



REPUBLIKA SLOVENIJA



STATISTIČNI URAD REPUBLIKE SLOVENIJE



www.stat.si

SLOVENSKA
DRŽAVNA
STATISTIKA

ZAUPANJA VREDNA
IN UPORABNIŠKO
USMERJENA

18. konferenca
Dnevi slovenske informatike

Registrski popis 2011 - informacijska podpora statističnim procesom

Mag. Danilo Dolenc
Milojka Krek
Rudi Seljak

Statistični urad Republike Slovenije

Portorož, 20. 4. 2011

Registrski popis 2011

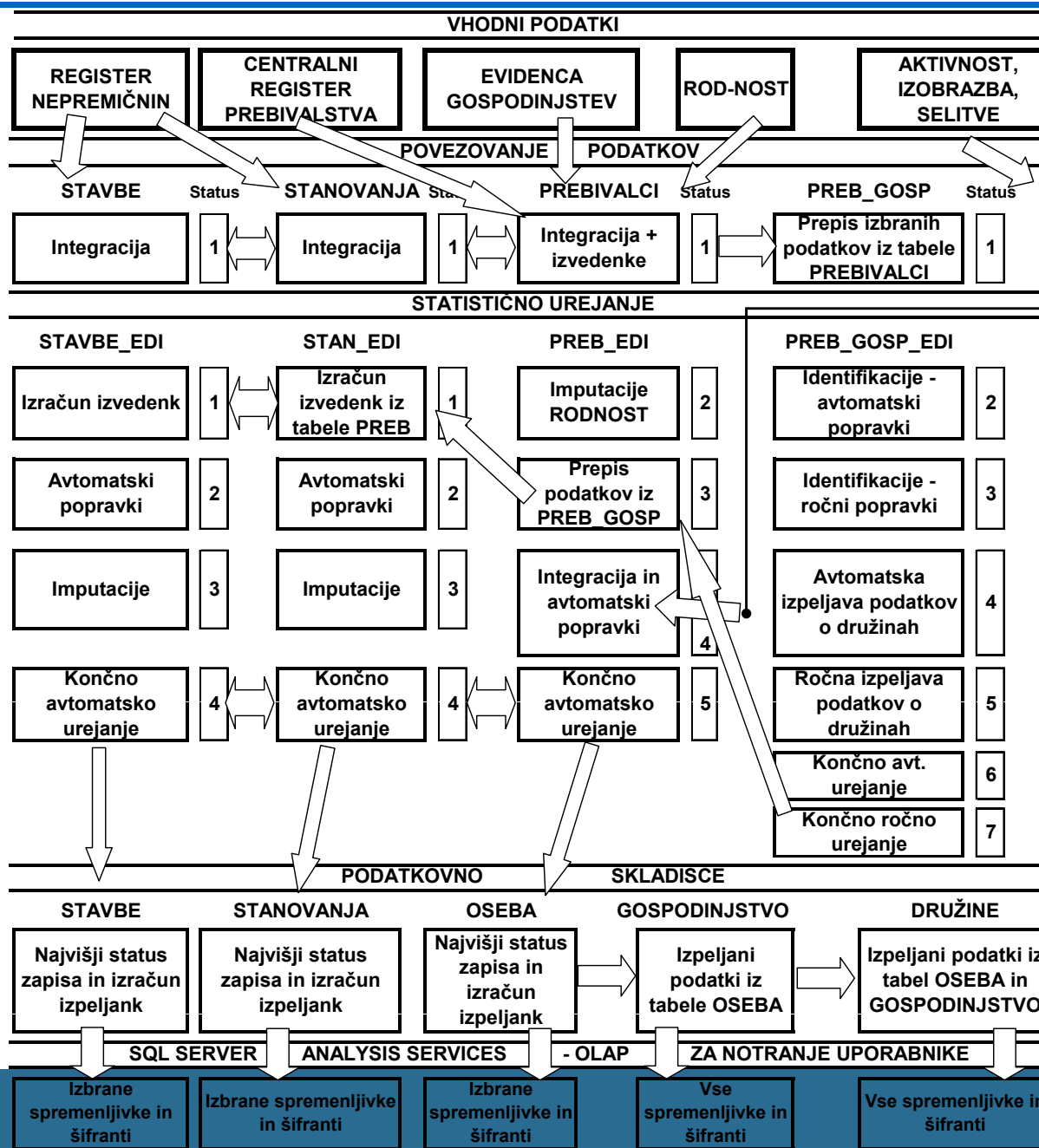
- Prvi popis brez terenskega zbiranja
- Vsi podatki pridobljeni iz
 - Administrativnih registrov
 - Različnih zunanjih baz podatkov
 - Rednih statističnih raziskovanj
- Projekt
 - Lastni resursi
 - Majhna skupina (6 oseb)
- Glavna prednost – majhni stroški
 - Proračunski prihranek vsaj 14 mio EUR

Pogoji za registrski popis

- Ustrezna zakonodaja
 - 32. in 33. člen Zakona o državni statistiki
 - Uredba EP in Sveta EU – prvič do sedaj
- Vzpostavitev novih / dopolnitev obstoječih administrativnih virov
 - CRP - številka stanovanja
 - E-gospodinjstva
 - Register nepremičnin (REN)
- Vsebina virov
 - Možnost pokritja obveznih vsebin Uredbe

Bistvo registrskega popisa

- Povezovanje podatkov
 - O prebivalcih iz različnih virov
 - Prebivalcev z gospodinjstvi
 - Gospodinjstev s stanovanji
 - Povezovalni znaki (identifikatorji)
 - **EMŠO / SID**
 - **Naslov s številko stanovanj**
 - **Številka gospodinjstva na naslovu**



Statistični proces - faze

- Priprava vhodnih datotek
 - Skupaj nekaj več kot 30 iz 20 virov
- Integracija vhodnih podatkov
 - Uskladitev identifikacij o osebah, gospodinjstvih in stanovanjih
- Statistično urejanje
- Izgradnja podatkovnega skladišča
- Dostop do podatkov (OLAP)

Posebnosti stat. procesa (1)

- Postopnost izvajanja procesov
 - Vhodni podatki niso na voljo istočasno
 - Trije veliki sklopi procesa, ki jim sledi objava
 - Podatki se po posamezni fazi ne smejo več spremeniti
- Ločen proces izpeljave gospodinjestev
 - Kompleksnost obdelav
 - Razlike v konceptih
 - Samo v tem podprocesu vmesnik za ročne popravke

Posebnosti stat. procesa (2)

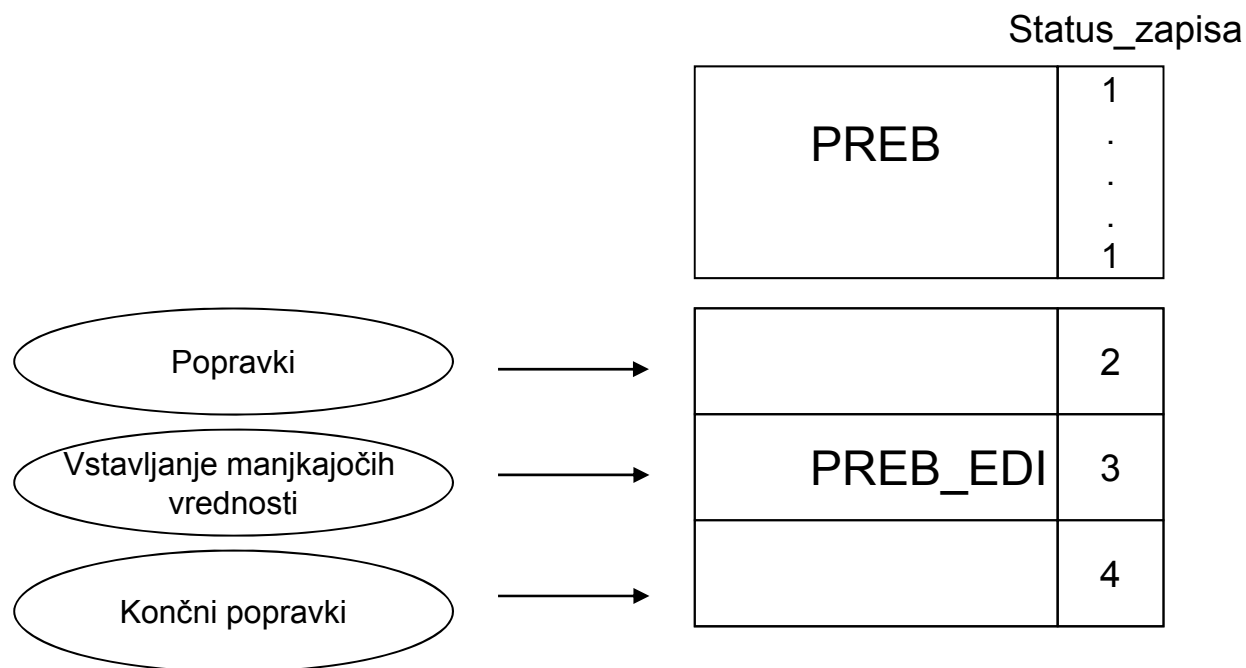
- Prevzem podatkov iz več virov
 - Pomožne tabele za izpeljavo podatkov
 - Izobrazba npr. iz 11 različnih virov
 - Tudi metodološko zahtevno
- Uporaba različnih statističnih metod
 - Kompleksnost obdelav
 - Avtomatski popravki (90 %)
 - Imputacije (9,5 %)
 - Ročni popravki (0,5 %) – izboljšanje kakovosti podatkov o gospodinjstvih

Posebnosti stat. procesa (3)

- Sledljivost spremembam
 - Vsak korak v procesu lahko poljubno mnogokrat ponavljamo. Ob vsaki ponovitvi pri nespremenjenih vhodnih podatkih ter nespremenjenih postopkih dobimo vedno isti rezultat.
- Ponovljivost postopkov
 - Popravki in dopolnitve ne “prekrijejo” starih vrednosti, ampak se vedno tvori nova verzija podatka.

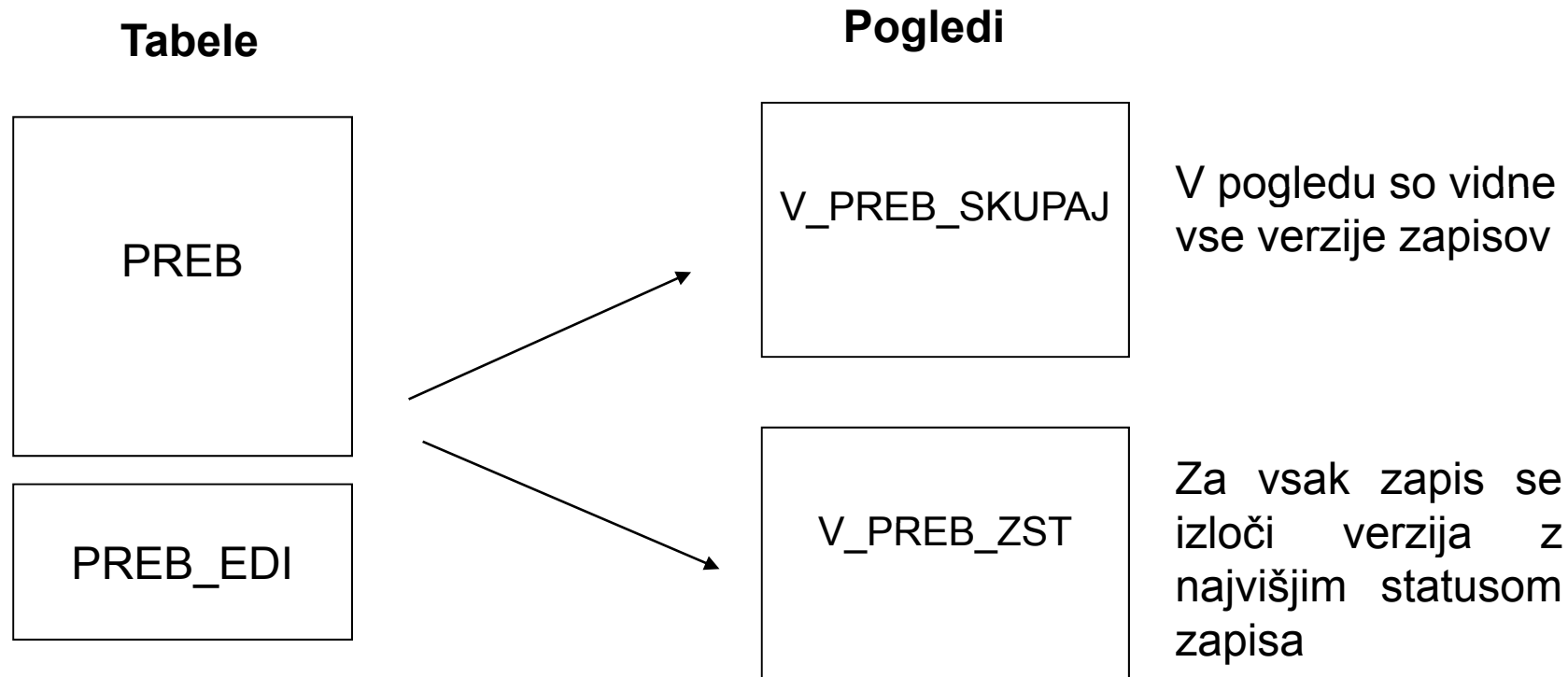
Osnovna organizacija baze

- Izhodiščni podatki so v tabeli, katere podatki se ne spreminjajo.
- Vse zapise, pri katerih smo spremenili vsaj en podatek, zapišemo v posebno tabelo ("EDI tabela")
- Verzije zapisov, nastale v različnih fazah procesa, označimo z vrednostjo statusa zapisa.



Tabele in pogledi

- Do vseh verzij podatkov lahko dostopamo s pomočjo pogledov.



Evidenca sprememb podatkov (1)

- Vsaka tabela s podatki ima "zrcalno" tabelo s statusi spremenljivk

Tabela s podatki

ID	V1	V2	...	Vn

Tabela s statusi spremenljivk

ID	V1_s	V2_s	...	Vn_s

Evidenca sprememb podatkov (2)

- Vedno, ko se spremeni vrednost podatka, se spremeni tudi vrednost pripadajočega statusa

Izhodiščni podatki

ID	V1	V2
111	300	01
112	2	Null

ID	V1_s	V2_s
111	21.11	31.11
112	21.12	Null

Podatki po popravkih

ID	V1	V2
111	3	01
112	2	Null

ID	V1_s	V2_s
111	22.13	31.11
112	21.12	Null

Podatki po vstavljanju manjkajočih vrednosti

ID	V1	V2
111	3	01
112	2	04

ID	V1_s	V2_s
111	22.13	31.11
112	21.12	41.14

Izvajanje postopkov

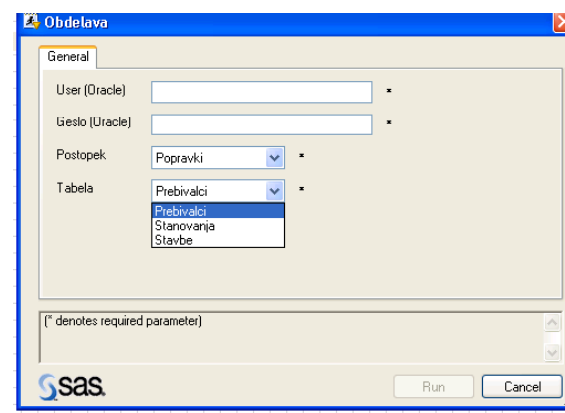
- V grobem lahko celoten proces razdelimo v tri dele:
 - Integracija podatkov; kompleksnejši popravki, pri katerih je potrebno naenkrat procesirati več zapisov (npr. vse zapise v določenem gospodinjstvu). Uporabijo se na mero pripravljeni programi.
 - Ročni popravki na individualnih enotah, ki jih vnašamo preko posebnega grafičnega vmesnika.
 - Avtomatski popravki, kjer so pravila postavljena izključno glede na podatke ene enote ter vstavljanje manjkajočih vrednosti. Uporabi se splošna aplikacija, ki se uporablja tudi za potrebe drugih raziskovanj (SAS). Aplikacija deluje na t.i. "metadata driven" pristopu.

Splošna aplikacije za popravke

- Navodila za izvajanje so podana v posebni bazi (MS Access). Za potrebe izvajanja ni potrebno posegati v osnovno programsko kodo.

Tabela	Spremenljivka	Pogoj	Nova_vrednost
PREB	V1	if V1 >100	round(V1/10)

- Postopek "poženemo" preko SAS – grafičnega vmesnika



Zaključne misli

- Objava podatkov že 29. aprila
- Slovenija se pridružuje majhni skupini držav, ki so registrski popis sposobne izvesti
- Kakovost vhodnih podatkov glavni predpogoj za kakovost končnih podatkov
- Pogostejša izvedba v prihodnje
 - Nekdanji popis, ki se je izvajal na 10 let, postaja povsem običajno statistično raziskovanje