

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

Fakulta životního prostředí

Katedra aplikované geoinformatiky a územního plánování



**Zhodnocení kvality dat OpenStreetMap
v Praze**

DIPLOMOVÁ PRÁCE

Vedoucí diplomové práce: Ing. Vítězslav Moudrý, Ph.D.

Diplomant: Bc. Jan Zídek

2016

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

Fakulta životního prostředí

ZADÁNÍ DIPLOMOVÉ PRÁCE

Jan Zídek

Regionální environmentální správa

Název práce

Zhodnocení kvality dat OpenStreetMap v Praze

Název anglicky

Quality assessment of the OpenStreetMap data in Prague

Cíle práce

- Literární rešerše využití VGI/OSM dat
- Porovnání kvality dat projektu OSM s oficiálními datovými zdroji na území hlavního města Prahy

Metodika

Účast veřejnosti na sběru a tvorbě prostorových dat existuje téměř celé století, ale teprve s rozvojem internetu (web 2.0) se tato forma sběru prostorových dat mohla globálně rozvinout do současné podoby. Dobrovolně sbíraná data jsou fenoménem posledních let, což se projevuje nejen v jejich širokém užití samotnými uživateli, ale i ve zkoumání možností jejich užití ve vědě, či aspektů, které vedou uživatele k jejich tvorbě. Autor zpracuje literární rešerši na téma UGC (User Generated Content) a VGI (Volunteered Geographic Information), přičemž se zaměří na projekt OSM (OpenStreetMap). Řešitel porovná kvalitu prostorové a popisné informace (např. kompletnost, sémantickou přesnost) v OSM s dostupnými oficiálními daty pro Prahu.

Doporučený rozsah práce

40 – 50 stran

Klíčová slova

Formulace klíčových slov je úkolem řešitele.

Doporučené zdroje informací

- Antoniou, V., & Skopeliti, A. (2015). Measures and Indicators of Vgi Quality: AN Overview. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1, 345-351.
- Fan, H., Zipf, A., Fu, Q., & Neis, P. (2014). Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science*, 28(4), 700-719.
- Flanagin, A. J., & Metzger, M. J. (2008). The credibility of volunteered geographic information. *GeoJournal*, 72(3-4), 137-148.
- Girres, J. F., & Touya, G. (2010). Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS*, 14(4), 435-459.
- Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial statistics*, 1, 110-120.
- Goodchild, M. F. (2007, January). Citizens as voluntary sensors: spatial data infrastructure in the world of Web 2.0. In *International journal of spatial data infrastructures research*.
-

Předběžný termín obhajoby

2015/16 LS – FŽP

Vedoucí práce

Ing. Vítězslav Moudrý, Ph.D.

Garantující pracoviště

Katedra aplikované geoinformatiky a územního plánování

Elektronicky schváleno dne 29. 3. 2016

doc. Ing. Petra Šímová, Ph.D.

Vedoucí katedry

Elektronicky schváleno dne 29. 3. 2016

prof. RNDr. Vladimír Bejček, CSc.

Děkan

V Praze dne 11. 04. 2016

Prohlášení

Prohlašuji, že jsem diplomovou práci na téma *Zhodnocení kvality dat OpenStreetMap v Praze* vypracoval samostatně, pod vedením Ing. Vítězslava Moudrého, Ph.D. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

V Praze dne 19. 4. 2015

.....

Poděkování

Děkuji Ing. Vítězslavu Moudrému Ph.D. za vedení mé diplomové práce a za veškeré poskytnuté rady, konzultace a připomínky týkající se jejího obsahu. Také děkuji všem, kteří se podíleli na korekci textu, mé rodině a všem ostatním, kteří mě při práci podporovali.

Abstrakt

OpenStreetMap (OSM) je v současnosti největším a objektivně nejúspěšnějším projektem ze všech pokusů o volně dostupnou alternativu ke komerčním mapovým databázím a projektům oficiálních agentur. V České republice má širokou podporu uživatelů, ale kvalita těchto dobrovolnických geografických dat (VGI) nebyla dosud zkoumána. Tato práce hodnotí z několika hledisek kvantifikovatelné vlastnosti budov v Praze. Cílem diplomové práce bylo vyhodnotit především kompletnost, sémantickou přesnost a tvarovou přesnost půdorysů budov. Data z OSM byla porovnána s Digitální mapou Prahy z Geoportálu Praha, jež je oficiálním polohopisným mapovým dílem popisujícím území hl. m. Prahy. Samotné zpracování dat bylo provedeno v programu ArcGIS od firmy ESRI. Výsledky ukazují, že OSM v Praze má vysokou sémantickou přesnost a také téměř 100% kompletnost z hlediska celkové plochy. Budovám však ve více než polovině případů chybí základní popisné informace. Z hlediska tvarové přesnosti jsou půdorysy budov v OSM určeny menším množstvím vrcholů, než půdorysy budov v datech z Geoportálu Praha a v některých případech chybí určité architektonické prvky.

Klíčová slova: OpenStreetMap, Půdorys budovy, Hodnocení kvality, VGI

Summary

OpenStreetMap (OSM) is the largest and objectively the most successful project from all attempts to open-source alternative to commercial map databases and projects of the official agencies. OSM is widely popular in Czech Republic, but the quality of this VGI (volunteered geographic information) dataset has not yet been studied. This master thesis evaluates several aspects of quantifiable characteristics of buildings in Prague. The objective was to primarily evaluate completeness, semantic accuracy and shape accuracy of building footprints. Data from OSM were compared with Digital map of Prague by Geoportal Prague, which is the official map resource covering the area of the city Prague. The data processing was done in ArcGIS by the ESRI company. The results show that OSM in Prague have high semantic accuracy and almost 100% completeness in terms of total area, but more than 50% of buildings are missing basic descriptive data. In terms of shape accuracy the footprints were identified by smaller amount of vertices than the footprints from Geoportal Prague and in some cases certain architectural details were missing.

Keywords: OpenStreetMap, Building footprint, Quality assessment, VGI

Obsah

1	Úvod	9
1.1	Cíle práce	10
2	Literární rešerše	11
2.1	Uživateli generovaný obsah (UGC).....	11
2.2	Dobrovolnická geografická data (VGI)	12
2.2.1	Vhodnost použití VGI dat	14
2.2.2	Indikátory vhodnosti použití VGI dat.....	14
2.2.3	„Bias“ ve VGI datech	15
2.2.4	Nepřesnosti v umístění získaných dat.....	16
2.3	Povaha a motivace uživatelů VGI	17
2.3.1	Typy a charakteristika uživatelů.....	17
2.3.2	Motivace uživatelů	19
2.3.2.1	Motivace uživatelů editovat své příspěvky	20
2.3.3	Negativní motivační faktory přispěvatelů.....	20
2.3.4	Kritika VGI	21
2.4	OpenStreetMap a další VGI Projekty	22
2.4.1	OpenStreetMap.....	22
2.4.1.1	OpenStreetMap v České republice.....	24
2.4.2	Další VGI mapové projekty	25
2.4.2.1	Wikimapia	25
2.4.2.2	Google Map Maker	25
2.4.2.3	Flickr	26
3	Metodika	27
3.1	Vstupní data.....	27
3.2	Relace mezi daty.....	27
3.3	Práce s daty.....	29
3.4	Kompletnost	29
3.4.1	Kompletnost OpenStreetMap v Praze.....	30

3.5	Sémantická přesnost	31
3.5.1	Sémantická přesnost OpenStreetMap v Praze.....	31
3.6	Tvarová přesnost.....	32
3.6.1	Algoritmus k nalezení relací 1:1.....	32
3.6.2	Metoda překrývajících se polygonů pro nalezení relace 1:1 mezi budovami v datových vrstvách.....	33
4	Výsledky	35
4.1	Kompletnost	35
4.2	Sémantická přesnost	39
4.2.1	Kvantifikace sémantické přesnosti	39
4.3	Bias	39
4.4	Tvarová přesnost.....	40
4.4.1	Kvantitativní posouzení metody překrývajících se polygonů k nalezení korespondujících budov v datových vrstvách.....	40
4.4.2	Tvarová přesnost OpenStreetMap v Praze.....	40
5	Diskuse	44
5.1	Srovnání práce s výsledky podobných vědeckých prací.....	44
5.2	Budoucí projekty a výzkumy v OpenStreetMap	45
5.3	Vhodnost použití dat z OpenStreetMap ve vybraných projektech.....	46
6	Závěr	48
7	Přehled literatury a použitých zdrojů	49

1 Úvod

Rapidní vývoj prostorových technologií v poslední dekádě podnítil širokou veřejnost k tomu, aby vytvářela prostorová data. (Goodchild, 2007a; Elwood, 2008). Spolu s popularizací internetu, osobních počítačů a přenosných zařízení s GPS (globální družicový polohový systém), jako jsou mobilní telefony a tablety, mohou lidé v současnosti snadno sdílet georeferencovaná data o pozorování světa (jako např. fotky či spatření divoké zvěře) na interaktivních tzv. geovizualizačních webových službách (Google Earth, Google Maps, Microsoft Visual Earth), na sociálních sítích (Twitter, Flickr, Instagram) nebo přispívat na různé vědecké projekty (Silvertown, 2009; Sullivan et al., 2009; Dickinson et al., 2012). Tyto technologie vytvořily nevídané možnosti pro běžného člověka, aby sdílel svá pozorování s celým světem. Pozorování uživatelů jsou poté nazývána VGI (Volunteered Geographic Information, „dobrovolnická geografická data“) (Goodchild, 2007a; Sui, 2008). Navíc tyto technologie vytváří příležitosti pro vědce pohodlně získat od uživatelů užitečné informace o geografických vlastnostech území nebo o unikátních jevech v území (Zhu et al., 2015).

VGI mají mnoho různých výhod. Zaprvé obsahují velké množství nedocenitelných informací o menších lokálních územích, která uživatelé (tedy lidé, kteří příslušná data vytvořili), jakožto „experti“ na svoje nejbližší okolí, sledovali po dlouhou dobu a získávali o nich informace (Goodchild, 2007b). VGI však mají potenciál odhalit informace i ohledně větších územích, pokud budeme schopni získat komplexní údaje od milionů či dokonce miliard lidí, kteří se na zemi pohybují. Navíc VGI je možno upravovat v reálném čase a reagovat tak na změny v území, které by např. satelity nebyly schopny dostatečně rychle zjistit, zatímco pro člověka žijícího v daném území je to poměrně snadné. V neposlední řadě je získávání VGI mnohem levnější, než sbírání dat tradičními vědeckými metodami. V mnoha případech lidé přispívají pouze v duchu altruismu a možného sebezviditelnění a ani neočekávají finanční kompenzaci, zatímco sběr dat pro profesionální účely je proces náročný na množství vyškolených odborníků, kteří se navíc musí držet striktních protokolů. I když to nemusí být na první pohled zřejmé, význam tohoto finančního faktoru může být v mnoha případech velmi důležitý pro úspěch projektu (Zhu et al., 2015).

Projekt OpenStreetMap je v současnosti největším a objektivně nejúspěšnějším geoinformatickým projektem ze všech pokusů o volně dostupnou alternativu k projektům oficiálních agentur (Goodchild et Li, 2012). Jeho úplně první verze byla spuštěna v červenci roku 2004 ve Velké Británii. OSM obsahuje geografická data pro všechny části světa, ve kterých je alespoň částečný dobrovolný zájem od internetových uživatelů (Haklay et Weber, 2008).

1.1 Cíle práce

Cílem práce je zjistit, jak jsou data o budovách z projektu OpenStreetMap kvalitní v porovnání s autoritativními datovými zdroji od oficiálních agentur. Toho bude docíleno tak, že se vyhodnotí jejich kvalitativní vlastnosti, jako např. kompletnost, tvarová přesnost či sémantická přesnost v Praze, oproti referenčním datům z Geoportálu Praha (Digitální mapa Prahy). Druhotnými cíly jsou další interpretace dat, porovnání výsledků s podobnými projekty a doporučení vhodného použití dat.

Obrázek 1: „Heat mapa“ stop po signálu GPS v OpenStreetMap v Evropě (Fisher, 2012)



2 Literární rešerše

2.1 Uživateli generovaný obsah (UGC)

Již na začátku této kapitoly je potřeba zmínit rozdíl mezi slovy “uživatel” (user) a “příspěvatel” (contributor). Z významu slova je příspěvatel ten, který data vytváří a uživatel je následně používá. V češtině však není slovo příspěvatel zdaleka tolik frekventované, a přestože se v některých případech lépe hodí při popisu UCG, VGI a jich příbuzných pojmů, nebude v této práci ve většině případů použito. U výše zmíněných termínů je totiž příspěvatel i uživatel často jedna a tatáž osoba a rozdíl se tak stírá. V zájmu konzistence textu bude užit převážně termín uživatel, zatímco slovo příspěvatel se objeví pouze v odůvodněných případech, kdy je jasné, že daná osoba s daty dále nepracuje či k nim ani pak nemá přístup.

UGC neboli užiteli generovaný obsah (User-generated content či řidčeji consumer-generated media) je definován jako „jakákoli forma obsahu, jakým je například blog, wiki, diskusní fórum, videozáznam, audiozáznam, chat, tweet, podcast, pin, digitální fotka, reklama a další formy médií, které byly vytvořeny užiteli onlinové služby nebo systému, často postupná na stránkách sociálních sítí“ (Chua et al., 2014)

Přestože užiteli generovaný obsah existoval již mnohem dříve, do mainstreamového používání se UGC dostal během roku 2005 (Chua et al., 2014) a zvedl zájem lidí o další publikování na webu a vytváření nového a unikátního mediálního obsahu. UGC má široké spektrum využití, včetně zpráv, výzkumu a řešení problémů, a reflektuje rozvoj produkce médií za pomoci nových technologií, které jsou přístupné a dostupné široké veřejnosti. Navíc díky kombinaci open-source dat a flexibilního licencování obsahu dále snižuje bariéry pro spolupráci mezi užiteli. V poslední dekádě získával UGC neustále na oblibě a více a více uživatelů se podílí na vytváření dat na sociálních sítích a fórech, na kterých tato data sdílí (EET, 2015).

Data z roku 2006 ukazují, že 35% uživatelů internetu ve Spojených státech amerických (USA) vytvořilo nějaký výše definovaný obsah a sdílelo ho online; 26-34% uživatelů tento obsah vytvořilo samo (fotky nebo videa); 32% ohodnotilo

nějaký výrobek, osobu nebo službu na onlineovém hodnotícím systému; 20% si vytvořilo osobní stránku na sociální síti, kterou mohou ostatní uživatelé vidět, a 8 % má osobní blog (Lenhart, 2006).

Když k tomu přidáme, že 70% všech dospělých lidí v USA je pravidelně online (HI 2007; Lenhart 2006), tak tato čísla naznačují, že se velké množství lidí nějak smysluplně podílí na vytváření internetového prostředí. Navíc tyto hodnoty neustále rostou - v roce 2013 dokonce více než polovina uživatelů internetu v USA vytvořila nějakou fotku a sdílela ji online a více než čtvrtina sdílela jimi vytvořené video (Statista, 2016).

Tento trend využívání UGC měl významný vliv i na geografické informace, které jsou upravovány nebo sdíleny online. Například satelitní a mapová služba Wikimapia v roce 2007 obsahovala skoro 7 milionů míst, které byly identifikovány a upravovány jejími uživateli (Flanagin et Metzger, 2008). Podobně pak nabízí Google Earth příležitosti pro uživatele, aby sdíleli data o lokálním prostředí pomocí fotografií a speciálních textových anotací. V prvních dvou letech přilákal Google Earth 200 milionů uživatelů (Google, 2007).

2.2 Dobrovolnická geografická data (VGI)

VGI je určitý typ prostorových dat, kterými přispívají dobrovolníci ve formě UGC. VGI nabízí možnost pro amatéry a pokročilé uživatele, aby sbírali a sdíleli geografické informace (Mohammadi et Malek, 2014). Množství těchto dat, kterými přispívají uživatelé, kteří nemají odborné vzdělání v geografii jako vědci, v poslední době neustále stoupá. (Goodchild 2007b). Mohou podpořit nebo obohatit autoritativní (od ověřených zdrojů) datové sady (Antoniou 2011), ale v mnoha oblastech jsou dokonce tato data jediná, která jsou k dispozici a mají tak potenciál k tomu, aby byla využívána v nejaktuálnějších geografických projektech. (Zhu et al., 2015) I v případě nejprostších metod sbírání dat tak mohou mít tyto sady komparativní výhodu (Craglia et al., 2008).

První náznaky tvorby geoinformatických dat od uživatelů ve vysoké kvalitě se začaly objevovat kolem roku 2004. Souřadnice různých geografických veličin bylo možné zachytit pomocí GPS nebo zaznačit na obrázcích s vysokým rozlišením, které nabízel

například Google Earth. Dobrovolníci z celého světa tak mohli začít vytvářet mapy svého nejbližšího okolí s vysokou znalostí místních poměrů a navíc efektivněji, než by to zvládl mapový expert ze vzdálené státní agentury. OpenStreetMap (OSM) se postupně stal nejznámější a také nejúspěšnější službou z množství různých projektů, které se snažily vytvořit alternativu k produktům oficiálních úřadů (Goodchild et Li, 2012).

Vyhodnocování kvality VGI dat je populární téma mezi akademiky a badateli. Běžnou praxí je pak porovnávání VGI datových sad a referenční datových sad za účelem zjištění kvality VGI dat kvantitativními metodami (Mohammadi et Malek, 2014). Byly navrženy různé postupy a indikátory pro zjištění kvality VGI dat a vzniklo několik metodologií k jejich porovnání s autoritativními datovými sety. Po dlouhou dobu měly výsadu vytvářet prostorová data státní mapové agentury jednotlivých zemí, případně komerční společnosti aktivní v této oblasti. Sběrání, modelování, spravování a aktualizace prostorových datových sad je často komplikované a časově náročné.

K tomu, aby bylo dosaženo co nejlepší možné kvality dat, byly vytvořeny standardizované procedury, které musí být důsledně dodržovány. Díky tomu jsou tyto sady opatřeny určitou úrovní garance kvality, která pramení z důvěryhodnosti hodnotící autority. Takové datové sady jsou však velmi drahé a pojí se s nimi restriktivní licenční podmínky (Antoniou et Skopeliti, 2015).

Přestože mají VGI datové sady podstatné rozdíly v produkčním procesu oproti autoritativním datům, má stále smysl používat zavedené a ozkoušené metody evaluace, abychom zjistili kvalitu prostorových dat vytvořených pomocí crowdsourcingu (novotvar, jenž označuje výsledek spolupráce blíže nespecifikované skupiny lidí, které bylo téma zadané jako všeobecná výzva). Komplexní nástroje na zpracování těchto dat popisují ISO standardy, které poskytují vodítka pro základní evaluaci charakteristik prostorových dat (Antoniou et Skopeliti, 2015).

Autoritativní data navíc v mnoha částech světa zastarávají a byla získána pomocí starších technologií, které jsou méně přesné, než technologie, které jsou dostupné současné široké veřejnosti. Například v USA je pro dobrovolníka snadné získat data, která jsou přesnější, než nejdetajnější data dostupná pro kontinentální státy

(v současnosti jsou tato autoritativní data v měřítku 1:24 000). V mnoha případech není žádný problém ukázat, že VGI jsou ve vyšší kvalitě než dostupná autoritativní data. (Goodchild et Li, 2012)

2.2.1 *Vhodnost použití VGI dat*

Je zajímavé, že jak se VGI datové sady stávají neustále detailnějšími a detailnějšími, přestává být jasné, jestli je použití autoritativních datových sad jako referenčních datových zdrojů nejlepší volbou pro evaluaci kvality. Jinak řečeno, jak poznamenávají Vandecasteele et Devillers (2015), je výzvou v různých částech světa vybrat tu kompletnější a přesnější datovou sadu, a to narušuje základní představu o vyhodnocovacích metodách kvality VGI.

Na druhou stranu kvalita VGI může být dosti nestálá. Časté změny, které dělají uživatele v důležitých attributech, mohou zhoršit celkovou kvalitu a použitelnost VGI datových sad (Antoniou et Skopeliti, 2015).

2.2.2 *Indikátory vhodnosti použití VGI dat*

Jelikož VGI je uživatelsky generovaný obsah, mnoho badatelů podporuje myšlenku, že by mohla existovat korelace mezi kvalitou VGI a demografickými ukazateli (např. Tulloch, 2008; Elwood, 2008). Tuto myšlenku podpořil empirickým výzkumem Mullen et al. (2014), který sledoval korelaci mezi demografickými údaji v území a *kompletností* (viz také kapitola 3.4.) a *poziční přesností* dat. Podobně Zielstra et Zipf (2010) ukázali, že území s nízkou hustotou populace (tj. zemědělské oblasti) mají přímý dopad na kompletnost VGI dat. I další práce naznačily, že hustota populace pozitivně koreluje s množstvím příspěvků, a tím ovlivňuje kompletnost dat (Zielstra et Zipf, 2010; Haklay, 2010) nebo poziční přesnost (Haklay et al., 2010).

V dalších empirických studiích bylo zjištěno, že sociální deprivace a nedostatečná socio-ekonomická situace v území následně ovlivňuje kompletnost a poziční přesnost v datech OpenStreetMap (Antoniou, 2011; Haklay et al, 2014). Podobně naznačují Girres et Touya (2010), že faktory, jako vysoký příjem a nízký věk populace, ústí ve vyšší množství příspěvků. Budhathoki et Nedović-Budić (2010) ve své práci podotýkají, že motivace uživatelů může ovlivnit jimi generované příspěvky.

OpenStreetMap data pro Prahu porovnávaná v této práci jsou ze zóny s velkým množstvím přispěvatelů s vysokým příjmem oproti zemím třetího světa (kde probíhá velké množství různých OSM projektů), proto se dá předpokládat, že budou relativně kompletní a pozičně přesná.

2.2.3 „Bias“ ve VGI datech

Prostorový *bias* (Sullivan et al. 2009; Munson et al. 2010) ve VGI datech je v určitých případech jeden z hlavních problémů, které musíme zohlednit, pokud vyhodnocujeme kvalitu prostorových dat (Flanagin et Metzger 2008; Bonter et Cooper 2012; Foody et al. 2013). Tento výraz by se dal přeložit jako „zaujatost“, „tendence“ nebo „předsudek“ uživatele.

Prostorová data mohou být „tendenční“ (*biased*) kvůli faktu, že pozorování dobrovolníků jsou oportunistická a *ad-hoc* ve své podstatě. To znamená, že je uživatel vytváří či poskytuje pouze za jedním účelem, jen z jeho úhlu pohledu nebo pouze tak, jak umí či chce a nezohledňuje další fakta.

Zhu et al. (2015) ve své práci ukazují *bias* na příkladu sbírání dat od místních obyvatel, kteří sledují divokou zvěř ve svém okolí. Místní obyvatelé však nesledují zvěř cíleně a všude, kde se zvířata pohybují. Na místo toho podávají informace o výskytech zvířat pouze tehdy, pokud ji zahlédnou, když jsou na cestě někam, a jdou dělat něco jiného. Cesty místních obyvatel nejsou náhodné ani pravidelné v tom, jak byly navrženy, ale jsou navrženy *ad-hoc*. V důsledku je pak pravděpodobné, že jsou prostorová data od místních obyvatel o pozorování zvěře ovlivněná faktorem *bias*. A pokud není takový prostorový *bias* správně zachycen a vyhodnocen, může nepříznivě ovlivnit závěry vytvořené na základě těchto VGI dat (Ponder et al. 2001; Reddy et Da 2003; Graham et al. 2004; Kadmon et al. 2004; Leitão et al. 2011). *Bias* může být dále ovlivněn takovými faktory, jako jsou: přístup k internetu, znalost jazyka, časové možnosti uživatele nebo jejich technické schopnosti (Ritchie et al., 2007). *Bias* se dle Antoniou et Skopeliti (2015) vyskytuje na všech úrovních granularity dat (od místních dat až po národní databáze) a ke stejnému výsledku došel i Quattrone et al. (2015), který zkoumal *bias* v datech na úrovni států. Pro vyhodnocení *bias* v datech na úrovni měst a obcí jsou podle Quattrone et al. (2015) však potřeba další výzkumy.

Antoniou et Schlieder (2014) také píší o problému shlukování uživatelů. Nejenom, že jednotliví uživatelé často preferují území jim známé a jimi oblíbené (*bias*), ale podobný problém lze nalézt i ve větším měřítku; tedy že některá území jsou často aktualizovaná, zatímco území s menším množstvím uživatelů či méně “oblíbená” nejsou aktualizovaná tak často.

Dalším problémem, nebo můžeme říci i výzvou, je hledání opatření či kompenzací *biasu* v prostorových datech. Pokud například pozorovali místní obyvatelé na nějakém místě zvěř velmi často, nemusí to nutně znamenat, že onu lokaci zvířata preferují. Může to například znamenat, že je místo dobře viditelné z několika cest, a proto pokaždé, když se tam zvíře vyskytne, tak ho vidí několik místních obyvatel zároveň. A na druhou stranu, pokud se zvířata vyskytnou na jimi oblíbeném místě, ale to je hůře viditelné pro místní obyvatele, tak je šance, že je někdo uvidí a onu skutečnost nahlásí, menší. Tento problém je pak potřeba nějak kompenzovat a řešit.

VGI jsou často *tendenční*, kvůli faktu, že ne každé místo na zemském povrchu je stejně dobře viditelné z cest, po kterých se místní obyvatelé pohybují. Obzvlášť pokud vezmeme v potaz nepravidelnost a nenáhodnou distribuci lidmi vytvářených cest a variabilitu terénních podmínek. Zhu et al. (2015) řešili problém tak, že nakumulované množství svědectví o umístění nějakého prvku (tedy frekvenci pozorování), použili ve výpočtech jako váhu, která kompenzuje *bias* způsobený různou viditelností z různých cest.

Kalkulace nakumulovaných svědectví (či „nakumulované viditelnosti“) každého umístění prvku zahrnuje tři kroky. Nejdříve se zpracuje analýza *viditelné oblasti* pro maximální viditelnost každého prvku z každé cesty (odhaduje se podle zkušenosti z praxe). Poté se odpovídající frekvence, se kterou uživatelé chodili po určité cestě, vynásobí s viditelností prvku z každé cesty. A za třetí se spočte suma viditelností ze všech cest dohromady, čímž získáme *nakumulovanou viditelnost* umístění tohoto prvku. Tato hodnota pak nepřímou úměrou vyvažuje přítomnost prvku s kumulativní viditelností daného prvku (Zhu et al., 2015).

2.2.4 Nepřesnosti v umístění získaných dat

VGI mají v některých případech vysokou přesnost umístění - například v datech z OpenStreetMap (Antoniou, 2011; Haklay et al, 2014; Girres et Touya, 2010) -

proto nemusí být přesnost umístění těchto dat až tak důkladně řešena. OSM data, která se vyskytují poblíž lidských sídel (např. cesty, ulice a budovy) a fyzické geografické prvky (např. řeky a jezera) mají vysokou přesnost zaměření, protože byly tyto prvky kontrolovány či digitalizovány z přesně georeferencovaných obrázků dálkového průzkumu s vysokým rozlišením. Avšak prostorové umístění VGI může být nepřesné, v závislosti na geografických prvcích ve sledovaném území, případně na dostupnosti zaměřovacích technologií (Zhu et al., 2015; Antoniou, 2011).

V případech, kde nejsou dostupné dostatečně kvalitní zaměřovací technologie nebo se geografické prvky často mění, je náročné zaručit dostatečnou kvalitu VGI dat. Například můžeme dostat informaci, že místní obyvatel, od kterého získáváme data v neprozkoumané oblasti, zahlédl zvěř „na tom a tom území“. To může být samozřejmě zaznačeno jako polygon (mnohoúhelník), ale neznamená to, že se zvěř vyskytla na každém místě v daném polygonu a rozhodně se nevyskytuje se stejnou pravděpodobností na různých místech v daném polygonu.

2.3 Povaha a motivace uživatelů VGI

2.3.1 Typy a charakteristika uživatelů

Příznivci VGI, jimiž jsou např. O'Reilly (2005), Cook (2008) a Tapscott et Williams (2010), tato uživateli vytvářená data označují jako revoluci, a většinou poukazují na jejich benefity. Naopak kritici VGI, jako McHenry (2004), Lanier (2006) a Keen (2008), jsou stejně tak hlasití ve svých názorech a „crowdsourcing“ reprezentují jako znepokojivý trend, který zvýší vliv amatérů na úkor legitimních odborných a profesionálních mediálních organizací.

V této polarizované debatě se autoři shodnou, že uživatelé mohou být rozděleni do pěti překrývajících se kategorií:

a) Začátečník - člověk bez formálního zázemí v oboru, který však má zájem, čas a ochotu nabídnout názor na danou problematiku.

b) Poučený amatér - někdo, kdo v sobě objevil zájem o problematiku, začal číst související literaturu, konzultovat to s kolegy a experty na specifická témata. Experimentuje s aplikací tématu, získává zkušenosti a předmět zájmu mu je blízký.

c) *Amatér expert* - někdo, kdo už může mít velké zkušenosti s problematikou, zaniceně ji praktikuje při různých příležitostech, ale nevydělává si tím na živobytí.

d) *Profesionál* - člověk, který studoval daný předmět a praktikuje ho. Svými zkušenostmi a znalostmi v oboru si vydělává a může být žalován, pokud by bylo dokázáno, že výsledky jeho práce, jeho názory a/nebo doporučení byly neadekvátní, nesprávné nebo urážlivé.

e) *Autorita v oboru* - někdo, kdo dlouhodobě a široce praktikoval daný předmět a ví se o něm, že dokáže podat detailní a přesné informace o problematice, a v minulosti předložil vysoce kvalitní produkty práce. O tuto reputaci se opírá a je možné, že by svou pověst ztratil, pokud by jeho důvěryhodnost i jen dočasně klesla.

Články od výše zmíněných podporovatelů VGI charakterizují průměrného uživatele obvykle jako *amatéra experta* nebo při nejhorším schopného *poučeného amatéra* a mají sklon ignorovat problém s právní odpovědností. Kritici VGI naopak poukazují na následky toho, pokud jsou přijata data od nekompetentního *začátečníka* nebo desinformovaného *poučeného amatéra* (Coleman et al., 2009).

Tito uživatelé dále spadají do několika kategorií:

a) *Ovlivněný trh* - jeho cílem je přispívat do komerční databáze nebo služby (např. TomTom, GARmin, Navteq, aj.)

b) *Sociální síť* - přispívá na stránky jakými jsou např. OpenStreetMap, Platial.com, a další

c) *Občanský/vládní* - svými příspěvky se prezentuje jako dotčený občan dané vesnice nebo města a podporuje nějaký tamní projekt nebo je to člen organizace v oblasti životního prostředí nebo člen skupiny bojující za práva zvířat

Je také potřeba zmínit, že není zaručené, že v čím vyšší kategorii se přispěvatel nachází, tím jsou data kvalitnější. Výše zmíněný *profesionál* může být velmi schopný v zacházení s příručním GPS, ale jeho znalost místního okolí a přírodních poměrů může být omezená. Na druhou stranu uživatel *začátečník* může vědět minimum o polohovacích systémech, ale za to být velmi dobře obeznámen s poměry v okolním prostředí, které se právě mapuje (Coleman et al., 2009).

2.3.2 Motivace uživatelů

Goodchild (2007b), uvádí, že propagace sama sebe je důležitým motivačním faktorem při aktivitě na internetu a ve své extrémní podobě může vést až k tomu, že uživatel o sobě uvádí na internetu úplně vše. A i přestože uživatel sdílí tato data pouze se svými známými a příbuznými, jsou často dostupná všem.

Pokud chceme lépe pochopit, proč jednotlivci přispívají geografickými informacemi, můžeme se poučit z projektů jako je Wikipedie a *freesource* (program zdarma ke stažení) nebo *open-source* (program volně přístupný ke stažení i k úpravě jeho součástí) komunitami uživatelů (*freesource* a *open-source software* - FOSS).

Empirické studie od Anthony et al. (2005), Kuznetsov (2006) a Schroer et Hertel (2009) čerpají z výzkumů sociálních hnutí, aby interpretovali, proč lidé přispívají na Wikipedii. Výzkum od Lakhani et Wolf (2005) poukazuje na to, že vývojáři FOSS jsou motivováni mírně jinak, než jak bylo zjištěno o uživatelích Wikipedie. Cook (2008) poskytuje informace obecně o tom, proč lidé aktivně přispívají všemi druhy online UGC.

Sjednocením a sumarizací těchto článků o uživatelích Wikipedie, FOSS a jiných UGC získáme následující údaje o jejich motivačních faktorech.

- a) *Altruismus* - přispívání pouze za účelem obohacení ostatních bez jistoty či vidiny zlepšení uživatelovy osobní situace.
- b) *Osobní nebo profesní zájem* - vytvoření příspěvku v rámci už existujícího pracovního poměru nebo součástí probíhajícího osobního projektu.
- c) *Intelektuální stimulace* - zvýšení uživatelových technických schopností, znalostí či zkušeností získaných skrz příspěvek.
- d) *Protektce nebo vylepšení osobního zájmu* - kde poskytnutí praktického řešení problému, který sdílela komunita, dává okamžitou návratnost uživateli v rámci společném zájmu.
- e) *Společenská odměna* - tím, že je uživatel součástí větší internetové sítě nebo komunity, získá - skrz kolaboraci, diskusi a vývoj - „pocit sounáležitosti

a společného cíle, který všechny spojuje“ (Kuznetsov, 2006) a pobídne ho to k další práci a následnému sdílení.

f) *Navyšování vlastní reputace* - příležitost pro registrované uživatele např. Wikipedie, aby si vytvářeli onlineové identity, které bude někdo respektovat, věřit jim a vážit si jich. Jejich pocit vlastní důležitosti se tím navýší.

g) *Prostředí pro kreativní a nezávislé sebevyjádření* - uživateli je nabídnuto místo, kde bude moci prezentovat svou kreativitu.

h) *Hrdost na své místo* - přidání informací o jedincově vlastní skupině nebo komunitě může být dobré pro vztahy s veřejností, turismus, ekonomický vývoj nebo prostě prokázání toho, že je jedincova ulice nebo podnik „na mapě“ (Coleman et al., 2009).

2.3.2.1 Motivace uživatelů editovat své příspěvky

Pokud už uživatel nějaká data vytvoří, není zaručené, že se jim bude věnovat i nadále a pravidelně je bude aktualizovat. Antoniou et Schlieder (2014) se ve své studii OSM zabývali význačnými územími v Londýně a aktivitou jejich uživatelů. Ukázalo se, že neexistuje vazba mezi tvůrci dat a prostorovými prvky. Ve skutečnosti se pouze 10% z těchto vysoce produktivních uživatelů vrací k cca 20% prvků, které v minulosti vytvořili. V práci je dále zmíněno, že některé části OSM nejsou dostatečně často aktualizované, přestože se většinová aktivita příspěvčatelů soustředí spíše na aktualizaci součástí, než na vytváření nových.

2.3.3 Negativní motivační faktory příspěvčatelů

Existují i negativní motivační faktory, které nejsou o nic méně významné, a také je potřeba je uvažovat. Ne všichni uživatelé musí zákonitě přidávat pouze objektivní a spolehlivé informace. Motivační faktory těchto uživatelů jsou jednodušší na identifikaci, než faktory pozitivní (Coleman et al., 2009).

a) *Potížista*: Škodící osoby či vandalové. Doufají v to, že vytvoří zmatení nebo skepsi ostatních uživatelů tím, že nahradí legitimní příspěvky nesmysly nebo urážlivým obsahem. Viegas et al. (2007) a Priedhorsky et Terven (2007) v jejich

pracích ukázali příklady z empirického výzkumu a následně charakterizovali a určili množství těchto příspěvků na Wikipedii a nabízí postupy, jak je rychle opravit.

b) Agenda: Nezávislé osoby nebo představitelé motivovaní postoji v dané komunitě, organizaci nebo kauze. Vysledováním jednotlivých uživatelů Wikipedie k jejich specifické IP adrese je nově vytvořený software WikiScanner indetifikoval a následně popsal praktiky konkrétních korporací, vládních institucí a zainteresovaných skupin v systematickém vytváření přehnaně zaujatých, nesprávných a/nebo zavádějících úprav Wikipedie. Ti záměrně vytvářeli příspěvky, které mluvili pouze o nich, členech konkrétní skupiny nebo jejich agendě (Borland, 2007).

c) Nenávistný a/nebo kriminální záměr: Jednotlivci s nenávistnými (a pravděpodobně nelegálními) zájmy konající za účelem zvýšení osobního zisku. V roce 2008 například neznámé osoby prolomily ochranu webu iReport.com a vložili tam nepravdivý příspěvek o hospitalizaci Steve Jobse, čímž způsobili krátkodobou fluktuaci akcií Applu, z níž poté benefitovali (Cohen, 2008).

2.3.4 Kritika VGI

Odborníci na komunikaci a informatiku upozorňují na to, že jakákoli informace, která je zveřejněna na webu, nemusela projít přes profesionála a v důsledku toho mohou být digitální informace náchylnější k milnému uspořádání, nekompletnosti, nepřesnosti či zastaralosti. (Flanagin et Metzger 2007; Metzger et al. 2003; Rieh et Danielson 2007).

Jak poznamenává Tulloch (2007) k programům, jejichž prostřednictvím jsou informace zpracovávány a v nichž jsou sledovány: „Mnoho z těchto nových aplikací je vyvíjeno jedinci, kteří nemají žádnou pracovní minulost v akademické sféře nebo zájem o tradiční akademické postupy, často ani nepřímo.“ Kombinace velkého množství digitálních informací a jejich snadné přístupnosti podnítily obavy ohledně jejich důvěryhodnosti, protože, jak tvrdí Rieh et Danielson (2007), kvůli této kombinaci je těžké zjistit, kdo data vytvořil, což následně vede k nejistotě v tom, jestli je možné jim věřit. K podobným závěrům došli i Fritch et Cromwell (2001) a Metzger (2007).

Kvalita těchto dat je tedy kolísavá, data jsou špatně zdokumentovaná, nejsou dodržovány vědecké principy při jejich získávání a pokrytí povrchu těmito daty je nekompletní. Goodchild et Li (2012) si myslí, že VGI data jsou jako podklad při vědeckém výzkumu nedostačující, ale že i tak mohou mít svou roli v brzké a průzkumné fázi výzkumu, a dají se podle nich vytvářet hypotézy. Je také jasné, že pokud by se dala zajistit vyšší kvalita a důvěryhodnost VGI dat, plynuly by z toho další výhody a možnosti využití.

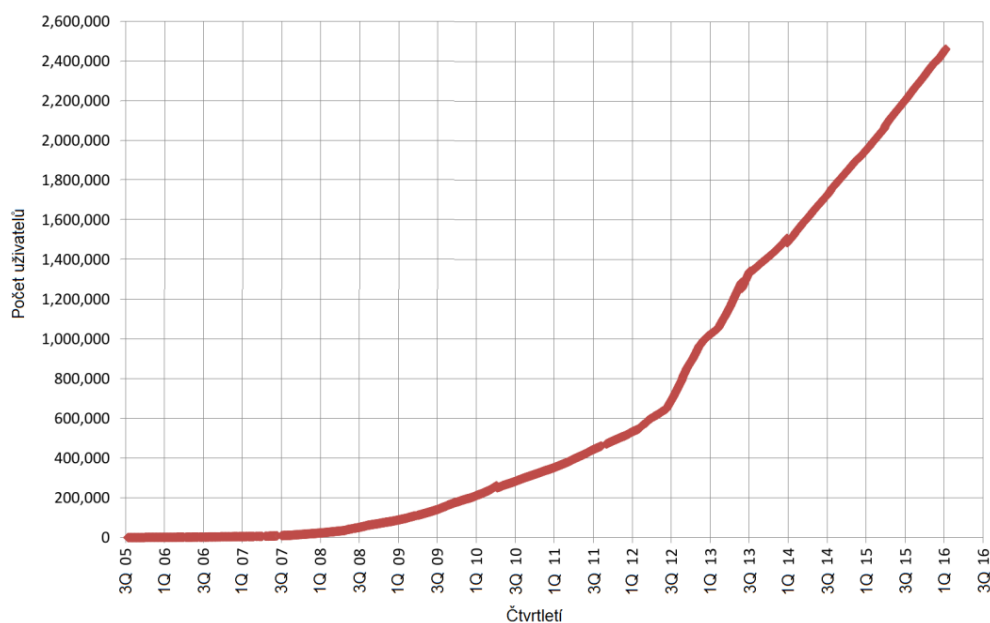
2.4 OpenStreetMap a další VGI Projekty

2.4.1 OpenStreetMap

Úplně první verze OpenStreetMap byla spuštěna v červenci roku 2004 ve Velké Británii a postupem času se stal nejpoblárnějším a nejúspěšnějším projektem, který využívá kolektivní spolupráce uživatelů spolu s koncepcí open-source.

OSM obsahuje geografická data pro každou část světa, ve které je alespoň částečný dobrovolný zájem od internetových uživatelů (Haklay et Weber 2008). Tato uživateli generovaná geografická data mají, jak již bylo zmíněno výše, velmi často vyšší přesnost umístění, než produkty státních mapovacích agentur (Haklay et al., 2010; Zhu et al., 2015).

Obrázek 2: Vývoj počtu registrovaných uživatelů OSM (OSM Wiki, 2016)

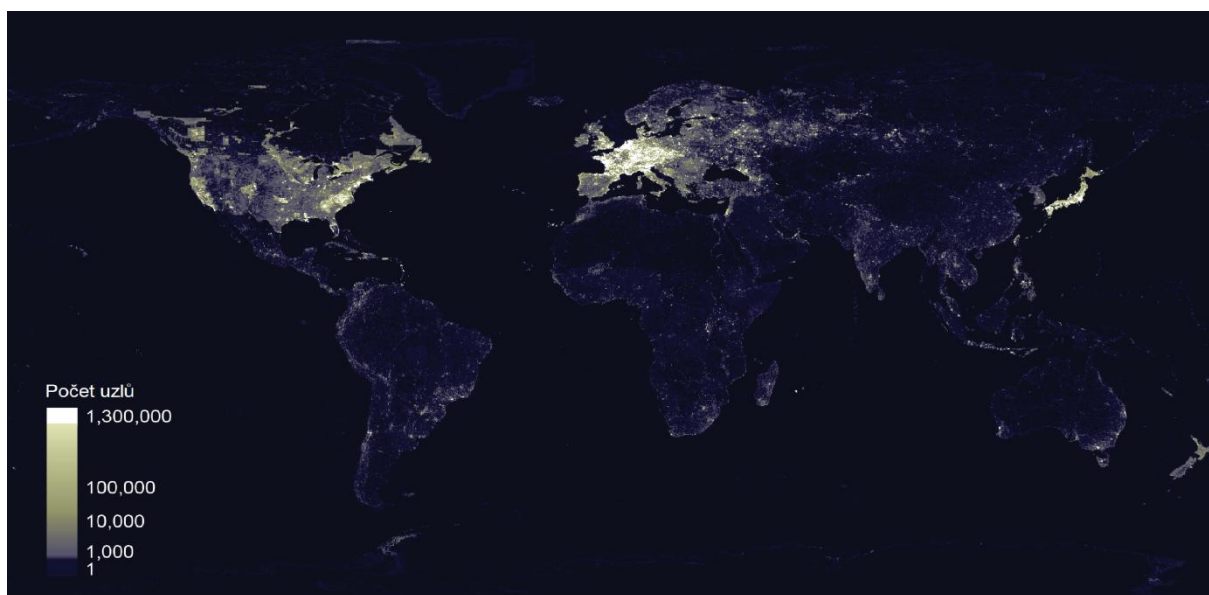


OSM je závislý na GPS datech od jednotlivců, kteří procházejí specifická území, aby byly vytvořeny dostatečně obsáhlé a aktuální mapy (Flanagin et Metzger, 2008). Jednotlivec je v tomto případě využit jako přepravce snímačů, jež zaznamenávají místní prostředí (Goodchild 2007b). Poté, co je zpracováno dostatečné množství datových bodů, mohou uživatelé sledovat, upravovat a používat výsledné mapy, v mnoha případech i velmi kreativně; například se z dat mohou renderovat 3D mapy nebo přidat sjezdovky pro lyžaře do horských území (Flanagin et Metzger, 2008).

Zakladatelem a správcem OSM je OpenStreetMap Foundation, jež je neziskovou organizací registrovanou ve Velké Británii. Projekt OpenStreetMap je opravdu celosvětovým fenoménem a je výsledkem společného úsilí více než dvou milionů uživatelů, jejichž postupný nárůst ukazuje obr. č. 2.

Není překvapením, že nejvíce údajů o území obsahují a nejčastěji aktualizovaná jsou území USA, Evropy a Japonska, jak zobrazuje obrázek č. 3.

Obrázek 3: Hustota prvků OpenStreetMap ve světě (IGOI, 2014)

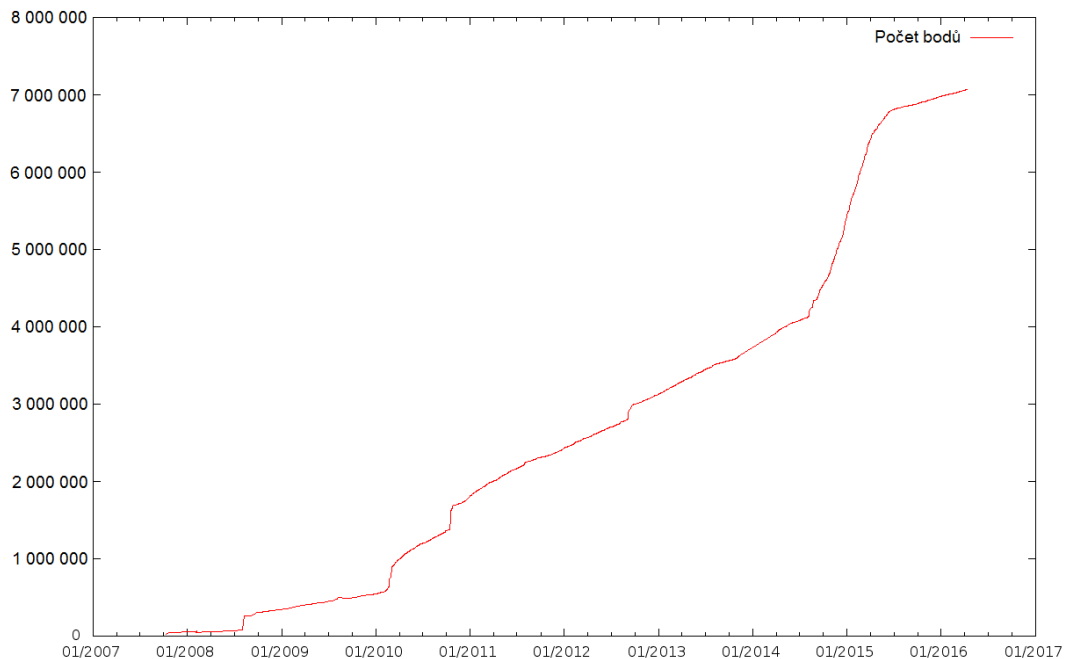


2.4.1.1 *OpenStreetMap v České republice*

V České republice se začal projekt OSM rozvíjet v průběhu roku 2008 a jeho převážná část je vytvořena čistě uživateli. To znamená, že neproběhlo nikterak velké množství importů z jiných geografických projektů. Prakticky na úplném počátku, v srpnu roku 2008, proběhlo „zalesňování ČR“ (GPSFM, 2008a) ze serveru www.uhul.cz, které se neobešlo bez problémů - některé vrstvy se překrývaly s vrstvami již vytvořenými či se špatně vykreslovaly, ale v Česku je komunita, která se OSM věnuje, velmi aktivní a projekt se dál úspěšně rozvíjel. Úplnou síť silnic 1. a 2. třídy poskytla Česku společnost Help Service - Remote Sensing a bylo potřeba domapovat hlavně silnice 3. tříd. Většina uživatelů v této době mapovala přes ortofotomapy UHUL či Yahoo maps (GPSFM, 2008b). Mapování silnic 3. tříd bylo dokončeno na začátku roku 2009.

Zpřístupnění leteckých snímků a další dat z komerčních či veřejných zdrojů přispělo ke zrychlení a zpřesnění mapových podkladů. Koncem roku 2010 se rozhodl Bing, patřící společnosti Microsoft, že uvolní používání jejich satelitních snímků pro účely obkreslování v OpenStreetMap. V roce 2014 poskytla pro mapování družicové snímky společnost Mapbox. Podle českých stránek projektu OSM (OSM CZ FAQ, 2016) mapuje většina uživatelů právě podle těchto zdrojů. Společnost Mapbox získala fotografie od firmy DigitalGlobe, se kterou již dříve spolupracovala na získávání snímků pro potřeby Humanitarian OpenStreetMap Teamu. V dubnu téhož roku proběhl poloautomatický import Adres z RUIAN (Registr územní identifikace, adres a nemovitostí). Cílem importu bylo především doplnění chybějících adresních bodů a provázání nových, již vytvořených bodů s položkami v databázi RUIAN (GPSFM, 2014). Na obrázku č. 4 je zobrazen vývoj celkového počtu bodů v OpenStreetMap v České republice.

Obrázek 4: Vývoj počtu bodů v OpenStreetMap v České republice (OSM ČR, 2016)



2.4.2 Další VGI mapové projekty

2.4.2.1 Wikimapia

WikiMapia je projekt spojující mapové informace z Google Maps s texty vkládanými pomocí technologie Wiki. Projekt byl založen 24. května 2006 ruskými internetovými podnikateli Alexandrem Korjakynem a Jevgenijem Saveljevem. Cílem projektu bylo vytvořit mapový zdroj, který bude zobrazovat celý svět. Koncem roku 2015 bylo popsáno 25 000 115 mapových objektů. I když není vyžadována registrace, bylo k datu 12. února 2015 na stránkách projektu zaregistrováno 1 579 206 uživatelů z celého světa (QUORA, 2015). Veškerý uživateli vytvořený obsah se stává intelektuálním vlastnictvím WikiMapie a je k dispozici pro nekomerční využití skrze rozhraní WikiMapia API (WikiMapia, 2016).

2.4.2.2 Google Map Maker

Google Map Maker (GMM) je v České republice dostupný od 16. července 2013. GMM byl původně určený hlavně pro rozvojové země, tedy tam, kde neměl Google profesionální mapové pokrytí. Později firma začala aplikaci uvolňovat i ve vyspělých zemích, kde naopak mohou zkušení uživatelé vylepšovat některé detaily a v předstihu aktualizovat chybějící bílá místa. Svými funkcemi se GMM podobá

OSM a dle komentářů jeho uživatelů, kteří s OSM v minulosti pracovali, je i práce v GMM obdobná (Čížek, 2013).

McDonough (2013) tvrdí, že je OSM snadněji uchopitelný než GMM a je pomocí něj jednodušší pro běžného uživatele najít další lidi ve svém okolí, kteří mají také zájem o geografii. GMM má své výhody, ale celková zkušenost s ním se podobá pohybování se v „městě duchů“, než v prosperující komunitě lidí.

2.4.2.3 Flickr

Přestože to nebylo původním záměrem tohoto projektu, webová stránka na sdílení fotografií - Flickr.com - také přispěla obrovským množstvím prostorových informací díky fotografiím specifických lokalit, jež nahráli uživatelé tohoto projektu. Potom co se postupně množství informací na stránce nahromadilo, tvoří obrázky relativně bohatý zdroj o umístění geografických objektů (Flanagin et Metzger, 2008).

3 Metodika

3.1 Vstupní data

Referenční autoritativní datová sada, jíž je v případě této práce Digitální mapa Prahy získaná z Geoportálu Praha - <http://www.geoportalpraha.cz/>, byla při vyhodnocování kvality dat považována za přesný odraz současného stavu, jelikož je Institut plánování a rozvoje hlavního města Prahy (který datový zdroj poskytuje) oficiální a důvěryhodnou institucí, kterou zřizuje hlavní město Praha (IPR Praha, 2016). Tento zdroj byl následně porovnán s OSM v geografickém informačním systému (GIS) na základě vybraných elementů hodnocení kvality.

OSM lze stáhnout ve formátu .osm přímo z oficiálních stránek, avšak pouze v omezené velikosti, aby nebyl příliš zatěžován hlavní server. Pro potřeby této práce však bylo potřeba stáhnout data pro celou Prahu, aby bylo následně možné vybrat nejvhodnější oblasti pro vyhodnocení kvality dat. Na webových stránkách <http://planet.osm.org/>, které jsou pravidelně aktualizovány a tamní data podléhají stejné licenci, jako data stažená z oficiálního serveru, je možné stáhnout najednou OSM data pro celou planetu, pro jednotlivé kontinenty, státy či metropolitní oblasti (pro Českou republiku jsou dostupné oblasti Praha a Brno).

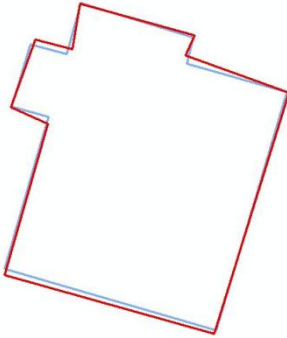
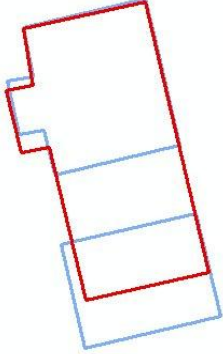
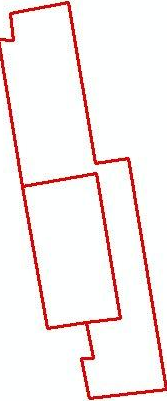
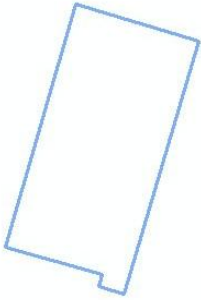
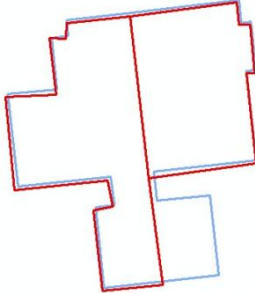
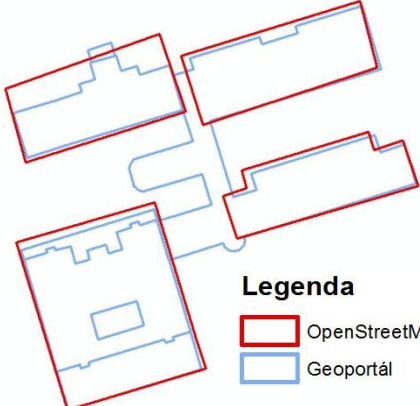
V této práci byl pro porovnání a vyhodnocení datových sad použit program ArcGIS od firmy ESRI (ESRI, 2016). OSM v něm však nejdou nativně otevřít a je tedy potřeba je konvertovat na *shapefile*, se kterým lze následně v ArcGIS pracovat. Toho bylo docíleno pomocí *open-source* programu QGIS.

3.2 Relace mezi daty

Oba datové zdroje obsahují velké množství polygonů, které mezi sebou mohou mít různé vztahy (relace). Jedná se vlastně o sémantickou přesnost dat, tedy zda se data v jednom zdroji vyskytují i v druhém, případně zda jsou nějak pozměněna oproti skutečnosti. V článku Fan et al. (2014), který vychází ze CityGML standardů pro OGC (Open Geospatial Consortium) (Gröger et al. 2008), najdeme následujících šest základních typů relací mezi daty (obr. č. 5):

- **1:1 relace:** Budova je sémanticky správně vyznačena.
- **1:N relace:** Jedna budova v OSM je agregátem N budov v referenčních datech.
- **1:0 relace:** Budova v OSM je sémanticky špatně zapsaná. Což znamená, že ve skutečnosti budovou není.
- **0:1 relace:** Tento případ je opakem relace 1:0. Tedy že se v OSM budova nevyskytuje.
- **N:1 relace:** Jedna budova v OSM je pouze částí budovy v referenčních datech.
- **N:M relace:** Budovy jsou nepřesně zakresleny s ohledem na sémantickou přesnost.

Obrázek č. 5: Možné relace mezi dvěma datovými sadami

Relace	1:1	1:N	1:0
Ilustrace			
Relace	0:1	N:1	N:M
Ilustrace			

Legenda
 OpenStreetMap
 Geoportál

3.3 Práce s daty

Data je před samotným porovnáním potřeba vhodně upravit. OSM data nejsou zobrazována v koordinačním systému *S-JTSK Krovak East North* (S-JTSK), který je běžný v České republice, ale ve *World Geodetic System 1984* (WSG1984). Data převedeme tak, že v *ArcToolboxu* otevřeme *Data Management Tools, Projections and Transformations* a následně ve *Feature* funkci *Project*. Tím vytvoříme úplně nový *shapefile*, který je potřeba do mapového dokumentu znovu nahrát.

Jelikož jsou pro potřeby práce porovnávány půdorysy budov, je potřeba vybrat z OSM pouze takové polygony, které ve skutečnosti představují budovy. V *atributové tabulce* se nachází sloupec „*building*“. Pro další práci s daty vytvoříme nový sloupec příkazem *add field*, který bude mít typ *double* (číselný). V novém sloupci (pojmenovaný *TruBuild*) pak pomocí funkce *Field Calculator* vytvořit číselnou hodnotu pro ta data, která nerepresentují budovy, tedy: $[building] = ""$. V případě, že polygon představuje budovu, se vypíše hodnota 0, pokud ne, tak se vypíše hodnota -1.

Dále byla použita funkce *Select by attributes*, do které byl následně zadán příkaz „*TruBuild*“ = 0. Poté se z vybraných *features* vytvoří samostatná vrstva: v rozbalovacím menu původní vrstvy, na řádku *Selection*, příkaz *Create layer from selected features*.

Nakonec je potřeba OSM vrstvu oříznout funkcí *clip*, aby obsahovala pouze data o budovách v kraji hl. m. Praha. S touto vrstvou budeme dále pracovat při vyhodnocování jednotlivých kvantitativních faktorů přesnosti OpenStreetMap.

3.4 Komplettnost

Komplettnost značí množství dostatek či spíše nedostatek OSM dat oproti autoritativním datům. Pro vyhodnocení komplettnosti dat používají odborníci různé metody. Mezi nejčastější patří:

a) *porovnávání v mřížce - grid* (Haklay, 2010; Zielstra et Zipf, 2010; Ludwig et al., 2011; Ciepluch et al., 2011; Forghani et Delavar, 2014),

b) *porovnání množství vlastností - number of features* (Girres et Touya, 2010; Jackson et al., 2013),

c) *porovnání celkové délky či celkové plochy* (Kounadi, 2009; Girres et Touya, 2010; Koukoletsos et al., 2011; Fan et al., 2014; Arsanjania et Vaz, 2015; Kalantari et La, 2015),

d) *„míra“ kompletnosti - completeness measure* (Mashhadi et al., 2014), případně

e) *kompletnostní index* (Arsanjani et al., 2015b).

3.4.1 Kompletnost OpenStreetMap v Praze

Pro vyhodnocení kompletnosti byla vybrána metoda porovnání OSM a Digitální mapy Prahy v mřížce - *grid*.

K vyhodnocení porovnání absolutního počtu budov mezi oběma datovými sety bylo potřeba navrhnout polygonovou síť, která obsáhne celou Prahu, a spojit ji s bodovou nebo polygonovou vrstvou s budovami. Výsledný *layout* pak po propojení a správném nastavení v každé své buňce obsahuje údaj o množství budov v ní.

Síť byla vytvořena pomocí funkce *fishnet*. Pro optimální přehlednost výsledného *layoutu* byly vytvořeny čtvercové buňky o straně 1000m, referenční vrstvou byla hranice Prahy. V atributové tabulce každé vrstvy s údaji o budovách, tedy ve vrstvě OSM a vrstvě Geoportálu, vytvoříme nové pole (*field*), do kterého vypíšeme požadovanou hodnotu pro každou budovu (v tomto případě hodnotu 1).

Poté je potřeba na vrstvu *fishnet* napojit data o budovách (nikoliv napojovat síť na data), čehož docílíme pomocí funkce *join*. Zvolíme možnost *Join data from another layer based on spatial location*, vybereme požadovanou vrstvu s budovami a následně zaškrtneme pole *sum* (data sečteme). Stejný postup zopakujeme pro druhou vrstvu budov. Ve vzniklých polygonových vrstvách nastavíme požadované zobrazení dat (*properties, symbology, graduated colors*).

Pro výpočet ploch v OSM a datech z Geoportálu byla použita totožná metoda, jako v případě porovnání absolutních množství budov, ale v atributové tabulce musel být předtím vytvořen nový sloupec (typ *double*) a pro každou budovu vypočtena její plocha. Toho bylo docíleno pomocí funkce *Calculate Geometry*, následně zvolena možnost *property - area* a požadovaný souřadnicový systém (S JTSK Krovak East North). Následně již bylo možné pomocí *join* napojit data na vytvořenou síť *fishnet* a vhodně zobrazit.

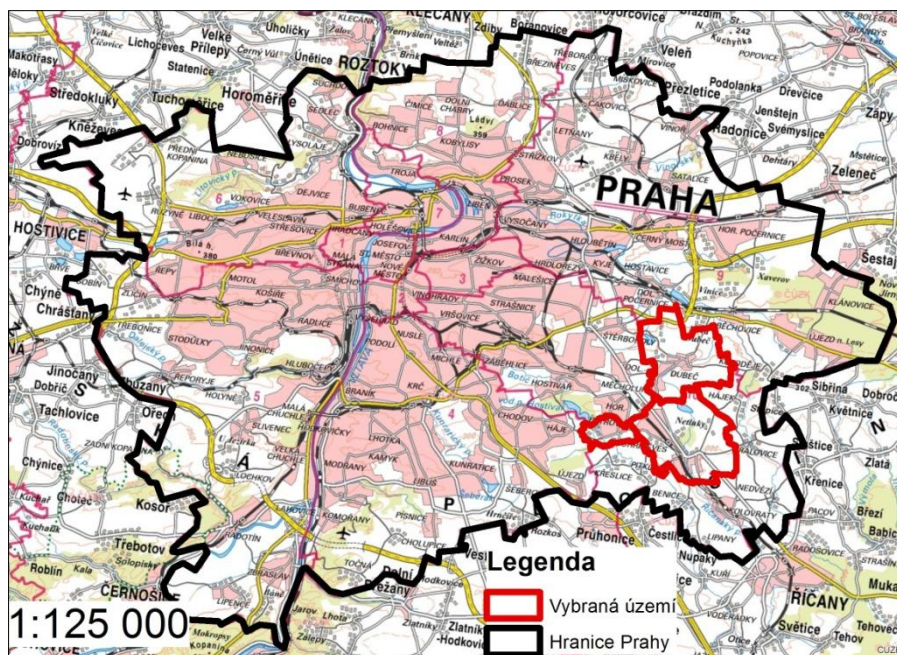
3.5 Sémantická přesnost

Sémantická přesnost značí, zda budova zobrazená v OSM existuje i ve skutečnosti a zda je správně označená. U dat z Geoportálu Praha přepokládáme 100% sémantickou přesnost. To znamená, že každá budova z autoritativního datového balíčku Geoportálu existuje i ve skutečnosti.

3.5.1 Sémantická přesnost OpenStreetMap v Praze

U dat z OpenStreetMap byla sémantická přesnost dat provedena vizuální inspekcí budov ve vybraných částech Prahy a jejich porovnání s ortofotomapou. Vybranými územími byly MČ Uhřetěves, Dubeč a Petrovice (viz obrázek č. 6), v nichž se vyskytuje 5535 OSM objektů (budov).

Obrázek č. 6 - Vybraná území pro vyhodnocení sémantické přesnosti



3.6 Tvarová přesnost

Pro vyhodnocení tvarové přesnosti dvou sad prostorových dat je potřeba pracovat pouze s prostorovými daty v *relaci 1:1* - tedy polygony, které označují právě jednu budovu v obou sadách. Tvarová přesnost se týká množství vertexů (vrcholů) jednotlivých polygonů a plochy korespondujících polygonů v obou sadách.

3.6.1 Algoritmus k nalezení relací 1:1

Nejdříve je potřeba korespondující polygony nalézt. Někteří výzkumníci (např. Girres et Touya, 2010) ve svých pracích „spojují“ polygony ručně pro dosažené stoprocentní přesnosti výsledků. Tento postup lze však praktikovat pouze pro omezené množství dat (v řádu stovek). V případě této práce, která se týká území celé Prahy, toto není možné, jelikož množství dat přesahuje sto tisíc. Je třeba nalézt vhodnou metodu navázání korelace mezi daty. Toho bylo dosaženo následující pomocí metody překrývajících se polygonů, která vychází z následujícího algoritmu, použitého v práci Fan et al. (2014) a inspirovaného prací Rutzinger et al. (2009):

Nechť je G_{osm} datová sada OSM a G_{ref} datová sada referenční. Pro půdorys budovy bud_{osm_x} v G_{osm} bude vyhodnoceno, zda se prolíná s polygonem v G_{ref} . V případě, že je nalezeno prolnutí s polygonem bud_{ref_y} , vykreslí se polygon $Pol_{intersect}$. Jelikož jsou budovy v OSM ve většině případů digitalizovány na základě mapy Bing (Goentz et Zipf, 2012; OSM CZ FAQ, 2016), je běžné, že se půdorysy budov v OSM a referenční datové sadě mírně liší a jsou částečně zkreslené kvůli použití specifického druhu senzorů mapového zdroje Bing. Práh pro posouzení, zda spolu budovy z obou sad korespondují, závisí z velké části na parametrech mapy Bing použité pro digitalizaci OSM. Rutzinger et al. (2009) objevil, že shoda (korespondence) dvou budov může být způsobena další sousední budovou, pokud je plocha překrytí menší než 30%. Tohoto výzkumu se drží i Fan et al. (2014). Práh pro vyhodnocení korespondence budov byl tedy nastaven na 30%. V případě, že platí

$$\frac{Pol_{intersect}}{\min(\text{plocha}(bud_{osm_x}), \text{plocha}(bud_{ref_y}))} > 30\%$$

pak nalezneme shodu mezi polygony. Tímto způsobem vyloučíme všechny případy *relací 1:0 a 0:1*. Fan et al. (2014) ve své práci dále pracuje i s ostatními typy *relací*, pro potřeby této práce stačí získat *relace 1:1*. Pro odfiltrování většiny¹ zbylých *1:N, N:1 a N:M relací* v datových vrstvách je potřeba z G_{osm} a G_{ref} vybrat pouze polygony, které obsahují celý $Pol_{intersect}$ (vyjma jeho hran, které kvůli tomu, jak jsou polygony v používaných vrstvách zakresleny, mohou být zároveň hranou sousedního polygonu).

3.6.2 Metoda překrývajících se polygonů pro nalezení relace 1:1 mezi budovami v datových vrstvách

Pro již upravené vrstvy OSM a Geoportálu z předchozího výzkumu získáme polygon pro další práci pomocí funkce *Intersect*. Pokud pracujeme s takto velkým objemem dat, tak si s ním program *ArcGIS* neumí poradit v případě, že vrstvy obsahují polygony s parametrem *NULL* či jinou vadou. Je tedy potřeba vrstvu OSM i Geoportálu před následnými kalkulacemi „ošetřit“ pomocí funkce *Repair geometry*. Následně již v *Intersect* vložíme do *Input features* obě vrstvy, *JoinAttributes - ALL*, *Output type* chceme polygon, necháme tedy *INPUT*.

Výsledná vrstva již příhodně obsahuje vypočtené plochy **původních (!)** polygonů ve své atributové tabulce, jelikož jsou ve výsledné vrstvě všechny atributy z původních vrstev zkopírovány. Nyní je potřeba vypočítat plochu nově vytvořených polygonů. Toho dosáhneme znovu tak, že v novém sloupci tabulky pomocí funkce *Calculate Geometry* zadáme *property - area*.

Poté v dalším sloupci *Perc_ploch*, ve kterém vypočteme procentuální podíl korespondujících budov, otevřeme *Field Calculator* a vybereme programovací jazyk *Python*. Necht' je plocha nově vytvořených polygonů ve výsledné vrstvě *Int_Plocha*, plocha polygonů z OSM *OSM_Plocha*, plocha polygonů z Geoportálu *Geo_plocha*, pak:

$$\text{Perc_ploch} = (!\text{Int_Plocha}! / \min(!\text{Geo_Plocha}!,!\text{OSM_Plocha}!)) * 100$$

¹ Vyhodnocení kvality metody v kapitole 4.4.1.

Z výsledných hodnot pak pomocí *Select by Attributes* vybereme všechny polygony, které se shodují ve více než třiceti procentech, tedy "*Perc_ploch*" > 30. Z tohoto výběru polygonů vytvoříme novou vrstvu *Intersect_30percent*.

Intersect_30percent bude referenční vrstvou k datům z OSM a Geoportálu. Použijeme *Select by Location* a vybereme prvky z OSM, *select features from: OpenStreetMap*, *Source layer: Intersect_30percent* a vybranou metodou bude *contains (Clementini) the source layer feature*, z nichž pak vytvoříme novou OSM vrstvu. Analogicky vytvoříme vrstvu pro Geoportál. V těchto vrstvách již máme budovy, které jsou v *relaci 1:1*.

Pro další práci s daty bylo potřeba v atributové tabulce vypočítat počet vertexů u každého polygonu. V novém sloupci je třeba otevřít Field Calculator, přepnout na Python a vložit následující vzorec:

VertexCount = !shape!.pointcount

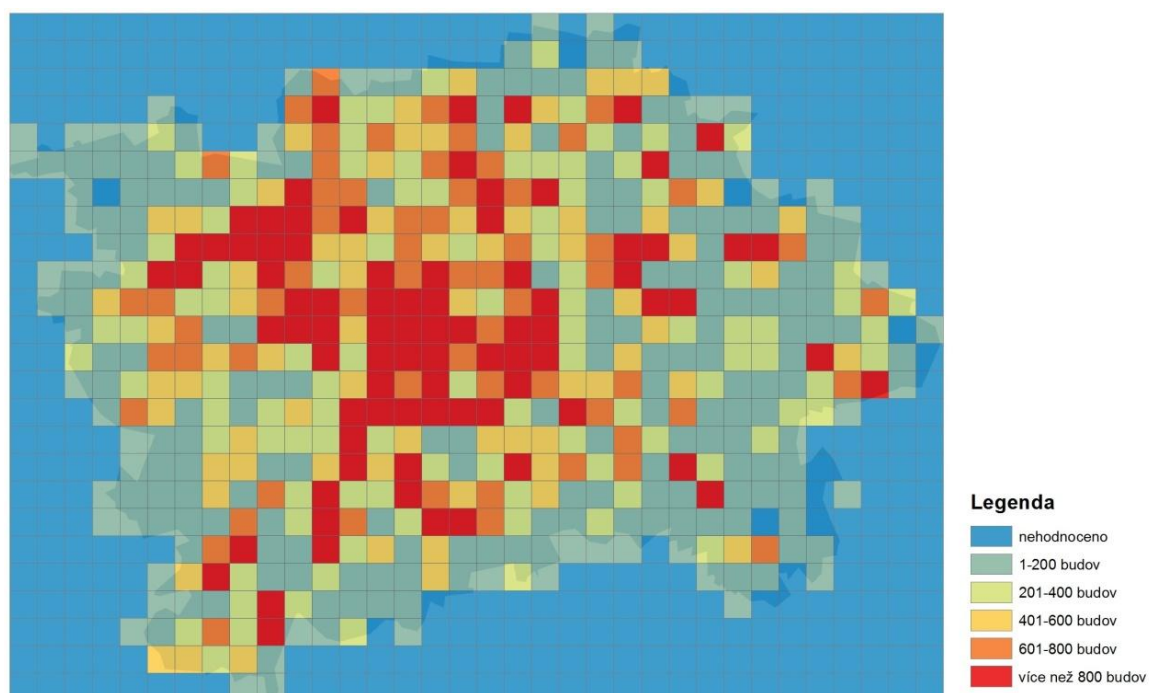
Poté je již možné data vhodně interpretovat. Velké množství informací o polygonech se pak dozvíme přímo v ArcGISu v *Symbology - Quantities*.

4 Výsledky

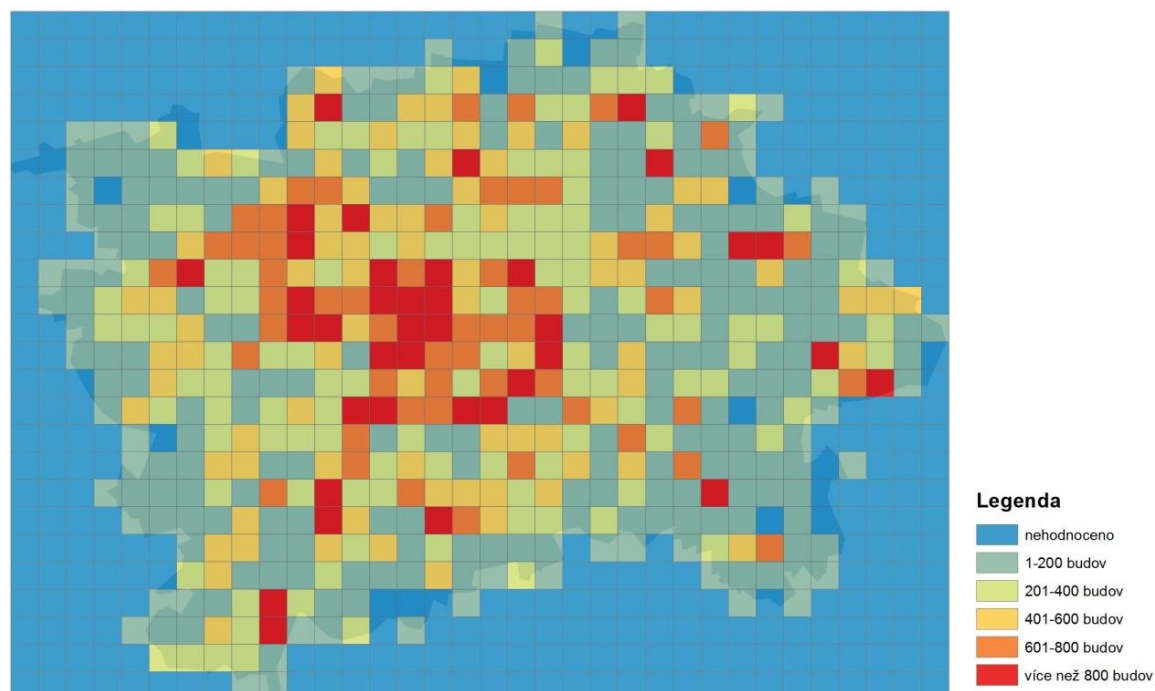
4.1 Komplettnost

V současné době obsahuje OSM 160 103 vyznačených budov, zatímco v datech z Geoportálu jich je 204 714, což se projevilo ve výsledných *layoutech* (obr. č 7 a obr. č. 8). Z tohoto hlediska je komplettnost dat OSM 78,21%.

Obrázek č. 7: Distribuce budov v Praze v buňkách 1x1 km z dat Geoportálu Praha



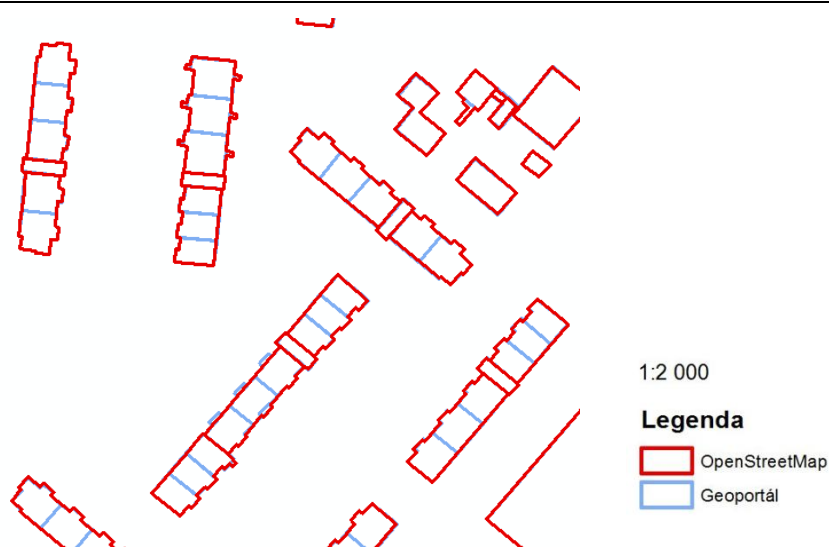
Obrázek č. 8: Distribuce budov v Praze v buňkách 1x1 km z dat OpenStreetMap



Dále oproti Geoportálu obsahuje OSM v atributové tabulce zanedbatelné množství dalších údajů o budovách, jakými jsou např. údaj o počtu pater, typ budovy a její název, což je vyhodnoceno v kapitole 4.2.1.

Výsledná hodnota 72,21% se neblíží sto procentům z velké části tím, že uživatelé OSM často nerespektují rozdělení budov v rámci bloku (obr. č. 9), tedy na jeden polygon v OSM ve skutečnosti (a v datech z Geoportálu) připadá více budov (*relace 1:N*). V dalších případech budovy v OSM úplně chybí (*relace 0:1*). Tento případ je podrobně popsán na konci této kapitoly, jelikož nalezneme ještě několik jeho podtypů.

Obrázek č. 9 - Nepřesné vyznačení řadových budov v OSM



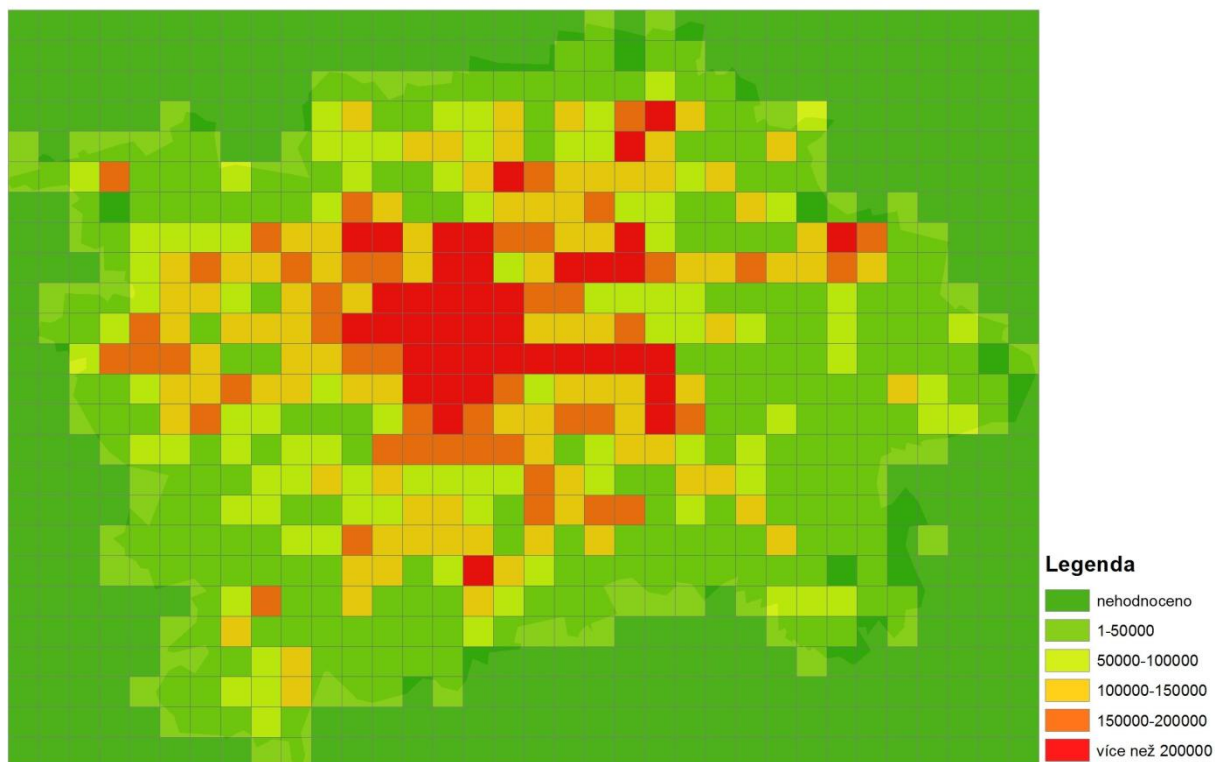
Tyto výsledky již nejsou ovlivněny nepřesným vyznačením řadových a jim podobných budov uživateli OSM (*relace 1:N*), protože součet ploch jednotlivých budov N se rovná ploše jedné budovy.

Celková plocha všech budov v OSM činí 37 199 156 m², což se rovná 37,2 km². Největší polygon má plochu 92 939 m² a je jím průmyslové centrum na Zličíně, druhou největší budovou je obchodní centrum na Černém Mostě, které ještě v datech z Geoportálu není dostavěno.

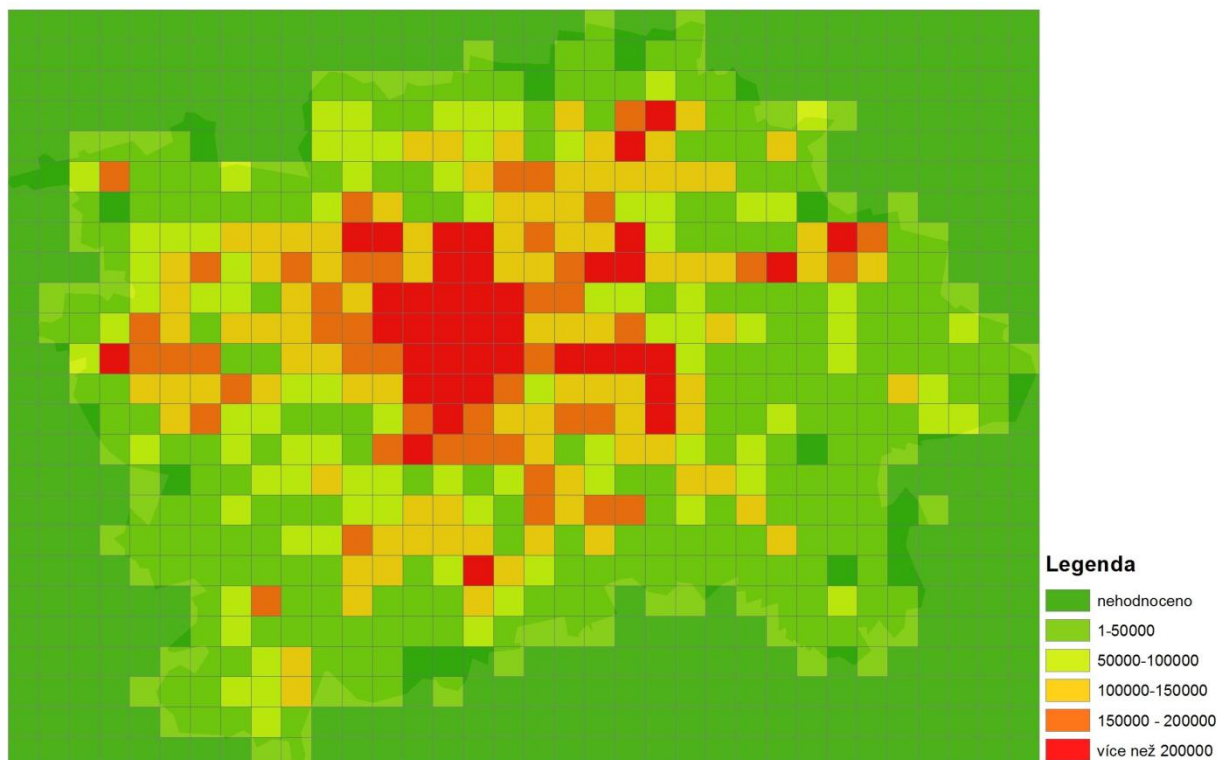
Celková plocha všech budov v Digitální mapě Prahy z Geoportálu činí 38 064 381 m² (38,1 km²), což je pouze o 2,28 procenta více, než v OSM. Největší budovou z hlediska plochy v datech z Geoportálu je hlavní budova na Letišti Václava Havla,

kteřá je v OSM rozdělená na několik částí (*relace N:1*). Dle tohoto faktoru činí kompletnost dat OSM 97,72%. Kartogramy (obr. č. 10 a obr. č. 11), které zobrazují celkovou plochu budov v každé buňce, jsou jak pro Geoportál, tak pro OSM téměř totožné.

Obrázek č. 10 - Celková plocha budov v buňkách 1x1 km v datech Geoportálu Praha (v m²)



Obrázek č. 11 - Celková plocha budov v buňkách 1x1 km v datech z OpenStreetMap (v m²)

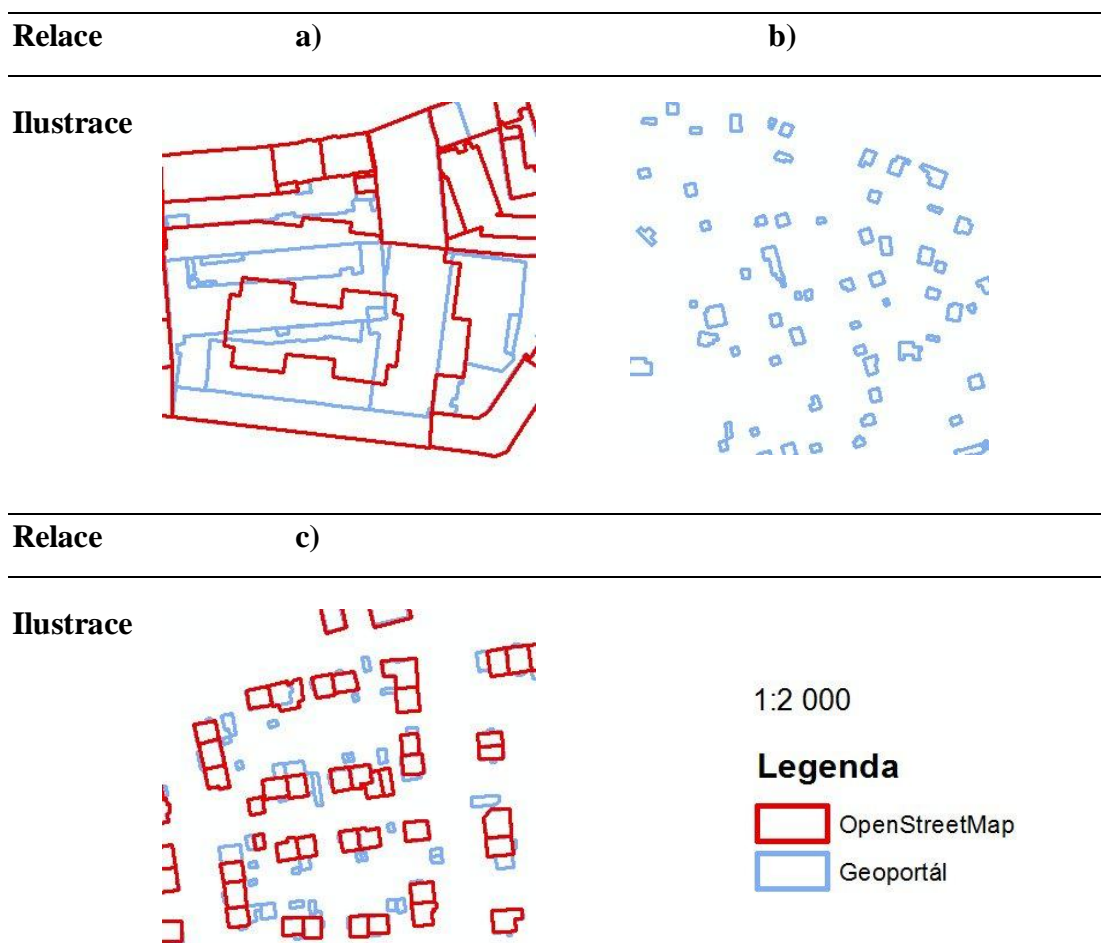


OSM tedy zachycuje převážnou většinu budov v Praze. Budovy, které zaznamenány nejsou a jsou v *relaci 0:1* k Digitální mapě Prahy, můžeme rozdělit v zásadě do tří typů, jež jsou zobrazeny na obrázcích č. 12 a), b), c).

- a) Budova je umístěna v uzavřeném dvoře, který je tvořen řadovými budovami a není k nim přístup či není z pohledu seshora vůbec vidět.
- b) Mnoho malých budov, jako jsou např. garáže, je velmi těžké identifikovat na ortofotomapách, ze kterých uživatelé OSM tvoří. Jejich střechy mají často nízký kontrast ve vztahu k terénu či okolním silnicím.
- c) V některých částech Prahy jsou vilky a přilehlé budovy (garáže) malé a obklopené z velké části stromy, což je problém při následné digitalizaci ortofotomapy.

Na obrázku č. 12 jsou modře vyznačeny budovy z Geoportálu, které se v OSM (červeně) nevyskytují, v měřítku 1:2000.

Obrázek č. 12 - Tři typy budov v OpenStreetMap, které mají ke Geoportálu *relaci 0:1*



4.2 Sémantická přesnost

Z průzkumu lze konstatovat, že jsou objekty v OSM sémanticky velmi přesně označené. Pokud byla již budova v OSM zakreslena, existuje i ve skutečnosti. Po porovnání s aktuální ortofotomapou vyšlo najevo, že v několika desítkách případů byly budovy ve skutečnosti již zdemolovány, ale v OSM nebyla tato skutečnost aktualizována. Tento fakt však není do sémantické přesnosti zahrnut. V několika jednotkách případů byl označen jako budova objekt, který se nepovažuje za budovu, a v pár případech nebylo na označeném místě nic, co by se stavbě podobalo, popřípadě bylo označeno parkoviště. Ze statistického hlediska je toto množství (~5520 z 5535) dostatečné k tomu, aby se dalo říci, že se sémantická přesnost z obecného hlediska OSM dat v Praze v tomto ohledu blíží 100%.

4.2.1 Kvantifikace sémantické přesnosti

Každý uživatel OSM má vlastní přístup, jak přispívá do celkové databáze, a data se tak zákonitě v některých aspektech liší. V kapitole 3.3. byla zmíněna OSM atributová tabulka, v níž se u sloupce *building* vyskytují hodnoty „yes“ (tedy pouze označení, že se opravdu jedná o budovu), ale také přesné určení budovy (např. *residential* - obytná budova nebo *church* - kostel). Přesné určení budovy má mnohem vyšší vypovídající hodnotu a je dobrým měřítkem kvality dat. V případě OSM dat pro hl. m. Prahu je 46,05% budov konkrétně určeno ($\text{celk. počet budov } 160\,103 - \text{počet budov s označením „yes“ } 86\,379 / 160\,103 = 53,95\%$; $100\% - 53,95\% = 46,05\%$). Pokud si vyznačíme na mapě všechny konkrétně určené budovy (*select by attributes*), lze po vizuální inspekci říci, že je tento trend v rámci celé Prahy konsistentní; tj. v každé čtvrti se vyskytují budovy konkrétně i obecně určené v přibližně stejném poměru.

4.3 Bias

Jak píše Zhu et al. (2015) i Antoniou (2011) OpenStreetMapová data jsou faktorem *Bias* (viz 2.2.2.) ovlivněna pouze minimálně, jelikož je většina dat o budovách překreslena uživateli z ortofotomap a k zaujatosti či tendencím není z principu prostor (viz také 2.2.3.). Projevuje se tu pouze výše zmíněný princip shlukování uživatelů. Data z centra města jsou, oproti datům na okraji Prahy, ve většině případů

velmi přesně zaměřena a častěji aktualizována, protože tímto územím prochází větší množství uživatelů a má větší zájem tato data spravovat (Antoniou et Schlieder, 2014). Také jsou tato data častěji zaměřena přímo v terénu a nikoli pouze překreslena z ortofotomap či podobných zdrojů.

4.4 Tvarová přesnost

4.4.1 Kvantitativní posouzení metody překrývajících se polygonů k nalezení korespondujících budov v datových vrstvách

Jelikož se analýza tvarové přesnosti týká pouze korespondujících budov (tedy v *relaci 1:1*) ve dvou datových sadách, je nutné znát kvalitu dat, která byla získána z použité metody. Bylo vybráno a manuálně vyhodnoceno obdélníkové území v Praze 1, kde se vyskytují samostatně stojící budovy i větší komplexy budov. V OSM části tohoto výběru se vyskytuje 1170 budov, v Geoportálu 1173 budov. V tabulce č. 1 jsou výsledky z průzkumu tohoto reprezentativního vzorku.

Tabulka č. 1 - Statistiky vzorku korespondujících budov

Statistika korespondujících budov po použití metody překrývajících se polygonů					
	Relace 1:N	Relace N:1	Relace N:M	Relace 1:1	Σ
Počet relací	9	1	3	1148	1160
Počet budov v OSM	9	2	11	1148	1170
Počet budov v Geoportálu	18	1	6	1148	1173

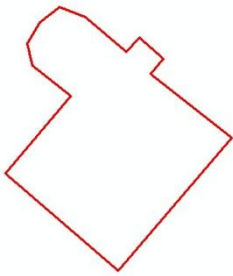
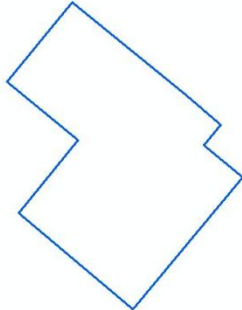

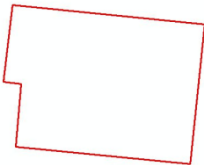
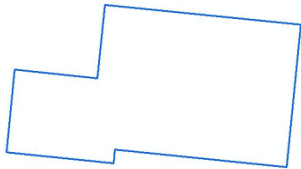

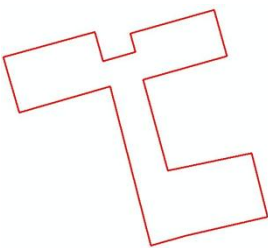
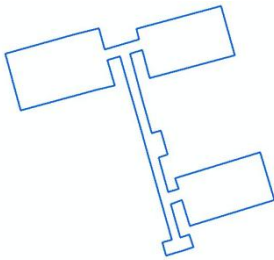

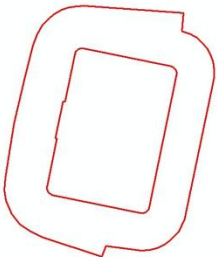
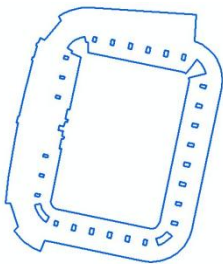

Z celkového počtu 1170 budov v OSM datové sadě bylo nalezeno 1148 budov, které jsou s budovami autoritativní datové vrstvy v *relaci 1:1*. Celková přesnost metody je tedy přibližně 98,12%.

4.4.2 Tvarová přesnost OpenStreetMap v Praze

Pro získání představy, jak vypadají typické půdorysy budov v *relaci 1:1* ve vyhodnocovaných datových zdrojích, jsou na obrázku č. 13 zobrazeny 4 příklady budovy v OSM, v Geoportálu a obrázek z *Bing maps*, ze kterého často uživatelé OSM překreslují. Je u nich také uveden rozdíl v počtu vertexů.

Pouze v několika málo případech je počet vertexů budovy v OSM vyšší, než počet vertexů budovy v Geoportálu, jako na obrázku č. 13 a), na němž má budova v OSM (červeně) komplikovanější půdorys než budova v Geoportálu (modře).

Obrázek č. 13 - Příklady typických půdorysů budov v OSM a Geoportálu

	Rozdíl vertexů	Budova v OSM	Budova v Geoportálu	Obrázek v Bing maps
a)	-4			
b)	2			
c)	24			
d)	214			

Nejčastěji jsou budovy v OSM o něco méně komplikované (viz obr. 13 b)). Podle výzkumu Fan et al. (2014) tomu tak může být převážně ze tří důvodů. Za prvé je náročné pro uživatele rozpoznat přesné architektonické detaily jen podle střechy z pohledu shora, za druhé je často důvodem nízké rozlišení mapy Bing v průběhu digitalizace a za třetí nemá mnoho uživatelů dostatek trpělivosti při digitalizaci, pokud je půdorys budovy opravdu komplikovaný. Na obr. 13 c) je půdorys budovy

v OSM oproti realitě velmi odlišný. Obecně platí, že čím je komplikovanější půdorys budovy, tím je větší rozdíl mezi OSM a Geoportálem. Poslední obr. 13 d) ukazuje pražský Eden, kde je rozdíl v množství vertexů značný. Uživatel OSM nemá reálnou šanci zjistit, jak je budova opravdu postavena a korektně ji digitalizovat.

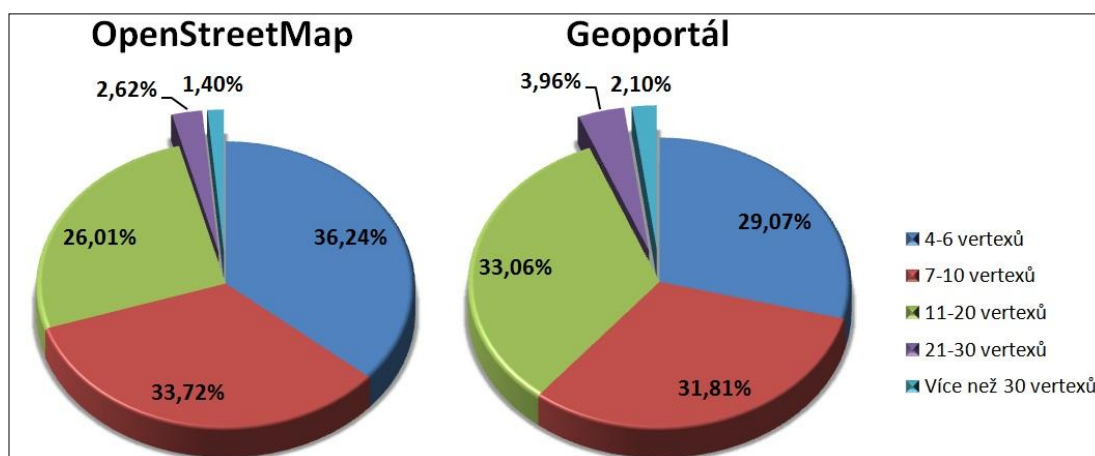
V tabulkách 2a) a 2b) jsou uvedeny statistiky pro tvarovou přesnost budov v *relaci 1:1*.²

Tabulka č. 2: Statistické porovnání datových sad Geoportálu a OpenStreetMap

a)	Statistická klasifikace Geoportál	b)	Statistická klasifikace OSM
	Celkový počet budov		151 136
	Celkový počet vertexů		1 403 443
	Minimum		4
	Maximum		347
	Průměr		9.26
	Medián		7
	Směrodatná odchylka		6.86

Pro účely tvarové přesnosti budov jsou důležité hodnoty průměrného počtu vertexů, mediánu a směrodatné odchylky. Průměrně mají budovy v *relaci 1:1* o 1,35 vertexu méně v OSM, než v Geoportálu. Vyšší směrodatná odchylka u budov v Geoportálu signalizuje větší vzájemné odlišnosti mezi typickými budovami v souboru zkoumaných budov. Jinými slovy, že je v Geoportálu větší množství

Obrázek č. 14- Porovnání skupin budov s typickým množstvím vertexů v obou sadách

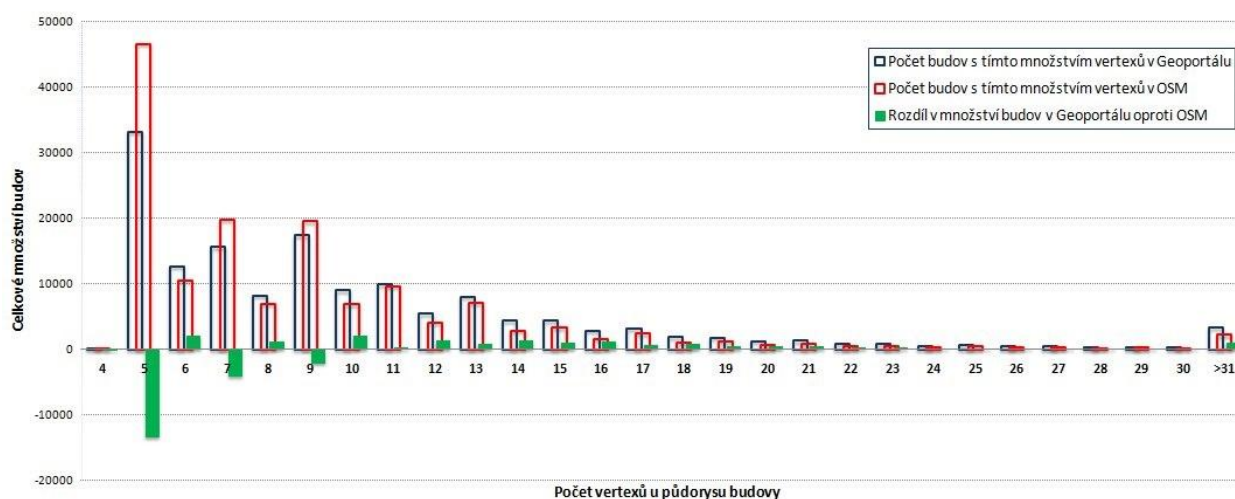


² Pokud by byly všechny relace mezi oběma datovými sadami v Praze vyhodnocovány manuálně, byl by samozřejmě celkový počet budov v *relaci 1:1* v obou sadách totožný. Reálně je tedy *relaci 1:1* o něco méně než celkový počet budov v tabulce a). Pokud použijeme koeficient 98,12% z kapitoly 3.4.4.3 dostaneme hodnotu, která mírně přesahuje 146 000 relací.

budov s vysokým počtem vertexů. Nižší směrodatná odchylka u OSM nám pak říká, že jsou si budovy v množství vertexů podobnější, než v Geoportálu. Mediánová budova v Geoportálu má také o dva vertexy více, než mediánová budova v OSM.

Grafické porovnání výše uvedeného trendu nabízí grafy a) a b) na obrázku č. 14. Budov, které jsou určeny velmi nízkým počtem vertexů (4-6), je v OSM poměrově o cca 7% více (!), než v Geoportálu. To v praxi znamená přes 10 000 budov. Budov určených středně vysokým množstvím vertexů (7-10) je v OSM stále o něco více (o necelá dvě procenta), avšak budov s vysokým množstvím vertexů (11-20) je již více v Geoportálu. Také budov s extrémně vysokým množstvím vertexů (přes 20) je více v Geoportálu. Podrobně je výskyt všech relevantních hodnot zobrazen na obrázku č. 15.

Obrázek č. 15: Vyhodnocení rozdílu v množství vertexů u půdorysů budov



5 Diskuse

5.1 Srovnání práce s výsledky podobných vědeckých prací

Pro vyhodnocení kompletnosti OpenStreetMap byly v této práci použity dvě metody na porovnání celkového množství objektů (budov) v datové sadě a následně byly graficky zobrazeny a porovnány v mřížce (*grid*). Z hlediska celkového množství budov byla zjištěna kompletnost OSM v Praze 78,21%, z hlediska celkové plochy budov 97,72%.

Fan et al. (2014) porovnával půdorysy budov v OSM v Mnichově s datovou sadou ATKIS (Amtliches Topographisch-Kartographisches Informationssystem), jež je vytvořena státním pracovním výborem Spolkové republiky Německo. Procento celkového množství budov v OSM se rovnalo 66,1%, avšak z hlediska celkové plochy bylo OSM dokonce o něco rozsáhlejší, než ATKIS (tedy více než 100%). V tomto ohledu jsou si výsledky z Prahy a z Mnichova velmi podobné.

Gires et Touya (2010) vyhodnocovali délku silnic a řek ve Francii a plochu jezer s francouzským datovým zdrojem BD TOPO® IGN. Celková délka silnic 1. třídy v OSM byla oproti oficiálnímu zdroji 45%, délka silnic 2. třídy 37%, řek 8% a celková plocha jezer 83%. V tomto roce již byly v České republice v OSM zaznamenány všechny silnice 1., 2. i 3. tříd a projekt OSM se dále rozvíjel.

Neis et al. (2010) a později i Zielstra et Zipf (2012) porovnávali OSM s komerční datovou sadou společnosti TeleAtlas, která dodává data do navigací. V pozdější práci bylo zjištěno, že celkové množství silnic v Německu v OSM je jen o 7% nižší, než v TeleAtlasu a v pěti největších městech v Německu je OSM dokonce obsáhlejší.

Z hlediska sémantické přesnosti (z obecného hlediska) je OSM v Praze téměř 100% přesné, tedy pokud je objekt zobrazen v OSM, existuje s nejvyšší pravděpodobností i ve skutečnosti. Ke stejnému závěru došel i Fan et al. (2014) pro data v Mnichově. Data z OSM v Praze však nejsou dostatečně popsána v atributové tabulce (tedy neobsahují tolik popisných informací, jako data z Geoportálu Praha). Celkové množství konkrétně určených budov (tj. přesné označení typu budovy) v OSM

v Praze dosahuje 46,05%. Fan et al. (2014) došel ve stejném ohledu k hodnotě 58,45% v Mnichově.

Bias se v OSM datech z Prahy prakticky nemá šanci měřitelně projevit, jelikož tímto faktorem není tento typ VGI obecně zatížen. Na toto téma uskutečnil výzkum Quattrone et al. (2015), který porovnával celé státy zobrazené v OSM. Zjistil, že velikost hrubého domácího produkt na obyvatele nemá větší vliv na kvalitu OSM dat. Projevuje se ale tzv. PDI (Power Distance Index) ve vztahu k množství aktualizací. Čím nižší je PDI v dané zemi, tím je vyšší pravděpodobnost, že budou uživatelé příspěvky v OSM častěji aktualizovat. Česká republika má na stupnici 0-100 PDI 57, tedy průměrnou hodnotu. Na nižší úrovni granuly dat (místní data) nebyla OSM dostatečně zkoumána.

Tvarová přesnost OSM v Praze byla vyhodnocena pro polygony v *relaci 1:1*, které byly nalezeny metodou překrývajících se polygonů, jež byla popsána v práci Fan et al. (2014) na základě poznatků z práce Rutzinger et al. (2009). Při průzkumu přesnosti metody došel Fan et al. (2014) v Mnichově k hodnotě 99,1%. V datech z Prahy byla po manuálním průzkumu menší části dat zjištěna přesnost metody 98,12%. Fan et al. (2014) však dále vyhodnocoval vlastnosti polygonů metodami, které přesahují náplň této diplomové práce. V základních údajích se ale data z Prahy a z Mnichova obecně shodují. Půdorysy budov v OSM jsou obecně určeny menším množstvím polygonů, než v autoritativních datových sadách (Geoportál, resp. ATKIS) a tento jev se ještě více prohlubuje u polygonů s vysokým množstvím vrcholů.

5.2 Budoucí projekty a výzkumy v OpenStreetMap

Do kolaborativních mapových projektů, jakým je i OSM, přispívají neustále dobrovolníci obrovským množstvím dat. Většina výzkumů řeší kvalitu dat a problémy s tím spojené, ale vývoj OSM v průběhu času ještě nebyl popsán (Arsanjani et al., 2015a). Tímto směrem by se tedy mohly ubírat další výzkumy. Jak bylo naznačeno v kapitole 2.2.2. a také kapitole 5.1., Quattrone et al. (2015) došel k závěru, že *bias* v prostorových datech není v publikovaných člancích a pracích dostatečně popsán na úrovni měst a obcí, a je potřeba dalších projektů na toto téma.

Nashromážděná OSM data jsou navíc heterogenní v různých směrech: prostorové vzory příspěvků, kvalita dat, postupy přispívání jednotlivých uživatelů a typ příspěvků se často velmi liší (Arsanjani et Bakillah, 2015).

5.3 Vhodnost použití dat z OpenStreetMap ve vybraných projektech

a) Mapování území postiženého katastrofou

Tomuto tématu se věnuje v posledních pár letech hodně pozornosti a je vhodné ho zmínit i v této práci. Území, jež bylo postižené živelnou katastrofou, se změní často k nepoznání a pro jeho znovuoobnovení je velmi důležité získat co nejrychleji prostorová data. OpenStreetMap je v těchto případech často využíván kvůli jeho rychlosti, ceně i celosvětové podpoře a často je dlouhou dobu jediným zdrojem informací o území. V poslední době probíhají také výzkumy na to, jak jsou dobrovolníci, kteří sbírají data v postižených územích, organizováni, strukturováni a jaká je jejich motivace (např. Poiani et al., 2016; Eckle et Porto De Albuquerque, 2015). I pro podporu obnovy budoucích postižených území bude OSM vhodným a využívaným nástrojem.

b) Geografické projekty profesionálů v oboru

Hlavním důvodem využití OSM v různých projektech po celém světě bývá především cena a snadná dostupnost. Při profesionálním výzkumu však nebývá financování, oproti jiným typům výzkumů, překážkou. V Praze je sice OSM z hlediska celkové plochy budov více než z 97% kompletní, ale budovy jsou často nepřesně zaměřené, neobsahují dostatečné množství popisných informací a obsahují další chyby. I z literární rešerše vyplynulo, že OSM bude muset ještě ujít dalekou cestu, než bude vhodný pro velké profesionální projekty (kapitola 2.3.4.).

c) Vzdělávání a studentské projekty

V rámci vědeckého výzkumu na sebe přitahuje OSM velkou pozornost. Jelikož za ním stojí velká a navíc stále rostoucí komunita lidí, badatelé z různých částí světa se zabývají výzkumem jeho součástí, kvalitou dat a dalšími problémy, jež se OSM týkají. Pro studentské projekty je vhodným testovacím subjektem a hlavně

v německy mluvících zemích je OSM velkým tématem a v jeho souvislosti či za jeho podpory vzniká velké množství programů a aplikací pod záštitou světových univerzit.

d) Mobilní aplikace

Některé aplikace pro tablety a mobilní telefony, jako např. aplikace pro běžce Endomondo (<https://www.endomondo.com>) nebo aplikace na plánování trasy TrailRunner (<http://trailrunnerx.com>), již úspěšně využívají OSM pro zobrazení území, ve kterém se uživatel pohybuje. Pro tyto případy a mobilní aplikace jim podobné je OSM vhodné. Různé navigace (např. Garmin) nabízí použití OSM jako volitelnou možnost, která je pro uživatele finančně výhodnější.

e) Webové stránky a webové aplikace

V Praze i dalších vyspělých městech po celém světě je OSM již dostatečně kompletní a obsahuje takové množství objektů, že je vhodné pro majitele podniků a provozovatele služeb, aby na svých webových stránkách používali OSM, jako alternativu, pokud nemohou nebo nechtějí používat placené služby.

6 Závěr

Tato práce nabízí posouzení kvality dat OpenStreetMap pro Prahu. Výsledky ukazují, že půdorysy budov v OSM mají vysokou kompletnost (97,72%) z hlediska celkové plochy území. Téměř celé území města je zmapováno a obsahuje budovy v OSM. Sémantická přesnost z obecného hlediska je téměř stoprocentní; tedy pokud je již budova zakreslena v OSM, existuje i ve skutečnosti. Avšak budovám chybí z velké části popisné informace, jakými jsou např. název budovy, typ, výška, aj. Budovy, které nejsou v OSM vůbec zmapovány, se dělí do tří typů. Budova prvního typu nemusí být na ortofotomapě, ze které uživatel vytváří digitalizovaná data pro OSM, z pohledu shora vůbec vidět. Druhý typ budovy při digitalizaci splývá s okolním terénem, jelikož má vůči němu nízký kontrast. Třetím typem jsou vilky a přilehlé garáže, které jsou z velké části obklopené stromy, což následně také vytváří problémy při digitalizaci.

Po posouzení tvarové přesnosti dat OSM bylo zjištěno, že půdorysy budov jsou určeny menším množstvím vrcholů oproti své předloze v autoritativní datové sadě (v průměru o 1,35 vrcholu) a v některých případech části budov zaznamenány vůbec nejsou. Tento trend nižší komplikovanosti půdorysů v OSM je obzvláště viditelný u budov s velkým množstvím vrcholů, které bývají zjednodušené, případně nekompletní z toho důvodu, že uživatel ani nemá možnost zjistit přesný tvar budovy pro digitalizaci.

VGI data mají i přesto potenciál k tomu, aby byla velmi významným zdrojem informací pro geografy k lepšímu poznání povrchu Země. Jsou často vysoce aktuální a díky motivaci uživatelů k tomu, aby sbírali data dobrovolně, jsou levnější než jakákoli jiná alternativa.

7 *Přehled literatury a použitých zdrojů*

Anthony D., Smith S., Williamson T., 2005: *Explaining quality in Internet collective goods: Zealots and good samaritans in the case of Wikipedia.* online: <http://web.mit.edu/iandeseminar/Papers/Fall2005/anthony.pdf>, cit. 13.4.2016.

Antoniou V., 2011: *User generated spatial content: an analysis of the phenomenon and its challenges for mapping agencies,* disertační práce, online: <http://discovery.ucl.ac.uk/1318053/>, cit. 28.10.2015.

Antoniou V. et Schlieder Ch., 2014: Participation Patterns, VGI and Gamification, příspěvek na konferenci, 8s.

Antoniou V. et Scopeliti A., 2015: *Measures and Indicators of VGI Quality: An Overview,* příspěvek na konferenci, 8s.

Arsanjani J J. et Bakillah M., 2015: *Understanding the potential relationship between the socio-economic variables and contributions to OpenStreetMap.* International Journal of Digital Earth, 8 (11), s. 861-876.

Arsanjani, J. J. , Helbich M., Bakillah M., Loos L., 2015a: *The emergence and evolution of OpenStreetMap: a cellular automata approach.* International Journal of Digital Earth Volume, 8 (1, 2), s.74-88.

Arsanjani J. J, Mooney P., Zipf A., Schauss A., 2015b: *Quality assessment of the contributed land use information from OpenStreetMap versus authoritative datasets* In: **Arsanjani, J. J., Zipf A., Mooney, P., Helbich M.:** *OpenStreetMap in GIScience: experiences, research, applications,* Springer Press, s. 37-58.

Arsanjani J. J. et Vaz E., 2015: *An assessment of a collaborative mapping approach for exploring land use patterns for several European metropolises.* International Journal of Applied Earth Observation and Geoinformation, 35, s. 329-337.

Bonter D. N. et Cooper C. B., 2012: *Data validation in citizen science: a case study from Project FeederWatch.* Frontiers in Ecology and the Environment, 10 (6), s. 305-307.

Borland J., 2007: *See Who's Editing Wikipedia - Diebold, the CIA, a Campaign.* Politics/Online Rights Blog, online: <http://www.wired.com/2007/08/wiki-tracker/>, cit. 12.4.2016.

Budhathoki N. R., Nedović-Budić Z., 2010: *Motives for VGI Participants.* Workshop 'VGI for SDI', Wageningen University, Nizozemsko.

Ciepluch B., Mooney P., Winstanley A., 2011: *Building generic quality indicators for OpenStreetMap.* 19th annual GIS Research UK (GISRUK), Portsmouth, Anglie.

- Cohen L., 2008:** *The rumor mill that won't stop running*. International Herald Tribune, online: http://www.nytimes.com/2008/10/14/technology/14iht-rumors.4.16949371.html?_r=0, cit. 13.4.2016.
- Coleman D. J., Georgiadou Y., Labonte J., 2009:** *Volunteered Geographic Information: the nature and motivation of producers*. International Journal of Spatial Data Infrastructures Research, Special Issue GSDI-11, Kanada, 20s.
- Cook S., 2008:** *Why Contributors Contribute*. Harvard Business Review, online: <http://usercontribution.intuit.com/w/page/18238302/The%20Contribution%20Revolution%20linked%20version>, cit. 13.4.2016.
- Craglia M., Goodchild M. F., Annoni A., Camara G., Gould. M., Kuhn W., Mark D., Masser I., Maguire D., Liang S., Parsons E., 2008:** *Next-Generation Digital Earth*. International Journal of Spatial Data Infrastructures Research, 3, s. 146-167.
- Čížek J., 2013:** *Google Map Maker konečně v Česku. Nyní může upravovat mapy každý*, Živě.cz, online: <http://www.zive.cz/bleskovky/google-map-maker-konecne-v-cesku-nyni-muze-upravovat-mapy-kazdy/sc-4-a-169717/default.aspx>, cit. 13.2.2016.
- Dickinson J.L., Shirk J., Bonter D., Bonney D., Crain R. L., Martin J., Phillips T., Purcell K., 2012:** *The current state of citizen science as a tool for ecological research and public engagement*. Frontiers in Ecology and the Environment, 10 (6), s. 291–297.
- Eckle M., Porto De Albuquerque, J., 2015:** *Quality Assessment of Remote Mapping in OpenStreetMap for Disaster Management Purposes*, příspěvek na konferenci, 10s.
- Elwood S., 2008:** *Volunteered geographic information: key questions, concepts and methods to guide emerging research and practice*. GeoJournal, 72 (3), s. 133–135.
- ESRI, 2016:** *Mapování bez hranic*, online: <https://www.arcgis.com/features/>.
- Everything Explained Today (EET), 2015:** *User-generated content explained*, GNU Free Documentation License, online: http://everything.explained.today/User-generated_content/, cit. 12. 2. 2016.
- Fan H., Zipfa A., Fub Q., Neisa P., 2014:** *Quality assessment for building footprints data on OpenStreetMap*, International Journal of Geographical Information Science, 28 (4), s. 700-719.
- Fisher E., 2012:** *OpenStreetMap GPS trace density in and near Europe*. online: <https://www.flickr.com/photos/walkingsf/8033247558/in/photostream/>, cit. 11. 4. 2016.

- Flanagin A. J. et Metzger M. J., 2008:** *The credibility of volunteered geographic information*, *GeoJournal* 72, s. 137–148.
- Flanagin A. J. et Metzger, M. J., 2007:** *The role of site features, user attributes, and information verification behaviors on the perceived credibility of Web-based information*. *New Media & Society*, 9(2), s. 319–342.
- Foody G.M., See L, Fritz S., Van de Velde M., Perger C., Schill C., Boyd D. S., 2013:** *Assessing the accuracy of volunteered geographic information arising from multiple contributors to an Internet based collaborative project*. *Transactions in GIS*, 17 (6), s. 847–860.
- Forghani M., Delavar M. R., 2014:** *A Quality Study of the Open Street Map Dataset for Tehran*. *IJGI: ISPRS International Journal of Geo-Information*, 3, s.750 – 763.
- Fritch J. W. et Cromwell R. L., 2001:** *Evaluating Internet resources: Identity, affiliation, and cognitive authority in a networked world*. *Journal of the American Society for Information Science and Technology*, 52 (6), s. 499–507.
- Girres J. F., Touya G., 2010:** *Quality assessment of the French OpenStreetMap dataset*. *Transactions in GIS*, 14 (4), s. 435-459.
- Goetz M. et Zipf A., 2012:** *Towards defining a framework for the automatic derivation of 3D CityGML models from volunteered geographic information*. *International Journal of 3-D Information Modeling*, 1, s. 496–507.
- Goodchild M., 2007a:** *Citizens as sensors: the world of volunteered geography*. *GeoJournal*, 69 (4), s. 211–221.
- Goodchild M., 2007b:** *Citizens as voluntary sensors: spatial data infrastructure in the world of Web 2.0*. *International Journal of Spatial Data Infrastructures Research*, 2, s. 24–32.
- Goodchild M. et Li L., 2012:** *Assuring the quality of volunteered geographic information*. *Spatial Statistics*, 1 (0), s. 110–120.
- Google, 2007:** *Introducing Google Earth outreach*, online: http://www.google.com/intl/en/press/pressrel/outreach_20070625.html, cit. 26.10.2015.
- GPSFM, 2008a:** *Openstreetmap zalesňuje ČR*, online: <http://gpsfreemaps.net/mapy/openstreetmap-zalesnuje-cr>, cit. 15.1.2016.
- GPSFM, 2008b:** *Jak pokračuje mapování silnic III*, online: <http://gpsfreemaps.net/openstreetmap/jak-pokracuje-mapovani-silnic-iii-tridy#comments>, cit. 15. 1. 2016.

- GPSFM, 2014:** *Import Adres z RUIAN do Openstreetmap*, online: <http://gpsfreemaps.net/openstreetmap/import-adres-z-ruian-do-openstreetmap>, cit. 15.1.2016.
- Graham C.H., Ferrier S., Huettman F., Moritz C., Peterson A. T., 2004:** *New developments in museum-based informatics and applications in biodiversity analysis*. Trends in Ecology & Evolution, 19 (9), s. 497–503.
- Gröger G., Kolbe T. H., Czerwinski A., 2008:** *OpenGIS city geography markup language (CityGML) encoding standard*, Open GeoSpatial Consortium, 16s.
- Haklay M. et Weber P., 2008:** *OpenStreetMap: User-Generated Street Maps*. IEEE Pervasive Computing, 7 (4), s. 12-18.
- Haklay M., 2010:** *How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets*. Environment and Planning B: Planning and Design, 37(4), s. 682-703.
- Haklay M., Basiouka S., Antoniou V., Ather A., 2010:** *How many volunteers does it take to map an area well? The validity of Linus' law to Volunteered Geographic Information*. Cartographic Journal, 47, s. 315–322.
- Haklay M., Antoniou V., Basiouka S., Soden R., Mooney P., 2014:** *Crowdsourced Geographic Information Use in Government*, Global Facility for Disaster Reduction & Recovery (GFDRR), World Bank: London, Velká Británie, s. 76.
- Harris Interactive, 2007:** *Four in five of all U.S. adults—an estimated 178 million—go online*, online: http://www.harrisinteractive.com/harris_poll/index.asp?PID=827, cit. 26.10.2015
- Chua T. S., Li J., Moens M. F., 2014:** *Mining user generated content*. Chapman and Hall/CRC, s. 7.
- IGOI, 2014:** *Content Density in OpenStreetMap*, Internet Geographies at the Oxford Internet Institute, online: http://geography.oii.ox.ac.uk/wp-content/uploads/2014/05/OpenStreetMap_Satellite.png, cit. 12.2.2016.
- IPR Praha, 2016:** *Institut plánování a rozvoje hlavního města Prahy*, online: <http://www.iprpraha.cz>, cit. 15.2.2016.
- Jackson S. P., Mullen W., Agouris P., Crooks A., Croitoru A., Stefanidis A., 2013:** *Assessing completeness and spatial error of features in volunteered geographic information*. ISPRS International Journal of Geo-Information, 2 (2): s. 507-530.

Kadmon R., Farber O., et Danin A., 2004: *Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models.* Ecological Applications, 14 (2), s. 401–413.

Kalantari M., La V., 2015: *Assessing OpenStreetMap as an Open Property Map.* OpenStreetMap in GIScience, Springer International Publishing, Švýcarsko, s. 255-272.

Keen A., 2008: *The Cult of the Amateur: How blogs, MySpace, YouTube, and the rest of today's user-generated media are destroying our economy, our culture, and our values* Doubleday, New York, USA, 236s.

Koukoletsos T., Haklay M., Ellul C., 2011: *An automated method to assess Data Completeness and Positional Accuracy of OpenStreetMap.* Proceedings GISRUK 2011, University of Portsmouth, Portsmouth, Velká Británie.

Kounadi, O., 2009: *Assessing the quality of OpenStreetMap data.* Advisor thesis, University College London, London, Velká Británie, 80s.

Kutnezov S., 2006: *Motivations of Contributors to Wikipedia.* New York University, online:
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.457.1402&rep=rep1&type=pdf>, cit. 12.2.2016.

Lakhani K., et Wolf R., 2005: *Why hackers do what they do: Understanding motivation and effort in free/opensource software projects.* In: **Feller J., Fitzgerald B., Hissam S., Lakhani K. [Eds]:** Perspectives in free and open source software. Cambridge, MIT, 570s.

Lanier J., 2006: *Digital Maoism: The Hazards of the New Online Collectivism,* The Edge online: https://www.edge.org/conversation/jaron_lanier-digital-maoism-the-hazards-of-the-new-online-collectivism, cit. 13.4.2016

Leitão P.J., Moreira F. et Osborne P.E., 2011: *Effects of geographical data sampling bias on habitat models of species distributions: a case study with steppe birds in southern Portugal.* International Journal of Geographical Information Science, 25 (3), s. 439–454.

Lenhart A., 2006: *User-generated content.* Pew Internet Project, online:
http://www.pewinternet.org/PPF/r/76/presentation_display.asp/, cit. 11.2.2016.

Ludwig I., Voss A., Krause-Traudes M., 2011: *A comparison of the street networks of Navteq and OSM in Germany.* Advancing Geoinformation Science for a Changing World, Springer-Verlag Berlin Heidelberg , s. 65-84.

- Mashhadi A., Quattrone G., Capra L., 2015:** *Impact of Society on Volunteered Geographical Information: the case of OpenStreetMap*. OpenStreetMap in GIScience: experiences, research, and applications, Springer International Publishing, Švýcarsko, s. 125-141.
- McDonough M., 2013:** *Google Map Maker vs. OpenStreetMap: Which Mapping Service Rules Them All?*, online: <http://www.digitaltrends.com/computing/google-map-maker-vs-openstreetmap-id-editor>, cit. 14.2.2016.
- McHenry R., 2004:** *The Faith-Based Encyclopedia*. Technology, Commerce and Society Daily, online: http://www.ideasinactiontv.com/tcs_daily/2004/11/the-faith-based-encyclopedia.html, cit. 13.4.2016.
- Metzger M. J., Flanagin A. J., Eyal K., Lemus D. R., McCann R., 2003:** *Credibility in the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment*. In: **Kalbfleisch P. [Ed]:** Communication yearbook 27, Mahwah, NJ: Lawrence Erlbaum, s. 293–335.
- Metzger M. J., 2007:** *Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research*. Journal of the American Society for Information Science and Technology, 58(13), s. 2078–2091.
- Mohammadi N. et Malek R. M., 2014:** *VGI and Reference Data Correspondence Based on Location-Orientation Rotary Descriptor and Segment Matching*. Transactions in GIS 19 (4), s. 619-639.
- Mullen W. F., Jackson S. P., Croitoru A., Crooks A., Stefanidis A., Agouris P., 2014:** *Assessing the impact of demographic characteristics on spatial error in volunteered geographic information features*. GeoJournal, 1 (19), s. 587-605.
- Munson M. A., Caruana R., Fink D., Hochanka W. M., Iliff M., Rosenberg K. V., Sheldon D., Sullivan B. L., Wood Ch., Kelling S., 2010:** *A method for measuring the relative information content of data from different monitoring protocols*. Methods in Ecology and Evolution, 1 (3), s. 263–273.
- Neis P., Zielstra D., Zipf A., Strunck A., 2010:** *Empirische Untersuchungen zur Datenqualität von OpenStreetMap – Erfahrungen aus zwei Jahren Betrieb mehrerer OSM-Online-Dienste*. Universität Heidelberg, Geographisches Institut, 6s.
- OpenStreetMap Czech FAQ, 2016:** *Cs:FAQ*, online: <http://wiki.openstreetmap.org/wiki/Cs:FAQ>, cit. 15.1.2016.
- OpenStreetMap ČR, 2016:** *Statistiky OpenStreetMap ČR*, online: <http://osm.kyblsoft.cz/statistiky/pocet-bodu.png>, cit. 12.2.2016.

- OpenStreetMap Wiki, 2016:** *Statistics tracking the remarkable growth of OpenStreetMap's database and community*, online:
<http://wiki.openstreetmap.org/wiki/Stats>, cit. 11.4.2016.
- O'Reilly T., 2005:** *What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software*. Blog - Tim O'Reilly, online:
<http://www.oreilynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>,
cit. 28.10.2015
- Poiani T. H., Rocha R., Degrossi L. C., Porto De Albuquerque, J., 2015:** *Potential of Collaborative Mapping for Disaster Relief: A Case Study of OpenStreetMap in the Nepal Earthquake 2015*, příspěvek na konferenci, 11s.
- Ponder W.F., Carter G.A., Flemons P., Chapman R.R., 2001:** *Evaluation of museum collection data for use in biodiversity assessment*. *Conservation Biology*, 15 (3), s. 648-657.
- Priedhorsky R. R. et Terveen L. G., 2010:** *The Value of Geographic Wiki*. Disertační práce, 124 s.
- Quattrone G., Capra L., De Meo P., 2015:** *There's No Such Thing as the Perfect Map: Quantifying Bias in Spatial Crowd-sourcing Datasets*. CSCW '15 Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, s. 1021-1032.
- QUORA, 2015:** *How many users does Wikimapia have?*, online:
<https://www.quora.com/How-many-users-does-Wikimapia-have>, cit. 12.2.2016.
- Reddy S. et Da, L.M., 2003:** *Geographical sampling bias and its implications for conservation priorities in Africa*. *Journal of Biogeography*, 30 (11), s. 1719–1727.
- Rieh S. Y. et Danielson D. R., 2007:** *Credibility: A multidisciplinary framework*. In: **Cronin B. [Ed]**, *Annual review of information science and technology*, 41, s. 307–364.
- Ritchie E. M., Silver J., Oshlack A., Holmes M., Diyagama D., Holloway A., Smyth G. K., 2007:** *A comparison of background correction methods for two-colour microarrays*. *Bioinformatics* 23 (20), s. 2700-2707.
- Rutzinger M., Rottensteiner F., Pfeifer N., 2009:** *A comparison of evaluation techniques for building extraction from airborne laser scanning*. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2 (1), s. 11–20.
- Schroera J., Hertel G., 2009:** *Voluntary Engagement in an Open Web-Based Encyclopedia: Wikipedians and Why They Do It*, *Media Psychology* 12 (1), s. 96-120.

- Silvertown J., 2009:** *A new dawn for citizen science*. Trends in Ecology & Evolution, 24 (9), s. 467–471.
- Statista, 2016:** *Statistics and facts about user-generated content in the U.S*, online: <http://www.statista.com/topics/1716/user-generated-content/>, cit. 10.3.2016.
- Sui D., 2008:** *The wikification of GIS and its consequences: Or Angelina Jolie's new tattoo and the future of GIS*. Computers, Environment and Urban Systems, 32 (1), s. 1–5.
- Sullivan B. L., Wood Ch. L., Iliff M.J., Bonney R.E., Fink D., Kelling S., 2009:** *A citizen-based bird observation network in the biological sciences*. Biological Conservation, 142 (10), s. 2282–2292.
- Tapscott D. et Williams A. D., 2010:** *Wikinomics: How mass collaboration changes everything*. Penguin Group, New York, USA, 368 s.
- Tulloch D. L., 2007:** *Many, many maps: Empowerment and online participatory mapping*. First Monday, 12 (2), s. 2-5.
- Tulloch D. L. 2008:** *Is VGI participation? From vernal pools to video games*. GeoJournal 72 (3-4), s. 161–171.
- Vandecasteele A., Devillers R. 2015:** *Improving volunteered geographic information quality using a tag recommender system: The case of OpenStreetMap* In: **Jokar Arsanjani J., Zipf A., Mooney P., Helbich M. [eds]: OpenStreetMap in GIScience: experiences, research, applications**, Springer International Publishing, Švýcarsko, s. 59-80.
- Viegas F. B., Wattenberg M., Kriss J., van Ham F., 2007:** *Talk Before You Type: Coordination in Wikipedia*. Proceedings of the 40th Annual Hawaii International Conference on System Sciences, 10s.
- WikiMapia, 2016:** *Terms of Service, Copyright Notice and Privacy Notice*, http://wikimapia.org/terms_reference.html, cit. 12.2.2016.
- Zhu A. X., Zhang G., Wang W., Xiao W., Huang Z.P., Dunzhu G.S., Ren G., Qin Ch.Z., Yang L., Pei T., Yang S., 2015:** *A citizen data-based approach to predictive mapping of spatial ariation of natural phenomena*. International Journal of Geographical Information Science 29 (10), s. 1864-1886.
- Zielstra D. et Zipf A., 2010:** *A comparative study of proprietary geodata and volunteered geographic information for Germany*. Proceedings of the Thirteenth AGILE International Conference on Geographic Information Science, Portugalsko, 15s.