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://urn.nsk.hr/urn:nbn:hr:131:407621

Rights / Prava: Attribution 4.0 International/Imenovanje 4.0 međunarodna

Download date / Datum preuzimanja: **2024-03-29**

Filozofski fakultet
Sveučilišta u Zagrebu

Repository / Repozitorij:

ODRAZ - open repository of the University of Zagreb
Faculty of Humanities and Social Sciences

dabar
DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI

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.

cessda

REPUBLIKA HRVATSKA
Ministarstvo znanosti i obrazovanja

CC BY

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

cessda

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: PRONALAŽENJE

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



- 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 webinari:

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



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



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 svrhe i/ili ograničavanje pristupa podacima samo na usku skupinu istraživača
- ispravno anonimizirani podaci mogu se koristi 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.*

Više informacija: Consent for data sharing / UK Data Service

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 Stati 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

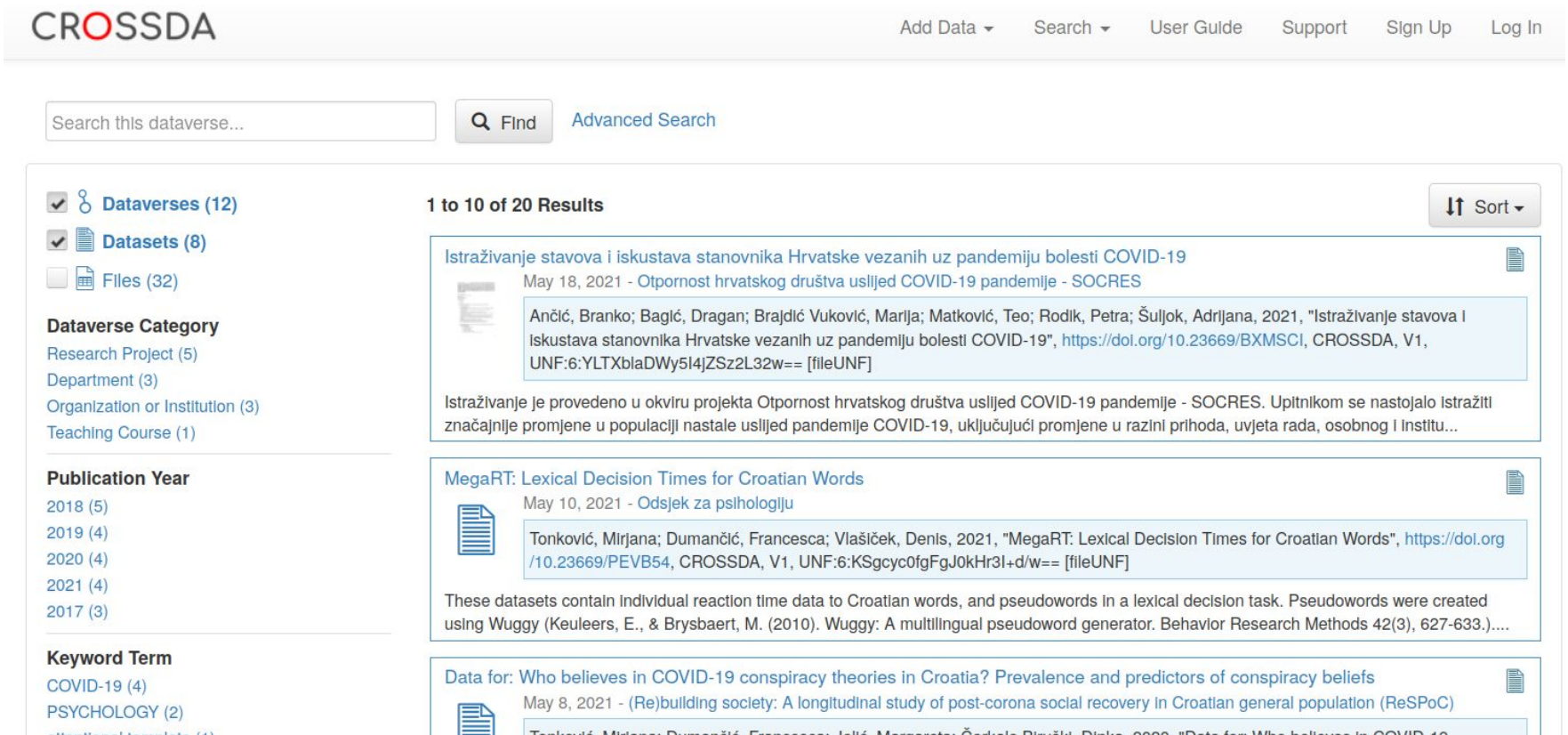


Procedura pohrane (4)

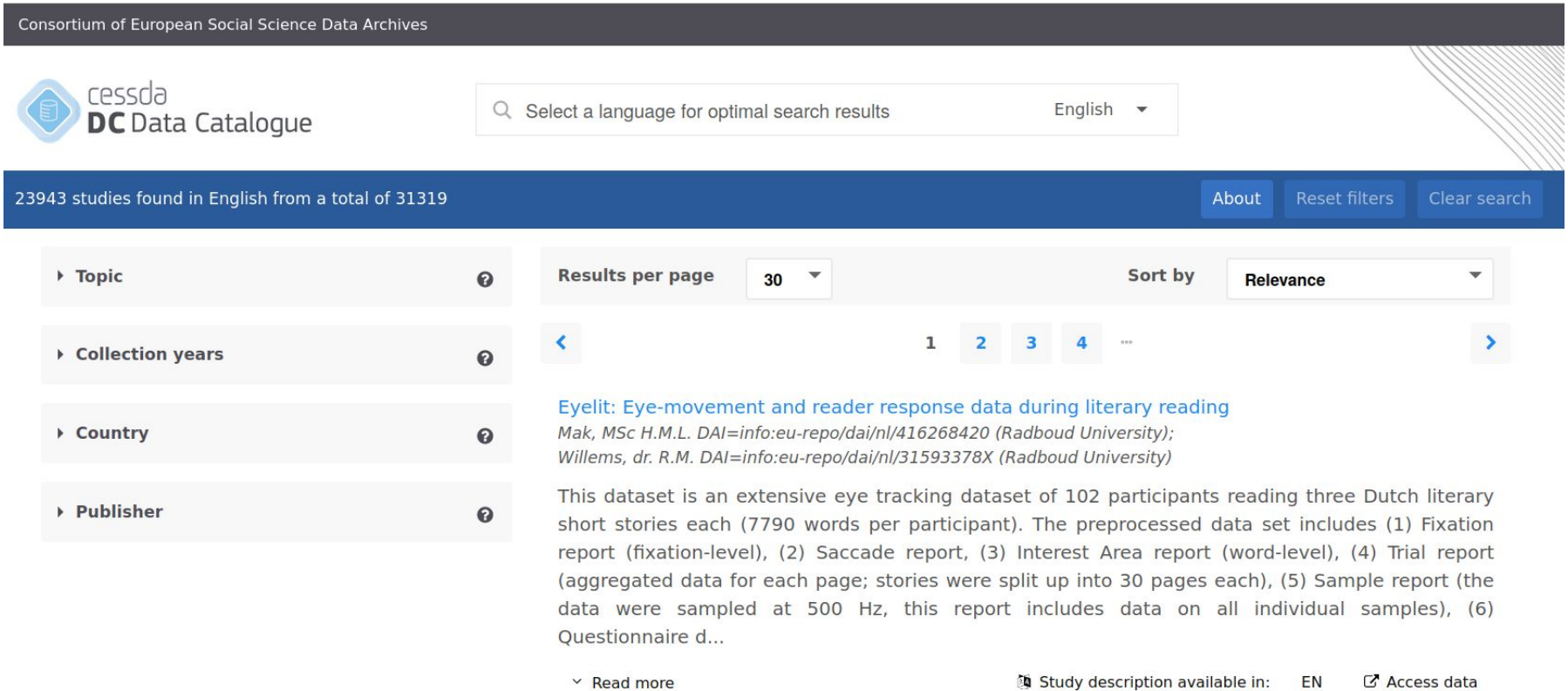
- osoblje arhiva preimenovat će datoteke u skladu s internim pravilima
  - dodaje se prefiks datotekama (arhivski broj + oznaka za vrstu datoteke), npr:
    - `cda1010_dat_SOCRES_RapidCATI_SUF11032021_hr.tab`
    - `cda1010_src_SOCRES_RapidCATI_priprema_kodiranja_ciscenje_hr.do`
- ukoliko je potrebno, 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

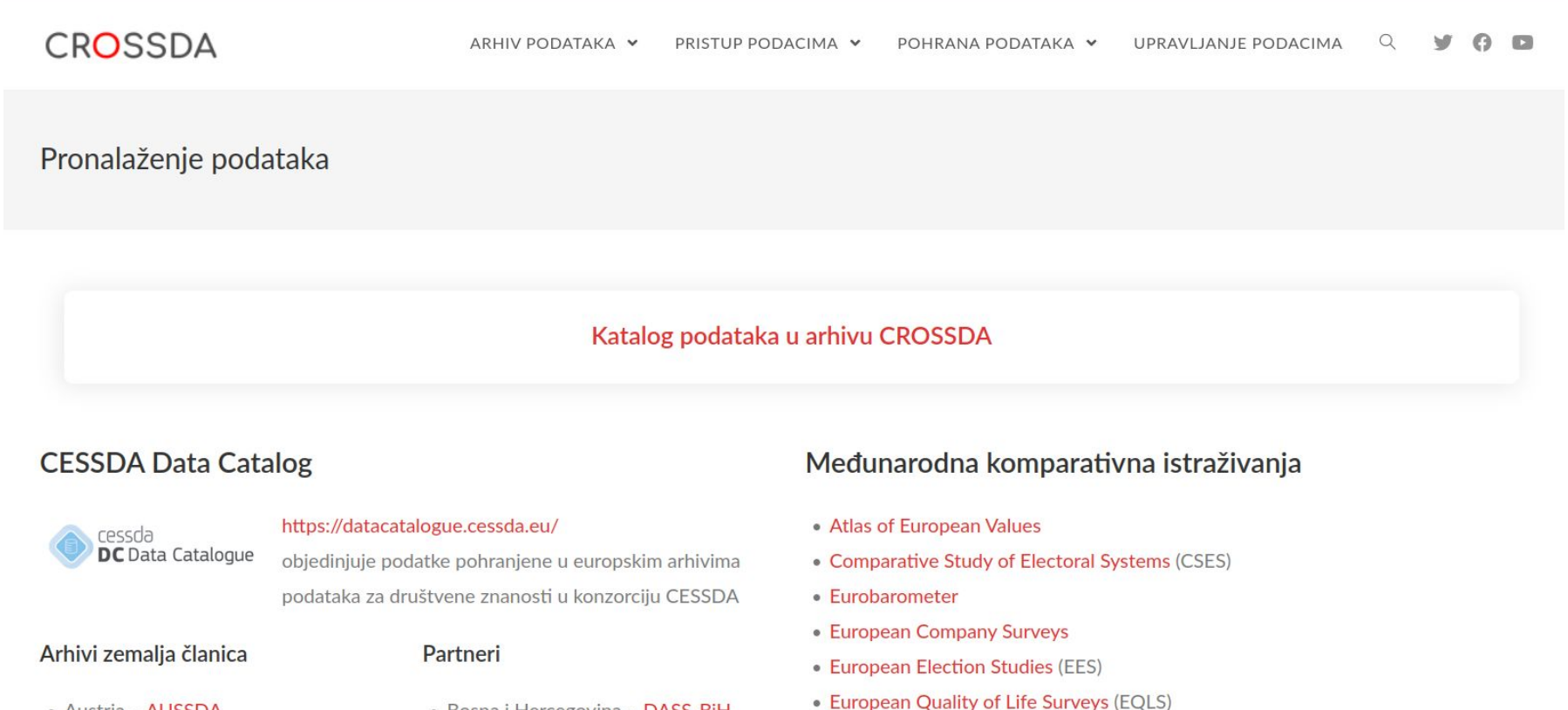
https://data.crossda.hr



https://datacatalogue.cessda.eu/



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



Hvala na pažnji!

arhiv.podataka@ffzg.hr

https://www.crossda.hr

@crossda_data