

# Proširenje osjetilnosti senzora u laboratoriju za obradbu i analizu podataka iz okoline (SSA@EDAL)

---

**Kalinić, Hrvoje**

**Data management plan / Plan upravljanja istraživačkim podacima**

*Publication year / Godina izdavanja:* **2023**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:166:162126>

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

*Download date / Datum preuzimanja:* **2024-12-26**

*Repository / Repozitorij:*

[Repository of Faculty of Science](#)



## Plan upravljanja istraživačkim podacima

Opće informacije		
	Ime i prezime predlagatelja	Izv. prof. dr. sc. Hrvoje Kalinić
	Matična organizacija	Sveučilište u Splitu, Prirodoslovno-matematički fakultet
	Naziv projekta	Proširenje osjetilnosti senzora u laboratoriju za obradbu i analizu podataka iz okoline (SSA@EDAL)
	Upravitelj podacima	
1.	Prikupljanje podataka i dokumentacija	
	Koje ćete podatke prikupljati, obrađivati, stvarati ili se ponovno njima koristiti? (navedite format, vrstu i opseg podataka)	<p>Na projektu planiramo raditi s dvije vrste podataka: prometnim podacima i oceanografskim podacima.</p> <p>Prometni podatci će se generirati na temelju simulacije na prometnom modelu grada Dublina u Irskoj. Podatci koji će se dobiti iz modela sadržavati će geolokciju s vremenskom oznakom i brzinom vozila koje daju vozila u pokretu (eng. Floating car dana, FCD). Vozilo se stoga ponaša kao senzor kretanja koristeći ugrađeni GPS prijemnik ili mobilni telefon. FCD podaci s rezolucijom od 2 sekunde će bit dokumentirani u CSV formatu za 11 različitih prometnih scenarija za tipični radni dan. Podaci (identifikacija vozila) u CSV formatu po stupcima su:</p> <ul style="list-style-type: none"> <li>• ID vozila</li> <li>• X koordinata vozila [m]</li> <li>• Y koordinata vozila [m]</li> <li>• Trenutna brzina vozila [m/s]</li> <li>• ID prometne trake na kojoj se nalazi vozilo</li> <li>• Vrijeme uzorkovanja podatka o vozilu iz simulacije [s]</li> </ul> <p>Potrebni prostor za pohranu neobrađenih podataka procjenjujemo na 25 GB</p> <p>Oceanografski podaci dohvaćaju iz javno dostupnih izvora – npr. iz <a href="https://www.oceancolour.org/">https://www.oceancolour.org/</a> ili <a href="https://cds.climate.copernicus.eu/">https://cds.climate.copernicus.eu/</a>. Gotovo svi podaci se izvorno nalaze u netCDF formatu. Iz njih „izrezujemo“ područje od interesa sa određenom rezolucijom i vremenskim uzorkovanjem. Primjerice, godišnji dnevni podaci koncentracije klorofila u Jadranu, uzeti sa širine 40° - 46° N i dužine 12° - 20° E, prostorne skale 1 x 1 km zauzimaju nešto manje od 650 MB, odnosno cca. 15GB za 20 godina. Satni podaci ERA5 reanalize vjetera (sjever - jug i istok - zapad komponente) prostorne rezolucije 0.25° x 0.25° (cca. 30x30km) na području šire Europe, od 30° - 40° N širine i -10° - 40° E dužine, za jednu godinu zauzimaju oko 1.7 GB, odnosno oko 70 GB za 40 godina.</p>

<p>Kako će se podaci prikupljati, obrađivati ili stvarati? (ukratko navedite metodologiju i procese osiguranja kvalitete te načine organiziranja podataka)</p>	<p>Prometni podaci su generirani pomoću mikroskopskog simulatora prometa SUMO (eng. Simulation of Urban MObility) koristeći prometni model grada Dublin. FCD podaci se izravno generiraju tijekom izvršavanja simulacije te se istovremeno pohranjuju u XML datoteku. Po završetku simulacije iz XML datoteke se generira završna CSV datoteka pomoću Python programskog jezika pogodna za daljnju obradu. Svaka CSV datoteka u nazivu sadrži broj koji predstavlja sjeme (slučajni broj) koji se koristi za inicijaliziranje generatora pseudoslučajnih brojeva koji osiguravaju različitost prometnih scenarija.</p> <p>Oceanografski podatci se iz Copernicus sustava dohvaćaju putem skripti koje definiraju varijablu, područje od interesa, vremenski raspon i prostornu rezoluciju. Na sličan način dohvaćamo podatke iz Oceancolour repozitorija. Postojanje skripti koje to automatski rade omogućava kontrolu kvalitete. Tipično radimo predprocesiranje podataka (primjerice filtriranje kopenenom maskom) te pohranjujemo međurezultate u obliku PKL formatu, a datoteku imenujemo u obliku <i>varijabla_godina.pkl</i>.</p>
<p>Koju ćete dokumentaciju i metapodatke ustupiti osim podataka? (navedite koje su informacije potrebne korisnicima kako bi mogli čitati i interpretirati podatke u budućnosti te koji će se standardi koristiti pri tumačenju podataka)</p>	<p>Prometni podatci sadržavat će popratni tekstualni dokument koji opisuje sve pojedinosti postupka simulacijskog eksperimenta i značajke generiranih FCD podataka uključujući identifikaciju svakog skup podataka prema jedinstvenom slučajnom sjemenu. Završni skup podataka pohranit će se u odabranom repozitoriju, popraćen dokumentom nazvanim README sa sadržajem svih datoteka kao i koja je konvencija korištena za njihovo imenovanje.</p> <p>Metapodatci za ocenaografske podatke su naziv korištene skripte za dohvat podataka, te područje, rezolucija i vremenski raspon.</p>
<p>2. Pravna i sigurnosna pitanja</p>	
<p>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)?</p>	<p>Simulacijski model grada Dublina „Dublin - Irish Motorway, National road and city centre“ korišten za generiranje podataka je javno dostupan na poveznici <a href="https://sumo.dlr.de/docs/Data/Scenarios.html">https://sumo.dlr.de/docs/Data/Scenarios.html</a>, a autore/stvaratelje modela može se citirati kako je navedeno u README dokumentu tog modela.</p> <p>Sirovi ocenaografski satelitski podatci već su javno dostupni u repozitorijima Copernicus i Oceancolour.</p>
<p>Kako će se regulirati pristup podacima i njihova sigurnost? Koji su potencijalni rizici koje treba uzeti u obzir? Kako ćete osigurati sigurnost pohrane</p>	<p>Podaci će se pohraniti u centraliziranom sustavu za pohranu kojim se službeno koriste Sveučilišne institucije u RH (Microsoft OneDrive). Pristup podacima omogućit će se preko pripadajućeg servisa za dijeljenje podataka u oblaku.</p>

	osjetljivih podataka?	
	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?	<p>Nema osjetljivih podataka.</p> <p>Simulacijski model je javno dostupan, a autorska prava vrijede kao za sve znanstvene radove. Generirani podaci su prikladni za dijeljenje. Podaci su dobiveni naknadnim varijacijama simulacijskih scenarija umjerenog mikroskopskog prometnog modela grada Dublina te kao takvi mogu biti korišteni za druge analize.</p> <p>Sirovi ocenaografski satelitski podatci već su javno dostupni u repozitorijima Coperinicus i Oceancolour. te prikladno citirani u objavljenim publikacijama.</p>
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> )?	Podatci će biti pohranjeni na dijeljenom serveru, te na lokalnom serveru/računalu na kojem se budu obrađivali.
	Koji je vaš plan čuvanja podataka? U kojim će se formatima čuvati?	Plan je čuvati podatke u izvorišnom obliku odnosno netCDF/CSV formatu. Međukorake čuvamo u PKL formatu. Podatke u pravilu čuvamo dok se sa njima radi, odnosno dok se ne objave publikacije. Trajnija pohrana i javna dostupnost ocanografskih podataka nije planirana s obzirom da im se može pristupiti kroz postojeće repozitorije. Trajnija pohrana i javna dostupnost prometnih podataka ovisit će o tome koliki overhead predstavlja održavanje tog sustava.
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?	Projekt ne predviđa dijeljenje podataka. Članovi tima imaju pristup podacima kroz interne mehanizme institucije.
	Ako postoje podaci koji se ne smiju dijeliti (prijavitelji vezani zakonskim, etičkim, autorskim pravila, povjerljivošću i sl.), pojasnite razloge ograničenja.	
	Potvrdite da ćete se koristiti digitalnim repozitorijem koji je u skladu s načelima <i>FAIR-a</i> .	Microsoft OneDrive nije prikladan repozitorij za FAIR princip.

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).	Microsoft OneDrive je komercijalan proizvod.
--	--

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 \(unizg.hr\)](#)