

Tvorba GML souboru odpovídajícího požadavkům INSPIRE pomocí programu HALE studio

Miroslav Fanta
CENIA, oddělení geoinformatiky
poslední aktualizace: 20.10.2020
miroslav.fanta@cenia.cz

Tento text předpokládá, že uživatel se již seznámil se základní způsoby práce s programem HALE studio, které jsou popsány v dokumentu INSPIRE_GML_HALEstudio1_CHMUmonitStanice_...

CORINE Land Cover

Data CORINE Land Cover se ve formátu GML dle požadavků INSPIRE skládají ze dvou částí (obě společně v jednom GML souboru)

- Land Cover Dataset
- Land Cover Units

Část Land Cover Dataset představuje „vnější obálku“, část Land Cover Units pak jsou vlastní polygony klasifikované do jednotlivých tříd pokryvu území dle klasifikace CORINE. Část Land Cover Dataset se nepodařilo vytvořit správně v programu HALE studio, ale bylo ji možno sestavit editací v programu Notepad++ na základě datového modelu a několika existujících vzorů. Část Land Cover Units lze vytvořit v programu HALE studio, následně ji je třeba v editoru Notepad++ pouze drobně upravit a spojit s částí Land Cover Dataset.

Příprava vstupních dat

Shapefile obsahující polygony pokryvu území klasifikované dle CORINE se načte do programu QGIS a v něm se provedou 2 následující kroky

- ověření, zda datová sada obsahuje 3. souřadnici. Pokud ano, je ji nutno odstranit a data exportovat do shapefile s pouze dvěma souřadnicemi (postup viz dokument „**Tvorba GML souboru odpovídajícího požadavkům INSPIRE pomocí sw HALE studio**“, kapitola „**Úprava vstupních dat do 2D, pokud obsahují souřadnici Z**“)
- export atributové tabulky do formátu csv: Klik pravým tlačítkem na vrstvu – Export – okno Uložit prvky jako... - otevře se okno „Uložit vektorovou vrstvu jako...“ – v rámečku Formát vybrat „Hodnoty oddělené čárkou [CSV]“ – doplnit název a místo uložení výstupního souboru – OK. Další zpracování a využití takto vytvořeného .csv souboru je popsáno níže v kapitole „**Vytvoření odkazů na jednotlivé Land Cover Units**“

Příprava části Land Cover Units

(vzorový soubor CORINE2_LCUonly1.hale)

Úvodní nastavení proměnných:

Variables

Specify project variables that can be used in certain transformation functions. Project variables may be overridden by system properties or environment variables (using a specific prefix).

ID_prefix	LandCoverUnit_	✖
ID_suffix	not_used_in_October2020	✖
ddCLCvocabularyRoot	http://dd.eionet.europa.eu/vocabulary/landcover/clc/	✖
eulID	http://inspire.ec.europa.eu/ids	✖
nilReasonValue	other:unpopulated	✖
projectNamespace	http://www.cenia.cz/CLC18_CZ/	✖
		+

Alignment:

CLC18_CZ_selection -> *Retype* -> **LandCoverUnit**

CODE_18 -> *Formatted string*: {{project:ddCLCvocabularyRoot}}{{CODE_18}} ->
-> **landCoverObservation.LandCoverObservation.class.href**

the_geom -> *Rename* -> **geometry**

ID -> *Formatted string*: {{project:ID_prefix}}{{ID}} -> **id** (= *gml_id*)

ID -> *Formatted string*: {{project:projectNamespace}}{{project: ID_prefix}}{{ID}} -> **identifier**
Assign: {{project:eulID}} -> **identifier.CodeSpace**

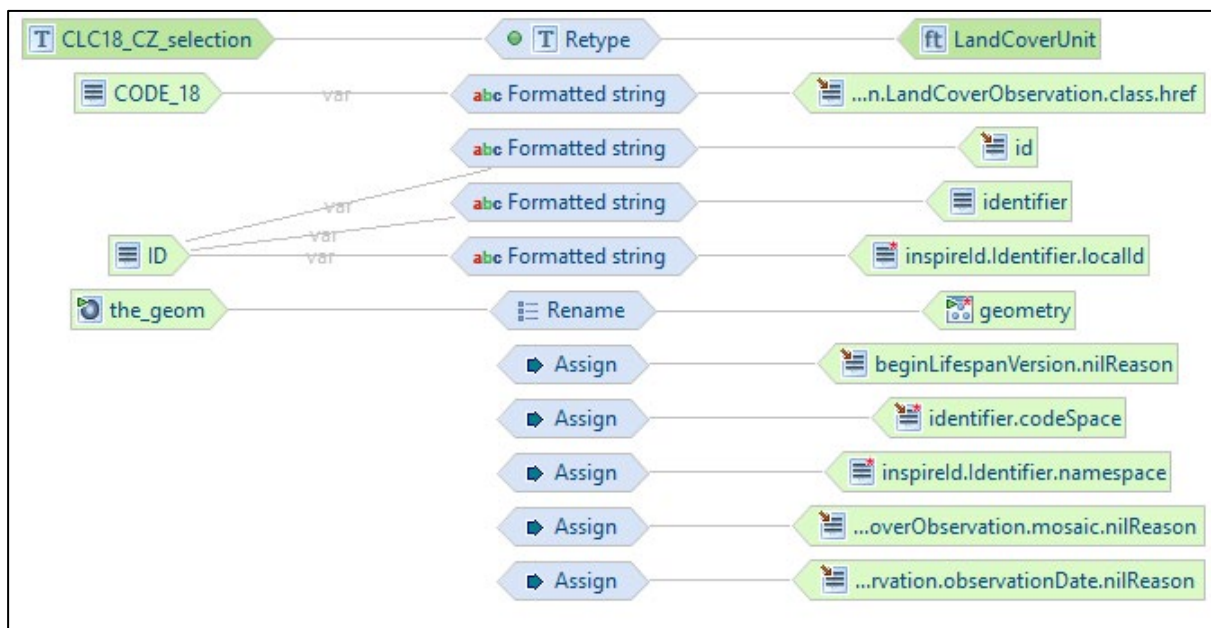
ID -> *Formatted string*: {{project:ID_prefix}}{{ID}} -> **inspireId.Identifier.localId**
Assign: {{project:projectNamespace}} -> **inspireId.Identifier.namespace**

Assign: {{project:nilReasonValue}} ->
-> **landCoverObservation.LandCoverObservation.mosaic.nilReason**

Assign: {{project:nilReasonValue}} ->
-> **landCoverObservation.LandCoverObservation.observationDate.nilReason**

Assign: {{project:nilReasonValue}} -> **beginLifespanVersion.nilReason**

Výsledný Alignment pro LandCoverUnit:



Tvorba výstupního gml souboru pro Land Cover Units:

output file name: např. clc18_LCUnits1.gml

namespace: http://www.cenia.cz/CLC18_CZ/ (nepoužívat proměnnou, URL nutno vypsát)

Při vlastní transformaci dat je u polygonových datových sad nutno nastavit **směr posloupnosti lomových bodů polygonů (winding order)**, neboť INSPIRE validátor vyžaduje posloupnost lomových bodů po směru hodinových ručiček (clockwise). Toto nastavení lze při exportu transformovaných dat provést na kartě **Coordinate reference system**:

Vytvoření odkazů na jednotlivé Land Cover Units

(vzorový soubor *CLC18_CZ_selection_2D.xlsx*)

Jako vstup se použije soubor typu .csv vytvořený jako export atributové tabulky shapefile se zdrojovými daty.

Otevřít excel – Data – Z textu/csv – vybrat příslušný csv soubor – ověřit zda je nastaven oddělovač čárka – Načíst – smazat všechny sloupce s výjimkou ID (identifikátor jednotlivých polygonů; hodnoty pravděpodobně CZ_nnn).

do sloupce před sloupcem ID vložit text:

<lcv:member xlink:href="http://www.cenia.cz/CLC18_CZ/LandCoverUnit_
(stejný pro všechny záznamy);

Uvedený řetězec souvisí s obsahem proměnných *projectNamespace* a *ID_prefix* použitých při tvorbě části **Land Cover Units** v programu HALE studio (viz výše). Sekvence http://www.cenia.cz/CLC18_CZ/ musí odpovídat obsahu proměnné *projectNamespace* a sekvence **LandCoverUnit_** musí odpovídat obsahu proměnné *ID_prefix*.

do sloupce za sloupcem ID vložit text:

"/> (počáteční uvozovky jsou součástí vkládaného textu !!!)

text je opět stejný pro všechny záznamy

pokud je při tvorbě části **Land Cover Units** v programu HALE studio použita proměnná *ID_suffix* (tj. textový řetězec doplněný do identifikátoru za originální identifikátor polygonu), musí být identický textový řetězec doplněný i na úplný začátek zde vytvářeného textu ve sloupci který následuje za sloupcem ID (ještě před úvodní uvozovky).

Následně se všechny 3 sloupce (před ID, ID a za ID) spojí v excelu funkcí **CONCAT** do jednoho sloupce a přeuloží do dalšího sloupce jako hodnoty. Ve výsledku takto získáme odkazy na jednotlivé Land Cover Units (= polygony území klasifikované dle CORINE), které využijeme při závěrečném vytváření finálního gml souboru, který bude obsahovat jak část Land Cover Dataset, tak i Land Cover Units.

Příklad výsledného řetězce bez použití *ID_suffix*:

<lcv:member xlink:href="http://www.cenia.cz/CLC18_CZ/LandCoverUnit_CZ_14"/>
(projectnamespace = http://www.cenia.cz/CLC18_CZ/; ID_prefix = LandCoverUnit_; ID_suffix
nepoužit; hodnota atributu ID (= identifikátor daného polygonu) ve vstupním shapefile = CZ_14):

Příklad výsledného řetězce při použití *ID_suffix*:

<lcv:member xlink:href="http://www.cenia.cz/CLC18_CZ/LandCoverUnit_CZ_14.abcd"/>
(ID_suffix = .abcd; ostatní viz výše)

Obsah řetězce mezi uvozovkami musí být pro každý polygon **identický** s textovým řetězcem uvedeným jako identifier v části Land Cover Units vytvořené programem HALE studio v sekvenci

<gml:identifier codeSpace=".....">zde_je_uveden_identifier</gml:identifier>

(jedná se o identifikátor, který má v datovém modelu pod sebou vnořený codeSpace)

Příprava části Land Cover Dataset

Použít vzorový soubor *LCD_template1.gml*

Zde použité identifikátory a namespace lze upravovat (viz komentáře ve vzorovém souboru *LCD_template1.gml*), ale takto se výsledný soubor gml správně zobrazí v QGISu a projde INSPIRE validátorem (testy „Interoperable data sets in GML (Guidelines for the Encoding of Spatial data version 3.3)).

Souřadnice uvedené v sekvencích *boundedBy* a *extent* představují obálku ČR – viz:

https://wiki.openstreetmap.org/wiki/Czech_Republic

Bounding box: exact borders ([12.09,51.06,18.87,48.55](#)), rounded ([12,51,1,19,48,5](#))

Do závěrečné části souboru pro Land Cover Dataset vložíme před řádek `</lcv:LandCoverDataset>` odkazy na jednotlivé Land Cover Units připravené dříve (viz kapitola vytvoření odkazů na jednotlivé **Land Cover Units** výše).

příklad odkazu na Land Cover Unit:

```
<lcv:member xlink:href="http://www.cenia.cz/CLC18_CZ/LandCoverUnit_CZ_14"/>
```

Spojení částí Land Cover Dataset a Land Cover Units

- vytvoříme bezpečnostní kopii gml souboru s Land Cover Units
- otevřeme gml soubor s Land Cover Units v editoru Notepad++
- smažeme prvních 9 řádků souboru, tj. řádky

```
<?xml version="1.0" ?>
<base:SpatialDataSet xmlns:sc="http://www.....">
  <base:identifier>
    <base:Identifier>
      <base:localId>clc18_LCUnits1</base:localId>
      <base:namespace>http://www.cenia.cz/CLC18_CZ/</base:namespace>
    </base:Identifier>
  </base:identifier>
  <base:metadata xsi:nil="true"></base:metadata>
```
- smažeme poslední řádek souboru, tj. řádek `</base:SpatialDataSet>`
- všechny výskyty řetězce `base:member` vyměníme hromadně za řetězec `gml:featureMember`
- celý vzorový gml soubor s Land Cover Dataset (LCD_template1.gml) včetně do něho vložených odkazů na Land Cover Units, ale **s výjimkou posledního řádku** vložíme na začátek souboru s Land Cover Units
- poslední řádek souboru s Land Cover Dataset vložíme na konec souboru s Land Cover Units (jedná se o řádek `</gml:FeatureCollection>`)

Takto vzniklý výsledný spojený gml soubor uložíme, přejmenujeme (např. na `clc18_CZ_selection.gml` či na `clc18_CZ_final.gml`), zkontrolujeme jeho zobrazení a atributy v QGISu a v INSPIRE validátoru (alespoň testy pro obecné části dat „**Interoperable data sets in GML (Guidelines for the Encoding of Spatial data version 3.3)**“)

Pozn:

INSPIRE validátor může v současnosti (říjen 2020) načíst maximální velikost gml souboru 50 MB. JRC pracuje na odstranění tohoto omezení.