



INSTITUT
PRO DOPRAVNÍ EKONOMII,
GEOGRAFII A POLITIKU

Zkušenosti z využití big data a připravovaný nákup nových datasetů

XIII. Seminář Telč 2018

v rámci projektu Nová mobilita – vysokorychlostní dopravní systémy a dopravní chování populace

Daniel Seidenglanz, Martin Kvizda

8. listopadu 2018



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



Obsah

1. **Úvod** – co víme o big data
2. **Naše zkušenost s big data** – případová studie V. Meziříčí
3. **Big data pro Novou mobilitu** – k čemu je chceme?

Cíl

- v kontextu projektu Nová mobilita vysvětlit, co chceme a co můžeme dostat;
- definovat parametry objednávky big data pro projekt.

1. Úvod – co víme o big data

- **zásadní přínos *big data*:**

Komplexní informace o všech typech pohybů v prostoru →
příležitost zkoumat nově a přesněji tradiční fenomény:

- prostorovost mobility
- temporalita mobility
- prostorové vztahy
- síťové analýzy
- kvantifikace poptávky po přepravě
- struktura segmentů poptávky

■ rizika použití:

- omezená hloubka informace (jen záznam pohybu) – schází motivace, identita, postoje, ...:
 - *trajectory data rich but activity information poor problem* (Gong et al., 2015; Chen et al., 2016)
 - *defective conclusion about individual behaviour drawn from data about aggregate behaviour* (Chen et al., 2016, 291; Freedman, 2002)
 - *human movement is not simply the result of mechanical process, but to a great extent influenced by a number of factors that vary from place to place and from population to population* (Chen et al., 2016, 292)

- techniky sběru a úpravy dat:
 - okamžiky detekce SIM karet a další technické záležitosti
 - posun informace – s čím výzkumník ve výsledku pracuje

- reprezentativnost dat (Patel et al., 2006; Wang et al., 2014, Chen et al., 2016):
 - *phone proximity*
 - vlastnictví více telefonů
 - míra penetrace

- velikost datasetů → náročnost na zpracování

- *new scientific, positivist, quantitative turn* (Schwanen, 2015, 2016, 2017)

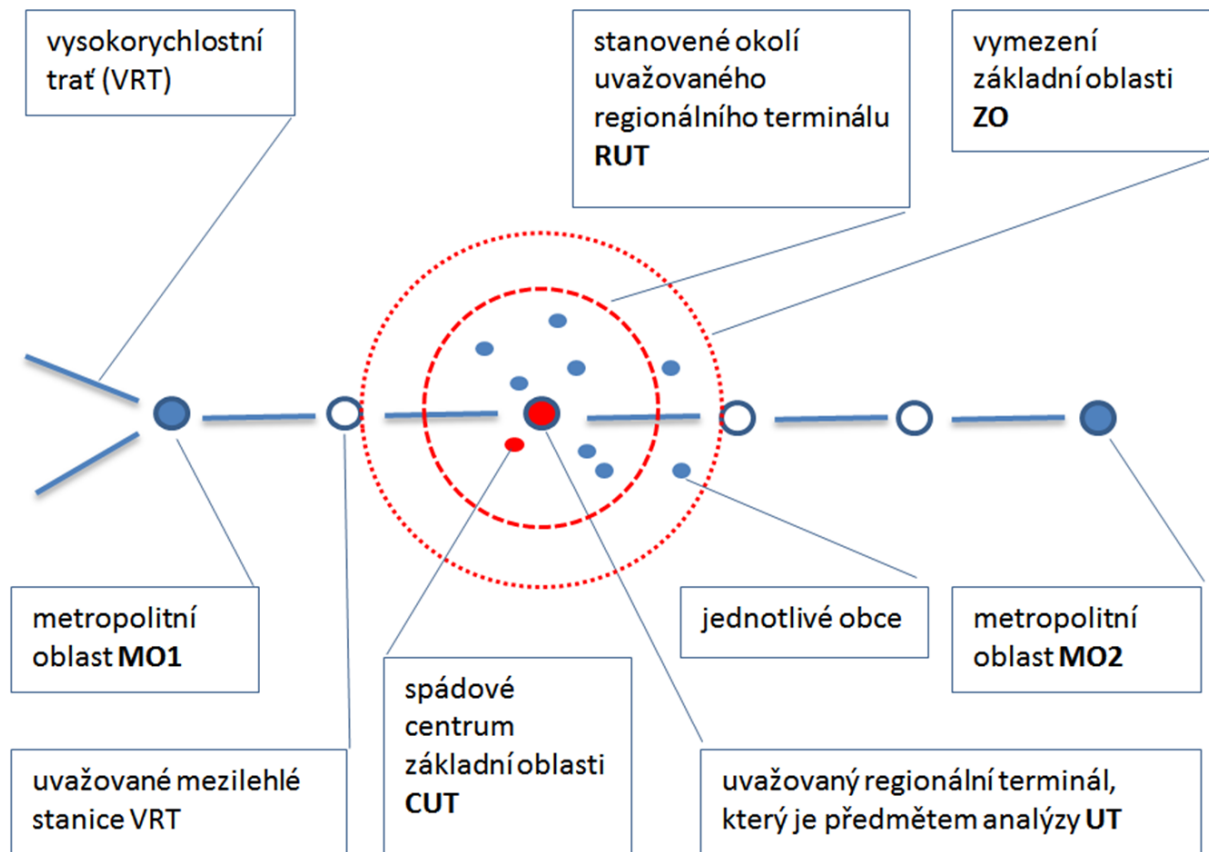
Praktické využití big data – sledovaný účel

- **města:** Řím (*Calabrese et al. 2011*), Abidjan (*Berlingerio 2013*), Boston (*Calabrese et al. 2013*), Budapešť (*Tettamanti et al. 2012*)
 - **rozsáhlé oblasti, omezená analýza:** Pobřeží slonoviny (*Liu et al. 2014*), Rakousko (*Horn a Kern 2015, Elias et al. 2016*), Senegal a Pobřeží slonoviny (*Gundlegård et al. 2016*)
 - **paralelní dotazníkové šetření** (*Bohte – Maat 2008, Chen et al. 2010, Gong et al. 2011, Liu et al. 2014, Calabrese et al. 2011, Calabrese et al. 2013, Chen et al. 2014*)
- **město Brno** – pohyb a přítomnost obyvatel
 - **ZSSK** – dojíždka do Bratislavy
 - **ČD** – linka Plzeň-Praha

2. Naše zkušenost s big data – případová studie V. Meziříčí

- projekt TAČR č. TD03000061 *Význam regionálních terminálů vysokorychlostních tratí pro hustotu dopravních proudů (2016 - 2017)*:
 - cíl projektu – vytvoření metodiky pro hodnocení socio-ekonomického významu regionálních terminálů připravovaných vysokorychlostních tratí (VRT)

- využití *big data* v projektu:
 - stanovení rozsahu zázemí potenciálního regionálního terminálu VRT ve Velkém Meziříčí
 - stanovení současné síly vazby k cílům potenciální VRT – velikost proudů mezi Velkomeziříčskem a Prahou a Brnem, respektive jejich metropolitními regiony
 - testování metodiky (!)



- atributy pořízených *big data*:
 - období:
 - měsíc bez sezónních výkyvů (březen, duben, říjen, listopad)
 - my – středa, 20. IV. 2016
 - prostorové atributy dat:
 - obec jako základní územní entita (eliminace lokalizační nepřesnosti *big dat*)
 - cesty typů: Velkomeziříčsko – Velké Meziříčí / Praha (Praha + PZ + PV) / Brno (Brno + BO)
 - časová specifikace cesty (odjezd a příjezd)
 - opakování cesty (anonymizovaný kód)
 - kvantifikátor *count* – přepočítání tržního podílu na celou populaci
 - 125 251 řádků

index	timefrom	timeto	KOD_OBEC_P_from	KOD_OBEC_P_to	id	count
77	3	5,25	584282	597007	63	2,63
86	3,25	6	581321	597007	72	2,92
129	4	5	582786	597007	115	2,31
245	4,75	6	584801	597007	220	2,73
313	5	6,5	583839	597007	283	2,31
327	5	7,5	583120	597007	297	2,71
382	5,25	6,5	584002	597007	352	2,05
436	5,5	6	597171	597007	403	1,74
539	5,75	6,75	584207	597007	506	3,79
540	5,75	6,75	583995	597007	507	2,64

Poznámky: **index** – číslo řádku v databázi

timefrom – okamžik zahájení cesty (časové údaje jsou převedeny do desetinné soustavy)

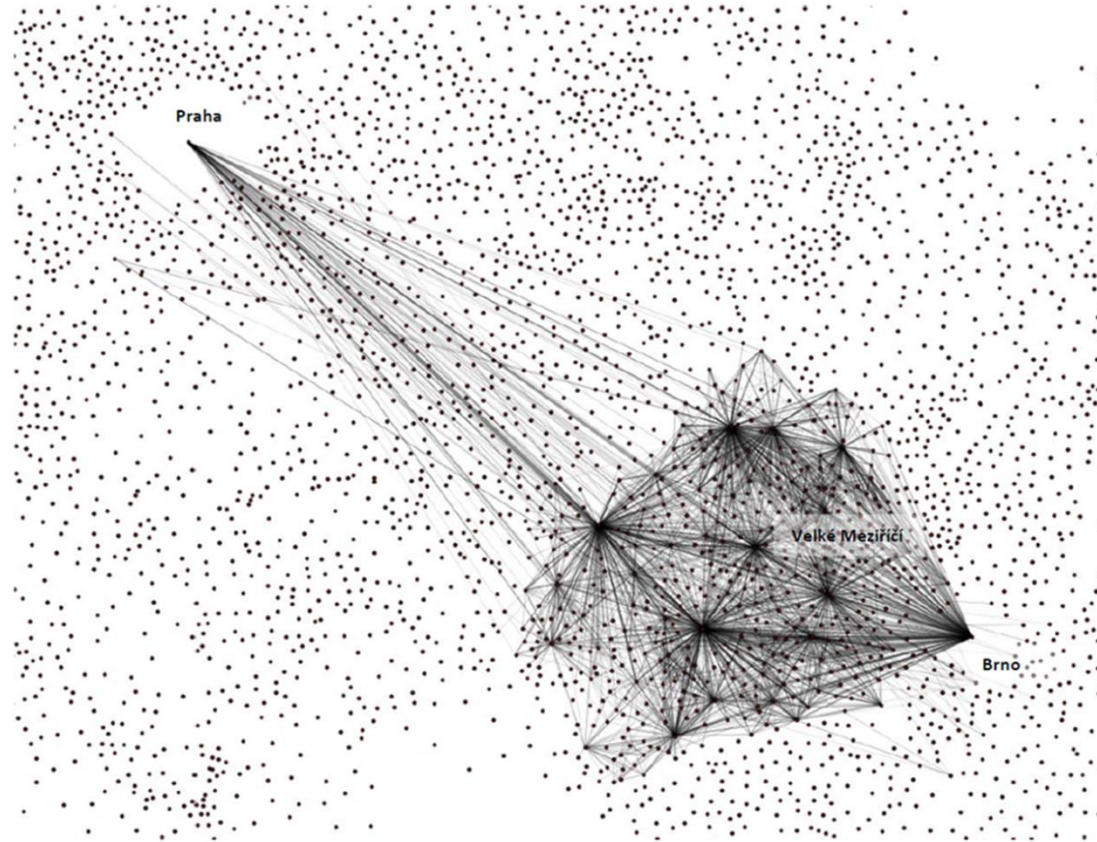
timeto – okamžik ukončení cesty (časové údaje jsou převedeny do desetinné soustavy)

KOD_OBEC_P_from – kód obce, v níž byla zahájena cesta, podle číselníku obcí CISOB (ČSÚ)

KOD_OBEC_P_to – kód obce, v níž byla ukončena cesta, podle číselníku obcí CISOB (ČSÚ)

id – anonymizovaný kód SIM karty

count – index udávající tržní podíl společnosti T-Mobile, index umožňuje přepočítání údajů na celkovou populaci



- výsledky analýz založených na zpracování *big data*

Tab. 2 RUT Velké Meziříčí – obce tvořící zázemí UT

	Počet obcí		Počet obyvatel		Počet cest do Velkého Meziříčí	
	(abs.)	(%)	(abs.)	(%)	(abs.)	k počtu obyvatel (%)
RUT Velké Meziříčí celkem	59	100,0	117 895	100,0	4 822	.
z toho samotné Velké Meziříčí:						
Velké Meziříčí	1	1,7	11 645	9,9	.	.
z toho obce přiřazené do RUT na základě relativního kritéria (poměr cesty/počet obyvatel):						
≥ 25,0 %	11	18,6	6 941	5,9	2 064	29,7
≥ 10,0 % and < 25,0 %	28	47,5	9 599	8,1	1 311	13,7
≥ 5,0 % and < 10,0 %	14	23,7	8 403	7,1	559	6,7
z toho obce přiřazené do RUT na základě absolutního kritéria (počet cest):						
Třebíč, Žďár nad Sázavou, Náměšť nad Osavou, Nové Město na Moravě, Bystřice nad Pernštejnem)	5	8,5	81 307	69,0	888	1,1

Pramen: ČSÚ (2017), databáze big dat

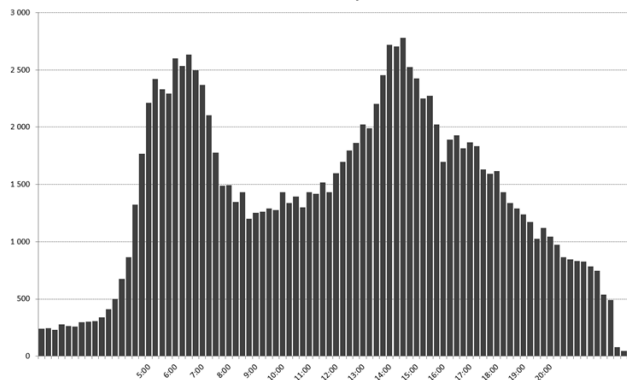
Tab. 3 RUT Velké Meziříčí – počet cest do Prahy a Brna a jejich metropolitních regionů

	Počet cest do:			
	Prahy (obec Praha)	pražského metropolitního regionu (obec Praha, okresy Praha- západ a Praha- východ)	Brna (obec Brno)	brněnského metropolitního regionu (obec Brno a okres Brno-venkov)
RUT Velké Meziříčí	1 189	1 505	4 451	7 510

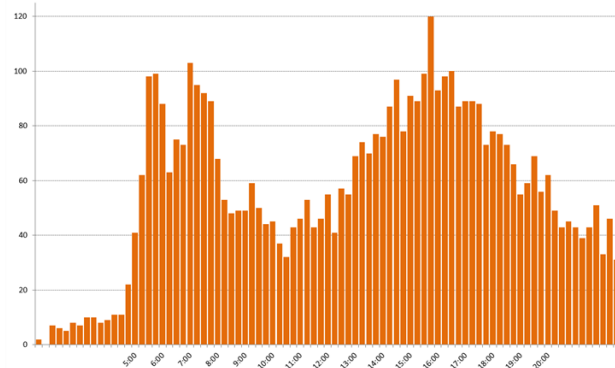
Poznámka: v tabulce jsou vyčísleny jen cesty v jednom směru, konkrétně z RUT Velké Meziříčí do Prahy a do Brna; cesty zpět uvedeny nejsou

Pramen: databáze big dat

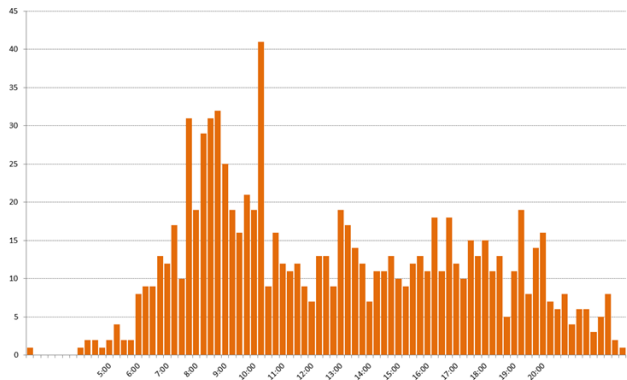
Velkomeziříčko - čas zahájení všech cest



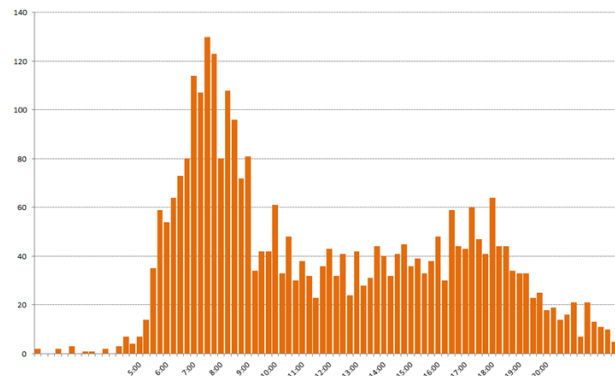
Velkomeziříčko - čas ukončení cest do Velkého Meziříčí



Velkomeziříčko - čas ukončení cest do Prahy



Velkomeziříčko - čas ukončení cest do Brna



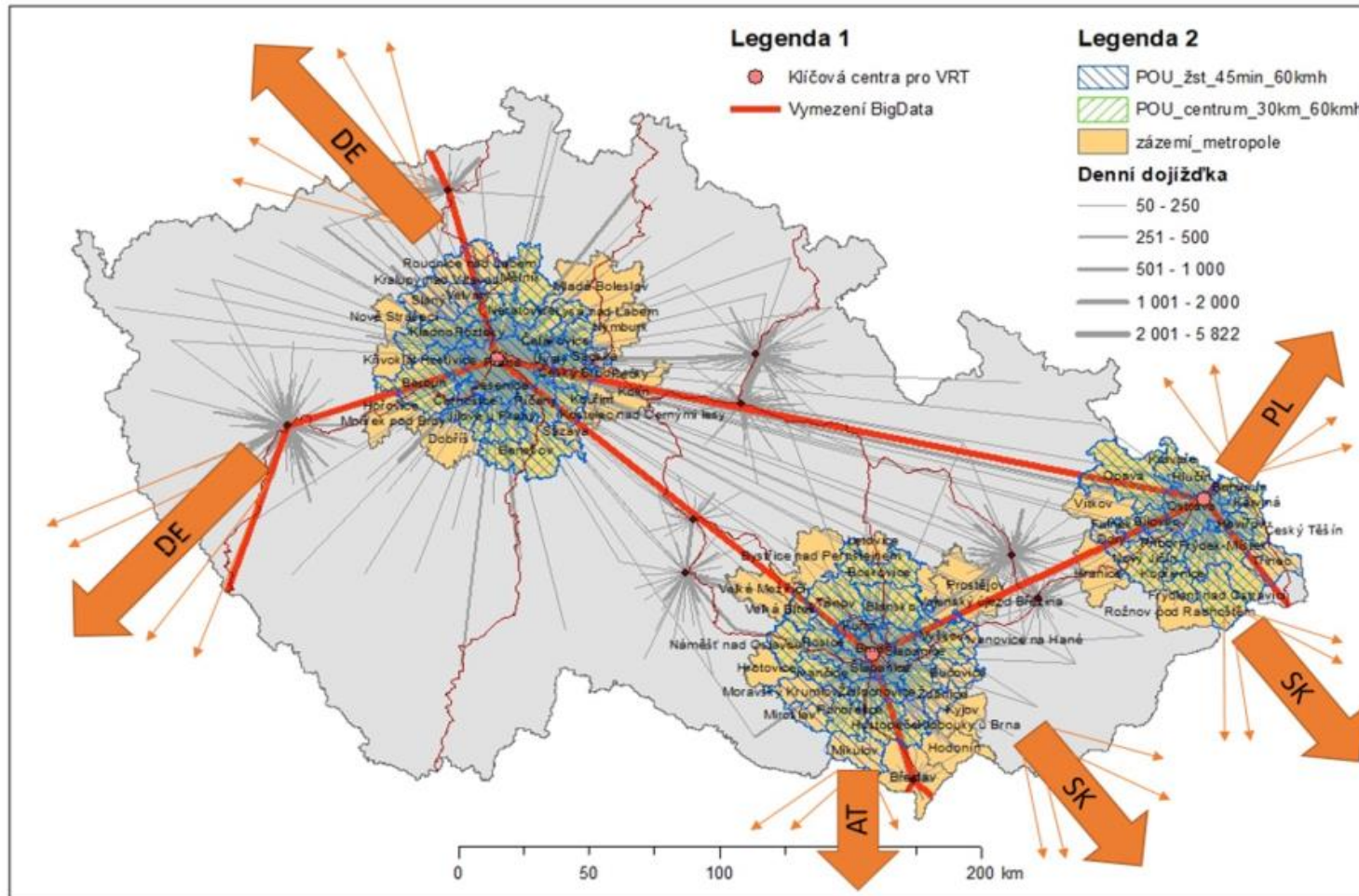
3. Big data pro Novou mobilitu – k čemu je chceme?

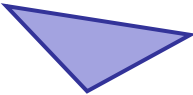
- nikoli konkurenční model (MD, SŽDC, Jaspers, ...)
- kombinace s ostatními metodami a jejich verifikace
- základní výzkum podle zaměření modulů:
 - kvantifikace poptávky po přepravě ve stopě budoucí VRT,
 - prostorové rozložení poptávky (kde cesty začínají a končí),
 - modální skladba poptávky, ...
- **spolupráce mezi moduly**
- **inovovaná metodika predikce poptávky po přepravě**
- **odhad přepravního potenciálu VRT v Česku**
- **predikce vlivu VRT na českou ekonomiku**

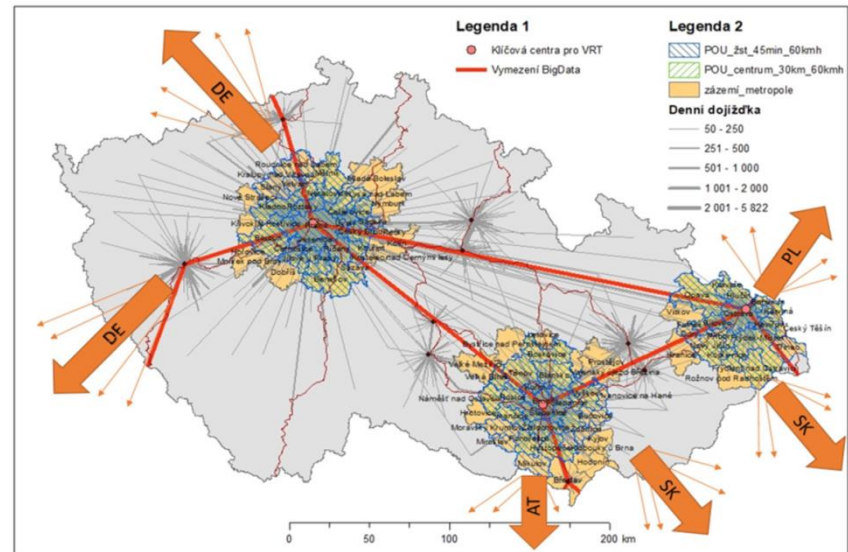
Objednávka big data – průběh přípravy

- **předmět:** měření skutečné mobility pomocí signalizačních dat mobilních operátorů (big data)
- garant nákupu: ESF MU (M. Kvizda)
- přímá návaznost na **moduly:** **A** (M. Kvizda), **B** (M. Marada), **D** (M. Kvizda), **E** (Z. Tomeš), **H** (D. Seidenglanz), **I** (D. Seidenglanz) a **K** (M. Šauer)
- **pracovní skupina:** M. Kvizda, Z. Tomeš, D. Seidenglanz, M. Šauer, V. Pařil, O. Repík a N. Voráčová
- **termíny:**
 - vyhlášení do konce XI. 2018
 - plnění do konce VI. 2019

Objednávka big data – prostorové vymezení



- **struktura dat** identická na trasách Praha – Ostrava, Praha – Brno, Brno – Ostrava
- vše, co projede po hraně  P – B – O, s identifikací počátku a konce cesty;
- vše, co přijede do Prahy z prostoru Německa;
- vše, co přijede do Brna z prostoru Rakouska a Slovenska;
- vše, co přijede do Ostravy z prostoru Polska a Slovenska.



Objednávka big data – minimální požadavky

- unikátní **identifikátor osoby** (anonymizovaný);
- identifikace **počátku a konce cesty**:
 - minimálně POÚ v rámci ČR,
 - stát v případě zahraničních cest;
- identifikace **země původu** vlastníka SIM;
- rozlišení **dopravního módu** (železnice, silnice);
- **čas zahájení cesty**;
- **čas ukončení cesty**;

- identifikace **opakování cest**;
- určení základních **demografických a socioekonomických** charakteristik cestovatelů;
- **časový rozsah sběru**: minimálně jeden typický týden s rozlišením jednotlivých dní
 - **varianta A**: kontrolní sběr za dva obdobné týdny
 - **varianta B**: jeden týden v turistické sezóně, jeden mimo sezónu
 - **varianta C**: 1 týden < období sběru ≤ 1 měsíc

- **opustili jsme:**

- monitoring nácestných stanic,
- alokaci jednotlivých cest do úrovně konkrétních spojů,
- rozlišení IAD x autobus,
- diferenciaci tras.

- **k diskusi ???**

- místo obvyklého pobytu (noc /pracovní den),
- místo práce (den),
- místo víkendového pobytu (noc / sobota),
- segment ekonomiky, kde jsou zaměstnání (např. dle NACE, pozice v zaměstnání podle vlastnictví firemního telefonu).

Objednávka big data – jak data vysoutěžíme?

- nabídka **souboru dat a metodiky**, která umožní co **nejlépe naplnit účel** výzkumu při dodržení **minimálních požadavků** a za **přijatelnou cenu**.

- dvojobálková metoda:
 - posouzení věcného naplnění a obodování,
 - s odstupem hodnocení ceny.

- opakování zakázky – opce na další nákup;

- ověřovací fáze – slyšení vítěze.

Úkoly:

- specifikovat účel výzkumu,
- stanovit parametry zvýšení kvality dat,
- určit váhy kritérií hodnocení,
- stanovit hodnotitele.

Úkol – specifikace účelu

Modul A - Analýza a modelování poptávky v osobní dopravě

- inovace metod analýzy a modelování existující poptávky po osobní dopravě;
- kvantifikace vnitrostátní přepravní poptávky mezi metropolemi a na ramenech do zahraniční (viz mapa).

Modul B - Faktory ovlivňující mobilitu

- identifikace skladby a síly definovaných dopravních proudů (viz mapa);
- identifikace účelu cesty: turistické a služební cesty.

Modul E - Analýza intermodální a intramodální konkurence v odvětví osobní dopravy

- identifikace modální skladby dopravních proudů mezi metropolemi (viz. mapa) v rozlišení použitého dopravního prostředku: IAD, autobus, vlak;
- kvantifikace cest v rozlišení na použitou službu a dopravce: ČD, RJ, FlixBus, atd.

Modul H - Výstavba systémů vysokorychlostní dopravy a související změny dopravní obsluhy a dostupnosti v prostoru ČR a střední Evropy

- identifikace počtu cest mezi metropolemi s poznáním skutečných počátků a konců jednotlivých cest (viz mapa);
- odhad změn v kvalitě dostupnosti, tzn. jak velké skupině cestujících se díky výstavbě VRT změní/zlepší dopravní dostupnost do národních či středoevropských metropolí;
- podklad k návrhu adekvátních konceptů/modelů dopravní obsluhy v soustavě zahrnutých středisek.

Modul K - Cestovní ruch

- odhad objemu a struktury turistických proudů do vymezených metropolí (Praha, Brno, Ostrava), které se uskuteční v rámci definovaných koridorů (viz mapa);
- identifikace původu a pohybu zahraničních návštěvníků, kteří se pohybují po trasách vymezených mapou.

Vynecháváme moduly D a I.

Úkol – parametry zvýšení kvality dat

- podrobná identifikace počátku/konce cest – obec, POÚ, okres, ...
- přesné stanovení času zahájení/ukončení cesty,
- dlouhý časový úsek analyzovaných dat (měsíc),
- ...

Úkol – určit váhy kritérií

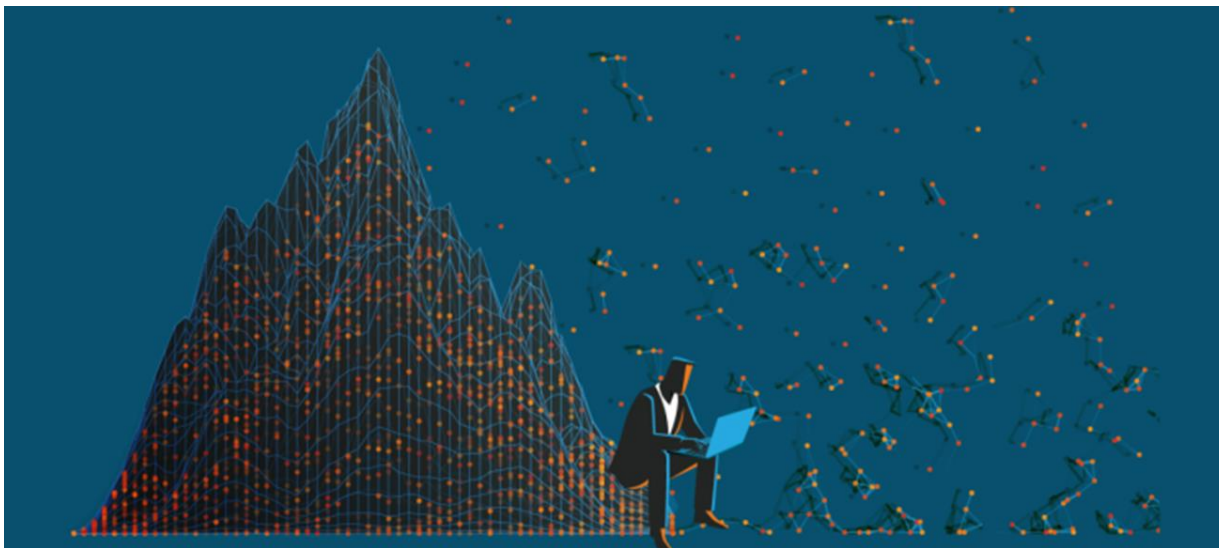
- odborná úroveň ➤ max. %
- rizika ➤ 10-20 %
- pokročilé řešení ➤ 20-30 %
- vlastnosti a schopnosti projektového manažera ➤ 5-10 %
- nabídková cena ➤ 20 %

Úkol – stanovit hodnotitele

- garanti dotčených modulů
- zástupci dopravců ČD a ZSSK

- osoby pověřené zadavatelem (MU): M. Kvizda, Z. Tomeš, D. Seidenglanz, M. Šauer
- externí (partneři): M. Marada, J. Chmelík, M. Farbiak

Děkuji za pozornost.



Nová mobilita – vysokorychlostní dopravní systémy a dopravní chování populace



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

Operační program Výzkum, vývoj a vzdělávání
výzva č. 26: Dlouhodobá mezisektorová spolupráce
CZ.02.1.01/0.0/0.0/16_026/0008430