

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	J4-60081
0.2	Naziv projekta	Analiza evolucijske zgodovine medonosne čebele za trajnostno izboljševanje populacij (BEEography)
0.3	Šifra vodje projekta	38985
0.4	Ime in priimek vodje projekta	Jana Obšteter
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 za populacijsko genetske analize uporabili javno dostopne podatke čebeljih genomov, shranjenih v različnih arhivih (NBGI, ENAA, Zenodo ...). Poleg tega bomo v projektu uporabili tudi javno dostopne podatke o paleoklimatskih in paleogeografskih razmerah.
1.2	Katere vrste podatkov boste ustvarili oz. ponovno uporabili in v katerih formatih bodo shranjeni?	V projektu bomo ustvarili podatke kratkih bralnih sekvenc iz celotnega genoma več sto čebeljih vzorcev. Surovi podatki bodo shranjeni v formatu fastq, končni rezultat analize pa bo shranjen v VCF formatu. Slednji formati so standardni formati hranjenja genomskih podatkov, ki jih pričakujejo

		tudi vsi programi za analizo. V projektu bomo za vse zbrane vzorce čebel ustvarili tudi meta podatke, ki bodo vključevali podatek o datumu in lokaciji vzorčenja, ter podvrsti. Slednje podatke bomo shranjevali v tekstovnem formatu CSV. V naslednjem koraku bomo iz podatkov ocenili predniške alele za vrsto <i>Apis mellifera</i>. Slednji bodo shranjeni v formatu CSV. Na podlagi pridobljenih genomskih podatkov in ocenjenjenih predniških alelov bomo v projektu ustvarili predniške grafe rekombinacij, ki bodo shranjeni v formatu ts (angl. <i>Tree sequence</i>). Slednji format učinkovito shrani genomske podatke ter je standardni format za rodoslovno analizo. Vsem vzorcem čebel vzorčenih v projektu bomo izmerili tudi fenotip, katerega bo v našem primeru predstavljala morfologija kril. Vsem vzorcem bomo slikali eno izmed kril, slike bodo shranjene v formatu PNG. V projektu bomo ustvarili tudi programsko kodo - tako v procesu analize kode kot tudi v procesu priprave enega izmed rezultatov projekta, simulatorja čebelje demografije. Vsa koda bo shranjena v dotičnih formatih (.R za programsko okolje R, .py za programsko okolje python).
1.3	Kakšen je namen ustvarjanja, zbiranja oz. ponovne uporabe podatkov in njihova povezava s cilji projekta?	Osrednji cilj projekta je analiza demografske zgodovine vrste <i>Apis mellifera</i>. Slednjo bomo analizirali in parametrizirali na podlagi analize genomskih podatkov različnih čebeljih podvrst, ki jih bomo pridobili v projektu ter javno dostopnih podatkov ustvarjenih v predhodnih raziskavah. Genomski podatki tako predstavljajo glavno orodje projekta. Analiza demografske zgodovine zahteva tudi, da imajo vsi vzorci poznane meta podatke – lokacijo in čas vzorčenja ter podvrsto oz. mitohondrialno linijo, kateri pripadajo. Validiranje demografskega modela bo vključevalo tudi potrjevanje verjetnosti modela, zgrajenega na genomskih podatkih, na podlagi paleoklimatskih in paleogeografskih podatkov. Med cilji projekta je tudi priprava bioinformatkega

		sosledja ter izgradnjo simulatorja čebelje demografske zgodovine. Za oba bomo v projektu ustvarili programsko kodo, ki bo omogočila analizo podatkov oz. bo omogočila uporabo rezultatov projekta.
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 (preglednice, fotografije) sd bodo shranjevali v oblaku OneDrive, ki ima funkcijo nazdora verzij. Vsa programska koda se bo shranjevala na repozitoriju github, ki je program za nazdor 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. Genomske podatke, 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 repozitoriju Zenodo (https://zenodo.org/), 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 meta-podatke za zbrane čebelje vzorce. Ti bodo vključevali vrsto organizma, ID organizma, lokacijo (mesto, država) in datum vzorčenja, informaciji o populaciji (podvrsta, mitohondrijska linija), 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. Meta-podatki bodo sledili BV-BRC standardu za genomske podatke (https://www.bv-brc.org/docs/system_documentation/data.html#genome-metadata)
3.1.3	Ali bodo metapodatki vsebovali ključne besede za izboljšanje najdljivosti in možnosti ponovne uporabe?	☒ Da ☐ Ne Ključne besede bodo vključevale področje dela, vrsto, vrsto analize.
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?	Podatki bodo javno dostopni takoj po analizi vseh načrtovanih vzorcev, ki bo potekalo še v zadnjem letu projekta, ter pre-objave znanstvenih člankov z rezultati, kar načrtujemo za zadnje leto projekta oz. kmalu po končanju projekta.
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 sicer standardnimi bioinformatскими sosledji oz. programi. Podatki bodo opremljeni z znanstvenim člankom, ki bo pojasneval 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?	Apis mellifera, short-read sequence data, genomic data, diversity study, museum samples, historical DNA, population genetics, subspecies, mitochondrial lineages.
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.
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?	Pri samem zbiranju bioloških vzorcev bomo sledili vnaprej pripravljenim oz. obstoječim protokolom. Pri dokumentiranju in hranjenju tekstovnih podatkov bomo v projektu uporabljali standardizirane zbirke in formate podatkov. Za

		zbiranje podatkov o vzorcih bomo s člani projekta vnaprej pripravili standardizirane tabele, ki jih bomo uporabili v vsakem krogu zbiranja oz. obdelave podatkov. Napake bomo omejili s kontrolami za vnos podatkov oz. z omejenimi prednastavljenimi možnostmi za posamezna polja. Za omejitev napak in povečanje kakovosti programske kode bomo pred analizami na celotnem setu podatkov programsko kodo testirali na pilotnem setu podatkov. Kakovost podatkov bomo merili s stopnjo popolnosti podatkov in pa rezultati pilotnih analiz.
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 na podatkih, ki jih boste ustvarili ali ponovno uporabili?	Lastnik raziskovalnih podatkov je vodja projekta, pred javno objavo bodo pravico do uporabe imeli tudi uporabniki, ki so prispevali vzorce in partnerji na projektu. 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, kot Zenodo, zato dodatnih stroškov za to ne pričakujemo. Pričakovan strošek bo strošek objave člankov v odprtem dostopu, 2000-3000€ na članek, stroške česa bomo krili ali iz sredstev projekta oz. bomo objavili v revijah, s katerimi ima Kmetijski inštitut Slovenije sklenjene konzorcijske pogodbe – v tem primeru dodatne stroška ne bo.
6.2	Kdo bo odgovorna oseba za ravnanje z raziskovalnimi podatki pri projektu?	Jana Obšteter

Uporabljeni viri:

- *Anotirana predloga načrta za ravnanje s raziskovalnimi podatki za projekte Obzorja Evropa*. CTK UL. Dostopno na: <https://dirrosdata.ctk.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