

Plan upravljanja istraživačkim podacima - EvoSat

Mravinac, Brankica

Data management plan / Plan upravljanja istraživačkim podacima

Publication year / Godina izdavanja: **2024**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:241:911371>

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

Download date / Datum preuzimanja: **2024-09-27**



Repository / Repozitorij:

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

Plan upravljanja istraživačkim podacima

Opće informacije		
	Ime i prezime predlagatelja	Dr. sc. Brankica Mravinac
	Matična organizacija	Institut Ruđer Bošković
	Naziv projekta	Rasvjetljavanje evolucije satelitnih DNA visokoprotočnim analizama satelitoma srodnih vrsta
	Upravitelj podacima	Brankica Mravinac, Brankica.Mravinac@irb.hr , voditeljica projekta
1.	Prikupljanje podataka i dokumentacija	
	Koje ćete podatke prikupljati, obrađivati, stvarati ili se ponovno njima koristiti? (navedite format, vrstu i opseg podataka)	<p>Projekt se zasniva na analizi satelitoma invertebratnih organizama, a istraživanja se provode na dva neovisna modelna sustava – vrstama kukaca iz roda <i>Tribolium</i> te vrstama školjkaša iz porodice Ostreidae. Tijekom projekta generirat ćemo i obrađivati dva različita tipa sirovih podataka, koji se odnose na oba modelna sustava:</p> <ol style="list-style-type: none"> 1. rezultate sekvenciranja DNA metodama nove generacije (Illumina i PacBio) kao i rezultate sekvenciranja Sanger metodom 2. slike metafaznih kromosoma izučavanih vrsta dobivenih konfokalnom mikroskopijom. <p>Za 1. kategoriju sirovi rezultati sekvenciranja bit će isporučeni u digitalnoj formi od strane centara u kojima će se vršiti usluga sekvenciranja i to u sljedećim formatima: rezultati Sanger sekvenciranja u fasta formatu, rezultati Illumina sekvenciranja u fastq formatu te rezultati PacBio sekvenciranja u BAM formatu. Procjenjujemo da ćemo za 1. kategoriju kumulativno generirati otprilike 8-9 TB podataka (500 MB podataka Sanger sekvenciranja, 20 GB podataka Illumina sekvenciranja te 8 TB podataka dobivenih PacBio sekvenciranjem).</p> <p>Za 2. kategoriju sirovi rezultati konfokalnih slika dobit će se i pohraniti u digitalnom obliku u TIFF formatu. Procjenjujemo da ćemo za pohranu konfokalnih slika trebati 1 TB prostora.</p>
	Kako će se podaci prikupljati, obrađivati ili stvarati? (ukratko navedite metodologiju i procese osiguranja kvalitete te načine organiziranja podataka)	<p>Rezultati sekvenciranja bit će organizirani na način da se pohranjuju u datoteke nazvane prema dogovorenim normama kako bi se nedvosmisleno mogao utvrditi naziv organizma kojem podaci pripadaju te vrsta i datum sekvenciranja kojim su podaci generirani. Skup podataka pratit će također i README dokumenti u kojima će se opisati hijerarhija direktorija, odnosno svi relevantni podaci za pojedini eksperiment koji je prethodio generiranju podataka. Daljnja analiza podataka primarno će se provoditi računalnim programima Geneious i RepeatExplorer. Program Geneious omogućava lokalnu pohranu generiranih rezultata (u različitim izlaznim formatima), dok se analize alatom RepeatExplorer odvijaju na serverskom poslužitelju Galaxy (https://repeatexplorer-elixir.cerit-sc.cz/) preko korisničkog računa. S obzirom na limitirani serverski prostor namijenjen pojedinom korisniku (200 GB), podaci dobiveni analizama „izvozit“ će se u lokalne direktorije osobnih računala pojedinih suradnika.</p> <p>Baze konfokalnih slika bit će hijerarhijski organizirane u datoteke poštujući sistematizaciju prema vrsti organizma->tipu analizirane satelitne DNA->datumu eksperimenta. Daljnje obrade slika radit će se u programima Adobe Photoshop, Inkscape i Gimpu. Proizvedene obrade slike pohranjivat će se inicijalno lokalno na računalima suradnika na projektu.</p>

		<p>Za sve eksperimente (izolacije DNA, odnosno fluorescencijske analize <i>in situ</i>), koji su prethodili krajnjem generiranju podataka, postojat će datirana zabilježba u laboratorijskim dnevnicima primarnih eksperimentatora s detaljnim opisom protokola prema kojem je eksperiment proveden te uz jasno istaknuta eventualna odstupanja od standardnog protokola, ako je do njih došlo.</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>Svi podaci bit će popraćeni dokumentacijom u vidu README dokumenta s objašnjenjima i detaljnim navođenjem imena vrste kojoj rezultati pripadaju, broja i razvojnog stadija analiziranih jedinki, metode sekvenciranja i imena platforme na kojoj su inicijalni rezultati/očitavanja (eng. <i>reads</i>) generirani. Procesirani podaci dobiveni daljnjim analizama bit će popraćeni README dokumentima s detaljnim objašnjenjima analiza, kao i specifikacijom parametara pod kojima su analize provedene. Prilikom objave radova za koje će biti potrebne kompleksnije, razvijene kodne knjige, iste će biti dostupne u skupu Github online alata koji omogućava i verzioniranje. Sirovi podaci koje budemo pohranjivali u javno dostupnim bazama podataka (primjerice, NCBI, ENA...) bit će pojašnjeni priloženim protokolima/kodnim knjigama dovoljnim za interpretaciju istih. Nadalje, skupovi podataka će u repozitorijima biti opisani metapodacima (detaljan opis svakog uzorka/stadija/tkiva/tipa sekvenciranja) i trajnim identifikatorima te će se na taj način omogućiti pretraživanje u databazama. U slučaju novo posloženih genoma ili opisanih sekvenci, iste će biti pripremljene kao dodatak radu u časopis na način na budu pohranjeni u repozitoriju.</p> <p>Za procesirane konfokalne slike bit će osigurana pojašnjenja priložena u vidu komentara, kojima će se specificirati eventualno poduzete intervencije u procesu analize slika, kao i njihovog procesiranja do publikabilne razine.</p>
2.	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 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>Pri izvedbi ovog projekta neće se kršiti etička načela. Također nismo ograničeni sporazumom o povjerljivosti.</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 osjetljivih podataka?</p>	<p>Pristup podacima koji su se koristili kako bi se dobili publikabilni rezultati bit će omogućeni po objavi rada u jednom u repozitorija ili na zahtjev s obzirom na veliku količinu podataka koja onemogućuje jednostavno dijeljenje preko mrežnih infrastruktura. Trajna pohrana velikih setova bit će izvršena na lokalnoj pohrani u najmanje dvije kopije kako bi se osigurala sigurna pohrana podataka i spriječio njihov eventualni gubitak.</p>

	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?	Sve podatke dobivene u sklopu istraživanja bit će licencirane licencom CC BY 4.0 po preporuci Europske komisije kako bi omogućili ponovno i ispravno korištenje podataka od strane drugih znanstvenika. Vlasnik podataka je Institut Ruđer Bošković, u kojem se realizira projekt i produciraju rezultati. S obzirom da nema osobnih i povjerljivih podataka osim intelektualnih dostignuća, nikakva ograničenja neće biti primjenjiva na podatke.
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>)?	Većina podataka čuvat će se kroz nekoliko razina/pohrana: 1) sirovi i procesirani podaci čuvat će se na lokalnim računalima sudionika projekta; 2) uz pohranu na lokalnim diskovima, s obzirom na veliku količinu prikupljenih podataka (>8 TB), kupljeni su dodatni vanjski tvrdi diskovi. Četiri diska od 5 TB služe za dodatnu sigurnosnu pohranu radnih podataka u tijeku analiza (dva za modul kukaca, dva za modul školjkaša). K tome., nabavljena su i dva diska velikog kapaciteta (8 TB) koja služe isključivo za pohranu sirovih podataka i finalnih procesiranih podataka – jedan je namijenjen za podatke vezane uz modul kukaca, a drugi uz modul školjkaša – za redovitu sigurnosnu pohranu podataka modula kukaca odgovorna je Brankica Mravinac, dok je za redovitu sigurnosnu pohranu podataka modula školjkaša odgovorna Eva Šatović Vukšić; 3) po potrebi, podaci od interesa izmjenjuju se između projektnih suradnika, a mogu se i pohranjivati na virtualnom poslužitelju MojOblak kojim raspolaze Institut Ruđer Bošković; 4) finalno, po objavi rezultata podaci će biti deponirani u depozitorijima javnih baza kao što su to NCBI, ENA, i FigShare.
	Koji je vaš plan čuvanja podataka? U kojim će se formatima čuvati?	Sigurnosne kopije podataka čuvat će se u komprimiranoj formi na vanjskim jedinicama za pohranu na neodređeni vremenski period, zajedno s metapodacima i kodnim knjigama kako bi ponovno korištenje podataka bilo moguće u bilo kojoj vremenskoj točki.
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?	Za ovaj projekt najprikladnije mjesto pohrane i dijeljenja sa znanstvenom zajednicom jesu javni repozitoriji National Center for Biotechnology Information (NCBI), European Nucleotide Archive (ENA) i FigShare, a korisnici će o postojanju tih podataka biti informirani u sklopu objavljenih publikacija. Također, temeljni i sirovi podaci bit će ustupljeni svim znanstvenicima koji traže dopuštenje korištenja, osobno ili na zahtjev preko za to osiguranih protokola za dijeljenje podataka (direktni FTP/SFTP transfer, ili putem programa za prijenos poput WinSCP ili FileZilla).
	Ako postoje podaci koji se ne smiju dijeliti (prijavitelji vezani zakonskim, etičkim, autorskim pravila, povjerljivošću i sl.), pojasnite razloge ograničenja.	Podaci neophodni za bilo koju publikaciju bit će dostupni u trenutku objavljivanja. Svi neobjavljeni podaci pohranit će se u repozitoriju na 24 mjeseca od završetka projekta.

	Potvrdite da ćete se koristiti digitalnim repozitorijem koji je u skladu s načelima <i>FAIR-a</i> .	Digitalni repozitorij koji će biti korišten prati sva 4 FAIR načela, Findable - dostupan na internetu poveznicama s objavljenih publikacija, Accessible - moguće je pristupiti uz stabilnu internetsku vezu, Interoperable - podaci se preuzimaju u standardnim formatima za pohranu podataka poput .zip, Reusable - svi metapodaci bit će objavljeni zajedno s glavnim dijelom podataka.
	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).	Potvrđujemo.

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 \(unizg.hr\)](#)