

SPHERE: Načrt ravnanja z raziskovalnimi podatki

0	Splošne informacije	
0.1	Šifra projekta	J7-70247
0.2	Naziv projekta	SPHERE - Vrednotenje okoljskih pojavov z informiranim globokim učenjem na podlagi podatkov opazovanja Zemlje
0.3	Šifra vodje projekta	37956
0.4	Ime in priimek vodje projekta	dr. Marko Bizjak
0.5	Ime in priimek osebe, ki je v RO zadolžena za podporo pri ravnanju z raziskovalnimi podatki	dr. Tamara Gomilšek, podatkovni skrbnik za ERC področje Naravoslovje in tehnika, UM FERl
0.6	Interna pravila RO za ravnanje z raziskovalnimi podatki	Politika odprte znanosti Univerze v Mariboru (POZUM). Spletna stran: https://www.um.si/raziskovanje/odprta-znanost/politika-odprte-znanosti-univerze-v-mariboru-pozum/
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 SPHERE uporablja obstoječe nacionalne in evropske podatkovne vire. Ponovno bomo uporabili oblake točk zračnega laserskega zajema, satelitske posnetke, fotogrametrične posnetke, meteorološke podatke in vektorske sloje o rabi tal ter stavbah. Podatki bodo uporabljeni za predstavitev realnega okolja, v katerem bomo vrednotili pojave.
1.2	Katere vrste podatkov boste ustvarili oz. ponovno uporabili in v katerih formatih bodo shranjeni?	V projektu bomo ustvarili kurirane podatke opazovanja Zemlje (GeoTIFF, GraphML) in tematske sloje pilotov (GeoTIFF). Ponovno bomo uporabili oblake točk zračnega laserskega skeniranja - LiDAR (LAS/LAZ), arhivske fotogrametrične in nemetrične posnetke (TIFF/JPEG), odprte satelitske podatke Copernicus in Landsat (GeoTIFF, NetCDF), meteorološke podatke ARSO (CSV) ter vektorske sloje o rabi tal in stavbah (SHP, GeoPackage). Gre torej za geoprostorske (rastrske, vektorske, 3D), numerične in slikovne podatke. Izbrani formati so odprti in standardni ter uveljavljeni v geoinformatiki in daljinskem zaznavanju, kar zagotavlja berljivost z odprtimi orodji ter omogoča deljenje in ponovno uporabo.
1.3	Kakšen je namen ustvarjanja, zbiranja oz. ponovne uporabe podatkov in njihova povezava s cilji projekta?	Ustvarjene in ponovno uporabljene podatke bomo uporabili za učenje, validacijo in primerjavo emulatorjev okoljskih pojavov, ki temeljijo na fizikalno informiranih nevronske mrežah. Ponovno uporabljeni viri (LiDAR, fotogrametrični in satelitski posnetki, meteorološki podatki) zagotavljajo realno predstavitev okolja, na novo ustvarjeni podatki pa referenco za merjenje točnosti napovedi.
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?	Med izvajanjem projekta bodo podatki shranjeni na podatkovnem strežniku projekta znotraj omrežja RO. Strežnik ima urejeno zaščito dostopa in redno varnostno kopiranje podatkov.
2.2	Kako boste izbrali podatke za dolgoročno hrambo?	Sledili bomo smernicam Politike odprte znanosti Univerze v Mariboru. Za dolgoročno hrambo bomo izbrali vse izvedene podatkovne zbirke, ki podpirajo objavljene ugotovitve in jih nameravamo shraniti v zaupanja vrednem javnem repozitoriju.
2.3	Ali bodo podatki shranjeni v zaupanja vrednem repozitoriju?	☒ Da ☐ Ne Podatki bodo dolgoročno shranjeni v Zenodo ali v Digitalni knjižnici Univerze v Mariboru – DKUM (https://dk.um.si/info/index.php/slo/), ki sta oba zaupanja vredna javna repozitorija.
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 Vse podatkovne zbirke bodo označene z DOI v primeru uporabe Zenodo ali s PID-i, ki bodo dodeljeni s strani repozitorija DKUM.
3.1.2	Kateri metapodatki bodo ustvarjeni in kateri metapodatkovni standardi bodo pri tem upoštevani?	Podatkovne zbirke bomo opisali z metaoznaki po standardu Dublin Core, shranjenimi v datoteki v formatu XML. Poleg tega bomo dodali tudi vse dodatne opise, ki jih ob oddaji podatkov zahteva repozitorij (na osnovi standarda DataCite Metadata Schema).
3.1.3	Ali bodo metapodatki vsebovali ključne besede za izboljšanje najdljivosti in možnosti ponovne uporabe?	☒ Da ☐ Ne V metapodatke bo vključena tudi množica (5-10) ključnih besed za boljšo najdljivost podatkov.
3.2	Zagotavljanje dostopnosti podatkov (A)	
3.2.1	Ali bodo vsi podatki odprto dostopni?	☐ Da ☒ Ne Pri določanju odprtosti podatkov bomo upoštevali načelo "odprti, kolikor je mogoče, in zaprti, kolikor je nujno". Če pri katerem od virov licenca ponovne distribucije ne dopušča, bomo namesto podatkov objavili metapodatke o viru in postopku dostopa.
3.2.2	Kdaj bodo podatki odprto dostopni in za koliko časa?	Podatki bodo privzeto na voljo ob objavi znanstvene publikacije. Naloženi bodo v zaupanja vreden javni repozitorij, kjer bodo dolgoročno shranjeni, skupaj z vsemi metapodatki. Če bodo nekateri podatki ključni za morebitno zaščito intelektualne lastnine (patenti), bomo takšne podatke objavili šele po zaključeni fazi zaščite intelektualne lastnine.

3.2.3	Na kakšen način bo v primeru omejitev pri uporabi omogočen dostop do podatkov med izvajanjem projekta in po njegovem zaključku?	Omejitev pri uporabi objavljenih podatkov ne predvidevamo, na voljo bodo v odprtem dostopu pod licenco CC0 ali CC BY.
3.2.4	Ali bo za dostop do podatkov oz. njihovo branje potrebna dodatna dokumentacija oz. informacija o ustreznih programski opremi?	☐ Da ☒ Ne Za branje podatkov je na voljo množica odprtokodnih rešitev v različnih programskih jezikih, npr. knjižnica GDAL in orodje PDAL, namizni program QGIS, Python knjižnice rasterio, geopandas in NetworkX, itd.
3.3	Zagotavljanje interoperabilnosti podatkov (I)	
3.3.1	Katere geslovnike oz. šifrantе boste uporabili pri pripravi podatkov in metapodatkov?	Na obravnavanem raziskovalnem področju še ne obstajajo standardni geslovniki ali ontologije, zato bomo podatke opisali s splošnimi standardi za metapodatke (Dublin Core) in s podrobnim opisom v datoteki »readme« v strojno berljivem formatu.
3.3.2	Ali boste primorani uporabiti manj poznane ali lastne geslovnike oz. šifrantе?	☐ Da ☒ Ne Podatkov ne bomo opisali z lastnimi geslovniki oz. šifrantmi.
3.4	Zagotavljanje ponovne uporabe podatkov (R)	
3.4.1	Na kakšen način boste zagotovili dokumentacijo, potrebno za ponovno uporabo podatkov?	Dokumentacija bo zagotovljena s priloženo datoteko »readme« (.TXT), napisano v angleškem jeziku. Zajemala bo naslov, avtorje, datum in verzijo, ključne besede, trajni identifikator in pogoje dostopa, licenco, nato pa podroben opis podatkov, njihovih struktur in velikosti, opis uporabljene programske opreme, postopka in parametrov, s katerimi so bili podatki generirani. Pri geoprostorskih nizih bomo dodatno navedli koordinatni sistem, prostorski obseg, prostorsko in časovno ločljivost ter izvirne podatkovne vire, iz katerih so bili izpeljani.
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 Vse podatkovne zbirke bodo na voljo pod licenco CC BY ali CC0.
3.4.3	Kakšne postopke zagotavljanja kakovosti podatkov boste uporabili?	Kakovost bomo zagotavljali z enotnimi protokoli za zajem in obdelavo podatkov, ki jih bodo uporabljali vsi sodelujoči na projektu, ter z dokumentiranimi postopki obdelave, ki odstranijo šum, zapolnijo vrzeli in poenotijo formate. Kakovost bomo ocenjevali s primerjavo napovedi z referenčnimi rezultati s standardnimi merami odstopanja.
4	Etični in pravni vidiki	
4.1	Ali obstajajo etična ali pravna vprašanja, ki bi	☒ Da ☐ Ne

	lahko vplivala na deljenje podatkov?	Projekt ne obdeluje osebnih podatkov in ne izvaja raziskav na ljudeh ali živalih, zato etičnih zadržkov ni. Pravno vprašanje predstavljajo pogoji uporabe vhodnih podatkovnih virov tretjih oseb in morebitna zaščita intelektualne lastnine.
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?	Avtorske pravice in lastništvo za ustvarjene podatke bodo pripadali organizaciji, ki bo te podatke generirala/zajela.
5	Drugi raziskovalni rezultati	
5.1	Ali boste poleg podatkov ustvarili ali ponovno uporabili tudi druge raziskovalne rezultate?	☒ Da ☐ Ne Poleg podatkov bomo ustvarili programsko opremo, diseminacijske materiale ter znanstvene publikacije. Tudi te rezultate bomo v največji možni meri naredili javno dostopne pod licenco CC BY oziroma ustrezno odprtokodno licenco in jih objavili na spletni strani projekta, v javnem repozitoriju kode oz. v certificiranem javnem repozitoriju.
6	Finančna sredstva	
6.1	Kakšni bodo stroški ravnanja s podatki in drugimi rezultati projekta po načelih FAIR in kako bodo kriti?	Stroške shranjevanja in obdelave podatkov med izvajanjem projekta ter stroške glavnega podatkovnega strežnika projekta znotraj omrežja RO za čas trajanja projekta in še 5 let po koncu projekta bo krila Fakulteta za elektrotehniko, računalništvo in informatiko Univerze v Mariboru. Za dolgoročno shrambo podatkov bomo uporabili repozitorija Zenodo ali DKUM, ki imata lastno financiranje. Stroški objav člankov v revijah z odprtim dostopom bodo kriti iz projektnih sredstev.
6.2	Kdo bo odgovorna oseba za ravnanje z raziskovalnimi podatki pri projektu?	Vodja projekta, dr. Marko Bizjak.