

---

# Využití geolokačních dat mobilních operátorů pro potřeby veřejné správy

---

Ministerstvo vnitra ČR



## Předmluva

Využití geolokačních dat ze signalizační sítě mobilních operátorů pro potřeby veřejné správy je dokumentem vytvořeným v rámci projektu „Zlepšení podmínek pro decentralizaci a dostupnost veřejné správy v území“. Tento projekt je realizován Ministerstvem vnitra a je finančně podpořen z Programu Řádná správa v rámci Fondů EHP a Norska 2014 – 2021 ([www.eeagrants.cz](http://www.eeagrants.cz)). Samotný projekt si klade za cíl prostřednictvím využití geolokačních dat optimalizovat strukturu výkonu veřejné správy a navrhnout úpravy územněsprávního uspořádání státu. Za tímto účelem budou sledovány vzorce prostorového chování populace, ze kterých bude následně možné odvodit optimální rozmístění služeb veřejné správy v území. Dokument má teoreticko-metodologický charakter a jeho obsahem jsou doporučené postupy pro pořizování a práci s geolokačními daty v českém prostředí. V rámci projektu bude zároveň sloužit jako hlavní podklad pro specifikaci požadavků pro zadávací podmínky veřejné zakázky na nákup geolokačních dat o meziobecní dojíždě za celé území Česka.

Na zpracování dokumentu se nepodílí pouze pracovníci Ministerstva vnitra, ale též zohledňuje zkušenosti stakeholderů z českého prostředí, a stejně tak znalostí českých i zahraničních expertů, kteří jsou do projektu zapojeni.

Tento dokument představuje základní vhled do problematiky práce s těmito daty, který je podložen současnou praxí z širokého spektra odborných pohledů. Geolokační data mají široké možnosti využití např. v oblastech územního plánování, regionálního rozvoje, lokalizace služeb, analýz mobility osob či plánování dopravní obslužnosti. Využívána jsou však i mimo oblast veřejné správy.

Jedná se o unikátní datový zdroj obsahující informace, které nejsou zjistitelné jinými (konvenčními) metodami. Nicméně i tento datový zdroj má své úskalí a je zapotřebí precizně definovat své požadavky na pořizovaná data, aby výsledné údaje byly relevantní k účelu, za kterým jsou pořizována. To je právě ambicí tohoto dokumentu a v tomto směru by měl být nápomocný potenciálním zájemcům o pořízení těchto dat.

## Obsah

1	Teoretický rámec .....	4
1.1	Úvod .....	4
1.2	Geolokační data mobilních operátorů .....	4
1.3	Síť buněk .....	5
1.4	Druhy technik záznamů .....	8
1.5	Formát dat a jejich specifikace.....	9
1.6	Definice prostorových vzorců chování .....	10
1.7	Mobilní operátoři na českém trhu .....	10
1.8	Zkušenosti a doporučená literatura .....	11
2	Definice požadavků na data .....	14
2.1	Definice sledovaného jevu a měřených ukazatelů .....	14
2.2	Zájmové skupiny a jejich definice.....	17
2.3	Typy sledovaných jevů .....	18
2.4	Typologie jedinců dle jejich pohybu/pobytu .....	20
2.5	Územní rozsah měření .....	21
2.6	Časová dimenze měření.....	22
2.7	Ochrana osobních a lokalizačních údajů .....	26
3	Problematické oblasti definice dat.....	27
3.1	Problémy spojené s územní podrobností měření .....	27
3.2	Problémy vycházející ze samotného fungování sítě.....	30
3.3	Specifické možnosti dodatečných informací ze sítě.....	32
3.4	Možnosti zjišťování tras pohybu .....	33
4	Příklad konkrétního nastavení specifikace dat.....	35
4.1	Účel pořízení dat .....	35
4.2	Předmět, způsob a forma plnění.....	35
4.3	Obecné požadavky na výslednou databázi .....	36

4.4	Metodologické požadavky .....	38
4.5	Měřené jevy/sledované ukazatele .....	39
4.6	Konečný formát výstupů .....	46
4.7	Licenční model.....	47
5	Závěr .....	48
6	Užité zdroje.....	49

# 1 Teoretický rámec

## 1.1 Úvod

Mobilní telefony se postupem času staly nedílnou součástí moderního osobního i pracovního života a jsou již mnohem více než jen nástroji pro komunikaci, pro sdílení a výměnu informací. Mobilní telefon či jiné mobilní zařízení je přirozenou součástí života takřka každého jedince v současné společnosti. Je jedním z osobních předmětů, které se snaží mít stále na blízku, jako například peněženku, klíče či hodinky. Lidé se tak snaží mít jej při svém přemísťování v prostoru stále při sobě či v blízkém dosahu. Mobilní telefony se tak pohybují po většinu dne se svými uživateli a mohou poskytovat informace o pohybu svých majitelů v prostoru a čase ať už se jedná o rozmezí hodin, dnů, týdnů či měsíců. Zkoumání aktivity mobilních telefonů tak může napomoci poskytnout lepší vhled do problematiky mobility obyvatelstva.

Údaje z mobilních sítí mohou být prostředkem k analýze sociálního prostředí, ke konstrukci prediktivních modelů lidské mobility a jsou též perspektivním zdrojem dat pro široké spektrum aplikací v sociálních vědách, urbánním plánování, krizovém managementu, analýze šíření epidemií (Šveda, Barlík 2018), cestovním ruchu, plánování a měření dopravy a v neposlední řadě také plánování veřejných služeb a sítě úřadů veřejné správy.

Skutečné objemy cestujících a přítomné populace v konkrétním místě nejsou na denní bázi k dispozici. Využitím tzv. geolokačních dat ze signalizační sítě mobilních operátorů lze však sledovat vzorce časoprostorového chování populace, ze kterých lze následně odvodit např. optimální rozmístění služeb veřejné správy v území (Bachir 2019).

## 1.2 Geolokační data mobilních operátorů

Princip metody využívání geolokačních dat mobilních telefonů je vcelku jednoduchý. Mobilní telefon či jiné mobilní zařízení se SIM kartou, pokud není vypnuté či mimo dosah signálu mobilní sítě, se automaticky připojuje k nejbližšímu vysílači provozovanému konkrétním mobilním operátorem. Díky tomu, **že je známo, ke kterému konkrétnímu vysílači je mobilní přístroj přihlášen a že je zároveň známo i prostorové umístění tohoto vysílače, lze snadno určit i přibližnou polohu připojeného zařízení** (poloha definována prostorem, který obsluhuje daný vysílač). Takzvaný *mobile positioning* tedy představuje lokalizaci mobilních přístrojů prostřednictvím sítě vysílačů. Tento systém představuje součást sítě GSM, díky níž používáme mobilní telefony pro komunikaci.

Na rozdíl od tradičních zdrojů dat o mobilitě jako je sčítání lidu, domů a bytů či celostátní sčítání dopravy, nebyl tento zdroj dat původně určen pro využití v socio-geografických analýzách a jejich využití je vedlejším produktem telekomunikačního provozu (Šveda, Barlík 2018). V rámci fungování systému GSM jsou data nazývána jako zbytková a jako taková nemají pro chod sítě mnoho dalšího využití.

Základem tohoto systému je **základnová převodní stanice (vysílač)**. Ta je v odborné literatuře často uváděna pod zkratkou **BTS<sup>1</sup>** (z anglického *Base Transceiver Station*). Úkolem BTS je vysílat a přijímat radiový signál. Na území Česka se nachází tisíce těchto stanic, viz statistika níže.

Přibližné počty buněk sítě GSM (pouze 4G sítě) v Česku<sup>2</sup>:

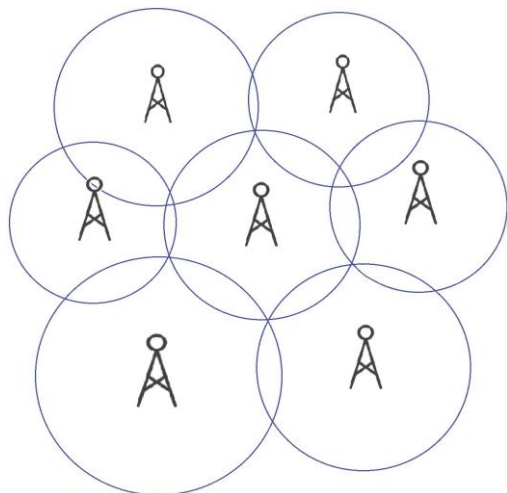
- T-Mobile: 31 932
- O2: 24 288
- Vodafone: 23 181

V českém prostředí je lokalizace stanic BTS veřejně dostupná. Prostřednictvím mapové služby GSMweb.cz je možné dohledat základní informace o každé stanici BTS v síti včetně její přesné lokalizace (GPS). Neposkytuje však informace o oblasti, která je touto stanicí obsluhována (blíže popsáno v následující podkapitole).

### 1.3 Síť buněk

Každá BTS vytváří tzv. buňku, servisní buňku, či servisní polygon. Tou se rozumí oblast pokrytá signálem jedné konkrétní BTS. Tuto oblast daná stanice obsluhuje (viz obr. 1). **Mobilní telefon/SIM kartu je možné lokalizovat v konkrétní buňce**. Nicméně v rámci buňky se může vyskytovat prakticky kdekoliv<sup>3</sup>. Platí tak, že čím je buňka rozlohou menší, tím je přesnost lokalizace přístroje v území vyšší.

Obrázek 1: Schematické znázornění rozmístění BTS v území



Zdroj: vlastní zpracování

signálu, tak i charakter reliéfu a zejména v urbanizovaných oblastech i zástavba. Proto je vytvoření

Z geografického hlediska je možné území rozdělit do oblastí (buněk), které obsluhují jednotlivé stanice (BTS). Buňky bývají nejčastěji zjednodušeně zobrazovány jako kružnice kolem BTS vzájemně se z části překrývající. Dále bývají pro zjednodušení vyjadřovány hexagonálními polygony, které **dokonale vyplňují prostor, aniž by se vzájemně překrývaly (na rozdíl od reality, kde naopak k překrývům dochází)**, a často také Voronoiovými diagramy (viz obr. 2, dole uprostřed). Tyto zjednodušené modely jsou utvářeny z důvodu složitého určení šíření rádiových vln, kdy do hry vstupuje jak kolísavá síla

<sup>1</sup> Zkratka je dále v textu využívána pro popis těchto stanic/vysílačů.

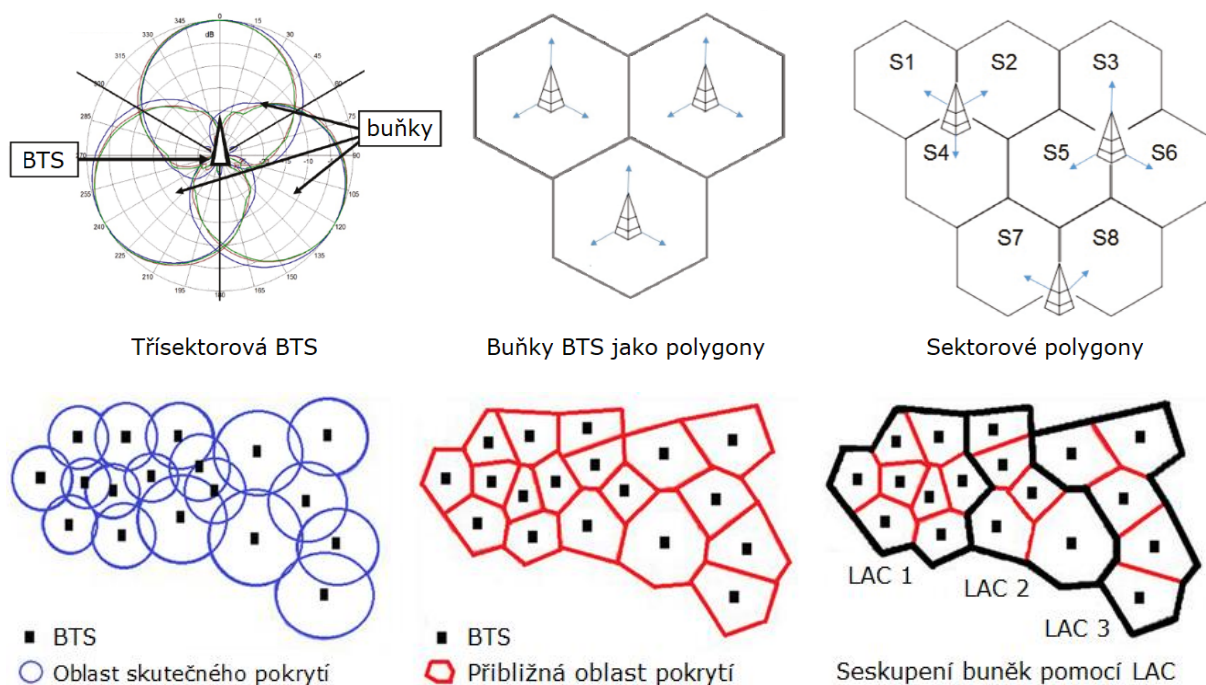
<sup>2</sup> Zdroj GSMweb.cz. Údaje jsou zde pro zjednodušení uváděny jen za 4G síť, která je u všech operátorů v Česku nejrozvinutější. Ostatní síť 3G či 2G mají počty buněk významně nižší a jsou doplňkem této sítě.

<sup>3</sup> Geolokační data mobilních operátorů tak neumožňují přesnou lokalizaci mobilního přístroje/SIM karty jako např. GPS lokalizace, ale její přesnost je omezena na konkrétní buňky příslušné k jednotlivým BTS.

realitě přesně odpovídajícího modelu prostorového rozložení buněk obtížné a výše uvedené zjednodušující metody se jeví jako vhodný kompromis (Bachir 2019).

Každá BTS je schopna pokrýt určité (omezené) území a obsloužit určitý (omezený) počet klientů (Šveda, Barlík 2018). Mobilní operátoři proto vytvářejí síť BTS tak, aby dosáhli co největšího pokrytí signálem a zajistili obslužnost všech uživatelů. **Mobilní telefon při komunikaci s přenosovou sítí neustále shromažďuje a vyhodnocuje údaje o počtu stanic ve svém okolí a síle jejich signálu.** Na základě aktuální nabídky dostupných BTS v okolí si pak pro komunikaci se sítí vybírá zpravidla tu nejbližší s nejsilnějším signálem. Při příliš vysokém zatížení sítě může dojít k situaci, kdy bude mobilní telefon připojen ke vzdálenější BTS se slabším signálem (Novák 2010). Často tak dochází k situaci, že je SIM karta připojena k jiné BTS, než by dle modelu servisních buněk měla být.

Obrázek 2: Různé možnosti zobrazení sítě BTS



Zdroj: Oliver a kol. (2015); upraveno autorem.

Pozn.: LAC = Local Area Code. Mnoho sousedních buněk bývá seskupeno do zón identifikovaných právě pomocí LAC.

Velikost každé buňky závisí jak na fyzicko-geografických podmínkách prostoru, tak na předpokládané míře využití vysílače. Dosah vysílače BTS může nabývat několika stovek metrů, ale také až 35 kilometrů. Zejména **ve venkovských řídko osídlených oblastech pokryje jeden vysílač BTS rozsáhlé území.** Zato **v centrech měst je síť vysílačů hustší,** neboť vedle zvýšených nároků na provozní kapacitu každé buňky zde hraje významnou roli faktor odstínění signálu městskou zástavbou.<sup>4</sup> Proto má ve městech svou BTS

<sup>4</sup> Jeden z mobilních operátorů působících na českém trhu se k této problematice vyjadřuje následovně: „Kvalitu signálu ovlivňuje řada obecných faktorů, jako jsou zástavba v lokalitě (na volném prostranství je větší rozptýl signálu) či reliéf krajiny, ale také faktory přímo u uživatelů – např. materiál oken, tloušťka zdi nebo typ zařízení.“ (T-Mobile)



takřka každá ulice či blok. Kromě urbanizovaných území je hustší síť i v jiných místech, kde dochází k vyšší koncentraci uživatelů, např. rekreační a turistické oblasti, významné dopravní tahy aj. Logicky jsou pak v místech s větší hustotou BTS servisní buňky menší a tím pádem je lokalizace zařízení přesnější.

S rostoucí vzdáleností od BTS klesají přijímané výkonové úrovně (laicky řečeno signál a kapacita) poměrně plynule, po určité vzdálenosti však prudce klesnou a ztratí se v šumu.

Obrázek 3: Různé podoby antén BTS



Zdroj: GSMweb.cz

Obecně platí, že všechny sousedící buňky se vzájemně překrývají, a tak je zabezpečeno souvislé připojení k síti, když je mobilní zařízení v pohybu. (Thuillier 2019).

K připojení telefonu/SIM karty k BTS může dojít jednak v důsledku **využívání služeb operátora** (1) nebo v rámci tzv. **periodického updatu** (2).

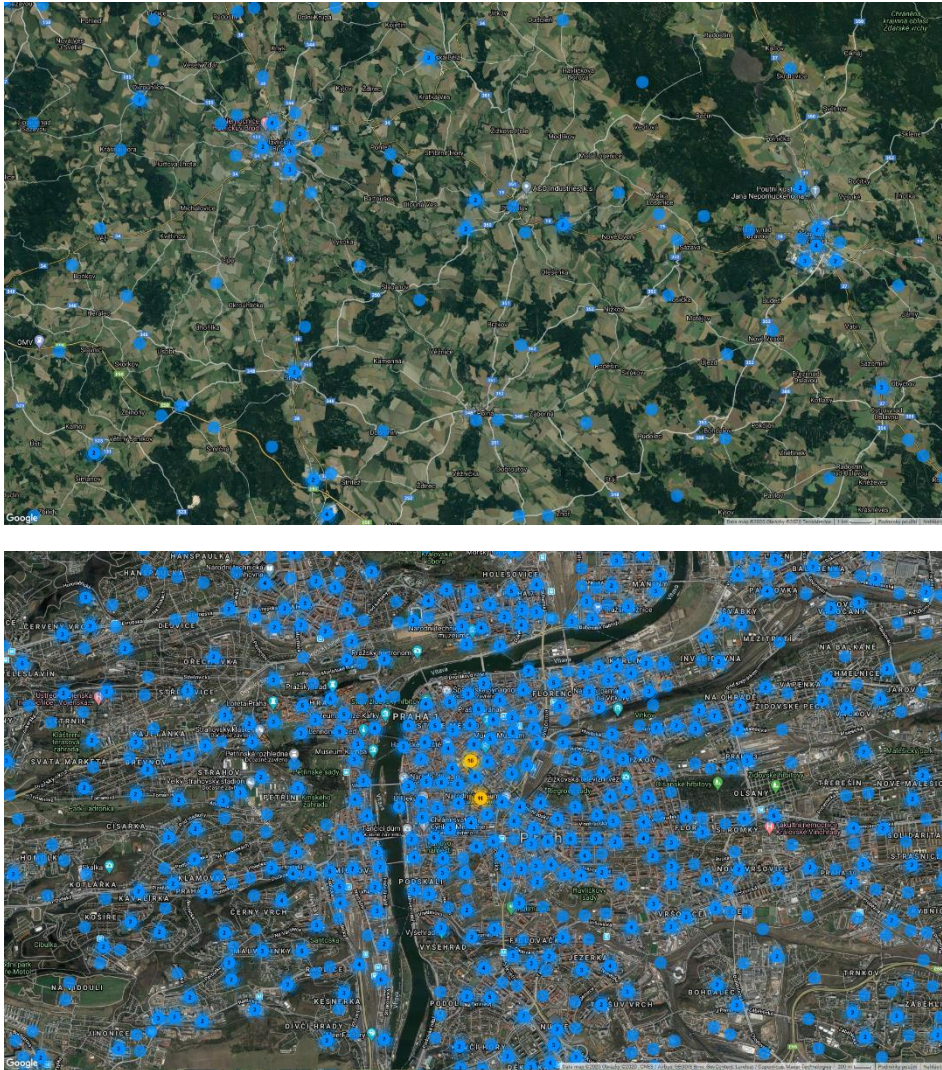
**Ad 1) služby operátora** zahrnují zprostředkování hovorů, kdy během samotného hovoru se neustále obnovuje připojení zařízení k BTS a v případě pohybu jedince může dojít i ke změně BTS. Dalšími službami může být zasílání zpráv SMS, MMS (jednorázové připojení k síti) či datové služby, během jejichž čerpání dochází rovněž k průběžnému (neustálému) připojování k síti.

**Ad 2) periodický update** je připojení, které v závislosti na typu sítě (2G, 3G, 4G, 5G) každé zařízení v určitých intervalech provádí. Není k tomu potřeba aktivní využívání služeb operátora. Dané zařízení vyhodnotí aktuální dostupnost signálu a připojí se k nejbližší („nejvýhodnější“) BTS (jednorázové připojení).



Z jednotlivých BTS, ke kterým se během dne zařízení postupně připojovalo, lze sestavit model časoprostorového chování každého uživatele.

Obrázek 4: Srovnání hustoty sítě BTS v oblasti Českomoravské vrchoviny a v Praze



Zdroj: GSMweb.cz

## 1.4 Druhy technik záznamů

Existuje několik způsobů lokalizačních technik. Poloha zařízení může být určena např. pomocí identifikace buňky, ke které je zařízení připojeno; triangulační vzdáleností od antény či mixem technik využívajících GPS. Pro potřeby této metodiky však stačí zúžit dělení na dva hlavní přístupy k lokalizaci prostřednictvím záznamů v signalizační síti, kdy dochází právě k identifikaci buňky: pasivní a aktivní přístup.

**Pasivní data** jsou automaticky ukládána do paměťových souborů mobilních operátorů. Tato data nejsou závislá na aktivitě uživatele. Proto poskytují celistvější informace než níže uvedená data aktivní. S tisíci

záznamů a vysokým časoprostorovým rozlišením zajišťují pasivní data dlouhodobé sledování a sbírání jedinečných údajů (Šveda, Barlík 2018). Jejich objem je větší z důvodů vyšší frekvence jejich sbírání. Pasivní záznamy mohou vznikat jakoukoli interakcí mezi mobilním zařízením a sítí, tj. využíváním služeb operátora i periodickým updatem (popsáno v předešlé kapitole). V dalších částech této metodiky jsou uvažována data pasivní.

**Aktivní mobilní polohovací data** představují data, kde je poloha zařízení určena pomocí speciálního dotazu pomocí rádiové vlny. Jedná se o alternativní způsob získávání dat mobilních operátorů. Tato data se týkají odchozích nebo příchozích hovorů, zpráv a též užívání mobilních dat, tedy úmyslné aktivní činnosti uživatele mobilního telefonu (proto aktivní data). Jde tedy o mobilní síťovou komunikaci mezi dvěma telefony, a to jak dokonanou, tak o pokusy o spojení mezi zařízeními. Každý záznam obsahuje anonymizované ID volajícího, časové razítko s dobou trvání, ID telekomunikačního zařízení (ID buňky) a typ hovoru – příchozí, odchozí, hlasový, textový (Bachir 2019).

## 1.5 Formát dat a jejich specifikace

Ačkoli geolokační data mají široké uplatnění napříč nejrůznějšími obory, primární záznam v síti mobilních operátorů sám o sobě má jen omezenou vypovídající schopnost.

Záznam se skládá ze tří informací:

1. identifikace konkrétní SIM karty
2. čas, ve kterém k aktivitě došlo
3. identifikace konkrétní BTS, která aktivitu provedla/zaznamenala (periodický update či aktivní užívání služeb operátora).

Schéma 1: Záznam v síti mobilních operátorů

Záznam v signalizační síti mobilních operátorů		
Identifikace jedince	Časová identifikace	Lokalizace
ID SIM karty	Čas záznamu	ID BTS

Z uvedených informací vyplývá, že konkrétní SIM karta se nacházela v dosahu konkrétní BTS v konkrétním čase (viz schéma 1). Každá SIM karta může učinit během dne desítky až stovky záznamů na různých místech. Kontinuálním sledováním lze zjistit pravidelnosti v prostorových vzorcích chování jednotlivých uživatelů. Samotná informace o pohybu SIM karet v záznamech neexistuje. Pohyb je odvozen od faktu, že v chronologicky navazujících záznamech se SIM karta přihlašuje k různým BTS. Tím je definován předpokládaný pohyb mezi jednotlivými buňkami. Za předpokladu, že se mobilní telefon

pohybuje společně se svým uživatelem, lze ztotožnit pohyb SIM karty<sup>5</sup> s pohybem konkrétního uživatele<sup>6</sup>.

## 1.6 Definice prostorových vzorců chování

Poskytovatel dat (primární zpracovatel), kterým je nejčastěji mobilní operátor, má k dispozici obrovské množství údajů z provozu mobilních zařízení. Jedná se o data, která jsou v primární podobě, ze které nelze jednoduše vyčíst informace o mobilitě obyvatelstva. Poskytovatel dat musí nejprve data anonymizovat, aby osobní identita vlastníka telefonu, a tedy i jeho časoprostorové chování, zůstaly zcela skryty. Poté již může být přistoupeno k tzv. agregaci dat, kdy se ze surových dat stávají data vhodná pro analyzování časoprostorových vzorců chování sledovaného vzorku populace.

Významným prvkem lidského života je opakovatelnost lidských činností. Pověšinou se **aktivity opakují v určité periodicitě**. Opakovatelnost se však neprojevuje pouze v časové rovině, ale i v prostorové, díky čemuž se **formují časoprostorové vzorce chování obyvatel**. Nepravidelné činnosti tak lze považovat za výjimečné, občasně či nahodilé jevy. Detailní specifikace požadovaných dat by měla odrážet v bližší definici sledovaných jevů (časové a územní) právě tyto pravidelnosti či nepravidelnosti (více kapitole 2).

Vzorce časoprostorového chování obyvatel lze sledovat prostřednictvím identifikace tzv. **významných míst výskytu**. Významné místo bývá definováno jako místo, kam se sledovaný jedinec s určitou pravidelností vrací a je tak pro něj z prostorového hlediska důležité. Vhodný koncepční rámec pro identifikaci významných míst je koncept kotevních bodů („*anchor points*“), které reprezentují hlavní body lidských aktivit. Nejčastěji se tak jedná o místa nezbytná pro určení dojížděkových vztahů – místo, kam vlastník mobilního telefonu dojíždí za zaměstnáním, a místo, kde tráví nejvíce mimopracovního času (domov). Zjednodušeně se jedná o identifikaci míst denní a noční lokalizace. Údaje o pohybech mezi těmito místy pomohou k pochopení a konstrukci modelu dopravních proudů a socio-ekonomického chování obyvatelstva (Šveda, Barlík 2018).

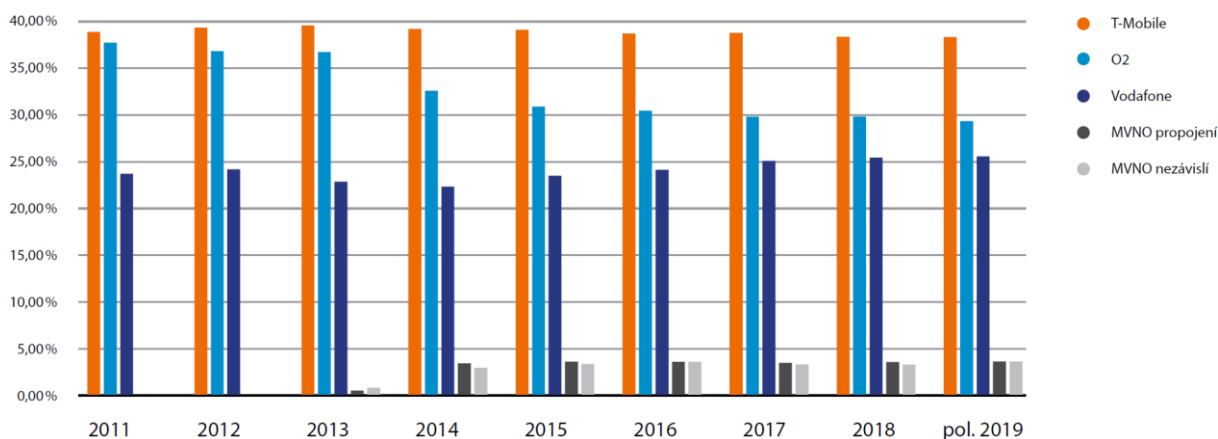
## 1.7 Mobilní operátoři na českém trhu

Dle poslední publikované výroční zprávy Českého telekomunikačního úřadu za rok 2019 činil na konci tohoto roku na území Česka celkový počet aktivních SIM karet takřka 15 milionů (meziroční nárůst zhruba o 208 tisíc). Již mnoho let českému trhu dominují tři největší poskytovatelé mobilních telekomunikačních služeb – T-Mobile, O<sub>2</sub> a Vodafone, kteří z 93 % pokrývají poptávku českých zákazníků.

<sup>5</sup> Problematika SIM karet, které nejsou součástí mobilních telefonů (technické SIM karty aj.) je řešena v příslušné kapitole 3.2.

<sup>6</sup> V textu tohoto dokumentu jsou proto pojmy „pohyb SIM karty“ a „pohyb uživatele“ používány jako synonyma.

Graf 1: Vývoj tržních podílů na základě celkového počtu aktivních SIM karet (v %)



Zdroj: Český telekomunikační úřad - Výroční zpráva za rok 2019

Pozn.: MNVO = operátor virtuální mobilní sítě

## 1.8 Zkušenosti a doporučená literatura

- Metodika „Limity využití mobilních sítí ve statistických šetřeních ČSÚ“

Cílem metodiky je definování základních úskalí práce s daty ze signalizační sítě, přičemž metodika sama obsahuje výčet základních oblastí, které musí v rámci projektu naplnit jak zadavatel, tak zpracovatel. Základním cílem metodiky je dosáhnout srovnatelnosti, a hlavně použitelnosti výstupů založených na datech ze signalizační sítě mobilních operátorů. Více informací [zde](#).

- Měření přítomného obyvatelstva města Brna (2016)

Datová sada od mobilního operátora obsahuje informace o počtu přítomného obyvatelstva v administrativní jednotce. Dataset obsahuje Kód ZSJ, den v týdnu, čas a počet přítomného obyvatelstva. Sběr dat byl proveden pomocí signalizační sítě mobilního operátora a byl měřen v září a říjnu roku 2016, na základě čehož se pak stanovil průměrný pracovní a víkendový den. Více informací [zde](#) a [zde](#).

- Pohyb obyvatelstva na základě dat mobilního operátora – Brněnská metropolitní oblast (2017)

Jaké jsou hlavní dopravní proudy v Brně? Odkud a kam obyvatelé Brna a jeho zázemí nejvíce cestují? Kolik cest se v Brně za den uskuteční? Na tyto a další otázky odpovídá analýza geolokačních dat mobilního operátora T-Mobile. Více informací [zde](#).

- ŠVEDA, M., BARLÍK, P. (2018): Daily commuting in the Bratislava metropolitan area: case study with mobile positioning data.

Předmět výzkumu: Analyzování vzorců mobility dojíždění v metropolitní oblasti.

Použitá metoda geolokace: pasivní

Odkaz na článek [zde](#).



- NOVÁK, J. (2010): lokalizační data mobilních telefonů: možnosti využití v geografickém výzkumu.

Předmět výzkumu: Prostorové organizace společnosti a mobilita v urbánním prostoru.

Použitá metoda geolokace: aktivní, pasivní

Odkaz na práci [zde](#).

- FRÍAS-MARTÍNEZ, V., SOGUERO RUIZ, C., FRIAS-MARTINEZ, E. (2012): Estimation of urban commuting patterns using cellphone network data.

Předmět výzkumu: Konstruování matic generovaných geolokačními daty, které zachycují stejné vzorce jako tradiční dojížděkové matice.

Použitá metoda geolokace: aktivní

Odkaz na článek [zde](#).

- AHRİPOVA, I. a kol. (2019): Mobile phone data statistics as a dynamic proxy indicator in assessing regional economic activity and human commuting patterns.

Předmět výzkumu: Analyzování úrovně ekonomické aktivity obyvatelstva v závislosti na aktivitě mobilních telefonů.

Použitá metoda geolokace: aktivní

Odkaz na článek [zde](#).

- YANG, X. a kol. (2018): Understanding the Spatial Structure of Urban Commuting Using Mobile Phone Location Data: A Case Study of Shenzhen, China.

Předmět výzkumu: Nalezení prostorových struktur na základě dojížděkových charakteristik vyplývajících z analýzy dat mobilních operátorů.

Použitá metoda geolokace: pasivní zaznamenávání polohy v hodinových intervalech

Odkaz na článek [zde](#).

- BACHIR, D. (2019): Estimating urban mobility with mobile network geolocation data mining.

Předmět výzkumu: Disertační práce se snaží odpovědět na široké spektrum otázek v rámci hlavní výzkumného cíle v podobě získávání dat pro potřeby určení míry městské mobility.

Použitá metoda geolokace: aktivní

Odkaz na článek [zde](#).

Další zdroje informací:

- OLIVER, N., MATIC, A., FRIAS-MARTINEZ, E. (2015): Mobile Network Data for Public Health: Opportunities and Challenges. Odkaz na článek [zde](#).
- THUILLIER, E. a kol. (2019): Clustering Weekly Patterns of Human Mobility Through Mobile Phone Data. Odkaz na článek [zde](#).
- CACERES, N., WIDEBERG, J. P., BENITEZ, F. G. (2007): Deriving origin–destination data from a mobile phone network. Odkaz na článek [zde](#).

## 2 Definice požadavků na data

Pro využívání geolokačních dat mobilních operátorů se uživatel musí vypořádat s řadou aspektů, které je zapotřebí pochopit, aby mohl jednoznačně **definovat své zadání/požadavky na data**. Tato kapitola v jednotlivých částech popisuje všechny aspekty, které by měl pořizovatel dat zvážit při formulaci zadání pro dodavatele, aby co nejkonkrétněji specifikoval požadovaná data. Pořizovatel si musí být zcela jist, **jaká data a za jakým účelem potřebuje**, přesně je definovat a následně je zahrnout do zadávací dokumentace jako předmět plnění. Obsah zadávací dokumentace je pak pro dodavatele (poskytovatele dat) závazný. Popsání níže uvedených aspektů do datové specifikace lze z pohledu uživatele považovat za přípravnou fázi vlastního procesu pořízení dat.

- Definice sledovaného jevu a způsobu měření
- Definice zájmové skupiny
- Typy sledovaných jevů
- Typologie jedinců
- Územní dimenze měření
- Časová dimenze měření
- Vypořádání se s technickými problémy fungování sítě
- Zvážení možnosti dodatečných informací o uživateli
- Požadavky na vysvětlení metodických postupů

### 2.1 Definice sledovaného jevu a měřených ukazatelů

Každý subjekt, který pořizuje geolokační data, má určitá očekávání a určitý důvod či motivaci k jejich využívání.

- OTÁZKA 1: Za jakým účelem data pořizujeme a čemu by měla sloužit?

Tato zdánlivě triviální otázka je stěžejní, jelikož stanovuje, na co mají data odpovědět. Definujeme tak účel analýzy/sledování.

- OTÁZKA 2: Jakým způsobem budeme s daty nakládat? (aplikované metody)

Ve druhé fázi definice požadavků je zapotřebí si ujasnit, jaké metodické postupy chceme použít. Metody by měly být relevantní, aby jejich výstup odpovídal na výzkumné otázky, definované v první fázi.

- OTÁZKA 3: Jaké jevy budeme sledovat?

V odpovědích na otázky 1 a 2 je specifikováno, co a jak chceme zkoumat. Z identifikace těchto kroků již samo vyplyne, jaká data potřebujeme (zjistíme konkrétní jevy, které je zapotřebí sledovat).



Schéma 2: Definice měřeného jevu



Požadavky na data jsou proto přímo výsledkem definice předmětu, objektu (co zkoumat) a metod studia (jak zkoumat). Požadovaná data musí mít takovou podobu, aby odpovídala stanoveným metodám a umožňovala provedení daných analýz. Otázka „**Jaká data potřebujeme?**“ tak není v tomto smyslu otázkou, ale spíše odpovědí, jelikož vyplývá z předchozích kroků (viz schéma 2).

Data ze signalizační sítě mobilních operátorů jsou sice cenným zdrojem, avšak i ona mají svá omezení (popsaná dále v textu). Je zapotřebí na ně brát zřetel a klást si jen takové výzkumné otázky, které je reálné těmito metodami zodpovědět.

#### Jak jevy měřit?

Samotný sledovaný jev je zprvu stanoven velmi obecně a je potřeba jej specifikovat do takové míry, abychom měli zcela konkrétní měřitelné ukazatele. V první řadě je zapotřebí k **obecně definovanému sledovanému jevu** nalézt relevantní **teoretický ukazatel**, jenž popisuje daný jev. Posledním krokem je stanovení **konkrétního měřitelného ukazatele**, jenž lze z dat, která jsou k dispozici, získat.

**1) obecně definovaný jev** (př. počet obyvatel lokality)

**2) teoretický ukazatel** (př. počet osob nocujících v dané lokalitě)

**3) konkrétní měřitelný ukazatel** (př. počet osob vyskytujících se v lokalitě od 0:00 do 4:00)

Konkrétní ukazatel (bod 3) musí být stanoven tak, aby co možná nejvíce odpovídal realitě, tj. sledovanému jevu (bod 1). Nicméně ukazatel zpravidla nikdy nepopíše sledovaný jev zcela přesně, vždy bude určitá **tzv. adekváční mezera**. Z výše nastíněného příkladu je zřejmé, že počet osob vyskytujících se v lokalitě mezi 0:00 – 4:00 v sobě nezahrne pouze osoby, které v lokalitě reálně žijí. To je právě adekváční mezera, kterou je zapotřebí vhodným nastavením ukazatele co možná nejvíce snížit, aby ukazatel co nejvíce popisoval sledovaný jev.

#### Praktický příklad:

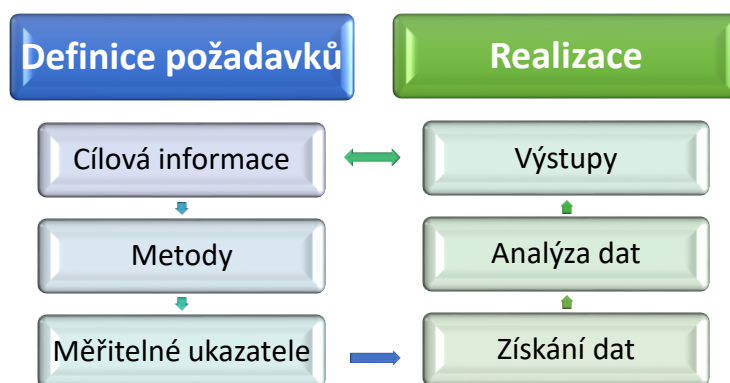
Město chce analyzovat možnosti vedení linek noční autobusové dopravy. Linky by měly obsluhovat lokality, kde je významná koncentrace osob v nočních hodinách (obecně definovaný sledovaný jev). Sledovaný kazatel lze tedy definovat jako místa výskytu osob aktivních v nočních hodinách (teoretický ukazatel). Konkrétně je pak sledována četnost výskytu unikátních SIM karet (osob) v každé servisní buňce ve městě v půlhodinových intervalech, přičemž jsou sledovány pouze SIM karty, u kterých byl zaznamenán pohyb v době mezi 23.00 – 5:00 (konkrétní měřitelný ukazatel).

Parametry měřitelného ukazatele je vhodné nastavit tak, aby výsledné údaje co nejvíce odpovídaly sledovanému jevu a nebyly započítávány SIM karty, které mohou zkreslit výpovědní hodnotu pro tyto účely (např. noční tranzit či noční směny přesahující měřený interval). Cílem je nastavit parametry tak, aby adekvátní mezera byla zanedbatelná.

#### Plánování výzkumu

Přípravná fáze v podobě definice požadavků na data (resp. Datová specifikace) je plně provázána s realizační fází výzkumu. **Jedná se o vstupní požadavek pořizovatele** pro poskytovatele dat (mobilního operátora), který na základě datové specifikace doručí zpracovaná lokalizační data. **Pouze kvalitní datová specifikace umožní získat validní výstupy.** Z tohoto důvodu je zapotřebí jí věnovat maximální pozornost a jednotlivé procesy detailně naplánovat a vzájemně provázat.

Schéma 3: Definice požadavků na data



Poskytovatel dat dá k dispozici přesně ta data, která si tým pořizovatele definuje ve svých požadavcích. Stanovené požadavky lze jen velmi obtížně měnit, z čehož plyne, že pořizovatel musí mít předem jasnou představu o svých požadavcích na data, o metodách, které na ně chce aplikovat a musí být schopen predikovat, jaké by měly být očekávané výstupy. V každé fázi tak musí být provázány jednotlivé procesy fáze plánování a fáze realizace (viz schéma 3).

ÚKOL: Specifikujte požadovaná data

- odpovězte na otázky 1, 2 a 3
- určete, co ve Vašem výzkumu budete chtít zjistit a jakými metodami při tom budete postupovat
- výsledkem úvahy jsou data, která potřebujete, abyste k požadovaným výsledkům za použití požadovaných metod došli

## 2.2 Zájmové skupiny a jejich definice

Definicí zájmové skupiny jednoduše myšleno, na koho je daný výzkum zaměřen. Dochází tak k faktickému zúžení celkového rozsahu zkoumaných SIM karet pouze na ty, které odpovídají zaměření výzkumu.

Velká část výzkumů však bývá zaměřena na studium **celé populace**, kde není zapotřebí **základní objem SIM karet** v síti jakoukoli filtrací omezovat (vyjma řešení technických nesrovnalostí v síti). Specifické výzkumy či detailní územní studie již potřebují z celkového počtu SIM karet **vyfiltrovat pouze ty, kterých se daný výzkum dotýká**. Výběr zájmových skupin musí být jednak věcně smysluplný a zároveň i technicky proveditelný. Parametry pro filtrování základního souboru tak musí být schopny definovat cílové jedince na základě specifické hodnoty dvou měřitelných jevů, tj. pobytu či pohybu<sup>7</sup> SIM karty/uživatele v určitém čase.

Praktický příklad:

Zájmové skupiny mohou mít (dle zaměření výzkumu) například následující podobu:

- pouze uživatelé, kteří se v den konání hudebního festivalu vyskytovali v buňkách BTS, které mají dosah do areálu festivalu.
- pouze uživatelé, kteří se alespoň 2x za měsíc vyskytnou v krajském městě a stráví v něm dohromady alespoň 2 hodiny
- pouze uživatelé, kteří s pravidelností alespoň 3x týdně v době od 00:00 do 04:00 vykazují pohyb v síti (aktivně se pohybují v noci)
- pouze uživatelé, kteří strávili v daném roce alespoň 100 dní v ČR

SIM karty, obsažené ve výsledném zájmovém souboru lze dále vnímat jako základní prvky měření. **Každá SIM karta má v měření stejnou váhu, stejný význam**. Z podstaty informace zaznamenané v síti nelze o jednotlivých SIM kartách zjistit bližší informace například o pohlaví, věku či vzdělání jedinců (problematice dodatečných informací ze sítě se detailněji věnuje podkapitola 3.3)

<sup>7</sup> Blíže k problematice měřitelných jevů (pohyb a pobyt) v následující kapitole „Typy sledovaných jevů“.

### Přepočet výsledných dat na populaci

Analyzované údaje se vztahují ke sledovanému vzorku SIM karet. Nicméně tyto údaje je poměrně problematické vtáhnout k celkové populaci. Vzorek SIM karet má vypovídající hodnotu zejména ve svých relativních hodnotách nikoli v absolutních. I v případě, že by poskytovatel dat analyzoval všechny aktivní SIM karty v Česku (14,7 mil k roku 2018), není jednoduché výsledky v absolutních hodnotách zobecnit na celou populaci. Zobecnitelnost je složitá i z důvodu, že vzorek zpravidla nadhodnocuje určité skupiny populace (zejména lidé v produktivním věku)<sup>8</sup>.

**K zobecnění je nutné využít vhodných matematických operací.** Ty zohledňují určité parametry, jako je zejména podíl daného operátora na místním trhu, dále např. lokální demografickou struktura. Vždy se však jedná **odhad či model s určitou mírou přesnosti**. Jedním z problémů je v tomto směru referenční rámec, tj., databáze údajů, ke kterým se mají data vztahovat. Geolokační data jsou alternativou ke konvenčním zdrojům, jako je evidence obyvatel či údaje ČSÚ. Nelze proto geolokační data vztahovat či kalibrovat k těmto datovým sadám, jelikož by došlo ke ztrátě unikátnosti informace. Nicméně odchylka výsledných hodnot od údajů v konvenčních zdrojích by neměla být značná.

V případě, že účel výzkumu nevyžaduje zobecnění na celou populaci, je vhodné se těmto přepočtům vyhnout a nezatěžovat tak data nepřesnostmi tím způsobenými. Pokud ovšem k přepočtu na populaci dochází, je naprosto nezbytné, aby poskytovatel dat zcela věrohodně a podrobně popsal, jakými operacemi s daty k těmto odhadům došel.

ÚKOL: Definujte zájmovou skupinu

- definujte parametry, pro vymezení souboru SIM karet, které budou základním souborem uživatelů sítě, o kterých chcete získat data
- zvažte, zda je nutné zobecnění výsledných hodnot na populaci, v případě že ano, požadujte podrobné vysvětlení metodického postupu.

## 2.3 Typy sledovaných jevů

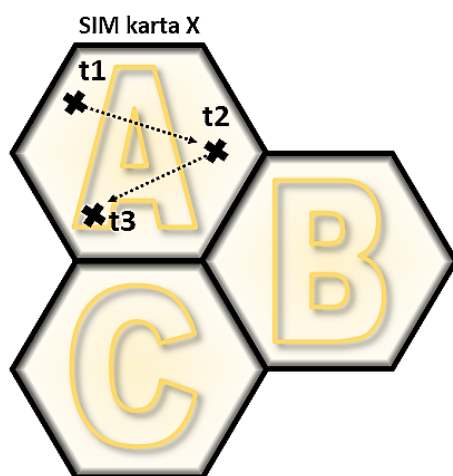
Data, která lze ze záznamů v signalizační síti získat, mají dvě základní podoby. Jsou orientovány na sledování **pobytu** (statický stav) **nebo pohybu**<sup>9</sup>. Při rozlišení těchto jevů je nutné vycházet ze zjednodušeného předpokladu – když se SIM karta přihlásí k určité BTS, máme **informaci o výskytu (pobytu) v dané lokalitě**.

<sup>8</sup> V populaci existují skupiny osob, které nemají možnost či schopnost využívat mobilní zařízení, tj. zejména děti přibližně do věku 10 - 12 let, některé skupiny seniorů či lidé se speciálními potřebami. Lze odhadovat, že osob bez mobilního zařízení může být i více než 15 procent všech obyvatel (přesné údaje však neexistují, jelikož je užívání mobilních telefonů sledováno pouze u osob starších 16 let).

<sup>9</sup> Zatímco pobyt je v datech obsažen přímo, pohyb v datech zaznamenan není, je pouze usuzován na základě kombinace údajů o pobytu v různých lokalitách.

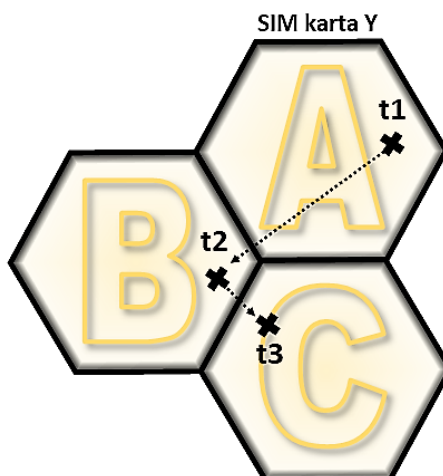
Schéma 4: Definice pobytu a pohybu

## Pobyt



ID SIM karty	Čas záznamu	ID BTS
X	t1	A
X	t2	A
X	t3	A

## Pohyb



ID SIM karty	Čas záznamu	ID BTS
Y	t1	A
Y	t2	B
Y	t3	C

Zdroj: vlastní zpracování

**Pobyt** – Z opakovaného záznamu u stejné BTS je evidentní, že SIM karta/uživatel v místě určitou dobu pobývá a lze přibližně i odvodit, jakou dobu zde pobýval/a.

Interval mezi dvěma po sobě následujícími připojeními ke stejné BTS je **časovým intervalem, po který minimálně v dané lokalitě SIM karta/uživatel pobýval/a**. Nelze však zjistit, co se mezi jednotlivými připojeními dělo<sup>10</sup>, což je jedno z hlavních omezení výpovědní hodnoty dat. Proto účel, za kterým jsou data pořizována, a jednotlivé výzkumné otázky nesmí vyžadovat časovou či územní přesnost nad rámec možností těchto dat.

**Pohyb** – Informace o pohybu není v datech přímo obsažena. Lze ji jen odvozovat z faktu, že **dvě po sobě následující připojení do sítě byla provedena různými BTS**. Z toho je zřejmé, že došlo ke změně lokality, ve které se SIM karta nachází a **muselo tedy dojít k jejímu pohybu**<sup>11</sup>. Bližší informace o pohybu v tomto směru neexistuje<sup>12</sup>.

<sup>10</sup> Není známo, jakou dobu se SIM karta v lokalitě vyskytovala ještě před prvním nebo po jejím posledním připojení k BTS. Zároveň není žádným způsobem zjistitelné, zdali mezi dvěma připojeními ke stejné BTS nedošlo k pohybu SIM karty mimo danou lokalitu (bez připojení k sousední BTS anebo pohybu v rámci jedné BTS).

<sup>11</sup> Pomineme-li fakt, že v síti může dojít ke změně BTS i bez nutnosti fyzického pohybu SIM karty jako v případě tzv. překmitávání aj. (popsáno dále v textu).

<sup>12</sup> Nelze s určitostí stanovit trasu, po které uživatel (SIM karta) cestoval, zjistíme jen přibližné místo počátku a konce cesty. Rovněž nelze s určitostí stanovit přesný čas pohybu, lze jen definovat časový interval mezi dvěma připojeními k síti, ve kterém k pohybu došlo. Nicméně i tak existují možnosti, jak informace o pohybu zpřesnit (bližší viz podkapitola 2.6).

Jak naznačuje schéma 4, fakt, že se SIM karta nachází po celou sledovanou dobu v jedné buňce, neznámá, že se nemůže fyzicky pohybovat, avšak během svého pohybu nepřekročí hranici buňky. Některé pohyby SIM karty v rámci buňky mohou být v realitě paradoxně i delší, než při pohybu mezi buňkami. Ukazuje se tak, data nejsou využitelná v přílišném územním detailu. Naopak hlavní silnou stránkou dat je jejich robustnost. Tyto přibližné informace o poloze nebo pohybu můžeme získat za ohromné množství uživatelů.

Sledovat lze tedy dva jevy – **pobyt** (stav) **či pohyb** – přičemž specifikace každého z nich může mít rozmanitou podobu. Který z jevů (pohyb či pobyt) je vhodné použít, závisí na účelu, za jakým jsou data pořizována.

ÚKOL: Definujte jevy, které chcete sledovat

- definujte, co pro účely vašeho projektu znamená pobyt
- definujte, co pro účely vašeho projektu znamená pohyb
- zvažte, zda podrobnost informace obsažené v geolokačních datech je pro účely Vašeho výzkumu dostatečná

## 2.4 Typologie jedinců dle jejich pohybu/pobytu

Jedná se o klíčový parametr geolokačních dat a návazných analýz mobility obyvatel, přičemž existuje mnoho přístupů vznikajících individuálně dle konkrétního metodického řešení úlohy/výzkumu.

Na základě prostorového chování SIM karet/uživatelů lze identifikovat prostorové vzorce každodenního chování. SIM karty/uživatele tak můžeme klasifikovat podle typu pohybu a jeho dílčích znaků (přizpůsobené konkrétnímu účelu).

Definice požadovaných typů pohybu osob a jejich parametrů musí být popsány v datové specifikaci. Nejčastěji bývají definovány následující typy osob:

### 1. **Rezident (RE): 1 man – 1 location**

Osoba pravidelně nocující v dané lokalitě. Lze ji tak považovat za bydliště. Základní parametr, který je zapotřebí určit každému jedinci, tj. každé SIM kartě.

### 2. **Dojíždějící za prací/do školy (DOP): 1 man – 1 location**

Osoby, které v dané lokalitě nebydlí, ale s denní pravidelností ji navštěvují a významnou část dne zde stráví. Lze předpokládat, že se jedná o dojížděku za prací či do školy. Místo je hlavním dojížděkovým cílem dané osoby.

#### *Podtyp - 2.1 Vyjíždějící (dynamičtí) rezidenti (DRE) za prací/do školy: 1 man – 1 location*

*Totožné s dojíždějícími za prací a do škol, avšak vztaženo k lokalitě jejich bydliště. Tato skupina osob je z pohledu zdrojové lokality podskupinou rezidentů (vyjíždějící rezidenti) z pohledu cílové*

*lokality však spadají do skupiny dojíždějících. Obdobně by šlo takové podtypy vyčlenit i v rámci dalších typů.*

### 3. Pravidelně dojíždějící za službami (DOS): 1 man – many locations

Osoby, které v dané lokalitě nebydlí, ale do daného místa dojíždí za službami, nákupy, kulturou, zdravotní péčí, sportem, návštěvy aj. Jedná se o dojížděku do jiného místa než za prací<sup>13</sup>. Každá osoba může mít více takových lokalit, a to i během jednoho dne.

*Podtyp - 3.1 Pravidelně vyjíždějící za službami*

*Totožná skupina jedinců, avšak z pohledu zdrojové lokality.*

### 4. Nepravidelně dojíždějící a náhodní návštěvníci (NEDo)

Osoby, které nejsou rezidenty, dojíždějícími za prací ani dojíždějícími za službami, nicméně se v dané lokalitě během sledovaného období vyskytnou.

### 5. Tranzitující (TT)

Osoby, pro které není lokalita cílovým místem v rámci jejich cesty. Pravidelně projíždějí, přestupují na jiný spoj, pohybující se v projíždějících dopravních prostředcích aj.

Je důležité definovat jednotlivé typy tak, aby k nim bylo možné přiřadit každou SIM kartu v síti. Jednotlivé typy představují ve své podstatě určité role. **Každý jedinec musí mít přiřazenou určitou roli**, přičemž během jednoho dne jich má zpravidla několik. A z pohledu různých lokalit může mít v jeden okamžik více rolí (např. vyjíždějící rezident = dojíždějící za prací).

ÚKOL: Definice typů pohybu

- Definujte vhodné typy pohybu pro potřeby vašeho projektu/výzkumu. Tj. proveďte kategorizaci uživatelů, dle jejich prostorového chování.
- Do datové specifikace popište charakteristiky/parametry daných typů pohybu.
- Pomocí zápisu v matici určete hodnoty, kterých by měl daný typ pohybu (daná kategorie) nabývat.

## 2.5 Územní rozsah měření

Kromě identifikace sledovaných jevů a zájmové skupiny (co měřit a koho se má měření týkat) je zapotřebí definovat rovněž územní parametry měření. Specifikace územních aspektů měření má dvě dimenze.

<sup>13</sup> Kategorii možné rozdělit dle intenzity dojížděky např. na pravidelnou a občasnou. Konkrétně lze definovat pravidelnou např. alespoň 1 návštěva týdně alespoň na 1 hodinu. Občasnou pak definovat kupříkladu alespoň 2 návštěvami s celkovým počtem 4 strávených hodin v lokalitě během jednoho měsíce (přesné definování vychází z účelu sledování).



- a) První definuje, **jaké území** má být zkoumáno. Jedná se o hranice, ve kterých má probíhat měření definovaných jevů a uživatelů (fakticky však vždy bude vymezeno pomocí BTS, které námi zvolenou oblast obsluhují).
- b) Druhá představuje **územní detail**, v jakém mají být data agregována. Z důvodu ochrany osobních údajů není možné poskytnout data o pohybu konkrétních osob. Data jsou tak agregována za zvolené územní jednotky<sup>14</sup>. Územní agregace tak stanovuje detail, za jaké územní celky mají být data dostupná, což je rovněž nazýváno jako **granularita**. Lze ji definovat prakticky libovolně, přičemž základní jednotkou je buňka BTS<sup>15</sup>. Nabízí se využívat pro agregaci územní jednotky administrativního členění státu (obce, obvody POÚ, ORP, okresy, kraje)<sup>16</sup>. Zde je však nutné zmínit nepřímou úměru mezi granularitou a validitou dat, kdy při vyšším územním detailu jsou data méně přesná (blíže v kapitole 3.1)<sup>17</sup>

Po definování zájmového území i územního detailu (jakýmkoli způsobem) pak přichází druhý krok, kdy jsou **určeny konkrétní BTS, které obsluhují definované území** (aktivita na straně operátora/poskytovatele dat). Každé vymezení je tak ve svém konečném důsledku určení konkrétních BTS, které budou do výzkumu zahrnuty.

ÚKOL: Definujte územní dimenzi měření

- definujte, v jakém území chcete měření provádět (celý stát, jeden kraj, konkrétní lokalita)
- definujte, za jaké územní jednotky chcete výstupy získat (obce, ORP, kraje či jiné)

## 2.6 Časová dimenze měření

Čas do specifikace dat vstupuje v různých podobách. Při definici požadovaných dat je zapotřebí určit:

- sledované/měřené časové období
- míru časového detailu sledování a intervaly pro agregaci záznamů
- způsob, jak zaznamenávat pohyb/pobyt v čase.

<sup>14</sup> Způsob, jakým k agregaci dojde je opět odvislý od účelu, za kterým jsou data pořizována. Sloučení dat do určitých územních celků je jen jedním z možných přístupů. Mimo to mohou být data agregována podle aktivity v určitém čase nebo dle specifického typu chování.

<sup>15</sup> Menší jednotku než příslušnost k buňce BTS nelze z dat získat.

<sup>16</sup> Kromě administrativního členění může být použito i jiné územní kritérium pro agregaci. Například bude-li účelem pořizování dat protipovodňová ochrana, může být agregace dat prováděna podle povodí jednotlivých toků. V tomto smyslu lze spatřovat velkou variabilitu využití získávaných dat

<sup>17</sup> Nejmenší územní jednotky administrativního členění (ZSJ, části obcí či obce) tak mohou být zatíženy velkou chybou, která se však při agregaci do vyšších jednotek snižuje.

### Sledované (měřené) časové období

Určuje časové období, kdy bude měření prováděno. Na rozdíl od výše uvedených aspektů se nezaobírá podstatou a způsobem měření. Zadává pouze rámec, kdy by k měření mělo dojít. Může se jednat o jeden den, vybrané dny, týden či několik po sobě následujících týdnů, vybrané dny/týdny v každém čtvrtletí roku apod.

Pokud jsou studie zaměřeny na mapování všeobecných prostorových vztahů, měla by být snaha měření **oprostit o sezónní a nahodilé jevy**, které by mohly zkreslit výsledky. Takovými jevy jsou například období školních prázdnin, státních svátků, lokální jednorázové akce s vysokým počtem účastníků a další<sup>18</sup>. Při konvenčních výzkumech mobility obyvatel je za typický den považována tzv. „běžná středa“ (oproštěná od sezónních vlivů). A to zejména z toho důvodu, že středy bývají nejméně ovlivněny specifickými událostmi v jízdních řádech. Nicméně geolokační data jsou natolik komplexní a zahrnují tak velké množství dat, že **je velmi složité určit některý den jako typický**. Avšak snaha vyhnout se zkreslení dat v důsledku očividně nestandardního chování velké části populace je i tak jednoznačně žádoucí.

### Časové intervaly

Údaje v síti nelze získat za konkrétní moment, avšak vždy jsou k dispozici za určitý (zvolený) interval. Data tak nereprezentují stav v daný moment, ale jsou agregována v určitých intervalech. V tomto směru je zapotřebí definovat dva parametry:

1. míra časového detailu, v jakém pohyby/pobyty osob sledovat a analyzovat.
2. intervaly, na které bude výsledná databáze strukturovaná.

Ad 1) detail měření – k periodickému updatu, kdy se SIM karta opakovaně hlásí do sítě mobilních operátorů, aniž by docházelo k její aktivitě (typu volání, SMS, datové přenosy), dochází v rámci sítě nejednotně<sup>19</sup>. Z toho důvodu nemůžeme v konkrétním čase říci, kolik je v síti uživatelů, lze jen **zvolit interval, za který údaj chceme získat**. Tento interval lze zvolit prakticky libovolně. Avšak v případě kratšího intervalu, než je např. 10 minut, narážíme na problém velkého množství SIM karet, které se v tomto intervalu nepřihlásí k síti ani jednou<sup>20</sup>. Je na místě zvážit, jak velká podrobnost je pro danou studii zapotřebí<sup>21</sup>.

<sup>18</sup> Je zapotřebí zohlednit, zda ve výzkumu bude, či nebude mít vliv pohyb vysokoškolských studentů, který je uskutečňován jen v určitých měsících v roce. Kromě těchto aspektů je rovněž nutné zvážit periodicky se opakující změny dopravního chování typické zejména pro víkendové dny a zcela specifický případ pátečních pohybů, které se zpravidla vymykají ostatním všedním dnům. Rovněž pondělní chování bývá často zkresleno, jelikož mnohdy se lidé z víkendových aktivit vrací až v pondělí.

<sup>19</sup> Intervaly jedné samostatné SIM karty jsou přibližně pravidelné a jsou odvislé od úrovně sítě, ke které se SIM karta připojuje (2G, 3G, 4G, popř. 5G). Při sledování většího počtu SIM karet tak lze vnímat intervaly, kdy k updatu dochází, jako nepravidelné či chaotické, avšak při sledování jedné SIM karty lze pravidelnosti nalézt.

<sup>20</sup> Tento problém lze ovšem vyřešit dále v textu popsáním příkladem, kdy SIM karta je počítána do té buňky, kde se k síti přihlásila naposledy až do momentu, kdy se do sítě opět přihlásí.

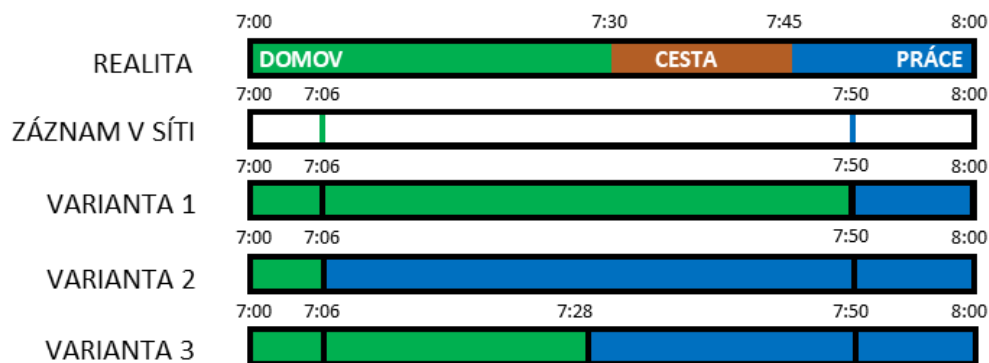
<sup>21</sup> Vzhledem k vypovídající hodnotě lze pro obecné studie pohybu obyvatel (nikoli detailní studie) doporučit sledované intervaly 30 minut, popřípadě delší.

Ad 2) intervaly pro agregaci dat (struktura databáze) musí být zvoleny tak, aby umožňovaly aplikaci plánovaných analytických postupů. Může se jednat např. o **časové sloty definující specifické části dne** či sloty zohledňující specifické chování v rámci týdne, měsíce, roku či individuální jednorázové akce. Definujeme tak intervaly, do kterých mají být výsledná data agregována a jak bude výsledná databáze strukturována<sup>22</sup>.

### Záznam pobytu/pohybu v čase

Způsob definování časové dimenze měřených jevů je stěžejní pro pochopení, co se za výslednými daty skrývá a jak je lze interpretovat. Schéma 5 ukazuje příklad jedné reálné situace, která je určitým způsobem zaznamenána v síti a v **důsledku různých přístupů může být výsledný pohyb v datech obsažen rozličně**.

Schéma 5 – Přístupy k měření polohy v časových intervalech mezi dvěma záznamy



Zdroj: vlastní zpracování

- Jedním z přístupů, jak definovat pobyt/pohyb v intervalech mezi záznamy, je **zachování příslušnosti k buňce do momentu, kdy nedojde k přihlášení k jiné BTS**. Takový přístup je vhodný pro případy, kdy je ve výzkumu kladen důraz na kontinuitu dat, a jsou orientované na mapování chování velkého množství jedinců z velkého území. Zaměřují se zejména na místo zdroje a cíle pohybu, nikoli na samotnou cestu. Tímto přístupem dochází skokově ke změně lokality buď na počátku, nebo častěji na konci intervalu (viz schéma 5 varianta 1 a 2).
- Mohou však existovat i studie, které jsou zaměřeny právě na identifikaci cestovních tras. Zde může být moment, kdy dojde ke změně lokality zásadní. **Stanovení momentu přesunu do středového intervalu** (varianta 3) tak zajistí vyšší pravděpodobnost přiblížení se reálnému času, kdy k přesunu došlo<sup>23</sup> (viz schéma 5).
- Popřípadě mohou být zvoleny jiné varianty, jejichž aplikace je již složitější. Moment, kdy se v takovém případě změní příslušnost k jedné či druhé buňce, pak může nastat v závislosti na

<sup>22</sup> Detail měření (ad 1) a časové úseky pro agregaci (ad 2) se mohou rovnat a většinou se tak i děje. Úseky pro agregaci nemohou být kratší než detail měření, který udává základní jednotku, za kterou jsou data v dané úloze/výzkumu analyzována.

<sup>23</sup> Nicméně ze schématu 7 je zřejmé, že ani tento moment nespadá do intervalu, kdy k cestě skutečně došlo.

zvolených parametrech prakticky kdykoli ve vymezeném intervalu<sup>24</sup>. Jedná se například o metody:

- zohledňující velikost jednotlivých buněk BTS.
- zohledňující čas, strávený uživatelem ve zdrojové buňce a v cílové buňce před a po uskutečnění pohybu<sup>25</sup>.
- prostřednictvím nástrojů síťové analýzy umožnit pohyb mezi buňkami jen po dopravní síti<sup>26</sup>.

#### Praktický příklad:

Studie monitorující počet obyvatel krajského města sledovala všechny SIM karty, které se během týdne objevily ve městě za účelem odhadu počtu stálých obyvatel a dojíždějících. Celý týden byl rozdělen do 30minutových intervalů (*detail měření*). Každé SIM kartě byla za každý interval přiřazena poloha, kde se vyskytovala převážnou část intervalu. Další datové operace již byly prováděny za tyto časové úseky. Na základě počtu nocujících osob v jednotlivých dnech – počet unikátních SIM karet v obci mezi 0:00 – 4:00 (*interval pro agregaci dat*) - byl následně určen předpokládaný počet obyvatel. Tento údaj byl pak porovnáván s údaji za další denní intervaly – ranní dopravní špička, pracovní doba, odpolední špička, víkendy aj. (*časové sloty pro agregaci dat*).

#### ÚKOL: Definujte časovou dimenzi měření

- zvolte metodu, jak technicky zaznamenávat pohyb v čase (podstata pohybu a jak dělit přechodový interval)
- zvolte míru časového detailu (nejnižší měřená časová jednotka)
- zvolte časové intervaly/sloty, do kterých budou časové jednotky agregovány (části dne, dny, všední dny vs. víkend)
- zvolte časový úsek měření (např. vybraný den, 4 po sobě následující týdny, jeden týden v každém čtvrtletí, každou první středu v měsíci aj.)
- za jaké územní jednotky chcete výstupy získat (obce, ORP, kraje či jiné)

Konkrétní parametry jsou odvislé od účelu pořizování dat.

<sup>24</sup> Způsob je vhodný pro menší vzorky (územně i populačně) a je zapotřebí z nich odfiltrovat případy, kdy kupříkladu uživatel vypne mobilní telefon na několik hodin (např. přes noc) a poté se připojí již na jiném místě.

<sup>25</sup> Přechodový interval by byl rozdělen mezi buňky A a B v závislosti na poměru stráveného času v buňce A před posledním přihlášením a stráveného času v buňce B po prvním přihlášení. Tato metoda umožní eliminovat vliv buněk, kterými na trase uživatel pouze projíždí a větší význam přikládá místům, kde se jedinec zdržuje delší dobu.

<sup>26</sup> V případě velmi úzce zaměřených studií, monitorující například pohyb po jedné komunikaci (dálnice, železniční trať aj.), lze data efektivně kombinovat s reálnými údaji o předpokládané době jízdy či s informacemi z jízdních řádů a získat bližší odhad, v jaké době k pohybu po dané trase došlo

## 2.7 Ochrana osobních a lokalizačních údajů

Přestože geolokační data ze signalizační sítě mobilních operátorů sledují primárně polohu SIM karet nikoli samotného uživatele, z podstaty věci tento údaj o jeho poloze vypovídá. Tyto údaje nejsou považovány za osobní údaje, jak by se mohlo zdát, nýbrž za **provozní a lokalizační údaje definované zákonem o elektronických komunikacích č. 127/2005 Sb., § 90 a § 91**. Tento zákon nejen provozní a lokalizační údaje definuje, avšak upravuje i nakládání s nimi.

Provozními údaji se rozumí jakékoli údaje zpracovávané pro potřeby přenosu zprávy. Lokalizačními údaji jsou údaje, které určují zeměpisnou polohu koncového uživatele (dle zmíněného zákona). Ke zpracování a nakládání s těmito údaji musí mít provozovatel sítě informovaný souhlas uživatele obdobně, jako je tomu v případě zpracování osobních údajů, přičemž musí být specifikován účel, ke kterému budou tyto údaje využity. Tento souhlas je ze strany uživatele odvolatelný.

Zároveň zákon stanovuje, že provozovatel sítě, který zpracovává a ukládá provozní údaje, včetně příslušných lokalizačních údajů, je **musí smazat nebo učinit anonymními, jakmile již nejsou potřebné pro přenos zprávy**, tj. jakmile již nejsou nezbytné pro samotné fungování sítě. Anonymizace dat je upravována zákonem o ochraně osobních údajů č. 101/2000 Sb., § 4 a §5. Anonymním údajem je chápán takový údaj, který buď v původním tvaru, nebo o jeho modifikaci nelze vztáhnout k určitému nebo určitému subjektu údajů. **Údaje tak musí být uchovávány, analyzovány či dále poskytovány v takovém formátu, kdy nelze konkrétní údaj ztotožnit s konkrétním uživatelem.**

Z výše uvedeného vyplývá, že otázka **ochrany osobních, provozních a lokalizačních údajů je primárně na straně provozovatele sítě popřípadě primárního zpracovatele (poskytovatele) těchto dat**. Ti jsou povinni údaje anonymizovat ještě před jejich prvotním zpracováním. Pořizovatel dat tak k těmto primárním neanonymizovaným údajům nemá a ani nemůže mít přístup, disponovat jimi či do nich nahlížet. Z podstaty výše zmíněného i sám provozovatel sítě s těmito údaji může disponovat až po jejich anonymizaci.

Termín lokalizační údaje, který je v textu tohoto dokumentu hojně užíván, není dle zákonné definice zcela přesným. Na základě výše zmíněného jsou data, která pořizovatel poptává, pouze jakýmsi statistickými údaji vycházejícími z dříve anonymizovaných lokalizačních dat. Nicméně pro terminologické zjednodušení je výraz „lokalizační data“ užíván i pro tyto statistické údaje od nich odvozené.

## 3 Problematické oblasti definice dat

### 3.1 Problémy spojené s územní podrobností měření

Celé úskalí definice granularity požadovaných dat tkví ve faktu, že síť BTS, tj. rozmístění jednotlivých vysílačů a jejich servisní buňky, nerespektuje žádné administrativní či fyzicko-geografické hranice.

#### Model servisních polygonů BTS

Kolem každé BTS existuje určitá oblast, která je obsluhována právě touto stanicí. Nicméně tyto oblasti tvoří navzájem se prolínající polygony a ani sami operátoři nejsou schopni zcela přesně identifikovat servisní oblast konkrétních BTS. Přestože jsou schopni vygenerovat model zobrazující servisní polygony každé BTS, jedná se vždy jen o model s větší či menší přesností. Jak již bylo představeno v podkapitole 1.3, věnované samotnému fungování sítě, může dojít k situaci, kdy je jedinec uprostřed konkrétního polygonu, a přesto se jeho SIM karta připojí k sousední BTS např. z důvodu přetížení konkrétní stanice či z důvodu změny komunikační služby. V tomto směru je zapotřebí po primárním zpracovateli záznamů (operátor) požadovat věrohodnou specifikaci postupu, jak byly polygony BTS stanoveny, přičemž nejjednodušším ale zároveň nejméně přesným přístupem je konstrukce Voronoiových polygonů, popsána v kapitole 1.3. Nicméně existují možnosti (a v praxi se běžně využívají) sestavit přesnější modely, které zohledňují například orientaci vysílače, předpokládanou sílu signálu, terén, vegetační pokryv či obecně land use popřípadě mohou využívat GPS a jiné lokalizační systémy pro zpřesnění odhadů dosahu jednotlivých vysílačů.

#### Agregace dat do územních jednotek

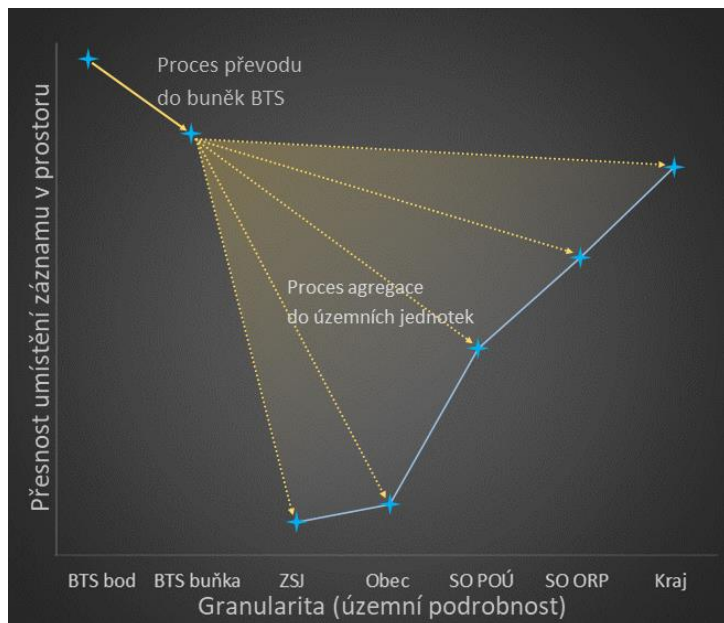
Jelikož síť buněk BTS je specifická a neshoduje se s žádnými územními jednotkami, je nezbytné agregovat data získaná za buňky BTS například do hranic správních obvodů obcí či jiných jednotek. Jak již bylo zmíněno, čím větší granularita, tím k větším nepřesnostem agregací dochází (viz schéma 6).

S určitostí dokážeme identifikovat, ke které BTS se SIM karta hlásí. Přesně dokážeme i určit příslušnost k servisnímu polygonu BTS, avšak ten sám už je vymezen s určitou nepřesností.<sup>27</sup> Při agregaci do územních jednotek je přesnost dat přímo úměrná územní podrobnosti (granularitě). **Čím větší jednotky zvolíme, tím budou výstupy přesnější<sup>28</sup>.**

<sup>27</sup> Pohraniční oblasti správních obvodů jsou méně osídlené, proto je ve výsledné databázi menší počet SIM karet, které je složité do jednotek umístit (nejméně jich je u největších jednotek).

<sup>28</sup> S rostoucí řádovostní úrovní je menší podíl osob žijících při hranicích obvodů vůči obyvatelům celého regionu, chybovost v umístění mezi jednotky proto neovlivňuje celkové výstupy tak zásadně (s rostoucí velikostí obvodů klesá podíl obyvatel hraničních oblastí).

Schéma 6: Přesnost dat při převodu záznamů na územní jednotky (závislost na granularitě)



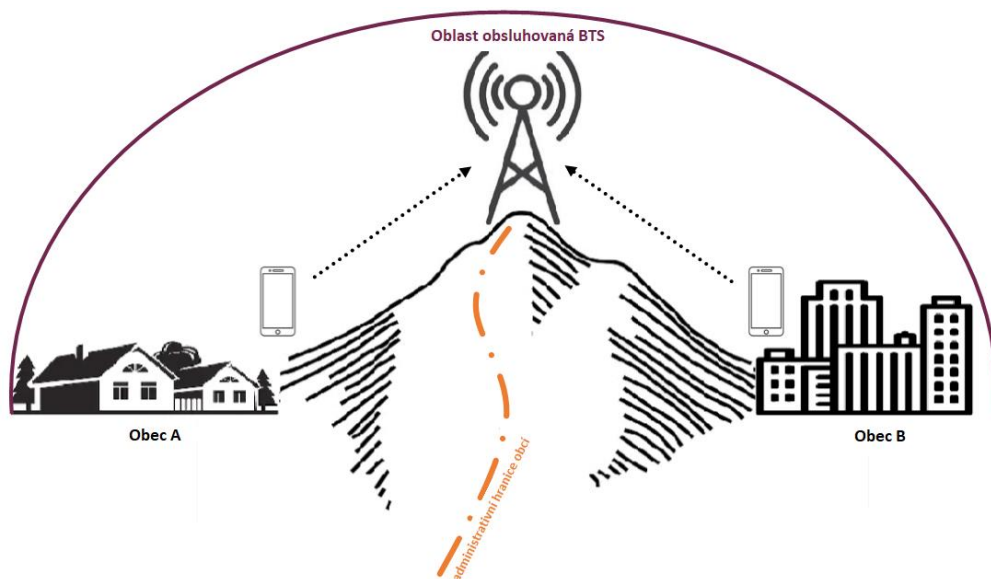
Zdroj: vlastní zpracování

Schéma 6 ukazuje dva procesy, které primární zpracovatel dat provádí, aby ze záznamu v síti získal prostorovou informaci: **převod bodu BTS na plochy buněk BTS, a agregaci/převod dat do požadovaných územních jednotek.** Každou z těchto úprav dochází ke snížení přesnosti a validity dat. Z toho důvodu je zapotřebí zvážit podobu výsledné databáze. V případě, že charakter výzkumu umožňuje pracovat s bodovými znaky (měřené hodnoty vztahovat k poloze BTS), je vhodné tímto způsobem postupovat. Pokud lze výzkum efektivně realizovat s daty za

jednotlivé buňky BTS, dochází k menším nepřesnostem vzhledem k nejednoznačnosti hranic těchto jednotek. V případě nutnosti převodu na administrativní či jakékoli jiné územní jednotky je nutné zvážit územní detail. Vždy se bude jednat o kompromis mezi přesností dat a jejich detailem.

Hlavním problémem převodu na velmi malé územní jednotky (ZSJ, obce) je skutečnost, že **jedna BTS může obsluhovat dvě a více samostatných obcí**, k čemuž na území Česka dochází často. V takových případech není možné určit, ve které obci se uživatel v daný moment nachází. Tento problém nesouladu hranic BTS a obcí je naznačen na schématu 7.

Schéma 7: Rozdělení buňky BTS mezi dvě obce



Zdroj: vlastní zpracování



Praktický příklad:

Schéma 7 zobrazuje případ, kdy jedna BTS obsluhuje dvě obce rozdílné velikosti. U obou obcí se jedná o celé jádrové části intravilánu obcí. V servisní oblasti této konkrétní BTS bylo v měřeném intervalu identifikováno 10 tis. SIM karet/uživatelů. Jakým způsobem budou zjištění uživatelé rozdělení mezi tyto dvě obce?

Primární zpracovatelé/poskytovatelé dat (povětšinou mobilní operátoři) mají na zajištění souladu polygonů BTS a hranic obcí určité metodické nástroje, které jsou součástí jejich jedinečného know-how. Nicméně i tak musí pořizovateli dat poskytnout jednoznačné a dostatečné vysvětlení, jak byl územní soulad jednotek zajištěn. Pořizovatel by měl na tomto vysvětlení trvat a požadovat jej.

Jak již bylo řečeno, nejjednodušší je (pokud to charakter výzkumu umožňuje) analyzovat data za buňky BTS či dokonce vztáhnout údaje k bodu, kde je lokalizována BTS, většinou to však není možné.

Možné řešení:

- Nejjednodušším a zároveň nejméně přesným řešením problému naznačeného schématem 7 je **rozdělení identifikovaných uživatelů mezi územní jednotky v poměru dle jejich počtu obyvatel** (určeného z konvenčních datových zdrojů). Avšak touto metodou lze dojít i ke zcela iracionálním výsledkům. V oblastech, kde na území ZSJ či obce zasahuje větší množství buněk BTS, které mohou zasahovat i do jiných územních jednotek, může být počet uživatelů v síti významně zkreslen oproti realitě.
- Další metodou je v těchto oblastech postupovat v menším územním detailu. Znamenalo by to **některé ZSJ či obce agregovat do větších celků a ty pak analyzovat společně**. V našem případě by tak obce A a B vystupovaly v daném datasetu jako jedna územní jednotka.
- Pokud je však zapotřebí, aby každá územní jednotka byla v databázi uvedena samostatně, je zapotřebí využívat sofistikovanějšího přístupu. Ten spočívá kupříkladu ve **využití hodnocení land use** (využití území/krajiny). Konstrukce detailního modelu využití krajiny nám umožní přesněji odlišit obydlené a neobydlené oblasti, popřípadě území s dalšími funkcemi. Ideální je tvorba *gridu* (územní mřížka) ve vysokém územním detailu. Každý čtverec takové mřížky má přidělený určitý pravděpodobný podíl SIM karet příslušné buňky, a to na základě charakteru zástavby, popřípadě adresných míst v daném čtverci <sup>29</sup>.

<sup>29</sup> Přičemž přesnost rozdělení je odvislá od přesnosti konstruovaného modelu.

Tím dojde ke zpřesnění rozdělování hodnot mezi jednotlivé obce. Nicméně vždy je zapotřebí mít na paměti, že se jedná o odhady výskytu osob uvnitř jedné buňky, tzn. pod základní jednotkou (rozlišovací úrovní) těchto dat. Vždy se tedy bude jednat pouze o odhad.

### Překmitávání

V místech překryvů buněk BTS dochází k tzv. překmitávání (přepínání) mezi BTS. Jedná se o situaci, kdy konkrétní SIM karta, aniž by sama byla v pohybu, z jistého důvodu změní při periodickém updatu nebo i při své aktivitě BTS, ke které se hlásí. Důvody ke změně mohou být různé, např. přetížení jedné BTS či vyhodnocení aktuálního signálu sousední BTS jako silnější apod. Může tak dojít k jevu, kdy se **uživatel ani jeho SIM karta fyzicky nepohybují, avšak v záznamech může být jejich aktivita detekována jako pohyb**. Paradoxně tak uživatel např. během noci, kterou stráví spánkem, může zaznamenat v důsledku překmitávání několikanásobné pohyby v síti. Tyto nežádoucí jevy je zapotřebí v záznamech identifikovat a zajistit, aby ve svém důsledku neovlivnily validitu výsledné databáze. To je úkolem poskytovatele dat/operátora.

Zajištění, aby technické aspekty typu překmitávání neovlivnily požadovanou databázi, je na straně poskytovatele dat, který by měl mít k tomuto účelu vyvinuté vhodné nástroje. I jejich fungování musí pořizovateli dat náležitě vysvětlit a prokázat jejich funkčnost.

ÚKOL: Definice aspektů, které mohou ovlivnit validitu dat

- Definujte konkrétní aspekty, které jsou důležité pro validitu dat
- formulujte jednotlivé body jako otázky, které je vhodné diskutovat přímo s poskytovatelem dat
- zahrňte do požadavků na data povinnost pro poskytovatele věrohodně vysvětlit veškeré datové operace, které s primárními záznamy učinil:  
tvorba modelu buněk BTS, převod buněk BTS na požadované územní jednotky (agregace dat), překmity a jiné rušivé jevy v síti.

## 3.2 Problémy vycházející ze samotného fungování sítě

Síť mobilních operátorů byla konstruována za jinými účely, než získávání dat o pohybu/pobytu jejich uživatelů. Z toho důvodu jsou ve zdrojových datech obsaženy mnohé nadbytečné údaje/záznamy či dochází k jevům, které ovlivňují validitu výsledných dat.

Nadbytečné údaje v síti - Prvním problémem, který musí primární zpracovatel dat řešit, je **redukce celkového počtu SIM karet**<sup>30</sup> s cílem získat danou informaci jen za ty SIM karty, jejichž prostorové chování lze ztotožnit s chováním jejich uživatele.

- Mezi nadbytečné lze v tomto směru považovat např. tzv. **technické SIM karty**<sup>31</sup>. Společným aspektem těchto SIM karet je právě jejich neztotožnitelnost s pohybem jejich uživatele. Lze předpokládat, že uživatel tato zařízení nemá po většinu dne při sobě, jako je tomu u mobilního telefonu. Navíc, uživatel k těmto technickým zařízením povětšinou ještě vlastní mobilní telefon se SIM kartou, jejíž pohyb se v záznamech již beztak projevuje. Primární zpracovatel (operátor) by měl disponovat nástroji, jak **tyto SIM karty identifikovat a z cílového souboru je vyloučit**, aby nezkrusovaly výslednou databázi. Síť mobilních operátorů by měla být schopna rozpoznat v jakém typu zařízení či dokonce i modelu se SIM karta nachází. Tyto SIM karty jsou často již prodávány jako technické a mají zvláštní podmínky například v rámci účtování za služby aj. Lze je rovněž identifikovat na základě faktu, že tato zařízení nejsou schopna volat ani přijímat hovory či jen přijímat SMS v určitém formátu. Technické SIM karty tak většinou nevykazují žádné známky aktivního užívání, jsou statické a nevykazují žádný pohyb (vyjma těch v automobilech).
- Obdobný problém v síti způsobují telefony s **duálními SIM kartami** či jen fakt, že uživatel s sebou nosí více aktivních telefonů/zařízení (osobní, pracovní)<sup>32</sup>. I s tímto problémem by měl primární zpracovatel dat (operátor) počítat a pokusit se eliminovat možnost zkruslení výsledných dat. Identifikovat tyto SIM karty je možné například tak, že v dlouhodobém měření **kopírují vzájemně svůj pohyb**.

Počet uživatelů v síti - Dalším technickým problémem je samotné fungování sítě, kdy k periodickému updatu jednotlivých zařízení dochází v různém čase. Nelze tak **zjistit, kolik je v konkrétním okamžiku v síti aktivních SIM karet/uživatelů**. Za použití intervalového měření získáme z jednotlivých měřených časových úseků rozdílné součty celkových uživatelů v síti. Jen z části jsou tyto výkyvy důsledkem technického fungování sítě. Počet uživatelů v síti mobilních operátorů během dne totiž kolísá i přirozeně. Zejména v nočních hodinách je významný pokles celkového počtu SIM karet v síti v důsledku vypnutí mobilních zařízení. Je tak v tomto směru obtížné data srovnávat za jednotlivá časová období nebo vůči jiným datovým zdrojům (konvenčním). Zmíněným výkyvům v datech se nelze nijak vyhnout, nicméně je potřeba s nimi při plánování analýz počítat<sup>33</sup>.

<sup>30</sup> Redukce je nutná již z toho důvodu, že dostupné údaje hovoří o přibližně 14,7 mil. aktivních SIM kartách na území ČR. Počet obyvatel ČR je přibližně 10,5 mil. přičemž v populaci jsou i skupiny osob, které mobilní telefon nevládní (malé děti, část seniorů atd.).

<sup>31</sup> Jedná se o SIM karty, které nejsou v mobilních telefonech, ale v určitých technických zařízeních (většinou "Smart" řešení). Za pomoci GSM ovladačů (obsahujících SIM kartu) tak lze například ovládat otevírání bran a vrat, alarm nebo vytápění domácnosti či závlahu zahrady. Mezi technické SIM karty můžeme řadit i karty vkládané do automobilů.

<sup>32</sup> Samotný problém s dualitou dat nastává pouze v případech, kdy jedinec vlastní dvě SIM karty stejného operátora.

<sup>33</sup> Například při nočním poklesu uživatelů sítě by mělo být u každé neaktivní SIM karty předpokládáno, že se nemění místo jejího posledního výskytu až do chvíle, kdy se ráno opět připojí do sítě (což s vysokou pravděpodobností bude na stejném místě).

Je na tomto místě zapotřebí zopakovat, že **veškeré operace, které primární zpracovatel se zdrojovými daty učiní, musí být jasně popsány**. Musí z nich být patrné, jak je zajištěna validita dat pro konkrétní účel zkoumání.

ÚKOL: Požadujte „metodický postup řešení“

Ve svých požadavcích na data vedle datové specifikace požadujte po poskytovateli dat (operátor) i souhrnnou metodiku. Ta by měla obsahovat:

- Popis dodávaných dat a metodické poznámky k jejich interpretaci.
- Metodický postup, jak byla data pořizována.
- Věrohodné vysvětlení všech operací, které s daty prováděl.
- Způsob, jakým se vypořádal s konkrétními problémy: agregace dat (časová, územní), model buněk BTS a jeho převod na územní jednotky, technické SIM karty, duální SIM karty, překmitávání a jiné rušivé jevy v síti apod.

Vysvětlení těchto bodů je možné požadovat buď při předávání dat, nebo je možné požadovat, aby metodické kroky definoval zpracovatel zakázky již při podávání své nabídky v rámci veřejné zakázky (vhodnější varianta). Metodické kroky se tak stanou předmětem posouzení kvality nabídky.

### 3.3 Specifické možnosti dodatečných informací ze sítě

Nelze získat informaci, která v datech obsažena není či informaci, která v datech sice je, ale je nemožné ji odlišit od ostatních nesouvisejících údajů. Jak již bylo řečeno, záznam v signalizační síti sám o sobě má zcela jednoduchou strukturu určující daného uživatele, čas a lokalitu.

Existuje však velké množství způsobů **dodatečných informací, které jsou schopny zvýšit vypovídací schopnost dat**. Vyžadují však buďto užití speciálních technických nástrojů nebo informaci dodávají jen za část sítě/uživatelů, popřípadě je informace nevěrohodná a zavádějící.

**Informace o elektronickém zařízení** – jednou z dodatečných informací, které lze k záznamu zjistit je **síla signálu během záznamu**. Síť mobilních operátorů by měla z technického hlediska umožňovat tuto informaci získat. Je ovšem nutné s konkrétním operátorem ověřit, zda takovou informaci v síti opravdu zaznamenává. Jak již bylo zmíněno, tato informace může být cenná při **zpřesňování polohy uživatele/SIM karty v rámci buňky BTS**.

Další možností, která již byla zmíněna v předešlých kapitolách, je získání **informace o zařízení, ve kterém se SIM karta nachází**. Přestože síť z technického hlediska umožňuje informaci zjistit, opět je nutné ověřit, zda konkrétní operátor tento údaj opravdu zaznamenává. Kromě již zmíněného využití<sup>34</sup> může znalost

<sup>34</sup> Identifikace SIM karet v technických zařízeních, nebo případy dvou SIM karet v jednom zařízení. To může napomoci zvýšení validity výsledného datového souboru, který nebude obsahovat nadbytečné a duplicitní informace.

zařízení včetně např. modelu, ve kterém je SIM karta vložena, dále indikovat i informace o samotném uživateli. Lze předpokládat, že uživatelé z vyšších příjmových skupin budou mít novější a dražší modely mobilních zařízení. Tudíž je pak z dat o pohybu uživatelů možné vyčíst určité rozdíly v prostorových vzorcích takových skupin.

**Informace o uživateli** – zjištění bližších informací o uživateli má velký potenciál pro rozšíření možností využití dat do různých sfér zkoumání. Těmito údaji jsou **pohlaví, věk, místo bydliště, průměrná měsíční útrata za služby** mobilních operátorů a spousta dalších informací. Tyto informace jsou však jen velmi těžko zobecnitelné na celkovou populaci, jelikož soubor uživatelů, ke kterým tyto údaje máme, je relativně malý a zároveň je zcela nereprezentativní (povětšinou budou nadhodnoceny skupiny uživatelů v ekonomicky aktivním věku)<sup>35</sup>. Informace o uživateli je možné využít zejména pro studie, které si nekladou za cíl zobecňovat výstupy, ale jde jim o **parciální zjištění určité informace za určité skupiny uživatelů**.

ÚKOL: Specifikace dodatečných informací

Zamyslete se, jaké dodatečné údaje o uživateli či specifikaci jejich mobilního zařízení jsou vzhledem k účelu Vašeho výzkumu potřebné. Mějte při úvahách na paměti sníženou vypovídací schopnost některých informací. Nespolehejte pouze na tvrzení poskytovatele, zjistěte o požadovaných údajích a jejich přesnosti a možnosti zobecnění bližší informace.

### 3.4 Možnosti zjišťování tras pohybu

Není možné zcela jasně určit, kde a v jakém konkrétním čase se SIM karta mezi dvěma záznamy k různým BTS pohybovala. **Evidentní je pouze to, že k pohybu došlo. Známe zdrojovou a cílovou BTS a interval, ve kterém byl pohyb uskutečněn.** I tak ovšem existují nástroje, jak tyto informace zjišťovat a tyto odhady co možná nejvíce zpřesňovat.

Informace je využitelná zejména při lokálních studiích či detailních studiích zaměřených na vytíženost konkrétních dopravních komunikací. V tomto velkém územním detailu je možné dojít ke věrohodným odhadům.

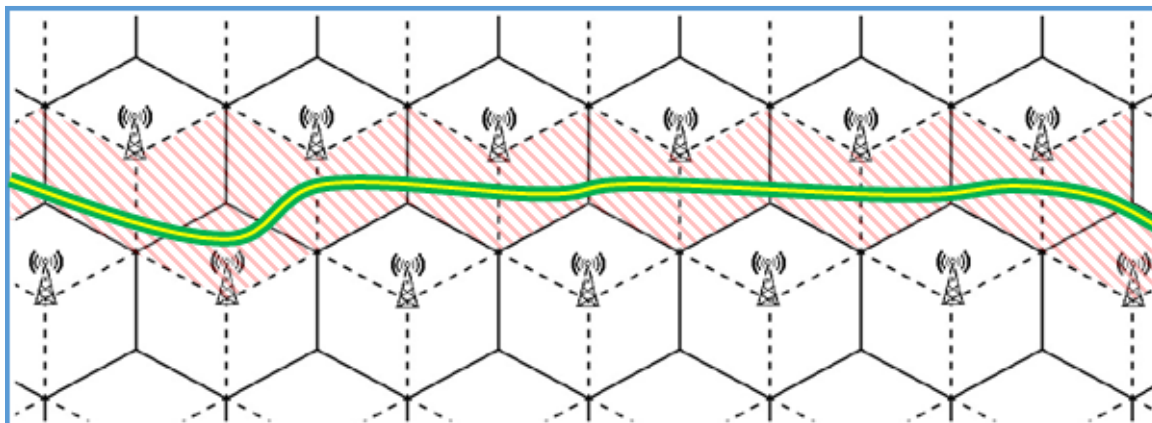
Řešení:

- Chceme-li zjistit objem cestujících na jedné konkrétní dopravní komunikaci, nejprve je nutné **identifikovat BTS podél trasy**, k nimž se při pohybu po konkrétní cestě může SIM karta hlásit. Pokud je pak v záznamech obsažena SIM karta, která se k těmto BTS v daném pořadí hlásí, lze odvodit, že jedinec se po dané trase opravdu pohyboval. Na své trase se SIM karta dozajista

<sup>35</sup> Velká část smluv zahrnuje více SIM karet (rodinné sdílení, firmy aj.). Nelze tak chování uživatele zapsaného ve smlouvě jednoznačně spojit s konkrétní SIM kartou. Pokud to možné je, i tak je obtížné jeho chování zobecňovat na celou populaci. Na příkladu rodinných smluv je zřejmé, že ve smlouvě bude uveden jeden z rodičů. Jeho prostorové vzorce chování jsou však specifické a v rámci rodiny je nelze nazvat jako reprezentativní.

nepřihlásí ke všem BTS, avšak jejich **pořadí a časové intervaly ve kterých se k BTS hlásí, indikují pohyb po dané trase** (viz schéma 8). V případě železnice lze za jistých okolností teoreticky odvodit i spoj, kterým jedinec cestoval (na základě času kdy se hlásí ke konkrétním BTS). V případě silniční dopravy je situace poněkud složitější<sup>36</sup>. Dopravní prostředek tak lze jen s určitou přesností odhadovat, což je další omezení výpovědní hodnoty dat, se kterým je potřeba počítat a zvážit jeho důsledky.

Schéma 8: Buňky BTS na trase dopravní cesty



Zdroj: vlastní zpracování

<sup>36</sup> Zde je obtížné jasně oddělit uživatele cestující hromadnou a individuální dopravou. V tomto případě může napomoci fakt, že v autobuse jede ve stejném čase velké množství osob, nicméně SIM karty cestujících mají periodický update v různých časech, a tudíž i k různým BTS. Pouze při přesunu mezi LA (location area) dochází k updatu všech SIM karet v podobných časech. I tak je ovšem obtížné určit, zda jedinec sedí v prostředku hromadné dopravy nebo cestuje osobním vozidlem jedoucím po stejné trase v přibližně stejný čas.

## 4 Příklad konkrétního nastavení specifikace dat

Následující kapitola popisuje konkrétní případ nastavení parametrů pro specifikaci požadovaných dat pro potřeby tvůrců tohoto dokumentu. Níže představená specifikace dat je připravena pro potřeby projektu s názvem „Zlepšení podmínek pro decentralizaci a dostupnost veřejné správy v území“.

### 4.1 Účel pořízení dat

Cílem projektu je na základě databáze údajů ze signalizační sítě mobilních operátorů popsat prostorové chování populace a z něj pak odvodit návrhy na optimální rozmístění služeb veřejné správy v území, které v sobě bude odrážet přirozené koncentrační procesy každodenní mobility občanů. Za tímto účelem je vypsáno výběrové řízení na Zpracovatele<sup>37</sup> nadlimitní veřejné zakázky.

Účelem této veřejné zakázky je výběr dodavatele pro nákup lokalizačních dat mobilních operátorů k realizaci výše zmíněného projektu.

Cílem realizace veřejné zakázky je získání databáze statistických údajů ze signalizační sítě o meziobecní dojížděcí. Na základě získaných dat Ministerstvo vnitra zpracuje analýzy prostorových vztahů v rámci denní mobility občanů Česka. Ty poslouží jako hlavní vstupy pro socioekonomickou regionalizaci, na základě jejichž výstupů budou formulována konkrétní doporučení pro organizaci veřejné správy v Česku.

### 4.2 Předmět, způsob a forma plnění

Předmětem plnění je **dodávka databáze statistických údajů**, které vycházejí z údajů ze signalizační sítě mobilních operátorů. Tato data jsou zpracována dle požadavků Zadavatele veřejné zakázky a reprezentují každodenní dojížděcí chování obyvatel jednotlivých obcí. Konkrétní požadavky na dodávanou databázi jsou specifikovány v příloze zadávací dokumentace veřejné zakázky<sup>38</sup>.

Plnění veřejné zakázky spočívá ve **dvou fázích dodávky požadovaných dat**:

- 1) **Testovací databáze za pilotní území** v rámci tzv. *Proof of Concept*
- 2) **Finální databáze za celé území Česka**

#### ad1) Testovací databáze za pilotní území

V první fázi dodá Zpracovatel požadovanou databázi pro území jednoho kraje, či souboru 8 sousedících SO ORP. Lokalitu navrhne dle svých možností Zpracovatel. Tato databáze slouží jako testovací vzorek (tzv. *proof of concept* – PoC), na kterém bude ověřována využitelnost dat pro stanovené účely. Cílem realizace PoC je vhodně nastavit parametry požadovaných dat pro finální databázi dodávanou pro celé

<sup>37</sup> Pro potřeby této kapitoly jsou užívány pojmy Zpracovatel ve smyslu zpracovatel zakázky, poskytovatel dat, primární zpracovatel dat a pojem Zadavatel ve smyslu pořizovatel da, zadavatel veřejné zakázky.

<sup>38</sup> V mírně pozměněné podobě je specifikace uvedena v následujících částech této kapitoly.



území ČR. V průběhu PoC jsou zástupci Zpracovatele součinní se zástupci Zadavatele ve snaze najít optimální nastavení parametrů.

## ad2) Finální databáze za celé území Česka

Po nalezení ideálního nastavení parametrů sledovaných ukazatelů v rámci PoC je možné přistoupit ke zpracování finální databáze. Tu dodavatel dodá ve stanoveném formátu, ve stanoveném čase, kompletní za celé území Česka dle nastavených parametrů.

V tomto směru rozlišujeme **fixní a variabilní parametry**.

**Fixní parametry** – jsou pevně nastaveny v zadávací dokumentaci jak pro testovací databázi PoC, tak i pro finální databázi za celé území Česka. Typově se jedná o parametry:

- Měřené ukazatele, územní detail (úroveň obcí), rozsah měření (počet měřených týdnů), způsob měření aj.

**Variabilní parametry** – jsou pevně nastaveny v zadávací dokumentaci pro testovací databázi PoC. V rámci PoC jsou tyto parametry testovány s cílem nalézt ideální nastavení pro finální databázi za celé území Česka. Typově se jedná o parametry:

- Parametry konkrétních ukazatelů např. typologie osob (rezident, dojíždějící, tranzitující), obdobně typologie cest (dojížděka za prací/do škol, za službami, časový detail měření aj.).

Při specifikaci konkrétního parametru v následujících podkapitolách je vždy určeno, zda se jedná o parametr fixní či variabilní.

## 4.3 Obecné požadavky na výslednou databázi

Zpracovatel při tvorbě výsledných databází statistických údajů musí postupovat dle níže zadaných požadavků. Níže zadané parametry jsou povětšinou fixními požadavky (není-li určeno jinak). Variabilní požadavky/parametry se týkají zejména specifikace sledovaných jevů.

**Objem dodávaných dat** – Statistické údaje v databázích musí vycházet z analýzy geolokačních dat (lokalizačních dat) mobilních operátorů tzn. ze záznamů v signalizační síti. Je žádoucí, aby soubor SIM karet, které má Zpracovatel k dispozici, byl co nejrobustnější. **Minimálně musí mít k dispozici základní soubor o rozsahu 3 miliony SIM karet reprezentativně rozmístěných po celém území Česka<sup>39</sup>**. Celkový soubor SIM karet musí umožnit analýzu na úrovni obcí. SIM karet tedy musí být dostatečné množství a musí být prostorově rozptýlené, aby popsaly 3 sledované jevy a všechny požadované ukazatele<sup>40</sup> za každou obec v Česku. *[fixní požadavky]*

<sup>39</sup> Hodnota stanovena na základě údajů českého telekomunikačního úřadu tak, aby každý z hlavních mobilních operátorů měl ve své síti dostatečný počet SIM karet ke splnění této podmínky. Fakticky je tím zajištěno, že Zpracovatel bude analyzovat data některého z těchto operátorů, popřípadě více operátorů.

<sup>40</sup> Specifikace 3 sledovaných jevů a požadovaných ukazatelů viz následující podkapitoly.

Objem dodávaných dat, přesněji rozsah primárního souboru SIM karet, které do analýz vstupují, bude předmětem hodnocení nabídky žadatelů. V tomto směru budou zvýhodněny nabídky, které budou kombinovat údaje více operátorů (ideálně všech na trhu).

Zpracovatel zakázky musí mít v době podání nabídky **zajištěn přístup k požadovaným údajům** ze signalizační sítě. Tzn., že jsou buď přímo subjektem, který získává zdrojová (lokalizační) data ze sítě (tj. mobilní operátor), nebo mají s tímto subjektem smlouvu o poskytnutí dat za účelem zpracování této zakázky.

Ze základního souboru analyzovaných SIM karet je nutné vyřadit technické a duální SIM karty<sup>41</sup>. V těchto případech je nutné využít specifických nástrojů, aby data o těchto SIM kartách neovlivňovala výslednou databázi.

**Územní dimenze** – Zadavatel požaduje dodání požadovaných statistických údajů za celé území Česka. Všechny údaje jsou požadovány za každou z 6 254 popř. 6 258 obcí<sup>42</sup>. Praha a ostatní územně členěná velká města Česka jsou ve výsledné databázi uvažovány jako jedna obec. **Granularita neboli územní detail je tedy požadován do úrovně obcí.** *[fixní požadavky]*

**Časová dimenze** – požadovaná statistická data jsou výsledkem analýz pohybu a pobytu SIM karet v signalizační síti. **Pohyb a pobyt SIM karet je sledován celkem ve dvou měřeních, každé měření sleduje období čtyř po sobě následujících týdnů.** *[fixní požadavky]* Období měření jsou vybírána tak, aby nebyla ovlivněna sezónními vlivy, jako jsou letní, jarní, či vánoční prázdniny, státní svátky apod. Jedno měření je požadováno v jarním období<sup>43</sup>, druhé v podzimním<sup>44</sup>. Nicméně vzhledem k současné epidemické situaci v Česku není v době vypisování veřejné zakázky jasné, jaká případná omezení pohybu osob budou v daných obdobích platit. Přesné vymezení termínů uskutečnění měření tak bude po vzájemné dohodě Zadavatele a Zpracovatele definováno ve smlouvě s ohledem na aktuální situaci, přičemž se obě strany budou pokoušet nalézt vhodné řešení, aby byla co nejméně ovlivněna výpovědní hodnota dat v důsledku nestandardních podmínek.

Mobilita SIM karet je sledována v pravidelných **intervalech, které nejsou delší než 60 minut.** *[fixní požadavky]* V každém analyzovaném intervalu je pobyt SIM karty/jedince přiřazen buď celý k té lokalitě, ve které strávil nejvíce času, popřípadě může být zvolena metoda rozpočtení mezi jednotlivé lokality v poměru stráveného času<sup>45</sup>.

<sup>41</sup> Problematika duálních a technických SIM karet je blíže popsána v kapitole 3.2.

<sup>42</sup> V Česku je v současnosti 6 254 samostatných obcí a 4 vojenské újezdy (bez stálého obyvatelstva). Sám Zpracovatel zvolí soubor buď 6 254, nebo 6 258 jednotek, dle nastavení svých nástrojů pro získávání dat.

<sup>43</sup> Ideálně od dubna (po velikonočních svátcích) přibližně do poloviny června (v roce 2021 jsou svátky 1. a 8. května v sobotu).

<sup>44</sup> V ideálním případě období září a října (nutné vhodně nastavit vzhledem ke svátkům 28 září a 28 října.)

<sup>45</sup> Zde popisované intervaly se týkají samotné analýzy pobytu SIM karet/osob, kterou realizuje Zpracovatel zakázky. Mohou být zvoleny různé metody zaznamenávání pobytu a rozličné intervaly (ne delší než 60 minut). Zvolená metoda i interval však nejsou závislé na intervalech výsledné databáze denního chodu přítomného obyvatelstva, jejíž parametry jsou definovány dále v textu.

## 4.4 Metodologické požadavky

Každý účastník výběrového řízení musí v rámci své nabídky předložit „Návrh řešení“, kde bude zcela jasně popsáno, jaký postup zhotovení zakázky a obsah databáze, který se zavazuje dodat.

Návrh řešení bude obsahovat i část nazvanou „metodologický postup řešení“, kde podrobně a věrohodně popíše způsob získávání dat a jednotlivé operace na nich učiněné.

**Metodický postup řešení** – účastník výběrového řízení jej podává současně s nabídkou jako součást Návrhu řešení. Metodický postup řešení bude obsahovat zcela konkrétní odpovědi na níže popsaných 5 tematických okruhů (TO). Těchto 5 částí bude předmětem hodnocení výběrové komise. Hodnocena bude jednak relevantnost a celková kvalita navrhovaného řešení a zároveň i technická proveditelnost daného řešení. Veškeré popisované nástroje a podstata jejich fungování musí být vysvětleny v dostatečné míře, aby bylo možné relevantnost řešení posoudit<sup>46</sup>.

**TO 1: Tvorba sítě servisních polygonů BTS** – předkladatel nabídky popíše zcela konkrétně, **jak dochází k vymezení sítě servisních polygonů BTS**. Jaké aspekty sítě zohledňuje (např. orientace vysílače, terén, půdní pokryv), jaké konkrétní nástroje k tomu využívá a jak je v modelu servisní sítě řešena problematika překryvu území obsluhovaných jednotlivými vysílači BTS (typově jedna lokalita v dosahu více vysílačů). Dále v této části vysvětlí, jak řeší problém **nesouladu sítě buněk BTS se správními obvody obcí**. Zcela konkrétně zde popíše, jak dochází k přerozdělení údajů z jednotlivé buňky sítě na území jednotlivých obcí, jakých metod a nástrojů je k tomu využíváno, v čem spočívá jejich funkčnost a které aspekty zohledňuje<sup>47</sup>. Musí být konkrétně uvedeno, jak bude řešen případ, **kdy jedna servisní buňka obsahuje intravilán dvou obcí**<sup>48</sup>. Jak je v takovém případě řešeno rozdělení údajů mezi tyto obce.

**TO 2: Technika zpracování záznamu** – zcela přesně musí být popsána technika, jak dochází ke **zpracování záznamů do časových intervalů**<sup>49</sup>. Jak se záznamy uskutečněné ve sledovaném intervalu (časovém snímku) projeví ve výsledných údajích? Jak jsou řešeny tzv. migrační intervaly<sup>50</sup>, tj. **čas mezi záznamy během kterého došlo k pohybu SIM karty**<sup>51</sup>? Řešení musí být jednoznačně popsáno a vysvětleno<sup>52</sup>.

**TO 3: Řešení problémů vycházejících ze samotného fungování sítě** – musí být popsáno, jak jsou řešeny konkrétní problematické jevy, jež se v rámci samotného fungování sítě projevují a mohou negativně ovlivnit validitu dat. V první řadě je to tzv. **překmitávání**, kdy se může **SIM karta změnit vysílač BTS, ke**

<sup>46</sup> Za dostatečné vysvětlení podstaty fungování nelze považovat např. tvrzení, že je využito matematického modelu. V takovém případě je zapotřebí podrobně popsat, jak model funguje, jaké proměnné do něj vstupují, popřípadě jaké proměnné jsou ve výstupech zohledněny.

<sup>47</sup> Například využití land use, adresních bodů, informací o typu zástavby, (pokud jsou využívány, vždy konkretizovat jakým způsobem), podrobně popsat i další využití metody, například využití gridu (konkretizovat jeho parametry a aspekty které zohledňuje) či kombinace s jinými systémy např. GPS.

<sup>48</sup> Pro tyto účely je vhodné využít schématu č. 7 kapitola 3.1, tzn. požadovat po uchazečích vysvětlení, jak jejich navržené metody vyřeší tento konkrétní příklad.

<sup>49</sup> Interval = maximálně 60minutové snímky, které jsou dále analyzovány, popřípadě je využito jiné techniky spočívající v rozdělení intervalu mezi jednotlivá místa (viz časová dimenze výše v této kapitole).

<sup>50</sup> Jedná se o období mezi dvěma po sobě následujícími připojeními ke dvěma různými BTS.

<sup>51</sup> Například je SIM karta považována za přítomnou v buňce dané BTS, dokud se nepřipojí k jiné BTS, popřípadě může být tento interval mezi připojeními rozdělen mezi příslušné buňky v určitém poměru.

<sup>52</sup> Pro lepší názornost problematiky tohoto tematického okruhu je možné využít modelového příkladu uvedeného ve schématu 5 podkapitoly 2.6.

kterému se hlásí, aniž by však k jejímu fyzickému pohybu došlo<sup>53</sup>. Jak je tento problém řešen a jak je zajištěno, aby překmitávání nenarušilo výpovědní hodnotu dat? Dále je pak v této části nutné vysvětlit, jak je řešena problematika **lokalit na okraji servisní sítě** (příhraniční oblasti, horské oblasti). V těchto lokalitách mohou jedinci často během dne zmizet z dosahu sítě a nejsou k jejich pohybu/pobytu jakékoli záznamy<sup>54</sup>. Jak jsou takové případy na základě zvolené techniky/metody přiřazovány ke konkrétním lokalitám a jak je zajištěno, aby tím nedošlo ke zkreslení dat v těchto oblastech? Obdobně se projevují i případy, kdy jedinec na určitou část dne vypíná své mobilní zařízení.

TO 4: Projekce na populaci – je nutné zcela jasně a podrobně vysvětlit, jak dochází k **přepočtu ze sledovaného vzorku (SIM karty) na statistické údaje vztažené k chování celé populace**. Jak je přepočet uskutečňován, jaké parametry do něj vstupují a jaké aspekty jsou zohledněny. K jakým ukazatelům jsou data vztahována popřípadě kalibrována a jak se výsledné údaje liší vůči konvenčním zdrojům obdobných dat. Jak zohledňuje navržené řešení přepočtu **problematiku reprezentativnosti vzorku**, kdy jsou mobilní zařízení využívána zejména částí populace s vysokou osobní mobilitou, naopak osoby nevlastnící mobilní zařízení jsou osoby s nízkou každodenní mobilitou (typově děti přibližně do 10 let, senioři aj.).

TO 5: Anonymizace údajů – data je nutné anonymizovat tak, aby nebylo možné spojit konkrétního jedince s konkrétním údajem. V rámci metodického postupu řešení je nutné popsat, **jak je anonymita dat zajištěna**. Dále je nutné popsat, jak zvolené postupy pracují s **údaji pod kritickou hodnotou pro zveřejnění**<sup>55</sup>. Zadavatel veřejné zakázky v tomto směru požaduje, aby v případě údajů pod kritickou hodnotou bylo odlišeno, zda hodnota daného ukazatele 0, nebo se jedná o nízkou nezveřejnitelnou hodnotu.

## 4.5 Měřené jevy/sledované ukazatele

Za strany Zadavatele je požadována dodávka 3 kompletních databází sledujících 3 jevy.

### 1) Data o počtu obvykle bydlicího obyvatelstva

Jedná se o tabulku obsahující statistické údaje o předpokládaném počtu rezidentů každé obce Česka.

Rezident – každá **SIM karta zahrnutá do výzkumu má přiděleno právě jedno místo, kde je její uživatel rezidentem**. [fixní požadavky]) Místo, kde je jedinec rezidentem, je lokalita, kam se pravidelně vrací a kde zpravidla nocuje po většinu týdne. Jedinec je rezidentem v té lokalitě, kde **tráví ve sledovaném 4týdenním měření nejvíce času v nočních hodinách určených intervalem 0:00 – 4:00**<sup>56</sup>. [variabilní parametr]

<sup>53</sup> Překmitávání má mnoho možných důvodů např. pobyt na okraji servisní buňky, vyhodnocení signálu sousední buňky jako silnějšího, momentální přetížení některých BTS aj.

<sup>54</sup> To se může týkat například osob pracujících v zahraničí, které na významnou část dne zcela zmizí ze sítě a opětovně se přihlásí až večer.

<sup>55</sup> Hodnota daného ukazatele je tak nízká, že nemůže být zveřejněna, aby nebylo údaj možné ztotožnit s konkrétní osobou (jedná se o případy, kdy se jedná o údaje v řádech jednotek).

<sup>56</sup> Z celkových 28 sledovaných dnů v době od 0:00-4:00 každá osoba strávila nejvíce času v určité lokalitě, kterou dále považujeme za domov.

Tabulka 1: Návrh výsledné databáze údajů o počtu rezidentů

Struktura dat požadovaných pro analýzu dojížděkových vztahů v Česku							
A	B	C	D	E	F	G	H
Data o obvykle bydlícího obyvatelstva - počty rezidentů							
Obec	ZUIJ obec	Počet rezidentů obce (počet obyvatel)	z toho				Druhé bydlení
			nevyjíždějící za službami ani za prací/školou	Občasně vyjíždějící za službami	intenzivně vyjíždějící za službami	vyjíždějící za prací/školou	
Říčany	123456	15 650	10 000	4 000	1 000	650	750
Praha	654321	1 253 641	1 200 000	40 000	10 000	3 641	57 000
Kolín	987654	31 154	25 000	4 000	1 000	1 154	1 500

Tabulka 1 zobrazuje návrh výsledné databáze a povinných údajů v ní obsažených. Tabulka obsahuje 6 254 řádků<sup>57</sup> (za každou obec Česka). **Ke každé obci je přiřazen počet rezidentů (sloupec C)**, tj. počet osob, pro které je daná obec určena jako rezidentní. Jedná se o základní referenční údaj dodávaných databází, reprezentující počet obvykle bydlícího obyvatelstva obce. Počet rezidentů každé obce je dále rozdělen na 4 podskupiny (sloupce D-G)<sup>58</sup>:

- Nevyjíždějící rezidenti (sloupec D) – rezidenti obce, kteří veškeré své aktivity uskutečňují na území obce, kde jsou rezidenty. Nemají jiné významné místo v prostoru (tzv. kotevní body – *anchor points*), než je jejich rezidenční obec. V jejich prostorovém chování neexistuje jiná obec, kde by se vyskytovali během 4 po sobě následujících týdnů alespoň dvakrát v celkovém objemu alespoň 4 hodin. *[variabilní parametr]*
- Občasně vyjíždějící rezidenti za službami (sloupec E) – rezidenti obce, kteří občasně vyjíždějí do jedné nebo více dalších obcí. Tito rezidenti mají mezi svými významnými jinou obec, kterou navštíví alespoň dvakrát během 4 po sobě následujících týdnů a stráví v ní dohromady alespoň 4 hodiny. Neexistuje u nich však žádný z intenzivnějších typů vyjížděky, tj. tito rezidenti nenavštíví žádnou obec více než čtyřikrát ve sledovaném období a netráví zde celkem více než 8 hodin. Sloupec E uvádí počet rezidentů, kteří zadaným parametrům vyhovují<sup>59</sup>. *[variabilní parametr]*
- Intenzivně vyjíždějící rezidenti za službami (sloupec F) – rezidenti obce, kteří z obce často a pravidelně vyjíždějí do jedné nebo více obcí. Tito rezidenti mají mezi svými významnými místy obec, kterou navštíví alespoň čtyřikrát během 4 po sobě následujících týdnů a stráví v ní dohromady alespoň 8 hodin. Neexistuje u nich však kotevní bod – obec, které by navštívili během sledovaného období alespoň 16krát a strávili zde alespoň 80 hodin. Sloupec F uvádí počet rezidentů, kteří zadaným parametrům vyhovují<sup>60</sup>. *[variabilní parametr]*

<sup>57</sup> V případě že Zpracovatel ve své síti obcí započítává i území vojenských újezdů jednalo by se o 6 258 řádků/obcí.

<sup>58</sup> Součet hodnot ve sloupcích D, E, F, G musí být roven hodnotě sloupce C.

<sup>59</sup> Osoby započítané do sloupce E nejsou započítávány do jiných sloupců. Je-li jejich vyjížděka do jiných obcí vyšší je jedinec zaznamenán mezi typy vyjíždějících s větší intenzitou, tj. bude sloupec F nebo G. Má-li osoba více významných míst se stejnou intenzitou, je v konkrétním sloupci zaznamenána jen jednou.

<sup>60</sup> Osoby započítané do sloupce F nejsou započítávány do jiných sloupců. Naplňuje-li osoba zároveň i parametry pro započítání do sloupce E, je započítána pouze do sloupce F. Naplňuje-li osoba zároveň i parametry pro započítání do sloupce G, je započítána pouze do sloupce G. Má-li osoba více významných míst se stejnou intenzitou, je v konkrétním sloupci zaznamenána jen jednou.

- Rezidenti vyjíždějící za prací a do škol (sloupec G) - rezidenti obce, kteří pravidelně vyjíždějí za prací a do škol do jiné obce. Tito rezidenti mají mezi svými významnými místy takovou obec, kterou navštíví během sledovaného období alespoň 16krát a stráví zde v součtu alespoň 80 hodin. Sloupec G uvádí počet rezidentů, kteří zadaným parametrům vyhovují <sup>61</sup>. [variabilní parametr]

Skupiny vyjíždějících ve sloupcích E, F a G svým vymezením odpovídají jednotlivým typům cest (typ 1, 2 a 3) definovaných pro databázi meziobecní dojíždky (viz dále v textu).

Každému rezidentovi je v této databázi přiřazena právě jedna z kategorií vyjíždky (sloupce D-G). Jednotlivé osoby často naplňují znaky několika zmíněných typů vyjíždky, a to i vůči různým obcím. V takovém případě je přiřazen ke skupině s nejintenzivnějším typem vyjíždky.

**Praktický příklad:**

Osoba jezdí pravidelně do jedné obce za prací (sloupec G), do jiné obce jezdí za nákupy (sloupec F) a do jiné obce jezdí na občasně návštěvy příbuzných (sloupec E). V takovém případě bude daná osoba označena dle nejintenzivnější vazby, kterou uskutečňuje, tj. bude označena jako vyjíždějící za prací, a tudíž započítána pouze do sloupce G.

Specifickým typem dojíždky je tzv. **druhé bydlení**, kdy rezidenti obce pravidelně nocují v jiné obci, kde se po určitou dobu zdržují a v daném místě se chovají obdobně jako místní rezidenti. Charakter druhého bydlení může mít různou intenzitu i periodicitu (většinou týdenní rytmus) a liší se v průběhu roku. Typově se jedná zejména o víkendové pobyty na chalupách či chatách, dojíždka studentů či pracovníků vícedenního charakteru a podobně.

Druhé bydlení (sloupec H) – tento sloupec obsahuje počty osob, které kromě své rezidentní obce pravidelně nocují i v jiné obci. Intenzita druhého nejčastějšího místa přenocování je alespoň 6 přenocování během sledovaného období. Za přenocování se považuje, že jedinec v daném místě stráví většinu intervalu 0:00-4:00. Zmíněných 6 přenocování v obci musí být uskutečňováno v pravidelné periodicitě, tj. všechna přenocování uskutečněna ve víkendovém období (PÁ-SO, SO-NE, NE-PO) nebo ve všední dny (PO-ÚT, ÚT-ST, ST-ČT, ČT-PÁ)<sup>62</sup>. Každá osoba může mít přiřazenu jen jednu obec druhého bydlení<sup>63</sup>. [variabilní parametr] **Sloupec H uvádí údaj o počtu rezidentů jiných obcí, pro které je daná obec místem druhého bydlení.**

<sup>61</sup> Osoby započítané do sloupce G nejsou započítávány do jiných sloupců (D, E či F), přestože mohou naplňovat u různých obcí parametry pro započítání do konkrétních sloupců.

<sup>62</sup> Tím dojde k částečnému odstínění rekreatů (velký počet přenocování v kuse), a odstínění nočních směnných provozů (směny v různých částech týdne).

<sup>63</sup> V případě, že u dané osoby jsou dvě obce, které splňují parametry druhého bydlení, za druhé bydlení je považována ta s větším počtem přenocování během daného období.

## 2) Denní chod počtu přítomného obyvatelstva

Jedná se o rozsáhlou databázi, která v každé obci Česka sleduje počet obvykle přítomného obyvatelstva během jednotlivých dnů v týdnu. Za každou z 6 254 či 6 258 obcí je v databázi k dispozici denně 24 hodinových intervalů, 7 dní v týdnu, tj. celkem 168 záznamů ke každé obci a každému sledovanému údaji. Databáze obsahuje statistické údaje na základě sledování během celého měření, tedy 4 po sobě následující týdny (celkově dvě taková měření). Základním sledovaným ukazatelem je typický počet osob přítomných v obci v daném časovém intervalu. Každý jedinec je za každý interval přiřazen pouze k jedné obci a to k té, kde stráví dominantní část intervalu. Počet typicky přítomných osob v obci je dále členěn dle jednotlivých atributů přiřazených konkrétním jedincům. Tabulka 2 zobrazuje návrh databáze a její jednotlivé náležitosti.

Tabulka 2: Návrh výsledné databáze údajů o denním chodu počtu přítomného obyvatelstva

Struktura dat požadovaných pro analýzu dojížděkových vztahů v Česku										
A	B	C	D	E	F	G	H	I	J	K
Denní chod počtu přítomných osob/obyvatelstva										
Obec	ZUII obec	den v týdnu	časový interval	Celkový počet přítomných osob	z toho					
					rezidenti	dojíždějící za prací/školou	intenzivní dojížděka za službami	občasná dojížděka za službami	Druhé bydlení	tranzitující a nahodilí návštěvníci
Říčany	123456	pondělí	00:00 - 01:00	15 231	13 514	610	326	438	112	231
Říčany	123456	pondělí	01:00 - 02:00	15 652	13 850	690	312	480	98	222
Říčany	123456	pondělí	02:00 - 03:00	15 134	13 451	584	351	385	105	258
Říčany	123456				....					
Říčany	123456	pondělí	23:00 - 24:00	15 205	13 389	681	321	435	121	258
Říčany	123456	úterý	00:00 - 01:00	15 286	13 405	685	326	451	103	316
Říčany	123456				....					
Praha	654321	pondělí	00:00 - 01:00	1 195 456	1 152 351	23 512	13 462	2 485	1 268	2 378
Praha	654321				....					
Kolín	987654	pondělí	00:00 - 01:00	31 154	27 201	1 523	658	845	364	563

Počet přítomných osob (sloupec E) – Jedná o statistický údaj o typickém (průměrném) počtu přítomných osob v dané obci v konkrétním intervalu vypočítaný na základě 4týdenního sledování. Přítomné obyvatelstvo je dále členěno na jednotlivé skupiny osob. Každému jedinci je v daném intervalu přiřazen určitý atribut<sup>64</sup>. [variabilní parametr]

Atributy přiřazované přítomným osobám v konkrétním intervalu odpovídají jejich vztahu k dané lokalitě (rezidentní obec, cíl dojížděky za prací/školou, cíl dojížděky za službami – intenzivní/občasná, obec druhého bydlení). Ty odpovídají typologii osob definovaných pro databázi obvykle bydlícího obyvatelstva (viz výše) a zároveň odpovídají typologii cest (typ 1, 2 a 3) definovaných pro databázi meziobecní dojížděky (viz dále v textu).

<sup>64</sup> Součet hodnot ve sloupcích F, G, H, I, J a K musí být roven hodnotě sloupce E.



Rezidenti (sloupec F): jedná se o osoby přítomné v konkrétním intervalu, pro které je lokalita rezidenční obcí (definice viz databáze obvykle bydlícího obyvatelstva). *[variabilní parametr]* Tato statistika umožňuje sledovat každodenní rytmus obce a určit, v jaké části dne a v jakém objemu rezidenti obec opouštějí a kdy do ní opět vracejí.

Dojíždějící za prací a do škol (sloupec G): jedná se o osoby přítomné v konkrétním intervalu, pro které je daná obec cílem dojížděky za prací či do škol. V tomto směru nehraje roli čas strávený jedincem v tento den v dané obci, vždy mu bude přidělen atribut dojíždějícího za prací. Vymezení viz databáze obvykle bydlícího obyvatelstva, popř. databáze meziobecní dojížděky. *[variabilní parametr]*

Intenzivně dojíždějící za službami (sloupec H): pro tyto osoby (přítomné v konkrétním intervalu) je daná obec cílem nepracovní dojížděky s vysokou intenzitou. Vymezení viz databáze obvykle bydlícího obyvatelstva, popř. databáze meziobecní dojížděky. *[variabilní parametr]*

Občasně dojíždějící za službami (sloupec I): tento atribut je přiřazován přítomným osobám, pro které je daná obec cílem nepracovní dojížděky nízké intenzity. Vymezení viz databáze obvykle bydlícího obyvatelstva, popř. databáze meziobecní dojížděky. *[variabilní parametr]*

Druhé bydlení (sloupec J): daný atribut je přiřazen přítomným osobám, pro které je daná obec místem druhého bydlení. Vymezení viz databáze obvykle bydlícího obyvatelstva, popř. databáze meziobecní dojížděky.

Nahodilé cesty a tranzitující (sloupec K): tato skupina přítomných osob nevyhovuje výše definovaným typům osob. Těmto jedincům, kteří jsou v daném intervalu přítomni, není možné přiřadit žádný z výše uvedených atributů. Povětšinou se bude jednat o jedince, kteří se v obci objeví nahodile, či tranzitující. Přestože jsou osoby v obci přítomné, nebyla obec pro ně identifikována jako významné místo pobytu (kotevní bod) za celé sledované období.

### 3) Data o meziobecní dojížděce

Jedná se o **OD matici směrově určené pravidelné denní meziobecní vyjížděky/dojížděky**. V databázi je identifikována **zdrojová obec, cílová obec a typ uskutečňované cesty**. Zadavatel požaduje statistiku **unikátních SIM karet**, které odpovídají stanoveným parametrům. Databáze tak reprezentuje počet osob, které daný typ cesty pravidelně uskutečňuje. Databáze obsahuje pro každou obec jak **hlavní vyjížděkový proud** daného typu cesty, tak i **všechny vedlejší vyjížděkové směry**, které zadaným parametrům vyhovují. Tabulka 3 zobrazuje povinné náležitosti databáze, které musí být Zpracovatelem naplněny.

Tabulka 3: Návrh výsledné databáze údajů o meziobecní dojíždě

Struktura dat požadovaných pro analýzu dojížděkových vztahů v Česku						
A	B	C	D	E	F	G
Data o meziobecní dojíždě						
zdrojová obec	ZUJ ZO	cílová obec	ZUJ CO	typ cesty	počet osob	počet osob (PO-PÁ, 8-17)
Říčany	123456	Praha	654321	1	1000	750
Říčany	123457	Praha	654321	2	2500	2100
Říčany	123458	Praha	654321	3	3500	2900
Říčany	123456	Kolín	987654	1	500	350
Říčany	123457	Kolín	987654	2	1500	1300
Říčany	123458	Kolín	987654	3	2500	2100
Kolín	987654	Říčany	123458	1	200	150
Kolín	987655	Říčany	123458	2	300	200
Kolín	987656	Říčany	123458	3	500	350

#### Typologie cest:

OD matice bude rozdělovat dojížděkové meziobecní vazby do 3 kvalitativních stupňů typů cest. Ty se od sebe liší významem a intenzitou:

- **Typ 1 = dojížděka za prací a do škol** – nejintenzivnější typ dojížděky, který lze připodobnit ke každodenní dojížděce za prací a do škol. Každý jedinec může mít přiřazenu pouze jednu lokalitu dojížděky za prací<sup>65</sup>. Tento typ dojížděky je stanoven vůči celému sledovanému období 4 po sobě následujících týdnů. Ve sledovaném období je **dané místo navštíveno alespoň 16krát a jedinec v dané lokalitě celkem stráví alespoň 80 hodin**<sup>66</sup>. [variabilní parametr] Fakticky údaj v této kolonce udává počet rezidentů obce, kteří mají předpokládané místo výkonu práce, popř. školu v cílové obci.
- **Typ 2 = intenzivní dojížděka za službami** – typ dojížděky, který lze připodobnit k pravidelné dojížděce za službami, a to zejména nákupy, dále pak kulturní, sociální, sportovní a jiné služby, popřípadě návštěvy příbuzných a přátel. Intenzita vyjížděky/dojížděky je v tomto případě nižší než u typu 1. Každá osoba může mít více takových míst a osoba je ve statistice započítána ke každému takovému směru<sup>67</sup>. Statistika tak v tomto případě uvádí počet osob (unikátní SIM karty), které na konkrétní relaci (obec A – obec B) uskutečňují daný typ pohybu. Tento typ je stanoven rovněž vůči celému sledovanému období 4 po sobě následujících týdnů. Ve sledovaném období je **dané**

<sup>65</sup> Součet počtu vyjíždějících za prací do jednotlivých směrů (tj. všechny směry s přiděleným typem 1 z dané obce) se musí rovnat počtu vyjíždějících za prací/školou (tabulka 1 sloupec G). Mírné odchylky mohou být způsobeny v případě existence nízkých hodnot údajů, tj. pod kritickou hodnotou pro zveřejnění.

<sup>66</sup> Odpovídá intenzitě dojížděky 4krát týdně po dobu 5 hodin v dané lokalitě, přičemž srovnávacím obdobím je celé sledované období 4 týdnů.

<sup>67</sup> Součet všech osob uskutečňujících typ cesty 2 z dané obce do různých směrů může paradoxně převyšovat počet obyvatel samotné obce, jelikož někteří jedinci mohou být započítáni za každý vyjížděkový směr, který uskutečňují.

místo navštíveno alespoň 4krát a jedinec v dané lokalitě celkem stráví alespoň 8 hodin<sup>68</sup>. [variabilní parametr]

- **Typ 3 = občasná dojízdka za službami** – nejslabší dojízdková vazba, u které lze identifikovat pravidelnost v chování a lze ji proto považovat za cílenou a periodickou dojízdku. Lze jej připodobnit k nedenní dojízdce za specifickými službami, jako jsou kulturní, sociální, sportovní a jiné služby, popřípadě lékařská péče či návštěva úřadů či návštěvy příbuzných a přátel. Obdobně jako v případě typu 2 může mít osoba více takových míst a osoba je ve statistice započítána ke každému takovému směru. Opět tak statistika uvádí počet osob (unikátní SIM karty), které na konkrétní relaci (obec A – obec B) uskutečňují daný typ pohybu (typ 3). Intenzita vyjíždky/dojíždky je v tomto případě nižší než u typu 2. Tento typ je stanoven rovněž vůči celému sledovanému období 4 po sobě následujících týdnů. Ve sledovaném období je **dané místo navštíveno alespoň 2krát a jedinec v dané lokalitě celkem stráví alespoň 4 hodiny**<sup>69</sup>. [variabilní parametr]
- **Sloupec G – počty osob uskutečňující cesty během obvyklé pracovní doby** – jedná se o specifickou část úlohy, která si klade za cíl zjistit, zda dojíždějící osoby mají reálnou možnost v cílové obci vyřídit své úřední záležitosti. V tomto sloupci jsou **počty osob, jejichž cesty (typu 1, 2 a 3) jsou uskutečňovány ve všední dny v době od 8 do 17 hodin**, přičemž pobyt v cílové lokalitě může být zahájen před tímto intervalem a ukončen po intervalu. Jinými slovy jsou do této statistiky počítány jen ty osoby, které daná kritéria pro jednotlivé typy cest splňují, přestože jsou započítávány pouze ty cesty/pobyty, jež byly alespoň z části uskutečněny ve stanoveném intervalu (PO-PÁ; 8-17 h.).

### Způsob sledování dojízdkových relací:

U každého jedince jsou identifikována významná místa v prostoru (kotevní body), která v daném období pravidelně navštěvuje. Nesledujeme tedy samotný pohyb, ale spíše pobyt a z něj odvozujeme vztahy mezi místy. Nehledá se tedy přímo vztah mezi danými místy, či trajektorie, kterou jedinec během sledovaného období v prostoru zaznamenal, ale pouze **významné body, kde určitou dobu strávil a opakovaně se tam vracel**. Pohyb jedince v každém dni je vztahován k místu, kde osoba strávila předešlou noc. Pokud tedy osoba v jeden den postupně navštíví více míst za různými účely (různé typy cest 1,2,3), je v údajích vždy vztahován pohyb vůči místu, kde jedinec strávil noc<sup>70</sup>.

Druhé bydlení – parametry jsou nastaveny tak, aby byla identifikována dominantní relace mezi rezidentní a pracovní obcí. Nicméně teoreticky může dojít k situaci, kdy jedinec ve svém dopravním rytmu navštěvuje určitou obec tak často, že ji lze považovat za cíl pracovní/školní dojíždky, avšak

<sup>68</sup> Odpovídá intenzitě dojíždky jednou týdně po dobu 2 hodin v dané lokalitě, přičemž srovnávacím obdobím je celé sledované období 4 týdnů.

<sup>69</sup> Představuje nejnižší intenzitu dojíždky, u které je prokazatelná periodicitu a nejedná se tak o nepravidelné, nahodilé cesty či pouze tranzitující osoby.

<sup>70</sup> Příklad: pokud rezident obce A pravidelně během cesty do práce v obci B zastaví v obci C a tráví tam určitý významný čas, bude zaznamenána relace obce A a B jako pracovní dojízdka a obce A a C jako dojízdka za službami. Relace mezi B a C, přestože k pohybu na této trase dochází, nebude v analýze zahrnuta.

dojíždka je uskutečňována z více lokalit (z míst druhého bydlení)<sup>71</sup>. V takovém případě je zapotřebí takovou dojíždkovou relaci přiřadit k místu, odkud je nejčastěji uskutečňována.

#### Extrémní případ

Osoba během sledovaných 28 dnů přenocovala 15krát v obci A (rezidentní) a 13krát v obci B (druhé bydlení). Tato osoba 16krát navštívila svou práci v obci C a strávila zde dohromady 80 hodin. Nicméně do práce jela pouze 3krát z obce A a 13krát jela z obce B. V takovém případě bude ve výsledné databázi u tohoto jedince zaznamenána dojíždka za prací jako relace B (druhé bydlení) – C (práce). Relace A – C bude následně zaznamenána jako typ 3.

V případě dalších typů cest (nepracovní vyjíždka/dojíždka intenzivní a občasná – typ 2 a 3), kde jeden jedinec může mít více cílových lokalit, je zaznamenávána každá taková relace. A to v případě cest z rezidentní obce i z místa druhého bydliště. Osoba tak může být u typů cest 2 a 3 započítána k různým zdrojovým obcím.

#### Praktický příklad:

Osoba je rezidentem v obci A a druhé bydlení má v obci B. Z každé z obcí jezdí za různými službami v okolních obcích C a D, přičemž intenzita dojíždky naplňuje kritéria cesty typu 2. Ve výsledné databázi bude jedinec započítán jako dojíždějící typu 2 z obce A do obce C a zároveň bude započítán jako dojíždějící typu 2 mezi obcemi B a D.

## 4.6 Konečný formát výstupů

Veškeré výstupní databáze, a to jak v rámci PoC, tak v rámci finální dodávky dat, budou zpracovány tak, aby bylo možné je otevřít a editovat v MS Excel. Samotné poskytované výstupy budou ve formátu .xlsx. Konkrétní podoba jednotlivých tabulek je naznačena v tabulkách 1, 2 a 3 a je po obsahové stránce (sledované ukazatele) pro Zpracovatele závazná. Případné změny podoby tabulek je možné po vzájemné dohodě zahrnout buď do smlouvy nebo je učinit v rámci PoC.

Ve všech zpracovaných tabulkách budou obce identifikovány podle totožného ukazatele, kterým bude kód obce (ZUJ), dle aktuálního číselníku obcí (CISOB) vedeného ČSÚ.

<sup>71</sup> Příkladem, sledovaná osoba může některé dny jezdit do práce z domova a občas z chalupy.

Dodání výsledných dat (finální databáze za celé území Česka) bude obsahovat tři soubory databází dle jednotlivých jevů (.xlsx formát) a mapovou přílohu ke každému ze sledovaných jevů. Mapová příloha bude obsahovat základní zobrazení sledovaných jevů prostřednictvím tematických map. Tyto mapy mají doprovodný charakter, a slouží k představení obsahu každé z databází. Konkrétní zobrazované ukazatele v mapách budou dohodnuty v rámci PoC.

#### 4.7 Licenční model

Zadavatel zakázky bude data dále analyzovat a připravovat vlastní výstupy. Ty budou spočívat jak v kartografické vizualizaci získaných údajů, tak ve výpočtech vlastních ukazatelů na základě těchto dat. Veškeré výstupy budou projektu volně dostupné veřejnosti, přičemž z kartografických výstupů budou dohledatelné konkrétní poskytované údaje za jednotlivé obce.

**Zadavatel je veřejným subjektem vázaným povinnostmi zajistit svobodný přístup k informacím<sup>72</sup>, přičemž veřejná zakázka je plně hrazena veřejnými zdroji. V případě odůvodněné žádosti ze stran krajů, konkrétních obcí či jiných správních orgánů Zadavatel dané údaje poskytne.**

Zadavatel požaduje možnost volného nakládání se zakoupenými databázemi, aby bylo možné dostát výše zmíněným závazkům. **Zadavatel se současně zavazuje, že data nebude dále přeprodat či jinak s nimi obchodovat.**

<sup>72</sup> Zákon č. 106/1999 Sb. o svobodném přístupu k informacím.

## 5 Závěr

Tento dokument přináší základní vhled do problematiky geolokačních dat mobilních operátorů. Jeho cílem je jednak představit tuto oblast široké veřejnosti a ukázat možnosti využití těchto údajů v rámci organizace nejen služeb veřejné správy, ale i celkových socioekonomických aktivit ve společnosti.

Dokument by měl sloužit případným zájemcům (zejména z oblasti veřejné správy), kteří uvažují o pořízení těchto dat. V textu naleznou potenciální zájemci užitečné informace, jak v celém procesu postupovat. Dochází tak tímto způsobem k transferu a sdílení zkušeností s přípravou veřejné zakázky na nákup geolokačních dat mobilních operátorů. Tato data mají velký potenciál využití, avšak je zapotřebí zcela precizně specifikovat své požadavky, aby byly údaje pro konkrétní účel využitelné. Vzhledem k finanční náročnosti nákupu těchto dat je vhodné své zkušenosti sdílet.

Postupně je v dokumentu možné se seznámit se samotnou podstatou geolokačních dat a fungování signalizační sítě. Dále pak jsou rozebírány konkrétní aspekty specifikace dat a problematické oblasti s tím spojené, které je vhodné s poskytovatelem dat prodiskutovat. Rovněž je v tomto směru poukazováno na jistá omezení tohoto datového zdroje. Následně je pak představen konkrétní příklad specifikace požadavků na data dodávaná v rámci veřejné zakázky jako součást projektu s názvem „Zlepšení podmínek pro decentralizaci a dostupnost veřejné správy v území“.

V budoucnu bude dokument doplňován přílohami, které budou popisovat další fáze realizace projektu. V tomto směru bude sdílena zkušenost s přípravou a samotným procesem výběrového řízení na poskytovatele dat, průběhem předběžných tržních konzultací s potenciálními Zpracovateli, dále pak zkušenosti s realizací fáze PoC a následně sdílení zkušeností s analýzou a kartografickou vizualizací získaných dat.

Údaje ze signalizační sítě mobilních operátorů mají velmi široké využití. Je však zapotřebí si klást správné výzkumné otázky, na které lze odpovědět, a precizně specifikovat požadovaná data, aby odpovídala účelu, za kterým jsou pořizována.

## 6 Užité zdroje

- AHRİPOVA, I. a kol. (2019): Mobile phone data statistics as a dynamic proxy indicator in assessing regional economic activity and human commuting patterns.
- BACHIR, D. (2019): Estimating urban mobility with mobile network geolocation data mining.
- CACERES, N., WIDEBERG, J. P., BENITEZ, F. G. (2007): Deriving origin–destination data from a mobile phone network.
- ČESKÝ TELEKOMUNIKAČNÍ ÚŘAD (2019): Výroční zpráva za rok 2019.
- GSMweb.cz (2020): Seznamy BTS, mapy BTS - nezávislý informační server.
- FRÍAS-MARTÍNEZ, V., SOGUERO RUIZ, C., FRIAS-MARTINEZ, E. (2012): Estimation of urban commuting patterns using cellphone network data.
- MAZOUCH, P. a kol. (2017): Limity využití mobilních sítí ve statistických šetřeních ČSÚ.
- NOVÁK, J. (2010): lokalizační data mobilních telefonů: možnosti využití v geografickém výzkumu.
- OLIVER, N., MATIC, A., FRIAS-MARTINEZ, E. (2015): Mobile Network Data for Public Health: Opportunities and Challenges.
- ŠVEDA, M., BARLÍK, P. (2018): Daily commuting in the Bratislava metropolitan area: case study with mobile positioning data.
- THUILLIER, E. a kol. (2019): Clustering Weekly Patterns of Human Mobility Through Mobile Phone Data.
- YANG, X. a kol. (2018): Understanding the Spatial Structure of Urban Commuting Using Mobile Phone Location Data: A Case Study of Shenzhen, China.