

Bogunović, Irena

Data management plan / Plan upravljanja istraživačkim podacima

Publication year / Godina izdavanja: **2024**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:187:603804>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2025-03-22**



Sveučilište u Rijeci, Pomorski fakultet
University of Rijeka, Faculty of Maritime Studies

Repository / Repozitorij:

[Repository of the University of Rijeka, Faculty of
Maritime Studies - FMSRI Repository](#)



Plan upravljanja istraživačkim podacima

Opće informacije		
	Ime i prezime predlagatelja	Irena Bogunović
	Matična organizacija	Pomorski fakultet Sveučilišta u Rijeci
	Naziv projekta	ENGLJSKE RIJEČI U HRVATSKOME JEZIKU: IDENTIFIKACIJA, AFEKTIVNO-SEMANTIČKO NORMIRANJE I ISPITIVANJE KOGNITIVNE OBRADE BIHEVIORALNIM I NEUROZNASTVENIM METODAMA
	Upravitelj podacima	Irena Bogunović; irena.bogunovic@pfri.uniri.hr
1.	Prikupljanje podataka i dokumentacija	
	Koje ćete podatke prikupljati, obrađivati, stvarati ili se ponovno njima koristiti? (navedite format, vrstu i opseg podataka)	<p>Podaci za proširenje korpusa će se prikupljati alatom <i>lacon-go-tiny-scrapy</i> (alat je dostupan na platformi GitHub). Prikupljat će se podaci s 10 novinskih portala u Hrvatskoj. Uz sirovi HTML, sprema se vrijeme preuzimanja stranice te s URL. Svi podaci će se spremati u JSON formatu te komprimirati gzip-om. Podaci će biti pohranjeni na lokalnom računaru. Predviđeni opseg podataka je 200 GB.</p> <p>Nakon obrade sirovih podataka, izradit će se novi korpus korištenjem Python skripti dostupnih na platformi GitHub. Rezultat će se sastojati od URL-a te očišćenog sadržaja HTML-a. Podaci će biti spremljeni u JSON formatu. Predviđeni opseg podataka je 10 GB.</p> <p>Iz očišćenog korpusa će se izdvojiti te agregirati sve riječi s pripadajućim frekvencijama. Lista riječi će biti pohranjena u .csv formatu te pohranjena na lokalnom računaru te repozitoriju Pomorskog fakulteta. Predviđeni opseg podataka je 20 MB.</p> <p>Iz liste riječi će se izdvojiti engleske riječi korištenjem Python skripti dostupnih na platformi GitHub. Lista riječi će biti pohranjena u .csv formatu te pohranjena na lokalnom računaru i repozitoriju Pomorskog fakulteta. Predviđeni opseg podataka je 2 MB.</p>
	Kako će se podaci prikupljati, obrađivati ili stvarati? (ukratko navedite metodologiju i procese osiguranja kvalitete te načine organiziranja podataka)	Sirovi HTML podaci će se prikupljati pomoću alata <i>lacon-go-tiny-scrapy</i> (alat je dostupan na platformi GitHub). Sirovi podaci će biti obrađeni pomoću programskog jezika Python. Grupiranje riječi te izdvajanje engleskih riječi će biti provedeno uz pomoć programskog jezika Python. Svi korišteni Python programi će biti ručno testirani na manjem opsegu podataka. Podaci iz svakog koraka će biti odvojeni u posebne direktorije te će se svi koraci moći ponoviti bez ugrožavanja kvalitete podataka iz prijašnjih koraka.
	Koju ćete dokumentaciju i metapodatke ustupiti osim podataka? (navedite koje su	Podaci korpusa će biti spremljeni u JSON formatu, dok će liste riječi biti spremljene u .csv formatu, što neće zahtijevati specijalizirani alat za pregled i analizu podataka.

	informacije potrebne korisnicima kako bi mogli čitati i interpretirati podatke u budućnosti te koji će se standardi koristiti pri tumačenju podataka)	
2.	Pravna i sigurnosna pitanja	
	Jeste li ograničeni sporazumom o povjerljivosti? Imate li potrebna dopuštenja za prikupljanje, obradu, čuvanje i dijeljenje podataka? Jesu li osobe čiji se podaci pohranjuju informirani o tome i jesu li dali privolu? Kojim ćete se metodama koristiti u svrhu zaštite osjetljivih podataka (GDPR - posebne kategorije osobnih podataka)?	Projekt je odobren odlukom etičkog povjerenstva Pomorskog fakulteta donesenom 11. 12. 2019. (602-04/19-01/53; 2170-57-19-01-1). Projekt poštuje sva ograničenja i zahtjeve kako je utvrđeno Zakonom o provedbi Opće uredbe o zaštiti podataka (NN 42/18). Kraj projekta nije vezan za sudionike, te navedeno projektno razdoblje ne uključuje istraživanje u kojem sudjeluju ispitanici.
	Kako će se regulirati pristup podacima i njihova sigurnost? Koji su potencijalni rizici koje treba uzeti u obzir? Kako ćete osigurati sigurnost pohrane osjetljivih podataka?	Podaci će se pohraniti na računalima voditelja projekta i u Laboratoriju te na platformi GitHub. Nema potencijalnih rizika.
	Kako ćete upravljati zaštitom autorskih prava i intelektualnog vlasništva? Tko će biti vlasnik podataka? Koje će se licencije primjenjivati na podatke? Koja će se ograničenja primjenjivati na ponovnu uporabu osobnih podataka?	Podaci će biti prikladni za dijeljenje te će se objaviti pod licencom <i>Creative Commons CC0</i> . Osobni podaci neće se rabiti.
3.	Pohrana i čuvanje podataka	

	Kako će podaci biti pohranjeni i kako će biti napravljena sigurnosna kopija podataka (<i>backup</i>) tijekom istraživanja? Koji su kapaciteti čuvanja podataka kojim raspolazete? Kojim se procedurama koristite za sigurnosnu kopiju (<i>backup</i>)?	Prilikom prikupljanja sirovih podataka, svi će se podaci automatski kopirati na <i>backup</i> disk. Korpus te liste riječi će biti spremljeni na dva različita diska te na Google Drive-u i repozitoriju Pomorskog fakulteta.
	Koji je vaš plan čuvanja podataka? U kojim će se formatima čuvati?	Podaci će trajno biti pohranjeni na digitalnom repozitoriju Pomorskog fakulteta u sklopu nacionalnog sustava Dabar (https://repository.pfri.uniri.hr/). Tablični podaci bit će pohranjeni u .csv datotekama, tekstualni u .docx i .pdf datotekama.
4.	Dijeljenje i ponovna uporaba podataka	
	Kako i gdje će se podaci dijeliti? Na kojem repozitoriju planirate dijeliti podatke? Kako će potencijalni korisnici doznati za podatke?	Podaci će se dijeliti na platformi Figshare. Potencijalni korisnici doznat će za podatke iz publikacija nastalih kao rezultat istraživanja. Konačnu verziju skupa podataka voditelj projekta podijelit će putem institucijskog repozitorija Pomorskog fakulteta Sveučilišta u Rijeci uspostavljenog u nacionalnom sustavu Dabar. Podaci će biti objavljeni pod CC0 licencom.
	Ako postoje podaci koji se ne smiju dijeliti (prijavitelji vezani zakonskim, etičkim, autorskim pravila, povjerljivošću i sl.), pojasnite razloge ograničenja.	Svi osobni podaci bit će anonimizirani u skladu s profesionalnim standardima i Zakonom o provedbi Opće uredbe o zaštiti podataka (NN 42/18). Svi ostali podaci smiju se dijeliti.
	Potvrdite da ćete se koristiti digitalnim repozitorijem koji je u skladu s načelima <i>FAIR-a</i> .	Digitalni repozitorij Pomorskog fakulteta dio je nacionalnog sustava Dabar, koji djeluje u skladu s načelima FAIR-a. Repozitorij Figshare također djeluje u skladu s načelima FAIR-a.
	Potvrdite da ćete se koristiti digitalnim repozitorijem koji održava neprofitna organizacija (ako ne, objasnite zašto ne možete dijeliti podatke na digitalnom repozitoriju koji nije komercijalan).	Za pohranu i dijeljenje podataka koristit će se Digitalni repozitorij Pomorskog fakulteta koji je sastavi dio nacionalnog sustava Dabar. Njime upravlja javna ustanova SRCE koja uslugu pruža besplatno, stoga potvrđujemo namjeru korištenja repozitorija koji održava neprofitna organizacija.

Ref:

[1] Celjak, D., Dorotić Malič, I., Matijević, M., Poljak, Lj., Posavec K. i Turk, I.: „Istraživački podaci - što s njima?“ [Istraživački podaci - što s njima? : priručnik o upravljanju istraživačkim podacima | Digitalni repozitorij Srca \(unizg.hr\)](#)