

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	J4-70164
0.2	Naziv projekta	Novi proteinski nanodelci za boj proti okužbam, odpornim na zdravila: oligolizini
0.3	Šifra vodje projekta	50490
0.4	Ime in priimek vodje projekta	Martina Durcik
0.5	Ime in priimek osebe, ki je v RO zadolžena za podporo pri ravnanju z raziskovalnimi podatki	V raziskovalni ustanovi ni osebe zadolžene za podporo pri ravnanju z raziskovalnimi podatki
0.6	Interna pravila RO za ravnanje z raziskovalnimi podatki	Raziskovalna ustanova ima interna pravila za ravnanje z raziskovalnimi podatki in sicer pravila za hranjenje podatkov, protokol za pisanje laboratorijskih dnevnikov z opisanimi informacijami, ki jih dnevnik mora vsebovati, ima dokumentirano bazo biološkega materiala ter pravila za shranjevanje materiala v bazo ter vodenje evidenc, pravila glede varovanja zaupnih informacij itd. Interni pravilniki niso javno objavljeni.
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 ponovno uporabili tako interne kot javno dostopne podatke kot izhodišče za bioinformatiko analizo, načrtovanje in primerjalno vrednotenje novih endolizinov oz. proizvedenih proteinov. To so: • aminokislinska in nukleotidna zaporedja znanih endolizinov ter bakteriofagnih hidrolaz iz javnih baz (UniProt, NCBI GenBank/RefSeq), • razpoložljive tridimenzionalne strukture sorodnih proteinov (PDB) in predhodno izračunani modeli (AlphaFold) za strukturno modeliranje in načrtovanje, • javno dostopni genomi ciljnih bakterijskih patogenov za analizo tarčnih struktur celične stene (NCBI, BacDive), • domenske/funkcijske anotacije iz InterPro/Pfam za identifikacijo katalitičnih in vezavnih domen. Ti podatki bodo uporabljeni kot vhodni podatki za bioinformatiko analize in načrtovanje proteinov, ne pa ponovno objavljeni kot lastni podatki.
1.2	Katere vrste podatkov boste ustvarili oz. ponovno	V projektu bomo ponovno uporabili in ustvarili naslednje vrste podatkov:

	uporabili in v katerih formatih bodo shranjeni?	- številčni podatki (preglednice): rezultati testov aktivnosti (MIC, cone lize, kinetika encimske aktivnosti), podatki čiščenja proteinov, parametri in metrike in silico analiz (ddG, RMSD/RMSF, rezultati dockinga) – format .xlsx, .csv; - besedilni podatki: nukleotidna in aminokislinska zaporedja, protokoli, SOP-i, zapisi iz elektronskega laboratorijskega dnevnika – format FASTA, GenBank (.gb), .txt, .pdf, .docx; - slikovni podatki: SDS-PAGE in western blot slike, vizualizacije strukturnih modelov, grafi – format .tiff, .png, .jpg; - interaktivni/strukturni podatki: surovi podatki sekvenciranja, 3D strukturni modeli in trajektorije molekulske dinamike – format FASTQ, .ab1, .pdb, .cif/.mmCIF, .xtc/.trr, .json/.yaml. Izbrani formati so, kjer je mogoče, odprti in splošno uveljavljeni standardni formati na raziskovalnem področju (FASTA, PDB/mmCIF, FASTQ, CSV), kar zagotavlja združljivost z uveljavljeno programsko opremo in repozitoriji. Nekateri formati so določeni z uporabljenimi laboratorijsko oz. računalniško opremo/programsko opremo in so standardni na tem področju.
1.3	Kakšen je namen ustvarjanja, zbiranja oz. ponovne uporabe podatkov in njihova povezava s cilji projekta?	Podatki predstavljajo osnovo za celoten raziskovalni proces, od identifikacije in načrtovanja proteinskih kandidatov do njihove eksperimentalne priprave in funkcionalnega vrednotenja. Bioinformatični podatki bodo uporabljeni za identifikacijo funkcionalnih domen, izbiro in načrtovanje proteinskih variant ter računalniško napoved njihovih lastnosti. Rezultati strukturnega modeliranja in molekulske dinamike bodo uporabljeni za razumevanje strukture, stabilnosti in potencialnega delovanja proteinov ter za njihovo računalniško podprto optimizacijo. Eksperimentalni podatki bodo omogočili preverjanje računalniških napovedi ter primerjavo različnih generacij proteinov glede na izražanje, stabilnost, biokemijske lastnosti in protibakterijsko aktivnost. Podatki bodo tako omogočili iterativni proces načrtovanja, izdelave, testiranja in nadaljnje optimizacije proteinskih kandidatov.
1.4	Kakšna je pričakovana velikost podatkov, ki jih	☐ 0–10 GB ☐ 10–100 GB

	nameravate ustvariti oz. ponovno uporabiti?	☒ 100–1000 GB ☐ >1000 GB
2	Shranjevanje in varnostno kopiranje podatkov	
2.1	Kje bodo podatki med izvajanjem projekta shranjeni in varnostno kopirani?	Podatki bodo med izvajanjem projekta shranjeni v varnem internem informacijskem okolju. Primarni eksperimentalni podatki, nastali neposredno med laboratorijskim delom, bodo dokumentirani v laboratorijskih dnevnikih (papirni zapisi) ter v elektronskem laboratorijskem dnevniku (ELN). Dostop do delovnih podatkov bo omejen na člane projektne skupine. Digitalni podatki, vključno z rezultati meritev, slikovnim gradivom, sekvencami, računalniškimi modeli, vhodnimi in izhodnimi datotekami bioinformatičkih analiz ter rezultati molekulske dinamike, bodo shranjeni na internih podatkovnih shrambah oziroma strežnikih z nadzorovanim dostopom. Uporabljeno bo načelo 3-2-1: • primarna kopija podatkov bo hranjena na delovnem mestu oziroma strežniku, • varnostne kopije bodo shranjene na ločenem sistemu, • ključni podatki bodo arhivirani na ločeni lokaciji. Dostop do neobjavljenih podatkov bo omejen na člane projektne skupine. Posebna zaščita bo namenjena podatkom, ki predstavljajo potencialno intelektualno lastnino, predvsem novim proteinskim sekvencam, optimiziranim variantam endolizinov, rezultatom načrtovanja proteinov ter podatkom, ki bi lahko predstavljali osnovo za patentno zaščito.
2.2	Kako boste izbrali podatke za dolgoročno hrambo?	Za dolgoročno hrambo bodo izbrani podatki, ki so ključni za razumevanje, preverjanje in ponovno uporabo rezultatov projekta, zlasti podatki, ki podpirajo znanstvene objave, pomembne računalniške in eksperimentalne rezultate ter podatki, povezani z morebitnimi rezultati intelektualne lastnine. Pri izboru bodo upoštevane: • zahteve financerja in veljavne politike odprte znanosti; • pogodbeni dogovori med projektnimi partnerji; • zahteve glede zaščite intelektualne lastnine; • znanstvena vrednost in možnost ponovne uporabe podatkov;

		• potrebe po dokazovanju izvora in sledljivosti raziskovalnih rezultatov. Podatki, ki niso potrebni za razumevanje rezultatov in nimajo dolgoročne raziskovalne, pravne ali poslovne vrednosti, se ne bodo nujno hranili trajno.
2.3	Ali bodo podatki shranjeni v zaupanja vrednem repozitoriju?	☒ Da ☐ Ne Podatki, ki ne bodo predstavljali poslovno občutljivih informacij ali intelektualne lastnine, bodo objavljeni v zaupanja vrednih repozitorijih, kot je: NCBI GenBank (specifični podatki - nukleotidna/aminokislinska zaporedja) ali Zenodo (splošni raziskovalni podatki). Navedeni repozitoriji omogočajo trajno hrambo in izvajanje digitalnega skrbništva ter dodeljujejo trajni identifikator (DOI oz. pristopno številko). Podatki, povezani z morebitnimi patentnimi prijavami ali komercialno relevantnimi rezultati, bodo do zaključka postopkov zaščite intelektualne lastnine ostali omejeni, trajno skrbništvo zagotovljeno znotraj varnega internega informacijskega okolja (točka 2.1).
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 Tisti podatki, ki bodo objavljeni v repozitorijih, bodo označeni s trajnim identifikatorjem, predvsem DOI. Pri sekvenčnih podatkih bodo uporabljeni tudi identifikatorji, ki jih dodelijo specializirane podatkovne zbirke, kadar bodo podatki vanje deponirani.
3.1.2	Kateri metapodatki bodo ustvarjeni in kateri metapodatkovni standardi bodo pri tem upoštevani?	Metapodatki bodo vključevali informacije o: • izvoru in namenu podatkov; • datumu nastanka; • raziskovalcu oziroma izvajalcu; • vzorcu oziroma proteinskem konstrukt; • uporabljenih eksperimentalnih metodah in pogojih; • uporabljeni opremi; • programski opremi in njenih različicah; • parametrih računalniških analiz in molekulske dinamike;

		• različici podatkov in povezavi z drugimi raziskovalnimi rezultati. Uporabljeni bodo metapodatkovni elementi, ki jih zahtevajo izbrani repozitoriji, ter standardi oziroma uveljavljene prakse za posamezne vrste bioloških podatkov. Kjer specifični področni standard ne bo primeren bodo metapodatki pripravljeni v strukturirani obliki.
3.1.3	Ali bodo metapodatki vsebovali ključne besede za izboljšanje najdljivosti in možnosti ponovne uporabe?	☒ Da ☐ Ne Metapodatki bodo vsebovali relevantne ključne besede, kot so <i>endolysin</i>, <i>phage lysin</i>, <i>polysaccharide depolymerase</i>, <i>antibacterial protein</i>, <i>protein engineering</i>, <i>protein nanoparticle</i>, <i>molecular dynamics</i>, <i>Klebsiella pneumoniae</i>, <i>Pseudomonas aeruginosa</i>, <i>Staphylococcus aureus</i> in <i>Staphylococcus epidermidis</i>
3.2	Zagotavljanje dostopnosti podatkov (A)	
3.2.1	Ali bodo vsi podatki odprto dostopni?	☐ Da ☒ Ne Podatki ne bodo v celoti odprto dostopni, predvsem zaradi varovanja intelektualne lastnine in morebitnega komercialnega potenciala rezultatov. Odprto bodo objavljeni podatki, ki ne predstavljajo poslovno občutljivih informacij ali ne ogrožajo zaščite intelektualne lastnine. Za ustrezne sekvenčne podatke bo po potrebi uporabljen NCBI GenBank, za druge raziskovalne podatke pa Zenodo ali drug ustrezen zaupanja vreden repozitorij. Podatki, povezani z morebitnimi patentnimi prijavami, novimi proteinskimi kandidati, optimiziranimi sekvencami ali drugimi komercialno relevantnimi rezultati, bodo do zaključka postopkov zaščite intelektualne lastnine ostali omejeni.
3.2.2	Kdaj bodo podatki odprto dostopni in za koliko časa?	Podatki brez IP omejitev bodo odprto dostopni najkasneje ob objavi pripadajoče znanstvene publikacije, v kateri bodo citirani s trajnim identifikatorjem, oziroma najkasneje ob zaključku projekta.

		Podatki, ki bodo predmet IP presoje, bodo objavljeni po zaključku postopkov zaščite intelektualne lastnine oziroma po presoji, da objava ne ogroža pravic do zaščite. Po objavi bodo podatki, kjer je to mogoče, trajno dostopni v repozitoriju, skupaj z metapodatki in DOI.
3.2.3	Na kakšen način bo v primeru omejitev pri uporabi omogočen dostop do podatkov med izvajanjem projekta in po njegovem zaključku?	Med izvajanjem projekta bo dostop do omejenih podatkov omogočen članom projektne skupine preko avtenticiranih uporabniških računov in nadzorovanih podatkovnih shramb. Dostop tretjih oseb do zaupnih oziroma poslovno občutljivih podatkov bo mogoč le na podlagi ustreznega dogovora med partnerji, po potrebi z uporabo sporazuma o zaupnosti (NDA) oziroma za fizične materiale sporazuma o prenosu materiala (MTA). Po zaključku IP presoje bodo podatki, ki jih bo mogoče javno objaviti, objavljeni v ustreznem repozitoriju.
3.2.4	Ali bo za dostop do podatkov oz. njihovo branje potrebna dodatna dokumentacija oz. informacija o ustrezni programski opremi?	☒ Da ☐ Ne Za nekatere podatke, predvsem rezultate molekulske dinamike in druge računalniške analize, bodo potrebne informacije o uporabljeni programski opremi, različicah programov in parametrih analiz. Podatkom bodo zato priložene informacije o uporabljenih orodjih, vhodnih podatkih in parametrih ter po potrebi datoteke README in analizne kode. Kjer bo mogoče in skladno z IP omejitvami, bodo skripte objavljene kot odprtokodna programska oprema.
3.3	Zagotavljanje interoperabilnosti podatkov (I)	
3.3.1	Katere geslovnike oz. šifrantе boste uporabili pri pripravi podatkov in metapodatkov?	Uporabljali bomo: • nomenklaturu Enzyme Commission (EC) za katalitično klasifikacijo endolizinov, • UniProt za funkcijske anotacije proteinov, • NCBI Taxonomy za opredelitev organizmov. • ...
3.3.2	Ali boste primorani uporabiti manj poznane ali lastne geslovnike oz. šifrantе?	☐ Da ☒ Ne

3.4	Zagotavljanje ponovne uporabe podatkov (R)	
3.4.1	Na kakšen način boste zagotovili dokumentacijo, potrebno za ponovno uporabo podatkov?	Podatki, namenjeni ponovni uporabi, bodo opremljeni z dokumentacijo, ki bo vključevala informacije o izvoru podatkov, datumu nastanka, metodologiji, eksperimentalnih pogojih, uporabljenih instrumentih in programski opremitvi ter definiciji ključnih spremenljivk in merskih enot. Podatkovnim zbirkam bodo po potrebi priložene datoteke README z opisom strukture podatkov, uporabljenih formatov, metod in postopkov analize.
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 Zaradi potencialne intelektualne lastnine vsi podatki ne bodo mogli biti objavljeni pod odprtima licencama. Podatki, ki bodo primerni za javno objavo, bodo po IP presoji opremljeni z ustrezno odprto licenco, npr. CC BY, kadar bo to mogoče. Za podatke, za katere odprta licenca ni primerna, bodo veljali pogoji uporabe, določeni v repozitoriju oziroma z dogovorom imetnikov pravic.
3.4.3	Kakšne postopke zagotavljanja kakovosti podatkov boste uporabili?	Kakovost podatkov bomo zagotavljali z: • standardiziranimi laboratorijskimi protokoli (SOP) za ekspresijo, čiščenje proteinov in teste aktivnosti, ki jih bodo upoštevali vsi sodelavci, • preverjanjem sekvenc konstruktov, • izvajanjem meritev v ponovitvah (biološke in tehnične ponovitve) ter uporabo ustreznih pozitivnih in negativnih kontrol pri testih aktivnosti, • redno kalibracijo in vzdrževanjem merilne opreme (spektrofotometri, itd.), • sprotnim dokumentiranjem eksperimentov v (elektronskem) laboratorijskem dnevniku, kar omogoča sledljivost, avtomatsko časovno žigosanje in preverjanje, • shranjevanjem vhodnih in izhodnih datotek računalniških analiz, • validacijo izračunov na testnih/referenčnih naborih podatkov, • notranjim pregledom podatkov znotraj projektne skupine ter usposabljanjem novih sodelavcev za dosledno beleženje in vnos podatkov.

4.	Etični in pravni vidiki	
4.1	Ali obstajajo etična ali pravna vprašanja, ki bi lahko vplivala na deljenje podatkov?	☒ Da ☐ Ne Glavno pravno vprašanje, ki vpliva na deljenje podatkov, je zaščita intelektualne lastnine – projekt izvaja zasebni zavod skupaj s konzorcijem akademskih in gospodarskih partnerjev, zato imajo posamezni podatki (predvsem novo načrtovane sekvence in konstrukti) komercialni potencial. O tem, kateri podatki so predmet omejitve, odloča interni postopek IP presoje (vodstvo zavoda/IP odbor), pravna vprašanja med partnerji pa dodatno ureja konzorcijska pogodba oz. sporazum (lastništvo, pravice do objave, postopek odobritve pred objavo).
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?	Lastništvo raziskovalnih podatkov in delitev pravic do njihove uporabe/objave bosta med partnerji urejena s konzorcijsko pogodbo oz. sporazumom, sklenjenim pred oz. ob začetku izvajanja projekta. Splošno izhodišče je, da je lastnik podatkov partner, katerega zaposleni oz. sodelavci so podatke ustvarili z lastnimi sredstvi/opremo; podatki, ustvarjeni skupno. Objava oz. zunanja izmenjava podatkov s strani posameznega partnerja je pogojena s predhodnim pregledom in potrditvijo vseh partnerjev. Za ponovno uporabljene, javno dostopne podatke (referenčna zaporedja, strukture, genomi iz UniProt, NCBI, PDB, AlphaFold DB ipd.) lastništvo ostaja pri izvornih avtorjih/bazah; uporabljamo jih skladno s pogoji uporabe posamezne javne baze in si jih ne lastimo.
5.	Drugi raziskovalni rezultati	
5.1	Ali boste poleg podatkov ustvarili ali ponovno uporabili tudi druge raziskovalne rezultate?	☒ Da ☐ Ne Poleg podatkov bomo v projektu ustvarili oz. ponovno uporabili tudi naslednje rezultate:

		- Digitalni rezultati: računalniški modeli proteinov/simulacije protokoli in standardni operativni postopki (SOP). - Fizični rezultati: ekspresijski plazmidni konstrukti in bakterijski sevi (E. coli) za produkcijo posameznih variant endolizinov, prečiščeni proteinski vzorci (endolizini) za nadaljnje testiranje. Digitalne rezultate (splošne postopke, protokole brez razkritja specifičnih novo načrtovanih zaporedij) bomo delili v skladu z načeli FAIR. Fizični materiali (ekspresijski plazmidi, sevi, proteinski vzorci) bodo hranjeni skladno z internimi postopki zavoda. Rezultati s potencialom komercialne uporabe bodo pred objavo ocenjeni z vidika zaščite intelektualne lastnine. Na voljo za izmenjavo z drugimi raziskovalnimi skupinami bodo po zaključeni IP presoji oz. vloženi patentni prijavi, praviloma na podlagi sporazuma o prenosu materiala (MTA) ali sporazuma o varovanju zaupnosti (NDA). Do takrat ostajajo interni rezultati projektne skupine.
6.	Finančna sredstva	
6.1	Kakšni bodo stroški ravnanja s podatki in drugimi rezultati projekta po načelih FAIR in kako bodo kriti?	Stroški bodo povezani s: - shranjevanjem in varnostnim kopiranjem podatkov (institucionalna IT infrastruktura), - uporabo elektronskega laboratorijskega dnevnika (licenca), - pripravo podatkov in metapodatkov za objavo, - morebitnim arhiviranjem v repozitorijih, - pripravo podatkov za zaščito intelektualne lastnine, kadar bo to potrebno. Stroški bodo kriti iz projektних sredstev in lastnih virov raziskovalne organizacije.
6.2	Kdo bo odgovorna oseba za ravnanje z raziskovalnimi podatki pri projektu?	Vodja projekta, Martina Durcik

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