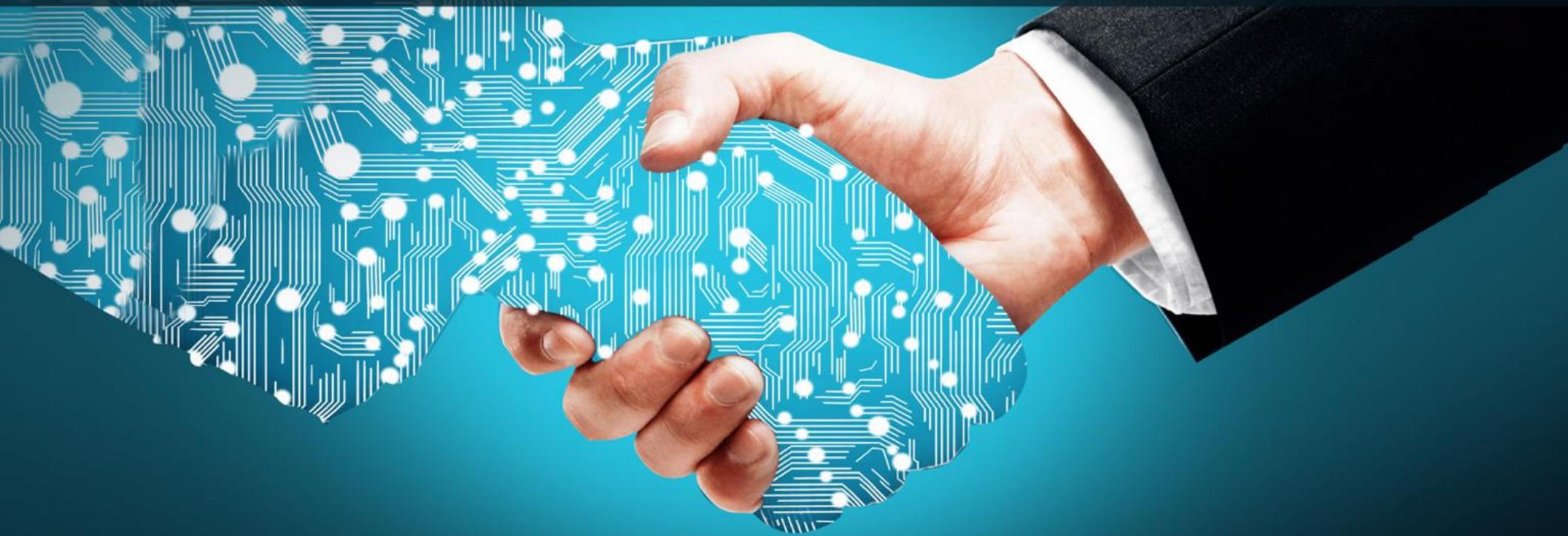


## Informatika v javni upravi 2016 "Digitalna preobrazba javne uprave - GaaS"



### Raziskovanje podatkovnih virov masivnih podatkov

## Vsebina predstavitve

---

- Viri masivnih podatkov
- Big Data implementacijska strategija
- Podatki mobilne telefonije
- Kombinirani viri podatkov
- Strganje podatkov za PDM (in cene)

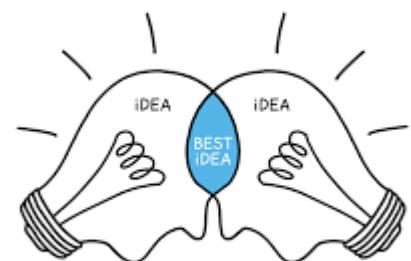
## Viri masivnih podatkov

V začetku 2016 je bila na SURS in izvedena vrsta delavnic s vsebinskimi metodologi z namenom „brainstorming-a“ idej o podatkovnih virih masivnih podatkov in statističnih domenah, kjer bi jih lahko uporabili. Na podlagi teh delavnic smo se odločili za raziskovanje naslednjih podatkovnih virov:

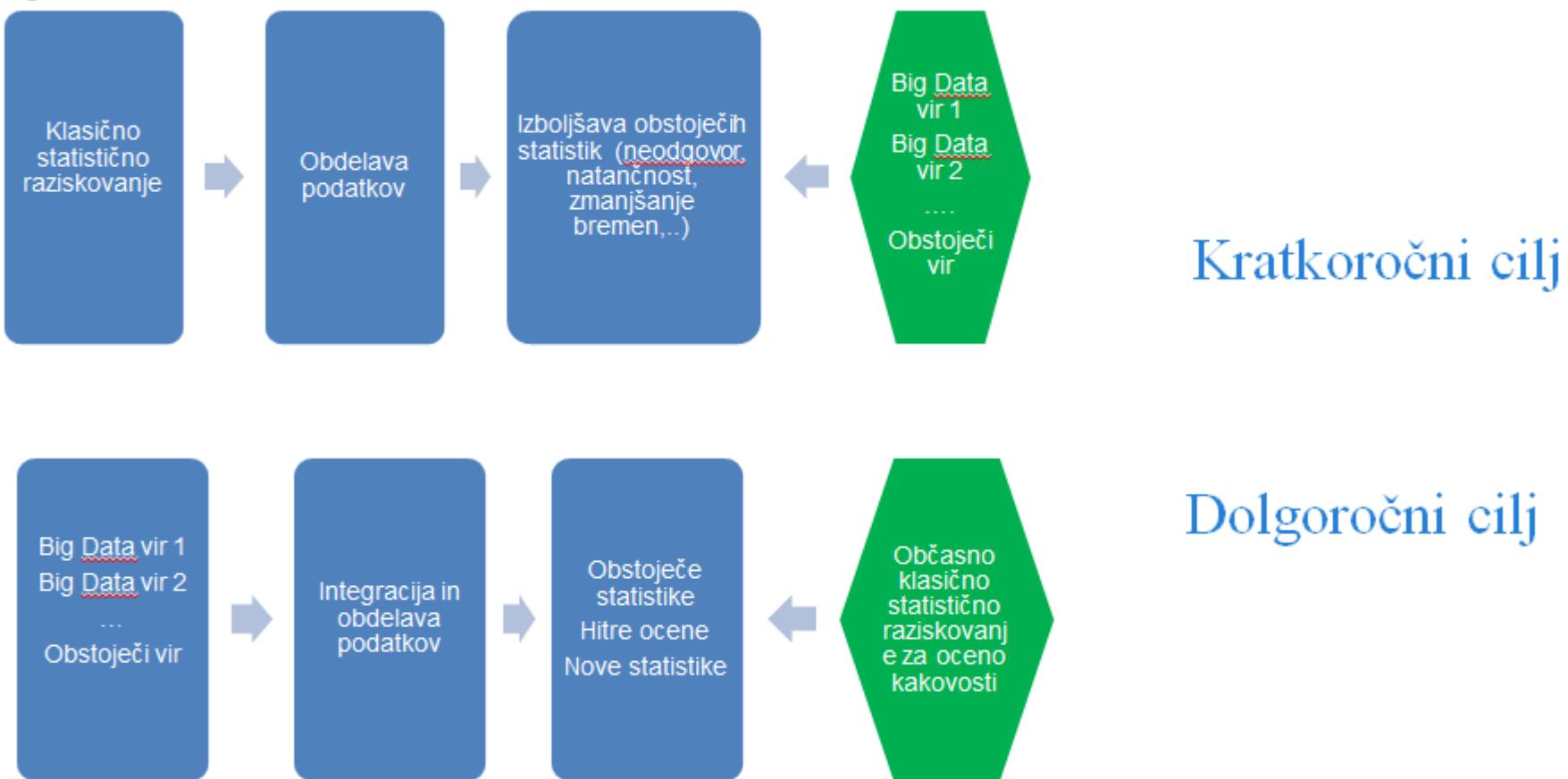
| PODATKOVNI VIR (I)   | NAMEN  |
|--|--|
| Podatki mobilne telefonije   | Statistike o mobilnosti prebivalstva                     |
| Kombinirani viri (Big Data, raziskovanja, administrativni in drugi viri) | Zgodnji ekonomski kazalniki                              |
| Strgani podatki spletnih strani  | Statistike prostih delavnih mest (PDM) in statistike cen |

Jasen cilj:

- Čimprejšnja uvedba BD v redno statistično produkcijo.

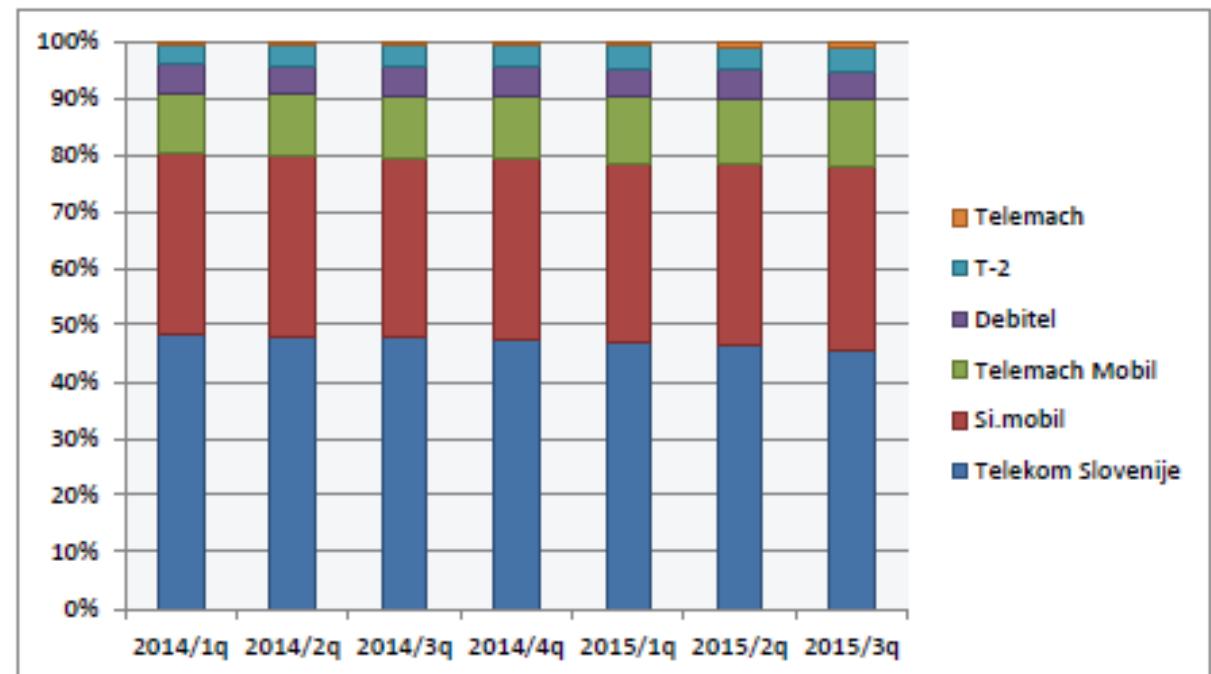


## Big Data implementacijska strategija na SURS



## Mobilna telefonija v Sloveniji

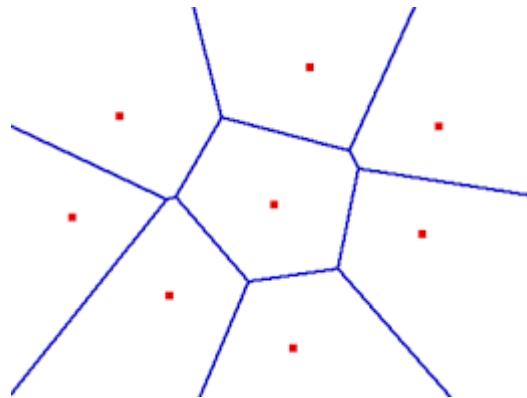
- 5 operaterji omrežja
- 1 ponudniki storitev
- 1 ponudniki paketov
- prvih pet ima podatke
- vsi so lahko pomembni



Source: AKOS

## Podatki

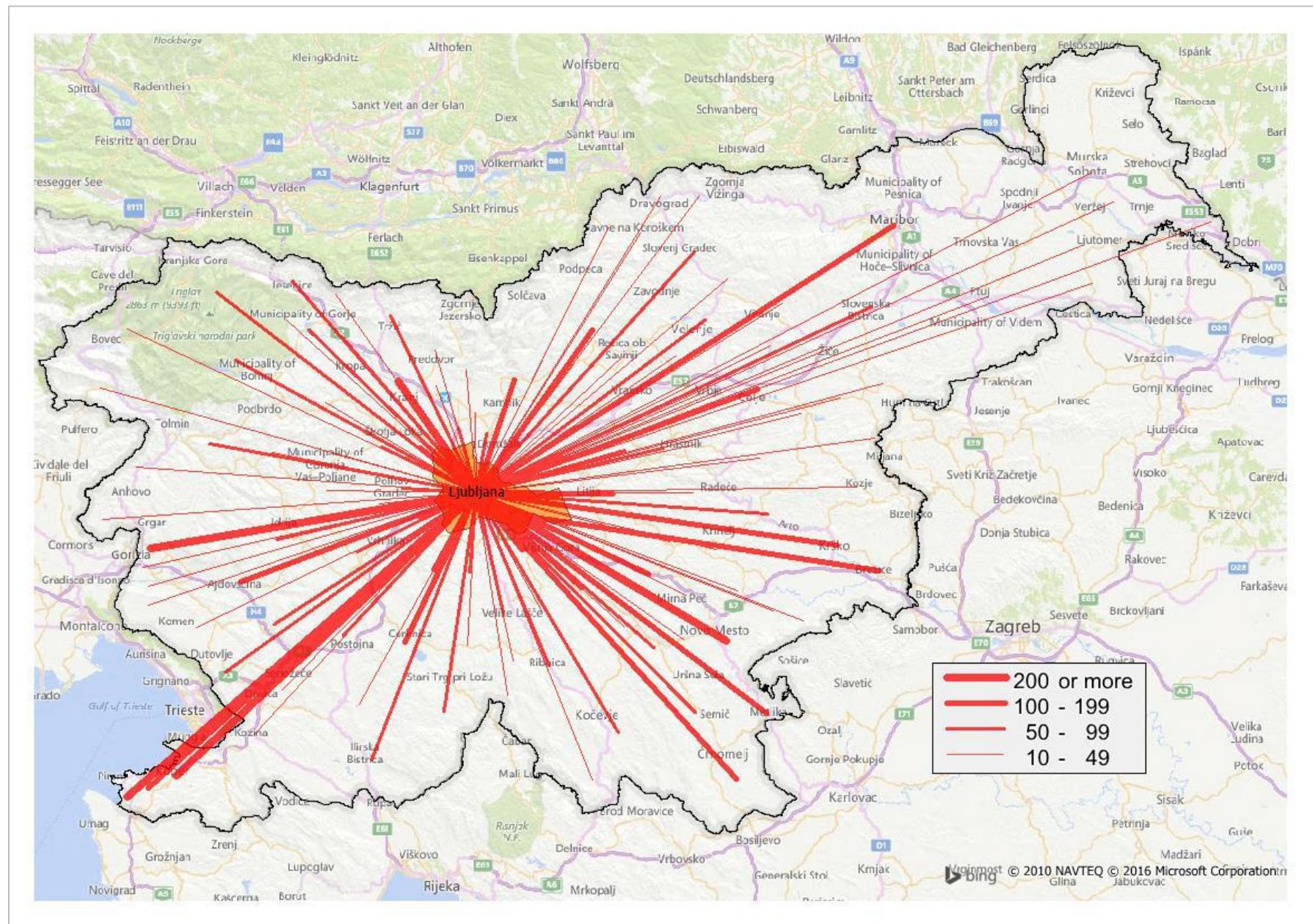
- 10 % uporabnikov opravi 75 % ali več dogodkov na eni BP
- 40 % opravi 50 % ali več dogodkov na eni BP
- 47 % opravi 75 % ali več dogodkov v eni „Voronoi občini“
- 82 % opravi 50 % ali več dogodkov v eni „Voronoi občini“



**a**  
110011101010011  
101010010110110  
1011INFORMATIKA01  
101V LJAVNI0101001010  
011011010UPRAVI02016

# Informatika v javni upravi: "Digitalna preobrazba javne uprave - GaaS"

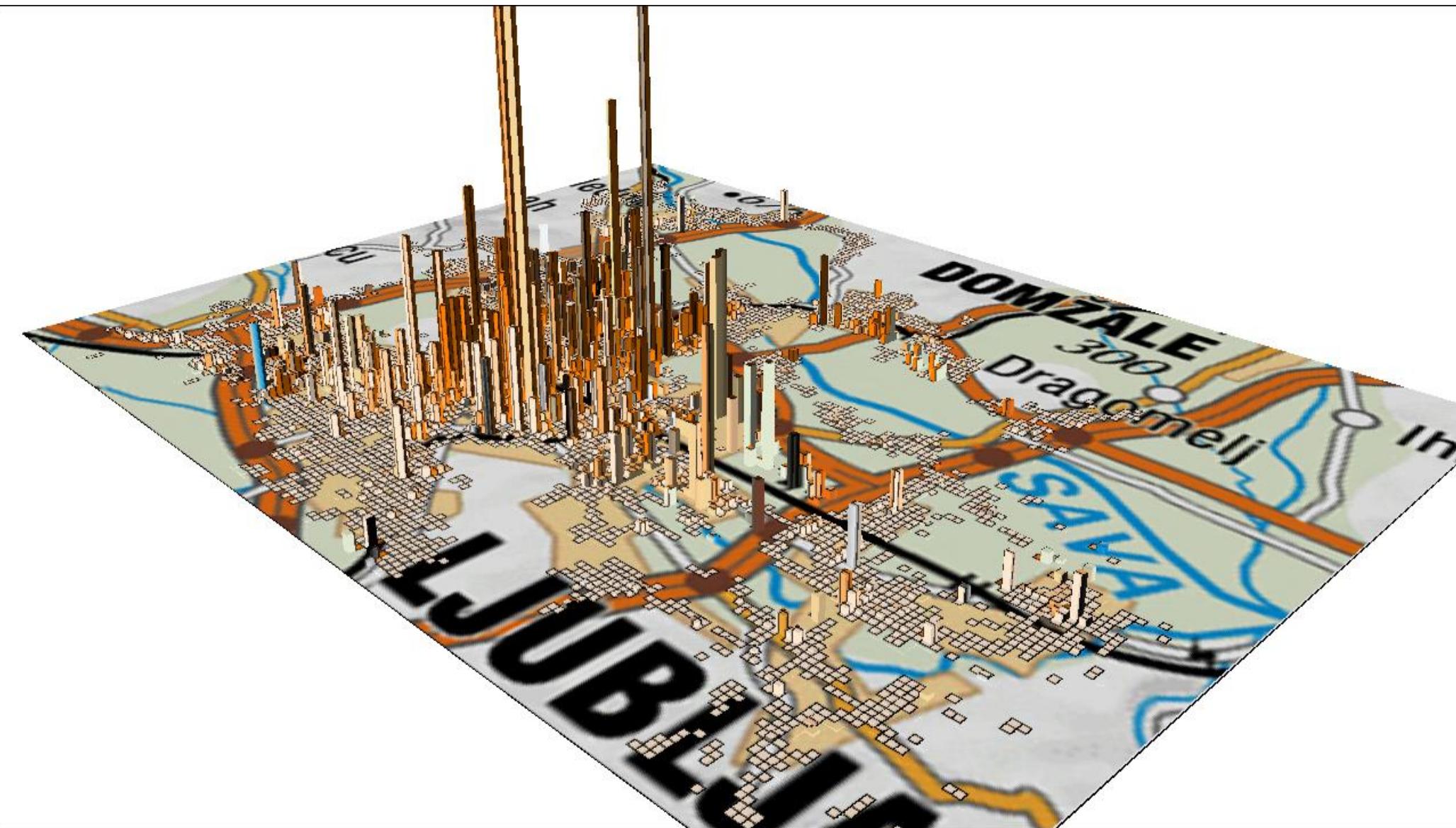
Kongresni center Brdo pri Kranju | 5. - 6. december 2016



**a**  
11001101010011  
1010100101101110  
1011INFORMATIKA01  
101V LJAVNI0101001010  
011011010UPRAVI02016

# Informatika v javni upravi: "Digitalna preobrazba javne uprave - GaaS"

Kongresni center Brdo pri Kranju | 5. - 6. december 2016

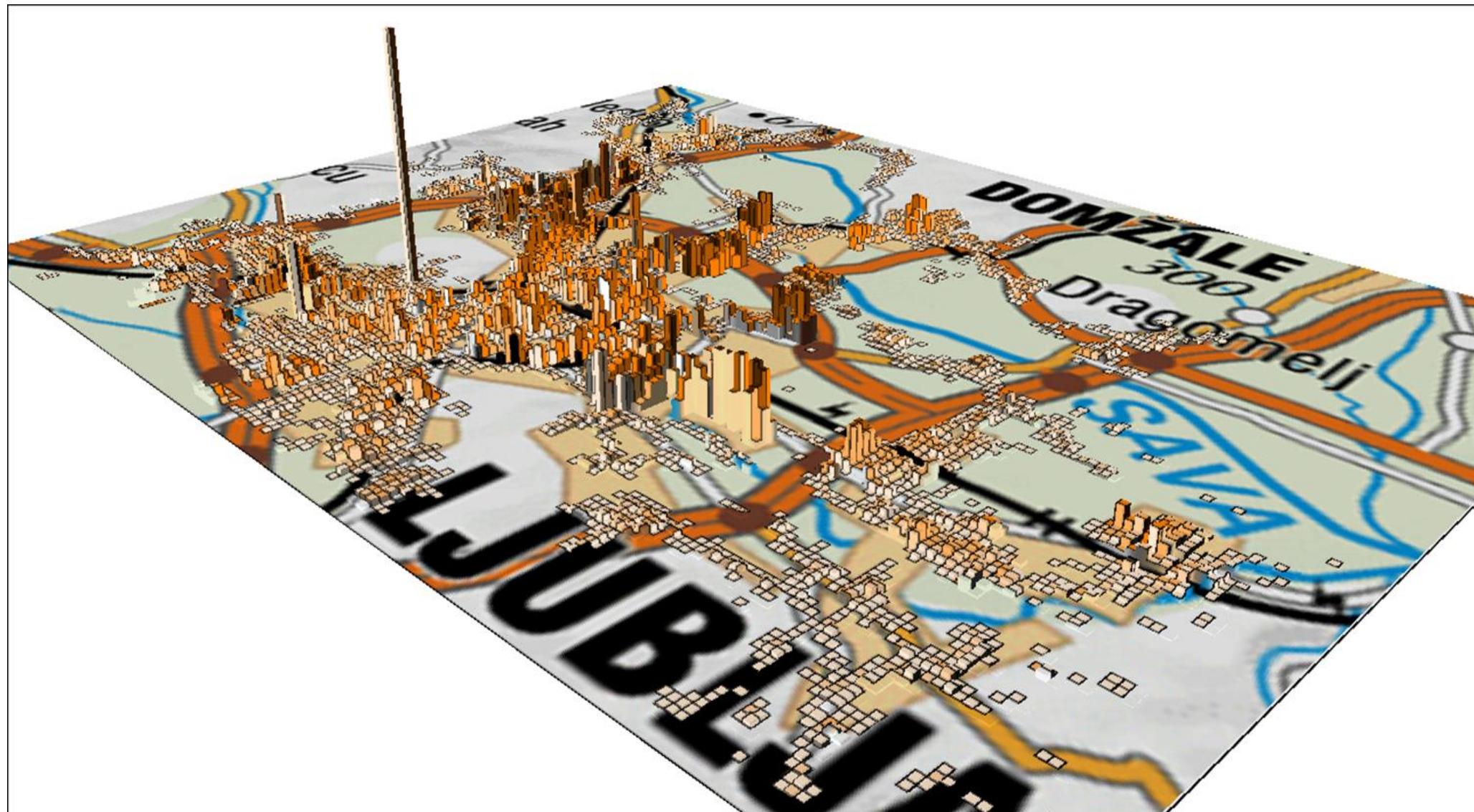




11001101010011  
101010010110110  
1011INFORMATIKA01  
101V LJAVNI0101001010  
011011010UPRAVI02016

# Informatika v javni upravi: "Digitalna preobrazba javne uprave - GaaS"

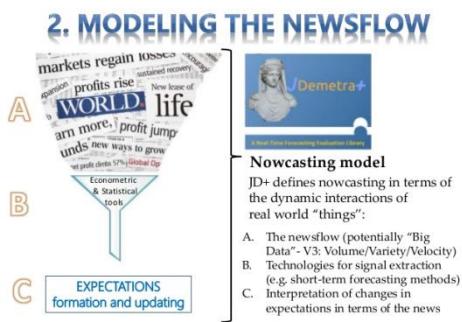
Kongresni center Brdo pri Kranju | 5. - 6. december 2016



## Zgodnji ekonomski kazalniki

Namen:

- raziskati in kombinirati vire podatkov za namene ocenjevanja zgodnjih ekonomskih kazalnikov
- Raziskati nowcasting metode



| Big Data                    | Registrski viri | SURS Raziskovanja      |
|-----------------------------|-----------------|------------------------|
| Prosta delovna mesta        | SRDAP           | Kratkoročne statistike |
| Prometni podatki            | DDV             | Poslovne tendence      |
| Podatki z družbenih omrežij | Plače           | Mnenje potrošnikov     |
| Bančne transakcije          | SPRS            | Zunanja trgovina       |
| Skenirani podatki cen       |                 | Poslovanje podjetij    |
| „News feed“                 |                 | Podatki o plačah       |
| Podatki mobilnih telefonij  |                 | Gradbena dovoljenja    |

- Raziskati ekonomske kazalnike, katerih ocene bi lahko pripravili bolj zgodaj (BDP, Kazalnik gospodarskega stanja v Sloveniji,...)

## Zgodnji ekonomski kazalniki (testni rezultati)

### Model:

#### 1. Metoda glavnih komponent (PCA)

- zmanjšanje dimenzije prostora podatkov
- izberemo nekaj prvih glavnih komponent

#### 2. Linearna regresija

- Y (spr., ki jo želimo oceniti): npr. prihodek od prodaje
- X1, ..., Xn (spr., s katerimi želimo oceniti): izbrane glavne komponente

**Podatki:** prihodek od prodaje za 973 podjetij IND-PN/M za vsak mesec v obdobju 2008 – 2015

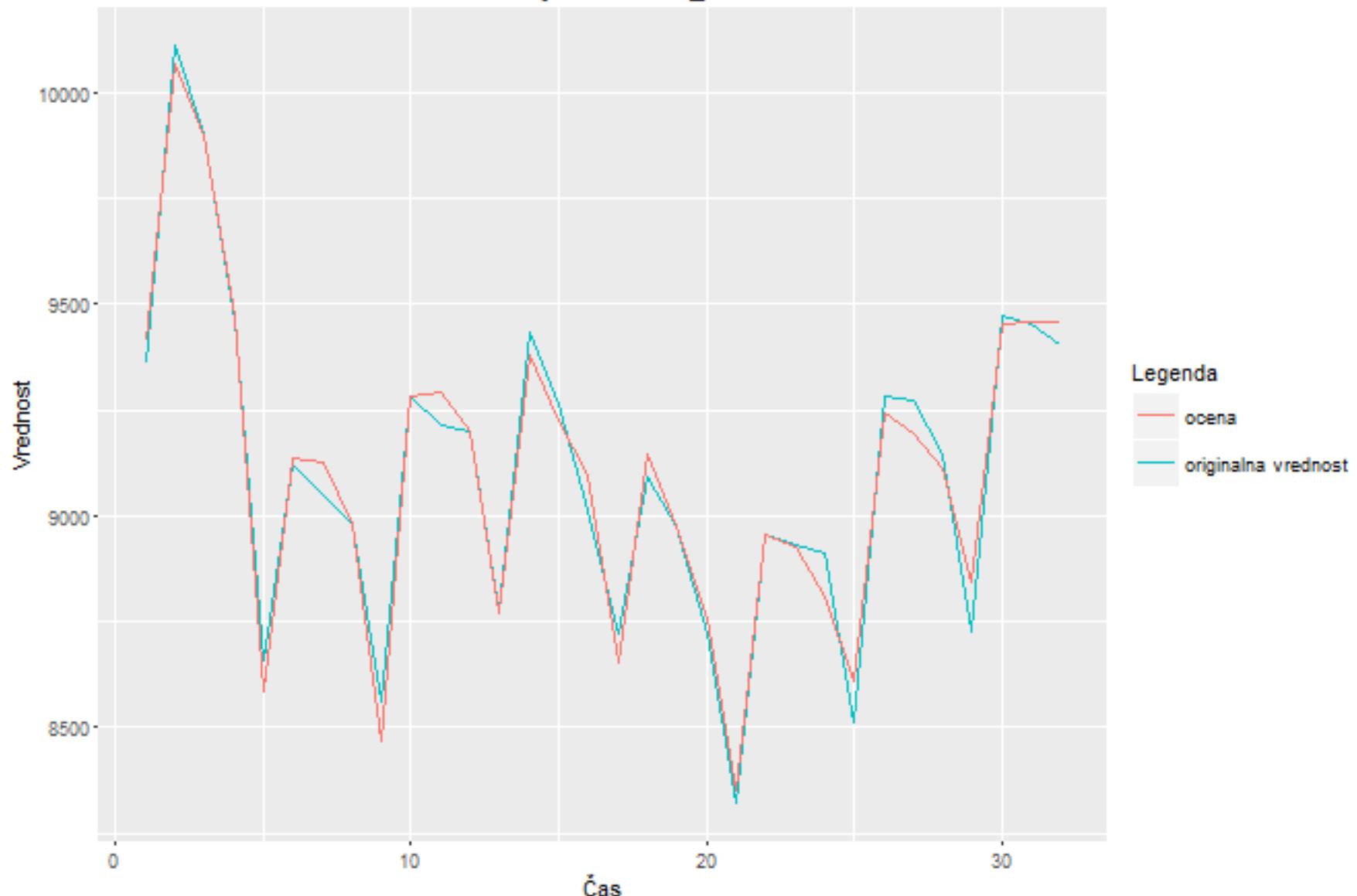
| obdobje | P001 | P002 | ... | P973  |
|---------|------|------|-----|-------|
| 2008M01 | 3526 | 214  | ... | 66519 |
| 2008M02 | 4252 | 332  | ... | 36012 |
| 2008M03 | 4111 | 411  | ... | 52447 |
| ...     | ...  | ...  | ... | ...   |
| 2015M12 | 5241 | 412  | ... | 71025 |

**Statistika, ki jo želimo oceniti:** Cetrtletni bruto domači prihodek v stalnih cenah (BDP\_SC)

### Programska oprema:

- priprava podatkov: SAS
- model in analiza: RStudio

Primerjava za BDP\_SC



Legenda

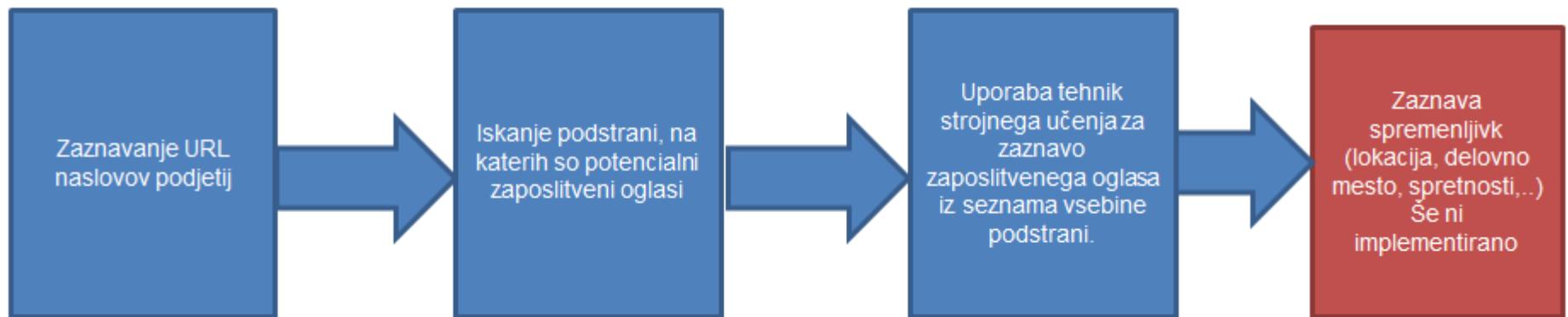
- ocena
- originalna vrednost

# Metodologija strganja spletnih podatkov portalov o PDM

Naloge:



## Metodologija strganja spletnih podatkov podjetij



## Pokritost: vzorec vs. postrgani&administrativni podatki

|                 | Poročani podatki | Postrgani&admin podatki | Zaposlitveni portali | Spletne strani podjetij |
|-----------------|------------------|-------------------------|----------------------|-------------------------|
| Število oglasov | 4321             | 2321                    | 1073                 | 262                     |
| Odstotek        | 100 %            | 54 %                    | 25 %                 | 6 %                     |

| Št. zaposlenih | Vprašalnik | BD viri | Odstotek |
|----------------|------------|---------|----------|
| 1              | 67         | 16      | 24 %     |
| 1 – 9          | 470        | 173     | 37 %     |
| 10 – 49        | 923        | 362     | 39 %     |
| 50 – 249       | 1681       | 744     | 44 %     |
| 250 +          | 1119       | 782     | 70 %     |

***Hvala za vašo pozornost !***

*Vprašanja?*

*Pripombe?*

*Predlogi?*