

# Plan upravljanja istraživačkim podacima POTATOSIGNALHUB

---

**Mihaljević, Snježana**

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

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

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:241:008033>

*Rights / Prava:* [Attribution-NonCommercial-ShareAlike 3.0 Unported/Imenovanje-Nekomercijalno-Dijeli pod istim uvjetima 3.0](#)

*Download date / Datum preuzimanja:* **2024-10-01**



*Repository / Repozitorij:*

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

## PLAN UPRAVLJANJA ISTRAŽIVAČKIM PODACIMA (PUP)

Opće informacije		
	Ime i prezime predlagatelja	Snježana Mihaljević
	Matična organizacija	Institut Ruđer Bošković
	Naziv projekta	Integrativna analiza signalnih puteva fitohormona uključenih u odgovor biljaka krumpira na infekciju viroidom vretenastoga gomolja krumpira
	Upravitelj podacima	Snježana Mihaljević, mihaljev@irb.hr
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>Tijekom projekta generirat ćemo tri različita tipa sirovih podataka:</p> <ol style="list-style-type: none"> <li>1. slike sa fluorescencijskog mikroskopa uzoraka listova krumpira</li> <li>2. podaci s real-time pcr uzoraka krumpira</li> <li>3. sirovi podaci transkriptomike uzoraka listova krumpira</li> </ol> <p>Svi podaci biti će pohranjeni u digitalnom obliku u formatu koji se dobije izravno s instrumenata (primjerice, <i>TIFF</i> obliku datoteka za slike dobivene fluorescencijskim mikroskopom; <i>MGXD</i> obliku za podatke dobivene na real-time pcr CFX96 Touch sustavu; <i>FASTAQ</i> obliku za sirove i obrađene podatke transkriptomike dobivene na Illumina platformi). Mjerenja i kvantifikacija slika/podataka (prikazi u obliku tablica) snimit će se u <i>EXCEL</i> obliku (za dugotrajnu pohranu, konvertirat će se u <i>CSV</i> oblik). Procjenjujemo da će za 1. kategoriju biti potrebno oko 4 GB, za 2. kategoriju 0.5 GB, a za 3. kategoriju oko 300 GB prostora.</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>Slike s fluorescencijskog mikroskopa pohranjuju se u datoteke, grupirane po tipu uzorka i kronološki. Obrada slika provodi se u aplikacijama sa slobodnim pristupom kao što su FIJI (Fiji Is Just ImageJ) i GIMP, ili programu Adobe Photoshop. Datoteke ćemo nazvati prema prije dogovorenim normama. Skup slika/podataka pratit će i Excel dokumenti s opisom uzoraka. Također, bit će zapisana i odstupanja od protokola i ostale korisne informacije.</p> <p>Analitički podaci za real time pcr prikupljaju se s instrumenta koji ih generira (BioRad CFX96 Touch) i obrađuje u matičnom programu (BioRad Maestro), u obliku Bio-Rad Gene Expression File (.mgxd).</p> <p>Sirovi podaci transkriptoma listova krumpira inficiranih viroidom PSTVd generirani su u NGS servisu i dostavljeni na tvrdom disku. Dodatna analiza podataka provodi se u programskom sučelju R, na računalima suradnika na projektu.</p> <p>Rezultati obrada podataka pohranjuju se lokalno na računalima suradnika na projektu, uz dodatno pohranjivanje na nekom od virtualnih poslužitelja (MojOblak Centra za znanstvene informacije IRB ili slično).</p> <p>U eksperimentima je uključena odgovarajuća kontrola, čime se osigurava valjanost podataka. U ovom projektu kontrolna skupina biljaka je inokulirana vodom (mock-inokulirane biljke), u odnosu na biljke inokulirane viroidom PSTVd. Dosljednost podataka procijenit će se usporedbom ponovljenih mjerenja.</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>Svi podaci bit će popraćeni detaljnim opisom svakog uzorka (vrsta, kultivar, organ, tretman, starost) i eksperimentalnog postupka (protokol za uzgoj biljaka, odnosno protokol za izolaciju RNA, sekvencioniranje, konstrukciju cDNA biblioteka). Prilikom trajnog pohranjivanja, sirovi podaci transkriptomike biti će popraćeni dodatnim objašnjenjima i u formatu prema standardima uobičajenim za odabranu javno dostupnu bazu podataka (European Nucleotide Archive (ENA), National Center for Biotechnology Information (NCBI) ili slično).</p> <p>Datoteka <i>README.txt</i> može se koristiti kao uhodani način za sve datoteke i mape koje obuhvaćaju projekt objašnjavajući kako su svi skupovi datoteka međusobno povezani, u kojem su formatu, te jesu li određene datoteke namijenjene zamjeni ostalih datoteka itd.</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)?	Pri izvedbi projekta PotatoSignalHub neće se kršiti etička načela. Projekt se provodi na biljkama isključivo u kontroliranim eksperimentalnim uvjetima, u Laboratoriju za kemijsku biologiju, IRB. Sporazum o povjerljivosti nije primjenjiv.
	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?	Podaci će se obrađivati i njima upravljati u zaštićenom nemrežnom okruženju koristeći se virtualnom desktop tehnologijom. Dio podataka bit će pohranjen u sustavu MojOblak kojim upravlja Odjelom za informatičku podršku IRB. Pristup podacima tijekom trajanja projekta imaju članovi istraživačkog tima. Pristup podacima koji su se koristili kako bi se dobili rezultati biti će omogućeni po objavi znanstvene publikacije u odabranom repozitoriju ili na zahtjev istraživačkoj grupi. Pristup podacima uključenim u izračun i dobivanje rezultata istraživanja objavljenih u znanstvenim publikacijama biti će javno dostupan s obzirom na standarde u polju istraživanja putem nekih od javnih baza podataka za istraživanje (npr. NCBI, ENA). Laboratorijski dnevnicu čuvaju se u laboratoriju glavnog istraživača. Dodatni elektronički podaci pohranit će se na računalu glavnog istraživača, koji tjedno izrađuje sigurnosne kopije. Trajna pohrana velikih setova biti će izvršena na vanjske diskove u najmanje dvije kopije kako bi se zaštitio integritet 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?	Ne očekuje se da će rezultat istraživanja dovesti do patenta. Ostali problemi intelektualnog vlasništva će se u rješavati prema preporukama Instituta Ruđer Bošković. Budući da podaci nisu podvrgnuti ugovoru, te se neće patentirati, objavit će se kao otvoreni podaci pod licencijom <i>Creative Commons</i> CCO po preporuci Europske komisije kako bi omogućili ponovno i ispravno korištenje podataka od strane drugih znanstvenika.
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)?	Jedna kopija radne verzije podataka biti pohranjena na lokalnim računalima kako bi se omogućio brz pristup i manipulacija s podacima. Podaci će se tijekom istraživanja s računala glavnog istraživača kopirati u nacionalni sustav za pohranu i dijeljenje podataka MojOblak (IRB) ili Puh ( <a href="https://www.srce.unizg.hr/puh">https://www.srce.unizg.hr/puh</a> ) koji članovima projektnog tima omogućava pristup aktualnoj verziji podataka i na kojem se dnevno automatizirano izrađuje sigurnosna kopija podataka. Osim toga, u laboratoriju imamo 3 vanjska diska od po 1 TB, što nam omogućuje da sve podatke pohranimo u najmanje dvije kopije. Najvažniji podaci biti će čuvani u najmanje tri kopije kako bi se mogli vjerno replicirati rezultati istraživanja, poglavito .fastq datoteke analize transkriptoma.
	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	Podatke ćemo trajno čuvati na virtualnom poslužitelju MojOblak IRB ili PUH (Srce) i također pohraniti u odgovarajući arhiv za podatke na kraju projekta. Tablične podatke čuvat ćemo u CSV obliku, a tekstualne u DOCX (Office Open XML) te PDF-A obliku. DOC oblik obavezno će se konvertirati u DOCX oblik. FASTQ datoteke biti će komprimirane putem GNU zip programa i pohranjene u GZ obliku. Očekivana količina podataka koja će se trajno pohraniti je oko 300 GB.

	(izraženo u MB/GB/TB)?	
4.	Dijeljenje i ponovna uporaba podataka	
	Kako i gdje će se se podaci dijeliti? Koji repozitorij će se koristiti za dijeljenje podataka? Kako će potencijalni korisnici doznati za podatke?	Podaci analize transkriptoma i dobiveni izračuni (tablice, grafički prikazi) biti će podijeljeni na javnim repozitorijima poput NCBI ili ENA, a korisnici će o postojanju tih podataka biti informirani u sklopu objavljenih publikacija.
	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 3 godine od završetka projekta.
	Potvrdite da ćete se koristiti digitalnim repozitorijem koji je u skladu s načelima FAIR-a.	Digitalni repozitorij koji će biti korišten podržava načela FAIR-a: Findable - dostupan na internetu poveznicama sa 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 biti će objavljeni skupa sa 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).	Jedna kopija digitalnih podataka će biti pohranjena na javno dostupnim nekomercijalnim baza Hrvatske znanstvene infrastrukture (Dabar, PUH).