

# Plan upravljanja istraživačkim podacima - PHATOME

---

**Kopriva, Ivica**

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

*Publication year / Godina izdavanja:* **2024**

*Permanent link / Trajna poveznica:* <https://urn.nsk.hr/urn:nbn:hr:241:042693>

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

*Download date / Datum preuzimanja:* **2024-09-08**



*Repository / Repozitorij:*

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

## PLAN UPRAVLJANJA ISTRAŽIVAČKIM PODACIMA (PUP)

Opće informacije		
	Ime i prezime predlagatelja	Ivica Kopriva
	Matična organizacija	Institut Ruđer Bošković
	Naziv projekta	Razvoj algoritama grupiranja podataka, regresije i izdvajanja značajki s primjenama u patologiji, oftalmologiji i metabolomics
	Upravitelj podacima	Ivica Kopriva
1. Prikupljanje podataka i dokumentacija		
	<p>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>	<p>Vrsta prikupljenih podataka određena je ciljevima projekta. Cilj C1 odnosi se na razvoj algoritama za grupiranje podataka i podacima prilagođenu linearnu regresiju. Dakle ovaj cilj se oslanja na prethodno prikupljene podatke.</p> <p>Cilj C2 odnosi se na semantičku segmentaciju hiperspektralne slike smrznute sekcije adenokarcinoma debelog crijeva u jetri bojane hematoksilin-eosinom. U tvrtki Photon etc. Montreal, Kanada (<a href="https://www.photonetc.com">https://www.photonetc.com</a>) snimljeno je 27 hiperspektralna slika smrznutih patohistoloških rezova prikupljenih intraoperativno od 14 pacijenata. Preparati su prethodno prikupljeni kroz HZZ projekt IP-2016-06-5235. Hiperspektralne slike su pohranjene u HDF5 formatu koji je općeprihvaćeni standard za spremanje i organizaciju velikih skupova podataka. Svakoj slici pridružena je binarna slika s tumorskim elementima dobivena glasanjem većine od tri medicinska eksperta.</p> <p>Cilj C3 odnosi se na semantičku segmentaciju RGB slike patohistoloških rezova bojanih hematoksilin-eosinom, s dijagnozama hepatocelularnog karcinoma, kolangiocelularnog karcinoma i adenokarcinoma debelog crijeva u jetri. Sve RGB slike, te pripadajuće binarne slike s tumorskim elementima dobivene glasanjem većine od tri medicinska eksperta biti će organizirane u zajedničku datoteku u HDF5 formatu.</p> <p>Cilj C4 odnosi se na metodologiju identifikacije metabolita prisutnih u malim udjelima u <sup>1</sup>H NMR spektrima urina 62 osobe sa dijabetesom tipa 2 i 62 osobe kontrolne skupine. Uz snimljene spektre uzoraka urina prikupljenih tijekom projekta IP-2016-06-5235, od ključne važnosti biti će biblioteka sa 290 do 330 <sup>1</sup>H NMR spektara metabolita očekivano prisutnih u humanom urinu, od koji će njih više od 100 biti indikativno za dijabetes tipa 2. I spektri uzoraka urina i spektri metabolita biti će spremljeni u standardnom FID (free induction decay) formatu, a također će biti organizirani u ASCII formatu i "mat" kako bi bili čitljivi za širi krug korisnika.</p> <p>Cilj C5 odnosi se na metodologiju poboljšanja kvalitete i izdvajanja značajki iz 3D OCT slike retine, a za dijagnostiku nekih čestih bolesti retine. Za tu će se svrhu koristiti 3D OCT slike raspoložive na javno dostupnim repozitorijima. Po potrebi koristiti će se i slike snimljene na Sveučilištu Soochow, NR Kina. U tome slučaju 3D OCT slike će biti u "mat" formatu s obzirom da će biti obrađivane u programskom okruženju MATLAB.</p>

<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>	<p>Vežano za cilj C2, patohistološki rezovi adenokarcinoma debelog crijeva u jetri prikupljeni su intraoperativno od 19 pacijenata u okviru HZZ projekta IP-2016-06-5235. U tvrtki Photon etc. Montreal, Kanada (<a href="https://www.photonetc.com/">https://www.photonetc.com/</a>) snimljeno je 27 hiperspektralna slika od 14 pacijenata. Slike su pohranjene u standardnom HDF5 formatu. Kroz predloženi projekt generirati će se korespondentne pseudo-RGB slike potrebne patolozima za označavanje tumorskih slikovnih elemenata. Koristeći pseudo-RGB slike tri medicinske eksperta će uz pomoć programskog alata razvijenog u okviru HZZ projekta IP-2016-06-5235 označiti tumorske slikovne elemente. Iz binarnih slika pojedinih označivača će glasanjem većine biti dobivene konačne labele za učenje i vrednovanje algoritama semantičke segmentacije hiperspektralne slike. Sve binarne slike biti će zajedno s hiperspektralnim slikama i pseudo-RGB slikama spremljene u HDF5 formatu. Na taj način će uz odgovarajuću dokumentaciju i spremanje na javno dostupan repozitorij podataka biti dostupne za dalju obradu ostalim znanstvenicima.</p> <p>Vežano za cilj C3, patohistološki rezovi hepatocelularnog karcinoma, kolangiocelularnog karcinoma i adenokarcinoma debelog crijeva u jetri biti će pribavljeni iz parafinskih blokova u repozitorijima Zavoda za citologiju i patologiju KBC Zagreb i KB Dubrava. Nakon bojanja hematoksilin-eosinom i fiksacije rezova na stalca u svjetlosnom mikroskopu biti će snimljene dvije RGB slike s tumorskim žarištem po preparatu (pacijentu). Snimanje će biti pri ukupnom optičkom pojačanju 400 puta (40 puta na objektivu i 10 puta na okularu). Koristeći programski alat razvijen u okviru HZZ projekta IP-2016-06-5235, dva patologa će označiti tumorske slikovne elemente. Iz binarnih slika pojedinih označivača će glasanjem većine biti dobivene konačne labele za učenje i vrednovanje algoritama semantičke segmentacije RGB slike. Sve binarne slike biti će zajedno s RGB slikama spremljene u HDF5 formatu. Na taj način će uz odgovarajuću dokumentaciju i spremanje na javno dostupan repozitorij podataka biti dostupne za dalju obradu ostalim znanstvenicima. Ako bude odobrena kupnja skenirajućeg mikroskopa (stavka 3.2 u financijskom planu projekta) biti će napravljena potpuno digitalizacija pripremljenih patohistoloških rezova.</p> <p>Vežano za cilj C4, biti će snimljeno 80 do 120 <sup>1</sup>H NMR spektara metabolita prisutnih u urinu osoba sa ili bez dijabetesa tipa 2. Metodologija pripreme uzorka i snimanja spektara je standardizirana i opisana u prethodnim publikacijama objavljenim u okviru HZZ projekta IP-2016-06-5235. Prema metodologiji opisanoj u istim publikacijama snimljeni su i <sup>1</sup>H NMR spektri urina 62 osobe s dijabetesom tipa 2 i 62 osobe kontrolne skupine. Spektri metabolita kombinirati će se sa bibliotekom od 210 <sup>1</sup>H NMR spektara metabolita snimljenih tijekom HZZ projekta IP-2016-06-5235 u novu biblioteku veličine 290 do 330 <sup>1</sup>H NMR spektara metabolita očekivano prisutnih u humanom urinu, od koji će njih više od 100 biti indikativno za dijabetes tipa 2. Ova biblioteka biti će raspoloživa u standardnom FID (free induction decay) formatu, a također će biti organizirana u ASCII formatu i "mat" formatu. To će nakon stavljanja biblioteke na javno dostupan repozitorij zajedno s odgovarajućom dokumentacijom omogućiti dalju uporabu za širi krug korisnika. Na ekvivalentan će se način postupiti i sa prethodno snimljenim <sup>1</sup>H NMR spektrima urina.</p> <p>Vežano za cilj C5, ako bude potrebno koristiti 3D OCT slike nekih čestih bolesti retine koje su snimljene na Sveučilištu Soochow, Suzhou, NR Kina, slike će se spremi u nekom od standardnih formata, npr ".mat" za 3D OCT sliku ili pojedinačni skanovi u JPEG formatu (vidi: <a href="https://data.mendeley.com/datasets/rscbjbr9sj/2">https://data.mendeley.com/datasets/rscbjbr9sj/2</a>).</p>
<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>	<p>Uz sve podatke proizvedene u okviru predloženog projekta biti će dan opis u PDF dokumentu. On će sadržavati informacije o njihovom prikupljanju i snimanju, te formatu i organizaciji u kojoj su spremljeni. Također, će se dati kod u programskom okruženju MATLAB koji će omogućiti pristup podacima te, posljedično, njihovu uporabi za dalju obradu.</p>
<p>2.</p>	<p>Pravna i sigurnosna pitanja</p>

	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)?	Svi podaci koji će se koristiti su ili prošli postupak anonimizacije ili će za vrijeme prikupljanja biti anonimizirani. Također, prethodno prikupljeni podaci u sklopu HZZ projekta IP-2016-06-5235 prikupljeni su uz potrebna dopuštenja uz suglasnost etičkog povjerenstva odgovarajuće ustanove. Podatci koji se planiraju prikupiti u okviru predloženog projekta (RGB slike patohistoloških rezova) biti će prikupljeni iz parafinskih blokova repozitorija Zavoda za citologiju i patologiju KBC Zagreb i KB Dubrava. S obzirom da je prikupljanje retrospektivno (uzorci su spremljeni u parafinske blokove nekoliko godina ranije) pismena informirani pristanak nije potreban. Podatci koji će se koristiti ili će biti prikupljeni za korištenje ne spadaju u kategoriju posebno osjetljivih podataka i često se koriste sa javno dostupnih repozitorija podataka.
	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?	Podatci koji će se koristiti ili će biti prikupljeni za korištenje ne spadaju u kategoriju posebno osjetljivih podataka i često se koriste sa javno dostupnih repozitorija podataka. Sigurnost pohrane će biti osigurana kroz pohranu podataka na javno dostupne repozitorije koji vode brigu o sigurnosti. Primjeri takvih repozitorija su "Fulir DATA" Instituta Ruđer Bošković ( <a href="https://data.fulir.irb.hr/">https://data.fulir.irb.hr/</a> ) i podatkovni repozitorij "Mendeley Data" ( <a href="https://data.mendeley.com/">https://data.mendeley.com/</a> )
	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?	Svim podacima koji će se koristiti na projektu biti će dozvoljen pristup bez ograničenja nakon što bude objavljen znanstveni rad u kojem se ti podaci koriste. Od budućih korisnika podataka se očekuje da citiraju izvorni rad u kojem su ti podaci prvi puta korišteni. Time se osigurava zaštita autorskih prava. Kako je već prethodno opisano svi podaci su ili će biti potpuno anonimizirani te u njima neće postojati osobni podaci.
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 će tijekom projekta biti pohranjeni na lokalnim računalima, a sigurnosne kopije će biti pohranjene na vanjskim memorijskim jedinicama. Očekivana količina podataka je: 15 GB za hiperspektralne slike smrznutih patohistoloških rezova adenokarcinoma debelog crijeva u jetri (cilj C2); 10 GB za RGB slike patohistoloških rezova hepatocelularnog karcinoma, kolangiocelularnog karcinoma i adenokarcinoma debelog crijeva u jetri; 5 GB za 3D OCT slike retine i 100 MB za <sup>1</sup> H NMR spektre metabolita i uzoraka urina kontrolne skupine i skupine sa dijabetesom tipa 2.
	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)?	Ne postoji vremensko ograničenje pohrane i čuvanja podataka. Formalno gledano ono će biti ograničeno politikom javnih podatkovnih repozitorija "Mendeley data" i/ili "Fulir DATA". Formati i veličine podataka su već prethodno navedeni.
4.	Dijeljenje i ponovna uporaba podataka	

<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>Nakon pohrane na javno dostupne repozitorije "Mendeley data" i/ili "Fulir DATA" podaci će biti svima dostupni bez ograničenja. Potencijalni korisnici će za podatke doznati kroz radove objavljene u međunarodno poznatim znanstvenim časopisima.</p>
<p>Ako postoje podaci koji se ne smiju dijeliti (prijavitelji vezani zakonskim, etičkim, autorskim pravila, povjerljivošću i sl.), pojasnite razloge ograničenja.</p>	<p>Neće postojati podaci koji se ne smiju dijeliti.</p>
<p>Potvrdite da ćete se koristiti digitalnim repozitorijem koji je u skladu s načelima <i>FAIR</i>-a.</p>	<p>Potvrđujem da ću koristiti digitalni repozitorij koji je u skladu s načelima <i>FAIR</i>-a.</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>Potvrđujem da ću se koristiti digitalnim repozitorijem koji održava neprofitna organizacija.</p>