

# Plan upravljanja istraživačkim podacima - HybridScale

---

**Davidović, Davor**

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

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

*Permanent link / Trajna poveznica:* <https://urn.nsk.hr/urn:nbn:hr:241:444579>

*Rights / Prava:* [Attribution-NonCommercial-NoDerivatives 4.0 International/Imenovanje-Nekomercijalno-Bez prerada 4.0 međunarodna](#)

*Download date / Datum preuzimanja:* **2024-10-30**



*Repository / Repozitorij:*

[Fulir DATA - Ruđer Bošković Institute Research Data Repository](#)

## Plan upravljanja istraživačkim podacima

Opće informacije		
	Ime i prezime predlagatelja	Davor Davidović
	Matična organizacija	Institut Ruđer Bošković
	Naziv projekta	Skalabilni algoritmi visokih performansi za buduće heterogene distribuirane računalne sustave
	Upravitelj podacima	Davor Davidović, Centar za informatiku i računarstvo, IRB, Bijenička cesta 54, 10000, Zagreb
1.	Prikupljanje podataka i dokumentacija	
	Koje ćete podatke prikupljati, obrađivati, stvarati ili se ponovno njima koristiti? (navedite format, vrstu i opseg podataka)	<p>Tijekom projekta svi podaci koji se koriste u istraživanjima se generiraju:</p> <ol style="list-style-type: none"> <li>1. Različiti tipovi ulaznih matrica potrebnih za istraživanje. Matrice su različitih veličina (10,000 – 600,000 broj stupaca i/ili redaka), pune, rijetke, sa strukturom, simetrične, nesimetrične. Matrice se generiraju vlastitim programima napravljenima u C++ i Python programskim jezicima. Format zapisa je binarni s vlastitim formatiranjem</li> <li>2. Grubi podaci generirani iz računalnih eksperimenata (simulacija) na lokalnom klaster računalu i vanjskim računalima (super računala i računalni klasteri). Podaci sadrže informacije o vremenima izvršavanja pojedinih dijelova programa/računalnog koda, konfiguraciji (ulaznih parametara) svake simulacije, brzini, preciznosti i točnosti izračuna. Podaci se spremaju u .csv formatu zapisa.</li> <li>3. Finalni rezultati analize performansi koji sadrže grafičke prikaze rezultata i analiza testiranja performansi računalnih kodova. Podaci se zapisuju u .pdf, .png, .jpg, .svg i .eps formatima.</li> <li>4. Ulazni podaci koji se dobiju iz drugih grupa i vanjskih suradnika bit će obrađeni i pohranjeni u skladu s njihovim PUIP-ovima.</li> </ol>
	Kako će se podaci prikupljati, obrađivati ili stvarati? (ukratko navedite metodologiju i procese osiguranja kvalitete te načine organiziranja podataka)	<p>Podaci koji se koriste u ovom projektu prikupljaju se na slijedeći način:</p> <ol style="list-style-type: none"> <li>1. Matrice različitih tipova i veličina generiraju se programima koji su razvijeni na projektu ili preko vanjskih servisa. Programi su pisani u C++ i Python. Matrice se mogu generirati na osobnim računalima ili na vanjskih serverima.</li> <li>2. Grubi podaci su podaci prikupljeni iz izlaznih podataka različitih računalnih aplikacija i računalnih softvera i biblioteka. Ovisno o pojedinoj aplikaciji podaci se ili direktno upisuju u izlazne datoteke (.csv) ili se podaci sa standardnog izlaza parsiraju i spremaju u csv file-ove.</li> <li>3. Finalni rezultati se generiraju koristeći Octave, Matlab, Python ili JupyterLab iz grubih podataka koristeći specijalizirane skripte za generiranje izlaznih slika. Podaci se generiraju na lokalnim računalima. Format, tip i vrsta podataka koji se generira ovise o načinu prikaza.</li> </ol>

		4. Podaci se prikupljaju direktno od partnera i/ili drugih grupa uz dogovor o načinu korištenja i obradi podataka.
	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)	<ol style="list-style-type: none"> <li>1. Informacije o tipu matrice se dijeli kroz pridruženi tekstualni file ili ime file-a formatirano na način da opisuje o kojoj vrsti i veličine matrice je riječ. Nije predviđeno da se ovi podaci dijele van istraživačke grupe.</li> <li>2. Grubi podaci sadrže zaglavlja koja opisuju stupce u .csv file-ovima te su sastavni dio datoteka.</li> <li>3. Završi rezultati ne sadrže posebne informacije jer se ne planiraju dijeliti zasebno nego u sklopu predavanja (prezentacije), poster i znanstvenih članaka u kojima su onda detaljno dokumentirani dobiveni podaci.</li> </ol>
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)?	Podaci generirani ili prikupljeni ovim projektom nisu ograničeni sporazumom o povjerljivosti te ne postoje nikakav etički ili pravna pitanja oko korištenja podataka. Podaci ne sadrže privatne podatke te nije potrebno provoditi zaštitu osjetljivih podataka (GDPR).
	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 i obrada podataka (matrice i grubi podaci simulacija) bit će moguć samo članovima istraživačke grupe koristeći secure shell protocol (ssh). Obrada i pristup će se odvijati u zaštićenom mrežnom okruženju. Pristup finalnim podacima bit će moguće široj znanstvenoj zajednici kroz repozitorij podataka Instituta Ruđer Bošković. Podaci neće biti dostupni u izvornom obliku nego kroz objavljene radove/prezentacije.
	Kako ćete upravljati zaštitom autorskih prava i intelektualnog vlasništva? Tko će biti vlasnik podataka? Koje će se licencije	Svi podaci (uključujući i računalni kod dobiven kao rezultat projektnih aktivnosti) bit će javno dostupni pod nekom od otvorenih licenci (npr. GNU GPLv3 ili Creative Commons CCO). Vlasnik podataka je Institut Ruđer Bošković i autori podataka (članovi istraživačke grupe).

	primjenjivati na podatke? Koja će se ograničenja primjenjivati na ponovnu uporabu osobnih podataka?	Ponovna upotreba podataka će biti moguća ali uz uvjet da se primjenjuje ista licenca i uz navođenje izvornih autora podataka.
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> )?	<ol style="list-style-type: none"> <li>1. Generirane matrice bit će pohranjene na mrežnim diskovima računalnih sustava (super-računala, hpc klasteri, mrežni serveri) kojima članovi grupe imaju pristup. Međusobno dijeljenje je putem dijeljenih datotečnih sustava određenog računalnog sustava. Sam raspodijeljeni računalni sustav brine se za repliciranje podatka. U slučaju gubitka podataka, podaci se lako mogu ponovno generirati koristeći pripremljene skripte/programe.</li> <li>2. Skripte/programe za generiranje podataka za svaki pojedini eksperiment bit će dostupni putem GitHub platforme za dijeljenje izvornog code-a. Sigurnosne kopije</li> <li>3. Završni rezultati bit će dostupni kroz repozitorij otvorenih podataka IRB (Fulir - <a href="http://fulir.irb.hr/">http://fulir.irb.hr/</a>), stranicama izdavača i javno dostupni repozitorijima (npr. Arxiv), te pohranjeni na lokalnim računalima i mrežnim diskovima na Institutu Ruđer Bošković.</li> </ol>
	Koji je vaš plan čuvanja podataka? U kojim će se formatima čuvati?	Generirane matrice se ne planiraju čuvati nakon završetka eksperimenta jer ih je moguće jednostavno ponovno generirati. Objavljene podatke trajno ćemo čuvati u institutskom repozitoriju (Fulir) te na javno dostupnim repozitorijima (Arxiv, Zenodo).
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?	Konačna verzija podataka dijelit će se kroz institucionalni repozitorij (Fulir) te ostale otvorene repozitorije (Arxiv, Zenodo) a konačne verzije računalnog koda na platformama na dijeljenje izvornog code (GitHub). Podaci pod točkom 1. se ne planiraju dijeliti. Riječ je o velikim količinama umjetno generiranih podataka koje zauzimaju veće količine podatkovnog prostora te ih je moguće jednostavno re-kreirati koristeći programe/skripte za generiranje podataka. Pristup skriptama/programima za generiranje ulaznih podataka će biti javno dostupne kroz servise za objavu izvornog code-a (npr. GitHub).
	Ako postoje podaci koji se ne smiju dijeliti (prijavitelji vezani zakonskim, etičkim, autorskim pravila, povjerljivošću i sl.), pojasnite razloge ograničenja.	Ne postoje takvi podaci.

	Potvrdite da ćete se koristiti digitalnim repozitorijem koji je u skladu s načelima <i>FAIR-a</i> .	Svi istraživački podaci bit će objavljeni u otvorenim repozitorijima poput Fulira, GitHub-a i Zenoda te će se tako osigurati da podaci budu u skladu s FAIR principima.
	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).	Koristiti će se svi javno dostupni repozitoriji koji zadovoljavaju standarde sigurnosti, no međutim, moguće je koristiti i komercijalna rješenja (npr. GitHub) koja zahtijevaju plaćanje od strane ovog projekta, u slučaju da će takav model pridonijeti boljoj sigurnosti podatak i boljoj funkcionalnosti za projekt. Bez obzira na model korištenja digitalnih repozitorija, osigurat će se da su rezultati/podaci javno i besplatno dostupni široj zajednici.

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 \(unizg.hr\)](#)