**PLAN UPRAVLJANJA ISTRAŽIVAČKIM PODACIMA (PUP)**

|  |  |  |
| --- | --- | --- |
| Opće informacije | | |
|  | Ime i prezime predlagatelja | Tanja Šegvić Bubić |
|  | Matična organizacija | Institut za oceanografiju i ribarstvo |
|  | Naziv projekta | Poboljšanje ekološke učinkovitosti kaveznog uzgoja morskih riba |
|  | Upravitelj podacima | Tanja Šegvić Bubić |
| 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) | Očekuje se da će se ovim projektom dobiti velika količina podataka o ekosustavu uzorkovanjem vode i sedimenta.  Prikupljeni podaci uključuju:  1. Tehnički podaci: • Molekularni standardni operativni postupak s detaljnim uputama za izvođenje složenih uzorkovanja, laboratorijskih protokola i obrade podataka na standardiziran i ponovljiv način • Metapodaci uzorka: detaljni metapodaci o tome gdje, kada, tko i kako su prikupljeni uzorci  2. Genetski podaci: • Sanger sekvence FASTA datoteka • Neobrađene sekvence: FASTQ datoteke s neobrađenim čitanjima sekvenci koje šalje platforma za sekvenciranje • Postprocesirane sekvence: FASTA datoteke s obrađenim i kvalitetno kontroliranim sekvencama. S njima su povezane datoteke parametara koje koristi softver koji obrađuje podatke • Klasterirane/OTU tablice: podaci obrađenog slijeda grupirani prema Operativnim taksonomskim jedinicama (OTU) • Dodijeljene OTU tablice i/ili ASV datoteke: OTU tablice s taksonomskim dodjelama i relativnim abundancijama  3. Podaci o fotografijama: • Fotografije visoke rezolucije uzoraka za DNA barkodiranje • Fotografije visoke rezolucije divljih i uzgojnih riba • Fotografije visoke rezolucije ribljih jaja i gonada  4. Evidencija pojavljivanja u CSV formatu: • Abundancija vrsta izravno promatranih morfološkom identifikacijom • Abundancija ribljih jaja sa svakog mjesta uzorkovanja • Fekunditet ribe s mjesta uzorkovanja • Biometrijske i morfometrijske vrijednosti riba  5. Podaci o okolišu u CSV formatu: fizikalno geokemijska mjerenja sedimenata prikupljenih na mjestima uzorkovanja |
|  | 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) | Na kraju razdoblja uzorkovanja provest će se obrada uzorka u skladu s protokolima koje su definirali stručnjaci projekta. Metapodaci uzoraka, podaci o fotografijama i podaci o promatranju vrsta te podaci o okolišu bit će prebačeni u uslugu pohrane i dijeljenja datoteka PUH. Podaci o neobrađenim sekvencama koje laboratorij za sekvenciranje dostavi kao FASTQ datoteke bit će učitani u ENA (Europska arhiva nukleotida), a pristupni brojevi ENA bit će dodani u PUH. R i python skripte bit će dostupne putem repozitorija Github ili Gitlab.  Snimljene fotografije analizirat će se pomoću softvera za obradu slika, stvarajući CSV datoteke koje sadrže ID-ove uzoraka, taksonomske dodjele i mjere brojnosti. |
|  | 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, *ReadMe* datoteke i sl.) | ReadMe datoteke bit će osigurane za sve skupove podataka, s glavnim informacijama o podacima, primijenjenoj metodologiji i interpretaciji podataka.  Mape slika sadržavat će indeksne datoteke koje sadrže meta podatke o svakoj datoteci i koristit će konvencionalno imenovanje i strukturu stabla |
| 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)? | Nije primjenjivo. |
|  | 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? | Svi članovi istraživačkog tima imat će pristup dobivenim podacima putem digitalnog repozitorija koristeći svoje individualne vjerodajnice. Podaci se neće slati trećoj strani prije objave, osim kao dokaz recenzentima projekta tijekom redovitog izvještavanja.  Neće se prikupljati niti pohranjivati osjetljivi podaci. |
|  | 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? | Podaci neće imati izravnu komercijalnu vrijednost. Nakon objavljivanja podaci će biti dostupni u digitalnom repozitoriju kao što je Zenodo i nacionalnim digitalnim akademskim arhivima i repozitorijima – DABAR i pod licencom Creative Commons (atribucija, dijeljenje pod istim uvjetima, nekomercijalno). |
| 3. | Pohrana i čuvanje podataka | |
|  | Kako će radne verzije podataka biti pohranjene tijekom projekta?  Kako će se napraviti sigurnosne kopije tih podataka (*backup*)?  Koja je očekivana količina podataka koja će se prikupiti i čuvati tijekom projekta (izraženo u MB/GB/TB)? | Različite verzije datoteka bit će jedinstveno identificirane korištenjem vlastite konvencije imenovanja. Najjednostavniji način identificiranja određene verzije bit će dodavanje numeričke oznake kao što je v1, v2\_6 ili datum (GGGGMMDD). Automatska kontrola verzije datoteke bit će omogućena korištenjem softverskih rješenja kao što je Github. Najmanje 3 potpuno odvojene kopije podataka (npr. računalo, prijenosni medij, oblak) omogućit će backup podataka.  Količina podataka:  - Datoteke FASTQ tekstualni format datoteke podataka sekvenciranja koji pohranjuje neobrađene podatke sekvence i kvalitetu - veličina cijelog skupa podataka bit će ispod 1 TB  - Fotografije će biti ispod 50 GB  - .csv datoteke bit će manje od 20 MB |
|  | 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)? | Lokalno će se podaci čuvati na nekoliko laboratorijskih računala i replicirati na računala članova tima. Podaci će također biti sačuvani na servisu za pohranu i dijeljenje PUH i IOF poslužitelju. Podaci će imati tjednu vanjsku sigurnosnu kopiju na licu mjesta.  Komponente skupova podataka bit će komprimirane i sačuvane kao takve. |
| 4. | Dijeljenje i ponovna uporaba podataka | |
|  | Kako i gdje će se podaci dijeliti? Koji repozitorij će se koristit za dijeljenje podataka? Kako će potencijalni korisnici doznati za podatke? | Nakon objavljivanja, podaci korišteni u ovom projektu bit će vidljivi s metapodacima, prepoznatljivi i locirani od strane  sredstva standardnog mehanizma identifikacije. Podaci će biti dostupni za sekundarna istraživanja putem online repozitorija. Molekularni podaci bit će dostupni u repozitoriju ENA i NCBI, kao i u Barcode of Life (BOLD).  Ostali podaci bit će dostupni kao suplementi objavljenih radova ili arhivirani u digitalnom repozitoriju, npr. Zenodo.  Osim što ćemo datoteke s neobrađenim podacima učiniti javno dostupnima, deponirati ćemo skripte za reprodukciju naših analiza na Githubu ili Gitlabu |
|  | Ako postoje podaci koji se ne smiju dijeliti (prijavitelji vezani zakonskim, etičkim, autorskim pravila, povjerljivošću i sl.), pojasnite razloge ograničenja. | Prema trenutnim saznanjima, ne. |
|  | Potvrdite da ćete se koristiti digitalnim repozitorijem koji je u skladu s načelima *FAIR-a*. | Da. |
|  | 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). | Da. |