

MIKROPODATKOVNI VIRI KOT PODLAGA ZA RAZISKOVANJE NA PODROČJU DEMOGRAFIJE V EVROPI IZZIVI IN MOŽNOSTI

Sebastian Kočar, Janez Štebe

Arhiv družboslovnih podatkov

Fakulteta za družbene vede (Univerza v Ljubljani)

Kardeljeva ploščad 5, 1000 Ljubljana, Slovenija

Tel: +386 1 5805 292

e-mail: sebastian.kocar@fdv.uni-lj.si

janez.stebe@fdv.uni-lj.si

POVZETEK

Olajšan dostop do mikropodatkov v zadnjih letih je odprl nove možnosti za kakovostnejše raziskovanje na področju demografije. Raziskovalcem so na voljo različni mikropodatkovni viri, ki se med seboj ločijo po metodi zbiranja podatkov (popisi, registrski podatki, anketni podatki) in temu ustrezni vsebinski pestrosti, ter po dostopnosti odvisno od tipa uporabnika (laični raziskovalci, študenti, doktorski študenti, registrirani raziskovalci). Uporabniki imajo različne pogoje dostopa, odvisno od zaščite možnosti razkritja posameznika ali organizacije. Dostopnost mikropodatkov na Statističnem uradu RS se nadgrajuje tudi v okviru sodelovanja z drugimi organizacijami in pri mednarodnih projektih. Eden izmed njih je projekt sodelovanja statističnih uradov in podatkovnih arhivov *Data without Boundaries*. Poudarek prispevka je na prikazu možnosti pooblaščenega dostopa do nezaščitenih mikropodatkov uradne statistike, uporabnih za različna področja mednarodne primerjalne demografije.

1 UVOD

Razne mednarodne ter nacionalne iniciative v zadnjih desetih letih so precej izboljšale dostopnost mikropodatkov, ki jih posedujejo statistični uradi in druge javne ustanove. Študenti, registrirani raziskovalci in drugi uporabniki imajo tako precej manj težav pri empiričnem raziskovanju demografskih značilnosti populacije, saj mikropodatki vsebujejo zelo podrobne klasifikacije kategorij spremenljivk in nudijo dodatne možnosti oblikovanja pregledov njihovih povezanosti. Na voljo je tako čedalje več virov za raziskovanje, od prosto dostopnih mikropodatkov do mikropodatkov, dostopnih le pod določenimi pogoji in le določenim tipom uporabnikov. O mikropodatkih govorimo, kadar imamo namesto vnaprej pripravljenih tabel z agregiranimi statistikami možnost dostopa in analize na ravni izvornih enot zbiranja podatkov, npr. posameznikov v vzorcu ali populaciji.

2 DEMOGRAFSKA PODROČJA IN MIKROPODATKI

Raziskovalcem so na razpolago mikropodatki, ki so uporabni za raziskovanje demografskih vprašanj na različnih vsebinskih področjih. Navajamo posamezne primere mikropodatkovnih virov, ki se jim bomo podrobneje posvetili v naslednjih poglavjih:

- *zgodovinska demografija* – popisi iz različnih časovnih obdobj, npr. zgodovinski popisi iz različnih držav na spletni strani IPUMS International [1]
- *socialna demografija* – npr. Evropska družboslovna raziskava [2]
- *ekonomska demografija* – npr. Anketa o delovni sili Statističnega urada Republike Slovenije [3]
- *politična demografija* – npr. sklop raziskav Politbarometer Centra za raziskovanje javnega mnenja in množičnih komunikacij [4]
- *teološka demografija* – npr. Popis 2002, ki ga je terensko izvedel Statistični urad Republike Slovenije [5]
- *jezikovna demografija* – npr. Popis 2002
- *biodemografija* – npr. mikropodatki Inštituta za varovanje zdravja Republike Slovenije
- *medicinska demografija* – npr. mikropodatki Inštituta za varovanje zdravja Republike Slovenije

Zdravstveni statistični podatki sicer veljajo za najbolj zaščitene osebne podatke, zato je dostop do njih izredno omejen, med drugim tudi zaposlenim znotraj inštitucij, katere jih zbirajo, analizirajo ali posedujejo. Dolgoročno pa ureditev dostopa do tovrstnih podatkov nedvomno predstavlja izziv za celotno raziskovalno skupnost.

3 VIRI DEMOGRAFSKIH MIKROPODATKOV

Ko govorimo o glavnih virih demografskih podatkov, je treba izpostaviti dve stvari. V zadnjih dvajsetih in več letih je »internetna revolucija« raziskovalcem bistveno izboljšala dostopnost določenih mikropodatkov. Druge pomembne

spremembe pa so se zgodile v zadnjih letih pri odpiranju inštitucij navzven za dostop raziskovalcev. Na voljo je tako kar nekaj virov podatkov, ki bi bili lahko podlaga kompleksnim demografskim analizam:

- *nacionalni popisi* kot eden najpomembnejših virov demografskih raziskovanj;
- *vzorci popisov*, ki sicer vsebujejo manjše število enot, vendar še vedno dovolj, da je mogoče ustrezno statistično sklepanje na celotno populacijo;
- *registrski podatki*, ki jih v okviru svojega programa raziskav uporabljajo nacionalni statistični uradi;
- *anketni podatki* uradnih ter akademskih raziskovanj, ki dosegajo visoka merila kakovosti, kot so, ustrezna velikost vzorca in prisotnost večjega števila demografskih in drugih spremenljivk;
- *drugi viri*...

Do nekaterih izmed mikropodatkov lahko dostopa kdorkoli, za dostop do drugih navedenih virov mikropodatkov pa gre registrirani raziskovalec skozi akreditacijski proces pri skrbniku mikropodatkov. Vire mikropodatkov je zato smiselno razdeliti glede na dostopnost, s čimer se ukvarjamo v naslednjem poglavju.

4 DEMOGRAFSKI PODATKI GLEDE NA DOSTOPNOST/TIP UPORABNIKA

Dostopne evropske mikropodatke na področju demografije lahko nadalje glede na navedene kriterije razdelimo v tri skupine:

- *Mikropodatkovni viri v prostem dostopu preko spleta* (namenjeni manj zahtevnim uporabnikom – študentom).
Na spletni strani je večinoma potrebna registracija, vendar je dostop običajno dodeljen brez omejitev.
- *Nezaščiteni mikropodatkovni viri Statističnega urada RS* (namenjeni registriranim raziskovalcem).
Do njih lahko dostopajo slovenski raziskovalci, vpisani v SICRIS-u, ter tuji raziskovalci. Pred dostopom je potrebna odobritev znotraj Statističnega urada, in sicer s strani Odbora za varstvo podatkov.
- *Evropski mikropodatkovni viri, dostopni tudi v sklopu projekta DwB* [6].
Podatki, ki so v sklopu omenjenega projekta testno dostopni preko posebnega protokola, so namenjeni evropskim raziskovalcem, vključno z doktorskimi študenti.

4.1 Viri demografskih mikropodatkov za dostop preko spleta

Za mikropodatkovne vire, dostopne preko spleta, na splošno velja, da lažje kot so dostopni, bolj so okrnjeni – tako glede števila in podrobnih vrednosti spremenljivk, kakor včasih tudi glede velikosti vzorca. Primerni so zlasti za študente, tako dodiplomske kot tudi podiplomske, ter za ostale manj zahtevne uporabnike. Izpostavili bi lahko naslednje vire mikropodatkov:

- *Arhiv družboslovnih podatkov* [7], ki razpolaga z več kot 500 raziskavami, mnoge izmed njih se lahko uporabi v demografskem raziskovanju, npr. Popis 2002 [8] – 5 % vzorec, baza podatkov vsebuje vse pomembne demografske spremenljivke;
- *IPUMS International* [9], ki razpolaga z mikropodatki popisov iz 68 držav, praviloma s 5 ali 10 % vzorci (do 25 % - nemški popis iz leta 1971);
- *IECM* [9], ki razpolaga z anonimiziranimi mikropodatki 55 evropskih popisov 19 držav;
- *Evropska družboslovna raziskava (European Social Survey)* in drugi viri podatkov iz mednarodnih družboslovnih anket, ki obsegajo vzorce več 10 tisoč enot in ob uporabi standardnih demografskih spremenljivk omogočajo kumulativno združevanje anketnih podatkov čez več valov izvedbe.

4.2 Viri demografskih mikropodatkov na Statističnem uradu RS

Na podlagi mednarodnih priporočil in ob vzpostavitvi postopkov za registracijo se je v zadnjih letih izboljšala dostopnost mikropodatkov na Statističnem uradu Republike Slovenije (SURS). Raziskovalci dostopajo do podatkov v varni sobi in preko oddaljenega dostopa, pri čemer je potrebna predhodna odobritev Odbora za varstvo podatkov, svetovalnega telesa generalne direktorice [10]. Statistični urad pri zagotavljanju čim boljših pogojev za raziskovalce sodeluje pri nekaterih mednarodnih projektih v partnerstvu z družboslovnimi arhivi podatkov po Evropi. Podobno sodelovanje se razvija tudi doma. Arhiv družboslovnih podatkov (ADP) sodeluje pri promociji uporabe in pripravi mikropodatkov ter metapodatkov (dokumentacije) raziskav, ki jih izvaja SURS. Nabor raziskav, do katerih ob odobritvi Odbora za varstvo podatkov lahko dostopajo raziskovalci, je velik, saj naj bi bili med drugim na voljo mikropodatki vseh raziskovanj, načrtovanih v letnih programih statističnih raziskovanj v zadnjih 12 letih. Izpostavimo lahko naslednje mikropodatke, primerne za podrobno demografsko raziskovanje:

- *Popis 2002*;
- *Popis 2011* [11];
- *mikropodatki, temelječi na registrskih podatkih*, npr. Centralni register prebivalstva, Statistični register delovno aktivnega prebivalstva;
- *anketni podatki*, npr. Anketa o delovni sili, ki je največje kontinuirano raziskovanje posameznikov in gospodinjstev z več kot 60 tisoč enotami v letnih bazah podatkov.

4.3 Evropski mikropodatkovni viri

Na evropski ravni raziskovalci s področja demografije lahko dostopajo do različnih demografskih podatkov, tako do podatkov Eurostata, kot tudi do nekaterih nacionalnih mikropodatkovnih virov. Poudarek enega od EU o.p. 7 projektov DwB – Data without Boundaries je med drugim tudi, kako raziskovalcem omogočiti dostop preko državnih meja do z vidika varovanja zasebnosti nezaščitenih mikropodatkov. Namen projekta je podpirati enak in

obvladljiv dostop raziskovalcev do mikropodatkov v Evropi, pri čemer je potrebno vzpostaviti harmoniziran infrastrukturni model storitev dostopa do podatkov, v katerem se ne bi zanemariło nacionalnega okolja in posebnosti zakonodaje, vseeno pa bi omogočili mednarodno sodelovanje. Trenutno so namreč podatki uradne statistike premalo uporabljeni v raziskovalne namene, razlogi za to pa izhajajo iz različnih nacionalnih, tehničnih, kulturnih in pravnih ovir. Projekt vključuje 12 delovnih sklopov, pri katerih sodelujejo organizacije kot so nacionalni statistični uradi in arhivi družboslovnih podatkov. Eden izmed skupnih ciljev je vzpostaviti spletni portal, na katerem bodo dostopne vse najpomembnejše informacije za raziskovalce v Evropi, vključno z metapodatki raziskav. Pri pripravi zasnov in vsebin za portal aktivno sodeluje tudi ADP. Cilj dela na tem delovnem sklopu je med drugim pripraviti ustrezno metapodatkovno shemo, pregled dostopnih podatkov uradne statistike ter metapodatke razpoložljivih raziskav, s tem pa podpirati evropske raziskovalce pri dostopu in delu z mikropodatki uradne statistike. V enem od delovnih sklopov, št. 3, katerega cilj je izboljšava pravne in informacijske varnosti ter ureditev statusa raziskovalcev pri dostopu do podatkov, sodeluje SURS.

V okviru projekta DwB pa je raziskovalcem preko razpisa testno omogočen tudi dostop do nezaščitenih podatkov uradne statistike [12], in sicer preko oddaljenega dostopa iz izbranih raziskovalnih podatkovnih centrov v Franciji, Nemčiji, na Nizozemskem in v Veliki Britaniji, ali v obliki gostovanja v varnih sobah. V letu 2012 sta za raziskovalce razpisana dva roka, 15. maj in 15. oktober, enako pa bo tudi v letih 2013 in 2014. Na razpis se lahko prijavijo akademski raziskovalci, vključujoč doktorske študente, ki lahko v primeru razpoložljivosti podatkov in odobritve prošnje dostopajo do podrobnih mikropodatkov na področjih sociologije, ekonomije, geografije in demografije. Raziskovalec mora do roka poslati prijavno dokumentacijo, ki vključuje opis njegovega raziskovalnega projekta, relevantnost teme, opis metodologije ter seznam podatkov, do katerih bi radi dostopal. Po morebitni odobritvi so raziskovalci deležni svetovanja pri dostopu do mikropodatkov, vključno s podporo pri nacionalnih akreditacijskih procesih, prevajanjem in financiranjem prevoza ter nastanitve. Od raziskovalca pa se na koncu pričakuje objava rezultatov raziskovanja ter kratko poročilo o obisku podatkovnega centra, namenjeno Evropski komisiji.

5 ZAKLJUČEK

Predstavljeni mikropodatkovni viri predstavljajo izhodišče za sveže in bolj kakovostno raziskovanje na različnih demografskih področjih. Dostopnost mikropodatkov se v

zadnjih letih povečuje, s tem pa se odpirajo nove dimenzije raziskovanja. Zato zaključujemo, da je smiselno v čim večji meri uporabljati navedene mikropodatke pri demografskem raziskovanju. Tovrstni podatkovni viri namreč omogočajo tako postavljanje novih vprašanj kot nove možnosti analiz in bolj relevantnih spoznanj. Večja dostopnost podatkov pa predstavlja tudi možnost prispevkov t.i. ljudske znanosti, se pravi, širše participacije laičnih raziskovalcev.

Viri

- [1] *IPUMS International*. Dostopno prek: <https://international.ipums.org/international/> (21. september 2012).
- [2] Jowell, Roger in skupina. 2009. *Evropska družboslovna raziskava 2008*. Datoteka podatkov. ESS Data Archive. Ljubljana: Arhiv družboslovnih podatkov. Dostopno prek: <http://adp.fdv.uni-lj.si/opisi/ess08/> (21. september 2012).
- [3] Statistični urad Republike Slovenije. 2012. *Delovno aktivno prebivalstvo po anketi o delovni sili*. Dostopno prek: http://www.stat.si/tema_demografsko_trg_adp.asp (21. september 2012).
- [4] Arhiv družboslovnih podatkov. 2012. *Seznam raziskovanj iz serije Politbarometer*. Dostopno prek: <http://adp.fdv.uni-lj.si/opisi/serija/PB/> (21. september 2012).
- [5] Statistični urad Republike Slovenije. 2002. *Popis 2002 – Rezultati Popisa 2002*. Dostopno prek: <http://www.stat.si/popis2002/> (21. september 2012).
- [6] *Data without Boundaries – DwB*. Dostopno prek: <http://www.dwbproject.org/> (21. september 2012).
- [7] *Arhiv družboslovnih podatkov*. Dostopno prek: <http://adp.fdv.uni-lj.si/> (21. september 2012).
- [8] Statistični urad Republike Slovenije. 2005. *Popis prebivalstva, gospodinjstev in stanovanj v RS v letu 2002 : Vzorec mikropodatkov za javno uporabo : 5-odstotni vzorec posameznikov, stanovanj in stavb*. Datoteka podatkov. Statistični urad Republike Slovenije. Ljubljana: Arhiv družboslovnih podatkov. Dostopno prek: <http://adp.fdv.uni-lj.si/opisi/popis02/> (21. september 2012).
- [9] *Integrated European Census Data*. Dostopno prek: <http://www.iecm-project.org/> (21. september 2012).
- [10] Statistični urad RS. 2012. *Osnovna pojasnila raziskovalcem o dostopu in uporabi statistično zaščitenih mikropodatkov*. Dostopno prek: http://www.stat.si/drz_stat_mikro.asp (21. september 2012).
- [11] Statistični urad Republike Slovenije. 2011. *Registrski popis 2011*. Dostopno prek: <http://www.stat.si/popis2011/> (21. september 2012).
- [12] *Data without Boundaries*. 2012. *Call for Research Proposals - Support for transnational access to official microdata - call for research proposals under the FP7 EU project "Data without Boundaries" (DwB)*. Dostopno prek: <http://www.dwbproject.org/access/call.html> (21. september 2012).