

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

FAKULTA STAVEBNÍ

STUDIJNÍ PROGRAM GEODÉZIE A KARTOGRAFIE

OBOR GEOMATIKA



DIPLOMOVÁ PRÁCE

NÁVRH ZÁSUVNÉHO MODULU QGIS PRO STAŽENÍ A ANALÝZU

DATOVÉ SADY LUCAS

Vedoucí práce: Ing. Martin Landa, Ph.D.

Katedra geomatiky

2021

Bc. Jaroslav ZEMAN

ZADÁNÍ DIPLOMOVÉ PRÁCE

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: Jaroslav Jméno: Zeman Osobní číslo: 459191

Zadávající katedra: Katedra geomatiky

Studijní program: Geodézie a kartografie

Studijní obor: Geomatika

II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce: Návrh zásuvného modulu QGIS pro stažení a analýzu datové sady LUCAS

Název diplomové práce anglicky: QGIS plugin for downloading and analyzing LUCAS data set

Pokyny pro vypracování:

Cílem diplomové práce je návrh a implementace zásuvného modulu pro program QGIS, který umožní na základě požadavku uživatele stažení vybraných dat z datové sady LUCAS (Land Use and Coverage Area frame Survey). Dále bude možné rozhodnout, zda mají být data časoprostorově agregována, nebo má být každé měření reprezentováno unikátním záznamem. Na závěr získá uživatel podrobnou analýzu stažených dat. Zásuvný modul bude vytvářen pomocí programovacího jazyka Python.

Součástí zadání je také navržení procesu harmonizace a validace datové sady LUCAS.

Seznam doporučené literatury:

Kurt Menke, G.: Mastering QGIS, Packt Publishing, 2015, ISBN: 9781784390068

Mark Pilgrim: Dive into Python 3, Apress, 2009, ISBN: 9781430224150

Marco Ballin, Mauro Masselli, Giulio Barcaroli, Marco Scarnò: Redesign sample for LUCAS 2018, 2018, doi:10.2785/132365

Jméno vedoucího diplomové práce: Ing. Martin Landa, Ph.D.

Datum zadání diplomové práce: 18.2.2021 Termín odevzdání diplomové práce: 17.5.2021

Údaj uveďte v souladu s datem v časovém plánu příslušného ak. roku

Podpis vedoucího práce

Podpis vedoucího katedry

III. PŘEVZETÍ ZADÁNÍ

Beru na vědomí, že jsem povinen vypracovat diplomovou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je nutné uvést v diplomové práci a při citování postupovat v souladu s metodickou příručkou ČVUT „Jak psát vysokoškolské závěrečné práce“ a metodickým pokynem ČVUT „O dodržování etických principů při přípravě vysokoškolských závěrečných prací“.

Datum převzetí zadání

Podpis studenta(ky)

Abstrakt

Obsah této diplomové práce je možné rozdělit do dvou částí. První část se zabývá harmonizací datové sady LUCAS a její publikací na mapovém serveru. Druhá část popisuje tvorbu Python balíčku `pyeumap.lucas` a samotného zásuvného modulu pro platformu QGIS, který tento balíček využívá.

Zásuvný modul poskytuje možnost na základě zvolených parametrů stáhnout pomocí OGC WFS služby body datové sady LUCAS. Po stažení dat jsou body zobrazeny v mapovém okně. Dále zásuvný modul umožňuje práci s těmito daty v podobě agregace tříd krajinného pokrytí.

Klíčová slova

QGIS, zásuvný modul, LUCAS, PostGIS, harmonizace, Python

Abstract

This diploma thesis has two main parts. The first part describes the harmonization of LUCAS dataset and its publication using map server. The second part describes how the Python package `pyeumap.lucas` and the QGIS plugin was created.

The plugin provides possibility to download LUCAS points using OGC WFS for the user. When points are downloaded, they are displayed in the map canvas. The plugin also allows to aggregate land cover classes.

Key words

QGIS, plugin, LUCAS, PostGIS, harmonization, Python

Prohlášení

Prohlašuji, že jsem diplomovou práci na téma „Návrh zásuvného modulu QGIS pro stažení a analýzu datové sady LUCAS“ vypracoval samostatně a veškeré použité zdroje jsem uvedl v kapitole „Použité zdroje“.

V Praze dne

Jaroslav Zeman

.....

.....

Poděkování

Na tomto místě bych rád poděkoval především vedoucímu své diplomové práce, Ing. Martinu Landovi, Ph.D., za mnohé rady, připomínky a vždy pozitivní přístup. Vážím si jeho důvěry, kterou ve mně vložil tím, že mi umožnil stát se členem týmu pracujícím na projektu Geo-harmonizer. Dále děkuji i všem ostatním účastníkům tohoto projektu. Dík patří také Bc. Petře Klémové za kontrolu jazyka v anglicky psané kapitole.

Obsah

Úvod	9
1 LUCAS.....	11
1.1 LUCAS.....	11
1.2 Sběr dat	12
1.3 Kontrola kvality	14
1.4 Klasifikace krajiny	14
1.4.1 Krajinné pokrytí	14
1.4.2 Využití půdy.....	16
1.5 Rešerše	16
2 Harmonizace	18
2.1 Přejmenování atributu	19
2.2 Změna hodnot	20
2.2.1 Atributy <i>lc</i> a <i>lu</i>	22
2.2.2 Atribut <i>obs_dist</i>	23
2.3 Převod na intervaly	23
2.4 Nevalidní hodnoty.....	24
2.5 Změna datových typů.....	26
2.6 Poloha bodu.....	28
2.7 Nové atributy.....	28
2.8 Zveřejnění dat	31
2.9 Automatizace	32
3 Použité technologie	33
3.1 QGIS	33
3.2 Python	34
3.3 PostgreSQL	35
3.4 GeoServer.....	35
3.5 OWSLib	36
4 Postup implementace.....	37
5 Balíček <i>pyeumap.lucas</i>	39
5.1 Sestavení dotazu.....	39
5.2 Vytvoření výstupů.....	41
5.3 Agregace land cover tříd	42
5.4 Jupyter Notebook	43

6	Zásuvný modul.....	44
6.1	Funkcionalita.....	46
6.1.1	Stážení dat.....	46
6.1.2	Agregace tříd.....	48
	Závěr	50
	Použité zdroje.....	53
	Seznam obrázků.....	55
	Seznam tabulek	56
	Seznam zkratk	57
	Seznam příloh.....	58
	QGIS plugin documentation	59

Úvod

Tato diplomová práce se zabývá tvorbou nástroje, který umožní snadný přístup k bodům datové sady LUCAS. Jejím primárním cílem je vznik zásuvného modulu pro platformu QGIS.

Práce vznikla v rámci mezinárodního projektu nazvaného „Geo-harmonizer: EU – wide automated mapping system for harmonization of Open Data based on FOSS4G and Machine Learning“. Na tomto projektu spolupracují odborníci z pěti zemí Evropské unie, a to Německa, Nizozemska, Chorvatska, Rumunska a České republiky. Českou republiku zastupuje ČVUT. Cílem výše zmíněného projektu je harmonizace prostorových dat napříč členskými státy Evropské unie a jejich následná publikace. Snahou je zpřístupnit data nejen odborníkům, ale i široké veřejnosti v takové formě, aby byla snadno použitelná. [1], [2]

V první části diplomové práce je představena datová sada LUCAS. Je zde popsáno, jaká data obsahuje a jaká je jejich využitelnost. Podstatné je znát historii sady LUCAS a její vývoj. Poprvé byla data pro LUCAS sbírána v roce 2006 na území jedenácti států Evropské unie. Od té doby probíhá měření každé tři roky. Nejaktuálnější data jsou tedy v současné době z roku 2018. Nejvíce jsou využívány atributy nazvané *lc* (*land cover* – krajinné pokrytí) a *lu* (*land use* – využití půdy), které především určují charakter krajiny v okolí daného bodu.

Druhá část diplomové práce popisuje způsoby, jakými byla data před jejich publikací upravována a harmonizována. Datová sada LUCAS se neustále vyvíjí. S tím souvisí změny, ke kterým docházelo při každém novém sběru dat. Jelikož od roku 2006 doposud (2018) bylo provedeno pět měření, je možné identifikovat značné množství změn. Aby bylo možné data z jednotlivých let porovnávat, bylo třeba je sjednotit.

Další část popisuje softwarové nástroje, které byly při práci využity. Nejvýznamnějším z nich byl programovací jazyk Python a jeho rozšíření pro program QGIS zvané PyQGIS, pomocí kterého byl vytvářen samotný zásuvný modul. Datová sada LUCAS byla uložena v objektově relační databázi PostgreSQL s rozšířením PostGIS, umožňující práci s prostorovými daty. V tomto prostředí byl využíván dotazovací jazyk SQL. Je zde popsán program QGIS, pro který byl zásuvný modul vytvářen a mapový server, na kterém byla data zveřejněna. V neposlední řadě je zde



zmíněn balíček jazyka Python OWSLib, jehož metody jsou využívány pro získání dat z mapového serveru.

Čtvrtá část obsahuje obecný popis implementace balíčku `pyeumap.lucas` a samotného zásuvného modulu.

V páté kapitole je představena funkcionality balíčku `pyeumap.lucas`. Dále je zde zmíněn Jupyter Notebook, jenž byl vytvořen jako názorná ukázka možností využití tohoto balíčku.

V závěrečné části je vysvětlena tvorba zásuvného modulu. Zároveň jsou tu zmíněny možnosti, jak zásuvný modul používat tak, aby uživatel obdržel požadovaná data.

1 LUCAS

1.1 LUCAS

LUCAS je zkratkou Land Use and Coverage Area frame Survey („Průzkum využití půdy a krajinného pokrytí“). Některé státy vytvářejí vlastní produkty mapující svá území, avšak tyto produkty nejsou vzájemně kompatibilní. Z toho důvodu není možné získat na jejich základě dobrou představu o stavu a vývoji krajiny v celé Evropské unii. Od roku 2006 vytváří Eurostat databázi bodů nazvanou LUCAS, popisující charakter krajiny na území Evropské unie. Díky této databázi je možné detailně analyzovat stav a změny v krajině napříč státy.

Vznik databáze LUCAS potvrdily Evropský parlament a Rada Evropy dne 22.5. 2000 rozhodnutím 1445/2000/ES „On the application of aerial-survey and remote-sensing techniques to the agricultural statistics for 1999 to 2003“ („O použití technik leteckého průzkumu a dálkového průzkumu Země pro zemědělské statistiky v letech 1999 až 2003“). První roky probíhaly přípravy a testování. Počínaje rokem 2006 je průzkum prováděn každé 3 roky. Dosud bylo provedeno pět měření, a to v letech 2006, 2009, 2012, 2015 a 2018. Na obrázku (Obr. 1) jsou znázorněny body měřené v roce 2018.



Obr. 1: Rozložení LUCAS bodů měřených v roce 2018

Při terénním šetření se zjišťuje nejen současný krajinný pokryv a využití území, ale také informace o životním prostředí. Pořizují se fotografie a u vybraných bodů se odebírají vzorky půdy, které jsou následně analyzovány v laboratoři. Postupně se rozšiřuje nejen počet sledovaných atributů (charakteristiky popisující území), ale také počet států, na jejichž území je měření prováděno. Tabulka (Tab. 1) zobrazuje počet států, na jejichž území bylo měřeno v jednotlivých letech a zároveň počty měřených bodů.

Tab. 1: Rozsah datové sady LUCAS

Rok	Počet států EU	Počet bodů
2006	11	168 402
2009	23	234 623
2012	27	270 272
2015	28	339 696
2018	28	337 854

Data LUCAS¹ se využívají v mnoha rozdílných oborech nejen pro ochranu přírody, lesní a vodní hospodářství, městské plánování, zemědělství, ale například i pro monitorování biologické rozmanitosti.

[3], [4]


1.2 Sběr dat

Měřené body jsou rozmístěny po celém zájmovém území (členské státy Evropské unie) v pravidelné mřížce 2 km od sebe. Body lze rozdělit do dvou skupin. První skupinu tvoří body, ke kterým se lze bez problému dostat, v tom případě jsou měřeny v terénu (in situ). Do druhé skupiny patří body, k nimž není přístup možný, případně by byl příliš náročný, tyto body jsou fotointerpretovány v kanceláři. V roce 2018 bylo necelých 240 000 bodů navštíveno v terénu a pouze cca 100 000 bodů interpretováno v kanceláři. Z těchto čísel vyplývá, že data popisující pouze necelou třetinu celkového počtu bodů nebyla určena terénním šetřením.


¹ <https://ec.europa.eu/eurostat/web/lucas/data/primary-data>

Průzkum v terénu provádí vyškolený pracovník, který s pomocí podrobného popisu jednotlivých sledovaných vlastností charakterizuje daný bod a jeho okolí. Údaje jsou zaznamenávány do dotazníku, který pro většinu atributů obsahuje výčet možností. Jednu z částí dotazníku, zabývající se krajinným pokryvem a využitím půdy, znázorňuje obrázek (Obr. 2). Tímto způsobem se předchází vyplňování nevalidních informací. Je snaha o to, aby všichni pracovníci, podílející se na sběru dat, používali stejné metody a výsledek byl tak objektivní.


3 LAND COVER AND LAND USE							
28	LC1 land cover 1:	31	LC2 land cover 2:	34	LU1 land use 1:	37	LU2 land use 2:
29	LC1 plant species:	32	LC2 plant species:	35	LU1 land use type:	38	LU2 land use type:
30	LC1 coverage (%):	33	LC2 coverage (%):	36	LU1 coverage (%):	39	LU2 coverage (%):




10%




25%



50%



75%



90%

40	Parcel area (ha):		
	1 <input type="checkbox"/> area < 0.1		
	2 <input type="checkbox"/> 0.1 ≤ area < 0.5		
	3 <input type="checkbox"/> 0.5 ≤ area < 1		
	4 <input type="checkbox"/> 1 ≤ area < 10		
	5 <input type="checkbox"/> area ≥ 10		
	N.R.		

Obr. 2: Ukázka části dotazníku pro sběr LUCAS dat [5]

Informace o bodech, které není možné navštívit fyzicky, jsou získávány fotointerpretací. Ta je prováděna pomocí ortofoto snímků, případně fotografií z terénu.

Každý bod je podrobně popsán velkým množstvím vlastností. Mezi zjišťované charakteristiky patří například poloha bodu, datum měření, krajinný pokryv, využití území, výška stromů, sklonitost terénu a mnohé další.

[3], [4]

1.3 Kontrola kvality

Data LUCAS jsou kontrolována různými způsoby ve třech fázích. Nejprve je automatizovaně zjišťována úplnost a konzistence dat. Následně jsou veškeré body kontrolovány vizuálně. Ve třetí fázi jsou body kontrolovány jednotlivě nezávislým kontrolorem kvality. Tímto způsobem je ověřována přesnost údajů u následujících bodů:

- 36 % bodů z celé datové sady
- prvních 20 % bodů měřených novým pracovníkem

Výše zmíněné kontroly odhalí určitou chybu, eventuálně chybějící informaci u zhruba 7 % z celkového počtu bodů. Nedostatky jsou oznámeny terénnímu pracovníkovi, který je následně opraví.

[3]

1.4 Klasifikace krajiny

Jak již bylo zmíněno, LUCAS využívá pro popis krajiny dva hlavní parametry, kterými jsou krajinné pokrytí (*land cover – lc*) a využití půdy (*land use – lu*). Pro tyto parametry byly vytvořeny nomenklatury, pomocí kterých je možno přesně definovat charakter daného místa. Jelikož se v okolí každého bodu může vyskytovat vícero krajinných typů, případně toto okolí může být využíváno vícero způsoby, určuje pracovník v terénu hlavní třídu (*lc1/lu1*) a vedlejší třídu (*lc2/lu2*). Zároveň stanovuje procentuální zastoupení těchto tříd na daném bodě a v jeho nejbližším okolí. Podrobný popis toho, co která třída obsahuje, je dostupný na webových stránkách Eurostatu [6].

1.4.1 Krajinné pokrytí

Krajinné pokrytí určuje fyzický a biologický pokryv zemského povrchu, včetně uměle vytvořených ploch, zemědělských oblastí, lesů, přirozených a částečně přirozených oblastí, mokřadů, vodních těles. LUCAS v současné době používá 8 kategorií, které se dále dělí na 29 tříd a 76 podtříd. Hlavní kategorie jsou označeny velkými písmeny A-H. Pro podrobnější dělení jsou používány kódy složené z písmene označujícího kategorii a čísla. Součástí zmíněných 76 podtříd jsou dvě specifické, jejichž označení je odlišné od ostatních. Těmito třídami jsou Bx1 a Bx2, které jsou

využívány pouze v případě, je-li bod fotointerpretován. Kód Bx1 představuje zemědělsky využívanou půdu, která je každoročně obdělávána, avšak není možné rozlišit plodinu, která zde v daný moment roste. Pokud by byl bod navštíven v terénu, byl by označen některým z kódů v rozmezí B11-B55. Třída Bx2 popisuje trvalé kultury. Stejně jako v případě třídy Bx1 není možné z ortofoto snímku přesně určit druh plodiny. V tomto případě by při podrobnějším průzkumu byla stanovena jedna ze tříd B71-B84. Obrázek (Obr. 3) vyobrazuje první a druhou úroveň dělení atributu *lc*. [6], [7]

Land cover			
A00	ARTIFICIAL LAND	A10	Roofed built-up areas
		A20	Artificial non-built up areas
		A30	Other artificial areas
B00	CROPLAND	B10	Cereals
		B20	Root crops
		B30	Non-permanent industrial crops
		B40	Dry pulses, vegetables and flowers
		B50	Fodder crops
		B70	Permanent crops: fruit trees
		B80	Other permanent crops
C00	WOODLAND	C10	Broadleaved woodland
		C20	Coniferous woodland
		C30	Mixed woodland
D00	SHRUBLAND	D10	Shrubland with sparse tree cover
		D20	Shrubland without tree cover
E00	GRASSLAND	E10	Grassland with sparse tree/shrub cover
		E20	Grassland without tree/shrub cover
		E30	Spontaneously re-vegetated surfaces
F00	BARE LAND AND LICHENS/MOSS	F10	Rocks and stones
		F20	Sand
		F30	Lichens and moss
		F40	Other bare soil
G00	WATER AREAS	G10	Inland water bodies
		G20	Inland running water
		G30	Transitional water bodies
		G40	Sea and ocean
		G50	Glaciers, permanent snow
H00	WETLANDS	H10	Inland wetlands
		H20	Coastal wetlands

Obr. 3: Třídy krajinného pokrytí (*lc*) [3]

1.4.2 Využití půdy

Atribut *lu* popisuje, jak je krajina využívána. Tímto způsobem můžeme určit, zda je dané území užíváno pro bydlení, průmysl, zemědělství, dopravu či další účely. LUCAS v současné době rozlišuje 4 hlavní *lu* kategorie, které se dělí do 16 tříd a ty následně do 40 podtříd. Kódy jednotlivých tříd jsou vytvořeny kombinací písmene „U“ a číselného označení. Obrázek (Obr. 4) znázorňuje rozdělení *lu* tříd první a druhé úrovně. [6]

Land use			
U100	PRIMARY SECTOR	U110	Agriculture
		U120	Forestry
		U130	Aquaculture and fishing
		U140	Mining and quarrying
		U150	Other primary production
U 200	SECONDARY SECTOR	U210	Energy production
		U220	Industry and manufacturing
U300	TERTIARY SECTOR, TRANSPORT, UTILITIES & RESIDENTIAL	U310	Transport, communication networks, storage, protection works
		U320	Water and waste treatment
		U330	Construction
		U340	Commerce, financial, professional and information services
		U350	Community services
		U360	Recreation, leisure, sport
		U361	Residential
U400	UNUSED AND ABANDONED AREAS	U410	Abandoned areas
		U420	Semi-natural and natural areas not in use

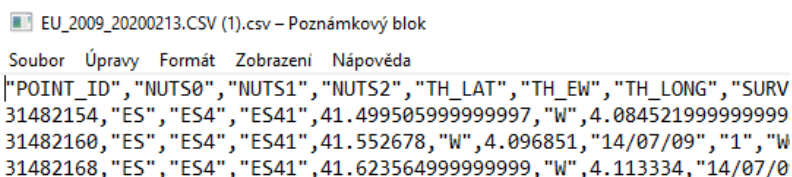
Obr. 4: Třídy využití půdy (*lu*) [3]

1.5 Rešerše

Přestože jsou data z datové sady LUCAS hojně využívána v mnoha různých oborech, nebyl v minulosti velký zájem o jejich harmonizaci a jejich snadnější zpřístupnění většímu množství uživatelů.

V současné době jsou data odpovídající jednotlivým rokům dostupná na webových stránkách tvůrce této datové sady, tedy Evropského statistického úřadu (Eurostat). Tato data však mají dva významné nedostatky. Prvním z nich je skutečnost, že nejsou harmonizována, což ztěžuje možnost jejich porovnávání napříč roky. Druhým nedostatkem je forma, ve které jsou dostupná. Data je možné stáhnout pouze ve formátu CSV. Názvy atributů jsou zaznamenány v prvním řádku dokumentu. Následující řádky obsahují jednotlivá měření, hodnoty jsou oddělené čárkou. Na obrázku (Obr. 5) je příklad části CSV souboru obsahujícího data měřená v roce 2009. Tento formát není pro prostorová data vhodný, proto je pro efektivní práci s nimi nutný jejich převod

do jiného formátu. Pro běžného uživatele je tento způsob náročný. Další nevýhodou uložení dat ve formátu CSV je, že i v případě potřeby pouze určité části datové sady je nutné stáhnout celý soubor a z něj data vybírat.


 EU_2009_20200213.CSV (1).csv – Poznámkový blok
 Soubor Úpravy Formát Zobrazení Nápověda
 ["POINT_ID", "NUTS0", "NUTS1", "NUTS2", "TH_LAT", "TH_EW", "TH_LONG", "SURV
 31482154, "ES", "ES4", "ES41", 41.499505999999997, "W", 4.0845219999999999
 31482160, "ES", "ES4", "ES41", 41.552678, "W", 4.096851, "14/07/09", "1", "W
 31482168, "ES", "ES4", "ES41", 41.623564999999999, "W", 4.113334, "14/07/0

Obr. 5: Ukázka části souboru s originálními daty

Výše popsané problémy částečně vyřešil tým, jehož součástí byli zaměstnanci Joint Research Centre, Eurostatu a GOPA Luxembourg. Způsob, jakým s daty pracovali, je popsán v článku *Harmonised LUCAS in-situ land cover and use database for field surveys from 2006 to 2018 in the European Union* [4]. Hlavním cílem jejich práce byla harmonizace LUCAS. Výsledná data je možno snadno porovnávat napříč roky. Byly vytvořeny nové atributy, kterými jsou například: rok měření daného bodu, počet opakovaných navštívení bodu a geometrie popisující polohu bodu. Hodnoty atributů, které jsou v původní databázi vyjádřeny kódem, byly převedeny na textový řetězec. Výhodou této změny je, že z hodnoty atributu je na první pohled zřejmá vlastnost, kterou popisuje, není tedy nutná tabulka interpretující jednotlivé kódy. Nevýhodou uložení hodnot atributů pomocí textových řetězců je obtížnější práce s nimi. Data² je možné stáhnout pomocí FTP (File Transfer Protocol).

Informace popisující body jsou uloženy v souborech *CSV*. Pouze geometrie se základní identifikací jednotlivých bodů jsou uloženy ve formátu *shapefile*. Pro plnohodnotné využívání veškerých informací je uživatel nucen data předzpracovat.

[4]

² https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/LUCAS/LUCAS_harmonised/

2 Harmonizace

V databázi LUCAS se vyskytuje mnoho nekonzistencí napříč roky. Jedním z důvodů, proč tomu tak je, je postupný vývoj a rozšiřování počtu měřených parametrů území. Tabulka (Tab. 2) zobrazuje jejich počet v jednotlivých letech.

Tab. 2: Počet LUCAS atributů

Rok	Počet atributů
2006	20
2009	44
2012	46
2015	59
2018	97

Vznik nových atributů v průběhu let nebyl jediným typem změny. Dalšími změnami, které v mnoha případech zneprůhlednily databázi jsou: zánik atributu (přestal být měřen), změna názvu atributu, změna hodnot, které daný atribut nabývá, a změna datového typu. Problémem, který se v databázi LUCAS vyskytuje v nemalé míře, je to, že některé atributy nabývají neplatných hodnot. Neplatnou hodnotou rozumíme takovou hodnotu, která není zmíněna v dokumentaci. Výše zmíněné nekonzistence a chyby byly identifikovány, a ty, které bylo možné opravit, byly opraveny. Snahou bylo veškeré změny vztahovat k verzi v současnosti nejaktuálnější, tedy k atributům z roku 2018.

Jako vzor pro harmonizaci dat sloužil v mnoha případech článek *Harmonised LUCAS in-situ land cover and use database for field surveys from 2006 to 2018 in the European Union* [4] popsáný v kapitole 1.5, jenž se touto problematikou zabývá. V následujících podkapitolách jsou popsány jednotlivé případy, ve kterých se přístupy k danému problému, využití pro tuto práci, s článkem shodují a také ty, ve kterých byl postup odlišný.

Před samotnou harmonizací bylo nutné data ve formátu CSV, stažená z Eurostatu, importovat do databázového softwaru. Následně pomocí SQL dotazů bylo možné provést podrobnou analýzu veškerých atributů v jednotlivých letech. Na základě této analýzy bylo definováno několik typů úprav, které byly implementovány do datové

sady. Příklady jednotlivých změn, které databázi upravovaly, jsou popsány v následujících podkapitolách.

2.1 Přejmenování atributu

V případě, že se lišil pouze název atributu, byl změněn do podoby, kterou měl v roce 2018. Taková změna byla provedena například u atributu *protected_area*, který byl přejmenován na *special_status*. Název *protected_area* byl používán v letech 2012 a 2015. V roce 2018 se již atribut s tímto názvem v datové sadě nevyskytuje, avšak nahradil ho atribut *special_status*, obsahující totožné hodnoty.

Atributy byly přejmenovávány stejným způsobem, jaký je popsán v článku [4], s výjimkou atributu *soil_crop*, vyskytujícího se v letech 2009, 2012 a 2015. Tento atribut byl přejmenován na *crop_residues*, což souvisí s vlastností, kterou popisuje. V letech 2009, 2012 a 2015 představuje tento atribut intervaly procentuálního zastoupení zbytků plodin na poli po sklizni. V roce 2018 je pomocí atributu *crop_residues* rozlišováno pouze, zda se na poli nějaké posklizňové zbytky plodin nacházely či nikoli. Jelikož dle mého názoru není podstatný podíl zmíněných zbytků, ale jejich výskyt, byly atributy spojeny do jednoho s názvem *crop_residues*. Intervaly procentuálního zastoupení používané do roku 2015 byly převedeny tak, že v případě zastoupení do 10 % byla nastavena hodnota 2 (žádné zbytky se na poli nenacházejí). Intervaly, které představovaly podíl větší než 10 %, byly změněny na hodnotu 1 (na poli se vyskytují posklizňové zbytky).

Pro automatické vygenerování SQL dotazů sloužících k přejmenování atributů byl pomocí jazyka Python vytvořen skript. Vstupem pro tento skript je CSV soubor obsahující změny názvů atributů. Obsah tohoto CSV souboru, a tedy i změny názvů atributů, zobrazuje tabulka (Tab. 3).

Tab. 3: Změny názvů atributů

Rok	Původní název	Nový název	Rok	Původní název	Nový název
2006	surv_date	survey_date	2012	photo_s	photo_south
2006	x_laea	th_lat	2012	photo_p	photo_point
2006	y_laea	th_long	2012	tree_height_srv	tree_height_survey
2009	area_size	parcel_area_ha	2012	tree_height_mat	tree_height_maturity
2009	lc1_pct	lc1_perc	2012	soil_plough	lndmng_plough
2009	lc2_pct	lc2_perc	2012	gps_alt	gps_altitude
2009	lc1_species	lc1_spec	2012	protected_area	special_status
2009	lc2_species	lc2_spec	2012	soil_stones	soil_stones_perc
2009	land_mngt	grazing	2015	area_size	parcel_area_ha
2009	obs_dir	obs_direct	2015	lc1_pct	lc1_perc
2009	photo_e	photo_east	2015	lu1_pct	lu1_perc
2009	photo_w	photo_west	2015	lu2_pct	lu2_perc
2009	photo_n	photo_north	2015	lc2_pct	lc2_perc
2009	photo_s	photo_south	2015	lc1_species	lc1_spec
2009	photo_p	photo_point	2015	lc2_species	lc2_spec
2009	tree_height_srv	tree_height_survey	2015	land_mngt	grazing
2009	soil_plough	lndmng_plough	2015	obs_dir	obs_direct
2009	gps_alt	gps_altitude	2015	photo_e	photo_east
2009	soil_stones	soil_stones_perc	2015	photo_w	photo_west
2012	area_size	parcel_area_ha	2015	photo_n	photo_north
2012	lc1_pct	lc1_perc	2015	photo_s	photo_south
2012	lc2_pct	lc2_perc	2015	photo_p	photo_point
2012	lc1_species	lc1_spec	2015	protected_area	special_status
2012	lc2_species	lc2_spec	2015	soil_plough	lndmng_plough
2012	land_mngt	grazing	2015	pi_extension	office_pi
2012	obs_dir	obs_direct	2015	trees_height_maturity	tree_height_maturity
2012	photo_e	photo_east	2015	trees_height_survey	tree_height_survey
2012	photo_w	photo_west	2015	soil_stones	soil_stones_perc
2012	photo_n	photo_north			

2.2 Změna hodnot

Pokud byly hodnoty atributu odlišné v jednotlivých letech a zároveň nabývaly platných hodnot, byly tyto hodnoty změněny tak, aby byly konzistentní. Tato změna byla realizována například u atributu *lc_lu_special_remark* a dokumentuje ji tabulka (Tab. 4). Z tabulky je zřejmé, že rozdílné kódové hodnoty popisují stejnou vlastnost.

Například v roce 2015 je dočasně zaplavené území (Temporary flooded) označeno hodnotou 10, avšak v roce 2018 je tato skutečnost zaznamenána hodnotou 9.

Tab. 4: Změny hodnot atributu *lc_lu_special_remark*

Původní kód (2015)	Popis (2015)	Nový kód (2018)	Popis (2018)
1	Tilled and/or sowed	2	Tilled/sowed
2	Harvested field	1	Harvested field
7	No remark	10	No remark
8	Not relevant	88	Not relevant
9	Temporary dry	8	Temporary dry
10	Temporary flooded	9	Temporary flooded

Jednotlivé případy, u kterých byly měněny hodnoty atributů, byly pečlivě posuzovány. Z důvodu, aby byla úprava provedena správným způsobem a nevznikly nové nepřesnosti, byla využita nejen dokumentace dostupná na stránkách Eurostatu [8], ale také výše zmíněný článek [4] a jeho výsledky. Převážná většina těchto změn byla totožná se změnami uvedenými v článku [4].

Dílní korekce byly zapisovány do pěti oddělených souborů podle roku, pro který měla být změna aplikována. Zápis byl prováděn ve formě SQL příkazů. Příklad části dokumentu, obsahujícího změny hodnot atributů pro rok 2015, je znázorněn na obrázku (Obr. 6).

```
*04b_harmonize_values_2015.sql – Poznámkový blok
Soubor Úpravy Formát Zobrazení Nápověda
UPDATE lucas2015 SET lc_lu_special_remark = '22' WHERE lc_lu_special_remark = '1';
UPDATE lucas2015 SET lc_lu_special_remark = '1' WHERE lc_lu_special_remark = '2';
UPDATE lucas2015 SET lc_lu_special_remark = '2' WHERE lc_lu_special_remark = '22';
UPDATE lucas2015 SET lc_lu_special_remark = '100' WHERE lc_lu_special_remark = '7';
UPDATE lucas2015 SET lc_lu_special_remark = '88' WHERE lc_lu_special_remark = '8';
UPDATE lucas2015 SET lc_lu_special_remark = '8' WHERE lc_lu_special_remark = '9';
UPDATE lucas2015 SET lc_lu_special_remark = '9' WHERE lc_lu_special_remark = '10';
UPDATE lucas2015 SET lc_lu_special_remark = '10' WHERE lc_lu_special_remark = '100';
```

Obr. 6: Ukázka souboru obsahujícího SQL příkazy pro změnu hodnot atributů

2.2.1 Atributy *lc* a *lu*

V případě atributů, popisujících krajinné pokrytí (*lc1*, *lc2*) a využití půdy (*lu1*, *lu2*), byly jejich hodnoty harmonizovány odlišně oproti ostatním. Pro vytvoření SQL příkazů byly využity tabulky vytvořené Ing. Tomášem Boučkem. Tyto tabulky byly vytvořeny pro atributy *lu* a *lc* jednotlivých let. V řádcích jsou zaznamenány třídy používané ve vybraném roce a ve sloupcích třídy z roku 2018. Pokud si dvě zvolené třídy odpovídají, je v buňce na průsečíku daného řádku a sloupce hodnota 1. V opačném případě obsahuje buňka hodnotu 0.

Na základě výše popsaných tabulek byly pomocí skriptu napsaného v jazyce Python automatizovaně vygenerovány SQL příkazy. Pokud se v jednom řádku tabulky vyskytovalo více hodnot 1 (danou třídu je možné převést do vícero tříd z roku 2018), nastavena taktéž u tříd, v jejichž řádku byly pouze hodnoty 0 (daná třída zanikla a v současnosti neexistuje žádná ekvivalentní třída). Z důvodu, aby byly zachovány i původní kódy, byly pro atributy *land cover* a *land use* vytvořeny odpovídající atributy *lc1_h*, *lc2_h*, *lu1_h*, *lu2_h*, do kterých byly zapsány harmonizované hodnoty.

Jiný způsob harmonizace těchto atributů je popsán v článku [4]. Zde je zřejmá snaha harmonizovat veškeré hodnoty za každou cenu. V případě, že se některá třída v průběhu let rozdělila do vícero podrobných tříd, je tato třída změněna na odpovídající třídu nadřazenou. Příkladem může být třída C12 „Coniferous forest“ z roku 2006. V roce 2018 existuje odpovídající třída C20 „Coniferous woodland“, avšak ta se dále dělí na podtřídy C21 „Spruce dominated coniferous woodland“, C22 „Pine dominated coniferous woodland“ a C23 „Other coniferous woodland“. Harmonizace popsaná v článku [4] spočívá ve změně třídy C12 na třídu C20. Tímto způsobem však dochází ke kombinaci *land cover* tříd různých úrovní. Třídy, které byly použity v některém měření v letech 2006, 2009, 2012, 2015 a nemají ekvivalent v roce 2018 (třída zanikla), nejsou nijak upraveny. Tato skutečnost znepráhledňuje daný atribut. Příkladem je třída *land use* U500 „Wetland“, která existovala pouze v roce 2006.

2.2.2 Atribut *obs_dist*

Atribut *obs_dist* v původní databázi obsahoval v roce 2006 kódy představující intervaly vzdáleností. V následujících letech byly v tomto atributu uloženy absolutní hodnoty vzdáleností každého teoretického bodu (význam termínu teoretický bod je popsán v kapitole 2.6) od jeho bodu měřeného. Harmonizace tohoto atributu byla provedena následujícím způsobem. V případě, že byly u daného bodu měřeny souřadnice bodu, na kterém byla prováděna observace, byla vzdálenost k teoretickému bodu dopočítána a do atributu uložena tato její nová hodnota. Pokud zmíněné souřadnice určeny nebyly, byl do atributu *obs_dist* vložen střed intervalu, jenž byl reprezentován původním kódem. Intervaly a odpovídající kódy popisuje následující tabulka (Tab. 5).

Tab. 5: Harmonizace atributu *obs_dist*

<i>obs_dist</i> (2006)		
Původní kód	Interval vzdáleností [m]	Střed intervalu (nová hodnota) [m]
1	0-3	1,5
2	3-50	24,5
3	50-100	75,0
4	> 100	101,0

2.3 Převod na intervaly

V tomto případě byl způsob úpravy podobný tomu, který byl popsán v kapitole 2.2. Změna se týkala atributů *lc1_perc*, *lc2_perc*, *lu1_perc*, *lu2_perc*, popisujících procentuální zastoupení jednotlivých *land cover* a *land use* tříd a atributu *soil_stones_perc*, který určuje, na jaké části území se nacházejí kameny. Na rozdíl od příkladu z kapitoly 2.2 nebylo možné použít jako referenční atribut ten z roku 2018, a to z následujícího důvodu. V roce 2018 bylo procentuální zastoupení určováno v absolutních procentech, nikoliv v procentuálních intervalech označených kódem, jak tomu bylo v předchozích letech. Jednotlivé hodnoty tedy bylo nutné agregovat do tříd tak, aby byla data konzistentní. Výhodou této agregace je zpřehlednění a jednodušší práce s daty. Tabulka (Tab. 6) popisuje použité intervaly převzaté z roku 2009 pro atribut *lc1_perc*.

Tab. 6: Hodnoty atributu *lc1_perc*

Kód	<i>lc1_perc</i> [%]
0	0
1	$1 \leq lc < 10$
2	$10 \leq lc < 25$
3	$25 \leq lc < 50$
4	$50 \leq lc < 75$
5	$75 \leq lc \leq 100$

Stejný postup je popsán i v článku [4]. Jediným rozdílem je způsob práce s nevalidními hodnotami, které se mimo jiné i ve výše zmíněných attributech vyskytují. Tento problém je popsán v následující kapitole.

2.4 Nevalidní hodnoty

Nejobtížnější korekcí byla úprava hodnot atributů, které byly nevalidní, tedy se vyskytovaly v databázi, přestože by v ní dle dokumentace být neměly. V mnoha případech byla touto nevalidní hodnotou 0 nebo prázdný řetězec. Každá neplatná hodnota byla detailně zkoumána. Především byl porovnáván počet bodů touto hodnotou označený s počty bodů, popsáných jednotlivými hodnotami stejného atributu v ostatních letech. Tam, kde bylo zřejmé, jakou hodnotu by měly zmíněné body nabývat, byla tato hodnota změněna, v opačném případě byla hodnota nahrazena kódem -1, reprezentujícím nevalidní hodnotu. Obě výše zmíněné eventuality byly využity například u atributu *wm_delivery*. Tabulka (Tab. 7) zobrazuje počty bodů v jednotlivých letech, kterým odpovídá daná hodnota atributu. Nevalidní hodnoty se vyskytují v roce 2009 a jsou označeny červeně. Z tabulky je zřejmé, že nevalidní hodnota 0 odpovídá hodnotě 8, jelikož v následujících letech vždy počet bodů označených hodnotou 8 značně převyšoval 200 000 bodů a zároveň obdobný počet bodů je v roce 2009 označen neplatnou hodnotou 0. U dalších nevalidních kódů, kterými jsou v daném případě 5, 6, 10 a 12, již není možné jednoznačně určit, co představují. Tyto kódy byly proto změněny na -1.

Tab. 7: Počet bodů atributu *wm_delivery* v jednotlivých letech

Kód	Počet bodů (2009)	Počet bodů (2012)	Počet bodů (2015)	Počet bodů (2018)
0	227 586	-	-	-
1	420	607	853	545
2	1 167	1 220	2 142	804
3	51	4 671	5 251	4 290
4	4 015	2 370	15 398	2 435
5	16	-	-	-
6	8	-	-	-
8	1 353	261 404	316 052	329 780
10	2	-	-	-
12	5	-	-	-

V databázi LUCAS bodů bylo identifikováno značné množství nevalidních hodnot. Jejich dalším příkladem mohou být atributy určující procentuální zastoupení *land cover* a *land use* tříd v roce 2018. V tomto roce obsahovaly atributy absolutní hodnoty procent. Jelikož se daný typ krajiny může nacházet maximálně na 100 % území, byly hodnoty větší než 100 opět nahrazeny kódem -1 pro nevalidní data.

Odlišný postup byl popsán v článku [4], podle kterého není převážná část neplatných hodnot nijak upravena. Tento přístup umožňuje uživateli pracovat i s hodnotami, u kterých není možné jednoznačně určit, jakou skutečnost představují. Uživatel je v tomto případě nucen si domýšlet, co daná hodnota reprezentuje. Domnívám se, že téměř každý, kdo bude data LUCAS využívat, bude mít zájem pouze o hodnoty, u kterých je jisté, co představují. Nevalidní hodnoty tedy databázi pouze znepřehledňují.

2.5 Změna datových typů

Další úpravou, kterou bylo nutné aplikovat na data, bylo nastavení odpovídajících datových typů jednotlivým atributům. Byl vytvořen skript pomocí jazyka Python, jehož vstupem byl CSV soubor, definující atributům jejich datové typy. Skript automaticky vygeneroval SQL dotazy sloužící ke změně těchto typů. Tabulka (Tab. 8) obsahuje názvy atributů spolu s příslušnými datovými typy. Jelikož před úpravou byl datovým typem všech atributů textový řetězec (string), obsahuje tabulka pouze atributy, jejichž datový typ měl být jiný.

Tab. 8: Změny datových typů atributů

Atribut	Datový typ	Atribut	Datový typ
bio_sample	integer	lc2_perc	integer
bulk0_10_sample	integer	lm_grass_margins	integer
bulk10_20_sample	integer	lm_plough_direct	integer
bulk20_30_sample	integer	lm_plough_slope	integer
car_ew	integer	lm_stone_walls	integer
car_latitude	float	lndmng_plough	integer
car_longitude	float	lu1_perc	integer
cprn_cando	integer	lu2_perc	integer
cprn_impervious_perc	integer	obs_direct	integer
cprn_lc1e_brdth	integer	obs_dist	float
cprn_lc1n	integer	obs_radius	integer
cprn_lc1n_brdth	integer	obs_type	integer
cprn_lc1s_brdth	integer	office_pi	integer
cprn_lc1w_brdth	integer	organic_sample	integer
cprn_urban	integer	parcel_area_ha	integer
cprnc_lc1e	integer	photo_east	integer
cprnc_lc1s	integer	photo_north	integer
cprnc_lc1w	integer	photo_point	integer
crop_residues	integer	photo_south	integer
erosion_cando	integer	photo_west	integer
eunis_complex	integer	point_id	integer
ex_ante	integer	soil_bio_taken	integer
feature_width	integer	soil_blk_0_10_taken	integer
gps_altitude	integer	soil_blk_10_20_taken	integer
gps_ew	integer	soil_blk_20_30_taken	integer
gps_prec	integer	soil_org_depth_cando	integer
gps_proj	integer	soil_std_taken	integer
grass_cando	integer	soil_stones_perc	integer
grassland_sample	integer	soil_taken	integer
grazing	integer	special_status	integer

inspire_plcc1	integer	standard_sample	integer
inspire_plcc2	integer	survey_date	date
inspire_plcc3	integer	th_ew	integer
inspire_plcc4	integer	tree_height_maturity	integer
inspire_plcc5	integer	tree_height_survey	integer
inspire_plcc6	integer	wm	integer
inspire_plcc7	integer	wm_delivery	integer
inspire_plcc8	integer	wm_source	integer
lc_lu_special_remark	integer	wm_type	integer
lc1_perc	integer		

Z tabulky (Tab. 8) je zřejmé, že datovým typem převážné většiny atributů je ve výsledné databázi celé číslo (integer). Tato skutečnost je způsobena tím, že vlastnosti atributů jsou vyjádřeny pomocí číselných kódů, jejichž význam je vysvětlen v dodatku této práce (Tab. 17).

Jiný přístup k datovým typům a s tím souvisejícím hodnotám, které atributy nabývají, je popsán v článku [4]. Skutečnosti, které hodnoty jednotlivých atributů popisují, mají v mnoha případech formu textového řetězce. Tento textový řetězec dává uživateli na první pohled lepší představu o tom, co daná hodnota vyjadřuje. Příkladem může být atribut *grazing*. Tabulka (Tab. 9) zobrazuje hodnoty, které atribut *grazing* nabývá v databázi vytvořené na základě článku [4], a pro porovnání i odpovídající hodnoty, obsahující tento atribut po úpravách provedených na základě této práce.

Tab. 9: Tabulka hodnot atributu *grazing*

grazing	
Hodnoty dle článku [4]	Hodnoty dle této práce
Signs of grazing	1
No signs of grazing	2
Not relevant	8

Nevýhodou zmíněného přístupu je především náročná práce s textovými řetězci. Například při výběru na základě atributového dotazu je uživatel nucen vypsát přesně celý řetězec a sebemenší chyba způsobí nefunkčnost dotazu. Další nevýhodou je

nepřehlednost atributové tabulky obsahující mnoho textových řetězců. Z těchto důvodů byly v této práci upřednostněny kódy, reprezentující jednotlivé vlastnosti.

2.6 Poloha bodu

Zcela zásadní pro práci s datovou sadou LUCAS je poloha jednotlivých bodů. Atributová tabulka rozlišuje tři různá umístění daného bodu. Atributy *car_latitude*, *car_longitude* obsahují souřadnice bodu, kde pracovník měřící daný bod zaparkoval auto. Tento atribut je nápomocný především při opakovaném měření bodu. Další dva způsoby určení polohy bodu jsou pro uživatele datové sady LUCAS užitečnější. Atributy *th_lat* a *th_long* určují polohu tzv. teoretického bodu. Teoretický bod je bod pravidelné mřížky, na kterém by mělo měření probíhat. Ne vždy je však možné dosáhnout teoretického bodu, z tohoto důvodu existují atributy *gps_lat* a *gps_long*, které obsahují souřadnice bodu, kde skutečně měření probíhalo. Tyto souřadnice by v ideálním případě měly být totožné se souřadnicemi bodu teoretického.

Způsob práce s atributy popisujícími polohu bodů je podrobněji popsán v následující kapitole.

2.7 Nové atributy

Poslední úpravou datové sady LUCAS bylo vytvoření nových atributů. Tyto nové atributy data zpřehledňují a usnadňují práci s nimi.

Nově přidány byly následující atributy:

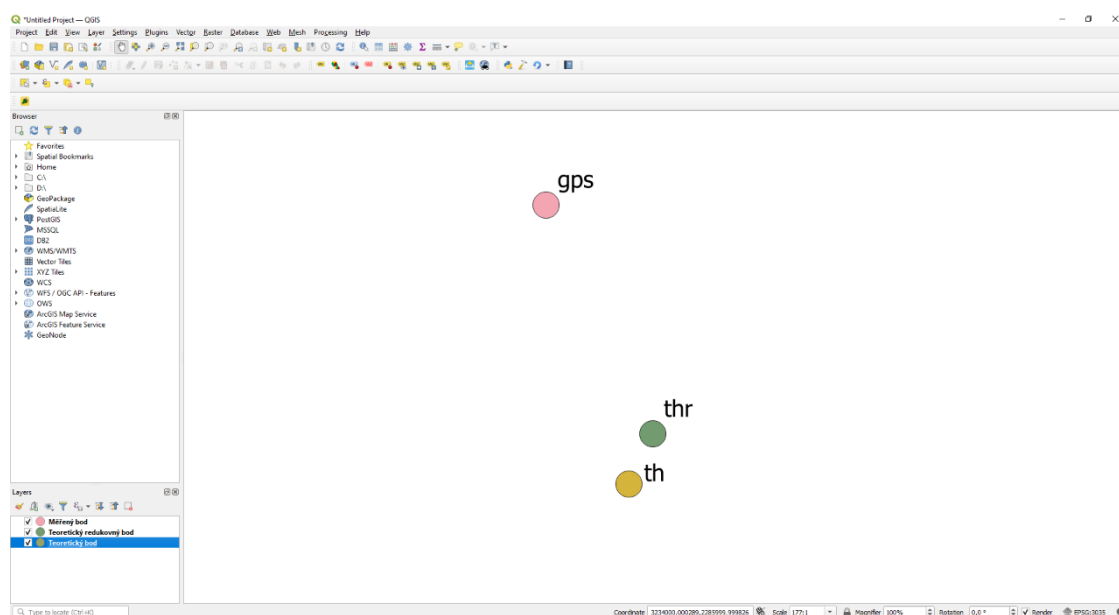
- *survey_year* – Rok, kdy byl bod měřen. Tento atribut má datový typ integer (celé číslo) a nabývá hodnot 2006, 2009, 2012, 2015 případně 2018. Vyskytuje se pouze u časově neagregovaných dat. V případě časoprostorově agregovaných dat by u mnoha záznamů tento atribut musel zaznamenávat více hodnot, a to pokud na daném bodě bylo měřeno opakovaně. Časoprostorová agregace dat je popsána v kapitole 2.8. Atribut je využíván v zásuvném modulu pro výběr bodů na základě roku observace.



- *lc1_h, lc2_h, lu1_h, lu2_h* – Harmonizované *land cover* a *land use* atributy. Tyto nové atributy byly vytvořeny z důvodu zachování původních *land cover* a *land use* tříd v databázi. V případě ostatních úprav byly měněny atributy přímo, jelikož by původní hodnoty neměly pro uživatele žádný význam a pouze by ho mátlly. Ostatní atributy se ve výsledné databázi vyskytují pouze v harmonizované podobě. Pouze u zmíněných atributů *land cover* a *land use* má smysl zachovávat i původní hodnoty. Jedním z důvodů je, že kódy, které byly používány před rokem 2018 a v roce 2018, se již nevyskytovaly (ani v podobě jiného kódu představujícího stejnou vlastnost), byly označeny jako nerelevantní. Stejným způsobem bylo nakládáno s kódy, které nebylo možné jednoznačně přeložit do nomenklatury používané v roce 2018. Hodnoty nových harmonizovaných atributů je tak možné efektivně porovnávat napříč roky. Pokud má uživatel zájem o specifickou vlastnost, zjišťovanou pouze v dřívějších letech (například u atributu *land use* byla hodnota U500 „Wetland“ určována pouze v roce 2006), dostane se k ní pomocí původních atributů *lc1, lc2, lu1, lu2*.
- *geom* – Geometrie polohy teoretického bodu určená ze zaokrouhlených hodnot atributů *th_lat, th_long* a atributu *th_ew*. Atribut *th_lat* obsahuje zeměpisnou šířku teoretického bodu a *th_long* zeměpisnou délku. Zeměpisné souřadnice bylo nutné zaokrouhlit, jelikož se jejich přesnost v různých letech lišila. V případě vytvoření geometrií z původních hodnot by se totožné body měřené v různých letech nenacházely na stejném místě, ale byly by mezi nimi posuny (Obr. 7) v řádu setin milimetrů až desítek centimetrů. Průměrný posun by byl cca 7 cm. Atribut *th_ew* definuje, zda se bod nachází na západní nebo východní polokouli. V případě bodů na západní polokouli bylo nutné před vytvořením geometrií změnit hodnoty zeměpisné délky na záporné.
- *dist_gps_th* – Vypočtená vzdálenost v metrech mezi bodem, na kterém probíhalo měření, a teoretickým bodem. Pomocí tohoto atributu je možné mimo jiné odhadnout, zda při opakované návštěvě daného bodu probíhala observace na totožném místě. V ideálním případě by měření mělo být prováděno přímo na teoretickém bodě, avšak v mnoha případech není na tento bod přístup možný přímo a měření tedy probíhá pouze v jeho blízkosti.

- *dist_th_thr* – Vypočtená vzdálenost v metrech mezi teoretickým bodem a teoretickým bodem, určeným ze zaokrouhlených souřadnic. Atribut popisuje případný posun mezi totožnými body vzniklými na základě nezaokrouhlených souřadnic. Jeho relativně malé, avšak velmi významné hodnoty, mimo jiné pro korektní vytvoření geometrií bodů, byly zmíněny výše při popisu atributu *geom*.

Obrázek (Obr. 7) dokumentuje výše popsané rozdílné polohy jednoho bodu. Ve skutečnosti jsou vzdálenosti mezi body velmi malé. Pro demonstraci bylo nutné nastavit měřítko mapového okna na hodnotu 177:1. Zároveň byl zvolen bod s co nejmenší vzdáleností měřeného bodu od bodu teoretického (*dist_gps_th*). Bylo tak učiněno proto, že ve většině případů je tato vzdálenost mnohonásobně větší než vzdálenost vzniklá redukcí souřadnic teoretického bodu (*dist_th_thr*). Při použití jiného bodu by polohy znázorněné na následujícím obrázku jako „th“ a „thr“ splývaly.



Obr. 7: Poloha LUCAS bodu

2.8 Zveřejnění dat

Po veškerých úpravách, spočívajících v korekci dat tak, aby byla konzistentní napříč roky, opravě chyb, změně datových typů a doplnění nových atributů, byla data zveřejněna na mapovém serveru. Pro zpřístupnění dat byl využit open source mapový server GeoServer. Vznikla tak OGC WFS (Web Feature Service) služba, pomocí které je možné k datům přistupovat.

Data byla publikována ve dvou formách, a to časoprostorově agregované a neagregované. Neagregovaná data jsou taková, u kterých každý záznam reprezentuje jedno měření. Časoprostorová agregace spočívá v propojení tabulek odpovídajících jednotlivým rokům pomocí atributu *point_id*. Tento atribut slouží jako unikátní označení každého bodu. Zároveň je v průběhu let neměnný, což znamená, že v datové sadě odpovídající jednomu roku nemohou být dva body se stejnou hodnotou *point_id*. K takovéto shodě může dojít pouze napříč roky, a to v případě, že se jedná o totožný bod. Každý záznam agregovaných dat představuje bod, u něhož jsou zapsány hodnoty všech měření, která na něm byla v průběhu let provedena. Tato forma umožňuje snadnější porovnávání dat napříč roky a sledování změn.

Dále byly pro obě výše zmíněné formy poskytovaných dat vytvořeny databázové pohledy, obsahující pouze omezené počty atributů. Atributy byly seskupeny do logických celků tak, aby si uživatel mohl vybrat skupinu atributů, o níž má zájem a výsledná atributová tabulka neobsahovala atributy, které by pouze překážely. Tímto způsobem bylo vytvořeno pět skupin (Tab. 11 - Tab. 16), a to „Land cover, Land use“, „Land cover, Land use, Soil“, „Forestry“, „Copernicus“ a „Inspire“. Využití těchto předem připravených skupin má velký význam především pro časoprostorově agregovaná data, neboť celá atributová tabulka obsahuje více než 500 atributů.

2.9 Automatizace

Ve chvíli, kdy vznikne nová databáze LUCAS bodů měřených v roce 2021, bude pravděpodobně obsahovat nekonzistence oproti databázi z roku 2018. Z toho důvodu bude nutné nová data opět harmonizovat. Proto, aby nová harmonizace byla co nejsnazší, byl celý proces automatizován. Automatizace byla užitečná i pro potřeby testování jednotlivých kroků. Bylo tak usnadněno opakované spuštění procesu harmonizace ve chvíli, kdy byla identifikována „nová“, dříve přehlédnutá chyba.

Jednotlivé části propojil Ing. Martin Landa, Ph.D. pomocí softwaru Docker. Docker je open source software vytvořený v roce 2013 firmou Docker, Inc. Umožňuje vytvořit izolované prostředí, ve kterém jsou definovány veškeré požadované parametry pro úspěšné spuštění jednotlivých skriptů. Takovými parametry jsou například verze používaných knihoven. To umožňuje snadnou replikaci veškerých úkonů s použitím téměř jakéhokoliv počítače. Zároveň je pomocí Docker vytvořen dokument spojující dílčí skripty. Tímto způsobem je možné použitím jednoho příkazu provést veškeré úkony, popsané v předchozích kapitolách, od nahrání CSV souborů s původními LUCAS daty do databáze přes úpravu a harmonizaci těchto dat až po zveřejnění pomocí mapového serveru. [9]

3 Použité technologie

3.1 QGIS



Obr. 8: Logo QGIS [10]

QGIS je open source geografický informační systém. Je uveřejněn pod licencí GNU GPL, což mimo jiné zaručuje, že veškerá díla odvozená od tohoto programu musí být publikována pod stejnou licencí, čímž je zajištěn jejich open source charakter. QGIS je možné používat na všech nejrozšířenějších operačních systémech, kterými jsou například Windows, Linux a macOS. Silnou stránkou Programu QGIS jsou zásuvné moduly (pluginy). Plugin je doplněk programu, který rozšiřuje funkčnost základního programu. Zásuvné moduly nevytváří pouze QGIS Development Team, ale také jeho uživatelé.

Tento program začal vytvářet Gary Sherman v roce 2002 pod názvem Quantum GIS. Od roku 2007 je QGIS projektem neziskové nevládní nadace Open Source Geospatial Foundation. Verze 1.0 vyšla na začátku roku 2009. V roce 2013 byl spolu s vydáním verze 2.0 změněn jeho název na současný QGIS. V současnosti je nejnovější verze 3.18.

[11]

3.2 Python

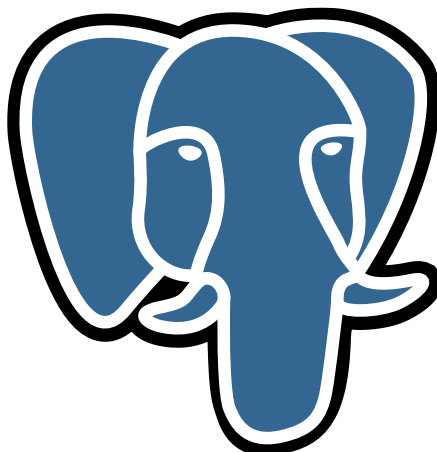


Obr. 9: Logo Python [12]

Python je objektově orientovaný programovací jazyk, který vytvořil v roce 1991 Guido van Rossum. V současné době patří mezi nejpoužívanější programovací jazyky. Je možné jej používat na mnoha operačních systémech (Windows, Linux, macOS, ...). Python je vytvářen jako open source pod licencí Python Software Foundation License. Existují dvě hlavní verze jazyka 2.x a 3.x, avšak vývoj verze 2.x byl ukončen v roce 2020. Od současné verze 3.9 je v plánu vydávat každý rok novou verzi označenou druhou číslicí názvu. Zároveň jsou průběžně vydávány verze opravující chyby, které označuje třetí číslice (např. 3.9.4). Zdrojové kódy jsou vystaveny na stránce <http://www.python.org/>. Velkou výhodou jazyka je množství modulů, které rozšiřují jeho použitelnost.

[13]

3.3 PostgreSQL



Obr. 10: Logo PostgreSQL [14]

PostgreSQL je objektově-relační databázový systém, který využívá a rozšiřuje jazyk SQL. Vývoj tohoto systému začal v roce 1986 na Kalifornské univerzitě. Je vytvářen pro všechny hlavní operační systémy, mezi něž patří Windows, Linux, macOS. Systém je vyvíjen jako open source pod licencí PostgreSQL License. PostgreSQL obsahuje doplňky, mezi které patří i PostGIS, umožňující efektivní práci s prostorovými daty.

[15]

3.4 GeoServer



Obr. 11: Logo GeoServer [16]

GeoServer je open source mapový server vytvořený pomocí programovacího jazyka Java. Je uveřejněn pod licencí GNU GPL. Je vyvíjen a testován velkou skupinou jednotlivců, ale také organizacemi z celého světa. Server umožňuje především publikaci prostorových dat.



GeoServer podporuje standardy vyvíjené mezinárodní organizací OGC (Open Geospatial Consortium). Mezi tyto standardy patří WFS (Web Feature Service) pro práci s vektorovými daty. Pro práci s rastry jsou využívány služby WMS (Web Map Service), WMTS (Web Map Tile Service) a WCS (Web Coverage Service). GeoServer je možné rozšiřovat také o možnost využívání WPS (Web Processing Service). Po publikaci dat na GeoServeru je možné k nim přistupovat pomocí webového rozhraní.

[17]

3.5 OWSLib

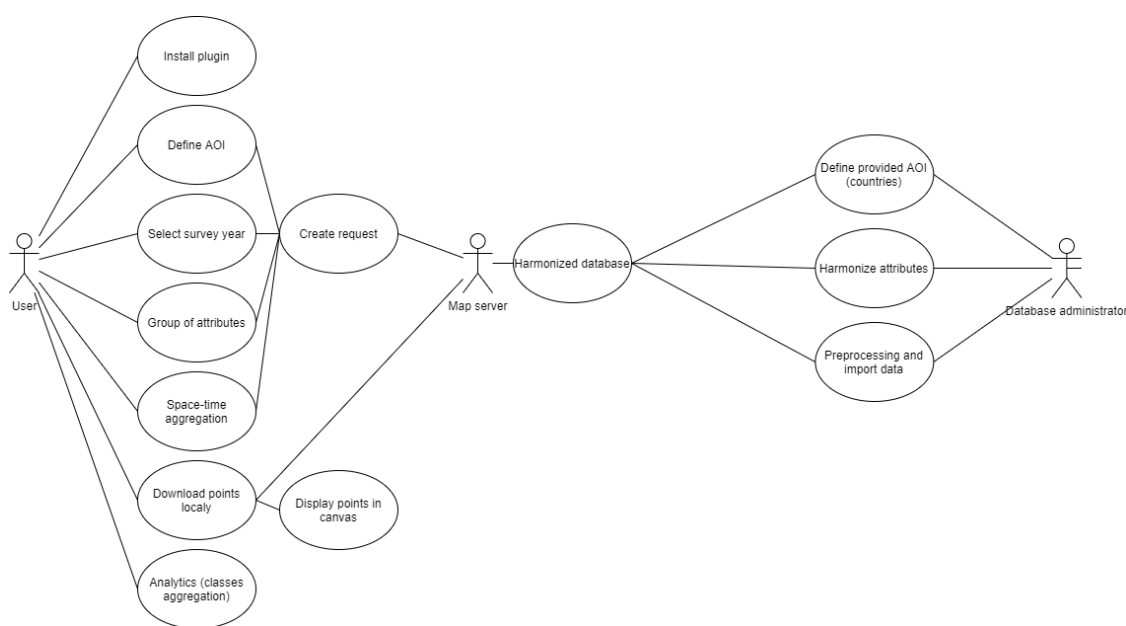
OWSLib je balíček umožňující pomocí jazyka Python pracovat s webovými službami vyvíjenými organizací OGC. Tento balíček byl původně vytvářen jako součást knihovny PCL (Python Cartography Library), avšak v roce 2006 se od knihovny oddělil a byl představen jako samostatný projekt.

Vývoj Python balíčku OWSLib stále pokračuje. Jeho nejnovější verze má označení 0.24.1.

[18]

4 Postup implementace

V této kapitole je popsán průběh implementace. Její realizaci je možné rozdělit do dvou částí, kterými jsou tvorba Python balíčku pyeumap.lucas a vývoj samotného zásuvného modulu pro platformu QGIS, jenž zmíněný balíček využívá.



Obr. 12: Diagram popisující průběh prací

Obrázek (Obr. 12) zobrazuje diagram obsahující tři aktéry, kterými jsou administrátor databáze LUCAS („Database administrator“), mapový server („Map server“) a uživatel zásuvného modulu („User“). Jsou zde znázorněny procesy, které musí zmínění aktéři vykonat tak, aby se požadované body dostaly ve správné formě z databáze k uživateli.

Úkolem administrátora databáze je především úprava dat, která byla detailně popsána v kapitole 2.

Mapový server, v tomto případě GeoServer, zabezpečuje publikaci dat pomocí WFS služby. WFS služba umožňuje snadný přístup pomocí webového rozhraní k vektorovým datům.

Třetím aktérem, znázorněným v diagramu, je samotný uživatel zásuvného modulu. Ten musí nejprve nainstalovat zásuvný modul do programu QGIS a v něm



definovat požadované vlastnosti, podle kterých budou vybrány odpovídající body. Následně jsou data uložena do zvoleného adresáře a uživatel s nimi může dále pracovat v prostředí QGIS, a to pomocí zásuvného modulu, případně pomocí mnoha funkcí, které program QGIS poskytuje.

5 Balíček pyeumap.lucas

Python balíček pyeumap.lucas byl vytvářen v rámci projektu Geo-harmonizer jako jedna ze součástí knihovny pyeumap. Knihovna pyeumap slouží primárně pro snazší zpřístupnění prostorových dat, pokrývajících území států Evropské unie, která se týkají životního prostředí, a implementaci funkcí umožňujících práci s těmito daty.

Balíček pyeumap.lucas umožňuje uživateli přistupovat k datům LUCAS a následně je zpracovávat. Funkcionalitu je možné rozdělit na tři hlavní části. První z nich zabezpečuje sestavení dotazu určeného pro GeoServer na základě požadavků uživatele. Druhá část požadavek zpracuje a odešle na GeoServer. Obdržená data jsou následně uložena na zvolené místo na disku. Třetí část umožňuje vlastní zpracování získaných dat, což představuje možnost agregace *land cover* tříd na základě předem definovaných pravidel.

Balíček pyeumap.lucas je navržen objektově. Jeho jednotlivé třídy jsou popsány v následujících kapitolách.

5.1 Sestavení dotazu

Sestavení dotazu zabezpečuje třída LucasReques. Proměnné této třídy definují parametry výsledného dotazu. Proměnnými, kterými uživatel dotaz ovlivňuje jsou:

- `bbox` – Určuje oblast, ve které se mají body nacházet, pomocí obdélníku definovaného souřadnicemi levého dolního rohu a pravého horního rohu.
- `propertyname`, `literal`, `operator`, `logical` – Jsou čtyři parametry, které umožňují výběr na základě atributového dotazu. Zadávají se pomocí nich následující vlastnosti:
 - `propertyname` – Název atributu, na základě kterého má být atributový dotaz vykonán.
 - `literal` – Hodnota zvoleného atributu.
 - `operator` – Funkce, která má být použita pro vykonání atributového dotazu, např.: `PropertyIsEqualTo` (vybere prvky, u kterých daný atribut



nabývá zvolené hodnoty), `PropertyIsEqualTo` (vybere prvky, u kterých daný atribut nenabývá zvolené hodnoty), `PropertyIsGreaterThan` (vybere prvky, u kterých daný atribut nabývá vyšší hodnoty, než je hodnota zvolená) atd. Tyto možnosti dává balíček OWSLib.

- `logical` – Logický operátor spojující atributové dotazy. Určuje, zda mají podmínky platit zároveň (*And*), případně zda postačí platnost jedné z podmínek (*Or*).
- `aoi_polygon` – Pomocí textového řetězce ve formátu *GML*, obsahujícího lomové body polygonu, určuje území, ve kterém se mají body nacházet.
- `years` – Je parametr, prostřednictvím kterého je možný výběr na základě let, během nichž byly body měřeny. Zadává se pomocí seznamu, jež může obsahovat některou z hodnot 2006, 2009, 2012, 2015, 2018, případně jejich libovolnou kombinaci.
- `st_aggregated` – Tento parametr může nabývat hodnot *True* (pravda), nebo *False* (nepravda) podle toho, zda mají být data časoprostorově agregována či nikoliv.
- `group` – Parametr `group` slouží k rozhodnutí, jaké atributy (Tab. 11 - Tab. 16) mají data obsahovat.

Nejdůležitější metodou třídy `LucasRequest`, která vytváří samotný dotaz, je metoda `build`. Tato metoda nejprve volá metodu `_check_args`, která zabezpečuje kontrolu toho, zda jsou vstupní argumenty zadány správně. Následně je dotaz postupně vytvářen pomocí metod OWSLib balíčku. Jednotlivé části dotazu jsou generovány na základě toho, jaké argumenty byly zadány. Pro spojení dílčích částí jsou využívány metody OWSLib balíčku *And* a *Or*. Na závěr je výsledný objekt převeden do formátu *XML*, který je spolu s názvem požadované vrstvy návratovou hodnotou metody `build`.

V případě dotazu vytvořeného kombinací atributového dotazu a minimálního ohraničujícího obdélníku (`bbox`) není výsledek správný. Tento problém byl opraven ve výsledném *XML* nahrazením chybných částí pomocí metody `replace`.

```

<ogc:And xmlns:ogc="http://www.opengis.net/ogc">
  <ogc:Or>
    <ogc:PropertyIsEqualTo>
      <ogc:PropertyName>survey_year</ogc:PropertyName>
      <ogc:Literal>2012</ogc:Literal>
    </ogc:PropertyIsEqualTo>
    <ogc:PropertyIsEqualTo>
      <ogc:PropertyName>survey_year</ogc:PropertyName>
      <ogc:Literal>2018</ogc:Literal>
    </ogc:PropertyIsEqualTo>
  </ogc:Or>
  <ogc:BBOX>
    <ogc:PropertyName>geom</ogc:PropertyName>
    <gml311:Envelope xmlns:gml311="http://www.opengis.net/gml" srsName="http://www.opengis.net/gml/srs/epsg.xml#3035">
      <gml311:lowerCorner>4624127 2998330</gml311:lowerCorner>
      <gml311:upperCorner>4650393 3013986</gml311:upperCorner>
    </gml311:Envelope>
  </ogc:BBOX>
</ogc:And>

```

Obr. 13: Kombinace prostorového a atributového dotazu (XML)

Obrázek (Obr. 13) zobrazuje výsledné XML vygenerované metodou build v případě, že mají být vybrány body, které byly měřeny v roce 2012 nebo v roce 2018 a tyto body se nacházejí uvnitř obdélníku definovaného souřadnicemi levého dolního rohu (4624127, 2998330) a pravého horního rohu (4650393, 3013986).

5.2 Vytvoření výstupů

O vytvoření výstupů se stará třída LucasIO. Nejpodstatnější metodou, kterou zmíněná třída obsahuje, je *download*. Metoda *download* zabezpečuje stažení dat na základě parametrů definovaných v dotazu vytvořeném třídou LucasRequest (viz kapitola 5.1). Pro stažení dat je stejně jako v případě vytvoření dotazu využíván balíček OWSLib. Požadavek je odeslán na GeoServer. WFS služba vrátí data, která jsou převedena na textový řetězec ve formátu GML. Následně je volána metoda *_load*, jejímž cílem je převést stažená data do formátu GPKG a uložit do přechodného adresáře.

Jelikož při stahování dat vzniká atribut nazvaný *gml_id*, který není relevantní a pouze znepřehledňuje atributovou tabulku, je pomocí metody *_postprocessing* z GPKG souboru odstraněn. Další funkcí této metody je odstranění přebytečných atributů v případě časoprostorově agregovaných dat. Časoprostorově agregovaná data obsahují každý atribut tolikrát, kolikrát bylo měření LUCAS prováděno (aktuálně pětkrát). V případě, že má uživatel zájem o časoprostorově agregovaná data měřená pouze v určitých letech, není žádoucí, aby obdržel i atributy popisující měření z ostatních let. Atributy odpovídající rokům, z nichž měření nebyla požadována, metoda *_postprocessing* odstraňuje. Tímto způsobem se znatelně zmenší velikost atributové tabulky a zpřehlední se.



Metoda `to_gml` převádí *GPKG* soubor uložený v přechodném adresáři do formátu *GML* a navrácí ho v podobě textového řetězce.

O uložení *GPKG* souboru na zvolené místo na disku se stará metoda `to_gpkg`. Vstupním parametrem této metody je cesta ke *GPKG* souboru, který se má vytvořit.

Úkolem další metody `to_geopandas` je pouze návrat dat ve formě datové struktury *GeoDataFrame* z knihovny *GeoPandas*. Tento formát usnadňuje práci s atributovou tabulkou pomocí jazyka *Python*.

Metody `num_of_features` a `is_empty` spolu souvisí. První z nich určuje počet prvků, které stažený soubor obsahuje, tento počet je návratovou hodnotou metody. Druhá metoda (`is_empty`) na základě zjištěného počtu prvků rozlišuje, zda soubor obsahuje alespoň jeden prvek. Pokud *GPKG* soubor není prázdný, vrací metoda hodnotu *False*. V opačném případě je metodou navraceno *True*.

5.3 Agregace land cover tříd

Třída `LucasClassAggr` slouží k agregaci *land cover* tříd. Metoda `_load_classes` na základě vstupního *JSON* souboru (Obr. 17) vytváří slovník obsahující agregační pravidla. Před tím, než je slovník vytvořen, probíhá kontrola formy i obsahu *JSON* souboru. Z *CSV* souboru jsou načteny veškeré validní *land cover* kódy, pomocí kterých je zjišťováno, zda agregační pravidla neobsahují kódy nevalidní. Dále metoda určuje, zda se kódy neopakují, což by vedlo k nejednoznačnosti agregačních pravidel.

O samotnou agregaci se stará metoda `apply`. Tato metoda prostřednictvím jazyka *Python* volá *SQL* dotazy pracující s atributovou tabulkou. *SQL* dotazy jsou využívány pro vytvoření indexu nad sloupečkem `lc1_h` (hlavní *land cover* třída harmonizovaná), pro přidání nového atributu `lc1_a` a zapsání agregovaných tříd na základě agregačních pravidel do tohoto atributu. Před spuštěním *SQL* dotazů je na základě názvu vrstvy rozhodováno, zda se jedná o časoprostorově agregovaná data či nikoliv. V případě časoprostorově agregovaných dat je index vytvořen nad všemi `lc1_h` sloupečky (za každý rok jeden) a taktéž je utvořen odpovídající počet nových atributů pro agregované třídy.

5.4 Jupyter Notebook

V rámci projektu Geo-harmonizer vznikl soubor, umožňující přehlednou demonstraci funkcionality pyeumap.lucas balíčku. Tvorbou tohoto souboru se zabýval především Ing. Martin Landa, Ph.D. Pomocí Jupyter Notebook je možné nejen psát zdrojový kód a spouštět ho, ale také vytvářet uspořádanou strukturu s komentáři jednotlivých částí. Výsledný dokument obsahuje ukázky mnoha možností, které dává uživateli balíček pyeumap.lucas, popsány v předchozích kapitolách 5.1, 5.2 a 5.3.

Na obrázku (Obr. 14) je znázorněna část Jupyter Notebooku. Tato část obsahuje vytvořený požadavek na základě zadaných souřadnic definujících ohraničující obdélník (bbox). Dále je zde ukázka předání požadavku objektu třídy LucasIO, který se stará o stažení dat. Zároveň se stažením dat je uživatel informován o počtu obdržených prvků.

Usage

1. Define request

Request is defined by `LucasRequest` object. Bbox filter can be enabled by `bbox` property. Currently only [EPSG:3035](#) is supported.

For testing purpose a request can be created by `build()` method.

```
In [2]: request = LucasRequest()
request.bbox = (4504276, 3020369, 4689608, 3105290)
request.build()

Out[2]: {'filter': '<ogc:BBOX xmlns:ogc="http://www.opengis.net/ogc"><ogc:PropertyName>geom</ogc:PropertyName><gml311:Envelope xmlns:gml311="http://www.opengis.net/gml" srsName="http://www.opengis.net/gml/srs/epsg.xml#3035"><gml311:lowerCorner>4504276 3020369</gml311:lowerCorner><gml311:upperCorner>4689608 3105290</gml311:upperCorner></gml311:Envelope></ogc:BBOX>'}
```

2. Download data based on request

LUCAS data is controlled by `LucasIO` (input/output) class. Data is downloaded by calling `download()` method based on prepared request. Number of downloaded features can be retrieved by `num_of_features()` method.

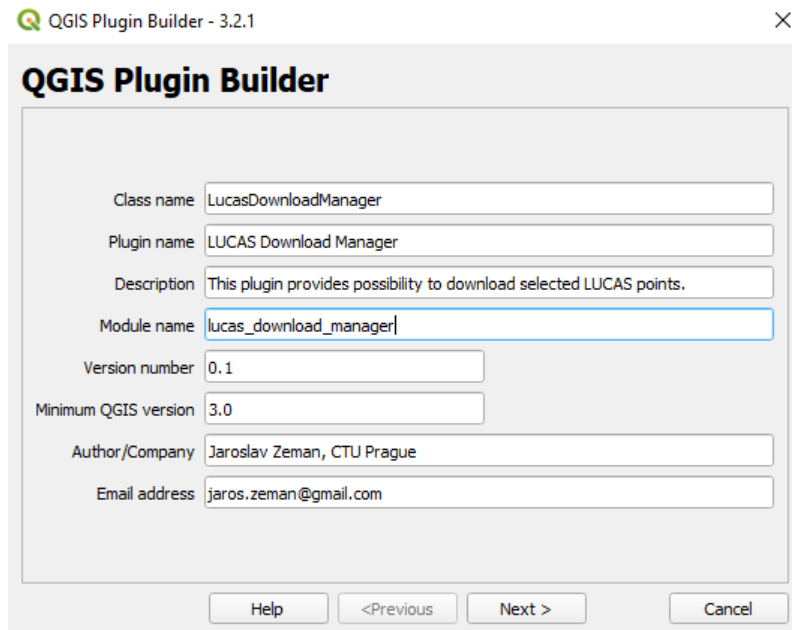
```
In [3]: lucasio = LucasIO()
lucasio.download(request)

print("Number of downloaded points:", lucasio.num_of_features())
```

Obr. 14: Ukázka části Jupyter Notebooku

6 Zásuvný modul

Kostra zásuvného modulu byla vytvořena pomocí zásuvného modulu QGIS Plugin Builder, dostupného v oficiálním QGIS plugin repozitáři. Do QGIS Plugin Builder bylo nutné nejprve zadat základní informace o vytvářeném zásuvném modulu. Mezi takové informace patří například název modulu, jeho umístění, stručný popis funkčnosti a další. QGIS Plugin Builder je zobrazen na následujícím obrázku (Obr. 15).



QGIS Plugin Builder - 3.2.1

QGIS Plugin Builder

Class name: LucasDownloadManager

Plugin name: LUCAS Download Manager

Description: This plugin provides possibility to download selected LUCAS points.

Module name: lucas_download_manager

Version number: 0.1

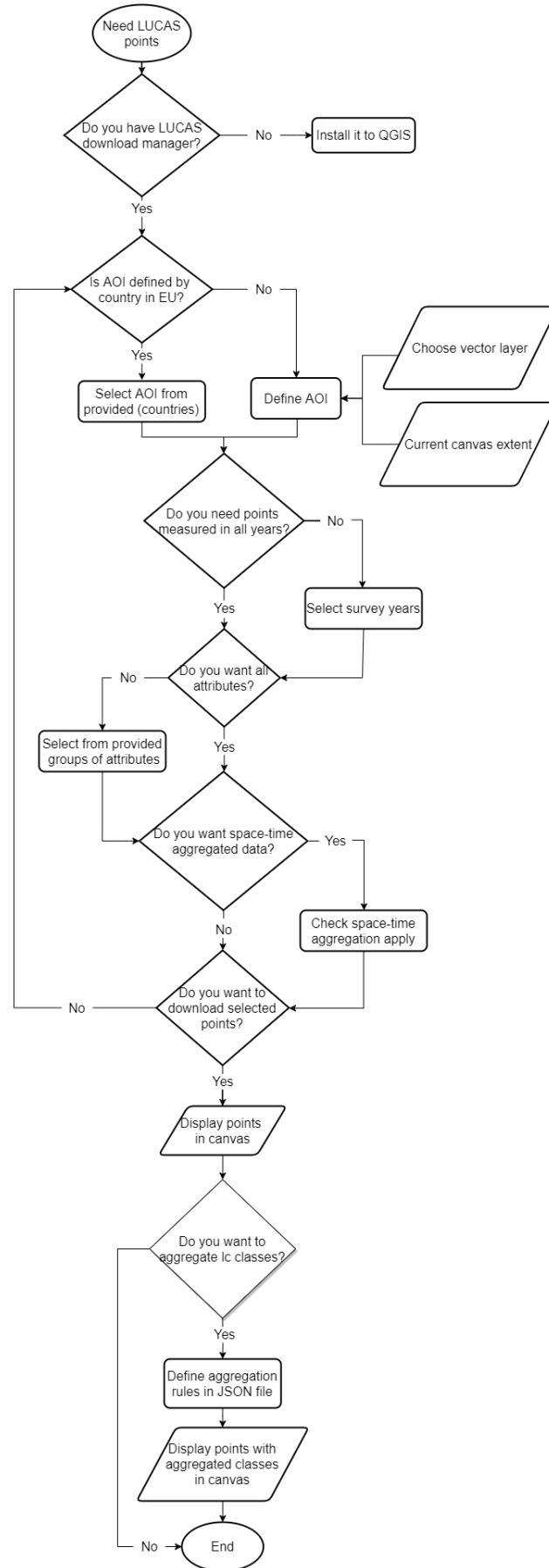
Minimum QGIS version: 3.0

Author/Company: Jaroslav Zeman, CTU Prague

Email address: jaros.zeman@gmail.com

Buttons: Help, <Previous, Next >, Cancel

Obr. 15: QGIS Plugin Builder



Obr. 16: Diagram popisující funkcionální zástupný modul

6.1 Funkcionalita

Funkcionalita zásuvného modulu z pohledu uživatele je podrobně popsána diagramem na obrázku (Obr. 16). Jsou zde znázorněny veškeré úkony, které je nutné vykonat proto, aby zásuvný modul korektně stáhnul a načetl požadovaná data do mapového okna.

Zásuvný modul (Obr. 18) je rozdělen do dvou částí, z nichž každá plní jinou funkci. První část, nazvaná „Download“, zajišťuje stažení požadovaných dat. Pomocí druhé části „Class aggregation“ je možné získaná data upravovat.

6.1.1 Stažení dat

Část „Download“ zásuvného modulu umožňuje uživateli definovat parametry požadovaných dat. Nejprve je nutné zvolit zájmové území („Area of interest“), na výběr jsou následující možnosti:

- **canvas** – Pokud uživatel zvolí tuto možnost, dochází ke stažení bodů nacházejících se na území, které aktuálně zobrazuje mapové okno. Podmínkou pro tuto eventualitu je, že mapové okno musí být v souřadnicovém systému označeném EPSG kódem 3035 (LAEA Europe).
- **country** – Po označení této možnosti je nutné v rozbalovacím seznamu zvolit země, na jejichž území se mají body nacházet. V tomto případě je využíván atribut *nuts0*, na jehož základě jsou odpovídající body vybírány.
- **vector layer** – Tato eventualita umožňuje vybrat jednu z aktivních vektorových vrstev a získat body uvnitř území pokrytého touto vrstvou. Restrikcí pro tento případ je možnost využití pouze polygonové vrstvy s omezeným počtem vrcholů. Důvodem, proč není možné použít polygon s libovolným počtem vrcholů je skutečnost, že GeoServer má nastavenou maximální délku požadavku. Požadavky odeslané na server, které tuto délku překročí, nejsou zpracovány a místo získání požadovaných dat je uživatel informován o chybě.

Druhým parametrem, pomocí kterého uživatel definuje vlastnosti bodů, o něž má zájem, je rok, kdy byl bod měřen. Do dnešního dne bylo provedeno pět měření LUCAS bodů, a to v letech 2006, 2009, 2012, 2015 a 2018. Uživatel má možnost pomocí

zaškrťovacích polí označit roky, ve kterých byly body měřeny. V případě, že není označen rok žádný, je s daty pracováno tak, jako by byly vybrány všechny roky.

V průběhu pěti měření, která byla doposud provedena, postupně narůstal počet charakteristik, popisujících jednotlivé body, a tím pádem i počet atributů. V případě, že by si uživatel stáhl body s veškerými atributy, obsahovala by atributová tabulka mnoho desítek sloupců a byla by značně nepřehledná. Ve většině případů zajímá uživatele pouze malé množství vlastností a ostatní tudíž překážejí. Z toho důvodu byly atributy rozčleněny do pěti skupin. Tyto skupiny jsou zobrazeny v následující tabulce spolu s počtem atributů, které obsahují.

Tab. 10: Skupiny atributů (Tab. 11- Tab. 16)

Skupina atributů	Počet atributů
Land cover, Land use	40
Land cover, Land use, Soil	55
Forestry	2
Copernicus	16
Inspire	8

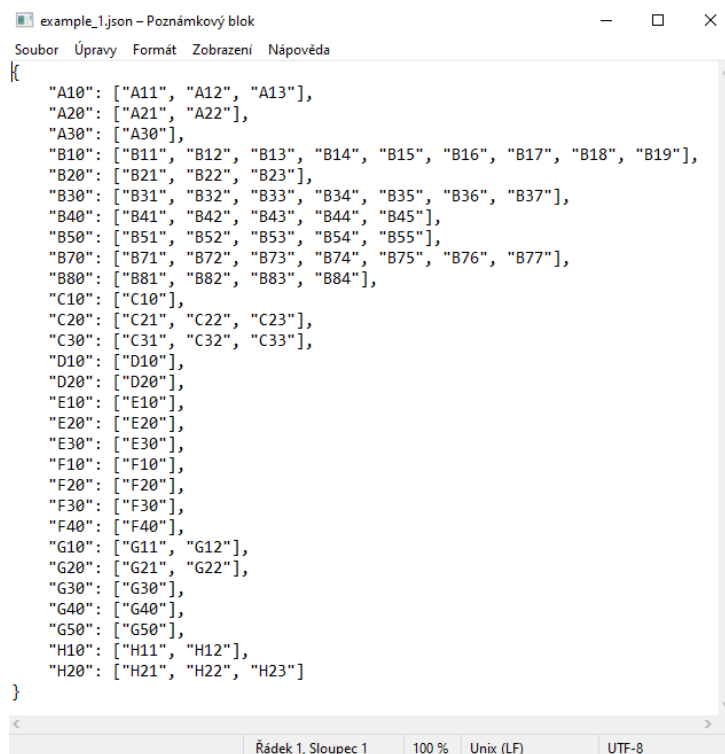
Je možné si zvolit jednu z výše uvedených skupin a výsledná atributová tabulka bude obsahovat krom povinných atributů, definujících především polohu bodu, pouze atributy z této skupiny.

Jelikož je velké množství bodů měřeno opakovaně, tedy v různých letech, je možné tato měření porovnávat a zjistit tak, jak se dané území vyvíjí v čase. V tom případě je vhodné, aby každý prvek v atributové tabulce odpovídal jednomu bodu a jednotlivé atributy popisovaly veškerá měření provedená na tomto bodě. Naopak, pokud uživatele zajímá stav krajiny pouze v určité době, je výhodnější, pokud každý prvek představuje jedno měření. Atributová tabulka pak obsahuje méně nežádoucích atributů, a je tak přehlednější. Parametrem, který tímto způsobem definuje charakter získaných dat, je časoprostorová agregace („Space-time aggregation“).

Následně je nutné definovat cestu, kam má být soubor se staženými daty uložen a poté pomocí tlačítka „Download“ spustit stahování. Po úspěšném stažení dat se body načtou do mapového okna a zásuvný modul zároveň informuje uživatele o jejich počtu.

6.1.2 Agregace tříd

Druhá část zásuvného modulu nazvaná „Class aggregation“ umožňuje uživateli agregovat harmonizované *land cover* třídy. Pravidla, na základě kterých mají být třídy agregovány, je nutné zapsat do *JSON* souboru. Tento soubor musí obsahovat tak zvané klíče, které reprezentují názvy nových tříd, a seznamy *land cover* tříd třetí úroveň, odpovídající jednotlivým klíčům. Na obrázku (Obr. 17) je příklad *JSON* souboru, který popisuje převod *land cover* tříd třetí úrovně na *land cover* třídy druhé úrovně.



```

{
  "A10": ["A11", "A12", "A13"],
  "A20": ["A21", "A22"],
  "A30": ["A30"],
  "B10": ["B11", "B12", "B13", "B14", "B15", "B16", "B17", "B18", "B19"],
  "B20": ["B21", "B22", "B23"],
  "B30": ["B31", "B32", "B33", "B34", "B35", "B36", "B37"],
  "B40": ["B41", "B42", "B43", "B44", "B45"],
  "B50": ["B51", "B52", "B53", "B54", "B55"],
  "B70": ["B71", "B72", "B73", "B74", "B75", "B76", "B77"],
  "B80": ["B81", "B82", "B83", "B84"],
  "C10": ["C10"],
  "C20": ["C21", "C22", "C23"],
  "C30": ["C31", "C32", "C33"],
  "D10": ["D10"],
  "D20": ["D20"],
  "E10": ["E10"],
  "E20": ["E20"],
  "E30": ["E30"],
  "F10": ["F10"],
  "F20": ["F20"],
  "F30": ["F30"],
  "F40": ["F40"],
  "G10": ["G11", "G12"],
  "G20": ["G21", "G22"],
  "G30": ["G30"],
  "G40": ["G40"],
  "G50": ["G50"],
  "H10": ["H11", "H12"],
  "H20": ["H21", "H22", "H23"]
}

```

Obr. 17: Soubor *JSON* definující agregaci tříd

Po vytvoření souboru s agregačními pravidly je nutné tento soubor načíst pomocí zásuvného modulu. Následně se jednotlivé třídy v zásuvném modulu zobrazí a po označení zvolené třídy jsou vypsány všechny „původní“ třídy, které do zvolené budou agregovány. Forma zobrazení je zde velmi přehledná a umožňuje snadnou kontrolu předem zformulovaných pravidel. Agregační pravidla se aplikují na *GPKG* soubor, jehož umístění je definováno v předchozí části „Download“. V daném souboru se vytvoří nové atributy, jejichž počet odpovídá počtu atributů *lc1_h* (harmonizované



hlavní *land cover* třídy). To znamená, že v případě časoprostorově agregovaných dat závisí počet nových atributů na počtu let, která uživatel označil v části „Download“. Například pokud byly označeny roky 2012, 2015 a 2018, obsahují získaná data mimo jiné atributy *lc1_h_2012*, *lc1_h_2015* a *lc1_h_2018* a při agregaci vzniknou nové atributy *lc1_a_2012*, *lc1_a_2015* a *lc1_a_2018*. V případě neagregovaných dat, u kterých každý záznam v atributové tabulce představuje unikátní měření, se vytvoří pouze jeden nový atribut *lc1_a*. Do nových sloupců atributové tabulky se zapíšou nově definované hodnoty odpovídající klíčům v *JSON* souboru. Pokud některý z *land cover* kódů nebyl pro agregaci použit (nebyl zapsán v souboru *JSON*), objeví se místo něj v atributu pro agregované třídy hodnota *NULL*.

Závěr

Cílem této diplomové práce byla harmonizace a validace datové sady LUCAS a její následné zpřístupnění uživatelům pomocí QGIS zásuvného modulu. LUCAS měření provedená v letech 2006, 2009, 2012, 2015 a 2018 měla být sjednocena s využitím dat z roku 2018 jako referenčních. Po úpravě dat mělo dojít k jejich publikaci na mapovém serveru pomocí WFS služby. Dále mělo dojít pomocí jazyka Python k vytvoření balíčku `pyeumap.lucas`, sloužícího především k vygenerování dotazu pro WFS službu a ke stažení dat. Zmíněný balíček měl být využíván zásuvným modulem do programu QGIS, vytvářející uživatelsky přívětivé prostředí pro přístup k bodům LUCAS. Zásuvný modul měl poskytovat analýzu stažených dat.

Vytyčené cíle se podařilo splnit. V podobě zásuvného modulu byl vytvořen nástroj, který významně usnadní běžným uživatelům přístup k datové sadě LUCAS, jenž je zdrojem mnoha hodnotných informací. V tuto chvíli poskytuje zásuvný modul uživateli pouze informaci o počtu stažených bodů, v budoucnu by měla být analýza dat podrobnější.

Pro harmonizaci databáze LUCAS bylo nutné provést velké množství úkonů. Bylo třeba se vypořádat nejen s odlišnými názvy atributů, různými kódy představujícími jednu skutečnost, ale i např. změnou způsobu zaznamenávání hodnot jednoho atributu v průběhu let. Významnou komplikací byla i přítomnost mnoha nevalidních hodnot u různých atributů. Z důvodu, aby úpravami dat nevznikly nové chyby a nepřesnosti, byla využívána nejen oficiální dokumentace LUCAS databáze, ale také článek *Harmonised LUCAS in-situ land cover and use database for field surveys from 2006 to 2018 in the European Union* [4] a jeho výstupy, které sloužily jako vzor. V mnoha případech byly využity obdobné postupy jako ty, popsané ve zmiňovaném článku, avšak lze identifikovat i zásadní rozdíly, charakterizované v kapitole 2. Významné odlišnosti jsou v přístupu k harmonizaci *land cover* a *land use* atributů, ve způsobu práce s nevalidními hodnotami, a především ve formě záznamu hodnot, které jednotlivé atributy nabývají. Po harmonizaci následovala změna datových typů atributů a tvorba atributů nových. Nově vytvořené atributy databázi obohacují a usnadňují práci s ní.

Pomocí softwaru Docker byl celý proces od nahrání CSV souborů, obsahujících originální LUCAS data, do databázového softwaru až po publikaci harmonizovaných

dat ve formě WFS služby na mapovém serveru GeoServer automatizován. Automatizace umožňuje snadné spouštění v izolovaném prostředí a dává tedy komukoliv možnost jednoduchým způsobem celý postup replikovat.

Balíček `pyeumap.lucas` byl vytvořen primárně pro získání požadovaných dat z GeoServeru a jejich následné zpracování. Pro sestavení požadavku na základě vstupních parametrů a jeho předání mapovému serveru je využíván balíček `OWSLib`, který umožňuje pomocí jazyka Python pracovat s webovými službami podporovanými OGC. Pomocí minimálního ohraničujícího obdélníku nebo textového řetězce ve formátu *GML*, obsahujícího lomové body polygonu, je možné určit, kde se mají uživatelem požadované body nacházet. Další možností je volba bodů na základě atributového dotazu. Speciálním případem atributového dotazu je výběr na základě roku měření. Dále je možné rozhodnout, zda mají být data časoprostorově agregovaná a jaké atributy mají výsledná data obsahovat. Problémy s implementací byly především v případě výběru na základě prostorového dotazu. Bylo obtížné předat dotazu polygon definovaný lomovými body, jelikož tuto možnost `OWSLib` balíček neumožňuje. Další komplikaci způsobila potřeba kombinace prostorového a atributového dotazu. Nepříjemností byla i skutečnost, že v době tvorby diplomové práce nejnovější verze `OWSLib` balíčku 0.23.0 nefungovala korektně. V průběhu stahování bylo třeba se vypořádat s atributem `gml_id`, který se vytváří automaticky při získání dat z mapového serveru. Balíček `pyeumap.lucas` umožňuje také zpracování stažených dat, a to prozatím pouze v podobě agregace hlavních *land cover* tříd. Pro demonstraci funkcionality `pyeumap.lucas` balíčku byl vytvořen Jupyter Notebook.

Poslední částí byla implementace samotného zásuvného modulu pro program QGIS. Zásuvný modul dává uživateli přehledné a intuitivní prostředí, pomocí kterého je možné přistupovat k harmonizované datové sadě LUCAS. Zásuvný modul využívá balíček `pyeumap.lucas` a přidává další možnosti, jak s daty pracovat. Tento modul je rozdělen do dvou částí. První z nich slouží pro definování požadovaných vlastností dat. Druhá část umožňuje agregaci hlavních *land cover* tříd. Pro ni je nutné vytvořit *JSON* soubor a v něm definovat agregační pravidla.

Veškeré dílčí části byly v průběhu práce mnohokrát testovány. V rámci dat byly porovnávány počty bodů původní datové sady s výslednou databází, kontrolována správnost aplikace jednotlivých harmonizačních kroků a korektnost časoprostorové

agregace. Bylo zjišťováno, zda byla data zveřejněna v požadované formě. Následně byly testovány i veškeré možné způsoby, jakými by uživatel mohl s balíčkem `pyeumap.lucas` a zásuvným modulem pracovat. Ověřena byla funkčnost všech možných kombinací požadavků, jsou-li na jejich základě uživateli navrácena odpovídající data. Tyto testy byly prováděny pouze manuálně a v budoucnu je bude nutné automatizovat.

Jelikož tato diplomová práce vznikla v rámci projektu Geo-harmonizer, budou se její výstupy nadále rozvíjet. Výsledná databáze, jejíž součástí je velké množství úprav, již byla mnohokrát kontrolována a nemělo by být třeba ji dále měnit. Změny a především doplnění jsou však možná v rámci `pyeumap.lucas` balíčku a také zásuvného modulu. Do balíčku `pyeumap.lucas` by bylo vhodné doplnit možnost výběru na základě více než jednoho polygonu, případně liniové a bodové vrstvy. Taktéž část, zabývající se zpracováním stažených dat, má určité rezervy. Balíček by mohl uživateli umožňovat více způsobů úprav než jen agregaci hlavních *land cover* tříd. Prostor pro zdokonalení je také na straně zásuvného modulu. Zásuvný modul by mohl zprostředkovávat překlad do dalších nomenklatur, jako například CORINE. V části zabývající se agregací *land cover* tříd by mohl být vytvořen nástroj pro interaktivní tvoření agregačních pravidel bez nutnosti jejich definování mimo QGIS v *JSON* souboru.

Po implementaci dalších úprav bude QGIS plugin zařazen do oficiálního QGIS repozitáře. Zároveň pro něj bude vytvořena i dokumentace, jejíž základem bude obsah dodatku této práce (QGIS plugin documentation). Jelikož zásuvný modul bude užitečný nejen pro uživatele z České republiky, ale z celé Evropy, potažmo celého světa, bude dokumentace vytvořena v anglickém jazyce, a tedy i dodatek této práce je psaný anglicky.

Použité zdroje

- [1] About Geo-harmonizer. Open Data Science Europe [online]. [cit. 2021-03-28]. Dostupné z: <https://opendatascience.eu/geoharmonizer-project/>
- [2] Geo-harmonizer: EU-wide automated mapping system for harmonization of Open Data based on FOSS4G and Machine Learning. INNOVATION AND NETWORKS EXECUTIVE AGENCY [online]. [cit. 2021-5-16]. Dostupné z: <https://ec.europa.eu/inea/en/connecting-europe-facility/cef-telecom/2018-eu-ia-0095>
- [3] LUCAS - Land use and land cover survey. EUROSTAT Statistics Explained [online]. [cit. 2021-03-28]. Dostupné z: https://ec.europa.eu/eurostat/statistics-explained/index.php/LUCAS_-_Land_use_and_land_cover_survey
- [4] D'ANDRIMONT, R., YORDANOV, M., MARTINEZ-SANCHEZ, L. et al. Harmonised LUCAS in-situ land cover and use database for field surveys from 2006 to 2018 in the European Union. Scientific Data [online]. 16.10. 2020, 7(352 (2020) [cit. 2021-03-29]. Dostupné z: [doi:https://doi.org/10.1038/s41597-020-00675-z](https://doi.org/10.1038/s41597-020-00675-z)
- [5] Land Use / Cover Area Frame Survey: Technical reference document C2 [online]. 2018 [cit. 2021-03-29]. Dostupné z: <https://ec.europa.eu/eurostat/documents/205002/8072634/LUCAS2018-C2-FieldForm-GD-Template.pdf>
- [6] Land Use / Cover Area Frame Survey: Technical reference document C3 [online]. 2018 [cit. 2021-03-29]. Dostupné z: <https://ec.europa.eu/eurostat/documents/205002/8072634/LUCAS2018-C3-Classification.pdf>
- [7] Krajinné pokrytí. ČÚZK: Geoportál [online]. [cit. 2021-5-16]. Dostupné z: [https://geoportal.cuzk.cz/\(S\(sgmfbk3yfcqbsovpwd1aet1x\)\)/Default.aspx?mode=TextMeta&text=INSPIRE_K_Pokryv&side=INSPIRE_dSady&head_tab=sekcce-04-gp&menu=421](https://geoportal.cuzk.cz/(S(sgmfbk3yfcqbsovpwd1aet1x))/Default.aspx?mode=TextMeta&text=INSPIRE_K_Pokryv&side=INSPIRE_dSady&head_tab=sekcce-04-gp&menu=421)
- [8] Land Use / Cover Area Frame Survey: Technical reference document C1 Instructions for Surveyors [online]. [cit. 2021-5-16]. Dostupné z:



- <https://ec.europa.eu/eurostat/documents/205002/8072634/LUCAS2018-C1-Instructions.pdf>
- [9] Docker overview. Docker Documentation [online]. [cit. 2021-5-16]. Dostupné z: <https://docs.docker.com/get-started/overview/>
- [10] Visual Style Guide. QGIS [online]. [cit. 2021-5-1]. Dostupné z: <https://www.qgis.org/en/site/getinvolved/styleguide.html>
- [11] Obsah - Školení QGIS pro začátečníky [online]. [cit. 2021-5-1]. Dostupné z: <https://training.gismentors.eu/qgis-zacatecnik/index.html>
- [12] The Python Logo. Python Software Foundation [online]. [cit. 2021-5-1]. Dostupné z: <https://www.python.org/community/logos/>
- [13] General Python FAQ: Python 3.9.4 documentation. Python - official website [online]. [cit. 2021-5-2]. Dostupné z: <https://docs.python.org/3/faq/general.html>
- [14] PostgreSQL - Wikipedie [online]. [cit. 2021-5-2]. Dostupné z: <https://cs.wikipedia.org/wiki/PostgreSQL>
- [15] PostgreSQL: About. PostgreSQL [online]. [cit. 2021-5-2]. Dostupné z: <https://www.postgresql.org/about/>
- [16] GitHub - geoserver: Official GeoServer repository [online]. [cit. 2021-5-2]. Dostupné z: <https://github.com/geoserver/geoserver>
- [17] About - GeoServer. GeoServer [online]. [cit. 2021-5-2]. Dostupné z: <http://geoserver.org/about/>
- [18] OWSLib 0.24.1 documentation: Introduction. OWSLib documentation [online]. [cit. 2021-5-16]. Dostupné z: <https://geopython.github.io/OWSLib/introduction.html>
- [19] MENKE, Kurt et al. Mastering QGIS. Birmingham: Packt Publishing, 2015. ISBN 9781784390068.
- [20] PILGRIM, Mark. Dive Into Python 3. Apress, 2015. ISBN 9781430224150.
- [21] OWSLib 0.24.1 documentation: Usege. OWSLib documentation [online]. [cit. 2021-5-16]. Dostupné z: <https://geopython.github.io/OWSLib/usage.html#wfs>

Seznam obrázků

Obr. 1: Rozložení LUCAS bodů měřených v roce 2018	11
Obr. 2: Ukázka části dotazníku pro sběr LUCAS dat [5]	13
Obr. 3: Třídy krajinného pokrytí (lc) [3]	15
Obr. 4: Třídy využití půdy (lu) [3].....	16
Obr. 5: Ukázka části souboru s originálními daty	17
Obr. 6: Ukázka souboru obsahujícího SQL příkazy pro změnu hodnot atributů	21
Obr. 7: Poloha LUCAS bodu.....	30
Obr. 8: Logo QGIS [10].....	33
Obr. 9: Logo Python [12].....	34
Obr. 10: Logo PostgreSQL [14]	35
Obr. 11: Logo GeoServer [16].....	35
Obr. 12: Diagram popisující průběh prací	37
Obr. 13: Kombinace prostorového a atributového dotazu (XML)	41
Obr. 14: Ukázka části Jupyter Notebooku	43
Obr. 15: QGIS Plugin Builder	44
Obr. 16: Diagram popisující funkcionalitu zásuvného modulu	45
Obr. 17: Soubor JSON definující agregaci tříd.....	48
Obr. 18: Plugin window and exmple of request	64
Obr. 19: Define output path	65
Obr. 20: Downloaded points displayed in canvas.....	65
Obr. 21: JSON file with aggregation rules	66
Obr. 22: Example of loaded JSON file into plugin.....	66
Obr. 23: Part of attritute table with aggregated land cover.....	67

Seznam tabulek

Tab. 1: Rozsah datové sady LUCAS	12
Tab. 2: Počet LUCAS atributů.....	18
Tab. 3: Změny názvů atributů.....	20
Tab. 4: Změny hodnot atributu lc_lu_special_remark.....	21
Tab. 5: Harmonizace atributu obs_dist.....	23
Tab. 6: Hodnoty atributu lc1_perc	24
Tab. 7: Počet bodů atributu wm_delivery v jednotlivých letech	25
Tab. 8: Změny datových typů atributů.....	26
Tab. 9: Tabulka hodnot atributu grazing	27
Tab. 10: Skupiny atributů (Tab. 11- Tab. 16).....	47
Tab. 11: Obligatory attributes	60
Tab. 12: Copernicus attributes	61
Tab. 13: Forestry attributes	61
Tab. 14: Inspire attributes	62
Tab. 15: Land cover and Land use attributes	62
Tab. 16: Soil attributes.....	63
Tab. 17: Explanation of attribute values	67

Seznam zkratek

AOI	Area of interest
CORINE	Coordination of Information on the Environment
CSV	Comma-separated values
EPSG	European Petroleum Survey Group
FOSS4G	Free and Open Source Software for Geospatial
FTP	File Transfer Protocol
GDAL	Geospatial Data Abstraction Library
GML	Geography Markup Language
GNU GPL	GNU General Public License
GPKG	GeoPackage
GPS	Global Positioning System
JSON	JavaScript Object Notation
LAEA	Lambert azimuthal equal-area projection
LC	Land Cover
LU	Land Use
LUCAS	Land Use and Coverage Area frame Survey
NUTS	Nomenclature of Territorial Units for Statistics
OGC	Open Geospatial Consortium
PCL	Python Cartography Library
SQL	Structured Query Language
WCS	Web Coverage Service
WFS	Web Feature Service
WMS	Web Map Service
WMTS	Web Map Tile Service
WPS	Web Processing Service
XML	Extensible Markup Language

Seznam příloh

Digitální přílohy

demo	Jupyter Notebook (balíček pyeumap.lucas)
deploy	Automatizace harmonizačních kroků
lc_lu_harmonization_tables	Harmonizační tabulky pro atributy <i>lc</i> a <i>lu</i>
lucas_download_manager	Zásuvný modul LUCAS Download Manager
pyeumap	Python balíček pyeumap.lucas

QGIS plugin documentation

The plugin LUCAS Download Manager has two main purposes: to download LUCAS data and to aggregate the land cover classes.

In the first part user must define required characteristics of points. Based on the request, plugin downloads harmonized LUCAS data.

The first step is to define the AOI (area of interest). Three possibilities are provided there:

- canvas – A bounding box defined by current map window is used.
- country – User can choose some of the 29 countries provided in combo box.
- vector layer – User should select one of the polygon layers loaded in the canvas. It must be layer with only one polygon feature. There is also a limit to the number of vertices.

The second option, which should be defined, is the year of the survey. There are five years: 2006, 2009, 2012, 2015, 2018 and the user can choose one of them or any combination of them.

The third step is the definition of group of attributes, which will contain the resulting file. The file always contains obligatory attributes, which define mainly the location of the point. Five groups of attributes are provided there. In the following tables there are names of these groups, the attributes they contain and a description of each attribute.

Tab. 11: Obligatory attributes

Obligatory	
Attribute	Description
point_id	Unique identifier
nuts0	NUTS level 0
nuts1	NUTS level 1
nuts2	NUTS level 2
nuts3	NUTS level 3
survey_date	Date of observation
survey_year	Year of observation
office_pi	Sample photo interpreted in office
ex_ante	Visited in the field
car_latitude	GPS car parking latitude
car_longitude	GPS car parking longitude
car_ew	GPS car parking east/west
gps_proj	GPS projection
gps_prec	GPS precision (average location error in meters)
gps_lat	GPS observation latitude
gps_long	GPS observation longitude
gps_altitude	GPS altitude
th_lat	Latitude of the theoretical point
th_long	Longitude of the theoretical point
dist_gps_th	Distance between observational and theoretical point - computed from coordinates [meters]
dist_th_thr	Distance between theoretical point and rounded coordinates of theoretical point - computed from coordinates [meters]
obs_dist	Distance between observational and theoretical point - original attribute (except year 2006) [meters]
obs_direct	Direction to theoretical point
obs_type	Observation type
obs_radius	Radius of observation circle

Tab. 12: Copernicus attributes

Copernicus	
Attribute	Description
cprn_cando	Copernicus survey was done
cprn_lc	Copernicus land cover on observation point
cprn_lc1n	Distance with the same copernicus land cover as observation point to north [meters]
cprnc_lc1e	Distance with the same copernicus land cover as observation point to east [meters]
cprnc_lc1s	Distance with the same copernicus land cover as observation point to south [meters]
cprnc_lc1w	Distance with the same copernicus land cover as observation point to west [meters]
cprn_lc1n_brdth	Percentage of next copernicus land cover visible on the photo in the north direction [%]
cprn_lc1e_brdth	Percentage of next copernicus land cover visible on the photo in the east direction [%]
cprn_lc1s_brdth	Percentage of next copernicus land cover visible on the photo in the south direction [%]
cprn_lc1w_brdth	Percentage of next copernicus land cover visible on the photo in the west direction [%]
cprn_lc1n_next	Next copernicus land cover (different from lc on observation point) in the north direction
cprn_lc1e_next	Next copernicus land cover (different from lc on observation point) in the east direction
cprn_lc1s_next	Next copernicus land cover (different from lc on observation point) in the south direction
cprn_lc1w_next	Next copernicus land cover (different from lc on observation point) in the west direction
cprn_urban	Point in urban area
cprn_impervious_perc	Percentage of area with no vegetation, using the birds-eye view [%]

Tab. 13: Forestry attributes

Forestry	
Attribute	Description
tree_height_survey	Height of trees at survey time
tree_height_maturity	Height of trees at maturity

Tab. 14: Inspire attributes

Inspire	
Attribute	Description
inspire_plcc1	Percentage of coniferous forest trees, using the birds-eye view [%]
inspire_plcc2	Percentage of broadleaved forest trees, using the birds-eye view [%]
inspire_plcc3	Percentage of shrubs, using the birds-eye view [%]
inspire_plcc4	Percentage of herbaceous plants, using the birds-eye view [%]
inspire_plcc5	Percentage of lichens and mosses, using the birds-eye view [%]
inspire_plcc6	Percentage of consolidated bare land, using the birds-eye view [%]
inspire_plcc7	Percentage of unconsolidated bare land, using the birds-eye view [%]
inspire_plcc8	Percentage of other land, using the birds-eye view [%]

Tab. 15: Land cover and Land use attributes

Land cover and land use	
Attribute	Description
lc1	Dominant land cover class (original)
lc1_h	Dominant land cover class harmonized (reference year 2018)
lc1_spec	Dominant land cover species class
lc1_perc	Percentage of coverage of dominant land cover class
lc2	Subsidiary land cover class (original)
lc2_h	Subsidiary land cover class harmonized (reference year 2018)
lc2_spec	Subsidiary land cover species class
lc2_perc	Percentage of coverage of subsidiary land cover class
lu1	Dominant land use class (original)
lu1_h	Dominant land use class harmonized (reference year 2018)
lu1_type	Dominant land use type class
lu1_perc	Percentage of coverage of dominant land use class
lu2	Subsidiary land use class (original)
lu2_h	Subsidiary land use class harmonized (reference year 2018)
lu2_type	Subsidiary land use type class
lu2_perc	Percentage of coverage of subsidiary land use class
parcel_area_ha	Area of the parcel to which the point belongs [hectare]
feature_width	Feature width
lm_plough_slope	Slope of ploughed field
lm_plough_direct	Plough direction
lm_stone_walls	Presence of stone walls
lm_grass_margins	Presence of grass margins
eunis_complex	EUNIS complex if there is agro-forestry
grassland_sample	Sample grassland module
grass_cando	Grassland survey was done
Grazing	Signs of grazing



Wm	Presence of water management
wm_source	Source of irrigation
wm_type	Type of irrigation
wm_delivery	Water delivery system
Indmng_plough	Signs of ploughing
special_status	Special status – protected/hunting
lc_lu_special_remark	Special remarks on land cover/land use
photo_point	Photo of point taken
photo_north	Photo of north direction taken
photo_east	Photo of east direction taken
photo_south	Photo of south direction taken
photo_west	Photo of west direction taken
crop_residues	Presence of crop residues
Transect	List of land cover codes

The group of attributes named „Land cover, land use and soil“ contains the same attributes as the group „Land cover and land use“ and in addition to that also attributes, which describe the soil. In the following tables only „Soil“ attributes are presented.

Tab. 16: Soil attributes

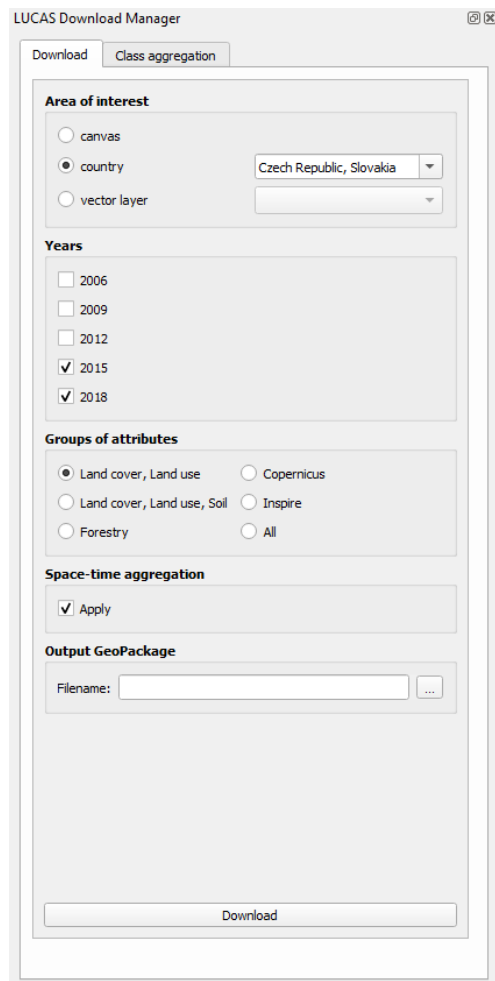
Soil (part of Land cover, land use and soil)	
Attribute	Description
soil_taken	Soil sample taken
erosion_cando	Erosion taken
bio_sample	Sample BIO SOIL module
soil_bio_taken	BIO SOIL survey was done
bulk0_10_sample	Sample BULK 0-10 module
soil_blk_0_10_taken	BULK 0-10 taken
bulk10_20_sample	Sample BULK 10-20 module
soil_blk_10_20_taken	BULK 10-20 taken
bulk20_30_sample	Sample BULK 20-30 module
soil_blk_20_30_taken	BULK 20-30 taken
standard_sample	Sample STANDARD SOIL module
soil_std_taken	Standard Soil survey was done
organic_sample	Sample ORGANIC SOIL module
soil_org_depth_cando	Organic Soil taken
soil_stones_perc	Percentage of stones on the surface

When the required group of attributes is defined, there is a possibility to choose whether the data should be space-time aggregated. Space-time aggregation means that

one record in the attribute table represents one point with all values measured in all years. On the other hand, when data are not space-time aggregated, every single record in the attribute table represents one survey. The difference is, when a point is measured in 2012, 2015 and 2018, if data are space-time aggregated, there is only one record in attribute table, if they are not space-time aggregated, there are three records.

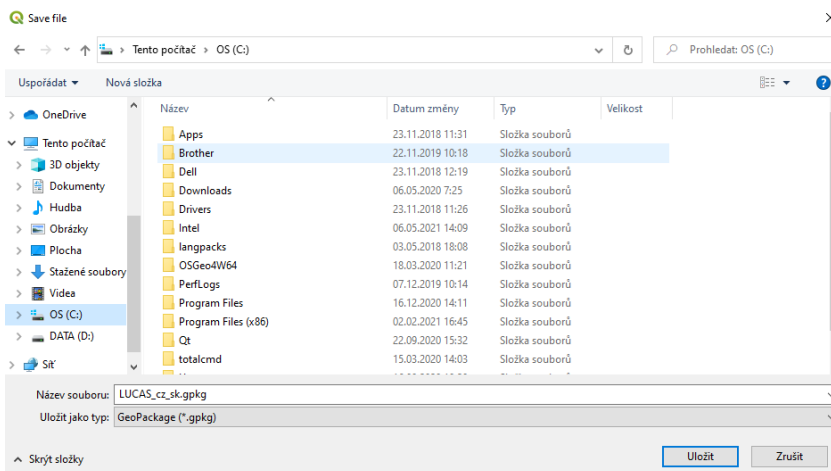
Then the user must choose the output path for the *GPKG* file. When the „Download“ button is pressed, the required data are downloaded into the selected directory, the *GPKG* file is loaded as a new layer into the map canvas and the number of downloaded features is displayed in the plugin window.

In the picture (Obr. 18) there is an example of the request. Points which will be downloaded are in the Czech Republic or Slovakia, measured in year 2015 or 2016. Attribute table will contain beside obligatory attributes only attributes connected to land cover and land use. Data will be space-time aggregated.



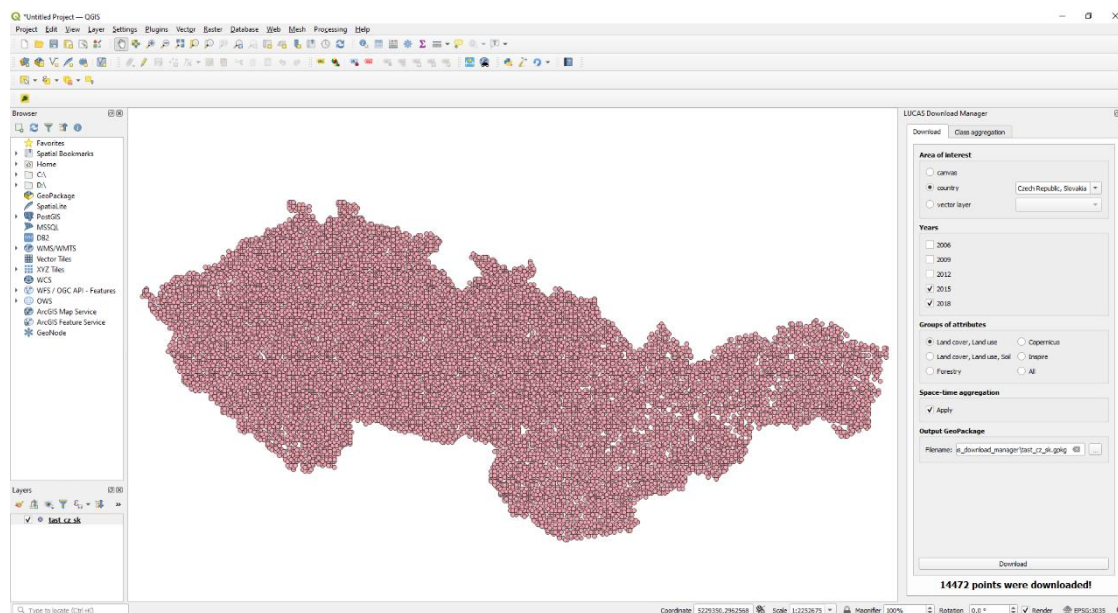
Obr. 18: Plugin window and exmple of request

Then it is necessary to define the output path. The picture (Obr. 19) shows how to do it.



Obr. 19: Define output path

Finally, all the required points are downloaded and displayed in the map canvas, and in the plugin window there is a number of downloaded features displayed (in the picture it is 14 472 features) (Obr. 20).



Obr. 20: Downloaded points displayed in canvas

The second part of the plugin LUCAS Download Manager is called „Class aggregation“. It aggregates harmonized land cover classes according to an aggregation rules, which must be defined in the *JSON* file. Two examples of such file are provided in the plugin directory. The user can modify them and use them or create a new one. There is one of the sample *JSON* file in the picture (Obr. 21).

```

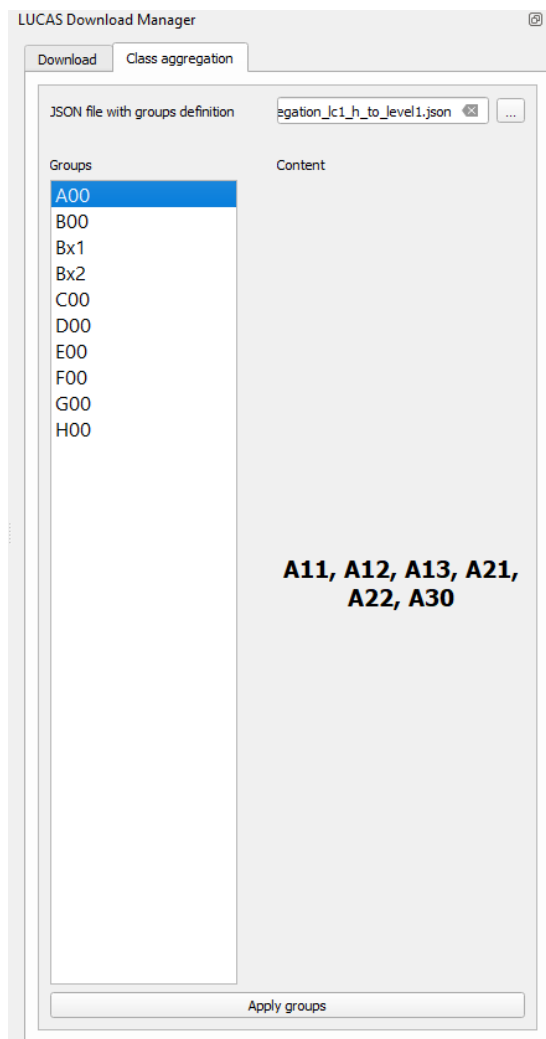
aggregation_lc1_h_to_level1.json – Poznámkový blok
Soubor Úpravy Formát Zobrazení Nápověda
{
  "A00": ["A11", "A12", "A13", "A21", "A22", "A30"],
  "B00": ["B11", "B12", "B13", "B14", "B15", "B16", "B17", "B18", "B19", "B20"],
  "Bx1": ["Bx1"],
  "Bx2": ["Bx2"],
  "C00": ["C10", "C21", "C22", "C23", "C31", "C32", "C33"],
  "D00": ["D10", "D20"],
  "E00": ["E10", "E20", "E30"],
  "F00": ["F10", "F20", "F30", "F40"],
  "G10": ["G11", "G12", "G21", "G22", "G30", "G40", "G50"],
  "H10": ["H11", "H12", "H21", "H22", "H23"]
}
  
```

Obr. 21: *JSON* file with aggregation rules

When the *JSON* file is loaded into the plugin, there are names of groups displayed. At the time when some of these names is clicked content of that group is displayed there as well. In the picture (Obr. 22) there is an example. There are ten groups (A00, B00, Bx1 etc.). Land cover classes A11, A12, A13, A21, A22 and A30 will be aggregated into the class A00.

The *GPKG* file, which is used for aggregation is defined in the previous part „Download“, so it is usually the last downloaded file.

When the „Apply Groups“ button is pressed, a new column for aggregated classes is created in the attribute table and they are written into it.



Obr. 22: Example of loaded *JSON* file into plugin

In the picture (Obr. 23) there is a part of attribute table with harmonized land cover attributes (*lc1_h_2015*, *lc1_h_2018*) and also aggregated land cover attributes (*lc1_a_2015*, *lc1_a_2018*). Aggregation rules defined in the file displayed in the picture were used there (Obr. 21).

	lc1_h_2015	lc1_h_2018	lc1_a_2015	lc1_a_2018
6613	NULL	B13	NULL	B00
6614	NULL	B32	NULL	B00
6615	NULL	C10	NULL	C00
6616	B13	B11	B00	B00
6617	NULL	E20	NULL	E00
6618	C31	C31	C00	C00
6619	NULL	E20	NULL	E00
6620	NULL	B11	NULL	B00
6621	NULL	B16	NULL	B00
6622	B11	B32	B00	B00
6623	C10	C10	C00	C00
6624	H11	H11	H00	H00

Obr. 23: Part of attribute table with aggregated land cover

While plugin is used, some issues can occur. These issues can be caused by the user, but also by something else (for example, when GeoServer does not working). When some problem happens, a message for the user is displayed using the QGIS message bar.

Values of most attributes are codes. The user needs to know the meaning of these codes. This is explained in the following tables.

Tab. 17: Explanation of attribute values

gps_proj	
1	WGS89
2	Gps problem
8	Not Relevant

lm_plough_direct	
1	Across the slope
2	Down the slope
3	Not Applicable
8	Not Relevant



gps_ew	
th_ew	
car_ew	
1	East
2	West
8	Not Relevant

lm_grass_margins	
1	No
2	Grass margin \leq 1 m width
3	Grass margin $>$ 1 m width
8	Not Relevant

obs_direct	
1	On the point
2	Look to the North
3	Look to the East
8	Not Relevant
-1	Not valid

lc_lu_special_remark	
1	Harvested field
2	Tilled/sowed
3	Clear cut
4	Burnt area
5	Fire break
6	Nursey
7	Dump site
8	Temporary dry
9	Temporary flooded
10	No remark
88	Not Relevant

parcel_area_ha	
1	area $<$ 0,5
2	$0,5 \leq$ area $<$ 1
3	$1 \leq$ area $<$ 10
4	area \geq 10
8	Not Relevant
-1	Not valid

wm_source	
1	Well
2	Pond/Lake/Reservoir
3	Stream/Canal/Ditch
4	Lagoon/Wastewater
5	Other/Not identifiable
8	Not Relevant
-1	Not valid

feature_width	
1	$<$ 20 m
2	\geq 20 m
8	Not Relevant
-1	Not valid

lm_stone_walls	
1	No
2	Stone wall not maintained
3	Stone wall well maintained
8	Not Relevant

wm_delivery	
1	Canal
2	Ditch
3	Pipeline
4	Other/Not identifiable
8	Not Relevant
-1	Not valid

grazing	
1	Visible signs of grazing
2	No sign of grazing
8	Not Relevant
-1	Not valid

special_status	
1	Protected
2	Hunting
3	Protected and hunting
4	No special status
8	Not Relevant

wm	
1	Irrigation
2	Potential irrigation
3	Drainage
4	Irrigation and drainage
5	No visible WM
8	Not Relevant

wm_type	
1	Gravity
2	Pressure: sprinkler irrigation
3	Pressure: micro-irrigation
4	Gravity/Pressure
5	Other/Not identifiable
8	Not Relevant
-1	Not valid

eunis_complex	
6	X06
9	X09
10	Other
11	Unknown
88	Not Relevant

photo_point	
photo_north	
photo_east	
photo_south	
photo_west	
1	Taken
2	Not taken
8	Not Relevant
-1	Not valid

office_pi	
ex_ante	
grassland_sample	
bio_sample	
bulk0_10_sample	
bulk10_20_sample	
bulk20_30_sample	
bulk20_30_sample	
standard_sample	
organic_sample	
0	False
1	True

obs_radius	
1	3 meters
2	20 meters
8	Not Relevant

lc1_perc	
lc2_perc	
0	0
1	$1 \leq lc < 10$
2	$10 \leq lc < 25$
3	$25 \leq lc < 50$
4	$50 \leq lc < 75$
5	$75 \leq lc \leq 100$
8	Not Relevant
-1	Not valid

Indmng_plough	
crop_residues	
cprn_cando	
cprn_urban	
grass_cando	
erosion_cando	
soil_blk_0_10_taken	
soil_blk_10_20_taken	
soil_blk_20_30_taken	
soil_std_taken	
soil_bio_taken	
soil_org_depth_cando	
1	Yes
2	No
8	Not Relevant

lu1_perc	
lu2_perc	
0	0
1	$1 \leq lu < 5$
2	$5 \leq lu < 10$
3	$10 \leq lu < 25$
4	$25 \leq lu < 50$
5	$50 \leq lu < 75$
6	$75 \leq lu < 90$
7	$90 \leq lu \leq 100$
8	Not Relevant
-1	Not valid

obs_type	
1	In situ < 100 meters
2	In situ > 100 meters
3	In situ PI
4	In situ PI not possible
5	Out of national territory
6	Out of EU28
7	In Office PI
8	Marine sea

soil_stones_perc	
0	0
1	$1 \leq soil_stone < 10$
2	$10 \leq soil_stone < 25$
3	$25 \leq soil_stone < 50$
4	$50 \leq soil_stone \leq 100$
8	Not Relevant

tree_height_survey	
tree_height_maturity	
1	< 5 m
2	≥ 5 m
8	Not Relevant
-1	Not valid

soil_taken	
1	Yes
2	Not possible
3	No, already taken
4	No sample required
8	Not Relevant

lm_plough_slope	
1	Flat
2	Gently sloping
3	Steeply sloping
4	Undulating
8	Not Relevant