

NAČRT ZA RAVNANJE Z RAZISKOVALNIMI PODATKI

OBRAZEC ARIS

Ta obrazec je namenjen pripravi načrta za ravnanje z raziskovalnimi podatki (NRRP) za raziskovalne projekte, ki jih (so)financira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije (ARIS), kot je določeno v 4. členu [Uredbe o izvajanju znanstvenoraziskovalnega dela v skladu z načeli odprte znanosti](#) (Uradni list RS, št. 59/23).

Raziskovalni podatki so opredeljeni kot zapisi o dejstvih (številčni podatki, besedilni, zvočni in slikovni zapisi), ki predstavljajo osnovno podlago za znanstveno raziskovanje in ki v okviru znanstvene skupnosti veljajo kot ustrezno sredstvo za preverjanje veljavnosti raziskovalnih spoznanj.

Prosimo vas, da izpolnite spodnji obrazec NRRP in ga posredujete ARIS **najkasneje v šestih mesecih od začetka izvajanja raziskovalnega projekta**. Priporočljivo je, da NRRP med izvajanjem raziskovalnega projekta po potrebi redno pregledujete in posodabljate. V primeru sprememb posodobljen NRRP priložite vmesnemu in zaključnem poročilu o rezultatih raziskovalnega projekta.

Pregled vsebine NRRP:

0. Splošne informacije
1. Povzetek in opis raziskovalnih podatkov
2. Shranjevanje in varnostno kopiranje podatkov
3. Zagotovitev podatkov na način FAIR
 - 3.1 Zagotavljanje najdljivosti podatkov (F)
 - 3.2 Zagotavljanje dostopnosti podatkov (A)
 - 3.3 Zagotavljanje interoperabilnosti podatkov (I)
 - 3.4 Zagotavljanje ponovne uporabe podatkov (R)
4. Etični in pravni vidiki
5. Drugi raziskovalni rezultati
6. Finančna sredstva

Uporabljene kratice:

- ADP – Arhiv družboslovnih podatkov
- GDPR – Splošna uredba o varstvu podatkov
- IT – informacijska tehnologija
- RO – raziskovalna organizacija
- ZVDAGA – Zakon o varstvu dokumentarnega in arhivskega gradiva ter arhivih
- ZVOP-2 – Zakon o varstvu osebnih podatkov

Navodilo za izpolnjevanje

Obrazec izpolnite tako, da vsebino vnašate v celice v skrajnem desnem stolpcu oz. tam označite eno od ponujenih možnosti. V teh celicah so sedaj v sivi barvi pisave navedeni razlage oz. navodila za vnos ustreznih podatkov in opisov. To pomožno besedilo lahko po vnosu vsebine izbrišete.

0	Splošne informacije	
0.1	Šifra projekta	J4-60075
0.2	Naziv projekta	 GENODRONE GenoTrot - Genotipizacija trotovske sperme za boljše razumevanje reproduktivnih lastnosti medonosne čebele
0.3	Šifra vodje projekta	30770
0.4	Ime in priimek vodje projekta	Ajda Moškrič
0.5	Ime in priimek osebe, ki je v RO zadolžena za podporo pri ravnanju z raziskovalnimi podatki	Lili Marinček
0.6	Interna pravila RO za ravnanje z raziskovalnimi podatki	https://nrrp.odprtaznanost.si/
0.7	Verzija NRRP	1.0
1	Povzetek in opis raziskovalnih podatkov	
1.1	Ali boste pri projektu ponovno uporabili že obstoječe podatke predhodnih raziskav?	☒ Da ☐ Ne V projektu bomo v okviru molekularnih analiz uporabili javno dostopne podatke čebeljih genomov, shranjenih v podatkovnih zbirkah (NCBI). Uporabili bomo podatke o nukleotidnih zaporedjih za določene gene, dele čebeljega genoma, celoten genom in transkriptomske podatke. Specifična nukleotidna zaporedja bomo uporabili pri primerjalnih analizah. Celoten čbwlji genom bo služil kot referenčno zaporedje na katerega bomo kartirali odseke nuklearne DNA, pridobljene v tem projektu.
1.2	Katere vrste podatkov boste ustvarili oz. ponovno uporabili in v katerih formatih bodo shranjeni?	V projektu bomo ustvarili naslednje tipe podatkov: – podatki enoceličnega sortiranja spermijev trotov, – genomi in sekvenčni podatki (Sangerjevo sekvenciranje, GBS, NGS, posamezno celično sekvenciranje – scDNA-seq), – podatki o mutacijskih stopnjah v zarodni liniji pred in po osemenitvi, – podatki o alelni raznolikosti gena <i>csd</i> v izbranih čebeljih družinah, – bioinformatične analize, simulacije in modeli, – spremljevalni metapodatki (izvor vzorcev, način osemenitve, pogoji parjenja, identifikacija družin,

		status družin) Tekom projekta GenoDrone bomo ustvarili podatke specifičnih nukleotidnih zaporedij iz več sto čebeljih vzorcev. Surovi podatki bodo shranjeni v formatih .ab1 (kromatogrami) (Sanger sequencing, T-RFLP genotyping), urejeni podatki pa v formatu .fasta. Vsi rezultati analiz bodo shranjeni v formatu .fasta in v formatu .geneious. V projektu bomo ustvarili tudi genomske podatke, ki bodo shranjeni v formatih .fastq, .bam in .vcf. Podatki elektroforeznih gelov bodo shranjeni v formatih .sgd, .tiff in .jpeg. V projektu bomo za vse zbrane vzorce ustvarili tudi meta podatke, ki bodo vključevali podatek o tipu vzorca, datumu in lokaciji vzorčenja, ter pripadajočih analizah v formatu .xlsx. Slednje podatke bomo shranjevali tudi v tekstovnem formatu CSV.
1.3	Kakšen je namen ustvarjanja, zbiranja oz. ponovne uporabe podatkov in njihova povezava s cilji projekta?	Projekt GenoDrone bo ustvarjal in obdeloval eksperimentalne, genomske, bioinformatične ter spremljevalne (metapodatkovne) podatke, ki so neposredno povezani z doseganjem obeh glavnih ciljev projekta: Cilj 1: Raziskati genomsko nestabilnost v spermijih troto medonosne čebele (<i>Apis mellifera</i>), s poudarkom na hipervariabilni regiji gena <i>csd</i>. Cilj 2: Razviti novo metodologijo za ocenjevanje uspešnosti nadzorovanega parjenja na podlagi analize variabilnosti gena <i>csd</i> v spermijih, shranjenih v spermateki matice. Namen ustvarjanja, zbiranja in ponovne uporabe zgoraj navedenih podatkov je zapolniti vrzeli v znanju o reproduktivni biologiji pri medonosni čebeli ter razviti zanesljivo in časovno učinkovito orodje za ocenjevanje uspešnosti nadzorovanega parjenja pri vzreji in selekciji čebel.
1.4	Kakšna je pričakovana velikost podatkov, ki jih nameravate ustvariti oz. ponovno uporabiti?	☐ 0–10 GB ☐ 10–100 GB ☐ 100–1000 GB ☒ >1000 GB
2	Shranjevanje in varnostno kopiranje podatkov	
2.1	Kje bodo podatki med izvajanjem projekta shranjeni in varnostno kopirani?	Tekom projekta bodo genomske podatke in rezultati analiz genomske podatke shranjeni na Elixir strežniku s sedežem na Univerzi v Ljubljani. Strežnik ima tudi varnostno kopijo. Dodatno se bodo podatki podvojeno shranjevali tudi na internem strežniku Kmetijskega inštituta Slovenije, za katerega skrbi interna IT služba ter ima varnostno kopiranje. Manjši podatki (sangerjevo sekvenciranje - kromatogrami, preglednice, fotografije) se bodo shranjevali v oblaku

		OneDrive, ki ima funkcijo nadzora verzij. Vsa programska koda se bo shranjevala na repozitoriju github, ki je program za nadzor verzij in tako shranjuje trenutne kot tudi predhodne verzije datotek.
2.2	Kako boste izbrali podatke za dolgoročno hrambo?	Vsi podatki, ustvarjeni v projektu, bodo javno dostopni in objavljeni v znanstvenih publikacijah z odprtim dostopom. Genomski podatki, tako surovi kot rezultati analiz, bodo dodatno objavljeni v javno dostopnih repozitorijih. Programska koda bo dostopna v repozitoriju github. Repozitorije za dolgoročno hrambo bomo izbirali po principu FAIR. Pri izbiri bomo izbirali certificirane repozitorije, ki bodo zagotavljali, da bodo vsi podatki i) opremljeni z DOI številko; ii) opremljeni z meta-podatki; iii) javno dostopni z odprtim dostopom; iv) opremljeni s primernimi ključnimi besedami.
2.3	Ali bodo podatki shranjeni v zaupanja vrednem repozitoriju?	☒ Da ☐ Ne Podatki bodo shranjeni v repozitorijih Zenodo (https://zenodo.org/) ali NCBI (ENA, GenBank, SRA), programska koda pa v repozitoriju github (https://github.com/).
3.	Zagotovitev podatkov na način FAIR	
3.1	Zagotavljanje najdljivosti podatkov (F)	
3.1.1	Ali bodo podatki označeni s trajnim identifikatorjem (PID)?	☒ Da ☐ Ne Vsi podatki bodo opremljeni z DOI številko. Programska koda bo opremljena z URL naslovom.
3.1.2	Kateri metapodatki bodo ustvarjeni in kateri metapodatkovni standardi bodo pri tem upoštevani?	V projektu bomo ustvarili metapodatke za zbrane vzorce. Ti bodo vključevali vrsto organizma, ID organizma, lokacijo (mesto, država) in datum vzorčenja, informacije o populaciji (podvrsta, mitohondrijska linija), podatek o analizah, podatek o postopku izolacije DNA, podatek o tipu genomskih podatkov (tehnologija priprava knjižnice in sekvenciranja), podatek o fenotipu oz. lokaciji, informacijo o projektu ter morebitne opombe. Metapodatki bodo ustvarjeni skladno s standardi MIxS (Genomic Standards Consortium), INSDC (ENA/BioSample), ter smernicami MINSEQE za visoko zmogljivo sekvenciranje. Uporabljeni bodo kontrolirani slovarji in ontologije (npr. NCBI Taxonomy, Sequence Ontology, OBI), s čimer bo zagotovljena interoperabilnost in skladnost z načeli FAIR.
3.1.3	Ali bodo metapodatki vsebovali ključne besede za izboljšanje najdljivosti in	☒ Da ☐ Ne

	možnosti ponovne uporabe?	Ključne besede bodo vključevale področje dela, vrsto, vrsto analize. Ključne besede: medonosna čebela, csd, reprodukcija, nadzorovano parjenje, vzreja, selekcija, genomika, posamezno celično sekvenciranje, spermateka.
3.2	Zagotavljanje dostopnosti podatkov (A)	
3.2.1	Ali bodo vsi podatki odprto dostopni?	☒ Da ☐ Ne
3.2.2	Kdaj bodo podatki odprto dostopni in za koliko časa?	Delo bodo podatki javno dostopni takoj po izvedeni analizi načrtovanih vzorcev, ob objavi znanstvene publikacije, v celoti pa ob zaključku izvajanja projekta. Označeni bodo s trajnim identifikatorjem.
3.2.3	Na kakšen način bo v primeru omejitev pri uporabi omogočen dostop do podatkov med izvajanjem projekta in po njegovem zaključku?	Ne obstajajo omejitve.
3.2.4	Ali bo za dostop do podatkov oz. njihovo branje potrebna dodatna dokumentacija oz. informacija o ustrezni programski opremi?	☒ Da ☐ Ne Za branje genomskih podatkov bo potrebna analiza s sicer standardnimi bioinformatičnimi sosledji oz. programi. Podatki bodo opremljeni z znanstvenim člankom, ki bo pojasnjeval njihovo analizo ter uporabnike usmeril do javno dostopne programske kode za njihovo analizo.
3.3	Zagotavljanje interoperabilnosti podatkov (I)	
3.3.1	Katere geslovnike oz. šifrante boste uporabili pri pripravi podatkov in metapodatkov?	<i>Apis mellifera</i> , short-read sequence data, genomic data, DNA, subspecies, whole genome sequencing, csd, complementary sex determiner, mating control, reproduction, single cell sequencing, sperm, spermatheca
3.3.2	Ali boste primorani uporabiti manj poznane ali lastne geslovnike oz. šifrante?	☐ Da ☒ Ne
3.4	Zagotavljanje ponovne uporabe podatkov (R)	
3.4.1	Na kakšen način boste zagotovili dokumentacijo, potrebno za ponovno uporabo podatkov?	Ponovna uporaba podatkov bo zagotovljena z objavo znanstvenih člankov, ki bodo natančno opisali postopke pridobitve in analize podatkov ter uporabnike usmerile do javno dostopne programske kode za analizo podatkov. Vsa programska koda bo javno dostopna. Slednje bomo zagotovili z objavo lastne programske kode kot tudi z uporabo odprtokodnih programov. Podatki bodo opremljeni tudi s podrobnimi meta-podatki.

3.4.2	Ali bodo vaši podatki javno dostopni in licencirani z odprtima licencama CC BY oz. CC BY-SA, da bo s tem omogočena čim širša ponovna uporaba?	☒ Da ☐ Ne
3.4.3	Kakšne postopke zagotavljanja kakovosti podatkov boste uporabili?	V projektu GenoTrot bo kakovost podatkov zagotovljena z uporabo standardiziranih laboratorijskih protokolov, zanesljivih molekularnih metod ter dokumentiranih bioinformatičnih analiznih postopkov. Biološki vzorci (spermiji trotov, spermiji iz spermateke, matice, troti, delavke) bodo obdelani v kontroliranih laboratorijskih pogojih. Pred sekvenciranjem bo izvedena kontrola kakovosti DNK (meritve koncentracije, čistosti in integritete). Vključene bodo negativne in pozitivne kontrole ter tehnične ponovitve, kjer bo to potrebno. Sekvenčni podatki bodo podvrženi standardnim postopkom kontrole kakovosti (quality trimming, filtriranje nizkokakovostnih odčitkov, odstranjevanje adapterjev). Analize bodo izvedene z uporabo uveljavljenih bioinformatičnih orodij, pri čemer bodo verzije programske opreme in uporabljeni parametri dokumentirani za zagotavljanje ponovljivosti. Vzpostavljen bo enoličen sistem identifikacije vzorcev, ki bo omogočal sledljivost od biološkega vzorca do končnega analiznega rezultata. Metapodatki bodo zapisani v standardizirani obliki in povezani s sekvenčnimi podatki.
4.	Etični in pravni vidiki	
4.1	Ali obstajajo etična ali pravna vprašanja, ki bi lahko vplivala na deljenje podatkov?	☐ Da ☒ Ne
4.2	Ali boste med izvajanjem projekta obdelovali oz. hranili osebne podatke?	☐ Da ☒ Ne
4.3	Ali bodo med projektom ustvarjene oz. ponovno uporabljene posebne vrste osebnih podatkov?	☐ Da ☒ Ne
4.4	Kako boste uredili lastništvo podatkov in morebitne avtorske pravice	Lastnik raziskovalnih podatkov je vodja projekta, pred javno objavo bodo pravico do uporabe imeli tudi uporabniki, ki so prispevali vzorce in partnerji na projektu.

	na podatkih, ki jih boste ustvarili ali ponovno uporabili?	Prehodno ustvarjeni podatki, ki jih bomo uporabili v projektu, so javno dostopni v javnih zbirkah in jih bomo ustrezno citirali.
5.	Drugi raziskovalni rezultati	
5.1	Ali boste poleg podatkov ustvarili ali ponovno uporabili tudi druge raziskovalne rezultate?	☐ Da ☒ Ne
6.	Finančna sredstva	
6.1	Kakšni bodo stroški ravnanja s podatki in drugimi rezultati projekta po načelih FAIR in kako bodo kriti?	Za objavo podatkov bomo uporabili brezplačen repozitorije, ki zadostujejo principu FAIR, zato dodatnih stroškov za to ne pričakujemo. Pričakovan strošek bo strošek objave člankov v odprtem dostopu, 2000-3000€ na članek, kar bomo krili ali iz sredstev projekta oz. bomo objavili v revijah, s katerimi ima Kmetijski inštitut Slovenije sklenjene konzorcijske pogodbe – v tem primeru dodatne stroške ne bo.
6.2	Kdo bo odgovorna oseba za ravnanje z raziskovalnimi podatki pri projektu?	Ajda Moškrič

Uporabljeni viri:

- *Anotirana predloga načrta za ravnanje s raziskovalnimi podatki za projekte Obzorja Evropa.* CTK UL. Dostopno na: <https://dirrosdata.ctl.uni-lj.si/raziskovalni-podatki/nacrt-ravnanja-z-raziskovalnimi-podatki/>.
- Bezjak, Sonja (ur.) (2024). *Spoznaj FAIR: Priročnik o odprti znanosti v Sloveniji.* Univerza na Primorskem. Dostopno na: <https://www.hippocampus.si/ISBN/978-961-293-328-9.pdf>.
- *Horizon Europe Data management plan template.* Dostopno na: <https://www.openaire.eu/images/Guides/HORIZON EUROPE Data-Management-Plan-Template.pdf>.
- *NWO Template Data management plan.* Dostopno na: <https://www.nwo.nl/en/research-data-management>.

Verzija dokumenta: 1.0

Datum: 15. 10. 2024