Plan upravljanja istraživačkim podacima

|  |  |  |
| --- | --- | --- |
| Opće informacije | | |
|  | Ime i prezime predlagatelja | Dijana Škorić |
|  | Matična organizacija | Prirodoslovno-matematički fakultet |
|  | Naziv projekta | IPS-2020-01-2960, Visokoprotočno Nanopore-sekvenciranje mikrobnih genoma u rješavanju problema dijagnostike i epidemiologije biljnih patogena |
|  | Upravitelj podacima | Dijana Škorić <dijana.skoric@biol.pmf.hr> |
| 1. | Prikupljanje podataka i dokumentacija | |
|  | Koje ćete podatke prikupljati, obrađivati, stvarati ili se ponovno njima koristiti? (navedite format, vrstu i opseg podataka) | Tijekom projekta generirat ćemo 5 kategorija podataka:   1. Fotografije 2. Fotografije s terena   Fotografije simptomatičnih, asimptomatičnih i drugih relevatnih biljaka (poput korova), izvora za navodnjavanje bit će snimljene na terenu prilikom uzorkovanja. Fotografije će biti pohranjene u .*jpg* formatu te će zauzet otprilike 20 GB memorije   1. Fotografije elektroforetskih gelova   Fotografije elektroforetskih agaroznih gelova (kontrola PCR produkata) bit će generirane na uređaju Amersham ImageQuant 800 te pohranjene u *.jpg* formatu. Ove će fotografije zazuzeti otprilike 10 GB memorije.   1. Fotografije inokuliranih biljaka   U svrhu validacije dobivenih rezultata, inukulirat ćemo eksperimentalne biljne domaćine u kontroliranim uvjetima (staklenik) te će progres eksperimenta biti zabilježen fotografijama. Fotografije će biti pohranjene u .*jpg* formatu te će zauzeti otprilike 10 GB memorije.   1. Numeričke tablice   Popisi uzoraka zajedno s pisanim bilješkama s terena i izmjerenim koncentracijama ekstrahirane RNA bit će spremljeni u obliku excel pregledne tablice. Također, podaci o koncentraciji nukleinskih kiselina mjereni tijekom analitičkih protokola (Nanodrop, Qubit) za Nanopore sekvenciranje bit će zabilježeni u obliku *excel* tablica, te dugoročno pohranjeni u *csv* formatu. Tablice će zauzeti otprilike 5 GB memorije   1. Sirove i obrađene sekvence, te popratni podaci   Podaci dobiveni Nanopore ili Sanger sekvenciranjem bit će pohranjeni u *fasta* ili *fastq* formatima. Prilikom bioinformatičke obrade sirovih sekvenci generirat ćemo veliki broj popratnih dokumenata u različitim formatima *(sam, bam, txt, gz, csv*, ...). Ovaj set podataka zahtjeva najveću količinu memorije procijenjenu na minimalno 1TB   1. Laboratorijski dnevnici i registratori (bilješke, *lateral flow* test (LFT) trakice)   Sve bilješke istraživača nastale tokom izvedbi laboratorijskih protokola i bioinformatiče obrade podataka bit će zapisane u laboratorijskim dnevnicima. Protokoli unaprijed pripremljeni ili stvarani tijekom laboratorijskog rada bit će pohranjeni u registratoru projekta. Trakice brze virusne detekcije (*lateral flow test*) bit će također pohranjene u registratoru projekta.   1. Porcije biljnog materijala   Ostaci biljaka s terena čuvat će se također u laboratoriju dvojako. Zamrznute porcije biljnog materijala čuvat će se u ledenici (-20°C) laboratorija. Dodatno, porcije tkiva bit će pohranjene kao liofilizirano tkivo u frižideru (4°C) u epruvetama s kalcijevim karbonatom. |
|  | Kako će se podaci prikupljati, obrađivati ili stvarati? (ukratko navedite metodologiju i procese osiguranja kvalitete te načine organiziranja podataka) | Prilikom samog terenskog uzorkovanja svaki će uzorak dobiti jedinstvenu oznaku (šifru) koja će ga pratiti od pohrane do konačne obrade i rezultata. Oznake uzoraka dodjeljivat će se na način da je jasno definirano sljedeće: jedinstveni broj uzorka, lokacija s koje je uzorak uzet, tip materijala (biljka-kratica vrste ili tipa biljke, voda, kukci), godina uzorkovanja te da li je uzorak bio simptomatičan ili ne (primjer: 36ViT21S označava uzorak br. 36 prikupljen na lokaciji Vidovec, tkivo rajčice iz 2021. godine s potencijalnim simptomima viroza.)  Prikazanim oznakama bit će obilježeni pohranjeni sirovi uzorci, fotografije s terena, LFT-trakice, izolirana RNA, sekvence.  Ostale tablice, popisi i fotografije bit će imenovani na način da sadrže kratak opis sadržaja, akronim projekta (NanoPhyto), datum stvaranja ili zadnje izmjene te inicijale istraživača koji dokument kreira.  Budući da je osnovni cilj projekta detekcija genoma biljnih patogena, konačne sekvence bit će detaljno bioinformatički obrađene te procesuirane kroz program Geneious software v10.2.6 (Biomatters Ltd., Auckland, New Zealand) za koji je kupljena licenca u sklopu projekta. Dodatno, prisustvo detektiranih virusa validirat će se RT-PCR procedurom sa specifičnim početnicama, Sanger sekvenciranjem amplikona, te inokulacijom u kontroliranim uvijetima.  Dio osiguranja kvalitete je provjera sekvenci od strane znanstvenika koji uređuju NCBI-bazu podataka sekvenci GenBank jer se za svaku sekvencu koja se obrađuje predviđa unošenje u GenBank zbog dobivanja pristupnog broja (Accession Number) pri čemu ona prolazi eksterne i objektivne postupke kontrole kvalitete. |
|  | 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) | Primarni način na koji ćemo ustupiti podatke javno biti će u znanstvenim publikacijama. Izvori i način prikupljanja i obrade podataka biti će detaljno objašnjeni u dijelu o materijalima i metodama, a rezultati će biti prikazani slikama, tablicama, plotovima te pojašnjeni tekstualno i legendama. Sekvence svih detektiranih virusa ili viroida bit će pohranjene online u Nacionalni centar za biotehnološke informacije (NCBI) s detaljima prikupljanja, obrade i anotacijama. Jedinstveni, dodjeljeni broj sekvence (tzv. accession number) bit će objavljen u sklopu publikacije te na taj način javno dostupan za pregledavanje i korištenje. Tijekom projekta bit će u izradi i diplomski rad studenta iz teme projekta na hrvatskom jeziku uz još detaljnije objašnjene postupke dolaska do podataka te uz objavljene dodatne rezultate koji nisu postali dio objave u znanstvenom radu. |
| 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)? | Budući da ne radimo s ljudskim ispitanicima ili životinjskim uzorcima, nemamo osjetljive podatke. Podaci će biti dostupni samo istraživačima uključenima u projekt do trenutka objave ili znanstvenog skupa.  Prilikom pisanja sažetka ili cjelokupnog rada, a prije same objave sve istraživače uključene u rad obavijestit ćemo o planovima o koautorstvu. Oni će moći sudjelovati u analizi podataka i stvaranju rukopisa, te će moći pregledati, dopuniti ili izmjeniti tekst prije no što se slože s konačnom verzijom tj. objavljenim materijalom.  GDPR se može odnositi na podatke koje imamo o uzgajivačima tj. objektima u kojima prikupljamo uzorke na terenu. Nijedan podatak koji se odnosi na njihove uzgojne objekte neće se objavljivati bez njihove privole. |
|  | 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? | Budući nemamo osjetljivih podataka, ne postoje drugi rizici pri regulaciji pristupa podacima osim da netko objavi rezultate prije nas. S ciljem izbjegavanja takvog rizika sirovi podaci će se obrađivati, a potom i pohranjivati isključivo tako da su dostupni unutar projektnog tima do trenutka objave. Po objavi znanstvenog rada, kako bi naši rezultati bili ponovljivi i provjerljivi poslati ćemo potrebne informacije svakom zainteresiranom znanstveniku. |
|  | 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? | Do trenutka javne objave rada kada su osigurana autorska prava svih osoba koje su doprinijele radu, podaci pripadaju nositeljici projekta, prof. Dijani Škorić te je potrebno njeno dopuštenje za uporabu podataka. Sve sekvence koje objavimo biti će javno dostupne na korištenje šire znanstvene javnosti u svrhu usporedbi i dodatnih analiza. |
| 3. | Pohrana i čuvanje podataka | |
|  | Kako će podaci biti pohranjeni i kako će biti napravljena sigurnosna kopija podataka (*backup*) tijekom istraživanja? Koji su kapaciteti čuvanja podataka kojim raspolažete? Kojim se procedurama koristite za sigurnosnu kopiju (*backup*)? | Podaci će biti podijeljeni s obzirom na istraživačke sezone te grupirani u nekoliko većih cjelina; (i) terenski popisi i fotografije, (ii) eksperimentalni dio, (iii) sekvence, (iv) validacija. Podaci će biti pohranjivani paralelno na kompjuterima poslijedoktorandice Grbin i prof. Škorić, te u zajedničkim folderima u cloudu. Radi konačne sigurnosne kopije, podaci će biti pohranjenju na projektni vanjski tvrdi disk nabavljen u tu svrhu na projektu. Dokumenti će biti uredno obilježeni kako je ranije navedeno u svrhu olakšanog snalaženja. |
|  | Koji je vaš plan čuvanja podataka? U kojim će se formatima čuvati? | Podaci će se čuvati na računalima poslijedoktorandice Grbin i prof. Škorić, na zajedničkom cloud-u te na vanjskom tvrdom disku (18TB). Privremeno, do završetka projekta, podaci o sekvencama biti će čuvani u na serveru fakulteta za Elektrotehniku i računarstvo u Zagrebu (FER). Po završetku projekta, konačne verzije podataka bit će pohranjene na repositoriju SRCA (<https://puh.srce.hr>) ili NCBI bazama ([https://www.ncbi.nlm.nih.gov/sra/SRX19817161[accn](https://www.ncbi.nlm.nih.gov/sra/SRX19817161%5baccn)] ). Formati u kojima će podaci dugoročno biti pohranjeni obuhvaćaju *pdf, csv, xlsx, fasta, fastq, .gz*. |
| 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? | Podaci o sekvencama (što su ujedno i ključni konačni rezultati ovog projekta) dostupni su brzo nakon pohrane u NCBI-u, pri čemu trenutak kada postaju dostupni ovisi isključivo o administratorima NCBI-a. Dijeljenje svih ostalih „radnih“ podataka prije objave rezultata bit će dostupno samo za članove projektnog tima. Nakon objave, svi relevantni dokumenti bit će dostupni na upit. Potencijalni korisnici će o našim podacima biti informirani na znanstvenim i stručnim skupovima te u objavljenim publikacijama. |
|  | Ako postoje podaci koji se ne smiju dijeliti (prijavitelji vezani zakonskim, etičkim, autorskim pravila, povjerljivošću i sl.), pojasnite razloge ograničenja. | Nemamo takvih podataka. |
|  | Potvrdite da ćete se koristiti digitalnim repozitorijem koji je u skladu s načelima *FAIR-a*. | Potvrđujemo da ćemo po završetku projekta finalne verzije ključnih rezultata pohraniti na repozitoriju u skladu s načelima FAIR. |
|  | 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 da ćemo po završetku projekta finalne verzije ključnih rezultata pohraniti na repozitoriju koje održava neprofitna organizacija. |

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)](https://repozitorij.srce.unizg.hr/islandora/object/srce:327)