



Scenariji uporabe DDI2.5 in trajni označevalci (npr. DOI) objavljenih mikropodatkov: Možnosti uporabe v okolju uradne statistike

Janez Štebe in Sonja Bezjak, Arhiv družboslovnih podatkov
Delavnica za raziskovalce SURS
Slovenija, 29. 11. 2021

Univerza v Ljubljani



[Creative Commons Priznanje avtorstva 4.0 International](https://creativecommons.org/licenses/by/4.0/)

Vsebina

1. Scenariji uporabe DDI2.5 (hišni raziskovalci in zunanji raziskovalci – uporabniki)
2. Prednosti objavljanja metapodatkov: načela FAIR
3. Praksa citiranja uporabe raziskovalnih podatkov v člankih
4. Pregled DDI2.5
5. Demo Dataverse repozitorija za vnos metapodatkov in generiranje DOI

Scenariji uporabe DDI2.5

1. Zunanji raziskovalci, ki delajo na mikropodatkih uradne statistike
 1. Pred objavo članka v znanstveni reviji zagotovi pregled, katere podatke je uporabil in kaj je na njih delal (koda), da je prišel do rezultatov → Kriza reproduktibilnosti
 2. Znanstvene revije zahtevajo Izjavo o razpoložljivosti podatkov in njihovo navajanje

2. Hišni raziskovalci SURS
 1. Mikropodatki serij in izbranih drugih mikropodatkov iz programa statističnih raziskav dokumentirani v SUF različici (*Scientific use file*) ali ScUF (Secure)
 2. Deljena prosto dostopna koda za dokumentacijo raziskav
 1. Primer <https://www.gesis.org/en/missy/materials/EU-SILC/setups>

Setups: EU-SILC

Routines to transform *.csv to SPSS/Stata system files

These setup files were prepared by [Heike Wirth](#).



Načela FAIR

- Najdljivi, Dostopni, Povezljivi in Uporabljivi podatki
Lastnosti podatkov, metapodatkov in dokumentacije

Primer enega od kriterijev: F1: (Meta) data are assigned globally unique and persistent identifiers

(glej [https://www.go-fair.org/fair-principles/f1-\)-meta-data-assigned-globally-unique-persistent-identifiers/](https://www.go-fair.org/fair-principles/f1-)-meta-data-assigned-globally-unique-persistent-identifiers/))

(+) Uradne statistike dobro izpolnjujejo nekatere vidike: reguliran dostop, urejene pravice, na voljo izčrpna dokumentacija, kot so metodološka poročila, poročila o kakovosti, inštrumenti

(-) Manj dobro: odsotnost preko spleta stalno dostopnih metapodatkov, ki omogočajo iskanje in pridobivanje mikropodatkov

Tu vskočijo metapodatki in podatkovni repozitoriji



Revije zahtevajo citiranje in DAS

- DAS Izjava o razpoložljivosti podatkov

Data and Code Availability Statement

It contains information about the sources of data used in the replication package, in addition to or instead of such detailed description in the manuscript. This is sometimes referred to as a “Data Availability Statement,” or if it also describes where additional code might be obtained, “Data and Code Availability Statements” (DCAS). A DCAS goes beyond a typical data citation, as it describes additional information necessary for the obtention of the data. These may include required registrations, memberships, application procedures, monetary cost, or other qualifications, beyond a simple URL for download which is typically part of Data Citations.

- https://social-science-data-editors.github.io/template_README/ ; Glej tudi Vilhuber, Lars. 2019. "Report by the AEA Data Editor." *AEA Papers and Proceedings*, 109: 718-29. DOI: **10.1257/pandp.109.718**
- Primer citiranja: *Minnesota Population Center. Integrated Public Use Microdata Series, International: Version 7.2 [dataset]. Minneapolis: University of Minnesota, 2019.* <http://doi.org/10.18128/D020.V7.2>



Primeri citiranja mikropodatkov SURS

- Priporočilo SURS:

vendar pa morajo ob vsaki uporabi podatkov in informacij navesti kot vir Statistični urad Republike Slovenije ali na kratko SURS.

Zgled:

– Vir: Statistični urad Republike Slovenije

ali

– Vir: SURS`

OBJAVE_RAZISKOVALCEV

https://www.stat.si/StatWeb/File/DocSysFile/11253/OBJAVE_RAZISKOVALCEV_objavljene_do_2020_SLO.pdf

Primeri navajanja podatkov:

1. primer

- Sources: Evidence of state aids and 'de minimis' aids, Ministry of Finance of the Republic Slovenia; own calculations

2. primer

- Za analizo vstopa diplomantov višješolskega oziroma visokošolskega študija na trg dela so uporabljeni mikropodatki Statističnega urada Republike Slovenije (v nadaljevanju SURS, 2013), ki vključujejo individualne podatke za generacije diplomantov od leta 2007 do vključno leta 2011. (...) Diplomantom omenjenih generacij so pripisani podatki Statističnega registra delovno aktivnega prebivalstva (SRDAP), ki vključujejo celotno zaposlitveno zgodovino posameznika.
- Statistični urad Republike Slovenije (2013). *Mikropodatki iz podatkovnih baz izobraževanja in trg dela*.

3. primer

- Our analysis builds upon the time use survey data conducted by national statistical offices in 2008 (Austria and Italy) and 2000 (Slovenia).
- Eurostat. (2009). *Harmonised European time use surveys*. Luxemburg: Publications Office of the European Union.

Pregled DDI2.5 in DDI/CESSDA kontrolirana besedišča

Prednosti z vidika zahtev revij in FAIR

- Določena enolična vsebina za citiranje
 - Elementi: Avtor oz. kolektivni avtor, naslov, leto, ustanova za objavo (meta)podatkov, DOI
- DDI/CESSDA kontrolirana besedišča
 - Izpolnjen vidik iskanja in interoperabilnosti, dostopnosti (meta)podatkov v mednarodnih katalogih:
<https://vocabularies.cessda.eu/>
- Informacija o dostopu do podatkov (kadar niso na voljo neposredno)

Primerjaj predlog klasifikacije režimov dostopa:

- S. Bender, J. Blaschke, H. Doll, A. Gordon, C. Hirsch, D. Hochfellner, J. Lane. (2019) The Annodata Framework: **Putting FAIR data into practice. Deutsche Bundesbank Technical Report 2019-03**

📄 Kako citiram to RAZISKAVO?

Statistični urad Republike Slovenije. (2013). *Anketa o delovni sili, 2010: Letna datoteka - anonimizirana* [Podatkovna datoteka]. Ljubljana: Univerza v Ljubljani, Arhiv družboslovnih podatkov. ADP - IDNo: ADS10P.
https://doi.org/10.17898/ADP_ADS10P_V1

Primeri različic dostopa do podatkov

- Statistični urad Republike Slovenije. (2013). *Anketa o delovni sili, 2010: Letna datoteka - anonimizirana* [Podatkovna datoteka]. Ljubljana: Univerza v Ljubljani, Arhiv družboslovnih podatkov. ADP - IDNo: ADS10P. https://doi.org/10.17898/ADP_ADS10P_V1

Omejitve

Podatki se izročajo izključno za izobraževalne in raziskovalne namene. Na voljo so pod licenco Creative Commons Priznanje avtorstva + Nekomercialno 2.5 Slovenija. Avtorske pravice pridržane. Arhiv izroča podatke uporabnikom samo za namen, ki ga posebej opredelijo, ob zagotovitvi spoštovanja profesionalnih etičnih kodeksov. Uporabnik se posebej zaveže, da bo skrbel za tajnost podatkov in opravljal analize brez poskusov identifikacije posameznika.

- Statistični urad Republike Slovenije. (2012). *Anketa o delovni sili, 2010: Letna datoteka - brez omejitev zaupnosti* [Podatkovna datoteka]. Ljubljana: Univerza v Ljubljani, Arhiv družboslovnih podatkov. ADP - IDNo: ADS10. Dostopno prek <http://www.adp.fdv.uni-lj.si/opisi/ads10/>

Omejitve

Podatki dostopni registriranim raziskovalcem v statistično-analitičen in znanstveno-raziskovalen namen.

Demo Dataverse repozitorija za vnos metapodatkov in generiranje DOI

<https://dataverse.adp.fdv.uni-lj.si/> --> Dodaj raziskavo; ob objavi se generira DOI

Navodila za scenarij 1:

- Obvezno vnos povezave na publikacijo;
- Dokumentacija uporabe podatkov, ki niso dostopni: doda informacijo o mestu, postopku in pogojih dostopa, imenu in naslovu datoteke, če ni dostopna lahko nadomesti s sintetično datoteko
- Priloži kodo programskih ukazov za pripravo in analizo podatkov

Navodila za scenarij 2:

- Izčrpen opis raziskave za namen iskanja in ponovne uporabe
- Povezave na dodatno dokumentacijo
- Elaborirani pogoji dostopa
- Enolična identifikacija fizičnih datotek podatkov za naročanje

Kontakt



Hvala za pozornost!
Vprašanja?

University of Ljubljana
Faculty of Social Sciences
Social Science Data Archive
Kardeljeva ploščad 5
1000 Ljubljana
Slovenia



www.adp.fdv.uni-lj.si



arhiv.podatkov@fdv.uni-lj.si



[Arhiv.Druzboslovnih.Podatkov](https://www.facebook.com/Arhiv.Druzboslovnih.Podatkov)



[@ArhivPodatkov](https://twitter.com/ArhivPodatkov)