

OTEVŘENÁ DOPRAVNÍ MAPA PRO EVROPU

Karel JEDLIČKA¹, Pavel HÁJEK¹, Jan JEŽEK¹, František KOLOVSKÝ¹, Daniel BERAN¹,
Tomáš MILDORF¹, Karel CHARVÁT², Dimitri KOZHUKH², Jan MARTOLOS³, Jan ŠTASTNÝ³

¹Západočeská univerzita v Plzni, Fakulta aplikovaných věd, Katedra geomatiky,
Univerzitní 8, Plzeň, 306 14, Česká republika

{smrcek, gorin, jezekjan, mildorf}@kgm.zcu.cz, {kolovsky, dberan}@students.zcu.cz

²Help Service – Remote Sensing spol. s r.o., Husova 2117, 256 01, Benešov, Česká republika

{dmitrii, charvat}@hsrs.cz

³EDIP s.r.o., Pařížská 1, 301 00, Plzeň, Česká republika

{martolos, stastny}@edip.cz

Abstrakt

Příspěvek seznamuje čtenáře s rozšířením datové sady „Otevřená dopravní mapa“ o intenzity dopravy kalkulované přes evropský kontinent. Po zpracování výpočtů intenzit dopravy na územích malého rozsahu (města, okresy, malé státy) pomocí komerčního desktopového softwaru, které byly zahrnuty v předchozích verzích této datové sady, byla motivace přejít s výpočty do většího územního rozsahu (až na celý evropský kontinent). Pro tento účel se použité desktopové řešení ukázalo jako nedostatečné, především kvůli zpracování velkého množství dat. Pro výpočet intenzit dopravy bylo tedy třeba přijít s jiným postupem prací a získat pro tyto práce odpovídající podkladová data. Výpočet intenzit dopravy vychází z demografických dat a silniční sítě, přičemž silniční síť pro Evropu již byla zpracována v dřívějších verzích Otevřené dopravní mapy. Demografická data za Evropu byla přejata ze statistického úřadu Evropské unie (EUROSTAT). Ta byla harmonizována do podoby vhodné pro vstup do výpočtu intenzit. K výpočtu intenzit dopravy bylo použito serverové řešení, konkrétně softwarová knihovna Apache Hadoop, pro kterou bylo třeba navrhnout a implementovat vhodný algoritmus pro výpočet intenzit na velmi rozsáhlém vzorku dat. Nakonec vypočtené intenzity dopravy prošly kalibrací na základě dopravně-statistických dat komise spojených národů UNECE, která poskytuje data o sčítání dopravy na mezinárodních silnicích Evropy.

Abstract

This contribution presents an extension of the Open Transport Map (OTM) by traffic volumes calculated across almost the whole European continent. After handling the calculation of the traffic volumes on small-size areas (such as cities, regions, small countries) using a commercial desktop software, which are incorporated in the previous versions of the OTM, there was a motivation to enlarge the calculated area, even to the whole Europe. The used desktop solution has been found insufficient for the European level traffic volumes calculation, mainly due to the limitations in calculation performance on such a big data. Therefore, a new approach for the traffic volumes calculation had to be deployed. And also relevant source data for the calculation had to be obtained. The calculation itself is based on demographical data of each particular Local Area Unit level 2 (LAU2) and road network. The road network of Europe has been taken from the already existing version of the OTM and the demographic data has been obtained from the statistical office of the European Union (EUROSTAT). After that, the demographic data had to be processed to the suitable form as an input for the traffic volumes calculation. The software library Apache Hadoop was used for the distributed calculation on a server. The proper algorithm for the computing of such a big data had to be created and implemented. At the end, the calculated volumes were calibrated to the data from traffic census from the main international roads of Europe provided by the United Nations Economic Commission for Europe (UNECE).

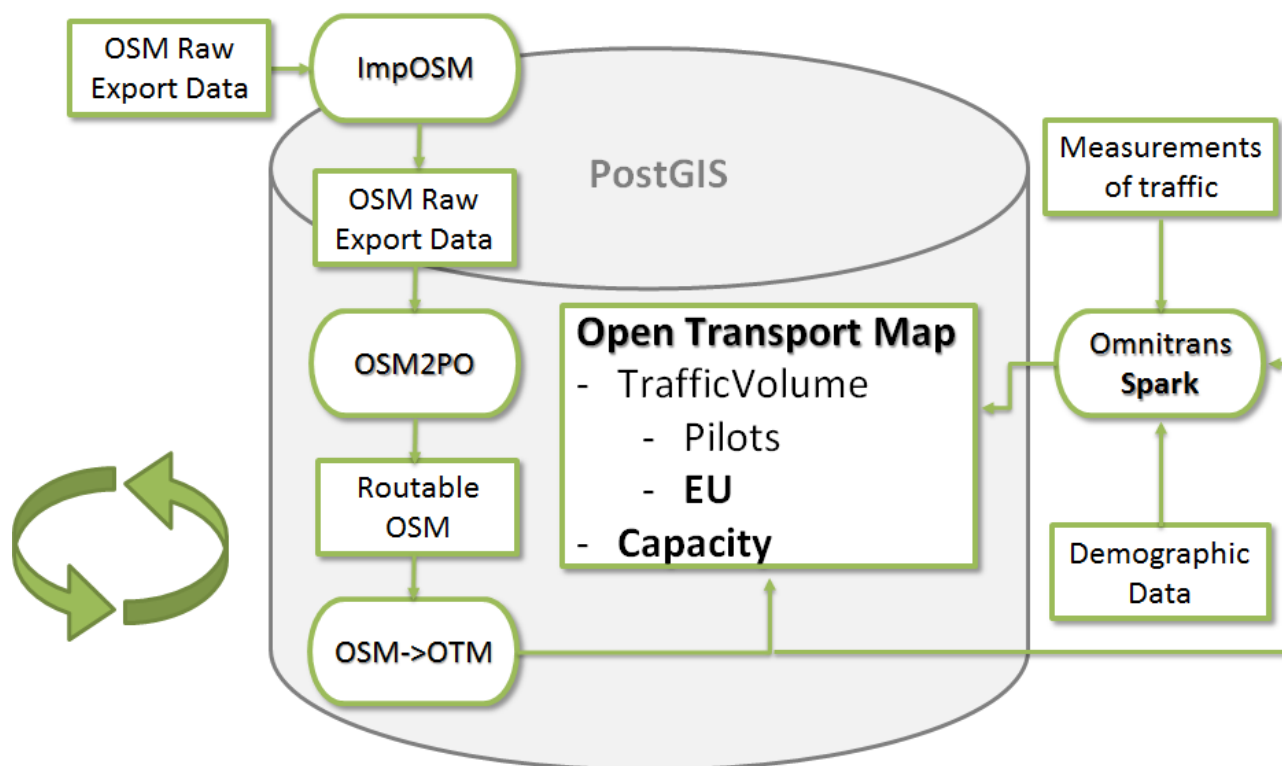
Klíčová slova: doprava; dopravní síť; intenzita dopravy; otevřená data; mapa; geografický informační systém; Evropa.

Keywords: transport; transport network; traffic volume; open data; map; geographical information system; Europe.

ÚVOD

Otevřená dopravní mapa (Open Transport Map - OTM) je datovou sadou, která obsahuje informace o dopravních intenzitách na evropské silniční síti. Na obrázku 1 níže jsou tučným písmem znázorněny témata, kterými se tento příspěvek bude převážně zabývat. Kroky vedoucí k vytvoření Otevřené dopravní mapy, implementovaný způsob datové harmonizace, postup výpočtu dopravních intenzit, volba kartografického zobrazení dopravních intenzit a ukázky konkrétních pilotních území jsou uvedeny ve Veeckman a kol. (2017), Jedlička a kol. (2015a), Jedlička a kol. (2015b), Jedlička a kol. (2016) a v Kozhukh a kol. (2015). V tomto článku bude rozebrán konkrétně postup výpočtu dopravních intenzit pro Evropu.

Tento text se zaměří na pravou část diagramu na obrázku 1 a to na postup zpracování podkladových dat pro výpočet intenzit dopravy a samotný výpočet dopravních intenzit nad silniční sítí Evropy. Výpočet dopravních intenzit je aplikován na silnice 3. třídy a vyšší, nikoliv na celou silniční síť. Rozdíl oproti výše uvedeným publikacím, ve kterých jsou prezentovány výpočty dopravních intenzit na pilotních územích typu městských aglomerací či malých států, je v řádově větším územním rozsahu (tj. násobně větším počtu vstupních prvků - úseků silnic), v nutném předzpracování dostupných demografických dat za Evropu a především v eskalované náročnosti výpočtu intenzit dopravy. Vzhledem k tomu, že se již pohybujeme, co do počtu zpracovaných prvků, v oblasti tzv. velkých dat (Big Data), bylo třeba přistoupit k serverovému řešení výpočtu dopravních intenzit, který je schopen takové množství dat zpracovat. I přes adaptaci výpočetního algoritmu z desktopové verze na serverovou, je tento výpočet časově i kapacitně velmi náročný. Dosažené výsledky, včetně jejich časové náročnosti, jsou uvedené v kapitole zabývající se výsledky experimentů.



Obr. 1: Diagram tvorby Otevřené dopravní mapy.

1. POSTUP VÝPOČTU DOPRAVNÍCH INTENZIT PRO EVROPU

Pro výpočet dopravních intenzit nad silniční sítí Evropy bylo třeba získat a zpracovat větší a zároveň více heterogenní množství podkladových dat, než tomu bylo v případech výpočtu intenzit pro předchozí pilotní území (ČR, Lotyšsko, pilotní území projektu OpenTransportNet¹). Jak je uvedeno na obrázku 1, pro výpočet dopravních intenzit jsou potřeba určitá vstupní podkladová data. Kromě samotné silniční sítě, nad kterou se výpočty realizují, se jedná o demografická data (Demographic Data) a pro kalibraci výpočtu intenzit pak data z dopravního sčítání (Measurements of Traffic). Harmonizace nalezených otevřených dat vhodných pro účely výpočtu dopravních intenzit je hlavním tématem této kapitoly.

Evropská silniční síť je v Otevřené dopravní mapě již od začátku jejího vzniku, tj. naplnění do databáze PostGIS, viz obrázek 1. V tomto ohledu tedy nebylo třeba tato data harmonizovat a předzpracovávat pro výpočet dopravních intenzit a z pohledu harmonizace jí nebude v dalších podkapitolách věnována další pozornost. Více informací o tomto tématu lze nalézt ve Veeckman et al. (2017), Jedlička et al. (2015a), Jedlička et al. (2015b) či v Jedlička et al. (2016). Jednotlivé části této kapitoly se budou zabývat zpracováním demografických dat pro Evropu, principem výpočtu dopravních intenzit na serveru a kalibrační vypočtených dopravních intenzit.

1.1. Demografická data a generátory dopravy

Data o počtu obyvatel je třeba definovat pro statisticky a rozsahem významné a vhodné územní celky. Z nich se poté vytváří tzv. generátory dopravy. Generátory dopravy jsou reprezentovány body, nesoucí informaci o tom, kolik obyvatel se v daném územním celku nachází. V dalším kroku postupu výpočtu dopravních intenzit jsou tyto generátory napojeny na jejich nejbližší úseky dopravní sítě.

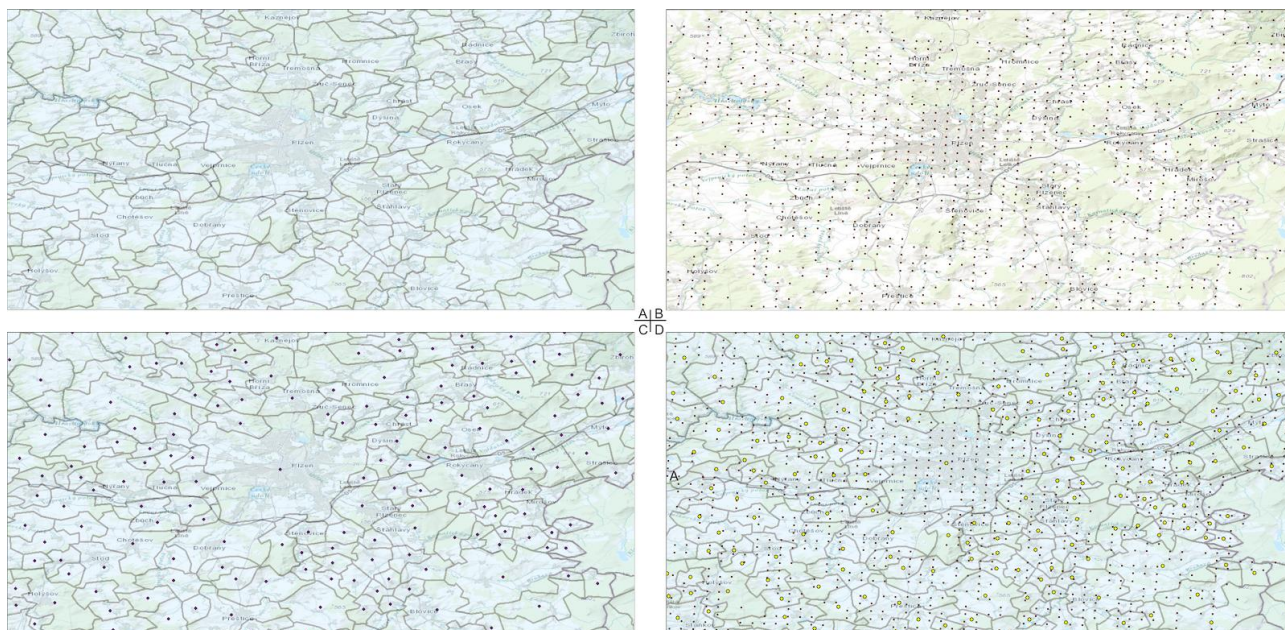
Z toho vyplývá, že tyto územní celky nesmí být ani příliš velké (v plošném slova smyslu), což by vedlo k přílišnému zjednodušení výpočtu a pro relevantnost výpočtu by takovýchto generátorů bylo málo. Na druhou stranu, příliš vysoká granularita těchto územních celků by byla statisticky více vypovídající, nicméně by náročnost výpočtu následně zvyšovala. Bylo třeba tedy zvolit vhodnou granularitu dat tak, aby bylo možné provést výpočet intenzit v limitním čase a zároveň, aby vybrané územní celky pokrývali co možná největším územním rozsahem Evropu.

Vhodnou reprezentací územního rozsahu byly zvoleny tzv. místní správní jednotky (Local Administrative Unit - LAU), které jsou dostupné ke stažení na stránkách EUROSTATu za celou Evropskou unii (viz Eurostat (2016a)). Konkrétně byly zvoleny správní jednotky LAU2 (dříve označovány jako NUTS 5). Jedná se o rozsahem o území označované v ČR jako obce. Tato data jsou z výše uvedeného zdroje dostupná pro státy EU a EFTA, nikoliv pro kandidátské země pro vstup do EU a pro potenciální kandidátské země, viz Eurostat (2016c). Vybraná datová sada je polygonová a obsahuje 119 396 záznamů. V těchto otevřených datech však chybí informace o počtu obyvatel žijících v jednotlivých správních jednotkách. Bylo třeba tedy nalézt vhodný doplňkový datový zdroj, který by takovou informaci obsahoval, nebo by z něj šla odvodit. Kýžená data jsou dostupná také na stránkách EUROSTATu (viz Eurostat (2016b)), jedná se o polygonovou datovou vrstvu ve formě pravidelného gridu pro celou Evropu (tj. i pro státy mimo EU, u kterých však data o populaci nejsou založena na zvoleném celoevropském sčítání lidu z roku 2011, ale na odhadech, viz Eurostat (2016b)). V této vrstvě je více jak 2,1 milionů záznamů, pohybujeme se tedy v oblasti tzv. velkých dat, viz Unwin et al. (2006). Bylo třeba tyto dvě datové vrstvy vhodně zkombinovat a vytvořit následně bodovou vrstvu generátorů dopravy jako jednu ze vstupních datových vrstev pro výpočet intenzit dopravy pro Evropu. Ukázka výše uvedených datových sad je na obrázku 2.

¹ viz <http://opentransportnet.eu/cs/>

Postup harmonizace výše uvedených datových sad byl následující:

- Zpracování dat nesoucích informace o počtu obyvatel - díky velikosti této vrstvy se přikročilo k použití serverového zpracování dat na místo zpracování dat na desktopu. Původní datová sada je polygonová a je rozdělená do dvou sad (jedna pro EU, druhá pro zbytek států Evropy, viz popis datové sady v Eurostat (2016b)). V tomto kroku byla data sloučena na základě unikátního identifikátoru a prostorového rozložení dat a následně převedena na bodovou vrstvu. Vznikla tak souvislá bodová datová sada statistických dat.
- Připojení statistických dat na územní rozsahy jednotlivých korespondujících územních celků (obohacení polygonů územních jednotek o informaci o počtu obyvatel v nich žijící) - statistický grid znázorňuje zjednodušenou reprezentaci tvaru Evropy, která v pobřežních částech nekoresponduje s rozsahem vybraných územních celků. Konkrétní prostorová distribuce hodnot o počtu obyvatel není v našem případě zásadní, kýžené je sčítání hodnot počtu obyvatel bodového gridu pro jednotlivá LAU. Statistické body spadající mimo území jednotlivých LAU2 byly napojeny na jejich nejbližší sousedy a následně došlo k sumarizaci údajů o počtu obyvatel jednotlivých LAU z prostorového průniku statistických bodů a polygonů správních jednotek.
- Tvorba generátorů dopravy - z výše obohacené polygonové vrstvy byly vytvořeny centroidy reprezentující generátory dopravy. Ty určují počet osob vyskytujících se v dané oblasti, jinými slovy předpokládáme, že daní obyvatelé se podílejí na silniční dopravě. Takto upravená data vstupují do výpočtu dopravních intenzit nad úseky silniční sítě.



Obr. 2: Ukázky zdrojových dat pro definování generátorů dopravy.

Pozn.: Popis obrázku 2 zobrazující Plzeň a okolí:

A - Polygony LAU2

B - Statistický grid

C - Centroid každého LAU s informací o součtu počtu obyvatel náležících mu

D - Polygony LAU překryté statistickým gridem a se znázorněnými centroidy polygonů (žluté body)

1.2. Výpočetní algoritmus, princip výpočtu dopravní intenzity

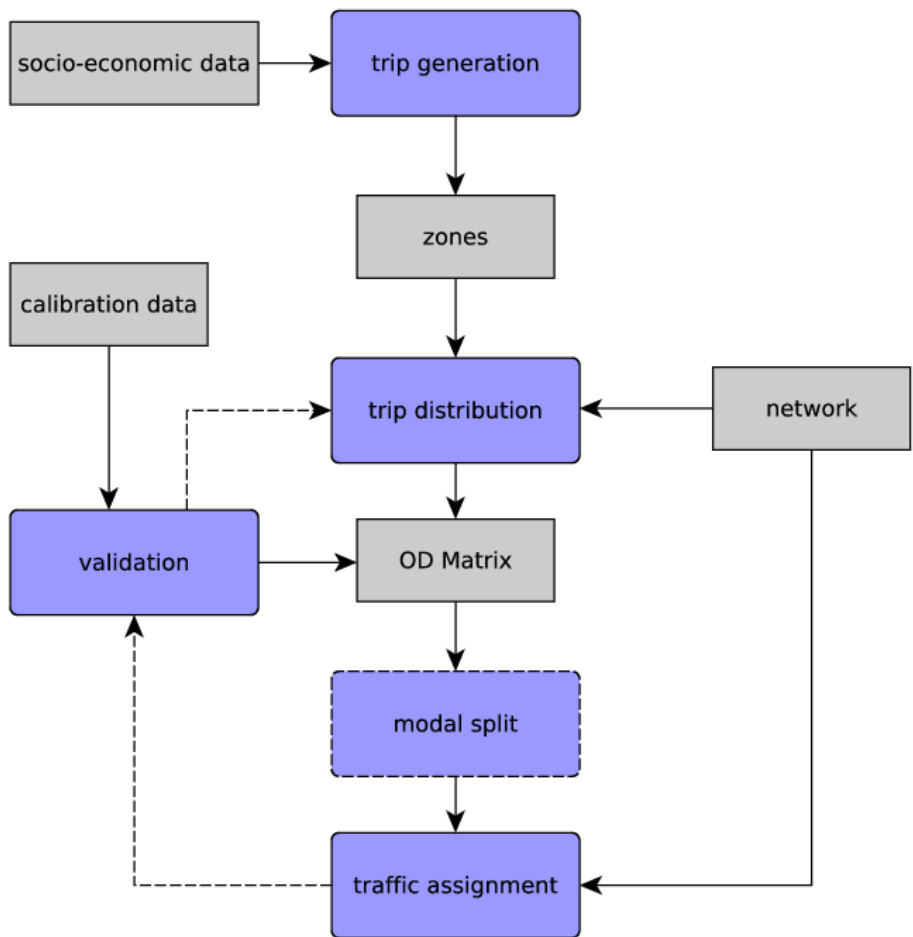
Při předchozích výpočtech intenzit dopravy byl použit software OmniTRANS firmy DAT.Mobility² a v něm implementovaný model jejich výpočtu (více informací viz články uvedené v úvodu této práce). Pro použití serverového řešení výpočtu bylo třeba implementovat sadu algoritmů založených na klasickém čtyř krokovém dopravním modelu (více informací o tomto modelu viz Ortúzar & Willumsen (2011)). Tato sada algoritmů byla kódovaná v jazyce Scala. Informace uvedené v této kapitole vychází především z Kolovský & Ježek (2017).

Postup výpočtu dopravních intenzit a jejich inkorporace do silniční sítě je uveden níže, zároveň je tento postup graficky vyjádřen na obrázku 3 (šedě jsou označeny vstupy a výstupy, fialově pak jednotlivé kroky postupu).

1. tvorba cest (trip generation) - jednotlivé cesty začínají a končí v generátorech dopravy. Ty mohou obsahovat nejen údaje o počtu obyvatel v daném generátoru, ale také jiná socio-ekonomická data jako procenta zaměstnaných, prodeje v korespondujících nákupních centrech, věk, počet dojíždějících osob do práce atd.
2. distribuce cest (trip distribution) - jedná se o nejdůležitější část postupu výpočtu intenzit dopravy. Cílem tohoto kroku je tvorba tzv. Origin-Destination Matrix - ODM, neboli Matice transportních (přepravních) vztahů. Tato matice obsahuje počet různých cest mezi jednotlivými generátory dopravy. Je postavena na myšlence, že počet cest mezi dvěma generátory závisí nepřímo úměrně na ceně překonané cesty (typicky v závislosti na vzdálenosti, čase, či nákladech). V tomto případě na nejkratší z cest a na myšlence, že většina osob si zvolí právě co nejkratší cestu. Pro nalezení nejkratších cest byl použit upravený Dijkstrův algoritmus.
3. modální rozdělení (modal split) - jde o alokaci počtu cest mezi více druhů dopravy. ODM je rozdělena do několika matic dle typu použité dopravy (auto, nákladní auto, veřejná doprava, apod.). Pokud v předchozích krocích nebyly jednotlivé druhy dopravy brány v potaz, tento krok se vynechává. V případě Otevřené dopravní mapy byl tento krok vynechán.
4. přiřazení dopravní intenzity (traffic assignment) - v tomto posledním kroku dochází k definování a přiřazení dopravní intenzity na jednotlivé úseky silniční sítě. Vstupem je tedy ODM a výstupem je počet vozidel projíždějících daným úsekem za jednotku času, konkrétně tzv. roční průměr denních intenzit (RPDI - aritmetický průměr denních intenzit dopravy všech dnů v roce, viz Bartoš & Martolos (2012)). Základní myšlenkou v tomto kroku je, že řidič využívá k přepravě pro něj optimální cestu. Dopravní intenzita tedy vzrůstá na daném úseku, který spojuje více generátorů dopravy. Byla zvolena metoda výpočtu, při které je cena překonání cesty konstantní v tom smyslu, že nezávisí na momentální dopravní intenzitě v místě a čase (tzv. all-or-nothing metoda, matematický popis metody viz Ortúzar & Willumsen (2011)).
5. kalibrace vypočtených dopravních intenzit (validation) - v tomto bodě dochází ke kalibraci ODM s využitím referenčních údajů o sčítání dopravy. Snahou je minimalizovat rozdíl objemu referenční a modelové dopravy pomocí numerické minimalizace objektové funkce.

Vstupem pro krok 1 je datová sada generátorů dopravy odvozená z demografických dat a pro krok 5 je vstupem datová sada kalibračních referenčních dat odvozená ze sčítání dopravy na mezinárodních evropských silničních tazích (viz další podkapitola). Algoritmus celého výpočtu dopravních intenzit je navržen tak, že tyto dvě datové sady jsou vstupními parametry procesu již při jeho iniciaci.

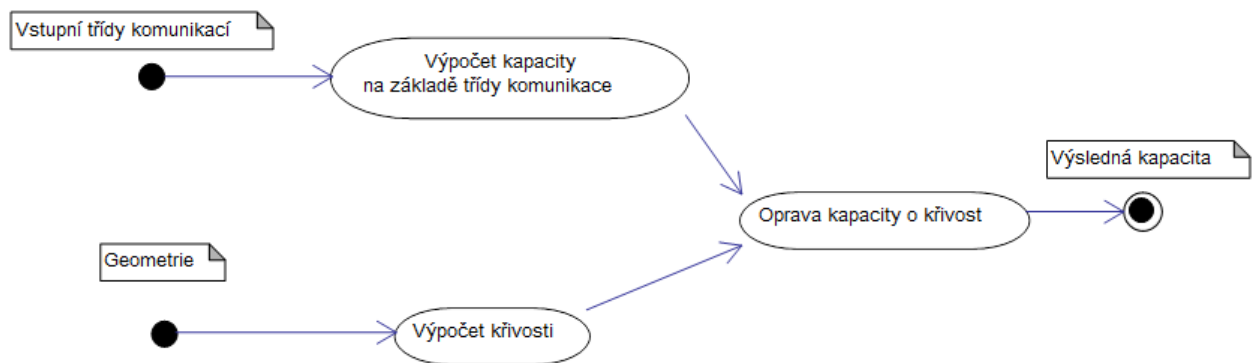
² <http://www.dat.nl/en/products/>



Obr. 3: Diagram postupu výpočtu dopravních intenzit (převzato z Kolovský & Ježek (2017)).

Nově implementovaný parametr pro výpočet dopravních intenzit

Jedním z nově implementovaných parametrů vystupujícím ve výpočtu dopravních intenzit je určení kapacity silničního úseku na základě třídy komunikace (s jako takovou bylo počítáno již v předchozích verzích Otevřené dopravní mapy), který je opraven o výpočet křivosti úseku na základě jeho geometrie (nově přidáný parametr). Diagram ukazující proces výpočtu kapacity komunikace je zobrazen na obrázku 4. Definování kapacity pro zájmové třídy komunikací je uvedeno v tabulce 1. Výpočet samotné křivosti je uveden ve vzorci 1.



Obr. 4: Postup určení kapacity úseku silniční sítě.

Tab. 1: Převod třídy komunikace na její kapacitu (převzato z Kolovský (2016)).

Třída komunikace (podle anglické notifikace OTM)	Kapacita [voz/24hod]
mainRoad	45 000
firstClass	15 000
secondClass	8 000
thirdClass	6 000

Vzorec 1: Výpočet křivosti (dle normy ČSN 73 6101)

$$K = \frac{\sum_{i=1}^j \gamma_i}{l}$$

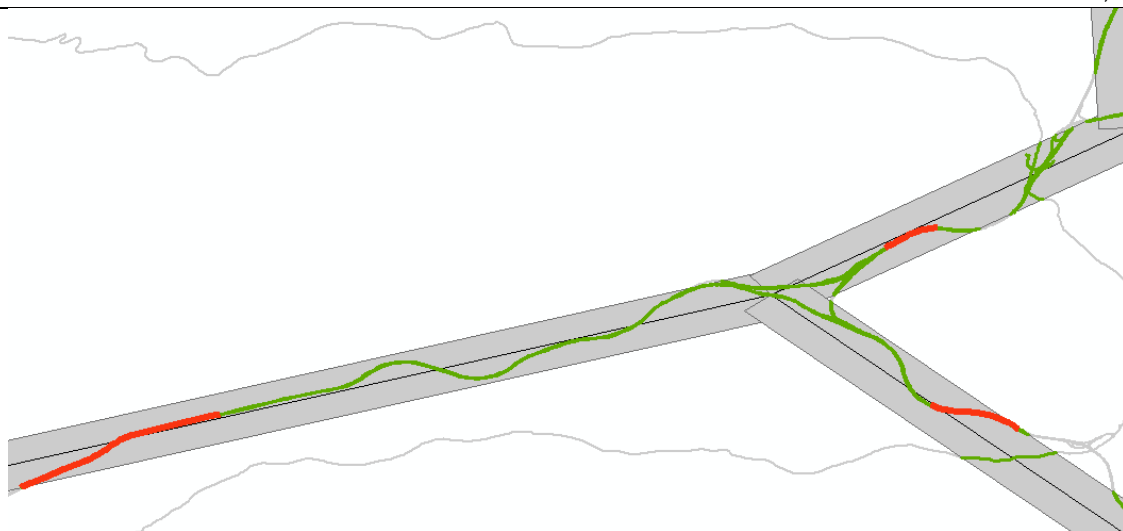
, kde γ_i je úhlová změna v situaci na dílčím úseku i v gradech, l je délka komunikace v kilometrech, j je počet dílčích úseků uvažované části komunikace.

1.3. Kalibrační data ze sčítání dopravy

Po získání vstupních dat pro výpočet dopravních intenzit (viz kapitola 1.1) a následném samotném výpočtu dopravních intenzit (viz kapitola 1.2), bylo třeba výsledné hodnoty intenzit porovnat s reálným vzorkem dat o průjezdu vozidel měřenými úseky, tzv. referenční datová vrstva (zapojení kalibrace viz obrázek 3 - čárkované linie). Kalibrace probíhá iteračním postupem numerické minimalizace objektové funkce. Pro kalibraci vypočtených intenzit dopravy byla zvolena otevřená datová sada celoevropského sčítání dopravy komise UNECE (The United Nations Economic Commission for Europe), viz UNECE (2016). Jako referenční rok sčítání byl zvolen rok 2005 a to z několika důvodů. I když jsou dostupná data ze sčítání z roku 2010, který je blíže k roku statistického sčítání obyvatel provedeného v roce 2011 (viz kapitola 1.1), je toto sčítání zpracováno do podoby GIS projektu a navíc se hodnoty sčítání mezi lety 2005 a 2010 zásadně nezměnily (autoři si jsou vědomi toho, že silniční síť v Otevřené dopravní mapě je aktuální k roku 2014). Datová sada s referenčními daty o sčítání dopravy je v liniové vektorové formě, kdy hodnota sčítání je implementována z bodu, kde se sčítalo na celou délku úseku. Sčítání dopravy nebylo prováděno na silnicích veškerých kategorií, ale jen na evropských mezinárodních silnicích. Vypočtené intenzity jsou tedy kalibrovány na silnicích vyšších tříd.

Tato data jsou však uložena v jiném datovém modelu, než je použit v Otevřené dopravní mapě pro její silniční síť, bylo tedy třeba provést datovou harmonizaci, která byla následující:

- Prostorový průnik referenčních dat s úseky Otevřené dopravní mapy - odlišný není jen datový model obou sad, ale také geometrie, která je v případě referenčních dat velice zjednodušená a není dostupná informace, z jaké datové sady pochází a jakým způsobem byla generalizovaná, byla-li. Byly vybrány ty referenční úseky, na kterých se vyskytovala hodnota RPDI.
- Pro tyto úseky byly vytvořeny obalové zóny (šedé oblasti), které byly použity pro výběr všech úseků sítě OTM, se kterou měly neprázdný průnik (zelené linie). Ze získaných kandidátů byl pro každý úsek obalové zóny měřeného úseku vybrán takový úsek sítě OSM, který byl celý nebo svou částí nejdelší pro dané dopravní měření (červené linie), viz obrázek 5. Šedé linie představují úseky Otevřené dopravní mapy mimo zájmové oblasti kalibrace, rovné černé linie jsou úseky kalibračních dat (tj. středové linie obalových zón).



Obr. 5: Napojení kalibračních úseků na síť Otevřené dopravní mapy

- Napojení hodnot RPD na konkrétní úseky Otevřené dopravní mapy (viz červené linie na obrázku 5).
- Dalším důležitým rozdílem v obou datových seditích bylo to, že v síti Otevřené dopravní mapy je většina řešených jízdních úseků reprezentována dvěma liniemi pro dva jízdní směry. Použitá kalibrační síť dopravního sčítání naopak pracuje jen s jednou linií pro oba směry. Pro maximální využití dopravních dat byly tedy pro všechny úseky Otevřené dopravní mapy, kterým byla přiřazena hodnota RPD a zároveň byly jednosměrné (atribut `direction = inDirection`), nalezena nejbližší sousední linie Otevřené dopravní mapy v opačném směru. Pro tento krok bylo nutné převést linie na body a hledat jejich nejbližší sousedy v této reprezentaci, protože hledáním nejbližšího souseda v liniové reprezentaci docházelo k přiřazení k linii nikoli paralelnímu, ale sériovému.

K dispozici bylo v konečném důsledku, po výše uvedených úpravách, 7 540 kalibračních sčítacích bodů, na kterých byla dostupná informace o RPD. Původní počet úseků, na kterých bylo provedeno sčítání dopravy byl 6 381. Vyšší počet použitých kalibračních bodů je způsoben použitým postupem uvedeným v posledním bodu výše popsané datové harmonizace. Tyto informace byly použity pro kalibraci vypočtených dopravních intenzit, která proběhla ve 20 iteračních krocích (viz další kapitola).

2. VÝSLEDKY EXPERIMENTŮ - NÁROČNOST VÝPOČTU DOPRAVNÍCH INTENZIT PRO EVROPU

Limitujícím prvkem pro výpočet dopravních intenzit se stal počet vstupních prvků a to především počet úseků silniční sítě. Příklady počtu vstupních prvků použité pro výpočet dopravních intenzit na desktopovém a serverovém zařízení jsou uvedeny v tabulce 2, konfigurace použitých výpočetních zařízení je uvedena níže. Jde o reálně použitá a spočtená data, nejedná se o simulace. Na obrázku 6 je zobrazen přehled úseků evropské silniční sítě, na kterých jsou spočteny dopravní intenzity.

Konfigurace desktopového zařízení:

standardní desktopové PC, 8 jádrové CPU, 12 GB RAM

Konfigurace serverového zařízení:

24 strojů: po 4 jádrech (96 jader) a 32 GB RAM (ze 128 GB RAM dostupných pro každý stroj)

Tab. 2: Limity výpočtu dopravních intenzit (desktop vs. server).

	Desktop	Server
Velikost sítě	Dopravní síť ČR, 60 054 hran	Dopravní síť Evropy, necelých 5 mil. hran (zredukováno ze 14 mil. ³)
Počet generátorů dopravy	9 365	156 812 ⁴
Velikost	87 703 225	829 051 938
Celkový výpočetní čas	4 hodiny	35 hodin

Byla snaha vypočítat podrobnější model pro dopravní síť ČR s více jak 22 000 generátory dopravy. Při použití výše zmíněného dopravně-inženýrského desktopového softwaru OmniTRANS spuštěného na výše uvedeném desktopu došlo k selhání tohoto výpočtu, což vedlo k nutnosti přesunutí výpočtů dopravních intenzit nad větším objemem dat na server (tento software nemá svoji serverovou verzi).

Paralelizace a časová náročnost výpočtu dopravních intenzit pro Evropu

Pro paralelní výpočty výpočetního systému Apache Spark je použit programovací model Map-Reduce. Ten je založen na dvou základních operacích - mapování (Map) a redukování (Reduce). Obě tyto operace jsou uskutečňovány v rámci clusteru paralelně. Operace Map je transformační funkcí. Pro každý záznam vstupních dat prování nějaký úkon, nějakou transformaci dat. Operace Reduce kombinuje záznamy ve vstupních datech a redukuje jejich počet na základě daných parametrů. Více informací v Dean & Ghemawat (2008) či v Kolovský & Ježek (2017)

Konkrétní časová náročnost pro výše definovanou úlohu za výše definovaných podmínek výpočtu na serveru (20-ti násobná iterace kalibrování výsledků) je následující:

- stanovení OD Matice: 1.3 h
- kalibrace OD matice: 19.9 h pro 20 iterací => 1 it/h
- přiřazení (assignment): 39 min

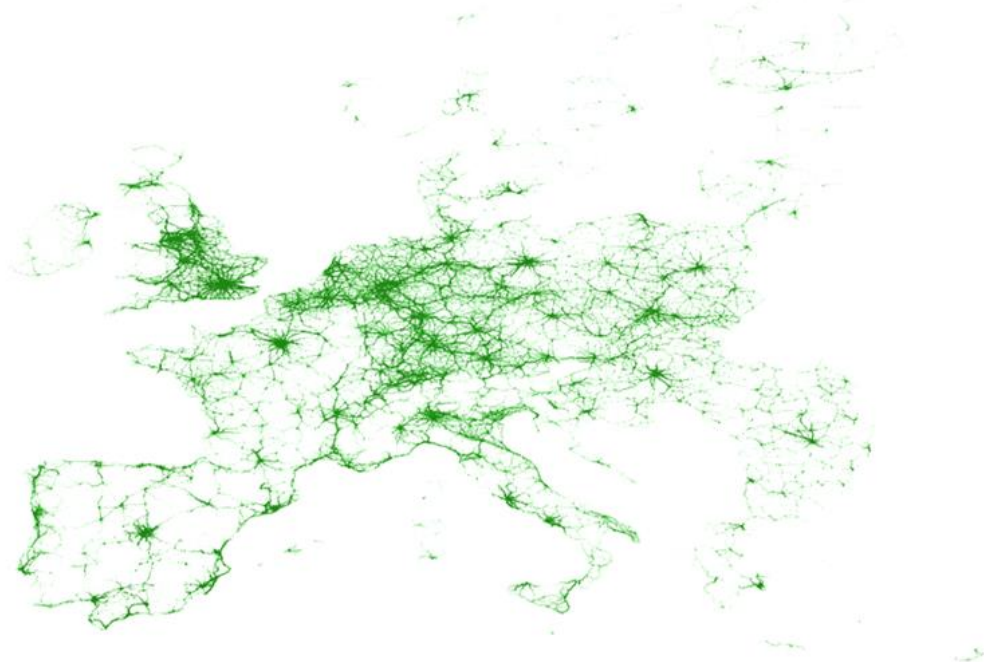
Výpočetní serverová kapacita byla zajištěna virtuální organizací Metacentrum⁵.

Výsledek výše uvedených serverových výpočtů je zobrazen na obrázku 6 níže. Jedná se o síť silnic 3. a vyšší třídy, vyskytující se na území dostupných generátorů dopravy (tzn. tam, kde byly definované Local Administrative Units).

³ byly redukovány uzly stupně 2 pomocí Depth-first search algoritmu (prohledávání do hloubky) z úseků 3. tříd silnic a vyšších

⁴ jedná se o počet generátorů po úpravě procesu zpracování demografických dat Evropy, viz kapitola 3

⁵ <https://metavo.metacentrum.cz/cs/>

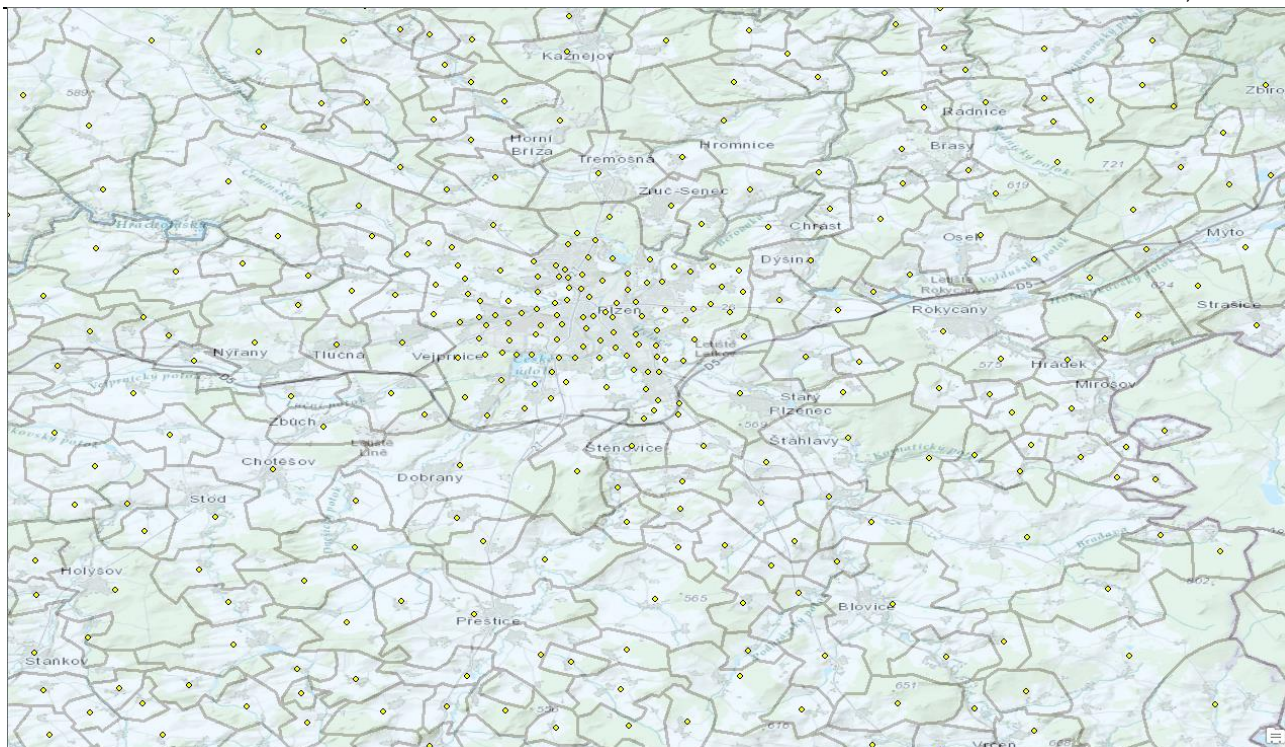


Obr. 6: Přehled evropské silniční sítě s vypočtenými dopravními intenzitami.

3. DISKUZE

V průběhu prací byla zjištěna řada situací, které bylo třeba vyřešit, abychom harmonizační procesy podkladových dat dotáhli do podoby, která vyprodukuje vstupní data vhodná pro navržený algoritmus výpočtu dopravních intenzit. Mezi nimi se nejzásadnějšími ukázaly následující:

Local Administrative Units 2 jsou definované po obcích. Jak je z obrázku 2 A (resp. D) vidět, pro město Plzeň se jedná o výrazně větší území, než je tomu u okolních obcí. Což pro statistické účely má svojí logiku. Pro účely výpočtu dopravních intenzit je však toto plošné rozdělení nevhodné. Pro každý polygon LAU2 je totiž vytvořen generátor dopravy a pro větší obce (cca nad několik desítek tisíc obyvatel) je distribuce obyvatel v takovém místě nedostatečná. A to také z toho důvodu, že silniční infrastruktura bývá v těchto oblastech velmi hustá a jeden bod reprezentující všechny obyvatele obce naprosto nereflexuje jejich rozdělení v obci. Došlo tedy k úpravě postupu tvorby generátorů dopravy. Na základě zjednodušení dopravní sítě v obcích nad 100 000 obyvatel byly definovány Thiessenovy polygony a na jejich podkladě bylo provedeno prostorové spojení se statistickou gridovou vrstvou. Z počtu necelých 120 000 původních generátorů dopravy bylo vytvořeno nových více jak 156 000 generátorů (viz tabulka 2 a kapitola 1.1). Oblast Plzně a okolí z obrázku 2 je po úpravě generátorů zobrazena na obrázku 7. Dalším problémem těchto administrativních jednotek je neúplné pokrytí evropského kontinentu (např. pro část Balkánského poloostrova, Ukrajina, Bělorusko, evropská část Ruska), takže tam, kde nejsou k dispozici tato data, nejsou generované dopravní generátory a tudíž ani není na úsecích silniční sítě v daném území spočtena dopravní intenzita, viz obrázek 6.



Obr. 7: Generátory dopravy použité pro výpočet dopravních intenzit pro Evropu - po úpravě.

U kalibračních dat představuje hlavní problém rozdílná geometrická přesnost, resp. míra generalizace u použitých kalibračních měření oproti geometrii sítě Otevřené dopravní mapy. Zvolený postup přiřazení RPDÍ se ukázal jako dobře použitelný, ale není stoprocentně úspěšný. Důležitá je například velikost použité obalové zóny. U příliš malé hrozí, že se měření nepřihodí žádnému úseku Otevřené dopravní mapy a u příliš velké zóny hrozí naopak přiřazení příliš vzdálenému a tedy chybnému úseku. Kombinace obalové zóny a výběru dle délky průniku se ukázala jako nejefektivnější. Ideální by samozřejmě bylo, kdyby toto přiřazení nebylo prostorové, ale atributové. Taková data ovšem v rámci Evropy nebyly v době řešení dopravních intenzit pro Evropu k dispozici.

Významným limitujícím faktorem pro výpočet dopravních intenzit jsou momentálně výše uvedená kalibrační data. Jejich počet je něco přes 6 000, což je obdobný počet použitý pro kalibraci výpočtu dopravních intenzit na území ČR. Jejich rozmístění není také nejvhodnější v souvislosti s konfigurací silniční sítě. Například v Itálii nejsou k dispozici ze sčítání dopravy UNECE naprosto žádné sčítací body. A v neposlední řadě způsob a kvalita napojení těchto kalibračních dat na správné úseky silnice je nedokonalá (tj. dochází k situaci napojení kalibračních dat na jiný úsek, než na kterém bylo sčítání prováděno - díky zvolenému způsobu přiřazování, viz kapitola 1.3). Toto je hlavním limitujícím prvkem přesnosti finálně vypočtených intenzit dopravy, změna vstupního počtu generátorů dopravy (viz výše) neměla takový zásadní vliv. Manuální přiřazení sčítacích bodů na geometrii sítě by vedlo k rapidnímu vylepšení dosažených výsledků, avšak za cenu násobně vyššího času zpracování dat.

Kromě výše uvedených záležitostí jsou zde další, se kterými je třeba počítat. Veškerá vstupní data jsou otevřená, kromě dat přejímaných z EUROSTAT, navíc jen omezeně ověřitelná. Pro kontext takto velkých dat je třeba zvolit vhodnou vizualizační platformu, která zvládne zpracovat a vizualizovat velké množství dat v reálném čase. Dále je dokumentace k algoritmu výpočtu dopravních intenzit implementovanému na serveru nekompletní, ale počítá se s jejím dokončením. V neposlední řadě, zůstává otázka financování infrastruktury pro udržitelný provoz vytvořené technologie.

ZÁVĚR

Tento příspěvek seznámil čtenáře s další fází tvorby Otevřené dopravní mapy. Tentokrát se jednalo o výpočet dopravních intenzit pro Evropskou silniční síť, konkrétně u 3. třídy silnic a vyšší a s tím spojené kroky. Mezi těmito kroky bylo získání a harmonizace demografických dat za Evropu, které slouží jako podklad pro tvorbu generátorů dopravy. Dále pak harmonizace kalibračních dat o sčítání dopravy na evropských hlavních tazích. V neposlední řadě následovala implementace algoritmu pro výpočet dopravních intenzit na silniční síti, kde výše uvedené datové sady (generátorů dopravy a kalibrační data) byly jejími vstupy. Tento algoritmus byl implementován jako serverové řešení proto, aby bylo vůbec možné takto velký objem vstupních dat zpracovat. V současné době můžeme říci, že jsme vytvořili kompletní workflow pro tvorbu Otevřené dopravní mapy od získání veškerých potřebných podkladových dat (silniční síť, demografická data, kalibrační data), až po výpočet dopravních intenzit nad evropskou silniční sítí. Při zapojení dostatečného výpočetního výkonu lze teoreticky provést výpočet intenzit dopravy na "libovolně velké" síti, včetně přepočtu vypočtených denních intenzit dopravy na hodinové.

PODĚKOVÁNÍ

Tato publikace byla podpořena:

- projektem LO1506 Ministerstva školství, mládeže a tělovýchovy ČR;
- projektem OTN - Open Transport Data CIP-ICT-PSP-PB 620533;
- projektem SGS-2016-004 Využití matematiky a informatiky v geomatice III.

Přístup k výpočetním zařízením patřícím do české Národní Gridové Iniciativy MetaCentrum NGI, na kterých byly prováděny serverové výpočty, byl umožněn díky projektu LM2015042 E-infrastruktura CESNET spadající pod program Projekty velkých infrastruktur pro VaVal.

LITERATURA

Bartoš, L., Martolos, J. (2012) Stanovení intenzit dopravy na pozemních komunikacích, Technické podmínky, II. vydání, Vydal: EDIP s.r.o., Hájkova 1203/32, 301 00 Plzeň, dostupné na adrese: http://www.pjpk.cz/data/USR_001_2_8_TP/TP189.pdf

Dean, J.; Ghemawat, S.: MapReduce: simplified data processing on large clusters. Communications of the ACM, ročník 51, č. 1, 2008: s. 107-113.

Eurostat (2016a) Land Administrative Units (LAU), dostupné na adrese: <http://ec.europa.eu/eurostat/web/nuts/local-administrative-units>

Eurostat (2016b) Population grids, dostupné na adrese: http://ec.europa.eu/eurostat/statistics-explained/index.php/Population_grids#The_GEOSTAT.C2.A02011_population_grid

Eurostat (2016c) STATISTICAL REGIONS FOR EU CANDIDATE AND EFTA COUNTRIES, dostupné na adrese: <http://ec.europa.eu/eurostat/web/nuts/statistical-regions-outside-eu>

Jedlička, K., Ježek, J., Kepka, M., Hájek, P., Mildorf, T., Kolovský, F., Beran, D. Dynamic Visualization of Volume of Traffic (2015a) In Papers ICC 2015. Brazílie: ICA, 2015. s. 1-13. ISBN: 978-85-88783-11-9

Jedlička, K., Mildorf, T., Charvát, K., Kozhukh, D., Charvát Jr., K., Martolos, J., Šťastný, J. (2015b) Benefits of Using Traffic Volumes Described on Examples in the Open Transport Net Project Pilot Regions, In AGRIS Online - Papers in Economics and Informatics, ISSN 1804-1930, VII, 2015, 2, p. 39-46

Jedlička, K., Hájek, P., Čada, V., Martolos, J., Šťastný, J., Beran, D., Kolovský, F., Kozhukh, D. (2016) Open Transport Map - Routable OpenStreetMap. In IST-Africa 2016 Conference Proceedings. Dublin: IIMC International Information Management Corporation Ltd, 2016. s. 1-11. ISBN: 978-1-905824-54-0

Kolovský, F., Ježek, J. (2017) Traffic Volume Modeling in Parallel Computing Environment, diplomová práce, Západočeská univerzita v Plzni, Katedra geomatiky (v tisku)

Kolovský, F. (2016) Výpočet kapacity komunikace – semestrální práce z předmětu Algoritmy prostorových analýz, Katedra geomatiky, Fakulta aplikovaných věd, Západočeská univerzita v Plzni, 2016

Kozhukh, D., Jedlička, K., Mildorf, T., Charvát, K., Charvát K., Jr, Martolos, J., Šťastný, J. (2015) Benefits of Using Traffic Volumes Described on Examples in the Open Transport Net Project Pilot Regions. *Agris On-line Papers in Economics and Informatics* 06/2015; ISSN 1804-1930 (Volume VII, Number 2, 2015):39-46.

Ortúzar, de Dios, J.; Willumsen, L. G. (2011) *Modelling Transport*. John Wiley & Sons, Ltd, 2011, ISBN 978-0-470-76039-0.

UNECE (2016) *Traffic Census 2005*, United Nations Economic Commission for Europe (UNECE), dostupné na adrese: http://www.unece.org/trans/main/wp6/e-roads_census_2005.html

Unwin, A., Theus, M., Hofmann, H. (2006) *Graphics of Large Datasets: Visualizing a Million*, 1 ed. Springer, July 2006.

Veeckman, C., Jedlička, K., De Paepe, D., Kozhukh, D., Kafka, Š., Colpaert, P., Čerba, O. (2017) *Geodata Interoperability and Harmonization in Transport: A Case Study of Open Transport Net*, In *Open Geospatial Data, Software and Standards*, 2017 (v tisku)