

Plan upravljanja istraživačkim podacima

Grahovac, Danijel

Data management plan / Plan upravljanja istraživačkim podacima

Publication year / Godina izdavanja: **2024**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:126:327096>

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

Download date / Datum preuzimanja: **2024-12-24**



mathos

Repository / Repozitorij:

[Repository of School of Applied Mathematics and Informatics](#)



PLAN UPRAVLJANJA ISTRAŽIVAČKIM PODACIMA (PUP)

Opće informacije		
	Ime i prezime predlagatelja	Danijel Grahovac
	Matična organizacija	Odjel za matematiku, Sveučilište J. J. Strossmayera u Osijeku
	Naziv projekta	Skaliranje u stohastičkim modelima (IP-2022-10-8081)
	Upravitelj podacima	Danijel Grahovac
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>U okviru projekta planirano je obrađivanje postojećih podataka o EEG signalima. Radi se o podacima prikupljenoj na djeci iz Ugande za vrijeme stanja kome izazvane cerebralnom malarijom. Osim toga, za svakog ispitanika prikupljeni su podaci o neurokognitivnom stanju šest mjeseci nakon buđenja iz kome. Navedeni podaci prikupljeni su u periodu između 2008. i 2015. godine i ustupljeni na korištenje projektnom timu od strane Prof. Alla Sikorskii, Michigan State University. Podaci sadrže EEG signale djece koji su snimljeni pomoću 19 elektroda postavljenih na određena područja na tjemenu. Format korištenih podataka je .edf, dok je pripadni opseg približno 3 GB.</p> <p>Osim toga, stvarat će se podaci temeljem računalnih simulacija, dok će se njihova obrada koristiti u smjeru demonstriranja svojstava i procjene parametara korištenih modela. U tu svrhu, kreirani podaci bit će spremni u .txt, .xlsx ili .nb formatu. Očekivani opseg podataka je u desecima gigabajta (GB).</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)	Prilikom stvaranja novih podataka putem simulacija koristiti će se programski paketi R i Mathematica, a koji su besplatno dostupni (R) ili su dostupni projektnom timu na matičnoj organizaciji (Mathematica). Također, vodit će se računa da kreirani podaci omogućuju reproducibilnost numeričkih rezultata što će osigurati kvalitetu dobivenih rezultata. Podaci će biti pažljivo organizirani u ovisnosti o tome koriste li se u svrhu procjene parametara, trajektorijskih svojstava modela, ili pak svojstava modela. Metodologija simuliranih i prikupljenih podataka podrazumijevat će standardne tehnike u aproksimacijama rješenja sustava stohastičkih diferencijalnih jednačbi (primjerice Eulerova metoda), rekurzivne metode, metode simulacija na temelju karakterističnih funkcija te Monte Carlo metode.
	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.)	Metapodaci podrazumijevaju dobivene procjene parametara korištenih modela na temelju kreiranih ili na korištenje ustupljenih podataka, a u svrhu evaluacije njihovih svojstava. Također, bit će korištene tablične i grafičke tehnike opisne statistike. Popratna dokumentacija bit će vođena u okviru korištenih kodova te njihovih pojašnjenja koji su doveli do metapodataka. Primjerice, kreirana softverska rješenja u programskom paketu R će u slučaju potrebe biti javno objavljena koristeći poznati R repozitorij CRAN. Metodologija obrade podataka bit će objašnjena u publikacijama koje će proizaći iz projekta.

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)?	Ne postoji ograničenje u vidu sporazuma o povjerljivosti. Podaci o EEG signalima ustupljeni su Prof. Alli Sikorskii kroz istraživački projekt američkog National Institutes of Health (R01 NS055349) na kojem je bila jedan od suradnika u svrhu istraživanja karakteristika i modeliranja prikupljenih EEG signala. A. Sikorskii slobodna je raspolagati podacima s ciljem provođenja istraživanja te ih ima dopuštenje dijeliti sa svojim suradnicima u okviru istraživačkih aktivnosti. Podaci su prikupljeni poštujući najviše standarde te su u potpunosti anonimizirani. Sukladno tome zaštita osjetljivih podataka u kontekstu GDPR-a je zagarantirana.
	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 podacima bit će dostupan samo onim članovima projekta koji sudjeluju na problemima koji uključuju njihovo izravno korištenje. To će biti osigurano putem korištenja zajedničkog privatnog repozitorija. Koristit će se repozitoriji najviše razine sigurnosti i standarda autentifikacije tako da će rizik kompromitacije podataka biti sveden na minimum. Osim toga, u upravljanju podacima prakticirat će se spremanje podataka na izvanmrežne lokalne sigurnosne kopije. Na taj način se izbjegavaju rizici te će biti osigurana sigurna obrada podataka.
	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?	Vlasnik podataka ustupio je na korištenje postojeće podatke projektnoj suradnici Alli Sikorskii koja je suglasna obrađivati ih s predloženim suradnicima u okviru projektnih aktivnosti. Na taj način ona će biti tretirana kao njihov indirektni vlasnik koji može dijeliti podatke sa suradnicima s ciljem istraživanja. S druge strane, simulirani podaci smatrat će se vlasništvom projektnog tima, odnosno dijela tima koji će raditi na njihovoj simulaciji. Voditelj projekta (D. Grahovac) vodit će posebnu pažnju o čuvanju i uporabi podataka kao njihov vlasnik.
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	Sve radne verzije podataka bit će paralelno pohranjene online kroz zajednički repozitorij te u vidu sigurnosne kopije na vanjske hard diskove. Navedena procedura osigurat će postojanje sigurnosnih kopija na više fizičkih i virtualnih mjesta. Očekivana količina podataka koje će se prikupiti i čuvati je u desecima gigabajta (GB).

	projekta (izraženo u MB/GB/TB)?	
	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)?	Završne verzije podataka dugotrajno će biti pohranjene u više izvanmrežnih kopija na vanjskim hard diskovima te na zajedničkom (online) repozitoriju. Formati podataka podrazumijevaju .xlsx, .txt, .nb, i .edf datoteke. Ukupna očekivana količina podataka mjerena je u desecima gigabajta (oko 30 GB).
4.	Dijeljenje i ponovna uporaba podataka	
	Kako i gdje će se podaci dijeliti? Koji repozitorij će se koristiti za dijeljenje podataka? Kako će potencijalni korisnici doznati za podatke?	Podaci će se među projektnim suradnicima dijeliti internetom putem online repozitorija OneDrive koji je dostupan matičnoj organizaciji. Podaci o EEG signalima nisu namijenjeni javnom dijeljenju. Informacije o podacima bit će dostupne kroz objavljene znanstvene radove koji se bave tematikom EEG signala te procjene parametara modela na kojima će raditi projektni tim. Na taj način će šira znanstvena zajednica imati informaciju o podacima i njihovoj primjeni u modeliranju. Nastojat će se koristiti objavljivanje u otvorenom pristupu kako bi vidljivost radova bila veća.
	Ako postoje podaci koji se ne smiju dijeliti (prijavitelji vezani zakonskim, etičkim, autorskim pravila, povjerljivošću i sl.), pojasnite razloge ograničenja.	Podaci o EEG signalima slobodno se mogu dijeliti među suradnicima na projektu u okviru istraživačkih aktivnosti usmjerenih analizi EEG signala. Podaci nisu namijenjeni javnom dijeljenju te ih nije planirano učiniti javno dostupnima s obzirom da su prikupljeni u okviru drugog projekta (National Institutes of Health projekt R01 NS055349).
	Potvrdite da ćete se koristiti digitalnim repozitorijem koji je u skladu s načelima FAIR-a.	Planirano istraživačke aktivnosti i upravljanje podacima bit će u skladu s FAIR načelima. Planirani OneDrive repozitorij osigurava transparentnost i usklađenost s načelima FAIR-a. Sva četiri aspekta načela FAIR-a (pronalaženje, dostupnost, interoperabilnost i upotrebljivost) osigurani su kroz planirani repozitorij. Preprint verzije publikacija dijelit će se na međunarodnim repozitorijima kao što je arXiv u skladu s politikama izdavača časopisa u kojima su radovi objavljeni.
	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).	U objavi preprint verzija publikacija koristit će se repozitorij arXiv koji održava neprofitna organizacija Cornell University Library. Za kreiranje softverska rješenja u programskom paketu R po potrebi će se koristiti repozitorij CRAN (The Comprehensive R Archive Network) koji održava R Foundation.