



PLAN UPRAVLJANJA ISTRAŽIVAČKIM PODACIMA

Jelena Banović*,
Aleksandra Bradić-Martinović

Institut ekonomskih nauka,
Centar za podatke u društvenim naukama RS,
Beograd, Srbija

Rezime:

U 21. veku razvoj tehnologije doneo je eksponencijalni rast količine podataka. Smatra se da je 90% trenutnih podataka u svetu generisano poslednjih pet godina. Nauka se, takođe, suočava sa novim mogućnostima kada su podaci u pitanju, posebno imajući u vidu da institucije koje finansiraju istraživanja, a pre svega se misli na istraživanja finansirana iz javnih fondova (*publicly funded*), postavljaju sve rigoroznije zahteve u pogledu čuvanja i deljenja podataka. Zbog toga je potrebno da se istraživači upoznaju sa formom Plana upravljanja podacima (*Data Management Plan*), njegovim značajem i elementima, kao i mogućnostima za automatsko generisanje finalne verzije plana upotrebom onlajn platformi.

Ključne reči:

podaci, nauka, istraživanje, planiranje, prikupljanje.

1. UVOD

Uspešna naučna istraživanja zasnivaju se na kompetentnosti. Jedna od dimenzija koja doprinosi kompetentnim istraživanjima i verodostojnim rezultatima je i **kvalitet podataka** koji se koriste u analizi. U tom smislu, razvoj informacione infrastrukture i digitalnih tehnologija iz korena je promenio pristup naučnim istraživanjima, koji počiva na dostupnosti informacija i podataka. Prema Berlinskoj deklaraciji [1] „Stara tradicija i nova tehnologija su se približile i omogućile da naučna istraživanja postanu javno dobro”. Uklanjanje barijera za pristup rezultatima istraživanja i podacima omogućuje brojne koristi, kao što su brže i jeftinije istraživanje, relevantnije obrazovanje, povećane šanse za obrazovanje siromašnih, ali i povezivanje čovečanstva u potrazi za znanjem.

Upravljanje podacima nije uvek jednostavan proces za koji je odgovoran sam istraživač. Često su uključeni veliki istraživački timovi, IT podrška, institucije koje se profesionalno bave podacima, finansijeri koji postavljaju uslove, eksterni data centri ili veb-servisi koji arhiviraju i dele podatke. Zbog toga je važno pažljivo isplanirati svaku fazu tokom životnog ciklusa podataka.

Cilj rada je podizanje svesti naučnih radnika u Srbiji o značaju podataka, kroz analizu formalnog dokumenta – Plana upravljanja podacima (*Data Management Plan*), za koji se uskoro može očekivati da će postati obavezni element projektnih prijava, posebno kada su u pitanju inostrani fondovi i projektni okviri, poput HORIZON2020.

Odgovorno lice:

Jelena Banović

e-pošta:

jelena.banovic@ien.bg.ac.rs



2. PLAN UPRAVLJANJA PODACIMA

Podaci u nauci nastaju kao rezultati istraživačke delatnosti naučnih radnika, istraživačkih timova i naučnih institucija i u daljim istraživačkim procesima egzistiraju kao fundamentalna znanja [2]. Da bi se podacima upravljalo na efektivan i efikasan način, tokom svih faza istraživačkog procesa i životnog ciklusa podataka, potrebno je predvideti i formalno obraditi svaki korak.



Grafik 1. Životni ciklus podataka

Izvor: Životni ciklus podataka [3]

Plan upravljanja podacima je zvaničan dokument koji prikazuje kako se podaci obrađuju za vreme i posle istraživačkog procesa. Plan odgovara na pitanja i daje jasne smernice koje određuju:

- ◆ metod prikupljanja podataka;
- ◆ način njihovog udruživanja;
- ◆ način na koji će se vršiti njihova analiza;
- ◆ sistem čuvanja podataka i mogućnosti dalje distribucije, identifikuju se nosioci prava, politike i procedure.

Potrebno je istaći da Plan nije ograničen na upotrebu samo u akademskim krugovima, već se može koristiti i za potrebe marketinških agencija, brokerskih i izdavačkih kuća, kao i svih ostalih organizacija koje u svom radu prikupljaju podatke bilo koje vrste. Kvalitetno upravljanje podacima nije cilj sam po sebi, već put ka otkrivanju novih znanja i usmerenje za naredna izučavanja. Plan čini da podaci budu sigurni, da se čuvaju u

svim fazama naučnog istraživanja, a kasnije, kada je proces istraživanja završen, obezbeđuje valjanu razmenu i dostupnost podacima. Plan olakšava proces istraživanja jer su njime unapred predviđeni mogući faktori i problemi koji bi eventualno usporili proces istraživanja, a jedan od glavnih ciljeva je osiguravanje kvalitetne naučne prakse. Dobar Plan je koncizan, informativan i precizan, jasno definiše sve stavke i razumljiv je kako istraživaču, tako i krajnjem korisniku (arhivu, finansijeru i sl.). Svaka institucija koja izrađuje Plan upravljanja podacima čini to prema svojim potrebama uz poštovanje opšte strukture koja obuhvata:

- ◆ opis istraživačkih podataka;
- ◆ standarde i metodologije prikupljanja;
- ◆ odgovore na etička pitanja;
- ◆ plan deljenja i pristupa;
- ◆ strategiju dugoročnog očuvanja podataka.

Prema *Finskom arhivu za čuvanje podataka iz oblasti društvenih nauka (Finnish social science data archive)*, kvalitetan plan upravljanja podacima daje odgovore na sledeća pitanja [4]:

- ◆ Koja je vrsta podataka koja se prikuplja i koji je sistem za njihovo prikupljanje korišćen?

Vrsta podataka koja se prikuplja uglavnom zavisi od tipa istraživanja. Najčešće su to upitnici, intervjui, fokus grupe, razni pisani materijali, arhivski materijali, zvanični dokumenti, medijski zapisi i slično. Metode prikupljanja podataka zavise od vrste podataka koji su neophodni za istraživanje kvantitativni podaci mogu biti prikupljeni putem intervjua, upitnika, merenjima, eksperimentima, a kvalitativni podaci se prikupljaju putem grupnih razgovora, snimanjem pojedinačnih intervjua, putem sastanaka kao audio ili video datoteke [5]. Pisani materijali se najčešće prikupljaju objavljivanjem zahteva za pisanje ili pozivnicom, a zatim se sakupljaju putem elektronske ili standarde pošte. Službena dokumenta se danas mogu dobiti putem interneta, a često i na zahtev u kom se naglašava da će se koristiti isključivo u istraživačke svrhe.

- ◆ Ko polaže prava na podatke?

Nosilac autorskih prava isključivo određuje kako se podaci mogu koristiti. Pitanje se jasnije odnosi na nosioca autorskog prava, ko ima pravo da odobri pristup podacima i koje politike i procedure se koriste za informisanje svih učesnika u istraživanju. Preporuka da istraživački timovi uvek pripreme sporazum o autorskom pravu nad podacima, što obuhvata i pravo njihove buduće upotrebe, kao i arhiviranja i deljenja.



- ◆ Da li je obezbeđena poverljivost i sigurnost podataka?

Poverljivost i bezbednost podataka je jedan od osnovnih postulata u istraživačkom okruženju. Sigurnost se garantuje tako što se lični podaci obrađuju samo do stepena potrebnog za istraživanje, a neovlašćeni pristup podacima se sprečava. Svi osetljivi podaci prolaze određene procedure kojima se smanjuje mogućnost od zloupotrebe i garantuje se bezbednost podacima. U kontekstu čuvanja podataka posebno se mora voditi računa o raspoloživosti podataka u digitalnom obliku, koji se jednostavno i brzo mogu distribuirati i umnožavati. Da bi se izbegle situacije navedene u prethodnom primeru potrebno je pažljivo anonimizirati podatke pre postupka diseminacije. Osnovni cilj anonimizacije je očuvanje privatnosti fizičkog ili pravnog lica [6]. Kroz ovo pitanje se definiše i pitanje skladištenja podataka, koje je takođe bitno zbog bezbednosti.

- ◆ Koji formati i softveri se koriste za skladištenje podataka?

Softveri koji danas postoje su prilagođeni za analizu i kvalitativnih i kvantitativnih podataka. U zavisnosti od njihovog tipa, određuje se i najpogodnija aplikacija. Preporuka je da se jedna kopija podataka uvek čuva u nezavisnom softveru ili u standardnom formatu, a to je najčešće XML format koji odoleva rapidnim promenama softverskih sistema.

- ◆ Da li je dostupna dokumentacija o obradi sadržaja?

Dokumentacija kreirana prilikom prikupljanja podataka, u različitim fazama istraživačkog procesa, predstavlja jedan od ključnih faktora za dugotrajnu zaštitu i kasniju ponovnu upotrebu tih podataka. Dokumentacija omogućava informisanje istraživača o projektu, njegovom toku i preprekama i pruža uvid u potencijalno ponovno korišćenje podataka, čime je životni vek podataka znatno produžen.

- ◆ Kakav je životni ciklus podataka?

Odgovor na ovo pitanje daje opis vezan za to šta dešava sa podacima nakon završetka projekta. Jasno je da ponovno korišćenje podataka u velikoj meri zavisi od mera za upravljanje podacima koje su korišćenje u procesu istraživanja, a to u najvećoj meri obezbeđuje DMP. Obezbeđuju se rešenja za širenje podataka i metapodataka, uslovi korišćenja itd.

3. STRUKTURA PLANA UPRAVLJANJA PODACIMA

Standardizovana struktura plana upravljanja podacima ne postoji, osim u slučajevima kada je usvojen kao obavezan deo projektne prijave i jednog istraživanja. Međutim, Interuniverzitetski konzorcijum za politička i socijalna istraživanja (*Inter-university Consortium for Political and Social Research – ICPSR*) nudi formu koja je prepoznata od naučne zajednice kao jedna kvalitetnija jer pokriva više segmenata istraživanja. U ovom slučaju, Plan upravljanja podacima sadrži sledeća polja [7]:

- ◆ **Opis podataka:** u ovom polju istraživači objašnjavaju prirodu i obim podataka koji će biti generisani;
- ◆ **Postojeći podaci:** polje sadrži strukturu već postojećih podataka, koliko su relevantni za projekat i na koji način će biti integrisani;
- ◆ **Format podataka:** opisuje format u kom će podaci biti generisani, čuvani i diseminirani;
- ◆ **Metapodaci:** polje sadrži opis metapodataka koji opisuju buduće podatke, kao i standard metapodataka;
- ◆ **Skladištenje i bekap:** opisuje se način skladištenja podataka, procedure za skladištenje, prikaz fizičkih i onlajn resursa;
- ◆ **Sigurnost:** polje obuhvata procedure i pravila za zaštitu podataka, definiše dozvole za njihovo korišćenje, ograničenja u korišćenju – embargo i slično;
- ◆ **Odgovornosti:** imena pojedinaca odgovornih za upravljanje podacima u istraživačkom projektu nalaze se u ovom polju;
- ◆ **Prava intelektualne svojine:** u ovom polju je neophodno jasno definisati osobu ili grupu istraživača koji će polagati pravo na podatke, način na koji će biti zaštićeni – definisanje pojma autorskog prava;
- ◆ **Pristup i deljenje:** definiše deljenje podataka, procedure pristupa, mehanizme za njihovu diseminaciju; da li će pristup biti u potpunosti otvoren, na zahtev korisnika ili za pristup samo određenih grupa i sl.;
- ◆ **Korisnici:** određuje se ko će biti korisnici podataka;
- ◆ **Selekcija i rokovi čuvanja:** sadrži opis podataka koje je potrebno arhivirati, uz opis postupka čuvanja ili eventualnog prestanka čuvanja podataka u budućnosti;



- ♦ **Etika i privatnost:** sadrži pristanak na privatnost, rukovanje podacima, zaštita učesnika u istraživanju itd;
- ♦ **Budžet:** sadrži troškove pripreme podataka, dokumentaciju za arhiviranje, prikaz načina na koji će troškovi biti izmireni; eventualni zahtevi finansijera;
- ♦ **Organizacija podataka:** podrazumeva kako će se podacima upravljati, kao će se vršiti kontrola, anonimizacija i sl.;
- ♦ **Garancija kvaliteta:** sadrži procedure za osiguranje kvaliteta podataka tokom trajanja projekta;
- ♦ **Zakonski uslovi:** podrazumevaju listu svih zakonskih uslova koji se moraju ispuniti u radu, kao i zakonski zahtevi finansijera;

Sva navedena polja su označena kao polja od velike važnosti i obavezna su.

4. PLATFORME ZA GENERISANJE PLANA UPRAVLJANJA PODACIMA

Dok se ne formira standardizovan obrazac plana upravljanja podacima, koji će se koristiti u naučnoj zajednici u Srbiji, na raspolaganju su brojni onlajn servisi koji omogućavaju automatsko kreiranje ovog dokumenta. Neki od njih su jako pouzdani i prilično detaljni, čime je omogućeno da se pokriju sve teme koje su bitne u istraživačkom ciklusu. Uzećemo kao primer platformu *DMPTool* na kojoj se može odabrati prototip plana koji je već predviđen za neke od vodećih svetskih institucija koje finansiraju naučna istraživanja (<https://dmptool.org/>). Praćenjem koraka ispunjavaju se sva neophodna polja i kao rezultat automatski se kreira plan koji se kasnije može dopunjavati. Nakon uspešno otvorenog naloga, najpre se nudi mogućnost unosa osnovnih informacija o samom projektu, poput naziva projekta, finansijera, kratkog opisa, informacije o rukovodiocu projekta. U delu *Plan Overview* daje se kratak pregled osnovnih polja koja će se u planu naći kako se ne bi izostavio nijedan bitan segmet. Izrada plana započinje u delu *Write plan* – izdvojene su sve teme koje plan sadrži. Zapoinju definisanjem uloga odgovornih lica, rukovodioca projekta, ostalih članova projektnog tima i čitave institucije.

Grafik 2. Kreiranje plana upravljanja podacima

Izvor: početna strana platforme *DMPTool* [8]

Zatim se definiše tip podataka koji se prikuplja i neophodno je detaljno ih predstaviti. Takođe, prilažu se informacije o dokumentaciji koja će nastati u toku rada, predstavljaju se procedure rada, metode za kasnije deljenje podataka – da li će budući korisnici potpisivati određene ugovore ili saglasnosti za upotrebu podataka ili će oni biti zatvorenog tipa zbog svoje senzitivnosti. Prednost platforme *DMPTool* je ta što kroz kratka objašnjenja i primere pruža korisnicima pomoć u kreiranju svake od navedenih pojedinosti. Kao rezultat dobija se detaljan plan koji je obuhvatio čitav proces rada – od početka istraživanja i prikupljanja podataka do načina i uslova za ponovnu upotrebu u nekim narednim istraživanjima.

5. PRIMENA U PRAKSI

Sve je više zahteva od strane finansijera naučno-istraživačkih projekata za izradom plana upravljanja, bilo prilikom prijave projekta, bilo nakon njegovog prihvatanja, a u kasnijim fazama projektnih aktivnosti. **Postoje tendencije da će se ovaj dokument uvrstiti u obaveznu dokumentaciju prilikom svake projektne prijave.** U cilju promovisanja „Otvorenog pristupa” mnogi finansijeri zahtevaju da istraživački podaci nastali kao rezultat projekata budu dostupni. Međutim, bitno je pomenuti da izrada ovog plana nije samo formalnost, potrebna da bi se udovoljilo zahtevima finansijera, već i vrlo korisna referenca koja pomaže da se predvide potencijalne poteškoće i unapred osmisle načini za njihovo uspešno otklanjanje. Amerika se može smatrati za lidera u ovoj oblasti, tako da plan upravljanja podacima nije



novina za njihovu istraživačku zajednicu, već postoji evidencija da su planove upravljanja podacima izrađivali pre više od deset godina. *Nacionalni institut za zdravlje* je u februaru 2003. godine usvojio **politiku deljenja podataka** za projekte koji se finansiraju sa više od 500.000 dolara. Za ovu vrstu projekata, u projektnoj prijavi mora biti predat i plan upravljanja podacima. Osnovni motiv izrade plana je stimulisanje istraživača da dele podatke u formatu koji bi sadržao detaljan opis kroz formiranje metapodataka [9]. Koristi od kreiranja ovog plana prepoznala je i NASA. Propisana je uredba kojim se definiše da se uz sve projekte koje će finansirati NASA dostavlja i plan upravljanja podacima, koji će jasno definisati način generisanja i deljenja podataka, ali pružiti i obrazloženje zašto određeni skupovi podataka neće biti dostupni javnosti [9]. *Kancelarija za nauku i tehnologiju* pri Beloj kući je 2013. godine sačinila memorandum koji zahteva generisanje plana u kome će biti opisani podaci dobijeni putem istraživanja koja finansira država; pri čemu je važno napomenuti da će podaci biti besplatno dostupni u roku od 12 meseci od njihove prve objave [9]. Osim toga, nekoliko ključnih zdravstvenih fondova u Velikoj Britaniji zahtevaju od istraživača da podnesu plan upravljanja podacima kao deo projektne prijave [9].

6. HORIZON 2020 I PLAN UPRAVLJANJA PODACIMA

HORIZON2020, kao jedan od vodećih evropskih finansijera programa za istraživačke delatnosti, prepoznao je značaj postojanja plana upravljanja podacima. Evropska komisija se zalaže za ideju da svi podaci budu javno dostupni, sa motivom da poboljša i unapredi pristup naučnim informacijama, a konačno smatra da pristup informacijama koje su već finansiranje iz budžeta HORIZON2020 ne treba duplo naplaćivati – prilikom prikupljanja i prilikom sekundarne upotrebe.

Osnovni zahtevi koje propisuje HORIZON2020:

- ◆ deponovanje istraživačkih podataka u repozitorijume (skladišta) i
- ◆ obezbeđivanje pristupa. [10]

Evropska komisija je podržala ideju i kroz lansiranje pilot projekta *Open Research Data Pilot* (ORD Pilot) koji ima za cilj da podatke koji su prikupljeni u okviru projekata finansiranih kroz okvir HORIZON2020 učini dostupnim, čime omogućuje njihovu ponovnu upotrebu, uz obaveznu zaštitu osetljivih podataka od nedobronamerne upotrebe. Za projekte koji su započeli pre 2017. godine izdvojeno je sedam oblasti u okviru

HORIZON2020-a koje su bile deo pilot projekta. Svi projekti koji su započeli nakon 1. januara 2017. godine u obavezi su da deponuju sve prikupljene podatke. Ako je jedan od osnovnih postulata deljenja podataka zaštita privatnosti, onda je jasno zašto ovaj pilot projekat služi za balansiranje između otvorenosti (dostupnosti) i zaštite. U programima koji su obuhvatali period od 2014. do 2016. godine, ORD Pilot je pokrivaio samo određena područja, a novi propisi obuhvataju sve tematske oblasti. ORD pilot omogućuje da podaci postanu lako dostupni, međusobno usaglašeni i da mogu ponovo da se dele u okviru naučne zajednice (findable, accessible, interoperable and reusable data – FAIR data). [10].

Sinteza 2018

Grafik 3. Forma za popunjavanje meta polja

Izvor: *Kreiranje Plana upravljanja podacima* [8]

Open Research Data Pilot zahteva:

- ◆ izradu plana upravljanja podacima tokom prvih šest meseci projekta, kao i njegovo naknadno ažuriranje;
- ◆ čuvanje u digitalnim repozitorijumima;



- ◆ adekvatan pristup podacima i
- ◆ pružanje informacija o alatima i instrumentima neophodnim za validaciju rezultata. [10]

Prema dopunskim smernicama od 26. jula 2016. godine, HORIZON2020, kao finansijer, propisuje da plan upravljanja podacima sadrži informacije o:

- ◆ načinu upravljanja informacijama tokom i nakon završetka istraživačkog projekta;
- ◆ informacije o tome koji podaci će biti prikupljeni, obrađeni i generisani;
- ◆ koji podaci će biti u slobodnom pristupu (*Open Access*) i
- ◆ kako će se prikupljeni podaci čuvati. [10]

Srbija je posredno pod uticajem ovih odluka i zahteva jer je priključena projektu 1. jula 2014. godine.

7. ZAKLJUČAK

Kao zaključak može se nametnuti pitanje da li je plan upravljanja podacima opcija ili potreba. Ako nauka počiva na hipotezi da je njena osnovna svrha deljenje i objavljivanje podataka dobijenih radom i zalaganje u toku nekog naučnog istraživanja, onda je jasno zašto u razvijenim zemljama Evrope i sveta kažu da je plan upravljanja podacima neophodnost. Zbog toga je najjednostavnije reći da je plan upravljanja podacima potreba tamo gde su naučni radnici spoznali potrebu da većina podataka bude slobodna, otvorena i javno dostupna. Svako naredno istraživanje time će biti unapređeno a nauka iz svih oblasti istraživanja podignuta na viši nivo. Što se tiče Srbije, može se reći da se istraživači još uvek upoznaju sa pojmom deljenja podataka, javno dostupnim podacima i planom za upravljanje podacima. Plan upravljanja podacima još uvek nije zaživeo u potpunosti, ali je evidentno da malim koracima ulazi u okvire naučno istraživačkog rada i postaje neophodnost. Time će se svi nedostaci i greške u radu svesti na minimum, a kvalitet rada i rezultata biće poboljšani, uz uštedu novca koji se ulaže u istraživanja.

ZAHVALNOST

Ovaj rad je deo istraživačkih projekata pod šiframa 47009 (Evropske integracije i društveno-ekonomske promene privrede Srbije na putu ka EU) i 179015 (Izazovi i perspektive strukturnih promena u Srbiji: Strateški pravci ekonomskog razvoja i usklađivanje sa zahtevima EU), finansiranih od strane Ministarstva prosvete, nauke i tehnološkog razvoja Republike Srbije.

LITERATURA

- [1] Berlinska deklaracija (2002). <https://openaccess.mpg.de/Berlin-Declaration>. Pristupljeno: 22.02.2018.
- [2] S. Filipi-Matutinović, Naučne informacije u Srbiji: Protok, dostupnost, vrednovanje. Beograd: Srpsko biološko društvo, Biološki fakultet, Univerzitetska biblioteka „Svetozar Marković, Narodna biblioteka Srbije, 2014.
- [3] L. Corti, V. Van den Eynden, L. Bishop, and M. Woollard, Managing and sharing data. Essex: UK Data Archive, 2011.
- [4] [FSSDA] Finnish Social Science Data Archive <http://www.fsd.uta.fi/aineistonhallinta/en/data-management-planning.html> pristupljeno: 20. 3. 2018.
- [5] W. Neuman, Social Research Methods: Qualitative and Quantitative Approaches. Essex: Pearson Education Limited, 2014.
- [6] A. Bradić-Martinović, A. Zdravković. „Zaštita privatnosti – anonimizacija podataka“. Preduzetništvo kao šansa: zbornik radova / V naučni skup USPON. Poslovni fakultet, Valjevo. str. 206-213, 2013.
- [7] [ICPSR] Inter consortium for political and social research <https://www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/> pristupljeno: 15.3.2018.
- [8] [DMPTool] <https://dmptool.org/plans/new> pristupljeno: 12. 3. 2018.
- [9] [DCC] – Data Curation Center <http://www.dcc.ac.uk/resources/data-management-plans>. Pristupljeno: 15.3.2018.
- [10] H2020 programme. Guidelines on FAIR Data Management in Horizon2020 http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf pristupljeno: 14.3.2018.