**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 |

Obsah obrázku text

Popis byl vytvořen automaticky

**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**

**Anotace**

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 modulu lucas a samotného pluginu, který tento modul využívá.

Zásuvný modul poskytuje možnost na základě zvolených parametrů stáhnout pomocí 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 land cover tříd.

**Klíčová slova**

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

**Annotation**

Překlad anotace…..

**Key words**

QGIS, plugin, LUCAS, PostGIS, harmonization, Python

**Obsah**

[Úvod 9](#_Toc70857502)

[1 LUCAS 11](#_Toc70857503)

[1.1 LUCAS 11](#_Toc70857504)

[1.2 Sběr dat 13](#_Toc70857505)

[1.3 Kontrola kvality 14](#_Toc70857506)

[1.4 Klasifikace krajiny 14](#_Toc70857507)

[1.4.1 Land cover 14](#_Toc70857508)

[1.4.2 Land use 15](#_Toc70857509)

[1.5 Rešerše 16](#_Toc70857510)

[2 Harmonizace 17](#_Toc70857511)

[2.1 Přejmenování atributu 18](#_Toc70857512)

[2.2 Změna hodnot 20](#_Toc70857513)

[2.2.1 Land cover a land use 21](#_Toc70857514)

[2.2.2 Observation distance 21](#_Toc70857515)

[2.3 Převod na intervaly 22](#_Toc70857516)

[2.4 Nevalidní hodnoty 23](#_Toc70857517)

[2.5 Změna datových typů 24](#_Toc70857518)

[2.6 Nové atributy 26](#_Toc70857519)

[2.7 Automatizace 26](#_Toc70857520)

[2.8 Zveřejnění dat 27](#_Toc70857521)

[3 Použité technologie 28](#_Toc70857522)

[3.1 QGIS 28](#_Toc70857523)

[3.2 Python 29](#_Toc70857524)

[3.3 PostgreSQL 29](#_Toc70857525)

[3.4 GeoServer 30](#_Toc70857526)

[4 Postup implementace 31](#_Toc70857527)

[5 Modul lucas 32](#_Toc70857528)

[5.1 Sestavení dotazu 32](#_Toc70857529)

[5.2 Vytvoření výstupů 34](#_Toc70857530)

[5.3 Agregace land cover tříd 35](#_Toc70857531)

[5.4 Jupyter Notebook 35](#_Toc70857532)

[6 Zásuvný modul 35](#_Toc70857533)

[6.1 Funkcionalita 38](#_Toc70857534)

[6.1.1 Download 38](#_Toc70857535)

[6.1.2 Class aggregation 40](#_Toc70857536)

[Závěr 41](#_Toc70857537)

[Použité zdroje 42](#_Toc70857538)

[Seznam obrázků 44](#_Toc70857539)

[Seznam tabulek 45](#_Toc70857540)

[Seznam zkratek 46](#_Toc70857541)

[Seznam příloh 47](#_Toc70857542)

# Úvod

Tato diplomová práce se zabývá tvorbou nástroje, který umožní snadný přístup k bodům datové sady LUCAS.

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“. Prací na tomto projektu se zabývají odborníci z pěti zemí Evropské unie, a to Německa, Holandska, 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]

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 land cover a land use, které především určují charakter krajiny v okolí daného bodu.

Druhá část diplomové práce popisuje způsoby, jakými byla data před 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. Nejvýznamnější z nich byl programovací jazyk Python a jeho rozšíření pro program QGIS zvaný PyQGIS, pomocí kterého byl vytvářen samotný zásuvný modul a modul lucas. Pro práci s datovou sadou LUCAS a její úpravu byl použit jazyk PostGIS a jeho rozšíření PostgreSQL. V neposlední řadě je zde popsán samotný program QGIS, pro který byl zásuvný modul vytvářen a mapový server, na kterém jsou data zveřejněna.

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

V páté kapitole je představena funkcionalita modulu lucas. Dále je zde zmíněn Jupyter Notebook, jenž byl vytvořen jako názorná ukázka možností tohoto modulu.

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

# LUCAS

## 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-snsing 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.

Obsah obrázku mapa

Popis byl vytvořen automatickyPři 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.

Obr. : Rozložení LUCAS bodů měřených v roce 2018

|  |  |  |
| --- | --- | --- |
| 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 |

Tab. 1: Rozsah LUCAS

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

[2], [3]

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

<https://ec.europa.eu/eurostat/web/lucas/overview>

## 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). Druhou skupinou jsou 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 necelých 100 000 bodů interpretováno v kanceláři. Z těchto čísel vyplývá, že data popisující pouze necelou třetinu celkového počtu bodů nebyla určena terénním šetřením.

Obsah obrázku stůl

Popis byl vytvořen automatickyPrůzkum 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, znázorňuje obrázek (Obr. 1). 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.

Obr. : Ukázka dotazníku LUCAS [4]

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. Je pořizována fotografie samotného bodu i jeho okolí. V některých případech je navíc odebírán vzorek půdy, který je následně podroben rozboru.

[2], [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 celého datasetu
* prvních 20 % bodů měřených novým pracovníkem

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

[2], [3]

## 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 v okolí každého bodu se může vyskytovat vícero krajinných typů, případně toto okolí může být využíváno vícero způsoby, určuje pracovník v terénu hlavní (*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.

### Land cover

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. Obrázek (Obr.:) vyobrazuje první a druhou úroveň dělení.

Obsah obrázku stůl

Popis byl vytvořen automaticky[5]

Obr. : Land cover třídy [2]

### Land use

Atribut využití půdy popisuje, jakým způsobem 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.: ) znázorňuje rozdělení LU tříd první a druhé úrovně.

[5]

Obsah obrázku stůl

Popis byl vytvořen automaticky

Obr. : Land use třídy [2]

<https://geoportal.cuzk.cz/(S(sgmfbk3yfcqbsovpwd1aet1x))/Default.aspx?mode=TextMeta&text=INSPIRE_K_Pokryv&side=INSPIRE_dSady&head_tab=sekce-04-gp&menu=421>

<https://ec.europa.eu/eurostat/statistics-explained/index.php/LUCAS_-_Land_use_and_land_cover_survey#The_LUCAS_survey>

## Rešerše

Přestože jsou data z datasetu LUCAS bodů 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 tohoto datasetu, tedy Evropského statistického úřadu (Eurostat) <https://ec.europa.eu/eurostat/web/lucas/data/primary-data>. Tato data 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*. Vzhledem k tomu, že se jedná o prostorová data, je nutné je pro efektivní práci s nimi převést do jiného formátu.

Výše popsané problémy částečně řešil 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 ([3]). 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. Data je možné stáhnou pomocí FTP (File Transfer Protocol). Dostupná jsou na adrese <https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/LUCAS/LUCAS_harmonised/>.

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.

[3]

# Harmonizace

Přestože je databáze LUCAS vytvářena a udržována organizací Eurostat, tedy statistickým úřadem Evropské unie, vyskytuje se v ní mnoho nekonzistencí napříč roky. Jedním z důvodů, proč tomu tak je, je postupný vývoj a rozšiřování počtu měřených parametrů území. Tabulka (Tab. 4) zobrazuje jejich počet v jednotlivých letech.

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

Tab. 2: Počet LUCAS atributů

Vznik nových atributů v průběhu let nebyl jediným typem změny. Dalšími změnami, které v mnoha případech znepřehlednily 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 v současnosti nejaktuálnější verzi, 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 ([3]) popsaný v kapitole 1.5, jenž se touto problematikou zabýval. V následujících podkapitolách jsou popsány jednotlivé případy, ve kterých se přístupy k danému problému, využité 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 dat. Příklady jednotlivých změň, které databázi upravovaly, jsou popsány v následujících podkapitolách.

## 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 nevyskytuje, avšak nahradil ho atribut *special\_status,* obsahující totožné hodnoty.

Atributy byly přejmenovávány stejným způsobem, jaký je popsán v článku [3], s výjimkou atributu *soil\_crop,* vyskytujícím 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.: ).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **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 |  |  |  |

Tab. 3: Změny názvů atributů

## 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 *lc\_lu\_special\_remark* a dokumentuje ji tabulka (Tab. 5). 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.

|  |  |  |  |
| --- | --- | --- | --- |
| 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 |

Tab. 4: Změna hodnot atributu lc\_lu\_special\_remark

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 [X], ale také výše zmíněný článek [3] a jeho výsledky. Převážná většina těchto změn byla totožná se změnami uvedenými v článku [3].

Obsah obrázku text

Popis byl vytvořen automatickyDílčí 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.).

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

### Land cover a land use

V případě atributů land cover (*lc1*, *lc2*) a land use (*lu1*, *lu*2) byly hodnoty těchto atributů 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), byla místo této třídy nastavena hodnota 8, představující nerelevantní data. Taktéž u tříd, v jejichž řádku byly pouze hodnoty 0 (daná třída zanikla a v současnosti neexistuje žádná ekvivalentní třída), byla nastavena hodnota 8. 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 [3]. 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 [3] 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 znepřehledňuje daný atribut. Příkladem je třída land use U500 „Wetland“, která existovala pouze v roce 2006.

### Observation distance

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 teoretického bodu od bodu měřeného. Harmonizace tohoto atributu byla provedena následujícím způsobem. V případě, že byly u daného bodu měřeny souřadnice bodu, na kterém bylo prováděno měření, byla vzdálenost k teoretickému bodu dopočítána a do atributu uložena tato 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. ).

|  |  |  |
| --- | --- | --- |
| **obs\_dist (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 |

## 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. 4) popisuje použité intervaly převzaté z roku 2009 pro atribut *lc1\_perc*.

|  |  |
| --- | --- |
| Kód | Hodnota *[%]* |
| 0 | 0 |
| 1 | 1 ≤ *lc* < 10 |
| 2 | 10 ≤  *lc* < 25 |
| 3 | 25 ≤  *lc* < 50 |
| 4 | 50 ≤  *lc* < 75 |
| 5 | 75 ≤  *lc* ≤ 100 |

Tab. 5: Hodnoty atributu lc1\_perc

Stejný postup je popsán i v článku [3]. Jediným rozdílem je způsob práce s nevalidními hodnotami, které se mimo jiné i ve výše zmíněných atributech vyskytují. Tento problém je popsán v následující kapitole.

## 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 jednotlivých hodnot stejného atributu v ostatních letech. V případě, že 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 hodnot, kterými jsou v daném případě 5, 6, 10 a 12, již není možné jednoznačně určit, co představují. Tyto kódy byly proto změněny na -1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 | - | - | - |

Tab. 6: Počet bodů atributu wm\_delivery v letech

V databázi LUCAS bodů bylo identifikováno značné množství nevalidních hodnot. 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 [3] 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.

## Změna datových typů

Další úpravou, kterou bylo nutné aplikovat na data, bylo nastavení atributům odpovídajících datových typů. Byl vytvořen skript pomocí jazyka Python, jehož vstupem byl *csv* soubor, definující atributům jejich datové typy. Skript automaticky vygeneroval SQL dotazy sloužící ke změně datových typů. Tabulka (Tab. ) 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.

|  |  |  |  |
| --- | --- | --- | --- |
| **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 |  |  |

Tab. 7: Změny datových typů atributů

Z tabulky (Tab. ) 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 příloze této práce.

Jiný přístup k datovým typům a s tím souvisejícím hodnotám, které atributy nabývají, je popsán v článku [3]. 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. ) zobrazuje hodnoty, které atribut *grazing* nabývá v databázi vytvořené na základě článku [3] a pro porovnání i odpovídající hodnoty, obsahující tento atribut po úpravách provedených na základě této práce.

|  |  |
| --- | --- |
| **grazing** | |
| **Hodnoty dle článku [3]** | **Hodnoty dle této práce** |
| Signs of grazing | 1 |
| No signs of grazing | 2 |
| Not relevant | 8 |

Tab. 8: Tabulka hodnot atributu grazing

Nevýhodou zmíněného přístupu je především náročná práce s textovými řetězci. Například při výběru na základě atributového dotazu je uživatel nucen vypsat 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.

## 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 – Ročník průzkumu, pro který byl bod měřen.
* geog\_th – Geometrie polohy teoretického bodu určená z atributů *th\_lat*, *th\_long* a *th\_ew*.
* geog\_gps – Geometrie polohy bodu, na kterém bylo měření prováděno určená z atributů *gps\_lat*, *gps\_long*, *gps\_ew*.
* geom\_gps
* geom\_thr – Geometrie polohy teoretického bodu určená ze zaokrouhlených hodnot atributů *th\_lat*, *th\_long* a atributu *th\_ew*.
* dist\_gps\_th – Vypočtená vzdálenost v metrech mezi bodem, na kterém probíhalo měření, a teoretickým bodem.
* dist\_th\_thr – Vypočtená vzdálenost v metrech mezi teoretickým bodem a teoretickým bodem, určeným ze zaokrouhlených souřadnic.
* dist\_thr\_grid -
* lc1\_h, lc2\_h, lu1\_h, lu2\_h – Harmonizované land cover a land use atributy.

## Automatizace

Ve chvíli, kdy vznikne nová databáze LUCAS bodů měřených v roce 2021, bude pravděpodobně tato databáze obsahovat nekonzistence oproti databázi z roku 2018. Z tohoto důvodu budou nutné nová data opět harmonizovat. Proto, aby nová harmonizace byla co nejsnazší, byl celý proces automatizován. Jednotlivé části propojil Ing. Martin Landa, Ph.D. pomocí softwaru Docker.

## 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 GeoServer. Vznikla tak 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*. Každý záznam agregovaných dat představuje bod, u něhož jsou zaznamenány 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 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, a to „land cover a land use“, „land cover, land use a soil“, „forestry“, „copernicus“ a „inspire“. Využití těchto předem připravených skupin má velký význam především v případě časoprostorově agregovaných dat, jelikož celá atributová tabulka obsahuje více než 500 atributů.

# Použité technologie

## QGIS



Obr. : Logo QGIS [7]

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 zaručen jejich open source charakter. QGIS je možné používat na všech nejrozšířenějších operačních systémech, kterými jsou například MS Windows, GNU/Linux a Mac OS X. Silnou stránkou Programu QGIS jsou zásuvné moduly (pluginy). Plugin je doplněk programu, který rozšiřuje funkčnost základního programu. Pluginy 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 jsou k dispozici tři verze, a to jedna dlouhodobá – stabilní a dvě krátkodobé, které se rychle vyvíjí a mění.

[8]

## Python



Obr. : Logo Python [9]

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.

[10]

## PostgreSQL



Obr. 8: Logo PostgreSQL [11]

PostgreSQL je objektově-relační databázový systém, který využívá a rozšiřuje jazyk SQL. Vývoj tohoto systému začal v roce 1986 na Kalifornské univerzitě. Je vyvíjen 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 například PostGIS, umožňující efektivní práci s prostorovými daty.

[12]

## GeoServer



Obr. : Logo GeoServer [13]

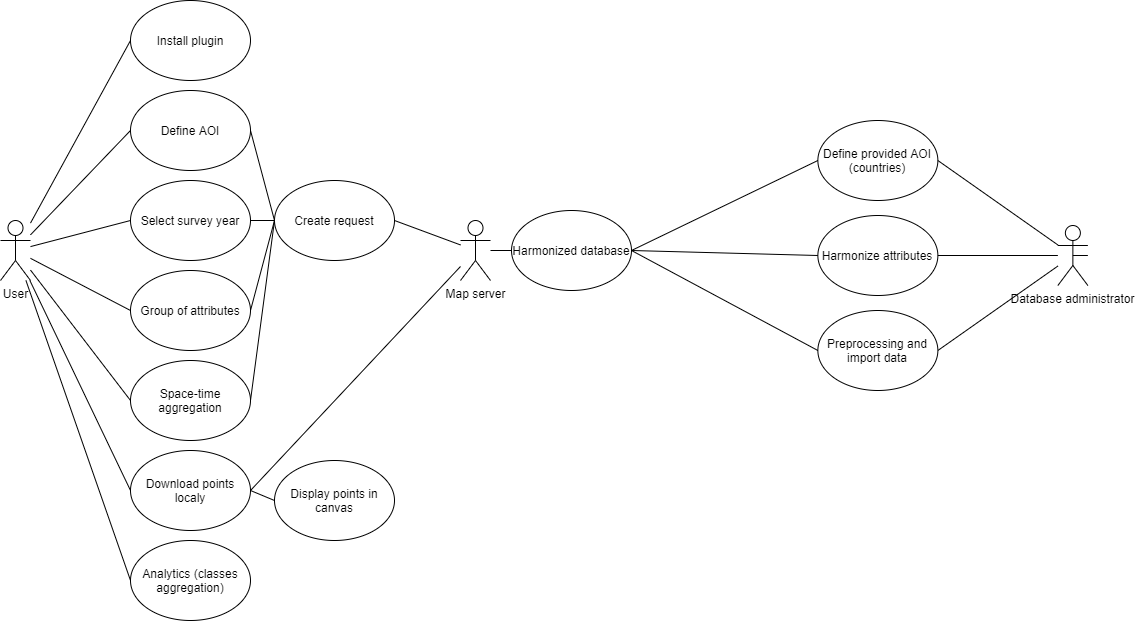
GeoServer je open source mapový server vytvořený pomocí programovacího jazyka Java. Je uveřejněn pod licencí GNU GPL. Je vyvíjen a testován velkou skupinou jednotlivců, ale také organizacemi z celého světa.

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 například o možnost využívání WPS (Web Processing Service). Po publikaci na GeoServeru je možné k datům přistupovat pomocí webového rozhraní.

[14]

# Postup implementace

V této kapitole je popsán průběh implementace. Práci je možné rozdělit do dvou částí, kterými jsou tvorba modulu lucas a vývoj samotného zásuvného modulu, jenž zmíněný modul využívá.



Obr. : Diagram popisující průběh prací

Obrázek (Obr. 10) zobrazuje diagram obsahující tři aktéry, kterými jsou administrátor databáze Lucas bodů, mapový server a uživatel zásuvného modulu. Jsou zde znázorněny procesy, které musí zmínění aktéři provést 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 Harmonizace.

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 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.

# Modul lucas

Modul lucas byl vytvářen v rámci projektu GeoHarmonizer 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.

Modul lucas umožňuje uživateli přístup k datům LUCAS a jejich následné zpracování. 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 zpracování získaných dat. Zpracování dat představuje možnost agregace land cover tříd na základě předem definovaných pravidel. Všechny zmíněné části jsou popsány v následujících kapitolách.

## Sestavení dotazu

Sestavení dotazu zabezpečuje soubor *request.py*. Tento soubor obsahuje jednu třídu, a to LucasRequest, jejíž proměnné definují parametry výsledného dotazu. Těmito proměnnými 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.
  + literar – 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.
  + logical – Logický operátor spojující atributové dotazy. Určuje, zda mají podmínky platit zároveň (*And*), případně zda postačí platnost jedné z podmínek (*Or*).
* polygoninter – Pomocí textového řetězce ve formátu *gml,* obsahujícího lomové body polygonu, určuje území, ve kterém se mají body nacházet.
* years – Je parametr, pomocí kterého je možný výběr na základě let, ve kterých byly body měřeny. Zadává se pomocí seznamu, jež může obsahovat některou z hodnot 2006, 2009, 2012, 2015, 2018, případně jejich libovolnou kombinaci.
* typename – Definuje název WFS vrstvy.
* staggregation – 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 mají data obsahovat.

Třída LucasRequest má tři metody. Nejdůležitější z nich, která vytváří samotný dotaz, je metoda *build*. Tato metoda nejprve volá metodu *\_chceck\_args*, která zabezpečuje kontrolu toho, zda jsou vstupní argumenty zadány správně. Následně je dotaz postupně vytvářen pomocí metod OWSLib knihovny. Jednotlivé části dotazu jsou generovány na základě toho, které argumenty byly zadány. Pro spojení dílčích částí jsou využívány metody OWSLib knihovny *And* a *Or*. Na závěr je výsledný objekt převeden do formátu *xml*, který je 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*.

Obr. 11: Kombinace prostorového a atributového dotazu (xml)

Obrázek (Obr. 10) zobrazuje výsledné *xml* vygenerované metodou build v případě, že mají být vybrány body, které byly měřeny v roce 2012 nebo 2018 a nacházejí se uvnitř obdélníku, definovaného souřadnicemi levého dolního rohu (4624127, 2998330) a pravého horního rohu (4650393, 3013986).

## Vytvoření výstupů

O vytvoření výstupů se stará soubor *io.py* a jediná třída, kterou obsahuje 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 využívána knihovna OWSLib. Požadavek je odeslán na GeoServer. WFS 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.

Metoda *to\_gml* převádí *gpkg* soubor uložený v přechodném adresáři a navrací textový řetězec ve formátu *gml*.

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. Funkcionalita spočívá ve vytvoření kopie souboru, dříve uloženého do přechodného adresáře, ve zvoleném adresáři a jeho přejmenování na požadovaný název.

Další metodou je *to\_geopandas*, jejímž úkolem je pouze návrat dat ve formátu *GeoDataFrame*. Tento formát usnadňuje práci s atributovou tabulkou pomocí jazyka Python.

Metody *num\_of\_features* a *is\_empty* spolu souvisí. První z nich opět pracuje s *gpkg* souborem uloženým v přechodném adresáři, který otevírá pomocí metody *OpenEx* z knihovny GDAL. Následně je určen počet prvků, které soubor obsahuje. Zmíněný počet je návratovou hodnotou metody. Druhá metoda na základě zjištěného počtu prvků určuje, zda soubor obsahuje alespoň jeden prvek. Pokud *gpkg* soubor není prázdný, vrací metoda hodnotu *False*. V opačném případě je metodou navráceno *True*.

## Agregace land cover tříd

Třetí částí modulu lucas je soubor *analyze.py*. Tento soubor obsahuje třídu LucasClassAggr, která slouží k agregaci land cover tříd.

Metoda *\_load\_classes* na základě vstupního *json* souboru 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 připojí *gpkg* soubor pomocí knihovny sqlite3 tak, aby bylo možné prostřednictvím Python volat 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 spouštěním 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 vytvořen odpovídající počet nových atributů pro agregované třídy.

## Jupyter Notebook

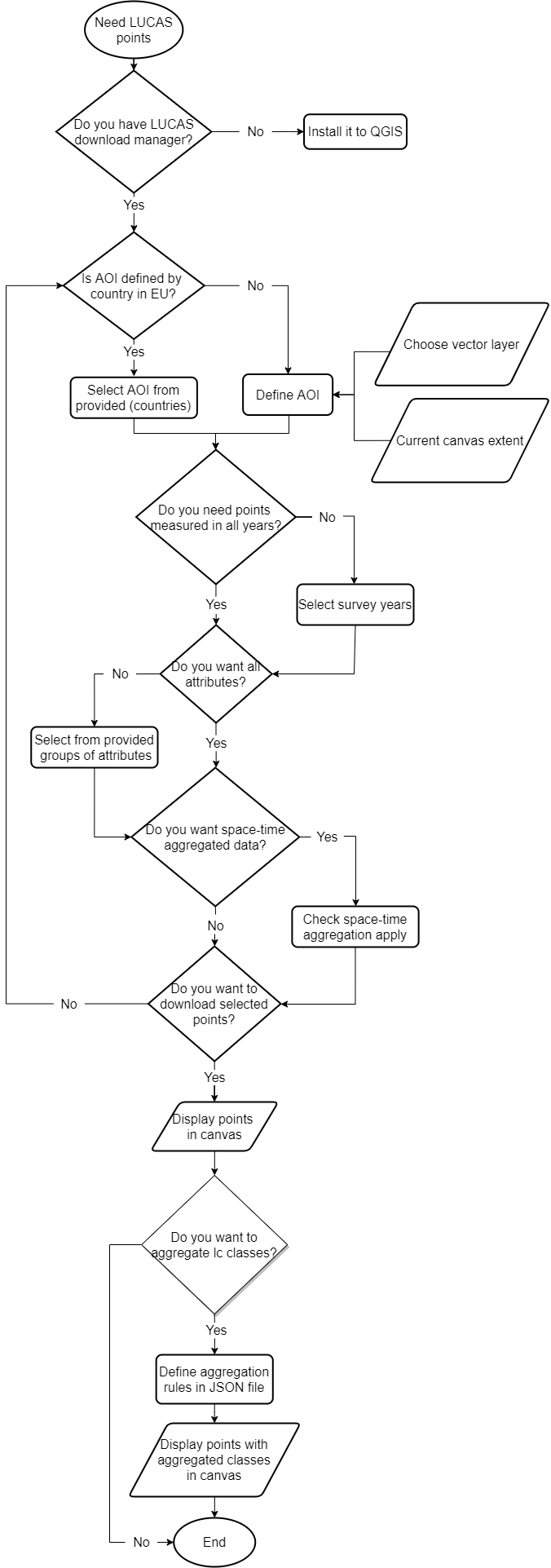
V rámci projektu GeoHarmonizer byl vytvořen soubor umožňující přehlednou demonstraci funkcionality lucas modulu v prostředí internetu. Tvorbou tohoto souboru se zabýval především Ing. Martin Landa, Ph.D. Pomocí Jupyter Notebook je možné nejen psát zdrojový kód a spouštět ho, ale také vytvářet uspořádanou strukturu s komentáři jednotlivých částí. Výsledný dokument obsahuje ukázky mnoha možností, které dává uživateli modul lucas popsaný v předchozích kapitolách 5.1, 5.2 a 5.3.

# Zásuvný modul

Kostra zásuvného modulu byla vytvořena pomocí QGIS pluginu Plugin Builder dostupného v oficiálním QGIS plugin repozitáři. Tento plugin vygeneroval po zadaní základních informací o zásuvném modulu jeho kostru. Mezi tyto informace patří například název modulu, jeho umístění, stručný popis funkčnosti a další. Plugin Builder Obsah obrázku text

Popis byl vytvořen automatickyje zobrazen na následujícím obrázku (Obr.:).

Obr. 2: Plugin Builder



Obr. 3: Diagram popisující funkcionalitu zásuvného modulu

## Funkcionalita

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

Zásuvný modul 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.

### Download

Tato část 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, jsou staženy body nacházející se v území, které aktuálně zobrazuje mapové okno. Podmínkou pro tuto eventualitu je, že mapové okno musí být v souřadnicovém systému označovaném EPSG kódem 3035 (LAEA Europe).
* country – Po označení této možnosti je nutné zvolit v „combo boxu“ 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 možnost umožňuje zvolit jednu z aktivních vektorových vrstev a získat body uvnitř území pokrytého touto vrstvou. Omezením pro tento případ je možnost využití pouze polygonové vrstvy s omezeným počtem vrcholů.

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 pomocí „check boxů“ má možnost označit roky, ve kterých byly body měřeny.

V průběhu pěti měření, která byla doposud provedena, 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 sloupečků a byla by značně nepřehledná. Ve většině případů zajímá uživatele pouze malé množství vlastností a ostatní tudíž překážejí. Z tohoto důvody byly atributy rozčleněny do pěti skupin. Tyto skupiny jsou zobrazeny v následující tabulce spolu s počtem atributů, které obsahují.

|  |  |
| --- | --- |
| **Skupina atributů** | **Počet atributů** |
| land cover a land use | 40 |
| land cover a land use a soil | 55 |
| forestry | 2 |
| copernicus | 16 |
| inspire | 8 |

Tab. 9: Třídy atributů

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 tomto 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 v určitém období, je výhodnější, pokud každý prvek představuje jedno měření. Atributová tabulka obsahuje méně nepotřebných 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 „Downoad“ spustit stahování. Po úspěšném stažení dat se body načtou do mapového okna a plugin zároveň informuje uživatele o jejich počtu.

### Class aggregation

Obsah obrázku text

Popis byl vytvořen automatickyDruhá část zásuvného modulu „Class aggregation“ umožňuje uživateli definovat pravidla, na základě kterých mají být agregovány land cover třídy. Pravidla 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.: ) je příklad *json* souboru, který popisuje převod land cover tříd třetí úrovně na land cover třídy druhé úrovně.

Obr. : 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 je vypsán její obsah. Forma zobrazení je zde velmi přehledná a umožňuje snadnou kontrolu předem formulovaných 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 vznikne pět nových atributů, které odpovídají rokům 2006, 2009, 2012, 2015 a 2018. V případě neagregovaných dat, u kterých každý záznam v atributové tabulce představuje unikátní měření, se vytvoří pouze jeden nový atribut. Do nových sloupečků atributové tabulky se zapíšou nově definované hodnoty odpovídající klíčům v *json* souboru. Zápis je proveden pomocí SQL příkazu.

# Závěr

Cílem této diplomové práce byla harmonizace a validace datové sady LUCAS a její následné zpřístupnění pomocí QGIS zásuvného modulu. LUCAS měření provedená v letech 2006, 2009, 2012, 2015 a 2018 měla být sjednocena, k čemuž měla být využita data z roku 2018 jako referenční. Po úpravě dat mělo dojít k jejich publikaci na GeoServeru pomocí WFS. Dále měl být pomocí jazyka Python vytvořen modul lucas sloužící především k vygenerování dotazu pro WFS a stažení dat. Zmíněný modul měl využívat zásuvný modul do programu QGIS, vytvářející uživatelsky přívětivé prostředí pro přístup k bodům LUCAS. Zásuvný modul měl poskytovat podrobnou analýzu stažených dat.

# Použité zdroje

[1] About Geo-harmonizer. *Open Data Science Europe* [online]. [cit. 2021-03-28]. Dostupné z: <https://opendatascience.eu/geoharmonizer-project/>

[2] 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

[3] 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

[4] *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>

[5] *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>

[6] BUCK, O., HAUB, C., WODITSCH, S., LINDEMANN, M. D., KLEINWILLINGHÖFER, L., HAZEU, G. W., KOSZTRA, B., KLEESCHULTE, S., ARNOLD, S., & HÖLZL, M. (2015). Analysis of the LUCAS nomenclature and proposal for adaptation of the nomenclature in view of its use by the Copernicus land monitoring services. EEA/EFTAS. Contract No 3436/B2015/RO-COPERNICUS/EEA.56195. . EEA - European Environment Agency. <https://land.copernicus.eu/user-corner/technical-library/LUCAS_Copernicus_Report_v2-2.pdf>

[7] Visual Style Guide. *QGIS* [online]. [cit. 2021-5-1]. Dostupné z: <https://www.qgis.org/en/site/getinvolved/styleguide.html>

[8] *Obsah - Školení QGIS pro začátečníky* [online]. [cit. 2021-5-1]. Dostupné z: https://training.gismentors.eu/qgis-zacatecnik/index.html

[9] The Python Logo. *Python Software Foundation* [online]. [cit. 2021-5-1]. Dostupné z: https://www.python.org/community/logos/

[10] 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>

[11] *PostgreSQL - Wikipedie* [online]. [cit. 2021-5-2]. Dostupné z: https://cs.wikipedia.org/wiki/PostgreSQL

[12] PostgreSQL: About. *PostgreSQL* [online]. [cit. 2021-5-2]. Dostupné z: https://www.postgresql.org/about/

[13] *GitHub - geoserver: Official GeoServer repository* [online]. [cit. 2021-5-2]. Dostupné z: https://github.com/geoserver/geoserver

[14] About - GeoServer. *GeoServer* [online]. [cit. 2021-5-2]. Dostupné z: http://geoserver.org/about/

# Seznam obrázků

[Obr. 1: Rozložení LUCAS bodů měřených v roce 2018 12](file:///D:\Skola_zaloha\DP\Text\DP.docx#_Toc70855439)

[Obr. 2: Ukázka dotaníku LUCAS [4] 13](file:///D:\Skola_zaloha\DP\Text\DP.docx#_Toc70855440)

[Obr. 3: Land cover třídy [2] 15](file:///D:\Skola_zaloha\DP\Text\DP.docx#_Toc70855441)

[Obr. 4: Land use třídy [2] 16](file:///D:\Skola_zaloha\DP\Text\DP.docx#_Toc70855442)

[Obr. 5: Ukázka souboru obsahujícího SQL příkazy pro změnu hodnot atributů 20](file:///D:\Skola_zaloha\DP\Text\DP.docx#_Toc70855443)

[Obr. 6: Logo QGIS [7] 28](file:///D:\Skola_zaloha\DP\Text\DP.docx#_Toc70855444)

[Obr. 7: Logo Python [9] 29](file:///D:\Skola_zaloha\DP\Text\DP.docx#_Toc70855445)

[Obr. 8: Logo PostgreSQL [11] 29](file:///D:\Skola_zaloha\DP\Text\DP.docx#_Toc70855446)

[Obr. 9: Logo GeoServer [13] 30](file:///D:\Skola_zaloha\DP\Text\DP.docx#_Toc70855447)

[Obr. 10: Diagram popisující průběh prací 31](file:///D:\Skola_zaloha\DP\Text\DP.docx#_Toc70855448)

[Obr. 11: kombinace prostorového a atributového dotazu (xml) 33](#_Toc70855449)

[Obr. 12: Plugin Builder 36](file:///D:\Skola_zaloha\DP\Text\DP.docx#_Toc70855450)

[Obr. 13: Diagram popisující funkcionalitu zásuvného modulu 37](file:///D:\Skola_zaloha\DP\Text\DP.docx#_Toc70855451)

[Obr. 14: Soubor json definující agregaci tříd 40](file:///D:\Skola_zaloha\DP\Text\DP.docx#_Toc70855452)

# Seznam tabulek

[Tab. 1: Rozsah LUCAS 12](#_Toc70855483)

[Tab. 2: Počet LUCAS atributů 17](#_Toc70855484)

[Tab. 3: Změny názvů atributů 19](#_Toc70855485)

[Tab. 4: Změna hodnot atributu lc\_lu\_special\_remark 20](#_Toc70855486)

[Tab. 5: Hodnoty atributu lc1\_perc 22](#_Toc70855487)

[Tab. 6: Počet bodů atributu wm\_delivery v letech 23](#_Toc70855488)

[Tab. 7: Změny datových typů atributů 25](#_Toc70855489)

[Tab. 8: Tabulka hodnot atributu grazing 26](#_Toc70855490)

[Tab. 9: Třídy atributů 39](#_Toc70855491)

# Seznam zkratek

LUCAS Land Use and Coverage Area frame Survey

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

JSON JavaScript Object Notation

LAEA Lambert azimuthal equal-area projection

LC Land Cover

LU Land Use

OGC Open Geospatial Consortium

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

**Nenalezena položka seznamu obrázků.**

**Digitální přílohy**