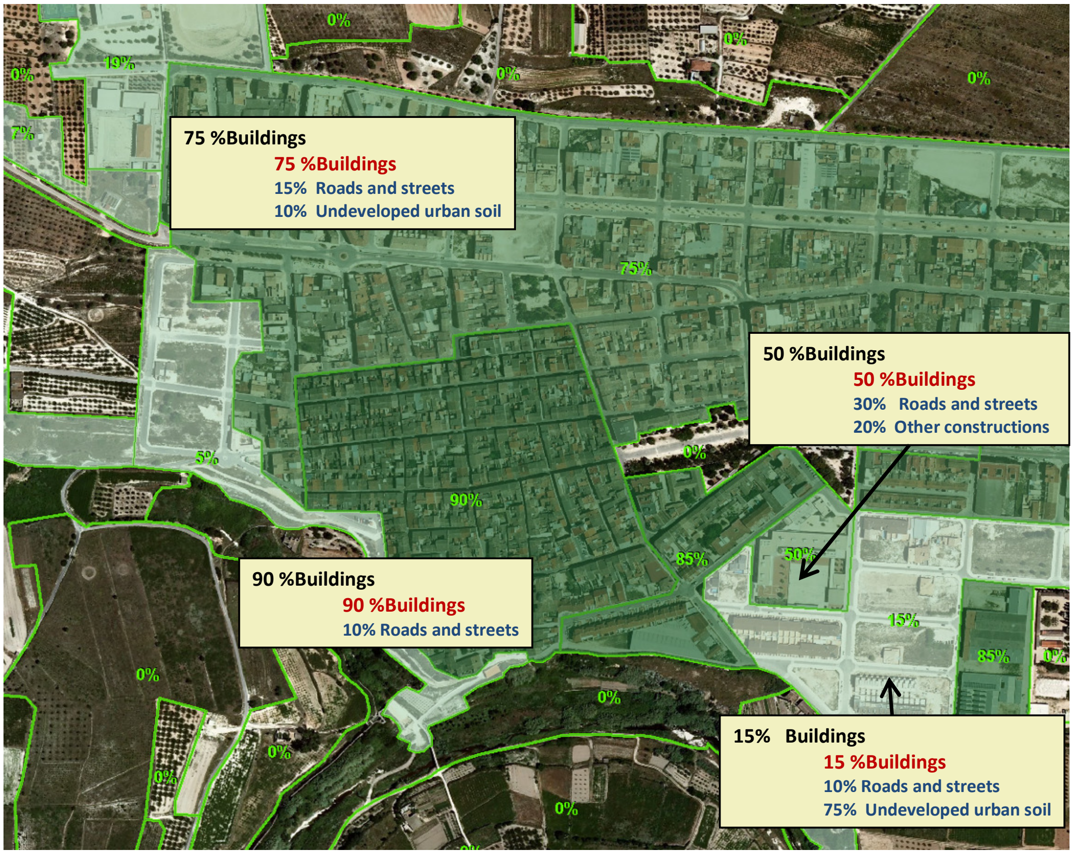


Figure 36: SIOSE polygons of an urban area, indicating the percentage of area covered by buildings

To model this essential part of the information on land cover and frame it in different INSPIRE datasets, the data type ParameterType / PercentageParameter is used. The attribute with the **name '% Buildings'** indicates the percentage of area covered by buildings in each SIOSE polygon, and can be used to produce the consequent SIOSE cloropleth map with this variable.

Example 4: Countable Parameter

When analyzing the need for water and nutrients in a plantation of tree crops, such as olive groves, is essential to consider the spacing of trees (and therefore the number of trees for a particular crop area). It is therefore a key parameter when characterizing an olive grove, both in terms of water needs, and in terms of agricultural production.

To provide this information, the data type ParameterType/CountableParameter can be used, with the name 'SpacingOfTrees', indicating the spacing of trees, as can be seen in the following SIOSE example.

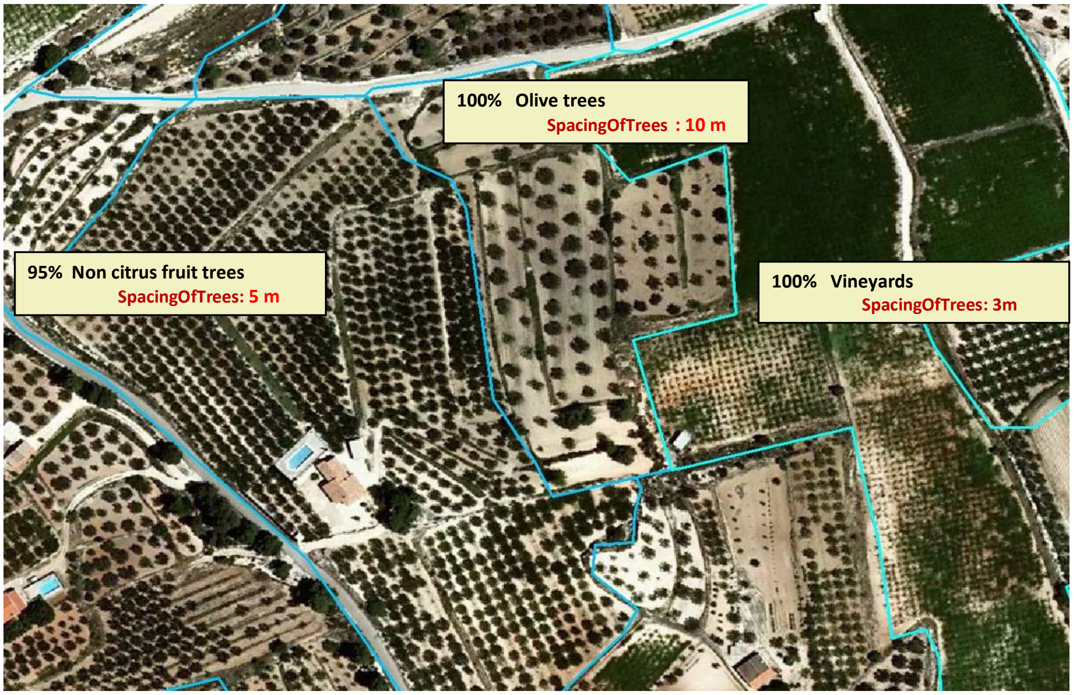


Figure 3: SIOSE polygons of an agricultural area, with the parameter related to the spacing of trees

Example 5: Percentage Parameter

According the examples of DLM-DE German or Spanish SIOSE, it is possible to obtain information on the density of a biophysical parameter, such as soil sealing, analyzing the contributions of the artificial landscape elements (buildings and structures, roads , etc.) for each occurrence of a specific geographical land cover. This possibility, as shown in the next examples, is a consequence of producing land cover information from more detailed geographical information that already exists in other geographic databases.

It is also possible to access such information with high biophysical component, and therefore obtainable from remote sensing data, from semi-automatic classification of satellite images. This is the case of High Resolution Layers proposed for GMES Land Services. These services, in fact, are considered efficient and homogeneous information to improve European databases, such as CORINE Land Cover, when there is no possibility to use detailed National inventories, according to the European requirements. Therefore, CORINE polygons for example, in addition to its class label, may have additional parametric information such as soil sealing, water content in the ground at a certain date, etc. Other very important advantage of using this density layers is monitoring the land cover change dynamics, which enable to focus on the changing area when updating land cover databases.

In this context the data type ParameterType/PercentageParameter is used again. In the next SIOSE example, this percentage parameter (**name: '% Soil Sealing'**) is used for describing and labelling the soil sealing density in each SIOSE polygon:

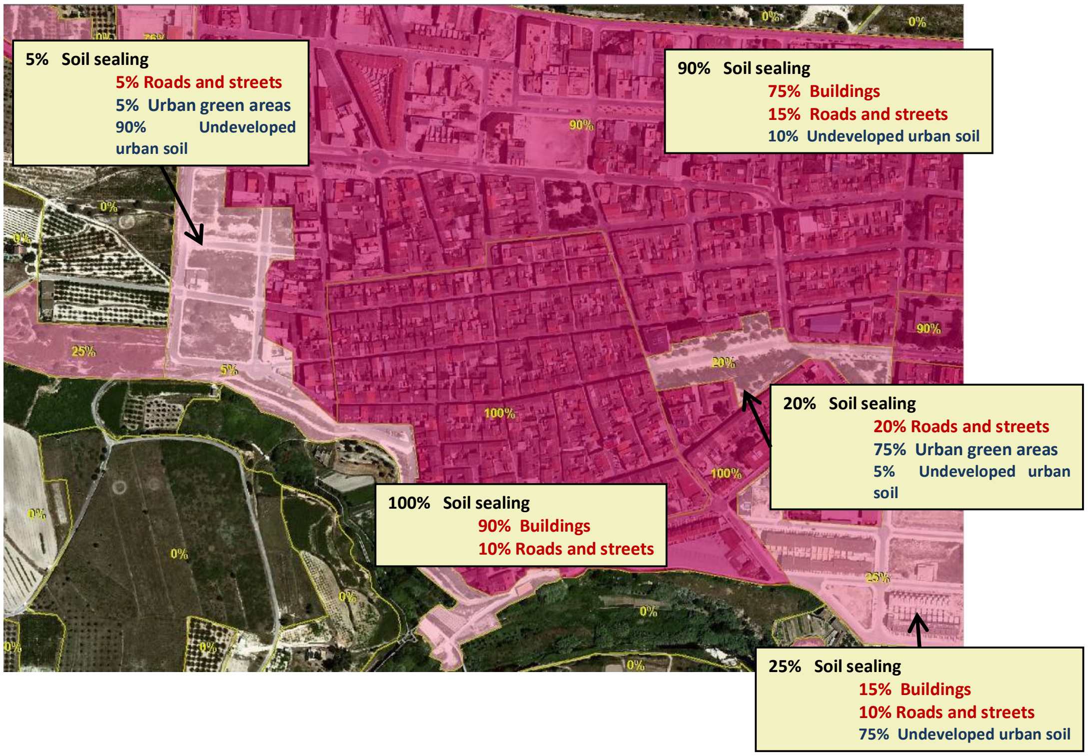


Figure 4: SIOSE polygons of an urban area, indicating the percentage of soil sealed

Annex F
(informative)
Example of legends for portrayal rules

F.1 CORINE Land Cover 2000 legend

The following table is extract *from http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2000-clc2000-100-m-version-9-2007/*.

Table 26 : CLC 2000 legend

GRID_CODE	LEVEL1	LEVEL2	LEVEL3	CLC_CODE	LABEL1	LABEL2	LABEL3	RGB
1	1	1	1	111	Artificial surfaces	Urban fabric	Continuous urban fabric	230-000-077
2	1	1	2	112	Artificial surfaces	Urban fabric	Discontinuous urban fabric	255-000-000
3	1	2	1	121	Artificial surfaces	Industrial, commercial and transport units	Industrial or commercial units	204-077-242
4	1	2	2	122	Artificial surfaces	Industrial, commercial and transport units	Road and rail networks and associated land	204-000-000
5	1	2	3	123	Artificial surfaces	Industrial, commercial and transport units	Port areas	230-204-204
6	1	2	4	124	Artificial surfaces	Industrial, commercial and transport units	Airports	230-204-230
7	1	3	1	131	Artificial surfaces	Mine, dump and construction sites	Mineral extraction sites	166-000-204
8	1	3	2	132	Artificial surfaces	Mine, dump and construction sites	Dump sites	166-077-000
9	1	3	3	133	Artificial surfaces	Mine, dump and construction sites	Construction sites	255-077-255
10	1	4	1	141	Artificial surfaces	Artificial, non-agricultural vegetated areas	Green urban areas	255-166-255
11	1	4	2	142	Artificial surfaces	Artificial, non-agricultural vegetated areas	Sport and leisure facilities	255-230-255
12	2	1	1	211	Agricultural areas	Arable land	Non-irrigated arable land	255-255-168
13	2	1	2	212	Agricultural areas	Arable land	Permanently irrigated land	255-255-000
14	2	1	3	213	Agricultural	Arable land	Rice fields	230-

					areas			230-000
15	2	2	1	221	Agricultural areas	Permanent crops	Vineyards	230-128-000
16	2	2	2	222	Agricultural areas	Permanent crops	Fruit trees and berry plantations	242-166-077
17	2	2	3	223	Agricultural areas	Permanent crops	Olive groves	230-166-000
18	2	3	1	231	Agricultural areas	Pastures	Pastures	230-230-077
19	2	4	1	241	Agricultural areas	Heterogeneous agricultural areas	Annual crops associated with permanent crops	255-230-166
20	2	4	2	242	Agricultural areas	Heterogeneous agricultural areas	Complex cultivation patterns	255-230-077
21	2	4	3	243	Agricultural areas	Heterogeneous agricultural areas	Land principally occupied by agriculture, with significant areas of natural vegetation	230-204-077
22	2	4	4	244	Agricultural areas	Heterogeneous agricultural areas	Agro-forestry areas	242-204-166
23	3	1	1	311	Forest and semi natural areas	Forests	Broad-leaved forest	128-255-000
24	3	1	2	312	Forest and semi natural areas	Forests	Coniferous forest	000-166-000
25	3	1	3	313	Forest and semi natural areas	Forests	Mixed forest	077-255-000
26	3	2	1	321	Forest and semi natural areas	Scrub and/or herbaceous vegetation associations	Natural grasslands	204-242-077
27	3	2	2	322	Forest and semi natural areas	Scrub and/or herbaceous vegetation associations	Moors and heathland	166-255-128
28	3	2	3	323	Forest and semi natural areas	Scrub and/or herbaceous vegetation associations	Sclerophyllous vegetation	166-230-077
29	3	2	4	324	Forest and semi natural areas	Scrub and/or herbaceous vegetation associations	Transitional woodland-shrub	166-242-000
30	3	3	1	331	Forest and semi natural areas	Open spaces with little or no vegetation	Beaches, dunes, sands	230-230-230
31	3	3	2	332	Forest and semi natural areas	Open spaces with little or no vegetation	Bare rocks	204-204-204
32	3	3	3	333	Forest and semi natural areas	Open spaces with little or no vegetation	Sparsely vegetated areas	204-255-204

33	3	3	4	334	Forest and semi natural areas	Open spaces with little or no vegetation	Burnt areas	000-000-000
34	3	3	5	335	Forest and semi natural areas	Open spaces with little or no vegetation	Glaciers and perpetual snow	166-230-204
35	4	1	1	411	Wetlands	Inland wetlands	Inland marshes	166-166-255
36	4	1	2	412	Wetlands	Inland wetlands	Peat bogs	077-077-255
37	4	2	1	421	Wetlands	Maritime wetlands	Salt marshes	204-204-255
38	4	2	2	422	Wetlands	Maritime wetlands	Salines	230-230-255
39	4	2	3	423	Wetlands	Maritime wetlands	Intertidal flats	166-166-230
40	5	1	1	511	Water bodies	Inland waters	Water courses	000-204-242
41	5	1	2	512	Water bodies	Inland waters	Water bodies	128-242-230
42	5	2	1	521	Water bodies	Marine waters	Coastal lagoons	000-255-166
43	5	2	2	522	Water bodies	Marine waters	Estuaries	166-255-230
44	5	2	3	523	Water bodies	Marine waters	Sea and ocean	230-242-255
48	9	9	9	999	NODATA	NODATA	NODATA	
49	9	9	0	990	UNCLASSIFIED	UNCLASSIFIED LAND SURFACE	UNCLASSIFIED LAND SURFACE	
50	9	9	5	995	UNCLASSIFIED	UNCLASSIFIED WATER BODIES	UNCLASSIFIED WATER BODIES	230-242-255

F.2 CORINE Land Cover 2006 legend

The following table is extract from *http://www.eea.europa.eu/data-and-maps/data/clc-2006-vector-data-version*.

Table 27 : CLC 2006 legend

GRID_CODE	CLC_CODE	LABEL1	LABEL2	LABEL3	RGB
1	111	Artificial surfaces	Urban fabric	Continuous urban fabric	230-000-

					077
2	112	Artificial surfaces	Urban fabric	Discontinuous urban fabric	255-000-000
3	121	Artificial surfaces	Industrial, commercial and transport units	Industrial or commercial units	204-077-242
4	122	Artificial surfaces	Industrial, commercial and transport units	Road and rail networks and associated land	204-000-000
5	123	Artificial surfaces	Industrial, commercial and transport units	Port areas	230-204-204
6	124	Artificial surfaces	Industrial, commercial and transport units	Airports	230-204-230
7	131	Artificial surfaces	Mine, dump and construction sites	Mineral extraction sites	166-000-204
8	132	Artificial surfaces	Mine, dump and construction sites	Dump sites	166-077-000
9	133	Artificial surfaces	Mine, dump and construction sites	Construction sites	255-077-255
10	141	Artificial surfaces	Artificial, non-agricultural vegetated areas	Green urban areas	255-166-255
11	142	Artificial surfaces	Artificial, non-agricultural vegetated areas	Sport and leisure facilities	255-230-255
12	211	Agricultural areas	Arable land	Non-irrigated arable land	255-255-168
13	212	Agricultural areas	Arable land	Permanently irrigated land	255-255-000
14	213	Agricultural areas	Arable land	Rice fields	230-230-000
15	221	Agricultural areas	Permanent crops	Vineyards	230-128-000
16	222	Agricultural areas	Permanent crops	Fruit trees and berry plantations	242-166-077
17	223	Agricultural areas	Permanent crops	Olive groves	230-166-000
18	231	Agricultural areas	Pastures	Pastures	230-230-077
19	241	Agricultural areas	Heterogeneous agricultural areas	Annual crops associated with permanent crops	255-230-166
20	242	Agricultural areas	Heterogeneous agricultural areas	Complex cultivation patterns	255-230-077
21	243	Agricultural areas	Heterogeneous agricultural areas	Land principally occupied by agriculture, with significant areas of natural	230-204-077

				vegetation	
22	244	Agricultural areas	Heterogeneous agricultural areas	Agro-forestry areas	242-204-166
23	311	Forest and semi natural areas	Forests	Broad-leaved forest	128-255-000
24	312	Forest and semi natural areas	Forests	Coniferous forest	000-166-000
25	313	Forest and semi natural areas	Forests	Mixed forest	077-255-000
26	321	Forest and semi natural areas	Scrub and/or herbaceous vegetation associations	Natural grasslands	204-242-077
27	322	Forest and semi natural areas	Scrub and/or herbaceous vegetation associations	Moors and heathland	166-255-128
28	323	Forest and semi natural areas	Scrub and/or herbaceous vegetation associations	Sclerophyllous vegetation	166-230-077
29	324	Forest and semi natural areas	Scrub and/or herbaceous vegetation associations	Transitional woodland-shrub	166-242-000
30	331	Forest and semi natural areas	Open spaces with little or no vegetation	Beaches, dunes, sands	230-230-230
31	332	Forest and semi natural areas	Open spaces with little or no vegetation	Bare rocks	204-204-204
32	333	Forest and semi natural areas	Open spaces with little or no vegetation	Sparsely vegetated areas	204-255-204
33	334	Forest and semi natural areas	Open spaces with little or no vegetation	Burnt areas	000-000-000
34	335	Forest and semi natural areas	Open spaces with little or no vegetation	Glaciers and perpetual snow	166-230-204
35	411	Wetlands	Inland wetlands	Inland marshes	166-166-255
36	412	Wetlands	Inland wetlands	Peat bogs	077-077-255
37	421	Wetlands	Maritime wetlands	Salt marshes	204-204-255
38	422	Wetlands	Maritime wetlands	Salines	230-230-255
39	423	Wetlands	Maritime wetlands	Intertidal flats	166-166-230
40	511	Water bodies	Inland waters	Water courses	000-204-242
41	512	Water bodies	Inland waters	Water bodies	128-242-230

42	521	Water bodies	Marine waters	Coastal lagoons	000-255-166
43	522	Water bodies	Marine waters	Estuaries	166-255-230
44	523	Water bodies	Marine waters	Sea and ocean	230-242-255

Annex G
(informative)
Existing land cover classification systems and LCML

G.1 LCML

ISO 19144-2 specifies a Land Cover Meta Language (LCML) expressed as a UML metamodel that allows different land cover classification systems to be described based on physiognomic-structural aspects and so defines a common reference structure for the comparison and integration of data from any generic land cover classification system (lccs). It also improves the harmonization and integration of spatial data sets with legends or nomenclatures developed from different land cover classification systems. Indeed, such systems are well established in ongoing mapping projects and cannot be easily changed.

This LCML approach provides a rigorous logical framework for the description of any land cover class. The key to describe any international, national or multi-national lccs in terms of the LCML is to use a compliant parametric approach, circumventing the traditional obstacles such as complex definitions, prefixed ranges of values and specific classification rules.

The main drawback of the LCML harmonization approach is that non physiognomic classification aspects of a lccs –such as land use- are not fully recognized by the LCML and that information will be partially lost in translation. Considering that the INSPIRE land cover theme is defined only by "physical and biological cover of the earth's surface", the resulting LCML translation can be considered as filtering the "pure" land cover information from an existing data set.

G.2 Translating a lccs database into its LCML version

The LCML provides a general framework of rules using independent diagnostic criteria. These lead to land cover metalanguage descriptor objects that are defined by a combination of a pre-defined set of land cover metalanguage elements, divided in two categories

- "basic metalanguage-elements" constitute the main physiognomic aspects of biotic and abiotic cover features organized in layers, for instance for biotic features trees, shrubs, herbaceous vegetation etc., and
- "metalanguage-element properties" further define the physiognomic aspect of each basic type metalanguage-element.

In the LCML model, further detail of the resulting land cover classes may be achieved by adding **optional** descriptive metalanguage-element characteristics not directly related to the physiognomic/structural characterization of the land cover but which assist in better describing the land cover class:

- LC_ElementCharacteristics" may be applied to a single basic metalanguage-element
- LC_ClassCharacteristics" relate to a whole Land Cover class, defined as the combination of single or multiple strata of single or multiple basic meta-elements

Figures 39 and 40 illustrate these basic elements, properties and characteristics through a practical example.

The metalanguage generates mutually exclusive land cover classes, with specific rules to deal with the all functional elements of the language (basic metalanguage-elements ,-properties and their relationships) and the different strata.

The process of translating an existing dataset for a given lccs into a dataset expressed in LCML involves 4 activities:

1. Perform a semantic analysis on the lccs class to understand its physiognomy and structure.
2. Design the appropriate LCML class for each lccs class, combining the applicable basic metalanguage-elements, metalanguage-element properties and optional metalanguage-element characteristics
3. Repeat activity 1 and 2 for each class in the original lccs
4. Apply class codes and parameter values at feature or polygon level.

G.3 Examples for CORINE LAND COVER

The above process can be illustrated with two examples of the CORINE Land Cover Classes.

CLC 213: Rice fields

The semantic analysis starts from the original CLC Description: "Land prepared for rice cultivation. Flat surfaces with irrigation channels. Surfaces periodically flooded."

The LCML translation logic of this description would consider two separate layers; one biotic and one abiotic. The biotic layer consist of the LCML element " herbaceous growth forms" "cultivated" with specific floristic name "rice". The second layer is composed by the element "natural water" "fresh" with a "persistence period" extend to the whole cultivation time.

The concept of this class CLC 213 allows for an unambiguous translation into LCML without loss of information. "Graminae", "Water Body" and "Periodic variation" are its basic elements, each with properties; "Cultivated and managed vegetation" and "Floristic aspect" are the optional characteristics of the "Graminae". ; "Water Salinity" is a characteristic of "Water Body" (Figure 40).

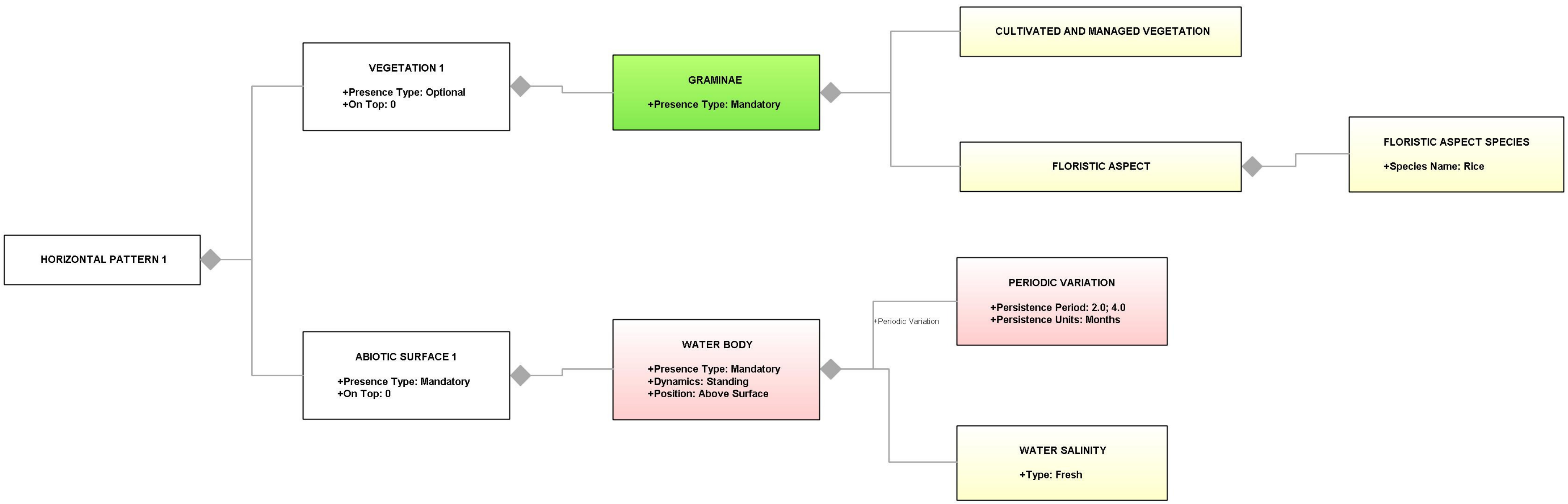


Figure 39: UML representation of the selection of LCML elements for translating CLC 213 Rice fields.

```
  <?xml version="1.0" encoding="UTF-8" standalone="no" ?>
- <LC_Legend xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" id="1" lccs_sort="0" uuid="be10d990-
     5660-11e1-ab66-001999808c3f" xsi:noNamespaceSchemaLocation="INSPIRE_Annex_LCML.xsd"
     xsi:type="LC_Legend">
  <name>New Legend</name>
  <description>Describe the legend</description>
- <elements>
- <LC_LandCoverClass id="2" uuid="c98596d0-5660-11e1-ab66-001999808c3f" xsi:type="LC_LandCoverClass">
  <name>Rice Fields</name>
  <description>Describe the land cover class</description>
  <map_code>lcc1</map_code>
- <elements>
- <LC_HorizontalPattern id="3" uuid="cd9b9440-5660-11e1-ab66-001999808c3f"
     xsi:type="LC_HorizontalPattern">
  <name>Horizontal Pattern 1</name>
  <description>Describe the horizontal pattern</description>
- <elements>
- <LC_Stratum id="4" ontop="0" uuid="1d273cc0-5662-11e1-ab66-001999808c3f" xsi:type="LC_Stratum">
  <name>Vegetation 1</name>
  <description>Vegetation 1</description>
- <elements>
- <LC_LandCoverElement id="5" uuid="233d9e10-5662-11e1-ab66-001999808c3f" xsi:type="LC_Graminae">
  <name>Graminae</name>
  <description>Describe a vegetation element of graminae type</description>
- <elements>
- <LC_Characteristic id="6" uuid="27a47287-43af-11de-bbd4-000cf147c442"
     xsi:type="LC_CultivatedAndManagedVegetation">
  <name>Cultivated And Managed Vegetation</name>
  <description>Describe the cultivated and managed vegetation</description>
  </LC_Characteristic>
- <LC_Characteristic id="7" uuid="27a47295-43af-11de-bbd4-000cf147c442" xsi:type="LC_FloristicAspect">
  <name>Floristic Aspect</name>
  <description>Describe the floristic aspect</description>
- <elements>
- <LC_Characteristic id="8" uuid="2c121880-5663-11e1-ab66-001999808c3f"
     xsi:type="LC_FloristicAspectSpecies">
  <name>Floristic Aspect Species</name>
  <description>Describe the floristic aspect species</description>
  <species_name>Rice</species_name>
……. /
 …….
- <LC_Stratum id="9" ontop="0" uuid="6ceae530-5663-11e1-ab66-001999808c3f" xsi:type="LC_Stratum">
  <name>Abiotic Surface 1</name>
  <description>Stratum 1</description>
- <elements>
- <LC_LandCoverElement id="A" uuid="8293f070-5663-11e1-ab66-001999808c3f" xsi:type="LC_WaterBody">
  <name>Water Body</name>
  <description>Describe a water body surface element</description>
- <elements>
- <LC_Characteristic id="B" uuid="27a2ebe2-43af-11de-bbd4-000cf147c442" xsi:type="LC_WaterSalinity">
  <name>Water Salinity</name>
  <description>Describe the water salinity characteristic</description>
  <type>Fresh</type>
</LC_Characteristic>
  </elements>
  <presence_type>Mandatory</presence_type>
- <periodic_variation id="C" uuid="860747b0-a39a-11e0-bd59-000cf147c442" xsi:type="LC_PeriodicVariations">
  <name>Periodic Variations</name>
  <description>Contains the elements of Periodic Variation</description>
- <elements>
- <LC_PeriodicVariation id="D" uuid="e3d58d80-5663-11e1-ab66-001999808c3f" xsi:type="LC_PeriodicVariation">
  <name>Periodic Variation</name>
  <description>Describe a periodic variation element</description>
  <persistence_units>Months</persistence_units>
  <persistence_period max="4.0" min="2.0" />
  </LC_PeriodicVariation>
  </elements>
  </periodic_variation>
  <dynamics>Standing</dynamics>
  <position>Above Surface</position>
  </LC_LandCoverElement>
…../…..
```


Figure 40: Extracts of the XML created by FAO LCCS3 software after translating "CLC 213 Rice fields"

<u>CLC class 243: Land principally occupied by agriculture, with significant areas of natural vegetation</u>

The semantic analysis starts from the original CLC Description: "Areas principally occupied by agriculture, interspersed with significant natural areas". This description, as is, could represent either of two spatial concepts: the CLC class reflects either a functional (occupation) unit or a “spatial heterogeneity” due to the scale of interpretation. As the rigorous LCML syntax separates between the two cases, a translator's choice is needed. Here, the first case has been considered and the class has been translated as “functional unit”. Still, a minimum percentage of allowed proportion between the different features composing this class should be considered. We acknowledge that such translation information on the land cover types and the correspondent percentage can often be derived from CLC4/CLC5 classes or the national mapping documentation

The LCML translation approach considers three horizontal patterns for CLC243. The first one composed by “herbaceous’ or “woody” crops, the second by natural vegetation, the third one by “water bodies”. However more horizontal patterns could have been added if wetlands and urban settlements must be considered. Considering class 243 as a “functional unit”, foresees clear and fixed ranges of proportional percentage between the different patterns.

The concept of this class CLC 243 carries some inherent vagueness that reflects in the LCML translation. Better ancillary information would allow a better translation, but "off the shelf", the class would be translated into three layers with arbitrary proportions.

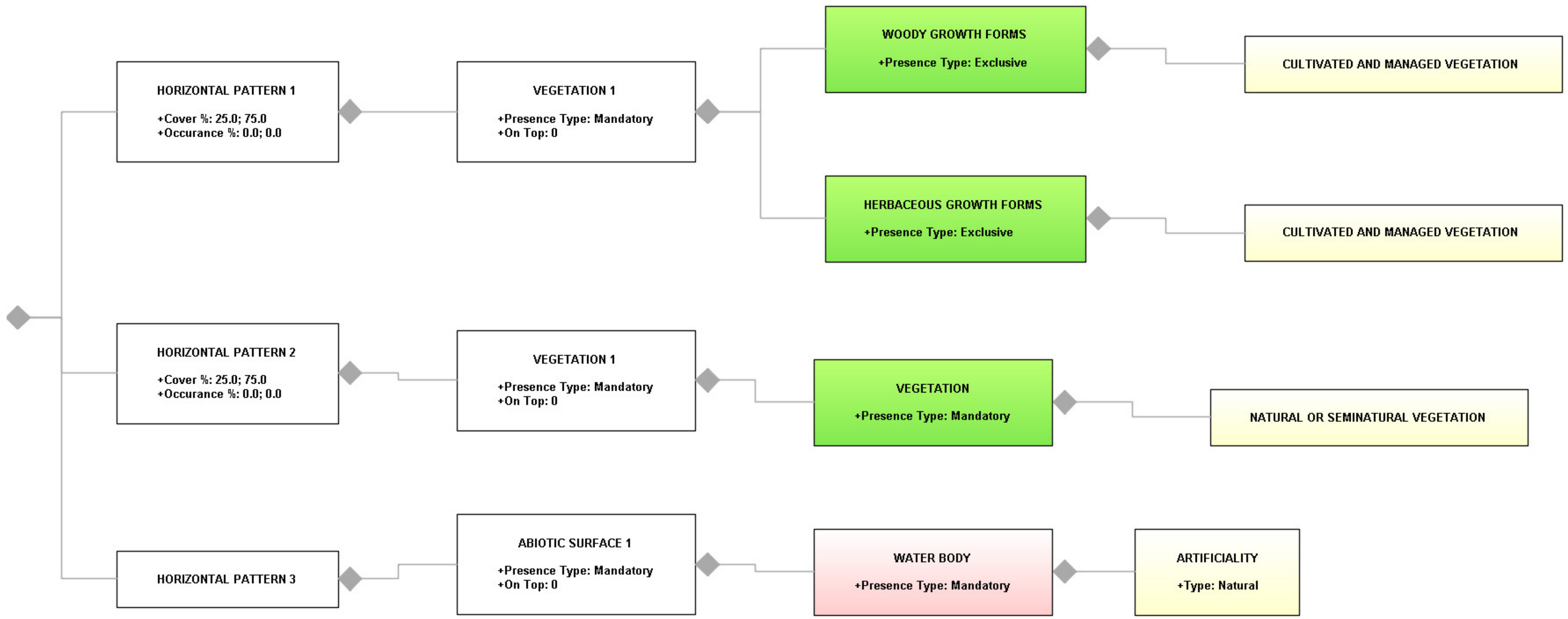


Figure 41: Selection of LCML elements for translating "CLC 243 Land principally occupied by agriculture, with significant areas of natural vegetation"

The above CLC examples are given only for illustrating this INSPIRE data specification on land cover. It is the responsibility of each owner of a regional Land Cover classification (e.g. CLC4 – CLC5) or the European Environment Agency (CLC3) to describe the classes in LCML terms.

G.4 Schema transformation by "semantic translation"

When the original lccs, the CLC legend or the optional INSPIRE land cover nomenclature are described in LCML terms, then the linguistically concise nature of LCML could support highly automated translation of the datasets from one classification into the other.
This would involve a schema transformation service to map the local implementation in a target schema. Such a service transforms data from the local schema towards the target schema according to appropriate schema mapping rules compliant with INSPIRE-level specifications. What follows focuses on the semantic aspects: i.e. the translation of the classification codes of the data set.
Such semantic schema transformation process from a local model to the LCML can be performed following a two-fold approach (Figure 42).

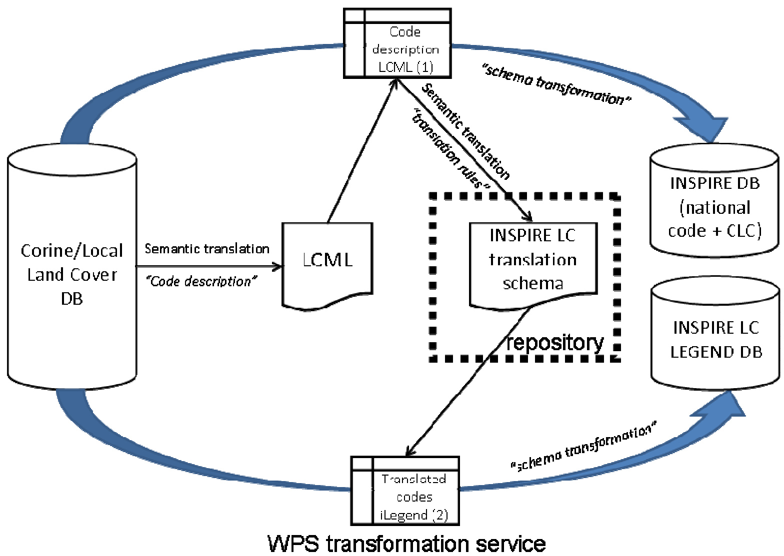


Figure 5: Schema transformation by semantic translation towards an INSPIRE compliant Land Cover Meta Language.

Any schema mapping between the MS application schema expressed in LCML and target schema (= INSPIRE LCML) should be performed at the semantic level first. As seen previously, this requires a good knowledge of the two schemas and requires decisions in order to split more general classes defined at the MS level into more detailed LCML classes, or on the contrary to aggregate or generalize classes (not polygons) where needed - all decisions are based on the appropriate classifiers or properties. The schema mapping expresses relations between source and target schema to be used for the data transformation itself. That mapping can be prepared manually, half-automatically or fully automatically as long as it is precise, unambiguous and complete and no data loss or inaccuracies occurs. The mapping rules between local and target schema are stored in the mapping repository, accessible for further data transformation requests.
Second, a transformation service needs to be set up to execute the data transformation based on that appropriate schema mapping available in the repository. The Commission Regulation for INSPIRE Transformation Services describes the Web Service interface on the abstract level and recommends to make an accessible (transformation) service available wrapped in an OCG Web Processing Service (WPS). From the software point of view, such service-based transformation could be implemented using any number of proprietary tools.
Such future automation remains challenging. Critical elements include the precise comprehension of the local land cover classification, its correct LCML description, the correct description of a target schema and a tool for the LCML schema mapping service for schema and data transformation. Still, each of these challenges has already been individually addressed during scientific projects.

Annex H
(informative)
INSPIRE "Pure Land Cover Components"

As a proposal for a European harmonized description of land cover information, the TWG LC suggests a collection of *Pure Land Cover Components* (PLCC) in the form of a code list. The purpose of this code list is not to provoke any extra mapping activities, but rather to give the data providers the opportunity to describe their data from the pure land cover point of view in a harmonized way and thus streamline comparison of outputs from various land monitoring initiatives. The concept behind that code list is the attempt to leave out any land use aspects as far as possible in describing the landscape. Together with the Land Use information which is stored in the HILUCS classes (developed by the TWG LU), a complete and complementary abstraction of landscape can be modeled.
Behind setting up the PLCC code list there has been going on a process of developing a modified land cover data model, which aims at suiting the future requirements of European land cover monitoring on continental as well as on national and regional level.
It also turned out that during the development of the data model behind the PLCC code list there is a strong comparability with the approach used in the ISO standard 19144-2 (Land Cover Meta Language LCML).
The PLCC code list is not meant to be applied as a mandatory way to describe land cover but as a recommendation which aims more at a good practice to increase interoperability and applicability. Furthermore, the forming of PLCC is open for future modifications, after incoming contributions of a broader user feedback and after data providers have collected experiences with it.

Minimum mapping unit (MMU):

The PLCC code list is meant to be scale and MMU-independent. The TWG LC is aware of the fact, that for the interest of the end user it might be of advantage to implement a certain minimum mapping unit, which could ensure a better geometrical comparability between neighboring countries, especially across borders, but also between different databases or nomenclatures. However, by introducing a certain MMU this could be a limiting factor to the philosophy of INSPIRE to give access to all relevant spatial data sets without limitations of scale. Similar to a requirement regarding data quality, also a required MMU could complicate the providing of land cover information or even exclude certain LC data from the INSPIRE community. Therefore it is not considered to demand a commonly defined MMU while applying the PLCC code list. Instead, each data provider is asked to enter his data as is, according to the MMU which result from his own mapping rules and technical guidelines.

Usage of the code list:

a) It can be used simply as a kind of nomenclature, and the LC data is entered into the model by application of the code list like a categorization and attaching only one single code from the code list to a certain land cover unit.

b) The code list can be used in a descriptive way, which gives room for attaching more than one land cover component from the code list to a single land cover unit, expressing that on a particular spot in landscape there exists more than one land cover component. This application of the code list would go in-line with the idea of the descriptive approach of the ISO standard 19144-2 (LCML).

c) Use the code list like mentioned in b), in a more elaborated way by not only mentioning more than one land cover component to be attached to a certain land cover unit, but also to enter a percentage value and structure description (like in LCML horizontal pattern/vertical stratum), which indicates the relative fraction and spatial composition of the considered land cover component inside the area extent of a definite land cover unit.

Attribution of land cover components:

The TWG has also discussed the issue of the possibility to open up the land cover component for further characterization through attribution (e.g. giving information about the water regime of an inland water course: perennial, periodic, episodic). However, it was decided to not go into the level of attribution. Hence, attribution would then be in the hands of the land monitoring community to bring forward the future process of data model development, but which is supposed to take place outside the INSPIRE legislative process.

Seasonality/Time Variability:

The PLCC code list in some cases may seem to be not suitable for tackling seasonality. This might be solved in a later stage of development with attributes and properties. If the data provider wants to describe seasonal variability in a proper way, he should then use LCML or any other classification system or nomenclature which suits better for that issue. Alternatively the data provider could enter a second observation date (e.g. during rainy season and dry season) connected to a second land cover component to express a high variability of land cover characteristic in the landscape due to seasonal effects (e.g. PLCC 003 and 009).

The harmonized characterization of seasonality aspects of land cover features, for example in vegetation or water bodies is being actually addressed within the works of the EAGLE Group, considering the past and present activities within GMES Land services (CLC & High Resolution Layers), and new initiatives for land cover monitoring (such as the descriptive approach of seasonality in ISO standard 19144-2 or seasonal attributes in object-oriented land cover databases).

Development of the Pure Land Cover Components:

The EAGLE Group has supported the TWG LC in developing a harmonized way to describe land cover through intensive and interactive contributions, which finally resulted in this proposed code list of PLCC. Since INSPIRE is a kind of "milestone" or a step wise approach in this process, it is expected that further developments of an enhanced way to describe and model land cover information are stimulated (be it a classification system or an object-oriented approach) and carried out by initiatives like EAGLE or HELM (Harmonised European Land Monitoring).

Owner and **Home** of the PLCC code list:

The PLCC Code List will be made available as a proposal on the EAGLE-website [http://sia.eionet.europa.eu/EAGLE].

It is still to be worked out who in the future will take over the ownership of the PLCC code list.

Table 28: Pure Land Cover Component (PLCC) – code list

Code No.	Component Name	Legend Color	Color Map (R/G/B)
001	Artificial constructions		(255/99/133)
002	Consolidated bare surface		(156/156/156)
003	Unconsolidated bare surface		(204/210/165)
004	Arable land		(255/255/168)
005	Permanent woody and shrubby crops		(247/200/100)
006	Coniferous forest trees		(68/150/0)
007	Broadleaved forest trees		(0/220/0)
008	Shrubs		(150/190/0)
009	Herbaceous plants		(202/242/77)
010	Lichens and mosses		(166/255/160)
011	Wetlands and marshes		(0/214/178)
012	Organic deposits (Peatland)		(156/127/120)
013	Chemical deposits		(227/212/255)
014	Intertidal flats		(173/138/167)
015	Fresh water course		(0/190/255)
016	Fresh water bodies		(90/214/255)
017	Salt or brackish water		(0/148/194)
018	Permanent snow and ice		(180/255/255)

Explanation of the Code List and **Examples**:

Land Cover represents the biophysical state of the real landscape, which means that it consists of natural, modified and artificial objects with their physiognomic properties and their spatial relationships. The main principle of the approach that has been applied in creating this list of PLCCs has been guided by the background questions "What do I interprete as land cover in the landscape from the view above (e.g. on satellite imagery)? How is the earth´s surface covered on a single specific spot of land? How can I order the landscape elements and form them into pure land cover components, regardless of its land use aspects as far as possible.

001 Artificial constructions:

All types of artificial man-made constructions with a sealed surface. It includes

- roof covered buildings (residential, commercial, industrial, transportation (train stations and airport terminals) etc.)
- other artificial constructions (e.g. dams, water sewage plants, power plants, dump sites)
- linear constructions (e.g. railway network, road networks).

It would exclude surfaces formed by bare surfaces (rock, sand, soil) under anthropogenous influence (e.g. quarries) or other man-made artificial vegetation covers (parks/gardens).

002 Consolidated bare surface:

Any type bare surface, formed by natural material and with a solid surface. It also may have been modified through man-made processes like on extraction sites.

It includes

- solid rock surface or hard pan without any further coverage of loose material
- quarries, extraction sites of rock formations.

It would exclude artificial solid surfaces like concrete or asphalt areas as part of any man-made infrastructures, which ought to be placed under *001_Artificial constructions*. Consolidated surface neither does contain salt surface due to water evaporation, which instead is placed under *013_Chemical deposits* (see below).

003_Unconsolidated bare surface:
Any type of bare unvegetated surface, formed by natural loose materials resulting from physical sedimentary processes (fluviatile, littoral, glacial/periglacial, aeolian, gravitative slope processes etc.)
It includes
- boulders, scree, pebbles, sand, silt, clay
- any kind of mixture of the above mentioned compartments (e.g. glacial moraines)
- also semi-natural areas, with a character of fallow land apparently out of use and lacking vegetation cover.

It may also contain very sparse vegetation spots; however, sparse vegetation cover should generally be modeled as a combination of one or more vegetated components (PLCC *006 - 012*) and bare surfaces (PLCC *002, 003 or 013)*.
It would exclude Bare soil in agricultural areas, which would be part of *004_Arable Land* (see below).

004_Arable land:
Land Cover Component strongly characterized by the aspect of land use. Agriculture has always been a category difficult to describe only from a pure land cover point of view as it is characterized by regular alternation of bare soil and crop cover.
It includes
- herbaceous crops (e.g. gramineae, different types of cereals, corn, wheat, barley, etc.)
- forbs (e.g. potatoes, tomatoes, strawberries, hop etc.)
- also bare soil in arable land, which is only temporarily uncovered with crop plants.

005_Permanent woody and shrubby crops:
Any type of multi-annual or permanent crop with woody or shrubby character. Usually a kind of planting pattern can be recognised.
It includes
- any type of fruit trees plantation (apple, cherry, nuts, oranges etc.)
- mixed fruit tree growing in an orchard pattern
- olive trees
- vineyards
- berry plantations and shrubs
- tree nurseries

It would exclude hop plantation, because only the structure of the planting is permanent, but not the crop itself, which belongs to 004_Arable Land.

006_Coniferous forest trees:
It includes coniferous trees, both deciduous (e.g. the larch) and evergreen species. Dwarf trees along the tree line (where habitat climate conditions have restricting influence on the growth form of trees) in mountainous or polar regions are considered here also as trees, not as shrub.

007_Broadleaved forest trees:
Any type of broadleaved trees. It may include also palm trees or other non-coniferous tree species. Dwarf trees along the tree line (where habitat climate conditions have restricting influence on the growth form of trees) in mountainous or polar regions are considered here also as trees, not as shrub.

008_Shrubs:
Any type of vegetation with woody character (ligneous stem) and with a growth form and height between herbaceous and trees. This class also includes dwarf shrubs (e.g. Erica spp.) making up heath vegetation.

009_Herbaceous plants:
All types of gramineous and forb vegetation.
It would exclude annual gramineous vegetation as crop type (cereals, corn, grains, etc.), which is placed under *004_Arable Land*.

010_Lichens and mosses:
All types of Lichens and Mosses. Mainly they would appear in habitats with restricted growing conditions for other plant species like low temperature, lacking of sunlight, very high soil moisture, or very dry conditions etc.

Mostly they would grow in association with other vegetation types. Applying this code list, it is therefore most likely to combine them with other PLCC, other than in polar or alpine regions where it can make up homogenous land cover.

011 Wetlands and marshes:

All types of wetland, which is under the influence of very high soil moisture due to high ground water level, high precipitation rates, due to frequent flooding and/or presence of surface water, which is shallow enough to allow vegetation cover over ground. This LC component makes no difference regarding the geographic location of the marshland.

It includes

- inland marshes (fresh or salt water)
- coastal salt marshes

Marshland is wet by definition and can stand alone. However, it does not give precise information by itself about the kind of present vegetation cover but describes instead the growing condition. Therefore is should be combined with some other LC components, e.g. Herbaceous. Also it is possible to add explicitly water as a contributing layer, which indicates that the marsh contain not only water-saturated soil but is also most of the time or regularly covered with surface water (either salty, brackish, or fresh). Most marshes are covered with herbaceous plants, but it is also possible to include shrubs or trees as part of the vegetation cover by using the referring PLCC *006 - 010*.

It would exclude occasionally flooded land, which by its character belongs to other landscape types and besides the temporal presence of surface water it cannot be classified as marshland.

012 Organic deposits (peatland):

Peat is a type of organic soil composed of incomplete decayed organic material because of lack of oxygen due to water saturated ground, which leads to stepwise accumulation of biomass. The frequency of peatland is greatest in regions with very humid climate, where the precipitation is much higher than the evaporation. Mire vegetation is adapted to the harsh conditions, for example with low content of oxygen in water and often rather acid soil.

Peatland is likely to be covered on its surface with vegetation,. In most cases the peat surface itself is not visible on the imagery, but instead the typical vegetation cover which is adapted to the habitat conditions of peatland. Similar to Marshes it describes the growing conditions of vegetation on the spot. Therefore it is considered to be a separate PLCC, which is best to be used in combination with other vegetation PLCCs.

It includes

- bogs (ombrotrophic) and fens (minerotrophic),
- bare peat with no vegetation cover on surface.

This PLC component can also be combined with *011_Wetland and marshes* (to express the water saturation) or with other vegetations PLC components *006 - 010*.

013 Chemical deposits:

Complementary to the organic deposit this is the category that contains all kinds of deposits/sediments, which result from chemical processes like evaporation of salt water with residuals due to mineral crystallization processes. No differentiation is made between natural and man-made chemical deposits.

It includes

- naturally occurring salt surface
- salines (for oceanic salt extraction)
- other crystalline loose chemical residuals (e.g. lime, gypsum, soda etc.) not yet having the character of solid geolocial stone formation.

014 Intertidal flats:

Typical transitional zone between the average high tide and low tide sea water line. It is under tidal influence and mostly covered with sand or fine alluvial mud/sea ooze on the ground, being twice a day covered and uncovered with water.

It includes

- coastal intertidal mud flats (salt water)
- intertidal zone along river estuaries.

It would exclude areas with occasionally exposed sea bottoms, caused by other conditions than tide (air pressure or water level variation due to heavy winds, e.g. in the Baltic Sea). They belong to the area of open sea water.

015 Fresh water course:

Any type of inland fresh water courses with linear character and/or principally flowing water. It includes

- Rivers and Streams; under different water regimes: perennial, periodic/seasonal, episodic/irregular; under low or high seasonal variation of water level; with various water course shapes like natural/braided river, controlled/regulated/channeled rivers
- Channels (purely artificial) for navigation or irrigation purpose.

016 Fresh water bodies:

Any type of inland water bodies, which have a still character. It includes

- natural lakes and ponds,
- water reservoirs (e.g. for drinking water supply, energy production, irrigation, fire extinction etc.)

No distinction is made between natural lakes and man-made retaining water bodies.

017 Salt or brackish water:

Any kind of salty or brackish water surface, regardless of geographical location or distribution.
It includes

- open sea water,
- coastal lagoons (salt or brackish),
- salty or brackish estuary zones of river mouths,
- inland salt or brackish water (e.g. in areas of geothermal activities or salty steppe lakes where evaporation is higher than water inflow.)

018 Permanent snow and ice:

It includes

- glaciers
- snow fields, which do not melt during warm summer period in between two winter seasons.

OAQ – Occasionally asked Questions/ Comments:

Why there is no Pure Land Cover Component like mixed forest?

Answer: it can be modeled with the two components *006_TConiferous forest trees* and *007_ Broadleaved forest trees* and entering a relative percentage value for area coverage of both components referring to the total area extend of a single geometric land cover unit.

How to describe something like Sparsely Vegetated Areas?

Answer: Use for example *002_Consolidated bare surface* in combination with another vegetation PLCC like *009_Herbaceous plants* and *010_Lichens and mosses* by also giving area percentage values.

Portrayal for code list:

This Color Map contains a legend color for each Pure Land Cover Component. When using several components of this code list to describe landscape, it is up to the data provider or end data user how to combine the colors. It can be done either in striped or dotted patterns, or only give color to the geometric unit according to the dominant component. In the last case, the information about a secondary or tertiary component can then still be entered in the model numerically with area percentages.

The R/G/B-Color Values of the PLCCs have been selected following good distinction on screen. When printed out on hard copy the colors may appear differently and not show good contrast anymore to each other. To avoid such effects it could be helpful to transform the R/G/B color values into CYMK.

Coherence between PLCC – CLC – LMCL

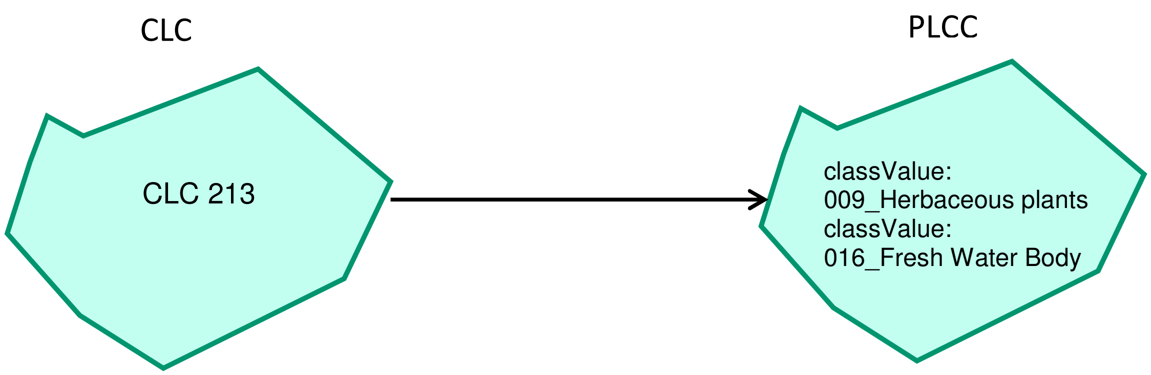
Expressing the two CORINE Land Cover (CLC) classes CLC 213 and CLC 243 translated into Land Cover Meta Language (LCML) in section F.3 into Pure Land Cover Component (PLCC) terms starts from their LCML translation.

CLC 213: Rice fields

The LCML translation of CLC 213 indicated that "Graminae", "Water Body" and "Periodic variation" are its basic elements, each with specific properties. "Cultivated and managed vegetation" and "Floristic aspect" are the optional characteristics of the "Graminae", while "Water Salinity" is a characteristic of "Water Body".

The first basic element is “graminae” which is translated 100% to the PLCC “009_Hherbaceous plants”. Other basic elements, properties and optional characteristics are ignored. The second element water body goes easily to PLCC “016_Fresh water body”.

Basic Element	PLC component
graminae	009_Herbaceous plants
water body	016_Fresh water body
periodic variation	n/a

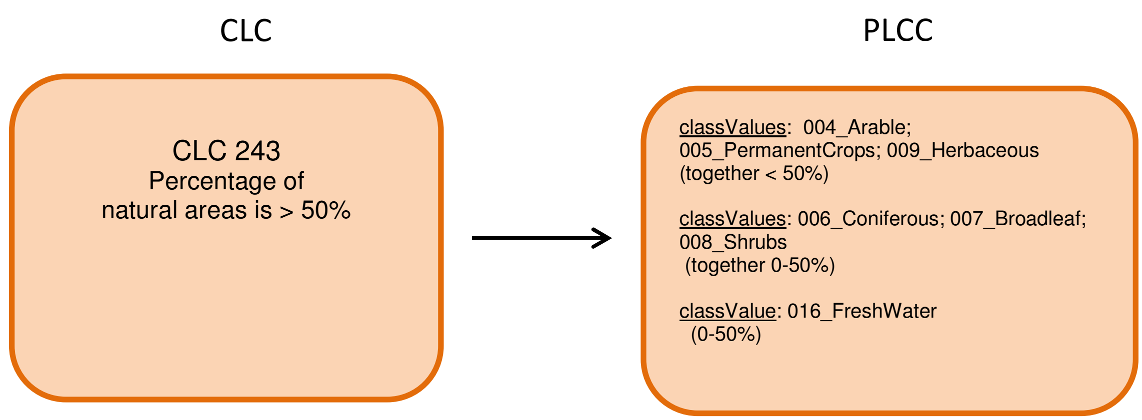


CLC class 243: Land principally occupied by agriculture, with significant areas of natural vegetation

The semantic analysis identified three horizontal patterns for CLC243; “herbaceous or woody crops”, “natural vegetation” and “water bodies”, considered as integral elements of a single “functional unit”. Such complex semantic interpretation requires clear and fixed ranges of proportional percentages between the three patterns. These patterns in turn represent one or more PLC components.

Horizontal Pattern	Contributing PLC components
herbaceous or woody crops	004_Arable_land, 009_Herbaceous_plants, 005_Permanent_woody_and_shrub_crops
natural vegetation	008_Shrubs, 006_ Coniferous_trees, 007_Broadleaved_trees
water bodies	016_Freshwater_bodies

The national CLC-based nomenclatures (using the 4th and 5th level of CLC hierarchy) and knowledge of the local conditions offer more information than is implied in the generic level 3 CLC class 243 illustrated above. This could obviously result in another set of classValues and better-known percentages. Absence of such information inevitably makes the determination of percentage for each horizontal pattern and its respective components arbitrary. However, the application of at least a minimum percentage of allowed proportion for each of the horizontal pattern is conditioned by the LCML translation.



Annex I
(informative)
Frequently Asked

1. Can CORINE Land Cover (CLC) be represented using the INSPIRE land cover data model?
Answer: Yes, CLC is covered by the core model

2. Why is CORINE Land Cover accepted in INSPIRE although it is violating the INSPIRE principle of clear separation between land cover and land use
Answer: CLC is a de facto harmonized pan-European land cover mapping initiative used by the environmental sector. Decisions about keeping, changing, modifying or abandoning CLC are up to the European Environmental Agency (EEA). The role of INSPIRE is to provide a tool for sharing existing data. Although CLC exceeds the boundary of the land cover theme, it is an existing data set that needs to be accommodated for by INSPIRE and this is done through the land cover data specification.

3. Why is not the CORINE Land Cover nomenclature made mandatory by the INSPIRE land cover data specification?
Answer: INSPIRE applies to all spatial land cover data in Europe. CORINE Land Cover is scale dependent and it is not a pure and exhaustive land cover nomenclature. CLC responds to specific needs linked to European land monitoring requirements in the 1980's and does not cover all current needs. Other nomenclatures exist for other uses.

5. Why does not INSPIRE enforce a single, standardized nomenclature for Europe?
Answer: Land cover mapping initiatives are developed for a multitude of different objectives. This is reflected in the broad range of nomenclatures and classification systems available. No single nomenclature or classification system covers all the different requirements by different users. The purpose of INSPIRE is to provide an infrastructure for exchanging data. INSPIRE must therefore provide a framework for exchanging land cover data using all different nomenclatures and provide mechanisms to compare and understand them. LCML is an example of such a mechanism (see Annex F).

6. Can the data specification be applied to the bottom of the ocean?
Answer: Yes. A mapping initiative implementing a mapping system and a nomenclature describing the bottom of oceans as a kind of land cover can use the INSPIRE Land Cover data specification.

7. Does the land cover data specification support 3D land cover units?
Answer: No.

8. Why are points allowed in the land cover data specification?
Answer: One way to observe land cover is to conduct a sampling survey where the observation units are conveniently represented as points. This allows the survey to use a detailed nomenclature, not attainable in a broad wall-to-wall survey. An example is the LUCAS survey carried out by Eurostat.

Annex J
(normative)
Encoding rules for TIFF and JPEG 2000 file formats

Introduction

This annex specifies how to use the TIFF or JPEG 2000 file formats for encoding the range set of grid coverages. Because pixel payload is not sufficient to construct a readable standalone image, additional descriptive information has to be packaged together in the same file, even if it is already provided somewhere else in GML. For this purpose, this part establishes schema conversion rules for all the coverage components of INSPIRE Application Schemas that have a corresponding element in the output TIFF or JPEG 2000 data structures. These conversion rules play an essential role in maintaining consistency between the different representations (i.e. GML, TIFF or JPEG 2000) of the same coverage information.

On the other hand, TIFF specifications and JPEG 2000 Standard offer many options and let some variables open for encoding image data. If this flexibility allows covering most applications, it leads, in turn, to a situation where disparate implementation platforms exist while being potentially incompatible. As a result, interoperability is often unlikely. In order to fill in this gap and to enable a controlled exchange of data across Europe, this annex draws up an implementation profile of TIFF and JPEG 2000 to constraint their usage within the scope of INSPIRE. It amounts to impose external format-dependent restrictions to the applicable values of the properties described in the INSPIRE application schemasThis annex specifies how to use the TIFF or JPEG 2000 file formats for encoding the range set of grid coverages. Because pixel payload is not sufficient to construct a readable standalone image, additional descriptive information has to be packaged together in the same file, even if it is already provided somewhere else in GML. For this purpose, this part establishes schema conversion rules for all the coverage components of INSPIRE Application Schemas that have a corresponding element in the output TIFF or JPEG 2000 data structures. These conversion rules play an essential role in maintaining consistency between the different representations (i.e. GML, TIFF or JPEG 2000) of the same coverage information.

On the other hand, TIFF specifications and JPEG 2000 Standard offer many options and let some variables open for encoding image data. If this flexibility allows covering most applications, it leads, in turn, to a situation where disparate implementation platforms exist while being potentially incompatible. As a result, interoperability is often unlikely. In order to fill in this gap and to enable a controlled exchange of data across Europe, this annex draws up an implementation profile of TIFF and JPEG 2000 to constraint their usage within the scope of INSPIRE. It amounts to impose external format-dependent restrictions to the applicable values of the properties described in the INSPIRE application schemas.

TIFF format

Format overview

The Tagged Image File Format (TIFF) is a binary file format for storing and interchanging raster images. Originally developed by the company Aldus (Adobe Systems), it is in the public domain since 1992, the year of the latest release of the specifications (revision 6.0 [TIFF]). TIFF has become a popular “de facto standard” for high color-depth digital images. It is widely used in image handling applications, covering various themes such as Land Cover.

TIFF specifications are divided into two parts. Part 1: Baseline TIFF defines all the features that every reader must support, while Part 2: TIFF Extensions provides additional format structures designed for specialized applications, that are not necessarily taken into account by all TIFF readers (e.g. JPEG or LZW compression, tiling, CMYK images).

As highlighted in the format name, the TIFF data structure is based on the definition of tags for describing the characteristics of images. To be more precise, a TIFF file contains an image file header pointing to one or several image file directory (IFD). The image file header fixes the technical properties of the file, such as the byte order (e.g. little-endian or big-endian) or the offset of the first byte. An image file directory holds the complete description of an image by means of fields or entries. Each IFD entry consists of a tag identifying the field, the field type (e.g. byte, ASCII, short int), the number of values and the values themselves or an offset to the values. The location of the actual image data within the file is given by the combination of information elements expressed in some fields.

INSPIRE TIFF profile for grid coverage data

This section lists the requirements and the constraints to be applied to the TIFF format when encoding INSPIRE Land Cover data sets in this format. It should be read in conjunction with the table in section 0 which provides more detailed information. Some of the rules presented here are directly inspired by the GeoTIFF Profile for Georeferenced Imagery [DGIWG-108] edited by DGIWG for the military community.

General rules

TG Requirement 1	Encoding of INSPIRE Land Cover data sets by using TIFF format shall conform to Baseline TIFF extended to LZW Compression.

NOTE Baseline TIFF is described in the part 1 of the TIFF specification 6.0 [TIFF], while the TIFF extension on LZW Compression is addressed in part 2.

TIFF files must be identified as such by network services by using a predefined Internet media type or MIME type.

TG Requirement 2	A file claiming to encode coverage elements in TIFF shall receive the *image/tiff* MIME type registered in RFC 3302.

NOTE The absence of the optional application parameter here does not necessarily imply that the encoded TIFF image is Baseline TIFF.

Data structure

Even though TIFF specifications allow describing multiple related images in a single file by using more than one Image File Directory (IFD), Baseline TIFF readers are not required to decode any IFD beyond the first one. In order to ensure alignment with Baseline TIFF, all indispensable information has to be included in the first IFD.

TG Requirement 3	The range set of the grid coverage shall be carried by only one Image File Directory (IFD), which is the first one in the TIFF file.

NOTE As a consequence, the different bands of a same image can not be split in separate IFDs.

TG Requirement 4	A TIFF file shall not contain more than two image file directories (IFD), the second being used to support a transparency mask.

The use of a second IFD is admitted for encoding an optional transparency mask, which is common for geographic raster data. This kind of ancillary information is useful at least for portrayal considerations. A transparency mask is a bi-level image matching pixel by pixel the image depicted in the first IFD. The pixel value 1 in the transparency mask means that the corresponding pixel in the image itself is significant. Conversely, the value 0 means that the corresponding pixel in the image holds a no data value (e.g. unknown, withheld). Typically, it must be made transparent when displaying the image.
The image file directory assigned to a transparency mask must receive the following TIFF tag values:

- BitsPerSample = 1

- Colormap: not used
- ImageDescription = ‘transparency mask’
- ImageLength = ImageLength of the first IFD
- ImageWidth = ImageWidth of the first IFD
- NewSubFileType: all bits equal 0, except bit 2 = 1
- PhotometricInterpretation = 4
- SamplesPerPixel = 1

Grid coordinate system

Baseline TIFF supports only one type of orientation for grid coverages, that is, one type of grid coordinate system.

TG Requirement 5	The origin of the grid coordinate system shall be the upper left corner of the grid coverage. The axis ‘row’ and ‘column’ shall be oriented downward and to the right.

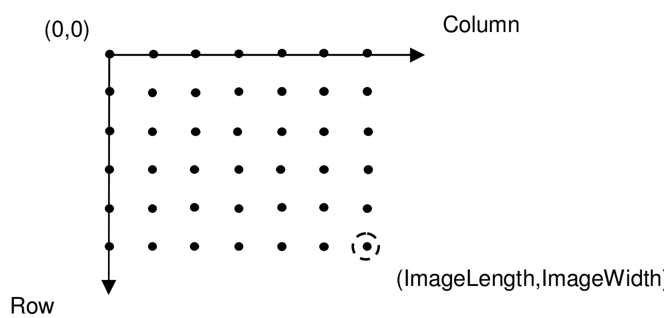


Figure 6: referenced grid as defined by Baseline TIFF

Range values

The Baseline TIFF specifications cover four image types: bi-level, greyscale, palette-colour and full-colour images. Multi-band images are allowed but not fully addressed: baseline TIFF readers are intended to skip over the extra components gracefully, using the values of the SamplesPerPixel and BitsPerSample fields.

Recommendation 1	The image data of a TIFF file should contain either 1 (bi-level, greyscale and palette-colour) or 3 bands (RGB).

NOTE Encoding multispectral images in TIFF is running the risk of losing a part of the coverage range set, since many software applications are not able to support more than three colours.

Open issue 1: The lack of a part of the coverage range set is a well-identified problem for the orthoimage delivery in the frame of the Control with Remote Sensing (CwRS) program of the MARS Unit of JRC. When data is delivered in TIFF, we occasionally receive only 3 out of the initial 4 four channels of the VHR satellite data (usually the colour infrared is the missing one). The lack of this information might be crucial for certain applications. In that respect we might think (in case of availability of multispectral data) to encourage the data producers to provide more than one RGB files, holding different band combinations – natural; colour infrared; false colour composite. It is a common practice in the frame of the CwRS, although it required additional efforts. Same delivery approach can be valid for JPEG2000 as well.

For better performances, it is preferable to encode the range values as arrays of type SHORT, BYTE or LONG, depending on the type of data.

TG Requirement 6	For imagery, the range values shall be expressed as unsigned integers coded on 8 or 16 bits, except for bi-level images which are 1-bit data. For other gridded data (e.g. elevation data, measured data), they shall be stored as 8 or 16-bits integers, signed or unsigned, or as 32-bits floating points.

NOTE If the original data do not satisfy this requirement, they will be converted in a representation using the next higher power of 2.

TG Requirement 7	In the case of multi-band images, the number of bits per component shall be the same for all the bands.

TG Requirement 8	In the case of multi-band images, the planar configuration shall be *Chunky* format, i.e. the bands are interleaved.

NOTE The range values of a same grid point in its different bands are stored contiguously. For instance, RGB data is stored as RGBRGBRGBRGB...

Compression

Data compression can be used within this profile to reduce the size of a file, provided that it does not distort the original range values after an encoding-decoding cycle. This condition allows, for example, ensuring the preservation of nil values.

TG Requirement 9	The range value data shall be either uncompressed or lossless compressed with packbit or LZW compression schemes.

NOTE As a TIFF extension, LZW compression is not supported by Baseline TIFF. However, it is included in this profile since its use is widespread, essentially for both its simplicity and its efficiency.

Internal tiling

The TIFF extension defined in section 15 of the specifications focuses on the way of laying out the image content into roughly square tiles. This method, as an alternative to the standard repartition of the range within separate strips, increases the access to data. However, it may cause some interoperability problems in the same time. It is therefore better not to use it and to restrict oneself to Baseline TIFF.

Open issue 2: Is there a strong requirement to allow internal tiling for TIFF files?

Mapping between TIFF and GML data structures

The following table indicates how to fill the content of TIFF tags for grid coverages in the context of INSPIRE. On the other hand, it gives the rules to be applied for ensuring the consistency of TIFF files with the Land Cover GML Application(s) Schema(s). It does not address the encoding of the possible transparency mask.

The columns *Tag name*, *Code*, *Type*, *Card.* and *Description* remind respectively the name, the code, the type, the maximum number of occurrences and the description of each Baseline TIFF tag within the meaning of the TIFF specification. The column *Obligation* informs if the tag is considered to be mandatory (M), conditional (C), optional (O) or inadequate (I). The column *Restricted values* specifies the values allowed for the tag in the context of INSPIRE. The column *Mapping to GML elements* establishes a correspondence between the tag values and the corresponding GML elements of the coverage whose type is one of those specified in the Generic Conceptual Model (e.g. RectifiedGridCoverage). N/A means not applicable.

Table 30. Baseline TIFF implementation profile and Mapping between TIFF tags and the associated object elements from the Land Cover GML Application Schema

Tag name	Code	Type	Card.	Description	Obligation	Restricted values	Mapping to GML elements (including restrictions)
Artist	315	ASCII	1	Person who created the image	O	-	N/A
BitsPerSample	258	Short	SamplesPerPixel	Number of bits per component	M	1 for bi-level images For imagery, constrained to **8** or **16** bits-per-pixel-per-band (e.g. 8 8 8 or 16 16 16 for RGB images). For other gridded data, 8, 16 and 32 bits-per-pixel-per-band	For each band *i*, rangeType.field[*i*].constraint.interval = "0 *2^BitsPerSample[i]-1*"
CellLength	265	Short	1	The length of the dithering or halftoning matrix used to create a dithered or halftoned bilevel file.	I	This field should be never used	N/A
CellWidth	264	Short	1	The width of the dithering or halftoning matrix used to create a dithered or halftoned bilevel file.	I	This field should be never used	N/A
ColorMap	320	Short	3*(2**BitsPerSample)	A colour map for palette colour images	C	Only for palette color images	N/A
Compression	259	Short	1	Compression scheme used on the image data	M	**1** for uncompressed data **5** for LZW compression **32773** for PackBits compression of greyscale and palette-colour data	N/A
Copyright	33432	ASCII	1..*	Copyright notice	O	-	N/A
DateTime	306	ASCII	20	Date and time of image creation	O	-	N/A NOTE the field DateTime should not be confused with the properties *phenomenonTime* and *beginLifespanVersion* that report other types of temporal information.
ExtraSample	338	Short	1..*	Description of extra components	C	Only when extra samples are present	N/A
FillOrder	266	Short	1	The logical order of bits within a byte.	O	**1** (default)	N/A

Tag name	Code	Type	Card.	Description	Obligation	Restricted values	Mapping to GML elements (including restrictions)
FreeByteCounts	289	Long	1	For each string of contiguous unused bytes in a TIFF file, the number of bytes in the string.	I	This field should be never used	N/A
FreeOffsets	288	Long	1	For each string of contiguous unused bytes in a TIFF file, the byte offset of the string.	I	This field should be never used	N/A
GrayResponseCurve	291	Short	2**BitsPerSample	For greyscale data, the optical density of each possible pixel value.	I	This field should be never used	N/A
GrayResponseUnit	290	Short	1	The precision of the information contained in the GrayResponseCurve	I	This field should be never used	N/A
HostComputer	316	ASCII	1..*	The computer and/or operating system in use at the time of image creation.	O	**-**	N/A
ImageDescription	270	ASCII	1..*	Description of the image subject.	O	-	N/A
ImageLength	257	Short or Long	1	The number of rows in the image.	M	-	domainSet.extent.high.coordValues[*0*]-domainSet.extent.low.coordValues[*0*]=ImageLength
ImageWidth	256	Short or Long	1	The number of columns in the image, i.e. the number of pixels per row.	M	-	domainSet.extent.high.coordValues[*1*]-domainSet.extent.low.coordValues[*1*]=ImageWidth
Make	271	ASCII	1	The scanner manufacturer.	O	-	N/A
MaxSampleValue	281	Short	SamplesPerPixel	The maximum component value used.	O	This field should be used only for statistical purposes	N/A
MinSampleValue	280	Short	SamplesPerPixel	The minimum component value used.	O	This field should be used only for statistical purposes	N/A
Model	272	ASCII	1	The scanner model name or number.	O	-	N/A
NewSubfileType	254	Long	1	A general indication of the kind of data contained in this subfile.	O	**0**	N/A
Orientation	274	Short	1	The orientation of the image with respect to the rows and columns.	M	**1** (default)	domainSet.extent.low.coordValues="0 0"

Tag name	Code	Type	Card.	Description	Obligation	Restricted values	Mapping to GML elements (including restrictions)
PhotometricInterpretation	262	Short	1	Colour space of the image data.	M	**1** for bi-level and greyscale images (0 is black) **2** for RGB images **3** for palette-colour images	N/A
PlanarConfiguration	284	Short	1	How the components of each pixel are stored.	M	**1** which means, for RGB data, that the data is stored as RGBRGBRGB…	rangeSet.fileStructure="Record Interleaved"
ResolutionUnit	296	Short	1	Unit of measurement for XResolution and YResolution.	M	**2** which means dpi (dot per inch)	N/A
RowsPerStrip	278	Short or Long	1	Number of rows per strip.	C Not used if tiling	It is recommended to choose this value such that each strip is about 8K bytes.	N/A
SampleFormat	399	Short	SamplesPerPixel	This field specifies how to interpret each data sample in a pixel.	M	**1** for imagery (unsigned integer data) 1 or 3 for gridded data	For imagery, for each band *i*, rangeType.field[*i*].constraint.interval[0] = "0"
SamplesPerPixel	277	Short	1	Number of components per pixel.	M	**1** usually for bi-level, greyscale and palette-colour images **3** usually for RGB images	rangeType.field.size()=SamplesPerPixel
SmaxSampleValue	341	Field type that best matches the sample data	SamplesPerPixel	The maximum value for each sample. This tag is used in lieu of MaxSampleValue when the sample type is other than integer.	I	This field should be never used	N/A
SminSampleValue		Field type that best matches the sample data	SamplesPerPixel	The minimum value for each sample. This tag is used in lieu of MaxSampleValue when the sample type is other than integer.	I	This field should be never used	N/A
Software	305	ASCII	1..*	Name and version number of the software package(s) used to create the image.	O	-	N/A
StripByteCounts	279	Short or Long	StripPerImage	For each strip, number of bytes in the strip after compression.	C Not used if tiling	**-**	N/A
StripOffsets	273	Long	StripPerImage	For each strip, the byte offset of that strip	C Not used if tiling	-	N/A

Tag name	Code	Type	Card.	Description	Obligation	Restricted values	Mapping to GML elements (including restrictions)
Thresholding	263	Short	1	For black and white TIFF files that represent shades of gray, the technique used to convert gray to black and white pixels.	I	This field should be never used	N/A
TileWidth	322	Short or Long		The tile width in pixels. This is the number of columns in each tile.	C if tiling	-	N/A
TileLength	323	Short or Long		The tile length (height) in pixels. This is the number of rows in each tile.	C if tiling	-	N/A
TileOffsets	324	Long		For each tile, the byte offset of that tile, as compressed and stored on disk.	C if tiling	-	N/A
TileByteCount	325	Short or Long		For each tile, the number of (compressed) bytes in that tile.	C if tiling	-	N/A
Xresolution	282	Rational		The number of pixels per ResolutionUnit in the ImageWidth direction.	M	**254**	N/A
Yresolution	283	Rational		The number of pixels per ResolutionUnit in the ImageLength direction.	M	**254**	N/A

In addition, the description of the coverage grid function must reflect the baseline ordering used by TIFF format to store the range values within a file. The following mapping must be applied:

coverageFunction.gridFunction.sequenceRule.type = “linear” AND coverageFunction.gridFunction.sequenceRule.scanDirection = “+2 +1”

Theme-specific requirements and recommendations

No further requirements or recommendations are defined for this theme.

JPEG 2000 format

Format overview

JPEG 2000 is a wavelet compression for storing and interchanging raster. Other wavelet compressions exist like ECW or MrSid. JPEG 2000 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information* in collaboration with ITU-T. The identical text is published as ITU-T Rec. T.800. First version was was published in 2000. JPEG 2000 is known as a very efficient format to distribute and access large imagery data.

JPEG 2000 Standard is defined by ISO 15444 serie (from 15444-1 to 15444-12). The two parts dealing with 2D still imagery and then interesting for INSPIRE are:

- ISO 15444-1: Core Coding System, defining how coders and decoders shall behave and how shall be structured a JPEG 2000 codestream. This part also defines JP2 format, the simpler wrapper for JPEG 2000 encoded data.
- ISO 15444-2: Extensions, defining extensions for JPEG 2000 codestream (new makers) and JPX format. This part deals with extended capabilities which are not useful for INSPIRE themes. That's the reason why this part is not used within INSPIRE.

JPEG 2000 is complex

- The JPEG 2000 codestream, which directly contains compressed data. This stream contains markers and segment markers which allow decoding and accessing data.
- The format which is the wrapper of the JPEG 2000 codestream. It is possible to only distribute the codestream (extension file .j2c), but to have a more comprehensive file, it's recommended to wrap this stream inside a format, whose the most common is JP2, described by Annex I of ISO 15444-1 (extension file .jp2) which adds some boxes describing encoded data.

The figure below shows the JP2 file structure :

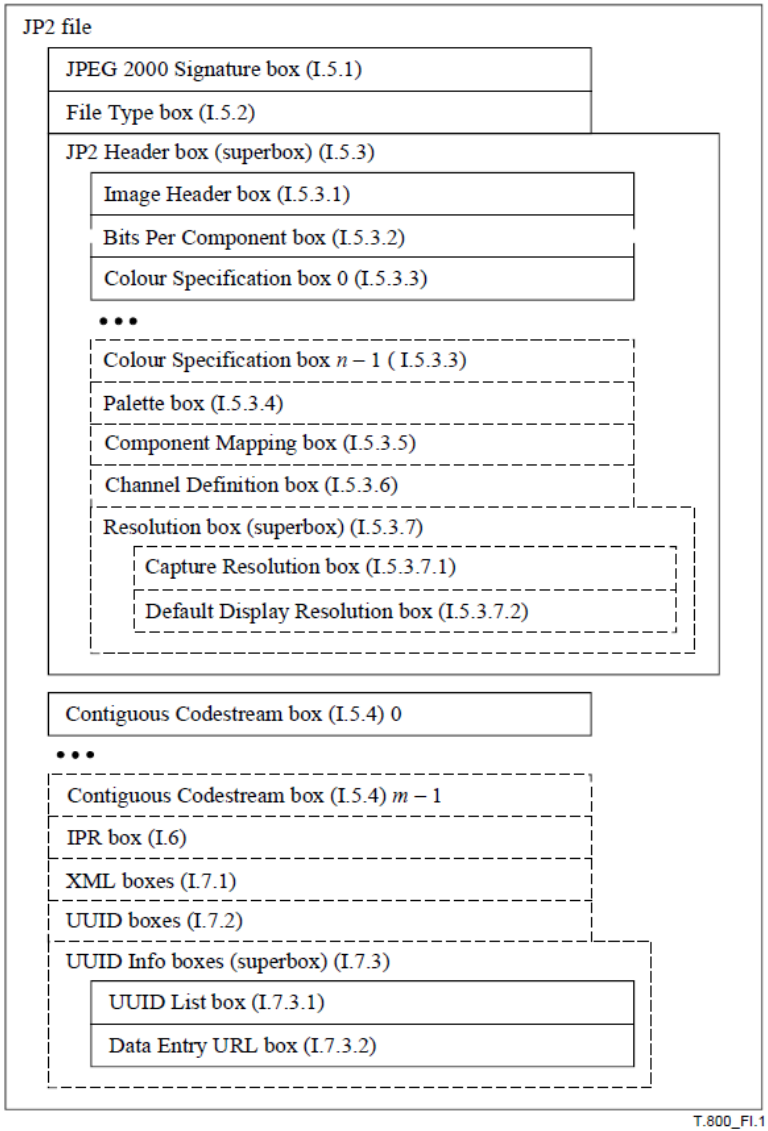


Figure 44 : JP2 file structure

JPEG 2000 profile for INSPIRE Land Cover data

This section lists the requirements and the constraints to be applied to JPEG 2000 when encoding INSPIRE Land Cover data sets in this format. It should be read in conjunction with the table in section 0 which provides more detailed information.

- **General rules**

TG Requirement 10	Encoding of INSPIRE Land Cover data sets by using JPEG 2000 shall conform to profile 1 of ISO 15444-1, extended by the use of boxes "association" and "label" (defined by JPX format in ISO 15444-2) necessary for GMLJP2 encoding (see GMLJP2 standard for more details).

JPEG 2000 files must be identified as such by network services by using a predefined Internet media type or MIME type

TG Requirement 11	A file claiming to encode coverage elements in JPEG 2000 shall receive the *image/jp2* MIME type registered in RFC 3745.

NOTE GMLJP2 uses extended boxes from JPX format, so it would suggest a *image/jpx* MIME type but GMLJP2 standard ask for image/jp2 as well because "association" and "label" boxes are just a very small part of JPX. Claiming conformance to JP2 allows GMLJP2 data to be supported by more visualisation softwares (some tools could stop reading JPEG 2000 files when seeing image/jpx MIME type). In the case of a software only compliant with ISO 15444-1 (reading strict JP2 files), the image and the GML (in the XML box) will be read and just the association between the two will be not interpreted.

So in both cases, pure JPEG2000 or GML embedded in JPEG2000 (GMLJP2), *image/jp2* MIME type shall be used.

Data structure

Even though JPEG 2000 Standard (and more precisely JP2 format) allows describing multiple codestreams in a single file by using more than one jp2c, only one is required to encode range sets of gridded coverages.

TG Requirement 12	The range set of the grid coverage shall be carried by only one jp2c box (one codestream per JPEG 200 file).

NOTE As a consequence, the different components of a same image can not be split in separate codestreams.

Grid coordinate system

JPEG 2000 Standard defines the origin of the grid coordinate system as being the upper left corner of the grid coverage. The axis 'X' and 'Y are oriented to the right and downward.

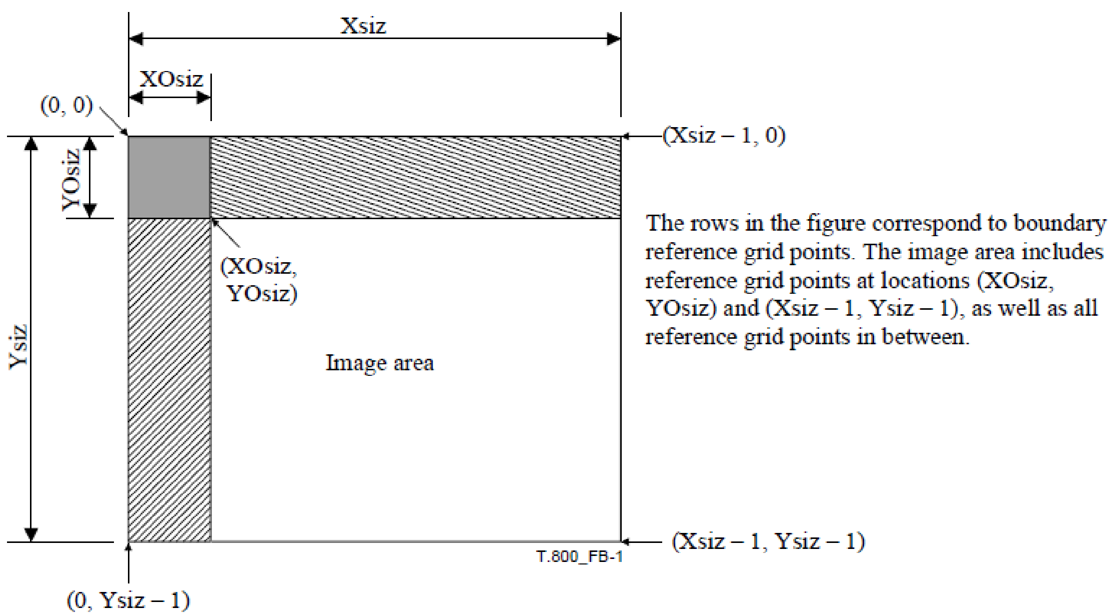


Figure 45 : referenced grid as defined by JPEG 2000

Source : ISO 15444-1

Range values

ISO 15444-1 allows a lot of image types with multiband composition. Within the scope of INSPIRE, following types are addressed: bilevel, grayscale, palette-color and full-color images (known as RGB images).

Recommendation 2 The image data of a JPEG 2000 file should contain either 1 (bilevel, greyscale and palette-colour) or 3 bands (RGB). An additional band for opacity could also be used.

The use of palette-colour in JPEG 2000 is restricted to a mapping to one component to RGB data.

Open issue 3: Is the use of palette-colour needed. It allows to compress RGB data on 1 band and provide a mapping table.

These components are described trough markers in the JPEG 2000 codestream (see Table 32) and boxes in the JP2 format (see Table 33).

TG Requirement 13 For imagery, the range values shall be expressed as unsigned integers coded on 8 or 16 bits, except for bi-level images which are 1-bit data. For other gridded data (e.g. elevation data, measured data), they shall be stored as 8 or 16-bits integers, signed or unsigned, or as 32-bits.

JPEG2000 (ISO 15444-1) does not allow to encode data as floats (only integers), but in general you can choose a Unit of measure for which your results are integers. For elevation, use centimeters (cm) instead of meters (m).

NOTE If the original data do not satisfy this requirement, they will be converted in a representation using the next higher power of 2.

TG Requirement 14 In the case of multi-band images, the number of bits per component shall be the same for all the bands.

Opacity channel

JP2 format allows to describe opacity channel (trough Channel Definition Box or CDEF) and then to display multiple files without overlapping issues. Opacity channel can be defined for RGB or greyscales images. The following table give example of an RGB image with alpha channel. CDF defines the 3 RGB components and then the alpha channel which applies to the all 3 RGB ones.

Table 31. Definition of opacity channel with CDEF box

<table>
<tr><td colspan="5">CDEF box</td></tr>
<tr><td>Hexadecimal</td><td colspan="3">Numeric conversion (2 bits)</td><td>interprétation</td></tr>
<tr><td>00 04</td><td colspan="3">4</td><td>4 components</td></tr>
<tr><td>00 00 00 00 00 01 |
00 01 00 00 00 02
00 02 | 00 00 00 03
00 03 00 01 | 00 00</td><td>0
1
2
3</td><td>0
0
0
1</td><td>1
2
3
0</td><td>component 0 is red R
component 1 is green G
component 2 is blue B
component 3 is an opacity channel related to all other component</td></tr>
<tr><td></td><td>Component number</td><td>signification</td><td>Color of the component</td><td></td></tr>
</table>

NOTE In this case, bit depth shall be defined for each of the four components trough the bpcc boxes in the JP2 Header Box.

JPEG 2000 allows defining opacity channel on more than 1 bit to have a scale of transparency. In pur case, we're just interested in full transparency and full opacity. So, within the scope of INSPIRE, it is recommended to code it on only 1 bit (0= transparent , 1=opaque)

Recommendation 3 In the case of an opacity channel, the bit depth should be 1-bit.

Compression

JPEG 2000 codestream allows both lossless and lossy compression. Lossless compression is important for some themes because you can't allow any loss of information. For example in Land Cover, or Land Use, you encode a code which represents a class. For other themes, a lossy compression without visual effect can be also interesting. JPEG 2000 lossy compression is very powerful with which you can compress imagery data by 1:10 or more without visual effect.

Open issue 4: Some TWGs may use lossy compression. If there is a need for requirement/recommendation, you can add it in the theme specific section

Internal tiling

JPEG 2000 allows internal tiling within the codestream. Profile 1 of ISO 15444-1 already requires no tiling (i.e. the image = 1 tile) or tiles size bigger than1024x1024 pixels. There is no further requirement.

Open issue 5: That means there is either 1 tile or more tiles bigger that 1024x1024? Maybe a clarification is needed.

Resolutions

JPEG 2000 codestream encode the full resolution image but has mechanisms to directly access (without any computation) particular sub-resolutions. So the JPEG 2000 file contains a pyramid of resolution. The number of decomposition Nd defines the smallest image you can access whose size is reduced by 2^{Nd} (compared to the full image). Profile 1 of ISO 15444-1 requires that the number of decomposition shall be such as: Height/2^{Nd} ≤128 pixels and Width/2^{Nd} ≤128 pixels (height and width of the full resolution image).

For example for a 2048x1024 image, the number of decomposition is 4, and the smallest thumbnail image is 128x64 pixels.

There is no further requirement.

Region of interest

When encoding in a lossy mode, JPEG 2000 allows to encode some image regions with better quality and then deteriorate the quality of other areas. This capability shall not be used.

TG Requirement 15 JPEG 2000 codestream shall not encode Region Of Interest (RGN marker segment).

Other options

JPEG 2000 allows more options:
- Quality layers, i.e. the capability to different levels of compressions within the same JPEG 200 file.
- Presence of markers, some allowing fast data access (TLM, PLT), other allowing error resiliency, ..
- Presence of precints and their size.
- Encoding order ; the codestream can be arranged in different ways depending the order you want the data to be decompressed.
- Color transformation, from RGB to three other decorrelated components (ICT or RCT transformations).

These choices depend on data size, data access (trough network services, via FTP, via USB stick, …) and then can't be made here.

Mapping between JPEG 2000 and GML data structures

The following table indicates how to fill the content of TIFF tags for grid coverages in the context of INSPIRE. On the other hand, it gives the rules to be applied for ensuring the consistency of JPEG 2000 files with the Land Cover GML Application(s) Schema(s). It does not address the encoding of the possible transparency mask.

As described by the Format overview section, the JP2 format contains the JPEG 2000 codestream. Both have elements that need to be consistent with GML. The first table deals with mappings between the JPEG 2000 codestream and GML, whereas the second table deals with mappings between JP2 boxes and GML elements.

The columns *marker/box*, *description*, *Type*, *Card.* and Conditions/*Values* remind respectively the marker code/box name, its description, its obligation/maximum number of occurrences allowed by JPEG 2000 standard (ISO 15444-1). The column *values* specifies the values allowed for the marker in the context of INSPIRE. The column *Mapping to GML elements* establishes a correspondence between these makers values and the corresponding GML elements of the coverage whose type is one of those specified in the Generic Conceptual Model (e.g. RectifiedGridCoverage). N/A means not applicable.

Table 32. mapping between markers in JPEG 2000 codestream and GML elements

Marker	Description	Card.	Values	Mapping GML
SIZ	Marker code (Image and tile size)	1		N/A
Lsiz	Length of marker segment	1		N/A
Rsiz	Denotes capabilities that a decoder needs to properly decode the codestream	1	0000 0000 0000 0010 Codestream restricted as described for Profile 0 from Table A.45	N/A
Xsiz	Width of the reference grid	1		= domainSet.RectifiedGrid.limits.GridEnvelope.high[0]
Ysiz	Height of the reference grid	1		= domainSet.RectifiedGrid.limits.GridEnvelope.high[1]
X0siz	Horizontal offset from the origin of the reference grid to the left side of the image area	1		= domainSet.RectifiedGrid.limits.GridEnvelope.low[0]
YOsiz	Vertical offset from the origin of the reference grid to the top side of the image area.	1		= domainSet.RectifiedGrid.limits.GridEnvelope.low[1]
XTsiz	Width of one reference tile with respect to the reference grid.	1		N/A
YTsiz	Height of one reference tile with respect to the reference grid	1		N/A
XTOsiz	Horizontal offset from the origin of the reference grid to the left side of the first tile	1		N/A
YTOsiz	Vertical offset from the origin of the reference grid to the top side of the first tile	1		N/A
Csiz	Number of components in the image	1	1 for greyscale imagery 3 for RGB data …	rangeType.field.size()
$Ssiz^i$	Precision (depth) in bits and sign of the ith component samples	1/component	x000 0000 to x010 101 Component sample bit depth = value + 1. x=0 (unsigned values) x=1 (signed values)	For each band i, rangeType.field[i].constraint.interval = “0 2^[$Ssiz^i$+1]-1”
$XRsiz^i$	Horizontal separation of a sample of ith component with respect to the reference grid	1/component	In most case $XRsiz^i$ =1 for all components	N/A
$YRsiz^i$	Vertical separation of a sample of ith component with respect to the reference grid	1/component	In most case $YRsiz^i$ =1 for all components	N/A

For each component i of the image, its size is defined by :

$Width^{i} = (Xsiz - X0siz)/XRsiz^{i}$
$Height^{i} = (Ysiz - Y0siz)/YRsiz^{i}$

Table 33. Mapping between boxes in JP2 format and GML elements

Box name	Type	Description	Card.	Conditions/Values	Mapping GML
JPEG 2000 Signature box	'jP\040\040'	The combination of the particular type and contents for this box enable an application to detect a common set of file transmission errors.	1	'<CR><LF><0x87><LF>'	N/A
File Type box	'ftyp'		1		N/A
BR		Brand. This field specifies the Recommendation | International Standard which completely defines this file.		'jp2\040' meaning is 15444-1, Annex I	N/A
MinV		Minor version. This parameter defines the minor version number of this JP2 specification for which the file complies.	1		N/A
CL		Compatibility list. This field specifies a code representing this Recommendation | International Standard, another standard, or a profile of another standard, to which the file conforms.	1..*	At least 'jp2\040'	N/A
JP2 Header box	'jp2h'		1		N/A
ihdr	'ihdr'	Image Header box	1		N/A
HEIGHT		Image area height	1	Ysiz – Y0siz	domainSet.RectifiedGrid.limits.GridEnvelope.high[1]-domainSet.RectifiedGrid.limits.GridEnvelope.low[1]
WIDTH		Image area width	1	Xsiz – X0siz	domainSet.RectifiedGrid.limits.GridEnvelope.high[0]-domainSet.RectifiedGrid.limits.GridEnvelope.low[0]
NC		Number of components	1	= Csiz	= rangeType.field.size() if no use of palette-colour data. If use of a colour palette NC=1, rangeType.field.size()=3.

		BPC		Bits per component	1	If the bit depth of all components in the codesteam is the same (sign an precision) = Ssiz^i	For each band i, rangeType.field[i].constraint.interval = “0 2^[Ssiz^i+1]-1” if no use of palette-colour data. If use of a palette colour, there is no relation.
		C		Compression type		7 (Other values are reserved for ISO use)	N/A
		UnkC		Colourspace Unknown.	1	0 (colourspace of the image is known and correctly specified in the Colourspace Specification boxes within the file) 1 (if the colourspace of the image is not known)	N/A
		IPR		Intellectual Property	1		N/A
	bpc^i		'bpcc'	Bits per component	Optional Required if component have different bit depth	x000 0000 to x010 101 Component sample bit depth = value + 1. x=0 (unsigned values) x=1 (signed values)	For each band i, rangeType.field[i].constraint.interval = “0 2^[Ssiz^i+1]-1”
	colr^i		'colr'	Each Colour Specification box defines one method by which an application can interpret the colourspace of the decompressed image data	1		N/A
		METH		Specification method	1	1 (Enumerated Colourspace) 2 (Restricted ICC profile) other values (Reserved for other ISO use)	N/A
		PREC		Precedence	1	0 (field reserved for ISO use)	
		APPROX		Colourspace approximation.	1	0	N/A
		EnumCS		Enumerated colourspace	1	16 (sRGB as defined by IEC 61966-2-1) 17 (greyscale) 18 (sYCC as defined by IEC 61966-2-1 Amd. 1) other values (Reserved for other ISO uses)	N/A

	pclr		'pclr'	Palette box. This box specifies a palette that can be used to create channels from components.	0..1		N/A
	cmap		'cmap'	Component Mapping box. The Component Mapping box defines how image channels are identified from the actual components decoded from the codestream.	0..1		N/A
	cdef		'cdef'	Channel Definition box	Optional		The description provided shall be consistent with the rangeType description
		N		Number of channel descriptions	1		= rangeType.field.size()
		Cni		Channel index	1/channel		N/A
		Typi		Channel type	1/channel	0 This channel is the colour image data for the associated colour. 1 (Opacity) 2 (Premultiplied opacity)	N/A
		Asoci		Channel association	1/channel	0 (This channel is associated as the image as a whole) 1 to (216– 2) This channel is associated with a particular colour as indicated by this value) 216– 1 This channel is not associated with any particular colour.	N/A
	res		'resd'		Optional		N/A
		resc		Capture Resolution box.	Optional		N/A
		resd		Default Display Resolution box.	Optional		N/A
Contiguou Codestream box			‘jp2c’	This box contains the codestream as defined by Annex A of ISO 15444-1.	1	Contains the encoded data in JPEG 2000.	
Intellectual property box			‘jp2i’	This box contains intellectual property information about the image.	Optional		N/A
XML Box			‘xml\040’	Box for XML formatted information to a JP2 file.	Optional		The place to provide GML within JPEG 2000 (see OGC standard for more details).

UUID box	'uuid'	Box for additional information to a file without risking conflict with other vendors	Optional		The place to provide GeoJP2 georeference. Shall be consistent with georeference given by : The origine of the grid : domainSet.RectifiedGrid.origin domainSet.RectifiedGrid.offsetVector
UUID info box	'uinf'	Box for providing access to additional information associated with a UUID.	Optional		N/A
UUID list box	'ulst'	This box specifies a list of UUIDs.	Optional		N/A
URL box	'url\040'	This box specifies a URL.	Optional		N/A

Theme-specific requirements and recommendations

No further requirements or recommendations are defined for this theme.

Example for Land Cover data :

For Land Cover data, as range values represent classification codes, compression shall be lossless.

JPEG 2000 codestream shall be compressed in a lossless mode.

For Land Cover, only a grid code is stored within the JPEG 2000 stream on one component.

JPEG 2000 codestream shall contain only one component with a bit depth of 8-bits and a greyscale colour space.