

# Plan upravljanja istraživačkim podacima projekta ALFABEATS

---

Jović, Alan

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

*Publication year / Godina izdavanja:* **2025**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:168:187934>

*Rights / Prava:* [Public Domain Dedication](#)/[Prenošenje u javno dobro](#)

*Download date / Datum preuzimanja:* **2025-02-06**



*Repository / Repozitorij:*

[FER Repository - University of Zagreb Faculty of  
Electrical Engineering and Computing repository](#)



## PLAN UPRAVLJANJA ISTRAŽIVAČKIM PODACIMA (PUP)

Opće informacije		
	Ime i prezime predlagatelja	Alan Jović
	Matična organizacija	Sveučilište u Zagrebu Fakultet elektrotehnike i računarstva
	Naziv projekta	Algoritmi za učinkovitu analizu biomedicinskih vremenskih nizova
	Upravitelj podacima	Alan Jović
1.	Prikupljanje podataka i dokumentacija	
	Koje ćete podatke prikupljati, obrađivati, stvarati ili se ponovno njima koristiti? (navedite formate, vrste i opseg svih podataka s kojima ćete raditi, a ne samo krajnji skup podataka koji će biti rezultat istraživanja)	<p>Prikupljat će se slobodno dostupni zapisi biomedicinskih vremenskih nizova iz internetskih izvora kao što su PhysioNet [<a href="https://physionet.org/">https://physionet.org/</a>], George B. Moody PhysioNet Challenges [<a href="https://moody-challenge.physionet.org/">https://moody-challenge.physionet.org/</a>], EEG-Datasets [<a href="https://github.com/meagmohit/EEG-Datasets">https://github.com/meagmohit/EEG-Datasets</a>] i drugih koji se nalaze u različitim formatima datoteka (npr. .txt, .csv, .edf).</p> <p>Također, prikupljat će se zapisi biomedicinskih vremenskih nizova iz lokalnih medicinskih ustanova: Poliklinike za prevenciju kardiovaskularnih bolesti i rehabilitaciju „Srčana“ [<a href="https://www.srcana.hr/">https://www.srcana.hr/</a>] i Klinike za psihijatriju "Vrapče" [<a href="https://bolnica-vrapce.hr/">https://bolnica-vrapce.hr/</a>] koji su pohranjeni u nekoliko formata datoteka.</p> <p>Ukupno se planira prikupiti najmanje sedam različitih skupova podataka od najmanje dvije vrste biomedicinskih vremenskih nizova (npr. EKG, EEG).</p>
	Kako će se podaci prikupljati, obrađivati ili stvarati? (ukratko navedite metodologiju i procese osiguranja kvalitete, načine organiziranja podataka te alate i instrumente kojima ćete se koristiti za prikupljanje i obradu)	<p>Podaci će se prikupljati na dva načina: 1) preuzimanjem datoteka iz baza podataka od internetskih izvora podataka i 2) dobavljanjem podataka na digitalnom mediju (hard disk, optički disk, USB memorija) od lokalnih medicinskih ustanova. Internetski izvori podataka sadržavaju samo anonimizirane zapise. Prikupljeni podaci od lokalnih medicinskih ustanova će biti prethodno anonimizirani od strane medicinskih stručnjaka za cjelokupne potrebe projekta.</p> <p>Jedan od ciljeva projekta je izgradnja i organizacija lokalne banke podataka za potrebe učenja i testiranja algoritama strojnog učenja. Stoga će se prikupljeni podaci na odgovarajući način (ovisno o vrsti podataka i izvoru podataka) organizirati po mapama u datotečnom sustavu.</p>
	Koju ćete dokumentaciju i metapodatke izraditi osim podataka? (dokumentacija mora sadržavati informacije i standarde potrebne korisnicima kako bi mogli samostalno čitati i interpretirati podatke u budućnosti, primjerice, kodne knjige, <i>ReadMe</i> datoteke i sl.)	<p>Izradit će se datoteka u kojoj će biti detaljno opisani svi prikupljeni skupovi podataka, uključujući navođenje izvora, vrste biomedicinskog vremenskog niza, primarnu namjenu skupa podataka (za kakvu vrstu analize će se koristiti), statističke podatke o zapisima (npr. broj zapisa, prosječno trajanje zapisa, frekvencija uzorkovanja zapisa). Za svaki zapis postojat će evidencija u vidu njegovog rednog broja (koda), poremećaja/zdravstvenih stanja koji su u njemu prisutni, demografskih podataka o pacijentu (ako je dostupno, npr. dob, spol) i medicinskih podataka o pacijentu (ako je dostupno, lijekovi koje pacijent uzima, postupke kojima je podvrgnut).</p>

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 obrađuju informirani o tome i jesu li dali privolu? Kojim ćete se metodama koristiti u svrhu zaštite osjetljivih podataka (GDPR - posebne kategorije osobnih podataka, navesti metode anonimizacije podataka)?	Projekt nije ograničen sporazumom o povjerljivosti, ali su lokalne medicinske ustanove ovlaštene za prikupljanje, obradu, čuvanje i dijeljenje podataka u skladu s etičkim normama u biomedicini. Stoga će prikupljeni podaci biti anonimizirani tako da članovima projektne grupe koji će provoditi analizu izvan ustanove gdje su podaci snimljeni neće biti poznati osobni podaci pacijenata. Liječnici na projektu će imati pristup svim informacijama o pacijentima kao i prikupljenim zapisima u ustanovi gdje rade, no oni osobne informacije neće odavati, već će s ostalim članovima grupe kao i s izgrađenim programskim rješenjem dijeliti samo anonimizirane zapise pacijenata, bez osobnih identificirajućih podataka, kako bi se onemogućilo saznavanje osjetljivih medicinskih informacija.
	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?	Pristup anonimiziranim podacima koji će biti prikupljeni unutar projekta imat će svi članove projektne skupine za potrebe izrade adekvatnih modela strojnog učenja. Podaci će se nalaziti na nabavljenoj radnoj stanici za analizu biomedicinskih vremenskih nizova i na lokalnim računalima članova projektne grupe na matičnoj organizaciji. Potencijalni rizik je neovlašteni pristup na ta računala i preuzimanje tih podataka. Osim sigurnosne lozinke za pristup operacijskom sustavu na svakom od tih računala i lozinke za udaljeni pristup radnoj stanici, sigurnost će se dodatno osigurati tako što će pristup mapi s podacima na svim računalima suradnika i na radnoj stanici biti kriptiran. Za vrijeme učenja modela strojnog učenja pristup će se po potrebi privremeno otvoriti kako bi se mogli izgraditi modeli.
	Kako ćete upravljati zaštitom autorskih prava i drugog intelektualnog vlasništva? Tko će biti vlasnik podataka? Koje će se licencije primjenjivati na podatke? Koja će se ograničenja primjenjivati na ponovnu uporabu osobnih podataka?	Zapisi iz internetskih baza podataka su javno dostupni tako da ne podlažu autorskim pravima i drugom intelektualnom vlasništvu, osim na način da ako se na temelju njih objavi istraživački članak, da se onda internetski izvor u njemu citira. Vlasništvo tih podataka ovisi od izvora do izvora, no često je vlasnik podataka sam internetski portal. Podaci iz lokalnih medicinskih ustanova vlasništvo su lokalnih medicinskih ustanova. Stoga, prikupljeni zapisi koji će biti organizirani u lokalnu banku podataka sa skupovima podataka o biomedicinskim vremenskim nizova neće biti vlasništvo matične organizacije niti Hrvatske zaklade za znanost, već će ti podaci ostati vlasništvo odgovarajućih izvora. Lokalna banka podataka koristit će se za potrebe projekta u svrhu izgradnje modela strojnog učenja, bez dijeljenja s vanjskim korisnicima izvan projekta.
3.	Pohrana i čuvanje podataka	
	Kako će radne verzije podataka biti pohranjene tijekom projekta? Kako će se napraviti sigurnosne kopije tih podataka ( <i>backup</i> )? Koja je očekivana količina podataka koja će se prikupiti i čuvati tijekom projekta (izraženo u MB/GB/TB)?	Podaci u svim verzijama, uključujući i konačnu verziju banke podataka, će se nalaziti na radnoj stanici za analizu biomedicinskih vremenskih nizova i na lokalnim računalima članova projektne grupe na matičnoj organizaciji. Tijekom projekta radić će se povremene sigurnosne kopije tih podataka na vanjski hard disk. Očekivana količina biomedicinskih podataka koja će se prikupiti na projektu je reda veličine nekoliko desetaka gigabajta.

<p>Kako će se završne verzije podataka dugotrajno pohraniti i čuvati (i nakon završetka projekta)? U kojim će se formatima čuvati podaci? Koja je očekivana količina podataka koja će se trajno pohraniti (izraženo u MB/GB/TB)?</p>	<p>Završna verzija banke podataka sačuvat će se po završetku projekta najmanje 5 godina na tri mjesta: na radnoj stanici za analizu biomedicinskih vremenskih nizova, na pomoćnom poslužitelju za koji je zadužen voditelj projekta i na jednom vanjskom hard disku. U slučaju kvara nekog od ova tri mjesta i nepovratnog uništenja podataka, osigurat će se zamjensko treće mjesto za čuvanje podataka (npr. novi poslužitelj ili novi vanjski hard disk). Kao i tijekom trajanja projekta, pristup mapi s podacima nakon završetka projekta bit će kriptiran. Podaci će se čuvati u izvornim formatima preuzetih zapisa. Očekivana količina biomedicinskih podataka koja će se trajno pohraniti je reda veličine nekoliko desetaka gigabajta.</p>
<p>4. Dijeljenje i ponovna uporaba podataka</p>	
<p>Kako i gdje će se podaci dijeliti? Koji repozitorij će se koristiti za dijeljenje podataka? Kako će potencijalni korisnici doznati za podatke?</p>	<p>Prikupljeni podaci se neće dijeliti s korisnicima izvan projektne grupe.</p>
<p>Ako postoje podaci koji se ne smiju dijeliti (prijavitelji vezani zakonskim, etičkim, autorskim pravila, povjerljivošću i sl.), pojasnite razloge ograničenja.</p>	<p>S obzirom da skupovi podataka dolaze iz različitih izvora, a vlasništvo podataka iz tih izvora nije od matične organizacije niti od Hrvatske zaklade za znanost, prikupljeni podaci se neće dalje dijeliti. Iznimno, podaci prikupljeni od lokalnih medicinskih ustanova će se dijeliti s vanjskim korisnicima samo u slučaju odobrenja medicinskih ustanova tijekom provedbe projekta.</p>
<p>Potvrdite da ćete se koristiti digitalnim repozitorijem koji je u skladu s načelima <i>FAIR-a</i>.</p>	<p>Potvrđujemo da će se za pohranu podataka koristiti digitalni repozitorij koji je u skladu s načelima <i>FAIR-a</i>.</p>
<p>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).</p>	<p>Potvrđujemo da će se za pohranu podataka koristiti digitalni repozitorij koji održava neprofitna organizacija. Podaci prikupljeni od lokalnih medicinskih ustanova će se dijeliti s vanjskim korisnicima samo u slučaju odobrenja medicinskih ustanova tijekom provedbe projekta.</p>