

Plan upravljanja istraživačkim podacima projekta ADEPT

Šegvić, Siniša

Data management plan / Plan upravljanja istraživačkim podacima

Publication year / Godina izdavanja: **2023**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:168:649051>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom](#).

Download date / Datum preuzimanja: **2024-11-04**



Repository / Repozitorij:

[FER Repository - University of Zagreb Faculty of
Electrical Engineering and Computing repository](#)



Plan upravljanja istraživačkim podacima

Opće informacije		
	Ime i prezime predlagatelja	Siniša Šegvić
	Matična organizacija	Sveučilište u Zagrebu Fakultet Elektrotehnike i računarstva
	Naziv projekta	Napredna gusta predikcija za računalni vid
	Upravitelj podacima	Siniša Šegvić
1.	Prikupljanje podataka i dokumentacija	
	Koje ćete podatke prikupljati, obrađivati, stvarati ili se ponovno njima koristiti? (navedite format, vrstu i opseg podataka)	<p>Naše projektne aktivnosti dovode do stvaranja sljedećih digitalnih sadržaja:</p> <ol style="list-style-type: none"> podrška za reprodukciju eksperimenata iz objavljenih članaka: izvorni kod, konfiguracijske datoteke, kratke upute za korištenje, skupovi slika i oznaka za učenje, validaciju i ispitivanje sirovi eksperimentalni podatci: napredovanje gubitka kroz epohe učenja, naučeni parametri, generalizacijski učinak, kopije podataka, predikcije modela itd
	Kako će se podaci prikupljati, obrađivati ili stvarati? (ukratko navedite metodologiju i procese osiguranja kvalitete te načine organiziranja podataka)	<p>Podrška za reprodukciju eksperimenata predstavlja primarni istraživački rezultat projekta. Organiziranje, održavanje i verziranje provodimo u privatnim projektima mrežnog servisa GitHub. Po objavljivanju znanstvenog rada, prvi autor uređuje javno dostupnu mrežnu stranicu s pročišćenom službenom inačicom izvornog koda, eksperimentalnog protokola i skupovima naučenih parametara.</p> <p>Prvi autor publikacije organizira čuvanje sirovih eksperimentalnih podataka nakon publiciranja, pri čemu trajanje čuvanja ovisi o količini podataka, dostupnom kapacitetu i kvarovima. Originalni podatci nalaze se na lokalnom datotečnom sustavu, dok pregledne sažetke smještamo u privatne tablice servisa Google Sheets.</p>
	Koju ćete dokumentaciju i metapodatke ustupiti osim podataka? (navedite koje su informacije potrebne korisnicima kako bi mogli čitati i interpretirati podatke u budućnosti te koji će se standardi koristiti pri tumačenju podataka)	<p>Službena podrška za reprodukciju eksperimenata može sadržavati izvorni kod, konfiguracijske datoteke, kratke upute za korištenje te skupove slika i oznaka za učenje, validaciju i ispitivanje.</p> <p>U nastavku su adrese mrežnih stranica sa službenom podrškom za reprodukciju eksperimenata iz ovog razdoblja:</p> <p>https://github.com/pb-brainiac/semseg_od https://github.com/UNIZG-FER-D307/universal_taxonomy https://github.com/jsaric/panoptic-swiftnet https://github.com/UNIZG-FER-D307/semisup-seg-efficient https://github.com/matejgrcic/DenseHybrid</p>

2.	Pravna i sigurnosna pitanja	
	<p>Jeste li ograničeni sporazumom o povjerljivosti? Imate li potrebna dopuštenja za prikupljanje, obradu, čuvanje i dijeljenje podataka? Jesu li osobe čiji se podaci pohranjuju informirani o tome i jesu li dali privolu? Kojim ćete se metodama koristiti u svrhu zaštite osjetljivih podataka (GDPR - posebne kategorije osobnih podataka)?</p>	<p>Većina naših istraživanja provodi se na javnim skupovima podataka (npr. ImageNet, Cityscapes, COCO, Vistas, ADE20K, VIPER, StreetHazards, Fishyscapes, SegmentMelfYouCan itd.) pa ne moramo briniti o dopuštenjima, privolama, sporazumima o povjerljivosti itd.</p> <p>Jedna iznimka je programska podrška za eksperimente vezane uz semantičko prognoziranje koje zbog ugovora s tvrtkom Rimac Technologies ne možemo objaviti javno.</p>
	<p>Kako će se regulirati pristup podacima i njihova sigurnost? Koji su potencijalni rizici koje treba uzeti u obzir? Kako ćete osigurati sigurnost pohrane osjetljivih podataka?</p>	<p>Pristup i sigurnost podataka reguliramo autorizacijskim protokolima mrežnog servisa GitHub. Kvarovi lokalnih sustava za pohranjivanje podataka mogu dovesti do gubitka sirovih istraživačkih podataka iako ne i sažetaka koji se nalaze na privatnim stranicama Google Sheets. Ovaj rizik nije praktično umanjivati jer se količina sirovih podataka mjeri u gigabajtima i terabajtima.</p>
	<p>Kako ćete upravljati zaštitom autorskih prava i intelektualnog vlasništva? Tko će biti vlasnik podataka? Koje će se licencije primjenjivati na podatke? Koja će se ograničenja primjenjivati na ponovnu uporabu osobnih podataka?</p>	<p>Zaštita autorskih prava može se osigurati međunarodnim patentima. Poznato je da su članci na vrhunskim konferencijama (CVPR, ICCV i ECCV za računalni vid te NeurIPS, ICLR, ICML i AAAI za umjetnu inteligenciju) dobri kandidati za patentne prijave. Međutim, percipiramo da su patenti u računarstvu korisni prvenstveno velikim tvrtkama koje imaju resurse za patentne tužbe. Zbog toga ne planiramo upravljati zaštitom autorskih prava osim ako se oko toga angažira industrijski partner.</p>
3.	Pohrana i čuvanje podataka	
	<p>Kako će podaci biti pohranjeni i kako će biti napravljena sigurnosna kopija podataka (<i>backup</i>) tijekom istraživanja? Koji su kapaciteti čuvanja podataka kojim</p>	<p>Podrška za reprodukciju eksperimenata iz objavljenih članaka pohranjuje se u privatnim i javnim projektima mrežnog servisa GitHub. Sažetci sirovih istraživačkih podataka čuvaju se u tablicama mrežnog servisa Google Sheets. Kopije tih sadržaja nalaze se na našim lokalnim poslužiteljima. Besplatni računi za sada pružaju dovoljan kapacitet za naše postojeće potrebe.</p>

	raspoložete? Kojim se procedurama koristite za sigurnosnu kopiju (<i>backup</i>)?	
	Koji je vaš plan čuvanja podataka? U kojim će se formatima čuvati?	Besplatni mrežni servisi za sada pružaju dovoljan kapacitet za trajno čuvanje podrške za reprodukciju eksperimenata u originalnim formatima (tekst, slike). Sirove istraživačke podatke čuvamo u obliku tekstnih i binarnih datoteka koliko je to moguće, ovisno o količini podataka te dostupnom kapacitetu i kvarovima lokalnog datotečnog sustava. Besplatni mrežni servisi za sada pružaju dovoljan kapacitet za trajno čuvanje preglednih sažetaka sirovih istraživačkih podataka.
4.	Dijeljenje i ponovna uporaba podataka	
	Kako i gdje će se podaci dijeliti? Na kojem repozitoriju planirate dijeliti podatke? Kako će potencijalni korisnici doznati za podatke?	Podršku za reproduciranje eksperimenata dijelimo preko mrežnog servisa GitHub. Potencijalni korisnici mogu pronaći mrežne adrese naših projekata u našim odgovarajućim znanstvenim radovima.
	Ako postoje podaci koji se ne smiju dijeliti (prijavitelji vezani zakonskim, etičkim, autorskim pravila, povjerljivošću i sl.), pojasnite razloge ograničenja.	Većinu javno dostupnih podatkovnih skupova ne smijemo redistribuirati. Umjesto toga navodimo mrežne adrese odgovarajućih službenih mrežnih stranica.
	Potvrdite da ćete se koristiti digitalnim repozitorijem koji je u skladu s načelima <i>FAIR-a</i> .	Mrežni servis GitHub usklađen je s načelima FAIR-a jer dodjeljuje trajne identifikatore, omogućuje odabir licencija te ima podršku za detaljan metapodatkovni opis.
	Potvrdite da ćete se koristiti digitalnim repozitorijem koji održava neprofitna organizacija (ako ne, objasnite zašto ne možete dijeliti podatke na digitalnom repozitoriju koji nije komercijalan).	GitHub je standardna platforma za dijeljenje istraživačkog izvornog koda i popratnih sadržaja u našoj istraživačkoj zajednici. Prema našim saznanjima, niti jedan neprofitni repozitorij ne nudi jednaku razinu usluge verziranja.

Ref:

[1] Celjak, D., Dorotić Malič, I., Matijević, M., Poljak, Lj., Posavec K. i Turk, I.: „Istraživački podaci - što s njima?“ [Istraživački podaci - što s njima?: priručnik o upravljanju istraživačkim podacima | Digitalni repozitorij Srca \(unizg.hr\)](#)