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**	J1-70036
0.2	**Naziv projekta**	Oblikovne funkcije za razpad B → ρℓν̄ v kromodinamiki na mreži
0.3	**Šifra vodje projekta**	34446
0.4	**Ime in priimek vodje projekta**	Luka Leskovec
0.5	**Ime in priimek osebe, ki je v RO zadolžena za podporo pri ravnanju z raziskovalnimi podatki**	Mreža podatkovnih svetovalcev in kontaktnih oseb Univerze v Ljubljani (kontakt: mojca.krmelj@fmf.uni-lj.si).
0.6	**Interna pravila RO za ravnanje z raziskovalnimi podatki**	Pravila in podpora UL za odprto znanost in ravnanje z raziskovalnimi podatki: Pravilnik o stabilnem financiranju UL, uredba o izvajanju znanstvenoraziskovalnega dela v skladu z načeli odprte znanosti in Zakon o znanstvenoraziskovalni in inovacijski dejavnosti (ZZrID). Na ravni FMF: https://www.fmf.uni-lj.si/sl/knjiznica/odprta-znanost/.
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 Ponovno bodo uporabljeni že ustvarjeni podatki ansambla C13, vključno s spektri in matričnimi elementi, ki so bili objavljeni 10.1103/PhysRevLett.134.161901. Uporabljeni bodo tudi že ustvarjeni, še neobjavljeni izvorni in vmesni podatki, nastali v sodelovanju med letoma 2019-2024. Ti podatki tvorijo del približno 600 TB zbirke za nadaljnje analize in preverjanje rezultatov.
1.2	**Katere vrste podatkov boste ustvarili oz. ponovno uporabili in v katerih formatih bodo shranjeni?**	Ustvarjeni in ponovno uporabljeni bodo številčni podatki kromodinamike na mreži: korelacijske funkcije, spektri, matrični elementi, parametri prilagoditev, kovariančne matrike in prehodne amplitude. Javno objavljeni reproducibilni sklopi bodo vsebovali tudi potrebne vhodne podatke, INDEX/README ter spremljajoče skripte. Podatki bodo shranjeni in objavljeni v obliki ASCII (.dat) ali HDF5 (.h5). HDF5 je primeren za strukturirane velike numerične podatke, ASCII pa za pregledne tabelirane rezultate. Oba formata sta uveljavljena v poteku dela projekta.
1.3	**Kakšen je namen ustvarjanja, zbiranja oz. ponovne uporabe podatkov in njihova povezava s cilji projekta?**	Podatki omogočajo določitev spektra sistema pionov, matričnih elementov prehodov težkega mezona v dvo-pionsko stanje in nato neskončno-prostorninskih amplitud. Na ansamblih z različnimi mrežnimi razdaljami in masami kvarkov bodo uporabljeni za ekstrapolacijo k fizikalni točki in kontinuumskemu limitu. Končni cilj je določitev oblikovnih funkcij razpada B -> rho l nubar z ovrednotenim proračunom negotovosti; analogna analiza razpada D -> rho l nubar služi kot preveritev.
1.4	**Kakšna je pričakovana velikost podatkov, ki jih nameravate ustvariti oz. ponovno uporabiti?**	☐ 0-10 GB ☐ 10-100 GB ☐ 100-1000 GB x >1000 GB
2	**Shranjevanje in varnostno kopiranje podatkov**	
2.1	**Kje bodo podatki med izvajanjem projekta shranjeni in varnostno kopirani?**	Delovne in glavne kopije bodo hranjene v računalniški infrastrukturi ATOS na Institutu Jožef Stefan, Odseku za teoretično fiziko. Dolgoročna hramba poteka na tračnih sistemih; stroške sistemov in kaset krije projekt. Neodvisne varnostne kopije so na računalniških centrih NERSC in SuperMUC-NG (LRZ, München). Dostop do delovnih podatkov je omejen na projektno skupino in urejen po pravilih gostiteljskih centrov.

2.2	**Kako boste izbrali podatke za dolgoročno hrambo?**	Celotna zbirka izvornih in vmesnih podatkov bo zaradi možnosti ponovne analize dolgoročno hranjena v upravljani HPC in tračni infrastrukturi. Za javni repozitorij bo izbran zmanjšan, vendar za preverjanje objavljenih zaključkov zadosten sklop: končni rezultati, kovariančne matrike, potrebni vhodni podatki, dokumentacija in spremljajoče skripte. Izbor temelji na znanstveni vrednosti, reproducibilnosti, obveznostih odprte znanosti ter dogovoru vseh sodelujočih imetnikov pravic.
2.3	**Ali bodo podatki shranjeni v zaupanja vrednem repozitoriju?**	☒ Da ☐ Ne Javno objavljeni podatkovni sklopi bodo shranjeni bodisi v Repozitoriju Univerze v Ljubljani (RUL, https://repozitorij.uni-lj.si/), ki je institucionalni repozitorij UL ali na zenodo (https://zenodo.org). Celotna približno 600 TB zbirka ostane v upravljani HPC in tračni hrambi ter ni primerna za prenos v javni repozitorij.
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 Javno objavljeni sklopi v RUL/zenodo bodo označeni s trajnim identifikatorjem, ki bo naveden tudi v povezanih publikacijah.
3.1.2	**Kateri metapodatki bodo ustvarjeni in kateri metapodatkovni standardi bodo pri tem upoštevani?**	Metapodatki javnega sklopa bodo vsebovali naslov, avtorje, verzijo, datum objave, šifro projekta J1-70036, licenco, povezavo do publikacije in opis vsebine. Datoteki README in INDEX bosta na ravni podatkov opredelili oznako ansambla, datum in izvor generiranja, strukturo datotek, tip podatkov, format ter pomen spremenljivk. Za citatne metapodatke bodo uporabljena polja RUL, usklajena z elementi DataCite; za notranje datoteke bo uporabljena dokumentirana področna nomenklatura projekta.
3.1.3	**Ali bodo metapodatki vsebovali ključne besede za izboljšanje najdljivosti in možnosti ponovne uporabe?**	☒ Da ☐ Ne V RUL/zenodo bodo navedene angleške ključne besede: lattice QCD, B -> rho l nubar, in podobno.
3.2	**Zagotavljanje dostopnosti podatkov (A)**	
3.2.1	**Ali bodo vsi podatki odprto dostopni?**	☐ Da ☒ Ne Celotna zbirka izvornih in vmesnih podatkov obsega približno 600 TB in ni v celoti odprto dostopna. Omejitve so tehnične: obseg zbirke, stroški dolgoročne hrambe in prenosov ter razpoložljiva infrastruktura. Podatki ne vsebujejo osebnih podatkov. Za vsako publikacijo bo v RUL/zenodo odprto objavljen zmanjšan sklop, ki omogoča preverjanje zaključkov. Celoten sklop podatkov pa bo na voljo na zahtevo.
3.2.2	**Kdaj bodo podatki odprto dostopni in za koliko časa?**	Javno objavljeni podatkovni sklopi bodo dostopni ob objavi vsake povezane znanstvene publikacije. V RUL bodo dolgoročno dostopni skupaj z metapodatki in 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?**	Za izvorne in vmesne podatke, ki niso javno objavljeni, je dostop možen na obrazloženo prošnjo vodji projekta. Prošnja se presodi glede na dogovor sodelujočih imetnikov pravic, razpoložljivo infrastrukturo, stroške prenosa in obseg zahtevanih podatkov. Če je dostop izvedljiv, se zagotovi ustrezen podatkovni podsklop oziroma dostop prek ustreznega centra.
3.2.4	**Ali bo za dostop do podatkov oz. njihovo branje potrebna dodatna dokumentacija oz.**	☒ Da ☐ Ne

	informacija o ustrezni programski opremi?	Javno objavljeni sklopi bodo imeli datoteki README in INDEX z opisom strukture, formata, spremenljivk, konvencij in povezane publikacije. Datoteke .dat so berljive z običajnimi besedilnimi orodji; za .h5 je potreben standardni bralnik oziroma knjižnica HDF5.
3.3	**Zagotavljanje interoperabilnosti podatkov (I)**	
3.3.1	**Katere geslovnike oz. šifrante boste uporabili pri pripravi podatkov in metapodatkov?**	Uporabljena bo ustaljena nomenklatura kromodinamike na mreži in fizike delcev: oznake ansamblov (npr. C13, D5, E5, D6), standardne oznake delcev in procesov ter standardne fizikalne količine in enote. Struktura HDF5 in vsebinske oznake bodo opisane v README/INDEX.
3.3.2	**Ali boste primorani uporabiti manj poznane ali lastne geslovnike oz. šifrante?**	☐ Da ☒ Ne Ne bodo uporabljene lastne ontologije ali šifranti. Projektne oznake ansamblov in datotek so notranji identifikatorji, katerih pomen bo izrecno opisan v README/INDEX.
3.4	**Zagotavljanje ponovne uporabe podatkov (R)**	
3.4.1	**Na kakšen način boste zagotovili dokumentacijo, potrebno za ponovno uporabo podatkov?**	Vsak javni podatkovni sklop bo vseboval README in INDEX. Dokumentacija bo navedla izvor podatkov, avtorje, datum in verzijo, strukturo datotek, format, uporabljene konvencije, pogoje dostopa ter povezano publikacijo. Končni rezultati bodo spremljani s kovariančnimi matrikami, potrebnimi vhodnimi podatki in spremljajočimi skriptami, kadar potrebno
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 Javno objavljeni podatkovni sklopi v RUL/zenodo bodo licencirani s CC BY 4.0 (oz. ekvivalentom). Tretje osebe jih bodo lahko uporabljale v skladu z licenco ob ustrezni navedbi avtorjev in PID.
3.4.3	**Kakšne postopke zagotavljanja kakovosti podatkov boste uporabili?**	Kakovost se zagotavlja z nadzorom popolnosti in berljivosti datotek, doslednim verzioniranjem ter dokumentiranjem izvora in strukture podatkov. Znanstveno kakovost rezultatov se preverja s stabilnostjo prilagoditev glede na analitične izbire, primerjavami med ansambli in celovitim proračunom negotovosti. Pred objavo sklope pregleda projektna skupina.
4.	**Etični in pravni vidiki**	
4.1	**Ali obstajajo etična ali pravna vprašanja, ki bi lahko vplivala na deljenje podatkov?**	☐ Da ☒ Ne Projekt ne odpira etičnih vprašanj ali pravnih omejitev, ki bi preprečevale objavo podatkovnega sklopa, potrebnega za preverjanje publikacij.
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?**	Pravice do podatkov, ustvarjenih v projektu, se urejajo z dogovorom vseh sodelujočih raziskovalcev in sodelujočih ustanov; med njimi sta tudi IJS in UL FMF. Pred javno objavo posameznega sklopa se uskladi soglasje imetnikov pravic. Javno objavljeni sklopi v RUL/zenodo bodo opremljeni z licenco CC BY 4.0.
5.	**Drugi raziskovalni rezultati**	
5.1	**Ali boste poleg podatkov ustvarili ali ponovno uporabili tudi druge raziskovalne rezultate?**	☒ Da ☐ Ne Poleg podatkov bodo po potrebi pripravljene spremljajoče skripte. Objavljeni bodo skupaj s podatki in dokumentacijo, kadar so potrebni za razumevanje ali ponovitev analize; ne bodo vzdrževani kot samostojen programski produkt.
6.	**Finančna sredstva**	
6.1	**Kakšni bodo stroški ravnanja s podatki in drugimi rezultati projekta po načelih FAIR in kako bodo kriti?**	Neposredni stroški ravnanja s podatki so stroški tračnih sistemov in kaset za dolgoročno hrambo. Posredni stroški vključujejo delo projektne skupine pri pripravi, dokumentiranju in objavi zmanjšanih podatkovnih sklopov. Ti stroški se krijejo iz projekta; uporabljena bo tudi obstoječa infrastruktura ATOS/IJS, NERSC in SuperMUC-NG.
6.2	**Kdo bo odgovorna oseba za ravnanje z raziskovalnimi podatki pri projektu?**	Luka Leskovec, vodja projekta.

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