

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

0	Splošne informacije	
0.1	Šifra projekta	J1-70052
0.2	Naziv projekta	Osvetlitev kemijskega prostora za zahtevne tarče (LUX: BLI-AS-MS)
0.3	Šifra vodje projekta	24400
0.4	Ime in priimek vodje projekta	Janez Ilaš
0.5	Ime in priimek osebe, ki je v RO zadolžena za podporo pri ravnanju z raziskovalnimi podatki	Tinka Leskovec, podatkovna svetovalka na UL FFA Datum sestanka: 28. 5. 2026, 26. 8. 2026
0.6	Interna pravila RO za ravnanje z raziskovalnimi podatki	Pravilnik o stabilnem financiranju znanstvenoraziskovalne dejavnosti UL, 25. in 26. čl., povzeto s spletne strani: Pravilnik-o-stabilnem-financiranju-znanstvenoraziskovalne-dejavnosti-UL-neuradno-precisceno-besedilo-velja-od-6.4.2024.pdf
0.7	Verzija NRRP	1.0 (datum objave: 31. 8. 2026)
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 omejenem obsegu ponovno uporabili javno dostopne in interne podatke iz predhodnih raziskav ter objavljene znanstvene literature. To bodo zlasti tridimenzionalne strukture bakterijskih in humanih topoizomeraz ter njihovih kompleksov z ligandi iz javnih podatkovnih zbirk, kemijske strukture in podatki o znanih ligandih oziroma zaviralcih topoizomeraz, objavljeni sintezni postopki in fotokatalitske reakcije ter predhodno razviti in validirani protokoli za sintezo, analitsko karakterizacijo, BLI in masnospektrometrične meritve. Za načrtovanje kemijskih knjižnic in učenje oziroma vrednotenje generativnih modelov bomo lahko ponovno uporabili tudi ustrezno licencirane podatke o kemijskih strukturah, reakcijah in lastnostih spojin iz javnih podatkovnih zbirk. Izvor, različica, datum dostopa, trajni identifikator oziroma pristopna oznaka ter licenca vsakega ponovno uporabljenega podatkovnega vira bodo dokumentirani. Ponovno uporabljeni podatki bodo jasno ločeni od podatkov, ustvarjenih v projektu.

1.2	Katere vrste podatkov boste ustvarili oz. ponovno uporabili in v katerih formatih bodo shranjeni?	V projektu bomo ustvarili in ponovno uporabili naslednje vrste raziskovalnih podatkov: 1. Računalniško in kombinatorično načrtovanje spojin z generativnimi modeli Ustvarjeni bodo podatki o vhodnih kemijskih strukturah in gradnikih, pravilih oziroma omejitvah kombinatoričnega načrtovanja, učnih, validacijskih in testnih množicah, konfiguraciji modelov, različicah kode, naključnih semenih, hiperparametrih, kontrolnih točkah modelov ter generiranih virtualnih knjižnicah. Dokumentirani bodo postopki filtriranja, odstranjevanja dvojnikov, standardizacije struktur, ocenjevanja kemijske raznolikosti, fizikalno-kemijskih lastnosti in sintezne izvedljivosti. <i>Formati:</i> SMILES (.smi, .csv, .txt), SDF/MOL/MOL2, JSON oziroma YAML za konfiguracije, .py in .ipynb za skripte in zvezke, po potrebi lastniški formati uporabljene programske opreme ter .png, .svg in PDF/A za prikaze. Kontrolne točke modelov bodo shranjene v izvornih formatih uporabljenega ogrodja, če bodo potrebne za ponovitev izračunov. 2. Podatki o kemijskih spojinah, knjižnicah in sintezi Ustvarjeni bodo zapisi o načrtovanih, generiranih in sintetiziranih spojinah, reakcijskih zmesih in knjižnicah, vključno s kemijskimi strukturami, izhodiščnimi spojinami, reagenti, katalizatorji, reakcijskimi pogoji, izkoristki, čiščenjem, identiteto in čistoto. Uporabljen bo enoten interni identifikator spojine oziroma reakcijske zmesi in šifrant, ki bo povezoval identifikatorje, strukture, SMILES/InChI/InChIKey, serije sinteze ter oznake v objavah. <i>Formati:</i> .cdxml, .smi, .sdf, .mol/.mol2, .csv, .xlsx, .txt, .docx in PDF/A. 3. NMR, HPLC in rutinski LC-MS podatki Za karakterizacijo izbranih vmesnih in končnih spojin bodo shranjeni izvorni in obdelani NMR, HPLC in LC-MS podatki ter pripadajoči metapodatki. <i>Formati:</i> izvorni NMR FID in obdelani podatki (.mnova), izvorni formati kromatografskih sistemov, izvozi v .csv oziroma .xlsx ter prikazi v .png in PDF/A. 4. Obsežni HPLC-HRMS podatki Orbitrap in njihova naknadna obdelava Pri analizi kompleksnih reakcijskih zmesi in afinitetno izbranih vzorcev bodo nastajale obsežne izvirne datoteke meritev v načinu full scan in po potrebi tandemskih MS meritev. Shranjene bodo izvirne datoteke instrumenta Thermo Scientific Orbitrap (.raw), zaporedja meritev, metode instrumenta, podatki o kalibraciji in kontroli kakovosti ter odprti izvozni formati, kadar bo pretvorba mogoča (npr. .mzML). Za vsako analizo bodo ohranjeni tudi parametri naknadne obdelave, različica programske opreme, sezname
-----	---	---

		iskanih elementnih sestav, masna toleranca, pravila za adukte in izotope, nastavitve detekcije vrhov, poravnave retencijskih časov, filtriranja in povezovanja signalov s spojinami oziroma reakcijskimi zmesmi. Rezultati bodo obsegali sezname značilk oziroma signalov z m/z, retencijskim časom, intenziteto, predlagano elementno sestavo, izotopskim vzorcem in, kadar bo na voljo, MS/MS-fragmentacijo. <i>Formati obdelanih podatkov:</i> .csv, .xlsx in/ali .json; skripte in delovni tokovi: .py, .ipynb oziroma izvorni formati programske opreme; grafični prikazi: .png, .svg in PDF/A. Izvirne datoteke .raw bodo vedno ohranjene nespremenjene, obdelani rezultati pa verzionirani in povezani z izvirno datoteko ter natančnim zapisom parametrov obdelave. 5. BLI podatki Pri biointerferometričnih meritvah bodo nastajali izvorni senzogrami vezave, podatki o referenčnih in kontrolnih meritvah, koncentracijah, serijah biosenzorjev, pripravi tarče in vzorcev, nastavitvah merilnega protokola ter temperaturi in časih posameznih korakov. Shranjeni bodo izvorni podatki instrumenta oziroma programske opreme (npr. .frd), izvoženi numerični podatki senzogramov v .csv ali .xlsx, parametri korekcije reference, poravnave, glajenja in prilaganja modela, izračunani odzivi ter po potrebi kinetični in ravnotežni parametri. Prikazi senzogramov bodo shranjeni v .png, .svg in PDF/A. Izvirni podatki bodo ohranjeni ločeno od vseh obdelanih in grafično prikazanih podatkov. 6. Podatki povezane platforme BLI-AS-MS in potrjevanja zadetkov Ustvarjeni bodo podatki, ki povezujejo oznako tarče, serijo BLI oziroma afinitetno-seleksijskega poskusa, reakcijsko zmes ali knjižnico, izvirno HRMS-datoteko, zaznani signal, predlagano strukturo, ponovno sintezo in rezultat potrditve vezave oziroma aktivnosti. Povezovalne preglednice in šifranti bodo shranjeni v .csv in .xlsx, odločitvena pravila, različice analiz ter razlogi za vključitev ali izključitev zadetkov pa v README (.txt ali .md) in drugi spremljevalni dokumentaciji. 7. Biokemijsko oziroma funkcionalno potrjevanje Za izbrane zadetke bodo nastajali izvorni numerični podatki encimskih oziroma drugih potrditvenih testov, krivulje koncentracijske odvisnosti, izračunani parametri in statistične analize. <i>Formati:</i> .csv, .xlsx, po potrebi .pzfx, ter grafični prikazi v .png, .svg in PDF/A. Lastniški formati so potrebni zaradi instrumentov in specializirane programske opreme. Kadar bo to tehnično smiselno in ne bo povzročilo izgube informacij, bodo podatki
--	--	---

		dodatno izvoženi v odprte oziroma široko uporabljane formate.
1.3	Kakšen je namen ustvarjanja, zbiranja oz. ponovne uporabe podatkov in njihova povezava s cilji projekta?	Namen ustvarjanja, zbiranja in ponovne uporabe podatkov je razviti integrirano platformo za razširjanje kemijskega prostora ter odkrivanje vezavnih ligandov za zahtevne biološke tarče. Podatki bodo nastajali v povezanem zaporedju, v katerem računalniško in kombinatorično načrtovanje z generativnimi modeli usmerja izbor gradnikov in zasnovo knjižnic; fotokatalitska sinteza ustvarja raznolike spojine in reakcijske zmesi; BLI in afinitetna selekcija zaznata vezavo na izbrane tarče; HPLC-HRMS Orbitrap omogoča iskanje in identifikacijo ustreznih signalov; ponovna sinteza in neodvisne meritve pa potrdijo identiteto in vezavo oziroma delovanje izbranih zadetkov. Podatki generativnega in kombinatoričnega načrtovanja bodo podpirali razvoj sintezne platforme in izbor kemijskega prostora. Sintezni in analitski podatki bodo omogočali optimizacijo fotokatalitskih reakcij, dokumentiranje sestave knjižnic ter preverjanje identitete in kakovosti spojin. BLI-senzogrami bodo zagotavljali neposredne podatke o vezavi, obsežni full-scan HRMS-podatki pa bodo omogočili naknadno, ponovljivo iskanje signalov in povezovanje vezavnega dogodka s kandidati v kompleksnih zmesih. Povezani podatki bodo uporabljeni za identifikacijo, ponovno sintezo in potrditev novih ligandov bakterijskih in humanih topoizomeraz ter za vrednotenje uporabnosti platforme pri drugih zahtevnih tarčah. Ohranjanje izvornih datotek, parametrov obdelave, kode, različic modelov in povezovalnih šifrantov je bistveno za sledljivost odločitev, ponovljivost naknadnega procesiranja in preverljivost objavljenih rezultatov.

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 Največji delež podatkov bodo predstavljale surove datoteke visokoločljivostne HPLC-HRMS analize v načinu neciljanega zajema podatkov (full scan Orbitrap HRMS), ki bodo nastajale med analizo kompleksnih reakcijskih zmesi, afinitetno izbranih vzorcev ter postopki identifikacije ligandov. Glede na predvideno trajanje projekta, obseg eksperimentalnega dela in delež neciljanih analiz ocenjujemo, da bo skupna količina ustvarjenih in hranjenih podatkov presegla 1 TB (do 10 TB). Poleg surovih datotek bomo hranili tudi pretvorjene podatke (npr. mzML), rezultate naknadne obdelave, sezname signalov, metapodatke, skripte za obdelavo podatkov, rezultate generativnih modelov ter surove in obdelane BLI senzograme.
2	Shranjevanje in varnostno kopiranje podatkov	
2.1	Kje bodo podatki med izvajanjem projekta shranjeni in varnostno kopirani?	Izvorni podatki bodo med nastankom shranjeni na lokalnih diskih računalnikov analiznih naprav oziroma na računalniški infrastrukturi, na kateri bodo izvedeni izračuni. Projektne kopije bodo urejeno prenesene v namensko mapno strukturo z nadzorovanim dostopom na infrastrukturi UL FFA. Dodatno bomo zagotavljali varnostno kopiranje na drug neodvisen nosilec ter v institucionalno okolje OneDrive/SharePoint. S tem bomo sledili načelu 3-2-1: tri kopije podatkov na dveh različnih vrstah nosilcev, od katerih bo ena na drugi lokaciji. Za velike Orbitrap .raw datoteke bo prenos in preverjanje varnostne kopije izvedeno čim prej po zaključeni sekvenci. Celovitost kopij bo preverjena najmanj s primerjavo velikosti datotek in, kjer bo izvedljivo, s kontrolnimi vsotami. Izvorne datoteke bodo nastavljene kot nespremenljivi vhod v nadaljnjo obdelavo; obdelani podatki, skripte, konfiguracije in rezultati bodo shranjeni ločeno ter verzionirani. Za izmenjavo med člani projektne skupine bomo uporabljali namensko skupino MS Teams/SharePoint z vnaprej določenimi pravicami dostopa. Za zelo velike datoteke se lahko uporabi varen institucionalni prenos, pri čemer javne storitve brez ustreznega nadzora dostopa ne bodo uporabljene za neobjavljene podatke.

2.2	Kako boste izbrali podatke za dolgoročno hrambo?	Izbor podatkov za dolgoročno hrambo bo temeljil na Pravilniku o stabilnem financiranju znanstvenoraziskovalne dejavnosti UL, zlasti 25. in 26. členu, na načelih FAIR ter na znanstveni vrednosti, preverljivosti in možnosti ponovne uporabe podatkov. Dolgoročno bomo hranili podatke, ki so potrebni za preverjanje in ponovitev objavljenih rezultatov: izvirne BLI-senzograme; izvirne Orbitrap .raw datoteke, neposredno povezane z objavljenimi ali potrjenimi rezultati; ključne kontrolne in slepe vzorce; odprte izvoze; končne sezname signalov in zadetkov; eksperimentalne protokole; metapodatke; šifrante; uporabljene skripte, konfiguracije in različice modelov; podatke o sintezi in karakterizaciji; ter obdelane in statistično ovrednotene podatke. Zaradi velike količine full-scan HRMS-podatkov bo izbor surovih datotek dokumentiran. Prednost bodo imele datoteke, povezane z objavami, potrjenimi zadetki, reprezentativnimi negativnimi rezultati, kontrolami in validacijo delovnega toka. Začasne kopije, predpomnilniki, podvojene delovne datoteke in vmesni izvozi, ki jih je mogoče zanesljivo ponovno ustvariti iz ohranjenih izvornih podatkov in dokumentiranega postopka, ne bodo predmet dolgoročne hrambe.
2.3	Ali bodo podatki shranjeni v zaupanja vrednem repozitoriju?	☒ Da ☐ Ne Ustrezno dokumentirani podatkovni nabori bodo objavljeni in dolgoročno shranjeni v Repozitoriju Univerze v Ljubljani (RUL), ki omogoča objavo raziskovalnih podatkov z metapodatki in trajnim identifikatorjem. Za programsko kodo, konfiguracije in večje podatkovne pakete se lahko uporabi tudi Zenodo, kadar je to primernejše glede na velikost in vrsto rezultata. Isti podatkovni nabor ne bo brez vsebinskega razloga podvojen v več repozitorijih; v drugem sistemu bo po potrebi objavljen le povezovalni metapodatkovni zapis. Če omejitve velikosti repozitorija ne bodo omogočale objave vseh izvornih Orbitrap .raw datotek, bodo odprto objavljeni najmanj metapodatki, reprezentativni oziroma z objavo povezani izvorni podatki, odprti izvozi, obdelani rezultati, skripte in natančen opis dostopa do preostalih dolgoročno hranjenih podatkov.
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 Zapisi v RUL bodo označeni s trajnim identifikatorjem vrste Handle. Podatkovnim in programskim objavam v Zenodu bo dodeljen DOI. Trajni identifikator bo naveden v znanstveni publikaciji, README dokumentaciji in drugih povezanih raziskovalnih rezultatih.
3.1.2	Kateri metapodatki bodo ustvarjeni in kateri metapodatkovni standardi bodo pri tem upoštevani?	Ustvarjeni bodo opisni, administrativni, tehnični in provenienčni metapodatki. Vključevali bodo naslov in opis nabora, ustvarjalce, institucijo, projekt J1-70052 in financerja, ključne besede, datum nastanka in objave, različico nabora, licenco, pogoje dostopa, povezano publikacijo oziroma programsko kodo ter formate datotek. RUL uporablja elemente standarda Dublin Core; v Zenodu bodo uporabljena njegova zahtevana metapodatkovna polja. Za računalniško načrtovanje bodo zabeleženi vir in različica podatkov, postopek standardizacije, delitev množic, različica kode in ogrodja, konfiguracija modela, hiperparametri, naključno seme, kontrolna vsota vhodnih datotek in kriteriji izbora. Za sintezo bodo zabeleženi identifikator serije, reagenti, pogoji, izvajalec, datum, uporabljeni protokol in povezava do spojine oziroma reakcijske zmesi. Za BLI bodo zabeleženi instrument in različica programske opreme, tip in serija biosenzorja, identiteta in priprava tarče, pufer, vzorec, koncentracija, temperatura, časi korakov, položaj na plošči, referenčne in kontrolne meritve ter parametri obdelave in prileganja. Za Orbitrap HRMS bodo zabeleženi instrument, različica programske opreme, metoda in sekvenca, polarnost, območje m/z, ločljivost, nastavitve vira, kalibracija, kromatografska metoda, vzorec, injekcija, serija, kontrolni vzorci, datum, izvajalec ter vsi parametri naknadne obdelave. Vsaka obdelana datoteka bo povezljiva z izvirno datoteko.
3.1.3	Ali bodo metapodatki vsebovali ključne besede za izboljšanje najdljivosti in možnosti ponovne uporabe?	☒ Da ☐ Ne Metapodatki bodo vključevali vsebinsko opredeljene ključne besede v slovenskem in angleškem jeziku. Predvideni izrazi so: farmacevtska kemija, organska sinteza, fotokemija, fotokataliza, kombinatorična kemija, generativni modeli, kemijske knjižnice, biointerferometrija, BLI, afinitetna selekcija, AS-MS, HPLC-HRMS, Orbitrap, full scan, masna spektrometrija, naknadna obdelava podatkov, topoizomeraze in odkrivanje zdravil. Ključne besede bodo vključene v repozitorijski zapis in README datoteko ter usklajene s terminologijo povezanih publikacij.
3.2	Zagotavljanje dostopnosti podatkov (A)	

3.2.1	Ali bodo vsi podatki odprto dostopni?	☐ Da ☒ Ne Vsi podatki ne bodo takoj in v celoti odprto dostopni. Začasna omejitev bo mogoča zaradi priprave znanstvenih objav, preverjanja kakovosti, varstva intelektualne lastnine in preprečevanja prezgodnjega razkritja rezultatov, ki bi lahko vplivalo na patentibilnost. Omejeni bodo samo podatki, za katere je omejitev utemeljena. Metapodatki bodo, kadar bo mogoče, objavljeni tudi v času embarga. Obsežnih izvornih Orbitrap .raw datotek morda zaradi tehničnih omejitev in nesorazmerne velikosti ne bo mogoče vseh neposredno objaviti v repozitoriju. V tem primeru bodo odprto objavljeni podatki, potrebni za preverjanje objavljenih zaključkov, skupaj z odprtimi izvozi, rezultati obdelave, skriptami in dokumentacijo; za preostale datoteke bo v metapodatkih naveden način utemeljenega dostopa. Po koncu postopkov zaščite in preverjanja bodo podatki odprti v največjem mogočem obsegu.
3.2.2	Kdaj bodo podatki odprto dostopni in za koliko časa?	Podatki, ki podpirajo znanstveno publikacijo, bodo, če ne bo pravnih, pogodbenih ali utemeljenih razlogov za embargo, odprto dostopni najpozneje ob objavi publikacije in v njej citirani s trajnim identifikatorjem. Pri potencialno patentibilnih rezultatih bo javna objava izvedena po vložitvi prijave oziroma po odločitvi, da zaščita ni potrebna. Podatki in metapodatki bodo v izbranem repozitoriju dostopni v skladu z dolgoročno politiko repozitorija. Če posameznih velikih izvornih datotek ne bo mogoče javno hraniti brez časovne omejitve, bodo trajno ohranjeni najmanj metapodatki, dokumentacija, odprti izvozi in informacije o možnosti dostopa.
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 bodo neobjavljeni podatki dostopni članom projektne skupine prek institucionalnega okolja z avtentikacijo in nadzorovanimi pravicami. Zunanji dostop bo mogoč po odobritvi vodje projekta oziroma skrbnika posameznega nabora ter po potrebi ob predhodni ureditvi zaupnosti, intelektualne lastnine in pogojev uporabe. Po zaključku projekta bo dostop do omejenih podatkov mogoče zahtevati pri kontaktni osebi, navedeni v metapodatkih. Zahteva bo presojena glede na pravne in pogodbene omejitve, varstvo intelektualne lastnine, namen uporabe, tehnično izvedljivost prenosa velikih datotek ter razpoložljivost dokumentacije. Projekt ne predvideva zbiranja osebnih ali posebnih vrst osebnih podatkov.

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 in ponovno analizo dela podatkov bo potrebna dodatna dokumentacija in specializirana programska oprema. Izvirne Thermo Scientific Orbitrap datoteke .raw zahtevajo ustrezna orodja proizvajalca ali združljivo programsko opremo; kadar bo mogoče, bodo dodani odprti izvozi .mzML. Izvirni BLI-podatki in senzogrami se analizirajo z namensko programsko opremo instrumenta, zato bodo priloženi tudi izvoženi numerični podatki in opis vseh korakov obdelave. Za .mnova, .cdxml in morebitne lastniške računalniške formate je potrebna ustrezna licenčna programska oprema. README bo za vsak nabor navedel ime proizvajalca in programske opreme, različico, operacijsko okolje, zahtevane dodatke, uporabljene parametre, strukturo map in postopek reprodukcije analize. Skripte, konfiguracije in okolja bodo deljena v strojno berljivih formatih, kadar to dovoljujejo licence. Dokumenti za dolgoročno branje bodo, kadar bo mogoče, izvoženi v .txt, .csv, .json, .mzML, .svg, .png in PDF/A.
3.3		
Zagotavljanje interoperabilnosti podatkov (I)		
3.3.1	Katere geslovnike oz. šifrante boste uporabili pri pripravi podatkov in metapodatkov?	Pri pripravi podatkov in metapodatkov bomo uporabljali uveljavljene identifikatorje in nadzorovane besednjake, kadar so primerni. Za biomedicinske pojme bomo uporabljali MeSH, za kemijske strukture standardizirane zapise SMILES, InChI in InChIKey, za proteine in encime uveljavljene identifikatorje podatkovnih zbirk ter dosledna imena tarč. Enote bodo zapisane v skladu s SI, analitski izrazi pa skladno z uveljavljeno terminologijo masne spektrometrije, kromatografije in biofizikalnih meritev. Ker projekt povezuje generativne modele, kombinatorične knjižnice, sintezne serije, BLI-poskuse in HRMS-sigale, bo uporabljen tudi lasten dokumentiran šifrant identifikatorjev spojin, reakcijskih zmesi, knjižnic, tarč, BLI-meritev, HRMS-sekvenc, izvornih datotek in potrjenih zadetkov. Šifrant ne bo nadomeščal standardnih identifikatorjev; zagotavljal bo sledljivo povezovanje podatkov med delovnimi sklopi in bo objavljen skupaj z ustreznim podatkovnim naborom.

3.3.2	Ali boste primorani uporabiti manj poznane ali lastne geslovnike oz. šifrante?	☒ Da ☐ Ne Lasten projektni šifrant bo potreben za povezovanje velikega števila virtualno generiranih spojin, reakcijskih knjižnic, vzorcev, BLI-senzogramov, HRMS-datoteke, zaznanih signalov in potrditvenih poskusov. Vseboval bo definicije polj, pravila poimenovanja, dovoljene vrednosti in povezave s standardnimi identifikatorji, kot so SMILES, InChIKey in identifikatorji tarč. Objavljen bo v odprtem strojno berljivem formatu .csv oziroma .tsv skupaj z README dokumentacijo.
3.4 Zagotavljanje ponovne uporabe podatkov (R)		
3.4.1	Na kakšen način boste zagotovili dokumentacijo, potrebno za ponovno uporabo podatkov?	Vsak objavljeni podatkovni nabor bo opremljen z README datoteko v formatu .txt ali .md. Ta bo vsebovala opis projekta in nabora, ustvarjalce, strukturo map, pomen datotek in stolpcev, enote, izvor podatkov, povezave med izvornimi in obdelanimi datotekami, uporabljene metapodatkovne standarde in geslovnike, programsko opremo in različice, postopke zbiranja in obdelave, kriterije vključitve ali izključitve signalov oziroma zadetkov, omejitve, licenco, priporočeni način citiranja ter kontaktno osebo. Za generativne modele bodo priloženi opis vhodnih množic, standardizacije, arhitekture oziroma vrste modela, hiperparametrov, semen in kriterijev izbora. Za BLI bodo dokumentirani protokol, referenčna korekcija in model prileganja. Za Orbitrap HRMS bodo dokumentirani zaporedje obdelave, parametri detekcije in filtriranja signalov ter povezava rezultatov z izvorno .raw datoteko. Kjer bo mogoče, bodo priloženi izvedljivi skripti, zvezki, konfiguracijske datoteke in primer zagona.
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 Odprto objavljeni raziskovalni podatki bodo praviloma licencirani z licenco CC BY 4.0. Programska koda bo objavljena pod ustrezno odprtokodno licenco, izbrano glede na uporabljene knjižnice in druge odvisnosti. Za ponovno uporabljene podatke in programsko opremo bodo spoštovane izvirne licence in pogoji uporabe. Podatki ali rezultati, ki so predmet embarga ali zaščite intelektualne lastnine, bodo licencirani šele ob odprtju dostopa.

3.4.3	Kakšne postopke zagotavljanja kakovosti podatkov boste uporabili?	Za zagotavljanje kakovosti, zanesljivosti, sledljivosti in ponovljivosti bomo: - redno vzdrževali, kalibrirali in preverjali delovanje raziskovalne opreme; - uporabljali standardizirane in dokumentirane protokole za fotokatalitsko sintezo, BLI, afinitetno selekcijo, HPLC-HRMS in potrditvene teste; - uporabljali slepe, referenčne in kontrolne vzorce ter tehnične ponovitve, kjer je to ustrezno; - preverjali identiteto in čistoto ponovno sintetiziranih spojin z neodvisnimi analitskimi metodami; - pri BLI dokumentirali kakovost nalaganja tarče, referenčno korekcijo, nespecifično vezavo, ponovljivost senzogramov in ustreznost izbranega modela prileganja; - pri Orbitrap HRMS spremljali kalibracijo, stabilnost retencijskih časov in mase, odziv kontrolnih vzorcev, prazne vzorce, prenos med injekcijami ter ponovljivost zaznave; - izvirne .raw datoteke ohranjali nespremenjene in vse obdelane rezultate povezali z različico programske opreme, skripte in konfiguracije; - različice seznamov signalov in zadetkov opremili z datumom, avtorjem, kriteriji filtriranja in razlogi za ročne posege; - generativne modele vrednotili na ločenih podatkovnih množicah ter dokumentirali standardizacijo, odstranjevanje dvojnikov, delitev podatkov, naključna semena, hiperparametre in merila kakovosti; - uporabljali enoten sistem poimenovanja in projektni šifrant za povezovanje spojin, zmesi, BLI-poskusov, HRMS-datotek in potrditvenih meritev; - hranili izvirne, obdelane in končne podatke v ločenih mapah z nadzorom različic; - ključne ročne prenose, izračune, identifikacije in povezave signalov s strukturami preverili po načelu pregleda drugega člana projektne skupine; - pred objavo izvedli notranji strokovni pregled podatkov, metapodatkov, kode in README dokumentacije ter upoštevali recenzijski postopek povezane znanstvene objave.
4.	Etični in pravni vidiki	

4.1	Ali obstajajo etična ali pravna vprašanja, ki bi lahko vplivala na deljenje podatkov?	☐ Da ☒ Ne Projekt ne predvideva osebnih podatkov, biološkega materiala človeškega izvora ali drugih podatkov, ki bi sami po sebi zahtevali etično odobritev. Na deljenje lahko začasno vplivajo pravice intelektualne lastnine, pogodbene obveznosti, licence ponovno uporabljenih podatkov in programske opreme ter omejitve pri razkrivanju modelov oziroma rezultatov pred vložitvijo morebitne patentne prijave. Te omejitve bodo obravnavane z nadzorom dostopa, embargom in predhodnim pregledom pred javno objavo.
4.2	Ali boste med izvajanjem projekta obdelovali oz. hranili osebne podatke?	☐ Da ☒ Ne V projektu ne bomo obdelovali ali hranili osebnih podatkov.
4.3	Ali bodo med projektom ustvarjene oz. ponovno uporabljene posebne vrste osebnih podatkov?	☐ Da ☒ Ne V projektu ne bodo ustvarjene ali ponovno uporabljene posebne vrste osebnih podatkov.
4.4	Kako boste uredili lastništvo podatkov in morebitne avtorske pravice na podatkih, ki jih boste ustvarili ali ponovno uporabili?	Lastništvo, upravljanje ter pravice do uporabe in objave raziskovalnih podatkov bodo urejeni s sporazumom o sodelovanju oziroma delitvi dela in sredstev ter z internimi pravili sodelujočih raziskovalnih organizacij. Lastnik podatkov bo praviloma organizacija, na kateri bodo podatki ustvarjeni, ob zagotavljanju pravic, potrebnih za izvedbo skupnega projekta, preverjanje rezultatov in pripravo objav. Za ponovno uporabljene podatke, programsko opremo, modele in zbirke spojin bodo upošteevane licence, pogoji uporabe in zahteve glede citiranja izvornih avtorjev oziroma upravljavcev. Pred odprto objavo kemijskih struktur, generiranih knjižnic, novih fotokatalitskih reakcij, potrjenih ligandov, kode ali modelov bo izveden pregled možnosti zaščite intelektualne lastnine. Prispevki sodelujočih k podatkom, kodi, modelom in publikacijam bodo dokumentirani.
5.	Drugi raziskovalni rezultati	

5.1	Ali boste poleg podatkov ustvarili ali ponovno uporabili tudi druge raziskovalne rezultate?	☒ Da ☐ Ne Poleg raziskovalnih podatkov bodo ustvarjeni tudi drugi raziskovalni rezultati: programska koda, skripte in računalniški delovni tokovi za kombinatorično načrtovanje, generativne modele in naknadno obdelavo HRMS-podatkov; konfiguracije in po potrebi kontrolne točke modelov; eksperimentalni protokoli za fotokatalitsko sintezo, BLI-AS-MS in potrjevanje zadetkov; ter fizične knjižnice spojin in ponovno sintetizirani potrjeni ligandi. Koda in delovni tokovi bodo, kadar to dovoljujejo licence in varstvo intelektualne lastnine, objavljeni v repozitoriju za kodo ter arhivirani v Zenodu ali Software Heritage s trajnim identifikatorjem in ustrezno odprtokodno licenco. Protokoli bodo objavljeni kot spremljevalna dokumentacija podatkovnega nabora oziroma v ustreznem repozitoriju. Za fizične spojine bo v publikaciji ali README opisan kontakt in pogoji morebitnega dostopa, ob upoštevanju razpoložljivosti, varnosti, pogodbenih obveznosti in intelektualne lastnine.
6.	Finančna sredstva	
6.1	Kakšni bodo stroški ravnanja s podatki in drugimi rezultati projekta po načelih FAIR in kako bodo kriti?	Stroški ravnanja s podatki bodo povezani z organizacijo, shranjevanjem, varnostnim kopiranjem, nadzorom celovitosti, dokumentiranjem, pretvorbo formatov, naknadno obdelavo, pripravo metapodatkov in objavo podatkov. Posebej pomembna bo hramba obsežnih full-scan Orbitrap .raw datotek, izvedenih podatkovnih zbirk ter različic analiz, skript in generativnih modelov. Na podlagi trenutno pričakovane velikosti 100–1000 GB je treba pred dokončno oddajo NRRP ponovno izračunati posredni strošek hrambe z aktualno različico orodja DSW Storage Cost Evaluator in z dejansko predvidenim obdobjem hrambe. V dokument zato namenoma ni vnesen neutemeljen natančen znesek. Priprava podatkov, pripadajočih metapodatkov, projektnega šifranta, README dokumentacije, odprtih izvozov in repozitorijskih objav bo del raziskovalnega dela v okviru projekta. Stroški bodo kriti iz sredstev projekta in obstoječe raziskovalne oziroma informacijske infrastrukture sodelujočih institucij. Ob posodobitvah NRRP bo ocena prilagojena dejanskemu obsegu Orbitrap podatkov in stroškom izbranega načina dolgoročne hrambe.

6.2	Kdo bo odgovorna oseba za ravnanje z raziskovalnimi podatki pri projektu?	Prof. dr. Janez Ilaš, vodja projekta, bo odgovoren za ravnanje z raziskovalnimi podatki celotnega projekta. Nosilci posameznih raziskovalnih nalog bodo odgovorni za pravilno ustvarjanje, poimenovanje, dokumentiranje, pravočasen prenos in preverjanje varnostnih kopij podatkov svojega delovnega sklopa. Tinka Leskovec, podatkovna svetovalka na UL FFA, bo zagotavljala podporo pri vprašanjih NRRP, repozitorijev, metapodatkov in načel FAIR.
------------	--	--

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>.
- Bezjak S., Grašič J., Korez B., Kverh B., Leskošek B., Romih T., Vipavc Brvar I., Zagorščak M., Županič A. (2025). *Priročnik o načrtovanju ravnanja z raziskovalnimi podatki*. Ljubljana: Založba Univerze v Ljubljani. [Dostopano DD. MM. LLLL]. Pridobljeno s: <https://nrrp.odprtaznanost.si/>
- *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 vprašalnika ARIS: 1.0 (z dne 15. 10. 2024)