

Priloga A: Načrt za ravnanje z raziskovalnimi podatki

0	Splošne informacije	
0.1	Šifra projekta	GC 0006
0.2	Naziv projekta	GeoAI: Geoprostorske informacijske tehnologije za odporno in trajnostno družbo
0.3	Šifra vodje projekta	24340
0.4	Ime in priimek vodje projekta	Anka Lisec
0.5	Ime in priimek osebe, ki je v RO zadolžena za podporo pri ravnanju z raziskovalnimi podatki	Krištof Oštir, član projektne skupine (UL FGG)
0.6	Interna pravila RO za ravnanje z raziskovalnimi podatki	UL: https://www.uni-lj.si/raziskovanje/odprta-znanost
0.7	Verzija NRRP	Različica 1.0
1	Povzetek in opis raziskovalnih podatkov	
1.1	Ali boste pri projektu ponovno uporabili že obstoječe podatke predhodnih raziskav?	☒ Da ☐ Ne <u>UL FGG, ZRC SAZU in GeoZS</u> bodo v raziskavo vključevali odprto dostopne podatke, predvsem satelitske posnetke Sentinel-1, Sentinel-2, Landsat; razpoložljive zgodovinske letalski posnetke za obdobje pred letom 1980; državne podatke zračnega laserskega skeniranja iz cikla 2015 in 2025; državni ortofoto in druge podatkovne vire nacionalne prostorske informacijske infrastrukture (e-prostor). <u>UL FGG in ZRC SAZU</u> bosta ponovno uporabljala podatke satelitskih časovnih vrst, referenčne množice za (pol-) nadzorovano klasifikacijo rastrskih podatkov in oblakov točk; deloma se bodo v raziskavo vključili podatki za učenje algoritmov strojnega in globokega učenja (državni ortofoto, podatki, zajeti z brezpilotnimi letalniki z različnimi senzorji, podatki terestričnega skeniranja). <u>UL FRI</u> bo ponovno uporabljal referenčne podatke za učenje algoritmov strojnega in globokega učenja (slike, oblaki točk) ter javne zbirke za učenje in vrednotenje napovednih modelov v okviru strojnega učenja.. <u>IJS</u> bo ponovno uporabili referenčne množice za (pol-) nadzorovano večciljno klasifikacijo slik in tabelarnih podatkov (CIFAR-10, CIFAR-100, ImageNet-MLT, OpenML-CC18 ...), matematične vire iz OEIS.org ter javne zbirke za odkrivanje enačb (npr. dynobench) in za vrednotenje LLM-ov.

		<u>GeoZS</u>: bo poleg odprtodostopnih prostorskih podatkov uporabljal podatke preteklih raziskav o geoloških strukturah podpovršja, vrtnah, 3D modele in tabelarične geološke podatke.
1.2	Katere vrste podatkov boste ustvarili oz. ponovno uporabili in v katerih formatih bodo shranjeni?	<u>UL FGG</u> bo ponovno uporabil in ustvaril podatke satelitskih časovnih vrst, multimodalno dopolnjene z drugimi satelitskimi viri; metapodatke in maske neveljavnih vrednosti; posebna pozornost bo na zajemu podatkov z brezpilotnimi letalniki (optični, multispektralni in hiperspektralni senzorji) in mobilnim skenerjem (geoSLAM), ki bodo georeferencirani in ustrezno semantično obogateni. <u>UL FGG, ZRC SAZU in GeoZS</u> bodo ponovno uporabili in ustvarili georeferencirane prostorske podatke (satelitski, letalski, terestrični); georeferencirane in ortorektificirane rastrske podatke (optični, multispektralni, hiperspektralni); georeferencirane oblake točk; segmentirane/označene prostorske podatke (rastrski, oblaki točk); časovne vrste satelitskih in letalskih posnetkov; prostorske podatkovne zbirke z rastrskimi in vektorskimi podatki. <i>Formati prostorskih podatkov so standardni in jih podpirajo orodja za prostorsko analitiko in strojno/globoko učenje (npr. LAS, GeoTIFF, GeoDatabase, GeoPackage, GeoJSON, CSV, GML, itd.).</i> <u>UL FRI in IJS</u>: Številčni – matrike lastnosti in rezultate klasifikacij (CSV, Parquet). Slikovni – RGB/PNG, JPEG. Besedilni – matematična zaporedja (TXT), opis benchmarkov (JSON). <i>Formati so odprti/razširjeni in jih podpirajo priljubljena orodja za strojno in globoko učenje (ML/DL).</i>
1.3	Kakšen je namen ustvarjanja, zbiranja oz. ponovne uporabe podatkov in njihova povezava s cilji projekta?	<u>UL FGG, ZRC SAZU in GeoZS</u>: Podatki omogočajo razvoj in nepristransko vrednotenje novih metod za predobdelavo prostorskih podatkov in časovnih vrst prostorskih podatkov, (pol)samodejno integracijo različnih prostorskih podatkov, napredno analitiko in verifikacijo večciljne klasifikacije prostorskih podatkov; Poudarek je na pripravi in semantični obogatitvi referenčnih prostorskih podatkov za strojno/globoko učenje ter na podatkih za verifikacijo rezultatov; Posebna pozornost bo namenjena aplikativnim področjem 3D/4D modeliranja grajenega okolja (UL FGG in ZRC SAZU), upravljanju tveganj pri naravnih nesrečah (ZRC SAZU in GeoZS), arheološkim aplikacijam (ZRC SAZU) in geološkemu modeliranju (GeoZS). Rezultati neposredno prispevajo k ciljem GeoAI na področju avtomatizacije integracije prostorskih podatkov in prostorske analitike.

		<u>UL FRI in IJS</u>: Podatki omogočajo razvoj in nepristransko vrednotenje novih metod samonadzorovanega učenja, semantične segmentacije, časovnih napovedi, pol-nadzorovane večciljne klasifikacije učenja. Rezultati neposredno prispevajo k ciljem GeoAI na področju razločljivosti in avtomatizacije prostorske analitike in znanstvenega sklepanja.
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 <u>UL FGG</u>: glede na obseg razpoložljivih prostorskih podatkov in oceno novo zajetih podatkov bo velikost presegala 1000 GB (slikovni podatki, 3D modeli, drugi prostorski podatki). <u>UL FRI</u>: predvidena poraba prostora naj bi ostala pod 100 GB, natančna ocena bo podana ob prvi posodobitvi NRRP. <u>IJS</u>: predvidena poraba prostora naj bi ostala pod 100 GB, natančna ocena bo podana ob prvi posodobitvi NRRP. <u>ZRC SAZU</u>: glede na obseg zgodovinskih prostorskih podatkov in oceno novo zajetih podatkov bo velikost presegala 1000 GB (slikovni podatki, časovne vrste, drugi prostorski podatki). <u>GeoZS</u>: glede na obseg obstoječih geoloških prostorskih podatkov in oceno novo zajetih podatkov bo velikost presegala 1000 GB (slikovni podatki, časovne vrste, drugi prostorski podatki).
2	Shranjevanje in varnostno kopiranje podatkov	
2.1	Kje bodo podatki med izvajanjem projekta shranjeni in varnostno kopirani?	V okviru naloge bomo upravljali veliko količino prostorskih podatkov in podatkovne nize učnih množic za strojno učenje. Podatke bomo imeli shranjene in jih bomo pretežno obdelovali v okviru računalniške infrastrukture partnerjev projekta, deloma bomo uporabljali tudi javno raziskovalno infrastrukturo, ki je na voljo v Sloveniji za raziskovalne namene (računalniške gruče UL FGG, UL FRI, IJS, infrastruktura VEGA): - <u>UL FGG</u>: na lastnih strežnikih UL FGG z dnevnim samodejnim varnostnim kopiranjem; za tekoče delo bodo raziskovalci uporabljali tudi lokalne računalnike in po potrebi varne oblačne storitve za deljenje z drugimi partnerji. - <u>UL FRI</u>: v interni shrambi UL FRI, ki je sestavni del računske gruče z avtomatskim varnostnim kopiranjem.

		- <u>IJS</u>: na varovanih strežnikih IJS z dnevnim samodejnim varnostnim kopiranjem. - <u>ZRC SAZU</u>: na lastnih strežnikih in računalnikih z dnevnim samodejnim varnostnim kopiranjem; za tekoče delo bodo raziskovalci uporabljali tudi varne oblačne storitve za deljenje z drugimi partnerji. - <u>GeoZS</u>: na lastnih strežnikih in računalnikih z dnevnim samodejnim varnostnim kopiranjem; za tekoče delo bodo raziskovalci uporabljali tudi varne oblačne storitve za deljenje z drugimi partnerji. Za obdelavo satelitskih posnetkov bomo uporabljali infrastrukturo CREODIAS v okviru Copernicus Data Space Ecosystem (CDSE), ker omogoča obdelavo na način »blizu podatkov« (manj prenosov velikih datotek in hitrejši pretok dela), neposreden in zanesljiv dostop do obsežnih podatkovnih arhivov Copernicus/Sentinel, ter enostavno skaliranje računske moči glede na potrebe projektnih faz (od testiranja do množične produkcije). Poleg tega infrastruktura podpira sodobne in standardizirane pristope (API-ji, avtomatizirani delovni tokovi, kontejnerji/orkestracija), kar povečuje ponovljivost in sledljivost rezultatov, olajša integracijo z obstoječimi orodji ter izboljša operativno robustnost (monitoring, upravljanje virov, varnost). Ključni razlogi so torej hitrost in učinkovitost obdelave, skalabilnost, standardizirana integracija, ponovljivost/sledljivost ter nižji operativni stroški in manjša kompleksnost upravljanja infrastrukture.
2.2	Kako boste izbrali podatke za dolgoročno hrambo?	Dolgoročno hranimo podatke, ki podpirajo objave ali imajo potencial za nadaljnjo znanstveno rabo – skladno z Uredbo o odprti znanosti in s smernicami ARIS. Za vse prostorske podatke bodo na voljo metapodatki, skladno z mednarodnimi standardi Podatki, ki dokazano prispevajo k razvoju in benchmarkingu metod strojnega učenja, bodo trajno arhivirani; manj relevantni vmesni rezultati se bodo po potrebi anonimizirali ali izločili.
2.3	Ali bodo podatki shranjeni v zaupanja vrednem repozitoriju?	☒ Da ☐ Ne Na področju objave rezultatov raziskav upoštevamo načelo odprtega dostopa do znanstvenih objav, ki ga je Agencija Republike Slovenije za znanstvenoraziskovalno in inovacijsko dejavnost začela uvajati že na podlagi Nacionalne strategije odprtega dostopa do znanstvenih objav in raziskovalnih podatkov v Sloveniji 2015-2020 (2015). Skladno z zahtevami ARIS bomo zagotovili odprti dostop do vseh recenziranih znanstvenih objav, ki so rezultat tega projekta. Kot rezultate raziskav štejemo

		znanstvene objave, kot so znanstveni članki in druge vrste recenziranih publikacij, raziskovalni podatki, programska oprema, ki je nastala kot rezultat raziskav, ter druge vrste rezultatov raziskav. Končne sklope podatkov in modele objavimo v Zenodo (DOI, trajna hramba) ter kodo v GitHub/Zenodo, ki omogočajo hrambo rezultatov raziskav v digitalni obliki.
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 Načrtovana je uporaba PID, ki jih ponujajo sami repozitoriji, v katerih bodo naloženi podatki (npr. Zenodo, ki ponuja DOI identifikator)
3.1.2	Kateri metapodatki bodo ustvarjeni in kateri metapodatkovni standardi bodo pri tem upoštevani?	Za vse prostorske podatkovne nize bodo pripravljene opisni podatki – metapodatki, skladno s standardom SIST EN ISO 19115-1:2015: Geografske informacije – Metapodatki in SIST EN ISO 19115-2:2009: Geografske informacije - Metapodatki - 2. del: Razširitev za podobe in mrežne podatke, kjer ločimo med obveznimi vsebinami (O) in izbirnimi vsebinami (I): - Naslov podatkovnega niza (O), - Referenčni datum podatkovnega niza (O), - Prostorski obseg (razsežnost) (O), - Prostorski referenčni sistem (O), - Povzetek opisa podatkovnega niza (O), vključno z vsebino in vrsto podatka, - Format zapisa podatka (O), - Struktura podatkovnega niza (I), - Vrsta prostorske predstavitve (I), - Vir, vključno z opisom metode zajema podatkov (O), - Prostorska ločljivost podatkovnega niza (I), - Kakovost (I), - Dodatne informacije o podatkovnem nizu (I), - Odgovorna oseba za podatkovni niz (I), - Spletni viri (I), - Jezik podatkovnega niza (O), - Kontaktna oseba za metapodatke (O), - Datum zapisa metapodatka (O), - Datum posodobitve zapisa metapodatka (I), - Pogoji uporabe/licenca (O). Dodatno bodo za podatke strojnega učenja uporabljeni metapodatki Dublin Core + razširitve iz ontologij OntoDM-core, OntoDT, OntoDM-KDD in OPTION.

3.1.3	Ali bodo metapodatki vsebovali ključne besede za izboljšanje najdljivosti in možnosti ponovne uporabe?	☒ Da ☐ Ne Zapisi v Zenodo/GitHub bodo opremljeni z opisnimi ključnimi besedami za lažje iskanje, ključne besede bodo tudi del metapodatkov za prostorske podatke.
3.2	Zagotavljanje dostopnosti podatkov (A)	
3.2.1	Ali bodo vsi podatki odprto dostopni?	☐ Da ☒ Ne Večina podatkov bo odprto dostopna. Izjema so osebni podatki, ki bodo zbrani le v določene namene in ustrezno anonimizirani (več v točki 4.2). <u>GeoZS</u> bo lahko uporabljal zaupne podatke, predvsem podatke iz preteklih raziskav rudišč ali geotermalne energije, ki se lahko še vedno smatrajo kot poslovna skrivnost. V tem primeru bo GeoZS pridobil ustrezna dovoljenja, žal ne bomo mogli zagotoviti popolno odprto dostopnost do takšnih podatkov.
3.2.2	Kdaj bodo podatki odprto dostopni in za koliko časa?	Ob objavi znanstvene publikacije; dostop brez časovne omejitve.
3.2.3	Na kakšen način bo v primeru omejitev pri uporabi omogočen dostop do podatkov med izvajanjem projekta in po njegovem zaključku?	Posebnih omejitev ni predvidenih; kjer veljajo, bo možen dostop na daljavo ob prijavi in soglasju vodje projekta. V primeru uporabe zaupnih podatkov bo protokol ravnanja, uporabe in shrambe določen skladno z internimi GeoZS predpisi o ravnanju z zaupnimi podatki (vezano na točko 3.2.1).
3.2.4	Ali bo za dostop do podatkov oz. njihovo branje potrebna dodatna dokumentacija oz. informacija o ustrezni programski opremi?	☐ Da ☒ Ne
3.3	Zagotavljanje interoperabilnosti podatkov (I)	
3.3.1	Katere geslovnike oz. šifrante boste uporabili pri pripravi podatkov in metapodatkov?	Geoinformacijske slovarje, INSPIRE, šifranti podatkovnih zbirk nacionalne prostorske informacijske infrastrukture, OntoDM-core/DT/KDD, OPTION; WordNet-ML pojmi; OpenML schema za tabularne podatkovne množice.
3.3.2	Ali boste primorani uporabiti manj poznane ali lastne geslovnike oz. šifrante?	☒ Da ☐ Ne Da, po potrebi bomo razširili geslovnik na področju geoinformatike in umetne inteligence na tem področju;

		uporabili in razširili bomo tudi lastne ontologije za strojno učenje in optimizacijo.
3.4	Zagotavljanje ponovne uporabe podatkov (R)	
3.4.1	Na kakšen način boste zagotovili dokumentacijo, potrebno za ponovno uporabo podatkov?	Način dokumentiranja še ni določen; pripravili bomo najmanj datoteko README z osnovnimi metapodatki.
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 dostopni nabori bodo licencirani CC0, obstoječi podatki in modeli, ki bodo osnovani na internih podatkovnih zbirkah bodo predvidoma licencirani z CC-BY-NC (prosta uporaba za nekomercialne namene z navedbo vira), podrobna pravila se bo določilo za vsak podatkovni niz.
3.4.3	Kakšne postopke zagotavljanja kakovosti podatkov boste uporabili?	Standardizacija protokolov, kalibracija instrumentov, sledljivost verzij podatkov/kode in statistične kontrole kakovosti. Kakovost prostorskih podatkov se bo izvajala skladno z ISO 19157-1:2023 standardom, kjer se bo z vzorčenjem preverjalo vse ključne elemente kakovosti podatkov. Tudi za ostale raziskovalne podatke in rezultate bomo izvajali kontrolo kakovosti z ustreznim vzorčenjem.
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 V okviru diseminacijskih in komunikacijskih dejavnosti bomo občasno zbirali tudi osebne podatke za namene organizacije dogodkov (registracija udeležencev), vendar za točno določen namen in samo za omejeno časovno obdobje (za čas organizacije dogodkov). Ti podatki se ne smejo uporabljati za druge namene in ne smejo biti pod nobenim pogojem javni. Udeležba pri teh dejavnostih je prostovoljna in zahteva predhodno registracijo, kjer se udeleženci strinjajo z deljenjem objav, kot so novice s fotografijami. Morebitne ankete in druge raziskave med uporabniki in raziskovalci bomo izvajali na način, da bomo zagotovili najvišjo stopnjo anonimnosti. Pri tem bomo strogo

		upoštevali zahteve GDPR in ZVOP, vključno s pridobivanjem privolitve za obdelavo osebnih podatkov.
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 so urejene v konzorcijskem sporazumu; avtorske in IP-pravice praviloma pripadajo posamezni raziskovalni organizaciji, ki podatke ustvari.
5.	Drugi raziskovalni rezultati	
5.1	Ali boste poleg podatkov ustvarili ali ponovno uporabili tudi druge raziskovalne rezultate?	☒ Da ☐ Ne Procesni modeli, programske kode
6.	Finančna sredstva	
6.1	Kakšni bodo stroški ravnanja s podatki in drugimi rezultati projekta po načelih FAIR in kako bodo kriti?	Posebni stroški niso predvideni; uporabili bodo obstoječe strežnike in infrastrukturo partnerskih organizacij.
6.2	Kdo bo odgovorna oseba za ravnanje z raziskovalnimi podatki pri projektu?	<u>UL FGG</u> : prof. dr. Anka Lisec <u>UL FRI</u> : izr. prof. dr. Luka Čehovin Zajc <u>IJS</u> : prof. dr. Sašo Džeroski <u>ZRC SAZU</u> : izr. prof. dr. Žiga Kokalj <u>GeoZS</u> : doc. dr. Mateja Jemec Auflič

Podpisana izjavljam, da so z vsebino Načrta za ravnanje z raziskovalnimi podatki seznanjeni vsi konzorcijski partnerji ter se z njo strinjajo.

Podpis:


Anka Lisec

(Zastopnik oz. pooblaščen oseb raziskovalne organizacije, ki je pri projektu v vlogi vodje konzorcija)

Verzija dokumenta: 1.0

Datum: 30. 12. 2025