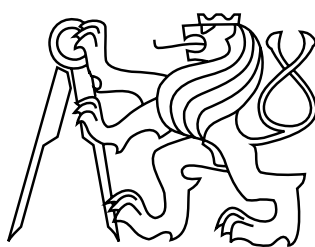


ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
FAKULTA STAVEBNÍ
OBOR GEODÉZIE, KARTOGRAFIE A GEOINFORMATIKA



BAKALÁŘSKÁ PRÁCE
ZAČLEŇOVÁNÍ GEOGRAFICKÝCH DATOVÝCH SAD
DO OPENSTREETMAP

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

červen 2016

Martin JÁKL



ZADÁNÍ BAKALÁŘSKÉ PRÁCE

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: Jákl	Jméno: Martin	Osobní číslo: 424013
Zadávající katedra: Katedra geomatiky		
Studijní program: Geodézie a kartografie		
Studijní obor: Geodézie, kartografie a geoinformatika		

II. ÚDAJE K BAKALÁŘSKÉ PRÁCI

Název bakalářské práce: Začleňování geografických datových sad do OpenStreetMap	
Název bakalářské práce anglicky: Integration of geographic datasets into OpenStreetMap	
Pokyny pro vypracování: Bakalářská práce se věnuje rešerši problematiky začleňování geografických datových sad kompatibilních s licenci Open Database License (ODbL) do databáze projektu OpenStreetMap (OSM). Cílem práce je vytvoření metodiky dávkových importů datových sad publikovaných jako tzv. "opendata" státními institucemi a samosprávou v ČR do OSM. Praktická část je zaměřena na problematiku začlenění dat IPR (Institut plánování a rozvoje hl. m. Prahy) a dalších relevantních zdrojů dat do projektu OSM.	
Seznam doporučené literatury: Bennett, J.: OpenStreetMap, Packt Publishing, Limited, 2010, ISBN: 978184719751 Ramm, F.: OpenStreetMap: Using and Enhancing the Free Map, UIT Cambridge, 2010, ISBN: 9781906860110 Pilgrim, M.: Dive Into Python, Createspace Independent Pub 2009, ISBN: 9781441413024	
Jméno vedoucího bakalářské práce: Ing. Martin Landa, Ph.D.	
Datum zadání bakalářské práce: 22.2.2016	Termín odevzdání bakalářské práce: 22.5.2016
Podpis vedoucího práce	Podpis vedoucího katedry

III. PŘEVZETÍ ZADÁNÍ

<i>Beru na vědomí, že jsem povinen vypracovat bakalářskou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je nutné uvést v bakalářské 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“.</i>	
Datum převzetí zadání	Podpis studenta(ky)

ABSTRAKT

Tato bakalářská práce se zabývá problematikou datových importů do projektu OpenStreetMap. Nejprve vysvětluje problematiku licencování dat OSM a nastiňuje myšlenku Opendata. Dále se zabývá možností začlenění datových sad z IPR Praha. Rozebírá nekompatibilitu licencí ODbL a CC BY-SA. V praktické části se zabývá vytvořením programu pro stahování dat z IPR Praha a importem do PostGIS databáze. Následně se zabývá rešerší a vlastním návrhem vhodných dat pro začlenění do projektu OpenStreetMap.

KLÍČOVÁ SLOVA

OpenStreetMap, import, IPR, Python, Opendata

ABSTRACT

This bachelor thesis deals with problems of data imports OpenStreetMap. First explains issues of licensing OSM data and outlines the idea of OpenData. It also deals with the possibility to integrate data sets from IPR Prague. It analyzes the incompatibility of licenses ODbL and CC BY-SA. The practical part deals with the creation of a program for downloading data from IPR Prague and importing them into PostGIS database. Subsequently deals with the researches and own draft of appropriate data for inclusion in OpenStreetMap.

KEYWORDS

OpenStreetMap, import, IPR, Python, Opendata

PROHLÁŠENÍ

Prohlašuji, že bakalářskou práci na téma „Začleňování geografických datových sad do OpenStreetMap“ jsem vypracoval samostatně. Použitou literaturu a podkladové materiály uvádím v seznamu zdrojů.

V Praze dne

(podpis autora)

PODĚKOVÁNÍ

Chtěl bych poděkovat vedoucímu mé bakalářské práce, Ing. Martinu Landovi, Ph.D., za odborné rady a pomoc při zpracování této práce. Dále bych chtěl poděkovat své rodině za projevenou podporu a trpělivost.

Obsah

1	Úvod	9
2	Teorie	10
2.1	OpenStreetMap	10
2.1.1	Vznik	10
2.1.2	Struktura dat	10
2.1.3	Změna licence	12
2.1.4	Licence ODbL	13
2.1.5	Zdroje dat	14
2.1.6	Importy	14
2.1.7	Talk-cz	15
2.1.8	Vykreslovače	15
2.2	Otevřená data	16
2.3	IPR Praha	18
2.3.1	Licence CC BY-SA 4.0	18
3	Použité technologie	19
3.1	PostgreSQL	19
3.1.1	PostGIS	19
3.2	Python	19
3.2.1	Knihovna argparse	20
3.2.2	Knihovna xmldict	20
3.2.3	Knihovna GDAL	20
3.3	QGIS	21
4	Praktická část	22
4.1	Geodatabáze PostGIS	22
4.2	Data IPR	22
4.2.1	Licenční problém	25
4.3	IprDownloader	26
4.3.1	IprDownloader	26
4.3.2	IprBase	27

4.3.3	IprPg	27
4.4	Ovládání	27
4.5	Navržená data pro import	28
4.5.1	Výstupy PID	29
4.5.2	Odpadní zařízení pro občany	30
4.5.3	Veřejné toalety	31
4.5.4	Parkovací automaty	32
4.5.5	Záchytná parkoviště P+R	32
4.5.6	Budovy 3D - výšky budov	33
5	Závěr	37
	Seznam zkratk	38
	Literatura	39
	Seznam příloh	41
A	Typy vykreslení OSM dat	42
B	Diagramy	45
C	Obsah CD	47

Seznam obrázků

2.1	Rozdělení prvků OSM, vlastní tvorba (zdroj: vlastní)	12
2.2	Příklad umístění licence (zdroj: http://www.openstreetmap.org/copyright)	13
2.3	Stupně otevřenosti dat, převzato z http://5stardata.info . . .	17
A.1	Standardní mapa (Standard)	42
A.2	Cyklistická mapa (Cycle Map)	42
A.3	Transport Map	43
A.4	MapQuest Open	43
A.5	Humanitární mapa (Humanitarian)	44
A.6	Projekt F4 (zdroj http://demo.f4map.com/)	44
B.1	Diagram postupu práce.	45
B.2	Diagram postupu práce pro získání výšky budov.	46

1 Úvod

K projektu OpenStreetMap (OSM) jsem se dostal již před několika lety. Zaujala mě možnost sám tvořit a upravovat „mapu“. Přidávat nové informace ze svého okolí vlastním sběrem dat a poté využívat takto vytvořenou mapu.

Postupem času jsem zjistil, co je, a co již není vhodné vkládat do mapy. Když mi bylo na Fakultě stavební ČVUT v Praze nabídnuto vypracovat bakalářskou práci na téma, které se dotýká problematiky datových importů OSM, neváhal jsem. Nedávno totiž Institut plánování a rozvoje hlavního města Prahy (IPR) uveřejnil všechna svá geografická data. Tím vznikla možnost začlenit další vhodná data do projektu OSM.

Obsahem této práce je čtenáře nejprve seznámit s projektem OpenStreetMap. Přiblížit mu vznik a vývoj projektu, užití a jeho licenci. Dále přiblížit problematiku datových importů do OSM a vysvětlit pojem otevřená data (Open Data).

V praktické části navrhnout aplikaci v programovacím jazyce Python, která by data poskytovaná IPR umožnila dávkově stáhnout a následně importovat do pracovní geodatabáze PostGIS a popsat ovládání vytvořené aplikace.

Následně provést rešerši zveřejněných dat IPR a porovnat je s daty, která již jsou v OSM. Navrhnout, jaká data by byla vhodná importovat. V průběhu práce konzultovat tento záměr s komunitou OSM. Nakonec vybraná data připravit k importu z geodatabáze PostGIS do OSM.

2 Teorie

2.1 OpenStreetMap

2.1.1 Vznik

OpenStreetMap (OSM) je projekt, jež vznikl s cílem vytvoření a sběru volně dostupných geografických dat a následně jejich možné vizualizace do topografických map. Projekt založil Steve Coast v červenci roku 2004 v Anglii. Jako inspirace mu posloužil projekt Wikipedia.

Zprvu projekt využívalo jen pár nadšenců, postupem času ale získal projekt popularitu. S nárůstem počtu uživatelů se zvyšoval i objem dat. Bylo tedy nutné navýšit kapacitu a zabezpečení serverů. Také bylo potřeba zlepšit síťové řešení a infrastrukturu. Tím je myšleno, že data byla rozdělena na více serverů.

V dubnu roku 2006 byla založena nadace OpenStreetMap Foundation pro financování projektu jako takového (zaměstnanců, běhu, zajištění bezpečnosti serverů atd.). [19]

2.1.2 Struktura dat

OSM data jsou nyní uložena v objektově-relační databázi PostgreSQL. [15]

Pro samotnou výměnu dat ale slouží souborový formát *.osm*, který je postaven na značkovacím jazyku XML. Jeho výhodou je pevně daná struktura a snadná orientace v kódu pro člověka. Nevýhodou je ovšem větší objem dat, který lze ale snížit kompresí.

Prvky (elementy) jsou v datovém modelu OSM rozděleny na:

- uzel (node)
- cesta (way)
 - neuzavřená
 - uzavřená
 - plocha (area)
- relace (relation)

Uzel (node)

Bodovým prvkem je v datovém modelu OSM tzv. uzel (node). Je definován jedinečným identifikátorem (**id**). Dále je ukládána verze uzlu, časový údaj, kdy byl do databáze přidán a v jaké změně k tomu došlo (**changeset**). Každý uzel má svoje souřadnice definovány v souřadnicovém systému WGS 84 (EPSG 4326). Dále je možné připojit různé atributy (tzv. tagy) tj. klíče s hodnotou (**tag k= v=**).

Příklad uzlu ve formátu *osm*:

```
<node id="2905214905" visible="true" version="1" changeset="22804106"
  timestamp="2014-06-08T06:57:20Z" user="Salamandr"
  uid="1708065" lat="50.1036981" lon="14.3897278">
  <tag k="natural" v="tree"/>
  <tag k="source" v="bing:ortofoto"/>
</node>
```

Cesta (way)

Označením pro liniový prvek je v datovém modelu OSM tzv. cesta (way). Každá cesta má svůj jedinečný identifikátor (**id**) a základní atributy (**visible**, **version**, ...), stejně jako je tomu u bodového prvku uzel. Je vždy určena dvěma nebo více uzly, ze kterých se skládá. Uvedena je pouze reference na id uzlů, nikoliv jejich souřadnice. Cestám lze taktéž přiřadit různé atributy.

Cesty se dále rozdělují na neuzavřené a uzavřené. Uzavřené cesty mohou tvořit plošné objekty (area). Těmto plošným prvkům lze přiřadit speciální atributy jako je například les (**landuse=forest**).

Příklad liniového prvku ve formátu *osm*:

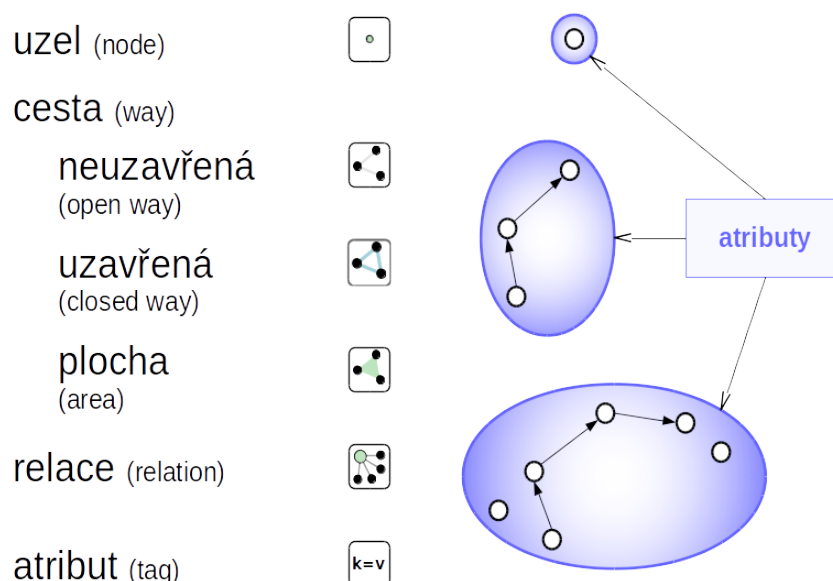
```
<way id="87249754" visible="true" version="2" changeset="34489106"
  timestamp="2015-10-07T11:52:41Z" user="Petr_Dlouhý" uid="17615">
  <nd ref="1014526199"/>
  <nd ref="1014525941"/>
  <nd ref="1014526337"/>
  <nd ref="1014526022"/>
  <nd ref="1014526277"/>
  <nd ref="1014525984"/>
  <tag k="highway" v="path"/>
  <tag k="source" v="bing:ortofoto"/>
</way>
```

Relace (relation)

Speciálním prvkem je relace (relation), do které lze zahrnout jeden a více prvků. Mohou se do nich spojit prvky stejného druhu (uzel a uzel), nebo různého (uzly,

cesty nebo i relace). Například dálnice je tvořena několika cestami (way), které jsou zahrnuty do společné relace, např. „Dálnice D1“. Nebo například turistická trasa může být relace sdružující jak liniové (cesty, pěšiny), tak i bodové (rozcestníky, vyhlídky, apod.) prvky.

Na wiki OpenStreetMap [11] je uvedeno, které atributy jsou pro daný druhu prvku vhodné (dle komunity uživatelů OSM). Pokud je daný atribut použit nevhodně, tak mohou být související prvky v některých vykreslovačích chybně zobrazeny, viz kap. 2.1.8.



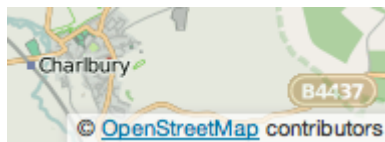
Obrázek 2.1: Rozdělení prvků OSM, vlastní tvorba (zdroj: vlastní)

2.1.3 Změna licence

Původně byla data OSM a generované grafické dlaždice distribuovány pod licencí Creative Commons AttributionShare Alike 2.0 (CC BY-SA 2.0).

Tato licence umožňovala užití (distribuci ale i editaci) díla (dat) pod podmínkou, že bude uveden zdroj OpenStreetMap.org ve viditelné části vytvořených mapových dlaždic [12].

V roce 2012 byla licence publikovaných dat změněna na Open Data Commons Open Database Licence (ODbL). Důvodem byla lepší licenční ochrana dat v datábázi. Původní licence (CC BY-SA) chránila data pouze pomocí autorského zákona. Obě licence jsou sice postaveny na principu „Uveďte autora“ a „Zachovejte



Obrázek 2.2: Příklad umístění licence (zdroj: <http://www.openstreetmap.org/copyright>)

licenci“. Rozdíl mezi nimi je ten, že pokud někdo pod licencí CC BY-SA vytvoří mapu (nebo jiné dílo), musí zachovat stejnou licenci. Nemusí však uvolnit přímo data, která pro tvorbu mapy použil.

Data OSM pod novou (ODbL) licencí, je možné použít nebo upravit, a následně je zveřejnit pod jakoukoli licencí za podmínky, že dojde k uvolnění souvisejících doplňků a vylepšení. [13].

Změna licence přinesla problém s daty, které byly poskytnuty projektu za podmínek předchozí licence (CC BY-SA 2.0). Bylo nutné se dotázat každého z dřívějších přispěvatelů dat, ať už právnických osob, tak i fyzických osob, zdali souhlasí s tím, aby jejich data byla distribuována pod novou licencí (ODbL). U přispěvatelů, kteří k tomuto nesvolili nebo se nevyjádřili, bylo nutné jejich příspěvky z databáze OSM vymazat. Tato situace nastala pouze ve zlomku případů (méně než 1 %). Nejvíce tato změna licencí ohrozila data v zemích jako je Polsko a Nový Zéland. [14]

V současnosti jsou tedy data OSM distribuována pod licencí ODbL a generované grafické dlaždice pod licencí CC BY-SA 4.0. [12]

2.1.4 Licence ODbL

Licence ODbL se zabývá právní ochranou databází, včetně veškerých autorských a příbuzných práv k databázi.

Licence ODbL umožňuje:

- kopírovat, distribuovat a užívat data
- vytvářet nová data z původních
- měnit původní data

Při použití databáze s licencí ODbL, je nutné připojit kopii přesného znění této licence. Při nakládání s databází ponechat nedotčena veškerá autorská práva

a uvést jejich zdroj (nebo autora). Veškeré odvozené databáze musí splňovat podmínky ODbL, tj. buď přímo licence ODbL nebo licence s ní kompatibilní. [7]

2.1.5 Zdroje dat

Snahou je, aby mapová data tvořili jedinci prvotním sběrem dat, tj. měřením vlastními GNSS (GPS, Glonass) přijímači a znalostí místních poměrů (uzavřené silnice, stezky atd.). Takto vzniklé dílo volně užívat k vlastním potřebám. Komunita přispěvatelů se zprvu pomalu, později poměrně rychle rozrostla a dnes čítá 3,7 milionů registrovaných uživatelů s alespoň jednou vytvořenou změnou v OSM. Z toho je 2,7 milionů účtů aktivních přispěvatelů.[16]

Takto vzniklé mapové podklady byly vhodné i pro další projekt, dnes již velice rozšířený a známý jako GeoCashing (GC). Projekt GC začal užívat mapové podklady OSM a zároveň jeho uživatelé začali sami přispívat do OSM.

Přispěvatelé dat do OSM musí respektovat licenci projektu ODbL. Tudíž i jejich zdroj dat musí splňovat podmínky této licence. Proto by měli všechny svoje změny, které v OSM vytvoří, řádně ozdrojovat atributem s klíčem `source`. V případě vlastního sběru dat se vyplňuje hodnotou `source=survey`, popřípadě uvést zdroj, odkud čerpali. Pokud tuto povinnost poruší a použijí zdroj, jež není kompatibilní s licenční politikou OSM, tak ostatní přispěvatelé do OSM tuto změnu, aby předešli sporům, odstraní. V tomto případě dojde k odstranění celé sady změn, byť by v ní byl pouze jeden prvek, jenž licenční podmínky poruší.

Druhým významným zdrojem dat jsou soukromé subjekty (společnosti). Většinou jde o podkladové zdroje dat, například ortofoto, které jsou vhodné pro obkreslování silničních sítí z leteckých nebo satelitních snímků a pod. Významným zdrojem těchto dat jsou mapy vyhledávače Bing od společnosti Microsoft. Ta svolila použít jejich letecké snímky většiny obydlené pevniny¹.

Třetím zdrojem dat a zároveň postupně dominujícím, co do jeho obsahu, jsou datové sady ze státního sektoru. Tyto datové sady jsou vzhledem ke své kvalitě a konzistenci velmi žádoucím zdrojem dat pro OSM.

2.1.6 Importy

Výraz import v tomto případě znamená začlenění většího množství dat z datové sady nebo datových sad z jednoho datového úložiště do jiného úložiště (např.

¹<http://wiki.openstreetmap.org/wiki/Bing>

databázových serverů). Při velkých objemech dat se využívá automatizace pomocí skriptů nebo programů, které samotný proces importu provedou.

Datové importy z veřejných databází do databáze OSM jsou velmi cenné. Otevřené geografické, ale i jiné, datové zdroje státních či veřejných institucí financovaných státem jsou často komplexní. Obsahují celistvý soubor dat, který daná instituce vyžaduje ke svému chodu. Jistá nevýhoda tu ale může být, data nemusí být vždy úplně aktuální. Některá data mohou být totiž zveřejněna i s větším časovým odstupem.

V rámci České republiky proběhlo již několik hromadných datových importů do OSM. Jak již bylo řečeno, časově nejnáročnější je samotná příprava importu. V případě importu dat do OSM, nejen napsání skriptu, ale i nutná diskuze tohoto záměru na diskuzní konferenci Talk-cz.

2.1.7 Talk-cz

Tato diskuze probíhá posíláním emailových zpráv do společné konference (diskuzního fóra). Uživatelům chodí emaily z probíhající diskuze. Pokud na nějaký chtějí reagovat, tak pošlou zprávu na emailovou adresu konference. Musí ale do předmětu zprávy (emailu) napsat Re: a předmět zprávy, na kterou reagují. Server tyto zprávy řadí pomocí předmětu a času. Diskuze je poté k dispozici na webových stránkách.²

2.1.8 Vykreslovače

Na hlavní stránce OSM (<http://www.openstreetmap.org>) je k dispozici mapová aplikace. Ta nabízí pět „základních“ přednastavených vrstev vykreslených z dat OSM. Pro vykreslení dat do grafických dlaždic se používá takzvané pseudo Mercatorovo zobrazení nebo Web Mercator (EPSG 3857). Toto zobrazení je vhodné pro celý povrch Země (až na velké zkreslení u pólů). První, kdo zobrazení Web Mercator použil, byla společnost Google pro svoje mapy Google Maps.[1]

- Standardní vrstva - vykresluje všechny prvky kartograficky „přiměřeně“ (viz obr. A.1/s42).
- Cyklomapa - zdůrazňuje cyklostezky, výškopis (viz obr. A.2/s42).

²<http://lists.openstreetmap.org/listinfo/talk-cz>

- Dopravní mapa - zdůrazňuje silniční a železniční síť (viz obr. A.3/s43).
- MapQuest Open - podkladová mapa pro potřeby GeoCaching (viz obr. A.4/s43).
- Humanitární mapa - zdůrazňuje důležité veřejné služby a potlačuje ostatní prvky (viz obr. A.5/s44).

Existují další projekty, jež kombinují data OSM a jiných zdrojů, s cílem tvorby tematických map. Například mapu turistických a cyklistických tras vykresluje pro celou Evropu projekt *mtbmap*.³

Dalšími zajímavými projekty jsou například ty, které k 2D polohopisných datům přidávají „třetí“ rozměr a vytvářejí tzv. 2.5D mapu. Většinou jde o 3D zobrazení budov, mostů (dle atributů), popřípadě i stromů (viz obr. A.6/s44).

2.2 Otevřená data

Myšlenka otevřených dat (open data) vznikla v USA. Pokud vzniknou jakákoli data z veřejných peněz, měla by tedy být veřejně přístupná. Některé studie uvádí, že tento jev měl kladný efekt na tamní ekonomiku.^{4 5}

Proto jako první hlavní zdroj pro OSM byly satelitní snímky povrchu Země (Landsat) a digitální model terénu (projekt SRTM) s rozlišením 30x30 m od NASA (pro pozdější vykreslení vrstevnic).

Trend otevírání dat se začal rozšiřovat zprvu do zemí západní Evropy (Velké Británie, Francie či Německa), ale i zemí mimo Evropu.[10]

V České republice se tomuto tématu věnuje Fond Otakara Motejla. Jedním z jeho hlavních projektů jsou právě Otevřená data. Fond spolupracuje s nadací Society Fund Praha a Rekonstrukce Státu. V rámci těchto uskupení je vyvíjen tlak na transparentnost veřejné správy, zveřejňování smluv a dat státních institucí, jelikož jejich získání a údržba byla placena z veřejných zdrojů.

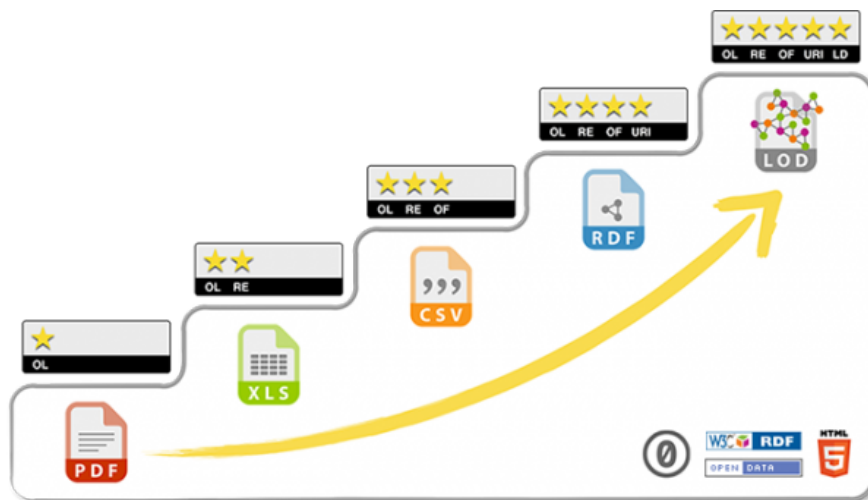
³mtbmap.cz

⁴<http://www.worldbank.org/content/dam/Worldbank/document/>

[Open-Data-for-Economic-Growth.pdf](#)

⁵<http://www.otevrenadata.cz/res/data/001/003498.pdf>

Otevřená data lze dělit na 5 stupňů, dle otevřenosti.



Obrázek 2.3: Stupně otevřenosti dat, převzato z <http://5stardata.info>

- Data jsou dostupná v WWW stránek s jasnými licenčními podmínkami. Není zde žádný požadavek na datový formát, a proto není brán jako dostatečný způsob zveřejňování dat. Těmito daty ku příkladu jsou Webové Mapové Služby (WMS). Jsou veřejně k dispozici, avšak nezveřejňují přímo data, ale jen obrázky z nich generované.
- Zveřejněná data jsou již ve strojově čitelném formátu, který je veřejně dobře znám. Nejčastěji se jedná o charakter tabulky. Musí umožňovat přístup k jednotlivým řádkům a obsahu buněk.
- Třetí stupeň navíc od druhého vyžaduje, aby k jeho čtení nebyl vyžadován speciální (uzavřený) software. Spadají sem tedy dokumenty formátů OpenOffice (Office Open XML, OpenDocument).
Pro prostorová data například formáty GML, KML nebo GeoPackage.
- U stupně otevřenosti 4 je v distribuované datové sadě povinnost zavést identifikaci entity ve tvaru Internationalized Resource Identifier (IRI).
- U nejvyššího možného stupně otevřenosti je vyžadováno, aby distribuce splňovala standardy propojených dat (Linked Data). [9]

Díky této iniciativě došlo v rámci České republiky ke zveřejnění dat státních úřadů s různým úspěchem a stupněm otevřenosti. Avšak tento trend zpřístupnění

a zveřejňování pozvolna pokračuje. Příkladem je Ministerstvo vnitra, Ministerstvo financí nebo Český úřad zeměměřický a katastrální (RÚIAN).

2.3 IPR Praha

IPR Praha je zkratka názvu pro Institut plánování a rozvoje hlavního města Prahy. Tento institut se věnuje urbanismu, architektuře a rozvoje města Prahy. Hlavním úkolem IPR je tvorba územního plánu hlavního města Prahy. Významným úkolem IPR je zajišťovat zpracování geografických informací. Od roku 2002 poskytuje na svých stránkách zdarma webové aplikace bez limitu využití. V rámci tzv. Pražského geoportálu došlo k jejich většímu využívání. Na základě platných Pravidel pro poskytování dat a výstupů z datových souborů a datového skladu Geografického byly ode dne 1. 4. 2015 zveřejněny datové soubory a další webové služby poskytované IPR. Data byla uveřejněna pod licencí CC BY-SA 4.0. [5]

2.3.1 Licence CC BY-SA 4.0

Licence Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) je zde uvedena ve zjednodušeném znění.

Uživatel smí:

- sdílet - rozmnožovat a distribuovat materiál prostřednictvím jakéhokoli média v jakémkoli formátu
- upravovat - remixovat, změnit a vyjít z původního díla

a to pro jakýkoliv účel, i komerční.

Poskytovatel licence nemůže odvolat tato oprávnění do té doby, dokud dodržujete licenční podmínky.

Uveďte původ — Je Vaší povinností uvést autorství, poskytnout s dílem odkaz na licenci a vyznačit Vámi provedené změny. Toho můžete docílit jakýmkoli rozumným způsobem, nicméně nikdy ne způsobem naznačujícím, že by poskytovatel licence schvaloval nebo podporoval Vás nebo Váš způsob užití díla.

Zachovejte licenci — Pokud budete toto dílo upravovat, pozměňovat nebo na něj navazovat, musíte svoje odvozená díla vystavovat pod stejnou licencí jako původní dílo.[3]

3 Použité technologie

3.1 PostgreSQL

PostgreSQL je open source objektově-relační databázový systém. Za dlouhou dobu vývoje se stal robustním a hojně využívaným řešením hlavně díky své spolehlivosti. Lze s ním pracovat ve všech známých operačních systémech. [17]

3.1.1 PostGIS

PostGIS je nadstavba databázového systému PostgreSQL. Rozšiřuje jej o funkce a datové typy, které umožňují uložení, manipulaci a správu geografických objektů. Mezi podporované geometrické typy patří bod (point), linie (linestring), nebo polygon (polygon). Dále je možno prvky sdružit do tzv. multiprvků (multiPoint, multiLineString nebo multiPolygon). Geometrie prvku určuje jeho polohu v daném souřadnicovém systému. Data je tedy možno transformovat do jiných PostGISem podporovaných souřadnicových systémů. Systém kromě základních operací jako výpočet délky, obvodu, plochy, podporuje i sofistikovanější funkce jako je např. výpočet obalové zóny a další.

Software PostGIS je šířen pod GNU General Public License (GPLv2).

3.2 Python

Programovací jazyk Python vznikl v roce 1991. Navrhl ho Guido van Rossum. Je vyvíjen jako open source projekt. Inspiroval se hlavně programovacím jazykem ABS, který byl přímo vytvořen pro výuku začátečníků v programování. Python je proto jeden z nejvýhodnějších programovacích jazyků pro začínající programátory, ale i přesto ho lze použít pro praktické programování.

Jedná se o jednoduchý programovací jazyk podporující různá programovací paradigmaty, hlavně objektové, imperativní, procedurální a v omezené míře i funkcionální. Je multiplatformní, lze jej tedy použít na různých operačních systémech. Klade velký důraz na syntaxi psaného kódu, využívá hlavně prázdné znaky v psaném kódu.

Programy nebo skripty lze psát libovolným textovým editorem. Vhodné je použít standardu PEP 8, který definuje správnou strukturu odsazení kódu.

V současné době je vydána stabilní verze 3.5.7, která oproti předchozí verzi 2.x mění některé již „zažité“ vlastnosti jazyka. Například nejviditelnější je změna funkce `print()`. Ve verzi 2.x stačilo napsat `print 'Ahoj'`. Ve verzi 3.x je již nutné řetězec vložit do závorek `print('Ahoj')`. Tyto dvě dnes používané verze 2.x a 3.x jsou navzájem nekompatibilní. [18] [20]

3.2.1 Knihovna argparse

Knihovna `argparse` je knihovna programovacího jazyka Python, která řeší konzolový vstup do aplikace a zpracovává ho pro další použití v programu.[2]

3.2.2 Knihovna xmltodict

Tato Python knihovna umí číst soubory ve formátu XML a zpracovat jejich obsah do datové struktury, která je jednodušší pro manipulaci v samotném programu. Data z XML zpracuje do slovníku (dictionary), kde klíčové slovo (key) reprezentuje hodnota (value) ve slovníku. Hodnotami mohou být dále i řady dalších klíčových slov (key).[21]

Příklad zpracování dat ve formátu XML.

```
<head>
  <many>ele1</many>
  <many>ele2</many>
</head>
```

Zpracování dat provedeme v Pythonu příkazem:

```
with open('example.xml') as fd:
    doc = xmltodict.parse(fd.read())
```

Objekt `doc` vypadá následovně:

```
OrderedDict([(u'head', OrderedDict([(u'many', [u'ele1', u'ele2'])]))]) }
```

Data v této struktuře lze již jednoduše procházet.

```
doc['head']['many'][0]
ele1
```

3.2.3 Knihovna GDAL

Jedná se o knihovnu pro práci s geografickými daty umožňující jejich čtení a zápis. Je napsána v programovacím jazyce C++. Je považována za jeden z hlavních open source projektů hojně využívaných ve sféře GIS. Podporuje velkou škálu

rastrových a vektorových formátů. Díky knihovně PROJ.4 podporuje celou škálu souřadnicových systémů a umožňuje transformaci dat mezi nimi. [4]

3.3 QGIS

QGIS je program vytvořený za účelem prohlížení a editace geografických dat. Jedná se o open source software. Je tedy možné, aby na jeho vývoji spolupracoval kdokoli. Vlastní uživatelské nástroje se píší ve formě tzv. zásuvných modulů, které je možné implementovat v jazyce C++ nebo Python. Tyto moduly si ostatní uživatelé mohou poté stáhnout a začlenit do své instalace QGISu. QGIS také umožňuje využívat jazyk Python v rámci vestavěné konzole. Lze tak vytvářet vlastní jednoduché skripty v jazyku Python a spouštět je přímo v QGISu.

Od svého vzniku byla vydána již spousta verzí tohoto programu. Dříve byly verze pojmenovávány podle měsíců planet Jupitera a Saturnu, později se začaly používat jména měst. V současnosti (duben 2016) je k dispozici verze 2.14 Essen.

4 Praktická část

4.1 Geodatabáze PostGIS

V rámci této bakalářské práce byla zřízena geodatabáze PostGIS na školním serveru *geo102*. Data z OSM jsou do školní databáze nahrávána pomocí programu *osm2pgsql*.⁶ Obsah databáze je každý den pravidelně aktualizován podle aktuálního stavu OSM. Tuto každodenní aktualizaci zajišťuje shellový skript, který je součástí přílohy.

Tabulky v databázi byly vytvořeny podle základního schématu. Pro každou třídu prvků (uzel, cesta a relace) je vytvořena samostatná tabulka. Pro uzly je vytvořena tabulka *czech_point*, pro cesty tabulka *czech_line* a pro plošné prvky (uzavřené cesty s atributy pro plochy) tabulka *czech_polygon*. Relace jsou řešeny tak, že pokud jsou tvořeny plochami, vytvoří z nich program *osm2pgsql* multipolygon a ten vloží do tabulky *czech_polygon*. Stejně jsou řešeny i relace tvořené cestami (tabulka *czech_line*), respektive body (*czech_point*). Ještě je dle schématu vytvořena tabulka *czech_roads*, která obsahuje všechny silnice v ČR.

Každý prvek má přiřazen, v rámci tabulky, jedinečné číslo *gid*. Dále je vytvořen pro každý prvek geometrie a pro každý atribut (dle schématu) stejnojmenný sloupec s hodnotou. Například bod s atributem *landuse=forest* má ve sloupci *landuse* hodnotu *forest*.

Takto vytvořená databáze bude použita pro analýzu dat IPR a jejich přípravu pro importu do OSM.

4.2 Data IPR

IPR na svém webu od jara roku 2015 zveřejňuje svá data v režimu otevřených dat.^{2.3} Vektorová data zveřejňuje ve formátech GeoJSON, DXF, GML a ESRI Shapefile. Dále je na výběr mezi dvěma souřadnicovými systémy S-JTSK (EPSG:5514) nebo WGS 84 (EPSG:4326).

Celkem IPR zveřejňuje (toho času) 96 datových setů, které jsou rozděleny do kategorií.⁷

⁶dostupné z <http://wiki.openstreetmap.org/wiki/Osm2pgsql>

⁷viz webové stránky IPR, www.geoportalpraha.cz/cs/opensdata

Pokud není řečeno jinak, jedná se o data ve vektorové formě. Těmito kategoriemi jsou:

- 3D model - Obsahuje 3D modely budov a mostů v Praze a digitální model povrchu (DMP), dále rastrově digitální model povrchu a terénu, absolutní a relativní výšky budov.

Z těchto dat by se pro OSM dala využít poloha budov ze souboru Budovy 3D. Budovy byly do OSM již importovány ze zdroje RÚIAN (Registr územní identifikace, adres a nemovitostí), není třeba je tedy importovat znovu.

- Digitální technická mapa Prahy - zde jsou k dispozici ve vektorové podobě všechny inženýrské sítě v Praze, technické budovy a parcelní hranice.

Do OSM se, možná zatím, nepřidávají inženýrské sítě. V současnosti lze ale do OSM přidávat zdroje veřejného osvětlení. Hranice města Prahy a jeho městských částí již v OSM jsou dostupná. Datovou sadu Technické budovy by bylo možné použít jako zdroj pro budoucí aktualizaci OSM. Průběhy hranic parcel lze již nyní získat z RÚIAN, který poskytuje ČÚZK.

- Doprava - obsahuje cyklistické trasy a značky, pěší trasy, parkovací zóny, automaty, P+R parkoviště a mapy zón PID. Cyklistické trasy v Praze lze do OSM importovat, jsou ale již velmi dobře zmapované samotnými uživateli. Datová sada obsahuje bodové značky, z těch by se některé mohly importovat. V případě datové sady Cyklistická doprava - značky, byla data prohledána. Dle IPR uváděné hodnoty 103 ve sloupci DRUH pro Bikesharing, bohužel žádné takové hodnoty neobsahovaly. Import dat z této datové sady nebyl proto dál brán v potaz. Bylo by možné do OSM přidat parkovací zóny (jako nový atribut u stávajících komunikací). Vhodná data jsou parkovací automaty a také P+R parkoviště.
- Geologie - obsahuje vektorové mapy radonového nebezpečí. Tato data se do OSM nepřidávají.
- Hluk - obsahuje hlukové mapy, a to ve dne a v noci. Oboje v rastrové formě. Tato data se do OSM nepřidávají. Dále obsahuje i protihlukové bariéry, které by bylo možné do OSM přidat.

- Kvalita životního prostředí - obsahuje datovou sadu Oblasti svozu komunálního odpadu. Tato data není vhodné do OSM začlenit. Dále ale obsahuje datovou sadu Odpadní Zařízení pro občany, která obsahuje informace o všech sběrných odpadu v Praze. Tato datová sada se jeví jako další vhodná pro import.
- Mapové podklady - obsahuje klady mapových listů různých měřítek (1:500 až 1:10 000). Klady mapových listů nejsou vhodná data pro OSM. Dále kategorie obsahuje i datové sady vrstevnic, a to po 5 m, 2 m a 1 m. Vrstevnice přímo OSM neobsahuje, avšak byly by velkým zpřesněním stávajícího zdroje vrstevnic. Nebo jako zdroj pro projekty, které kombinují OSM s jinými daty.⁸
- Občanská vybavenost obsahuje pouze datovou sadu Veřejné toalety. Jistě jsou již toalety v Praze částečně zmapovány, ale i tak by bylo vhodné je doplnit. Jsou tedy vybrána jako další vhodná data pro import.
- Ochrana přírody a krajiny, obsahuje data o ochranných pásmech památných stromů. Bohužel ne samotné stromy. Vhodná data pro import jsou z datové sady Přírodní parky a Významný krajinný prvek - registrovaný. Z těchto datových sad by se mohly doplnit, nebo aktualizovat data v OSM.
- Ortofoto obsahují letecké snímky (rastr) Prahy s rozlišením až 5 cm na pixel jak ve viditelném spektru světla, tak i v infračerveném.
Ortofoto je vhodné pouze jako podkladový zdroj, například pro mapovací aplikaci JOSM.
- Ovzduší obsahuje vektorové mapy znečištění ovzduší a také zdroje znečištění. Tato data se v OSM nemapují, jedině co by se mohlo přidat do OSM jsou bodové zdroje, a to například značkou pro komín. Dále jsou v této kategorii i bonity, a to z různých hledisek (ovzduší, půdy, osvit, atd.). Tato data nejsou vhodná pro import do OSM.
- Platný územní plán. V této kategorii jsou datové sady Veřejně prospěšných staveb (plošné, liniové a bodové). Datová sada VPS obsahuje také informace o P+R u stanic metra. Tyto informace pocházejí z územního plánu, a tedy nemusí být realizovány. Po prozkoumání a porovnání s datovou sadou Záchytná

⁸Projekt mtbmap.cz nebo <http://mapa.prahounakole.cz/>

parkoviště P+R, lze konstatovat, že jsou všechny záznamy obsaženy již v této datové sadě.

- Socioekonomická data obsahuje pouze vektorovou mapu ceny pozemků. Data o ceně pozemků se do OSM nepřidávají.
- Technická infrastruktura - vodní hospodářství, obsahuje záplavová území Q20, Q50 a Q100 a území zaplavené povodněmi v roce 2013. Informace o záplavových územích se do OSM nepřidávají. Dále jsou v této kategorii datové sady Protipovodňové ochrany, které obsahují údaje o všech dočasných protipovodňových zdech. Jelikož se jedná o dočasné překážky, a to ještě jen v době povodně, nemá smysl je do OSM přidávat.
- Urbanismus - Z této kategorie by se mohly pro OSM využít informace z datové sady Stavební uzávěry - dopravní. Bylo by ale nutné udržovat aktuálnost informací o plánovaných dopravních uzavírkách. Dále by bylo ještě možné přidat do OSM počet pater budov z datové sady Podlaživost.

4.2.1 Licenční problém

Jak již bylo řečeno výše, IPR svá data zveřejnil a stále zveřejňuje (v době tvorby této práce) pod licencí CC BY-SA 4.0, viz kapitola 2.3.1.

Jelikož jsou ale data OSM distribuována pod licencí ODbL (v1.0), která není kompatibilní s licencí CC BY-SA, tak neumožňují začlenit data z IPR.

Na některá data (informace), jako například místní názvy, ulice atd. se nemohou vztahovat autorská práva. Neboť nevyžadují žádnou kreativitu pro jejich „vytvoření“. Ovšem na soubor (databázi) těchto dat vytvořený podle daných kritérií již může být právní ochrana uplatněna. A právě proto vznikla licence ODbL, která se zaměřuje na ochranu databáze jako celku.

U dat chráněných autorským právem a danou licencí se vyžaduje jejich vzdání. V některých zemích, kde se nelze úplně vzdát autorských práv, se vyžaduje omezení autorských práv na nejnižší možnou míru. Autor souhlasí, že nebude vymáhat svoje autorská práva, po dobu, kdy jsou data uložena v databázi. [8]

IPR byl kontaktován, zdali by mohla být jeho data použita do OSM. Dle vyjádření vyplynulo, že IPR by neměl problém s použitím svých dat, jelikož jsou

zveřejněna pod myšlenkou Opendata. Byla navržena možnost udělit od IPR pro OSM licenční výjimku.

Vznikla by tu však situace, kdy by byla data IPR k dispozici pod licencí CC BY-SA a v OSM (sice modifikovaná) pod licencí ODbL. Tohoto dualismu by mohl využít někdo, komu by licence CC BY-SA neumožňovala jeho záměr. Mohl by si totiž data obstarat z databáze OSM, sice modifikovaná, ale již pod novou licencí (ODbL), která by mu jeho záměr umožnila.

Hlavní rozdíl je totiž v tom, že pod licencí CC BY-SA sice může data různě měnit a upravovat, ale poté když je zveřejňuje, musí zachovat licenci CC BY-SA. Nemusí ale uvolnit (zveřejnit) data, která k nim přidal. V případě licence ODbL se tato situace diametrálně liší. Uživatel opět smí data upravovat nebo k nim přidávat jiná data, ale poté může svoje dílo distribuovat pod jinou licencí, za podmínky, že uvolní veškerá data, která k tomu použil.

Vhodnější by tedy bylo, změnit licenci dat zveřejňovaných IPR na ODbL.

4.3 IprDownloader

Jak již bylo řečeno, tak IPR distribuje geografická data ve formě datových souborů stažitelných z webových stránek organizace. Kromě samotných datových souborů je k dispozici kanál ve formátu ATOM (<http://opendata.iprpraha.cz/feed.xml>). Pomocí něho se lze také dostat ke všem distribuovaným datům. Jelikož by bylo zdlouhavé stahovat všechny tyto soubory ručně, byl v rámci práce vytvořen program, který umožňuje stahování a také import dat primárně do geodatabáze PostGIS. Skript byl napsán v programovacím jazyce Python (3.2) a využívá knihovny GDAL.

Zdrojový kód programu je rozdělen do tří souborů: `IprDownloader.py`, `IprBase.py` a `IprPg.py`.

4.3.1 IprDownloader

Skript `IprDownloader.py` parsuje vstupní údaje a zpracovává je k dalšímu použití v programu, viz 3.2.1. Základními parametry programu jsou `--alike`, `--crs`, `--format`, `--outdir`, `--d`. Dalšími vstupními parametry mohou být údaje o databázi, kam se mají data importovat. Jde o `--dbname`, `--dbschema`, `--dbhost`, `--dbport`, `--dbuser`, `--dbpassword`.

4.3.2 IprBase

Skript `IprBase.py` definuje třídu `IprDownloader` se všemi potřebnými funkcemi. Většina funkcí slouží pro otevírání, čtení a parsování XML souborů, viz kapitola 3.2.2. Obsahuje funkci pro hledání, která prochází všechny záznamy v ATOM souboru a porovnává, zda název definovaný parametrem `alike` je obsažen v názvu souboru `dat`. Poté najde dle nastavení (`crc`, `format`) daný soubor a stáhne jej do definovaného adresáře.

4.3.3 IprPg

Skript `IprPg.py` definuje třídu `IprDownloaderPg`, která dědí všechny metody rodičovské třídy `IprDownloader` a definuje další funkce potřebné pro správný import stažených dat do databáze PostgreSQL. Přesněji stažený soubor zkontroluje, jestli není komprimován v archivu `*.zip`, pokud ano, tak vytvoří nový adresář, který má stejný název jako archiv. Do nově vytvořeného adresáře poté extrahuje všechna data z archivu.

4.4 Ovládání

Při spuštění vypíše program všechna nalezená data, která IPR dává k dispozici. Pro hledání a označení, která data se mají stáhnout, slouží parametr `--alike`. Pokud hledaný řetězec obsahuje mezeru, tak je nutné jej umístit do uvozovek. Program rozlišuje velká a malá písmena, háčky a čárky.

Příklad: `$ iprdownloader.py --alike 'Technická mapa'`

Dále jsou k dispozici nastavení.

- souřadnicový systém `--crs`. Podporován je S-JTSK a WGS-84. Kromě názvu lze použít i EPSG kódy (5514 a 4326). Přednastavený souřadnicový systém je S-JTSK.

- datový formát souborů `--format`. Mezi podporované vektorové formáty patří:

`json` JavaScript Object Notation

`dxg` AutoCAD DXF

`gml` Geography Markup Language

`shp` ESRI Shapefile

Výchozím formátem je `shp` .

Pokud uživatel chce stahovat rastrová data, je potřeba změnit předdefinovaný formát souboru na adekvátní rastrový formát např. `tif` atd.

- adresář na lokálním počítači `--outdir`. Je předdefinován adresář `data` ve složce, kde je uložen samotný program.

Pro stažení dat stačí přidat parametr `--download` .

Příklad:

```
$ iprdownloader.py --alike 'Technická mapa' --crs 4326 \
--format gmp --outdir IPR
```

V případě importu stažených souborů do databáze PostGIS je nutné zadat minimálně název databáze `--dbname`. To v případě, pokud je databáze PostGIS umístěna na lokálním počítači. Pokud je databáze umístěna na vzdáleném počítači, je nutné zadat jeho adresu `--dbhost` a port `--dbport`. Pokud je ještě vyžadován autorizovaný přístup, použije se přístupové jméno `--dbuser` a heslo `--dbpassword`. Dále je možné zvolit si schéma v databázi `--dbschema`. Pokud je definován název databáze, není již potřeba používat parametr `--download` pro stahování dat.

Příklad:

```
$ iprdownloader.py --alike 'Technická mapa' \
--dbname pgis_osm_jakl --dbschema IPR --dbhost geo102.fsv.cvut.cz \
--dbport 5432
```

4.5 Navržená data pro import

Po analýze všech dat byla vyhodnocena vhodná data pro začlenění do OSM.

- Výstupy PID, obsahuje vstupy/výstupy z metra.
- Odpadní zařízení pro občany, dle IPR obsahuje sběrný odpadu.
- Cyklistická doprava - značky, dle technické dokumentace by mělo obsahovat bodové značky stojanů Bikessharing.
- Veřejné toalety
- Parkovací automaty

- Záchytná parkoviště P+R

Vybraná data byla pomocí IprDownloader.py stažena a naimportována do školní databáze PostGIS. Tato školní databáze obsahovala i aktuální data OSM. V programu QGIS byla data pomocí SQL SELECTu filtrována, na jejich základě byly vytvořeny nové tabulky. Data byla filtrována za použití doporučených atributů na české OSM wiki. [11]

Vybraná data IPR byla prohledána a porovnána s aktuálním stavem v OSM. Byly navrženy vhodné atributy, které by se ze zdrojových dat daly naplnit. Vybraná data s navrženými atributy byla uveřejněna na Talk-cz.⁹

4.5.1 Výstupy PID

V datech od IPR bylo obsaženo celkem 353 záznamů. Jednalo se o výstupy z metra (dveře, schodiště a výtahy).

V OSM databázi byly v tabulce `osm.czech_points` nalezeny záznamy s hodnotou

```
railway = subway_entrance
```

Vytvořená tabulka z dat OSM obsahovala 236 záznamů. Po porovnání dat bylo zřejmé, že nejvíce jsou v OSM zmapovány výstupy v centru, kde jsou posuny v řádech decimetrů až metr. V okrajových částech města nejsou zmapovány všechny výstupy, nebo chybí výstupy pro celé stanice. Protože body vstupů do metra jsou většinou součástí linií, není vhodné body mazat a vytvářet nové. Bylo potřeba najít mezi stávajícími body v OSM a IPR adekvátní dvojice. To znamená, že se jedná o body, které reprezentují stejný objekt v realitě.

Ani editace stávajících elementů není příliš přínosná. Výstupy z metra jsou minimálně 3 metry široké a případné změny by byly v řádů cm až dm. Bylo by ale vhodné přidat výstupy z metra tam, kde nejsou v databázi OSM. Jedná se o stanice metra Zličín, Luka, Kačerov, Rajská zahrada a Černý most. Celkem se tedy jedná pouze o 26 bodů, které by bylo snadnější přidat ručně, protože výstupy jsou dostatečně zmapované, pouze občas chybí v uzlech atribut `railway = subway_entrance`.

⁹<https://lists.openstreetmap.org/pipermail/talk-cz/2016-April/013887.html>

4.5.2 Odpadní zařízení pro občany

Data od IPR obsahují seznam 36 sběrných dvorů s velmi detailním popsáním (příjem, otevírací doba, zřizovatel).

Z OSM tabulky `osm.czech_point` byly záznamy filtrovány podle hodnot

```
amenity = recycling
recycling_type = centre
```

Bylo nalezeno celkem 11 záznamů (bodů), z toho bylo 6 bodů (z IPR) ve vzdálenosti 100 metrů od bodů z OSM. Těchto 6 bodů bude vhodné pouze přesunout. Zbylé body je nutné v databázi OSM nově vytvořit.

Z dostupných dat z tabulky byly navrženy tyto atributy

```
amenity = recycling
recycling_type = centre
opening_hours =
fee =
operator =
source =
source:loc = IPR
```

Navržené atributy by se naplnily z tabulky dat od IPR. Otevírací doba by byla vložena do `opening_hours`. Hodnoty v tomto atributu by byly formátovány dle doporučeného zápisu.¹⁰ U atributu `fee` bylo navrženo `free of charge for the Prague citizens`, tedy zdarma pro občany Prahy. Hodnota `operator` by byla ze stejnojmenného sloupce z dat od IPR. Atribut `source` by byl odvozen ze sloupce POSKYT.

Dále byly navrženy atributy

```
recycling:paper = yes
recycling:plastic = yes
recycling:metal = yes
recycling:wood = yes
recycling: ... = yes
```

Podle toho co zde lze recyklovat, dle hodnoty ve sloupci `odpriprije`.

¹⁰http://wiki.openstreetmap.org/wiki/Key:opening_hours

4.5.3 Veřejné toalety

Datový soubor Veřejné toalety obsahuje jednu tabulku dat. V této tabulce je celkem 247 záznamů. U každého záznamu je uvedena lokalita a provozovatel. Dále je uvedena otevírací doba, zdali je nutné zaplatit poplatek, jestli je bezbariérový a zdroj těchto dat.

Záznamy odpovídající toaletám byly v OSM vyhledány podle atributu

```
amenity = toilets
```

Dle těchto atributů bylo nalezeno na území Prahy 185 záznamů v OSM. Při porovnání polohy záznamů (bodů) z OSM a z IPR se hledaly nejdříve body z IPR do vzdálenosti 5 m. V této oblasti bylo nalezeno naprosté minimum bodů. Proto byla použita vzdálenost 20 m, v tomto případě bylo nalezeno již více bodů z IPR v okruhu bodu z OSM. Pokud by se do vzdálenosti 20 m od bodu OSM nacházel pouze jeden bod z IPR, bylo by vhodné opravit jeho pozici (geometrii) a atributy. Ale v případě, že bylo nalezeno více bodů z IPR, bylo nutné jeden nich vybrat a upravit jeho polohu (geometrii) a ostatní body nově vytvořit v OSM. Protože bylo těžké rozhodnout, který bod přesunout a které body nově vytvořit, byl zvolen postup vytvořit nové body v OSM pro všechny záznamy (body) z datasetu IPR.

Na základě rešerše dat z IPR a doporučených atributů byly vybrány tyto

```
acces = yes/permissive  
fee = no/yes  
wheelchair = yes/no  
opening_hours =  
operator =  
description =  
source = IPR  
source:location = IPR
```

Pokud by hodnota byla `fee = yes` připojil by se atribut

```
fee:price =
```

a jeho hodnota by byla cena v Kč.

Hodnota u `description` by byla převzata ze sloupce `lokalita`. U atributu `operator` byla navržena hodnota `Hlavní Město Praha`, ale jen u bodů, které

mají v tabulce (data od IPR) ve sloupci `Operator` hodnotu `HMP`. U zbývajících bodů (záznamů) by tento atribut neměl žádnou hodnotu.

4.5.4 Parkovací automaty

Datová sada Parkovací automaty obsahovala celkem 452 záznamů (bodů) parkovacích automatů. Je zde uveden pouze `TYP` a `Poskytovatel` dat. Nelze tedy určit zdali automat přijímá mince, bankovky (českou nebo cizí měnu) a nebo platební karty.

Pro porovnání byly vyhledány body z OSM s atributy

```
amenity = vendings_machine
vending = parking_tickets
```

Podle těchto parametrů bylo nalezeno pouze 8 bodových prvků v databázi OSM. Pouze dva záznamy z OSM jsou v blízkosti k bodům z IPR. Jeden je vzdálený do 5 metrů a druhý do 20 metrů. Nelze tedy určit, zda oba (IPR a OSM bod) odkazují na stejný objekt, nebo na dva různé. Proto se původní body nebudou přesouvat, a ani rušit. Nové prvky půjde zpětně rozlišit podle atributu `source:loc`.

U nově přidáných uzlů byly navrženy tyto atributy

```
amenity = vendings_machine
vending = parking_tickets
source = IPR
source:loc = IPR
```

Do hodnoty `source` by byla použita hodnota ze sloupce `POSKYT`. Hodnota u atributu `source:loc` by byla použita `IPR`.

4.5.5 Záchytná parkoviště P+R

Datová sada Záchytná parkoviště P+R obsahoval 16 záznamů. V tabulce bohužel nejsou žádné jiné údaje, kromě polohy uložené v geometrii a poskytovatel dat ve sloupci `poskyt`.

Z dat OSM byly vyhledány body pomocí atributu

```
amenity = park_ride
```

Tento atribut je stále ve fázi schvalování, tudíž nebyly nalezeny žádné záznamy. Proto bylo vyzkoušeno hledání pomocí obecnějšího atributu


```
amenity = parking
```

Na území Prahy bylo nalezeno, toho času, 182 záznamů (bodů). Bylo vyzkoušeno hledání s přidáním atributem `name`, který by obsahoval P+R. Byl nalezen pouze jeden záznam.

Dále bylo vyzkoušeno hledání bodů s atributy

```
amenity = parking
park_ride = yes
```

Na území Prahy byl nalezen pouze jeden prvek.

Je tedy navrženo importovat všechny prvky z databáze IPR. Navržené atributy jsou

```
amenity = parking
park_ride = yes
source = HMP-URM
source:loc = IPR
```

4.5.6 Budovy 3D - výšky budov

V komentářích k návrhům v Talk-cz bylo dále doporučeno přidání výšek budov ze souboru dat Budovy 3D.

```
building:height
```

V návrhu bylo doporučeno označení pro zdroj tohoto importu, pro případné pozdější úpravy a opravy.

```
source:building:level =
```

Při rešerši dat uložených ve školní geodatabázi bylo zjištěno, že data stahovaná a ukládaná z OSM neobsahují potřebné údaje. Přesněji ukládaná tabulka neobsahovala údaje o výšce budov (sloupec `height` a sloupec `building:height`). Proto bylo nutné základní schéma pro `osm2pgsql` změnit tak, aby se tyto údaje také ukládaly. Upravené schéma bylo uloženo a bylo nastaveno jako výchozí.

Také se ukázalo, že bude vhodné, aby se ukládal údaj, jestli není polygon vytvořen z relace budovy, protože v rámci této relace (budovy) mají definovanou

výšku (`height`, nebo `building:height`) pouze její části, ze kterých je složena. K tomuto rozeznání byl použit atribut `building:part`.

V sekci Budovy 3D je dále uloženo 128 datových sad. Tyto datové sady obsahují budovy jen z malé oblasti, protože by datová sada budov pro celé území Prahy byla moc velká (až v řádů GB). Každá datová sada je přibližně obdélníková oblast z území Prahy a má svůj originální název.

V rámci této práce bude pracováno s datovou sadou `BD_Prah73`. Pro větší přehledost byl přidán diagram v příloze (viz obr. B.2/s46). Po vybrání této jedné datové sady (oblasti) byla vytvořena minimální ohraničující oblast okolo ní a přidán přesah o 100 metrů na každou stranu. Byl zvolen tento postup, aby se z dat OSM nevyřadily polygony, které by nebyly celým objemem uvnitř oblasti (jedné datové sady Budovy 3D).

Pomocí tohoto obdélníku byly vybrány budovy z OSM a poté byla vytvořena nárazníková zóna (`buffer`) s hodnotou 3 m pro každou budovu. Tento postup byl zvolen jelikož budovy z IPR a OSM nemusí mít přesně stejnou polohu a tvar (polygonů). Tento dočasný dataset byl nazván `bud73_OSM`.

Pro velikost nárazníkové zóny kolem budov z OSM se provedla analýza. Modelováno bylo s hodnotami po 0.5 m od 0.5 m až do vzdálenosti 5 m. Při malé velikosti nárazníkové zóny nastávala situace, že v této zóně nebyla nalezena žádná budova (polygon budovy) z datasetu IPR. Mohlo by se zdát, že čím větší nárazníková zóna kolem budov z OSM, tím lépe. Avšak, čím byla velikost nárazníkové zóny zvolena větší, tím více se stalo, že v této nárazníkové zóně „uvízlo“ více budov (polygonů) z datasetu IPR. Bylo tedy nutné zvolit „zlatou střední cestu“ tj. 3 m.

Další možností jak zamezit, aby v nárazníkové zóně kolem budov z OSM „uvízlo“ více polygonů budov (z datasetu IPR), je spouštět skript vícekrát a postupně volit větší a větší nárazníkovou zónu. Skript je napsán tak, aby vybral a vytvořil nárazníkovou zónu jen těm budovám (polygonům), které nemají atribut `building:part` ani `building:height`. Ovšem tento postup spouštět skript vícekrát je samozřejmě také více časově náročné.

Dále byl proveden průzkum dat z datové sady `BD_Prah73` ze souboru Budovy 3D (z IPR). Bylo zjištěno, že polygony, které vytvářejí 3D model jedné budovy, mají stejnou hodnotu ve sloupci `id_bud`. V 3D geometrii jsou uloženy souřadnice

x,y a z vrcholů polygonu. Polygony se stejnou hodnotou `id_bud` lze dále dělit dle hodnot ve sloupci `typ`.

Těmito hodnotami jsou

```
dilci plocha stresni kruhove plochy
komin
sikma stresni plocha
svisla obvodova stena
vikyr_stresni nadstavba
vytah_vetrani_klimatizace
vodorovna stresni plocha
vyznacena vez na strese
zakladna vikyre, stresni nadstavby
zakladova deska
.. a další
```

Poté byly z datové sady `BD_Prah73` vybrány všechny polygony, které měly ve sloupci `typ` obsahují frázi `stresni`. Byly vybrány pouze polygony tvořící střechu, protože bylo potřeba získat 2D polygon, který by nejlépe reprezentoval budovu v rovině „xy“. Geometrie těchto vybraných 3D polygonů byly opraveny funkcí `ST_MakeValid()` a následně „přetvořeny“ na 2D polygony funkcí `ST_Force_2D()`. Vynechala se jen Z-ová souřadnice z geometrie. Kvůli některým situacím, kdy k sobě polygony přesně nedoléhaly, bylo nutné kolem nich vytvořit nárazníkovou zónu (buffer). Jako dostačující byla zvolena hodnota 0.1 m. Následně byly tyto polygony takzvaně „rozpuštěny“. To znamená, že polygony, které se dotýkají minimálně jednou společnou stranou, se sloučí do jednoho polygonu. Rozpuštěny byly polygony navzájem jen ty, které měly stejnou hodnotu ve sloupci `id_bud`. Tím se docílilo, že každá budova bude reprezentována jenom jedním 2D polygonem. Tato dočasná datová sada byla nazvána `bud73_IPR`.

Následně byly pro každý polygon z `bud73_OSM` hledány všechny polygony z `bud73_IPR`, které by byly uvnitř. Na tuto operaci byla použita funkce

```
ST_Within( bud73_ipr.geom , bud73_osm.geom )
```

Tímto byly nalezeny prvky z dat IPR, které jsou „přibližně na stejném místě“ jako budovy z OSM. Do tabulky o dvou sloupcích byly uloženy pouze údaje `gid`

pro budovy z OSM a `id_bud` polygonu budovy (z dat IPR). Tato dočasná tabulka byla nazvána `gid_id`.

Dále tedy bylo možné k těmto prvkům (`gid` OSM budov) přiřadit výšku podle hodnoty `id_bud`, kterou mají všechny polygony tvořící jednu budovu stejnou. Nejprve musely být vybrány pouze ty záznamy `gid`, které jsou v tabulce `gid_id` pouze jednou. To jest polygon budovy z OSM, ve které byl nalezen pouze jeden polygon z datasetu IPR. Následně k nim byly připojeny výšky podle hodnoty `id_bud`. Pro nalezení výšky dané budovy ¹¹ byly použity všechny polygony, které ji tvoří. V rámci jedné budovy (podle stejné hodnoty v `id_bud`) byla výška budovy vypočtena jako rozdíl nejvyššího bodu a nejnižšího bodu budovy. Přesněji byla vypočtena takto

```
select id_bud,
       max(maxZ) - min(minZ) as Height
from(   select id_bud,
              ST_ZMax(geom) as maxZ,
              ST_ZMin(geom) as minZ
        from   ipr.bd3_prah73
        ) as max_min
group by id_bud
```

Ve skriptu je tato dočasná tabulka označena `osm_height`.

Nakonec se získané hodnoty výšek budov přidají do původního datasetu (tabulky) `praha_building_osm`. Dočasné tabulky `bud73_osm`, `bud73_ipr`, `gid_id` a `osm_height` jsou nakonec vymazány.(viz příloha C/s47)

¹¹dle http://wiki.openstreetmap.org/wiki/Cs:Simple_3D_Buildings

5 Závěr

V bakalářské práci bylo popsáno základní dělení prvků v databázi OpenStreetMap, licencování a hlavní zdroje dat. Dále byl vysvětlen termín Opendata (otevřená data). Při řešení licenční otázky byl zjištěn problém. Licence dat OSM a IPR nejsou kompatibilní. Nelze tedy v současné době začlenit data IPR do databáze OSM. V rámci práce byly popsány hlavní rozdíly mezi těmito licencemi. Problém byl řešen tak, že se vyčká na změnu licence u dat IPR, a v tomto mezechase se import připraví, aby se poté mohl už jen provést.

Pro snadnější práci s daty byl vytvořen program, který umožňuje hledání všech distribuovaných dat IPR a dále jejich stažení a následný import do geodatabáze PostGIS. Program byl napsán v jazyce Python.

Všechna data zveřejněná IPR byla prozkoumána (v programu QGIS) a bylo posouzeno, zdali jsou či nejsou vhodná začlenit je do databáze OpenStreetMap. Po vybrání vhodných dat pro začlenění do databáze OSM byla provedena rešerše aktuálního stavu v databázi OSM. Na základě toho vzešly čtyři datové sady vhodné pro import. Jmenovitě Parkovací automaty, Parkoviště P+R, Odpadní zařízení pro občany a Veřejné toalety. Následně z nich byla vytvořena a upravena data s ohledem doporučených atributů OSM. Na diskuzním fóru Talk-cz byl tento záměr zveřejněn a v diskuzi dále bylo navrženo, zabývat se ještě importem výšek budov.

Po změně licence u dat zveřejňovaných IPR se samotný import může provést v programu `osmosis`. Jako zajímavé by se do budoucna jevílo zpracovat 3D modely budov z datasetu Budovy 3D. Což by znamenalo navrhnout program (skript), který by vytvořil modely budov v OSM.

Seznam zkratek

OSM	OpenStreetMap (<i>Open Street Map</i>)
IPR	Institut plánování a rozvoje hlavního města Prahy
RUIAN	Registr územní identifikace, adres a nemovitostí
PID	Pražská integrovaná doprava
CC-BY-SA	Uveďte původ-Zachovejte licenci (<i>Creative Commons Attribution-ShareAlike</i>)
ODbL	(<i>Open Database License</i>)
GIS	Geografický informační systém (<i>Geographic Information System</i>)
GNSS	Globální družicový polohový systém (<i>Global Navigation Satellite System</i>)
GPS	Globální polohovací systém (<i>Global Positioning System</i>)
GLONASS	Glonass (<i>Globalnaja navigacionnaja sputnikovaja sistéma</i>)
WGS 84	Světový geodetický systém 1984 (<i>World Geodetic System 1984</i>)
S-JTSK	Systém jednotné trigonometrické sítě katastrální
GC	Geocaching (<i>Geocaching</i>)
SQL	SQL (<i>Structured Query Language</i>)
XML	Rozšiřitelný značkovací jazyk (<i>Extensible Markup Language</i>)
GDAL	Geoprostorová datová knihovna (<i>Geospatial Data Abstraction Library</i>)
QGIS	Quantum GIS
JOSM	Java OpenStreetMap Editor

Literatura

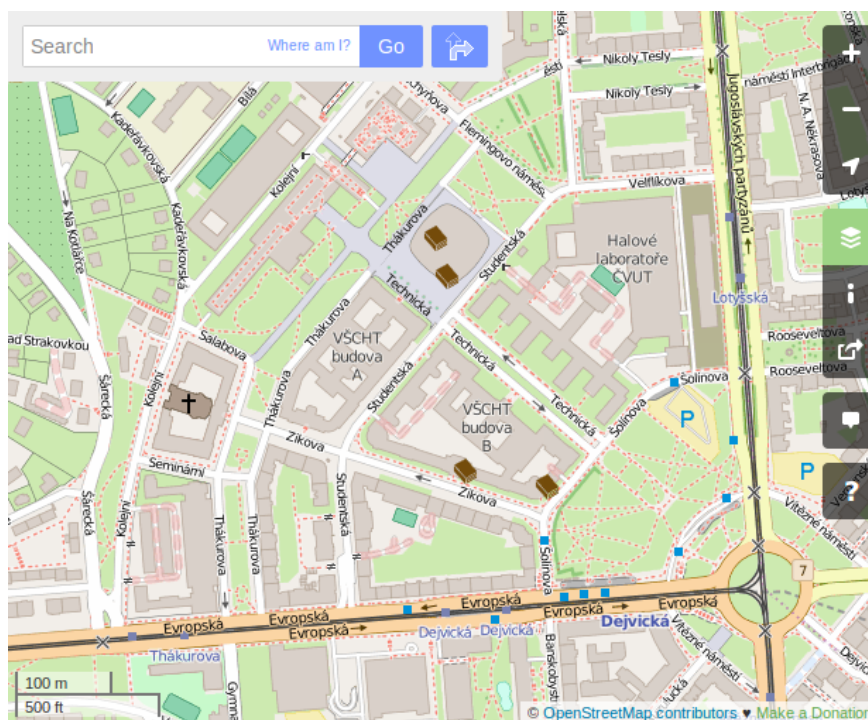
- [1] BATTERSBY, Sarah E. et al. Implications of Web Mercator and Its Use in Online Mapping. [online]. 05-05-2016. s. 92. Dostupné z: http://cegis.usgs.gov/projection/pdf/Battersby_ImplicationsofWebMecatorandItsUseinOnlineMapping.pdf.
- [2] *Argparse Tutorial* [online]. Python Software Foundation, 10-04-2016. Dostupné z: <https://docs.python.org/2/howto/argparse.html>.
- [3] *Creative Commons licence* [online]. Creative Commons, 15-05-2016. Dostupné z: <https://creativecommons.org/licenses/by-sa/4.0/legalcode/>.
- [4] *GDAL/OGR FAQ* [online]. 08-04-2016. Dostupné z: <http://trac.osgeo.org/gdal/wiki/FAQ>.
- [5] *Co IPR Praha dělá?* [online]. Institut plánování a rozvoje hlavního města Prahy, 30-03-2016. Dostupné z: www.iprpraha.cz/clanek/343/co-ipr-praha-dela.
- [6] *IPR PRAHA - základní info* [online]. Institut plánování a rozvoje hlavního města Prahy, 05-04-2016. Dostupné z: http://www.iprpraha.cz/uploads/assets/dokumenty/obecne/brozura_ipr_1509.pdf.
- [7] *Open Database License (ODbL) v1.0* [online]. Open Knowledge Foundation, 15-05-2016. Dostupné z: <http://opendatacommons.org/licenses/odbl/1.0/>.
- [8] *License/Why CC BY-SA is Unsuitable* [online]. OpenStreetMap Foundation, 30-08-2012. Dostupné z: <http://trac.osgeo.org/gdal/wiki/FAQ>.
- [9] *Stupně otevřenosti datových sad* [online]. Otevřená data, 15-05-2016. Dostupné z: <http://opendata.gov.cz/standardy:stupne-otevrenosti>.
- [10] *European countries making clear progress with Open Data* [online]. not known, 05-05-2016. Dostupné z: <http://blogs.worldbank.org/opendata/european-countries-making-clear-progress-open-data>.

- [11] *Map Features, česká verze* [online]. OpenStreetMap Foundation, 23-03-2016. Dostupné z: https://wiki.openstreetmap.org/wiki/Cs:Map_Features.
- [12] *Copyright and Licence* [online]. OpenStreetMap Foundation, 10-03-2016. Dostupné z: <http://www.openstreetmap.org/copyright>.
- [13] *We Are Changing The License* [online]. OpenStreetMap Foundation, 10-04-2016. Dostupné z: http://wiki.osmfoundation.org/wiki/License/We_Are_Changing_The_License.
- [14] *Automated redactions complete* [online]. 28-04-2016. Dostupné z: <https://blog.openstreetmap.org/2012/07/26/automated-redactions-complete/>.
- [15] *OpenStreetMap moves to PostgreSQL* [online]. Paul Ramsey, 30-04-2016. Dostupné z: <http://blog.cleverelephant.ca/2009/04/openstreetmap-moves-to-postgresql.html>.
- [16] *OpenStreetMap statistic* [online]. altogetherlost.com and Pascal Neis (neis-one.org), 08-04-2016. Dostupné z: <http://osmstats.neis-one.org>.
- [17] *PostgreSQL: about* [online]. The PostgreSQL Global Development Group, 14-04-2016. Dostupné z: <http://postgresql.org/about/>.
- [18] *General Python FAQ* [online]. Python Software Foundation, 01-04-2016. Dostupné z: <http://docs.python.org/2/faq/general.html>.
- [19] *Wikipedie, anglická a česká verze* [online]. Wikipedia Foundation, 10-03-2016. Dostupné z: <https://www.cs.wikipedia.org/wiki/OpenStreetMap>.
- [20] *Python* [online]. Wikipedia Foundation, 12-04-2016. Dostupné z: <https://cs.wikipedia.org/wiki/Python>.
- [21] *XML parsing* [online]. A Kenneth Reitz Project, 01-04-2016. Dostupné z: <http://docs.python-guide.org/en/latest/scenarios/xml/>.

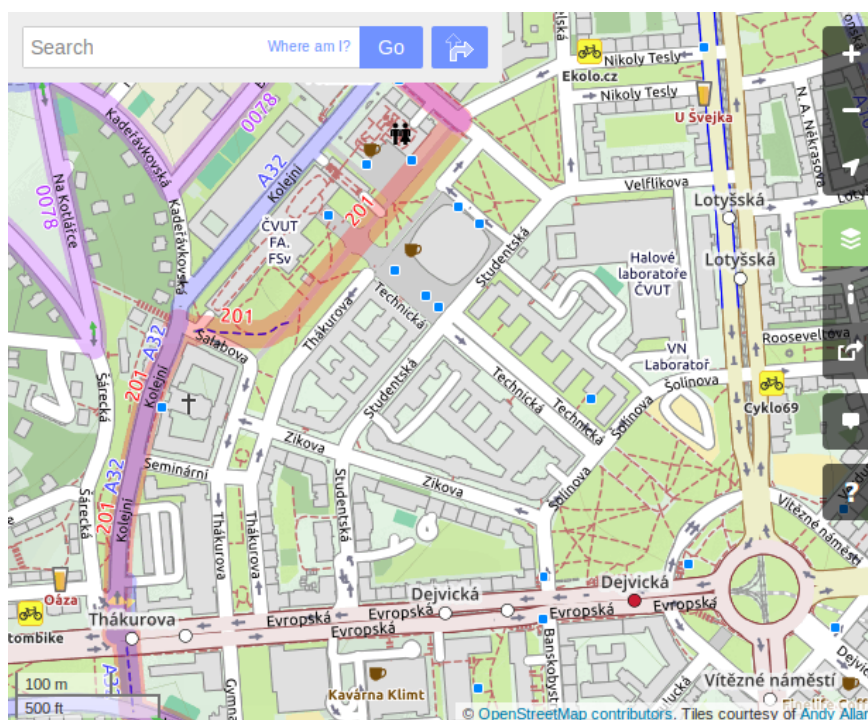
Seznam příloh

A Typy vykreslení OSM dat	42
B Diagramy	45
C Obsah CD	47

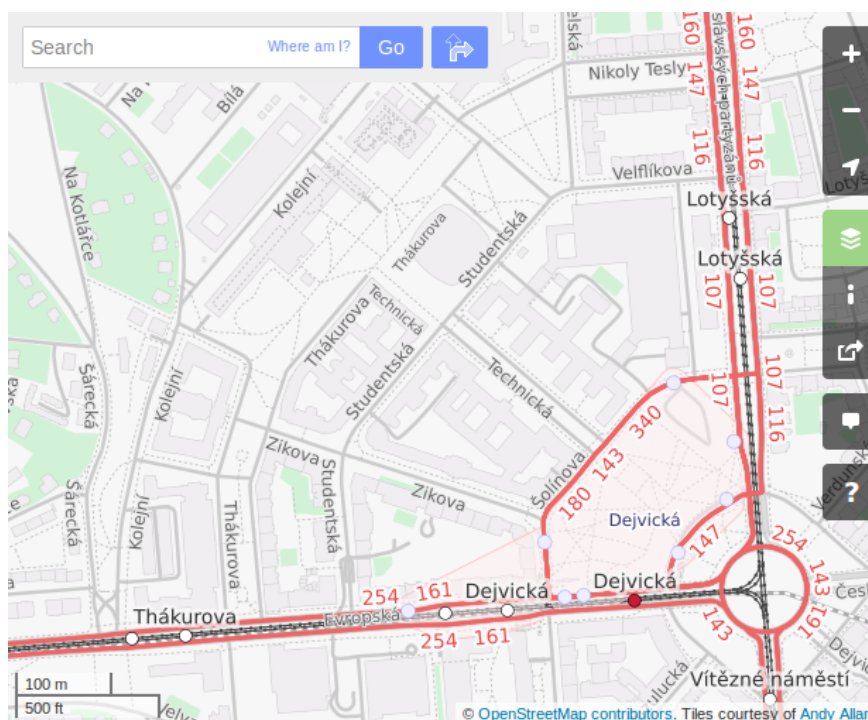
A Typy vykreslení OSM dat



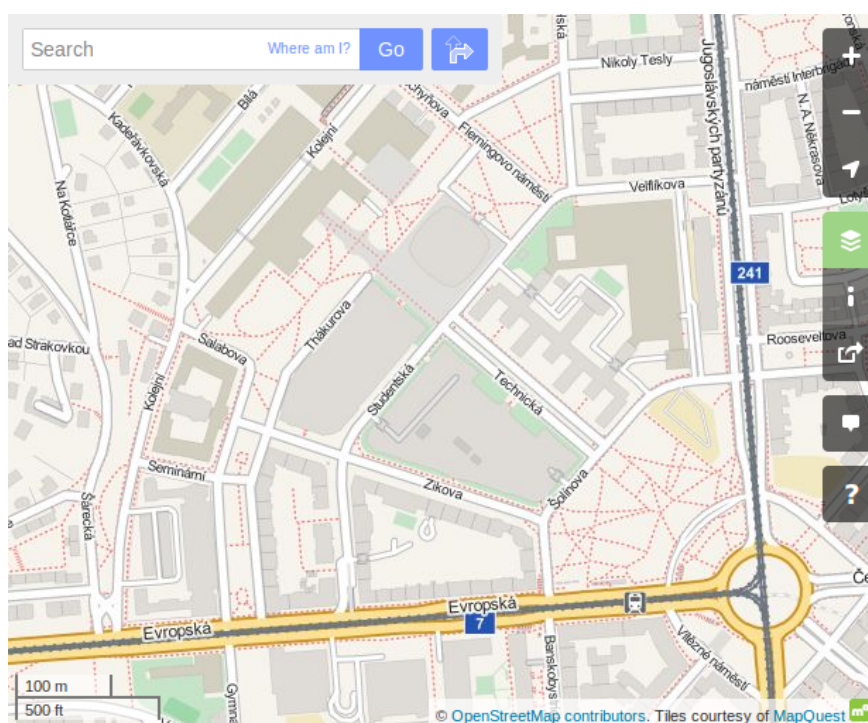
Obrázek A.1: Standardní mapa (Standard)



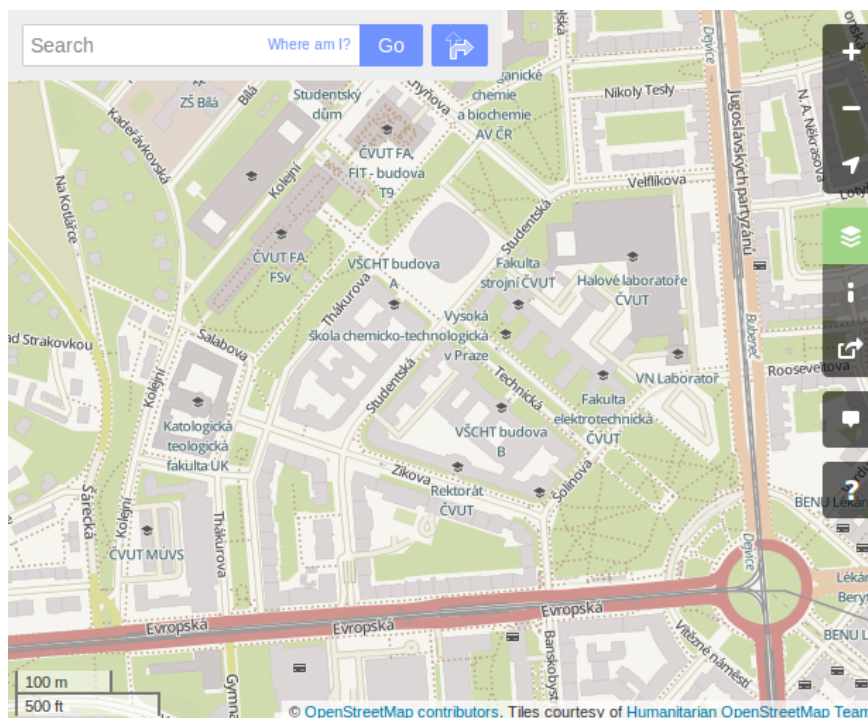
Obrázek A.2: Cyklistická mapa (Cycle Map)



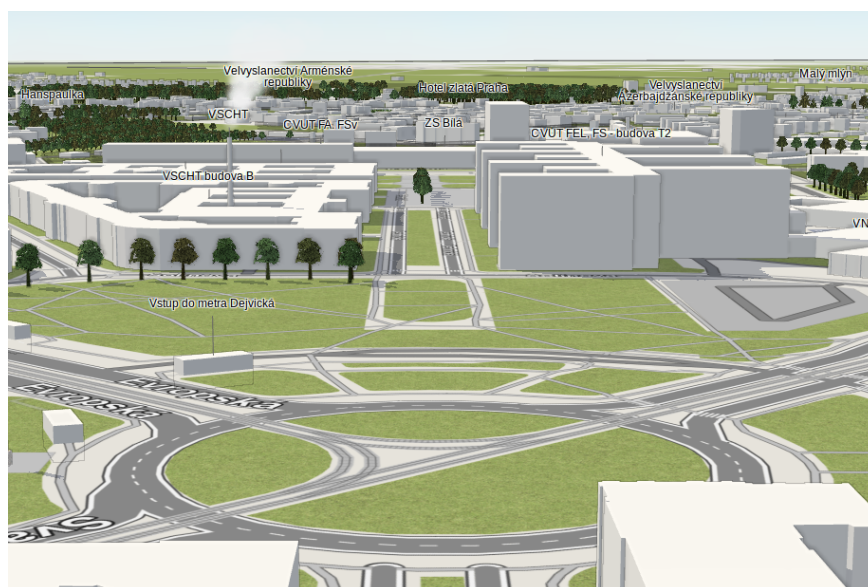
Obrázek A.3: Transport Map



Obrázek A.4: MapQuest Open

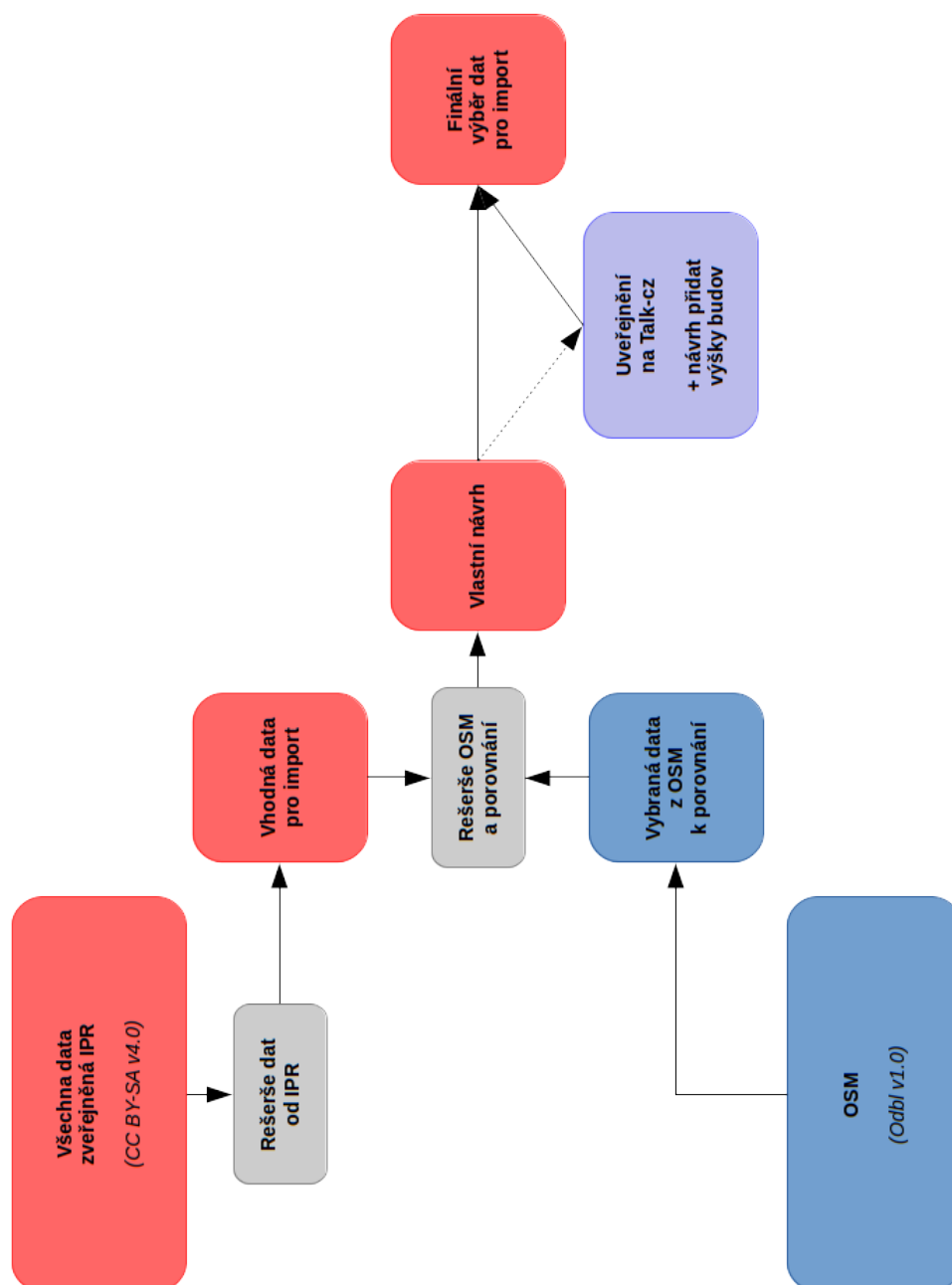


Obrázek A.5: Humanitární mapa (Humanitarian)

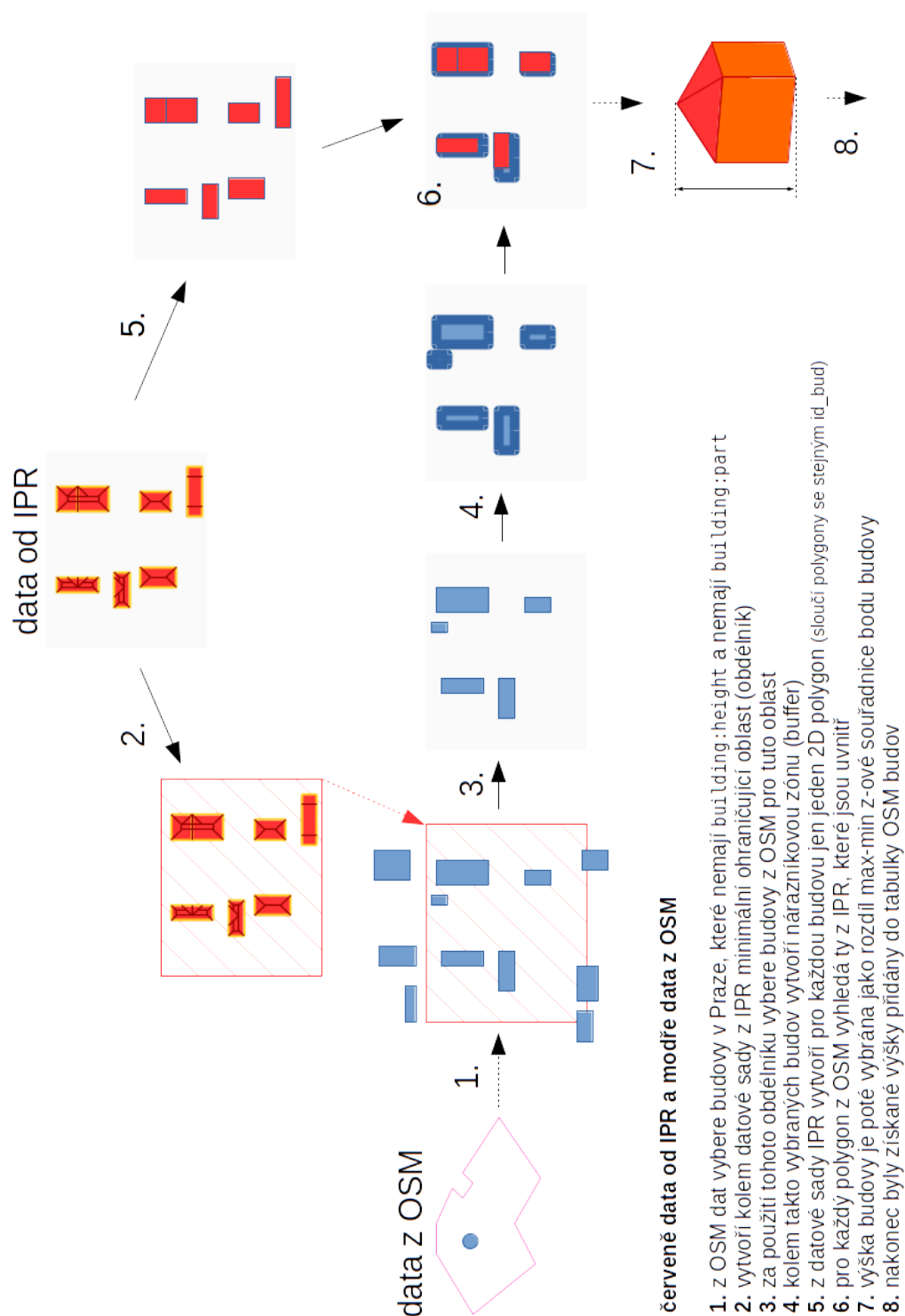


Obrázek A.6: Projekt F4 (zdroj <http://demo.f4map.com/>)

B Diagramy



Obrázek B.1: Diagram postupu práce.



Obrázek B.2: Diagram postupu práce pro získání výšky budov.

C Obsah CD

.	
src	
iprdownload	zdrojové soubory programu
IprBase.py	
IprPg.py	
Iprdownload.py	
pg	zdrojové kódy shell skriptu
pgis_osm_bp.style	vlastní schema dat z OSM
upgrade_pgis_osm_bp.sql	
sql	zdrojové kódy sql dump
buildins.sql	
park_and_ride.sql	
parkomat.sql	
recycling_centre.sql	
toilets.sql	
text	
latex	zdrojové soubory textu této práce
martin-jakl-bp-2016.xls	anotace práce
martin-jakl-bp-2016.pdf	tento text
zadani-oficialni.jpeg	naskenované oficiální zadání práce