

Pohrana i pronalaženje podataka u arhivu CROSSDA i konzorciju CESSDA

Glavica, Marijana

Educational content / Obrazovni sadržaj

Publication status / Verzija rada: **Accepted version / Završna verzija rukopisa prihvaćena za objavljivanje (postprint)**

Publication year / Godina izdavanja: **2021**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:131:407621>

Download date / Datum preuzimanja: **2021-11-17**



Repository / Repozitorij:

[ODRAZ - open repository of the University of Zagreb](#)
[Faculty of Humanities and Social Sciences](#)



CROSSDA

Pohrana i pronalaženje podataka u arhivu CROSSDA i konzorciju CESSDA

Marijana Glavica, mglavica@ffzg.hr
Hrvatski arhiv podataka za društvene znanosti
Filozofski fakultet Sveučilišta u Zagrebu

24. svibnja 2021.



REPUBLIKA HRVATSKA
Ministarstvo znanosti i
obrazovanja



Što je CROSSDA?

- **nacionalni javni servis** koji se bavi dugoročnim očuvanjem istraživačkih podataka i omogućavanjem pristupa istraživačkim podacima u području društvenih znanosti
- dio je **europske istraživačke infrastrukture** - Konzorcija europskih arhiva podataka za društvene znanosti **CESSDA ERIC**, čime se osigurava poslovanje po najvišim europskim standardima i dobrim praksama
- djeluje pri Filozofskom fakultetu u Zagrebu koji je Ministarstvo znanosti i obrazovanja imenovalo nacionalnom koordinacijskom ustanovom za CESSDA ERIC; glavni financijer MZO



Consortium of European Social Science Data Archives

<http://cessda.eu>

Podaci u arhivu CROSSDA

- **kvantitativni i kvalitativni** podaci nastali u empirijskim istraživanjima u području društvenih znanosti
- mogućnost **ponovne upotrebe** u novim istraživanjima
 - opsežna istraživanja složenih fenomena na nacionalno reprezentativnim uzorcima
 - organizacijski i financijski zahtjevni poduhvati
 - pružaju odgovore na pitanja na koje se u pravilu ne odgovara unutar samo jednog znanstvenog rada - veliki potencijal za ponovno korištenje!
 - za razumijevanje društva često je nužna usporedba povijesnih i suvremenih društvenih fenomena (komparativna i longitudinalna istraživanja, analiza trendova i slično)
- **transparentnost** istraživačkog procesa
- mogućnost **reproduciranja** rezultata iznesenih u objavljenom znanstvenom radu
- "otvorena znanost" i obaveze koje uvode časopisi

Sadržaj

I. dio: UPRAVLJANJE

preduvjeti za pohranu
podataka

plan upravljanja
podacima

informirani pristanak

II. dio: POHRANA

procedura pohrane

razine prisupa

izjava o pohrani

metapodaci

III. dio: PRONALAZENJE

katalozi: CROSSDA,
CESSDA i drugi izvori

uvjeti korištenja
podataka

Upravljanje podacima

Preduvjeti za pohranu podataka

- dobro organizirane datoteke s podacima i popratnom dokumentacijom u kojima se (vi sami) možete snaći
- dovoljno dokumentacije za razumijevanje sadržaja datoteka s podacima
- podaci očišćeni i pripremljeni za analizu
- anonimizirani podaci
- istraživanje provedeno u skladu s etičkim načelima i pravnim propisima (GDPR)
- razjašnjena prava vlasništva nad podacima i popratnom dokumentacijom

Plan upravljanja podacima

- alat za strukturiranje upravljanja podacima koji nastaju ili se koriste unutar istraživačkog projekta (Data Management Plan, DMP)
- izrađuje se na početku projekta, ažurira tijekom cijelog projekta
- EK već nekoliko godina zahtjeva od istraživača izradu DMP-a
- HRZZ od jeseni uvodi obavezu izrade DMP-a



cessda

DMEG Data Management
Expert Guide

<https://cessda.eu/dmeg>

Data Management Expert Guide



1. Plan



2. Organise & Document



3. Process



4. Store



5. Protect



6. Archive & Publish



7. Discover



8. Contributors



- namijenjen istraživačima u području društvenih znanostima
- obuhvaća upravljanje podacima tijekom cijelog istraživačkog projekta, od faze planiranja do arhiviranja i objave podataka
- sadrži popis pitanja (ček-listu) za izradu plana upravljanja podacima:
 - [.pdf](#)
 - [.docx](#)

Organizacija podataka

- organizacija varijabli
 - struktura foldera
 - imenovanje datoteka
 - verzioniranje
 - formati podataka
- Povezani CROSSDA webinar:

[Kako organizirati podatke u istraživačkom projektu?](#)

[Rad s tabličnim podacima](#)
- trivijalan zadatak ili ozbiljan organizacijski problem?
 - savjet: jedna osoba u projektu neka bude zadužena za organizaciju podataka

Dokumentacija

- informacije o projektu / istraživanju
- metodološki izvještaj: opis uzorka i uzorkovanje, korišteni instrumenti, korišteni softver za prikupljanje podataka, opis postupka prikupljanja podataka, upute za anketare i kontrola anketara, odaziv, utežavanje
- kodne knjige (značenje varijabli i kodova)
- upitnici, pokazne kartice
- bibliografski zapisi o publikacijama u kojima su korišteni podaci
- dokumentacija o transformacijama datoteke s podacima

Čišćenje i priprema za analizu

- pod očišćenim podacima podrazumijevamo ispravljene podatke, očišćene od neispravnih, nepotpunih, nepravilno oblikovanih ili duplih zapisa podataka.
- sve promjene u podacima u odnosu na sirovu verziju trebale bi biti evidentirane, najbolje uz pomoć sintakse statističkog programa koji je za to korišten



11:00 - 14:00

**PER RSPERA AD ASTRA: PRIPREMA PODATAKA
POMOĆU PROGRAMSKOG JEZIKA R**

Petak , Online via Zoom

**VIŠE
INFORMACIJA**

Anonimizacija

- iz podataka su uklonjeni svi **direktni i indirektni** identifikatori uz pomoć koji je moguće otkriti konkretnog pojedinca čiji su odgovori prikupljeni u istraživanju
- potpuna anonimizacija često dovodi do velikog smanjivanja informacijske vrijednosti podataka
- sve promjene u podacima trebale bi biti evidentirane, najbolje uz pomoć sintakse statističkog programa koji je za to korišten

28
SVI

13:00 - 15:00

**UVOD U ANONIMIZACIJU KVANTITATIVNIH
PODATAKA**

Petak , Online via Zoom

**VIŠE
INFORMACIJA**

Informirani pristanak (1)

- postupak u kojem istraživač pruža sudioniku istraživanja informacije o istraživanju, tako da sudionik može dobrovoljno i informirano odlučiti hoće li prihvatiti ili odbiti suradnju
- ako se prikupljaju osobni podaci - privola (GDPR)
 - obavezno u pisanom obliku ili kao zvučni zapis
 - formulacije trebaju biti nedvosmislene i specifične
 - privola se daje jasnom potvrdnom akcijom, najbolje granularno
- ako se ne prikupljaju osobni podaci (anonimne ankete) - etički pristanak
 - sudionik pristaje ili ne pristaje sudjelovati u istraživanju na temelju informacija koje mu je istraživač pružio

Više informacija: [CESSDA Data Management Expert Guide](#)

Informiranje o korištenju podataka (1)

Razmislimo o ovim formulacijama:

- *"Prikupljeni podaci objavit će se samo u sumarnom obliku u izvještajima i znanstvenim radovima."*
- *"Vaši će odgovori biti dostupni isključivo niže potpisanim istraživačima za potrebe znanstvenih analiza."*
- *"Isključivo anonimni podaci mogu biti proslijeđeni trećim osobama samo u znanstvene svrhe."*

Kako ste vi u svojem posljednjem istraživanju informirali sudionike o korištenju podataka? Zašto vam je bitna takva formulacija?

Informiranje o korištenju podataka (2)

- istraživačima je bitno poručiti sudionicima da će se s njihovim podacima postupati odgovorno i tako zadobiti povjerenje sudionika - zbog toga se ističe znanstvena svrha i/ili ograničavanje pristupa podacima samo na usku skupinu istraživača
- ispravno anonimizirani podaci mogu se koristiti i u svrhu nastave, ili možda za donošenje javnih politika
- možda želite omogućiti pristup podacima jer smatrate da transparentnost pridonosi kvaliteti vašeg znanstvenog rada - želite da bilo tko, sa što manje prepreka, može provjeriti rezultate koje ste objavili u članku
- i podaci koji nisu anonimizirani mogu se pružiti na korištenje i drugim istraživačima ako sudionik na to pristane i ako se pristup podacima drži pod kontrolom
- povjerenje se gradi iskrenošću - bolje je sudionicima iskreno reći da će se njihovi podaci koristiti na siguran način i u različite svrhe

Formulacije u pristanku koje omogućuju dijeljenje podataka

za istraživanja u kojima se prikupljaju osobni podaci (podatke nije moguće anonimizirati tako da se ne izgubi njihova informacijska vrijednost):

- *Razumijem da će moji osobni podaci poput imena, email adrese i broja telefona biti dostupni samo osobama koje provode ovaj projekt.*
- *Slažem se da moji odgovori budu pohranjeni u Hrvatskom arhivu podataka za društvene znanosti.*
- *Razumijem da će drugi istraživači imati pristup ovim podacima samo ako se obavežu da će čuvati povjerljivost informacija sadržanih u podacima.*

Informacije sudionicima o dijeljenju podataka

za istraživanja u kojima se ne prikupljaju osobni podaci

- *Anonimizirani podaci bit će trajno pohranjeni u Hrvatskom arhivu podataka za društvene znanosti kako bi bili dostupni u znanstvene svrhe ili za korištenje u podučavanju studenata.*
- *Vaši odgovori su anonimni što znači da nije moguće otkriti koja je osoba dala koji odgovor. Takvi podaci bit će trajno pohranjeni u Hrvatskom arhivu podataka za društvene znanosti i javno dostupni kako bi se povećala iskoristivost i odjek informacija koje ste nam u ovom upitniku pružili.*

Autorska prava

- tko je vlasnik podataka? tko ima pravo dati dozvolu za pohranu i pružiti podatke na korištenje?
 - autor
 - financijer?
 - ustanova pri kojoj je autor zaposlen?
- kako financijeri ili ustanove stječu pravo raspolaganja podacima?
 - isključivo pisanim putem!
 - ugovor o financiranju projekta
 - ugovor o radu
 - izjava o prijenosu prava
- pohrana i objava sekundarnih podataka koje ste koristili u istraživanju nije moguća bez dozvole vlasnika

Pohrana i objava podataka

"Što je moguće otvorenije, po potrebi zatvoreno."

- svi podaci u društvenim znanostima ne mogu biti potpuno otvoreni
 - podaci sadrže informacije pomoću kojih je moguće identificirati identitet pojedinaca
 - istraživači su sudionicima u istraživanju obećali da će se podaci koristiti samo u određene svrhe
- ipak, podatke je moguće je pohraniti i objaviti! tri osnovna mehanizma:
 - informirani pristanak (etički i pravni)
 - anonimizacija
 - reguliranje pristupa
- o reguliranju pristupa brinu arhivi podataka
 - CROSSDA trenutno pohranjuje i objavljuje samo podatke koji ne sadrže osobne identifikatore
 - sljedeći korak u razvoju arhiva CROSSDA je implementacija mehanizama kontrole pristup podacima koji sadrže osobne podatke (udaljeni pristup ili sigurna soba)
- moguće je objaviti dvije verzije podataka - jednu koja je "potpuno" anonimizirana, drugu koja sadrži osobne identifikatore i dostupna je uz registraciju i kontrolu

Reguliranje pristupa podacima u arhivu CROSSDA

A. Slobodno dostupni podaci (otvoreni pristup)

- dostupni pod uvjetima licence Creative Commons Autorstvo 4.0 Međunarodna (CC BY 4.0)

B. Podaci dostupni uz registraciju

- **Podaci dostupni samo za znanstvene svrhe**
 - priprema i provedba znanstvenih istraživanja (projekti, radovi, doktorske disertacije)
 - provjera rezultata iznesenih u znanstvenom radu (recenzenti ili drugi istraživači)
 - izrada poslijediplomskih specijalističkih radova i diplomskih radova na razini diplomskih studija
- **Podaci dostupni za znanstvene svrhe i za korištenje u nastavi**
 - osim u znanstvene svrhe, dozvoljeno je i korištenje u nastavi (npr. vježbe iz statistike, studentski seminarski radovi, završni radovi na razini preddiplomskih studija)

Procedura pohrane (1)

- istraživač šalje upit o pohrani na arhiv.podataka@ffzg.hr
- dogovara se sastanak s osobljem arhiva, razgovor o:
 - pripremi podataka
 - [razinama pristupa podacima](#)
 - [uvjetima korištenja podataka](#)
 - sljedećim koracima
- potpisuje se izjava o pohrani
 - istraživač daje dozvole za korištenje podataka i arhivu i krajnjim korisnicima
 - voditelj arhiva potvrđuje da su podaci zaprimljeni i da će s podacima postupati kako je dogovoreno u izjavi
 - predložak izjave: [za otvorene podatke](#) | [samo u znanstvene svrhe](#)

Procedura pohrane (2)

- istraživač šalje zahtjev za otvaranjem korisničkog računa u Dataverse-u
 - [prijaviti se](#) u sustav korištenjem **AAI@EduHr** korisničkog računa
- osoblje arhiva dodjeljuje istraživaču potrebne ovlasti u Dataverse-u
- istraživač kreira skup podataka i prenosi datoteke s podacima i popratnom dokumentacijom u Dataverse
 - podaci u SPSS-u ili Stata automatski se konvertiraju u otvoreni format (tekstualna datoteka s podacima + metapodaci u DDI formatu)
 - za skup podataka rezervira se DOI broj
- istraživač unosi metapodatke o skupu podataka u Dataverse
 - [upute za izradu metapodataka](#)

Procedura pohrane (3)

- osoblje arhiva pregledava sadržaj predanih datoteka
 - jesu li sve varijable i kodovi za vrijednosti varijabli jasno dokumentirani?
 - jesu li u podacima sadržane vrijednosti koje nisu navedene u kodovima?
 - je li dostavljena sva potrebna dokumentacija za razumijevanje skupa podataka?
 - je li postignuta odgovarajuća razina anonimizacije?
- osoblje arhiva šalje istraživaču izvještaj o skupu podataka sa savjetima za doradom
- istraživač ispravlja podatke i nadopunjuje dokumentaciju prema savjetima i dostavlja novu verziju datoteka



Procedura pohrane (4)

- osoblje arhiva preimenovat će datoteke u skladu s internim pravilima
 - dodaje se prefiks datotekama (arhivski broj + oznaka za vrstu datoteke), npr:
 - cda1010_dat_SOIRES_RapidCATI_SUF11032021_hr.tab
 - cda1010_src_SOIRES_RapidCATI_priprema_kodiranja_ciscenje_hr.do
- ukoliko je potrebno, datoteke se konvertiraju u formate prikladnije za korištenje i/ili dugoročnu pohranu
- osoblje arhiva kontrolira i nadopunjuje metapodatke (ključne riječi iz tezaurusa, predmetna klasifikacija i drugo)
- prije objave, zapis se šalje na provjeru
- nakon što istraživač odobri promjene, osoblje arhiva objavljuje skup podataka

Pronalaženje podataka

https://data.crossda.hr

CROSSDA

Add Data ▾

Search ▾

User Guide

Support

Sign Up

Log In

Search this dataverse...

Find

Advanced Search

 **Dataverses (12)**

 **Datasets (8)**

 **Files (32)**

Dataverse Category

Research Project (5)

Department (3)

Organization or Institution (3)

Teaching Course (1)

Publication Year

2018 (5)

2019 (4)

2020 (4)

2021 (4)

2017 (3)

Keyword Term

COVID-19 (4)

PSYCHOLOGY (2)

International treaties (1)

1 to 10 of 20 Results

Sort ▾

Istraživanje stavova i iskustava stanovnika Hrvatske vezanih uz pandemiju bolesti COVID-19

May 18, 2021 - Otpornost hrvatskog društva uslijed COVID-19 pandemije - SOCRES



Ančić, Branko; Bagić, Dragan; Brajčić Vuković, Marija; Matković, Teo; Rodik, Petra; Šuljok, Adrijana, 2021, "Istraživanje stavova i iskustava stanovnika Hrvatske vezanih uz pandemiju bolesti COVID-19", <https://doi.org/10.23669/BXMSCI>, CROSSDA, V1, UNF:6:YLTxbldWY5l4jZSz2L32w== [fileUNF]

Istraživanje je provedeno u okviru projekta Otpornost hrvatskog društva uslijed COVID-19 pandemije - SOCRES. Upitnikom se nastojalo istražiti značajnije promjene u populaciji nastale uslijed pandemije COVID-19, uključujući promjene u razini prihoda, uvjeta rada, osobnog i Instituta...

MegaRT: Lexical Decision Times for Croatian Words

May 10, 2021 - Odsjek za psihologiju



Tonković, Mirjana; Dumančić, Francesca; Vlašiček, Denis, 2021, "MegaRT: Lexical Decision Times for Croatian Words", <https://doi.org/10.23669/PEVB54>, CROSSDA, V1, UNF:6:KSgyc0fgFgJ0kHr3l+d/w== [fileUNF]

These datasets contain individual reaction time data to Croatian words, and pseudowords in a lexical decision task. Pseudowords were created using Wuggy (Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. Behavior Research Methods 42(3), 627-633.)....

Data for: Who believes in COVID-19 conspiracy theories in Croatia? Prevalence and predictors of conspiracy beliefs

May 8, 2021 - (Re)building society: A longitudinal study of post-corona social recovery in Croatian general population (ReSPoC)



Tonković, Mirjana; Dumančić, Francesca; Jakić, Marija; Čuček, Dinko, 2021, "Data for: Who believes in COVID-19

https://datacatalogue.cessda.eu/

Consortium of European Social Science Data Archives



Search Select a language for optimal search results

English

23943 studies found in English from a total of 31319

About

Reset filters

Clear search

Topic



Results per page

30

Sort by

Relevance

Collection years



1

2

3

4

...



Country



Publisher



Eyelit: Eye-movement and reader response data during literary reading

Mak, MSc H.M.L. DAI=info:eu-repo/dai/nl/416268420 (Radboud University);

Willems, dr. R.M. DAI=info:eu-repo/dai/nl/31593378X (Radboud University)

This dataset is an extensive eye tracking dataset of 102 participants reading three Dutch literary short stories each (7790 words per participant). The preprocessed data set includes (1) Fixation report (fixation-level), (2) Saccade report, (3) Interest Area report (word-level), (4) Trial report (aggregated data for each page; stories were split up into 30 pages each), (5) Sample report (the data were sampled at 500 Hz, this report includes data on all individual samples), (6) Questionnaire d...

Read more

Study description available in:

EN

Access data

Ostali izvori - <https://www.crossda.hr/pronalazenje-podataka/>

CROSSDA

ARHIV PODATAKA ▾

PRISTUP PODACIMA ▾

POHRANA PODATAKA ▾

UPRAVLJANJE PODACIMA



Pronalaženje podataka

Katalog podataka u arhivu CROSSDA

CESSDA Data Catalog



<https://datacatalogue.CESSDA.eu/>

objedinjuje podatke pohranjene u europskim arhivima podataka za društvene znanosti u konzorciju CESSDA

Arhivi zemalja članica

Partneri

Međunarodna komparativna istraživanja

- [Atlas of European Values](#)
- [Comparative Study of Electoral Systems \(CSES\)](#)
- [Eurobarometer](#)
- [European Company Surveys](#)
- [European Election Studies \(EES\)](#)
- [European Quality of Life Surveys \(EQLS\)](#)

Hvala na pažnji!

arhiv.podataka@ffzg.hr

<https://www.crossda.hr>

@crossda_data