

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	Z2-70066
0.2	Naziv projekta	LANTERN-PPG: Large-scale Attentive Network for Transparent Explainable Representations of Non-invasive PPG
0.3	Šifra vodje projekta	50929
0.4	Ime in priimek vodje projekta	Gašper Slapničar
0.5	Ime in priimek osebe, ki je v RO zadolžena za podporo pri ravnanju z raziskovalnimi podatki	Lokalno na odseku E9 je to Junoš Lukan, ki se udeležuje skupnih dogodkov na nivoju IJS
0.6	Interna pravila RO za ravnanje z raziskovalnimi podatki	https://zic.ijs.si/odprta-znanost/#NRRP
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?	☐ <u>Da</u> ☐ Ne Uporabili bomo obstoječe javne podatkovne množice na področju PPG (npr. MIMIC-III in derivate, VitalDB, PulseDB, itd.) in remote PPG (npr. PURE, UBFC-rPPG, COHFACE, VIPL-HR-v2, MMSE-HR, MAHNOB-HCI, MMPD, etc.)
1.2	Katere vrste podatkov boste ustvarili oz. ponovno uporabili in v katerih formatih bodo shranjeni?	Uporabili bomo že obstoječe video in tabelarične podatke v povezavi s PPG in rPPG. Konkretno so tabelarični signali tipično v formatih CSV, MAT, medtem ko so video podatki najpogostejše v formatu MP4.
1.3	Kakšen je namen ustvarjanja, zbiranja oz. ponovne uporabe podatkov in njihova povezava s cilji projekta?	Uporabljeni podatki bodo ključni za učenje in validacijo predlaganih temeljnih modelov (r)PPG. Za to potrebujemo širok nabor po možnosti raznolikih podatkov, kjer bomo uporabili vse znane odprte podatkovne množice, kot navedeno. Svojih podatkov ne bomo zbirali. Če želimo v skladu s cilji projekta naučiti splošen in robusten razločljiv model, potrebujemo dovolj in dovolj raznolike podatke.
1.4	Kakšna je pričakovana velikost podatkov, ki jih nameravate ustvariti oz. ponovno uporabiti?	☐ 0–10 GB ☐ 10–100 GB ☐ 100–1000 GB ☒ <u>>1000 GB</u>
2	Shranjevanje in varnostno kopiranje podatkov	
2.1	Kje bodo podatki med izvajanjem projekta shranjeni in varnostno kopirani?	Med izvajanjem projekta bodo podatki hranjeni na odsečnih podatkovnih strežnikih, ki so dostopni le raziskovalcu in notranji IT službi. Dodatnih kapacitet ne predvidevamo, saj imamo cca. 100TB shrambe namenjene podatkom.
2.2	Kako boste izbrali podatke za dolgoročno hrambo?	Podatke bomo predvidoma hranili za trajanje projekta oz. učenje modelov. Gre za javno dostopne množice z lastnimi

		licencami, ki tipično dovoljujejo ponovno uporabo (npr. creative commons). Dolgoročna hramba ni predvidena.
2.3	Ali bodo podatki shranjeni v zaupanja vrednem repozitoriju?	☐ Da ☐ <u>Ne</u> Kot omenjeno bomo za rabo v projektu začasno podatke hranili na lokalnem strežniku z omejenim dostopom, gre pa v splošnem za prosto dostopne podatke, ki se hranijo v znanih repozitorijih (naprimer PhysioNet) in je tako deljenje že vzpostavljeno in uveljavljeno.
3.	Zagotovitev podatkov na način FAIR	
3.1	Zagotavljanje najdljivosti podatkov (F)	
3.1.1	Ali bodo podatki označeni s trajnim identifikatorjem (PID)?	☐ <u>Da</u> ☐ Ne Večina oz. vsi podatki v teh javnih množicah so vezani na znanstvene objave, ki imajo DOI. Poleg tega imajo tudi standarden URL.
3.1.2	Kateri metapodatki bodo ustvarjeni in kateri metapodatkovni standardi bodo pri tem upoštevani?	Uporabili bomo metapodatke, ki so v podatkovnih množicah že prisotni (npr. frekvenca vzorčenja, datum, seja snemanja, itd.). Dodatnih predvidoma ne bomo ustvarjali.
3.1.3	Ali bodo metapodatki vsebovali ključne besede za izboljšanje najdljivosti in možnosti ponovne uporabe?	☐ Da ☐ <u>Ne</u>
3.2	Zagotavljanje dostopnosti podatkov (A)	
3.2.1	Ali bodo vsi podatki odprto dostopni?	☐ <u>Da</u> ☐ Ne Uporabljeni podatki so vsi javni in odprto dostopni - na ta način jih bomo tudi sami pridobili za uporabo. Včasih je potrebno opraviti kakšen tečaj ali podpisati kakšen obrazec, kar smo že opravili (npr. v okviru PhysioNeta za določene zbirke)
3.2.2	Kdaj bodo podatki odprto dostopni in za koliko časa?	So že in bodo predvidoma tako tudi ostali.
3.2.3	Na kakšen način bo v primeru omejitev pri uporabi omogočen dostop do podatkov med izvajanjem projekta in po njegovem zaključku?	Omejitve so v skladu z licencami, ki jih te javne podatkovne množice imajo, vendar so tipično relativno permisivne (npr. CC 4.0), tako da se ne predpostavlja omejitev pri dostopu do podatkov. Naučeni modeli ne bodo izpostavljeni podatkov v surovi obliki, uteži pa bodo predvidoma javno dostopne na linku za deljenje (npr. github, če bo velikost omogočala)
3.2.4	Ali bo za dostop do podatkov oz. njihovo branje potrebna dodatna dokumentacija oz. informacija o ustrezni programski opremi?	☐ Da ☐ <u>Ne</u> Dodatna ne, le v skladu z branjem formatov (CSV, MAT, MP4)
3.3	Zagotavljanje interoperabilnosti podatkov (I)	

3.3.1	Katere geslovnike oz. šifrante boste uporabili pri pripravi podatkov in metapodatkov?	Ne bomo ustvarjali dodatnih metapodatkov. Nekateri v teh podatkih že obstajajo.
3.3.2	Ali boste primorani uporabiti manj poznane ali lastne geslovnike oz. šifrante?	☐ Da ☐ <u>Ne</u>
3.4	Zagotavljanje ponovne uporabe podatkov (R)	
3.4.1	Na kakšen način boste zagotovili dokumentacijo, potrebno za ponovno uporabo podatkov?	Za že obstoječe javne množice je dokumentacija na voljo na repozitorijih, kjer so hranjeni. Za uteži modela bomo pripravili osnovne podrobnosti, kot so datumi, avtorji, kratki primeri kode kako to uporabiti, README markdown.
3.4.2	Ali bodo vaši podatki javno dostopni in licencirani v skladu z odprto licenco CC0, da bo s tem omogočena čim širša ponovna uporaba?	☐ Da ☐ <u>Ne</u> Podatke bodo lahko uporabljale tretje osebe, saj so prosto dostopni in jih delijo drugi avtorji. Kar se tiče naših uteži (ponovno, ne gre za surove podatke ampak samo numerične anonimizirane vektorje in matrike brez jasnega pomena za človeka) bodo deljene v skladu z ustrezno licenco creative commons (potencialno CC0).
3.4.3	Kakšne postopke zagotavljanja kakovosti podatkov boste uporabili?	Za surove podatke bomo uporabili algoritme, ki ocenjujejo njihovo kvaliteto prek t.i. SQI = signal quality index. Naprimer pri signalih PPG gre v tem primeru za periodičnost, konsistenco, zaznavanje šuma in anomalij, itd.
4.	Etični in pravni vidiki	
4.1	Ali obstajajo etična ali pravna vprašanja, ki bi lahko vplivala na deljenje podatkov?	☐ Da ☐ <u>Ne</u>
4.2	Ali boste med izvajanjem projekta obdelovali oz. hranili osebne podatke?	☐ Da ☐ <u>Ne</u> Sicer video posnetki vsebujejo obraze ljudi, ampak kot omenjeno so ti deljeni v skladu z licencami in mednarodno zakonodajo (soglasja, itd.).
4.3	Ali bodo med projektom ustvarjene oz. ponovno uporabljene posebne vrste osebnih podatkov?	☐ <u>Da</u> ☐ Ne Pri video je razbrati rasno ali etnično poreklo. To je raziskovalno pomembno, saj je odprto vprašanje tudi robustnost rPPG glede na barvo kože. Kot omenjeno je to del obstoječih množic in se bomo držali pravil, licenc ter zakonodaje pri uporabi in začasni lokalni hrambi.
4.4	Kako boste uredili lastništvo avtorskih pravic in pravic	Kar se tiče uteži modelov, ki bodo edini ustvarjeni IP, se bodo avtorske pravice standardno oz. v skladu s pogodbo

	intelektualne lastnine podatkov, ki jih boste ustvarili ali ponovno uporabili?	delile med IJS in ARIS glede na podpisano oz. standardno prakso. V skladu s smernicami stremimo k odprtemu dostopu.
5.	Drugi raziskovalni rezultati	
5.1	Ali boste poleg podatkov ustvarili ali ponovno uporabili tudi druge raziskovalne rezultate?	☐ Da ☐ Ne Primarni output našega projekta bodo AI modeli (uteži nevronske mreže). Te uteži planiramo deliti in omogočiti ponovno uporabo v raziskovalni skupnosti v skladu s FAIR.
6.	Finančna sredstva	
6.1	Kakšni bodo stroški ravnanja s podatki in drugimi rezultati projekta po načelih FAIR in kako bodo kriti?	Ne predvidevamo dodatnih stroškov. Lokalni podatkovni strežnik že obstaja in se uporablja v okviru vsakodnevnega dela na odseku in IJS.
6.2	Kdo bo odgovorna oseba za ravnanje z raziskovalnimi podatki pri projektu?	Gašper Slapničar, vodja projekta, v posvetovanju z lokalnim predstavnikom za odprto znanost (odsečno Junoš Lukan, inštitutsko kdorkoli je tisti, h kateremu me usmeri odsečni predstavnik)

Uporabljeni viri:

- *Anotirana predloga načrta za ravnanje s raziskovalnimi podatki za projekte Obzorja Evropa*. CTK UL. Dostopno na: <https://dirrodata.ctl.uni-lj.si/raziskovalni-podatki/nact-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

Številka: 6311-94/2024-1

Datum: 15. 10. 2024