Plan upravljanja istraživačkim podacima

|  |  |  |
| --- | --- | --- |
| Opće informacije | | |
|  | Ime i prezime predlagatelja | Dijana Oreški |
|  | Matična organizacija | Fakultet organizacije i informatike |
|  | Naziv projekta | SIMON: Inteligentni sustav za automatsku selekciju algoritama strojnog učenja u društvenim znanostima |
|  | Upravitelj podacima | Dunja Višnjić |
| 1. | Prikupljanje podataka i dokumentacija | |
|  | Koje ćete podatke prikupljati, obrađivati, stvarati ili se ponovno njima koristiti? (navedite format, vrstu i opseg podataka) | Prikupljat ćemo i obrađivati strukturirane tablične podatke iz javno dostupnih repozitorija. Podaci će biti pohranjeni u .xslx formatu.  Shodno tome, obrađivat ćemo sljedeće vrste podataka:   1. Cjeloviti skupovi podataka iz javno dostupnih repozitorija. 2. Tablica s meta-značajkama svih skupova podataka.   S obzirom da će skupovi podataka varirati u broju instanci i broju atributa, veličine datoteka će biti u rasponu od nekoliko MB do nekoliko GB. |
|  | Kako će se podaci prikupljati, obrađivati ili stvarati? (ukratko navedite metodologiju i procese osiguranja kvalitete te načine organiziranja podataka) | Podaci će nastati preuzimanjem iz javno dostupnih repozitorija. Nakon preuzimanja, podaci će se pohraniti u dva direktorija, ovisno o domeni (poslovanje ili obrazovanje). Podaci će se pročistiti i pripremiti primjenom alata SAS JMP. Za pročišćene i pripremljene podatke izračunat će se meta-značajke svih skupova podataka primjenom Python biblioteka. Temeljem meta-značajki kreirat će se meta-podaci za sve skupove podataka. |
|  | 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) | Uz svaki skup podataka kreirati će se i skup metapodataka. Metapodaci će sadržavati obvezne elemente:  Izvor: URL na javno dostupni skup podataka  Naslov: slobodni tekst  Godina objave  Vrsta izvora: skup podataka  Opis: slobodni tekst sa svim informacijama potrebnim za interpretaciju podataka  Broj instanci  Broj atributa |
| 2. | Pravna i sigurnosna pitanja | |
|  | 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)? | Svi podaci koji će se koristiti su javno dostupni. Uz svaki skup podataka biti će poveznica na izvor podataka – javno dostupni repozitorij. |
|  | 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? | Podaci će za vrijeme trajanja projekta biti pohranjeni u sustavu Gooogle drive. U okviru servisa je osigurano redovito pohranjivanje sigurnosnih kopija podataka. |
|  | 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? | Vlasnici podataka biti će vidljivi na poveznicama uz svaki skup podataka. |
| 3. | Pohrana i čuvanje podataka | |
|  | Kako će podaci biti pohranjeni i kako će biti napravljena sigurnosna kopija podataka (*backup*) tijekom istraživanja? Koji su kapaciteti čuvanja podataka kojim raspolažete? Kojim se procedurama koristite za sigurnosnu kopiju (*backup*)? | Podaci će za vrijeme istraživanja biti pohranjeni u sustavu Google Drive. Redovito se izrađuju sigurnosne kopije. Kapacitet čuvanja podataka nadilazi potrebe ovog projekta. Kopija pojedinih podataka će se čuvati na računalima istraživača koji rade na pojedinim isporukama. |
|  | Koji je vaš plan čuvanja podataka? U kojim će se formatima čuvati? | Podatke ćemo čuvati najmanje tri godine nakon završetka projekta. Podaci će se čuvati u XLSX formatu. |
| 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? |  |
|  | Ako postoje podaci koji se ne smiju dijeliti (prijavitelji vezani zakonskim, etičkim, autorskim pravila, povjerljivošću i sl.), pojasnite razloge ograničenja. | Koristit ćemo samo javno dostupne podatke s poveznicama na izvore podataka. |
|  | Potvrdite da ćete se koristiti digitalnim repozitorijem koji je u skladu s načelima *FAIR-a*. | DA. |
|  | 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). | DA. |

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)](https://repozitorij.srce.unizg.hr/islandora/object/srce:327)