

## Izazovi kreiranja i ponovne upotrebe ekoloških podataka

Aleksandra Bradić-Martinović<sup>1</sup>,  
Larisa Malić<sup>2</sup>, Jelena Banović<sup>1</sup>

Pregledni rad  
UDC:

### UVOD

Ekologija je naučna disciplina čiji se rezultati u velikoj meri zasnivaju na zajedničkim naporima istraživača. Ova oblast je tokom proteklog veka doživela nagli napredak, koji se ogleda i u tome što se od kratkoročnih projekata i posmatranja koji su izvodili pojedinci ili manji istraživački timovi, prešlo na dugoročne multidisciplinarnе projekte koji integrišu različite skupove podataka uz upotrebu sofisticiranih analitičkih pristupa. Ekološka istraživanja često zahtevaju i interakciju sa drugim disciplinama, kao što su geologija, okeanografija, klimatologija i slične, ali i različitim poljima, na primer epidemiologija ili ekonomija [7]. Na osnovu izloženog, nije teško zaključiti da su naučni podaci skupi za proizvodnju, pri čemu mogu biti imati i ogromnu vrednost u budućnosti, ukoliko se uzmu u obzir specifična vremena i mesta, koja je nakon izvesnog vremena nemoguće ponoviti. U tom smislu, oni su dragoceni za naučnu zajednicu, ali i nenaučnu zajednicu u koju spadaju kreatori javnih politika. Istraživanja o naučnim praksama vezanim za podatke koncentrisana su na „velike nauke“, kao što je fizika ili na saradnju u oblastima poput biodiverziteta, međutim jednako je važno i razumevanje naučne prakse u pogledu upravljanja podacima proučavanja naučnih oblasti u kojima mali timovi kreiraju opservacije dugoročnog, multidisciplinarnog i međunarodnog značaja [2].

U savremenim uslovima, istraživanja u oblasti ekologije kreiraju veliku količinu različitih podataka. Taj obim se eksponencijalno povećava upotrebom linearnih akceleratora, senzornih mreža, satelita, seizmografa i sličnih uređaja, tako da se postavlja pitanje mogućnosti njihovog iskorišćenja. Upotreba informacione tehnologije ima potencijal da omogući upravljanje ovim podacima kroz dve dimenzije: da obezbedi njihovu analizu i čuvanje za potrebe istraživačkih timova, ali i njihovo deljenje sa širom istraživačkom zajednicom (*Designated Community*).

Adrese autora: <sup>1</sup>Institut ekonomskih nauka, Centar podataka u društvenim naukama Zmaj Jovina 12, Beograd, <sup>2</sup>Institut za književnost i umetnost, Beograd

Rad primljen: 28. 04. 2017.

Rad prihvaćen: 14. 07. 2017.

Druga dimenzija posebno dobija na značaju u poslednje vreme, jer ukoliko se ovi podaci mogu čuvati u formi u kojoj ih je moguće ponovo upotrebiti, onda se mogu i deliti preko distributivnih mreža. Na taj način, podaci postaju važan krajnji proizvod istraživanja, koji ujedno dopunjuju tradicionalnu ulogu naučnih publikacija.

Cilj ovog rada je da ukaže na značaj i vrednost podataka koji se dobijaju u procesu ekološkog istraživanja, kao i na mogućnosti njihovog deljenja i upotrebe od strane šire naučne zajednice. U prvom delu rada predstavljene su karakteristike ekoloških podataka i životni ciklus podataka, sa osvrtom na aktuelnu situaciju. U drugom delu rada prikazali smo prednosti koje se ostvaruju kada se podaci učine dostupnim, ali i razloge zbog kojih se istraživači nerado upuštaju u ovaj proces.

### 1. KARAKTERISTIKE EKOLOŠKIH PODATAKA

Istraživački podaci su činjenični delovi informacija koji se koriste za proizvodnju i proveru ispravnosti rezultata istraživanja. Generalno, podaci se mogu svrstati u pet kategorija [1].

- *Opservacioni*: podaci koji su vezani za tačno određeno vreme i mesto i koji nisu zamenljivi (terenske opservacije, očitavanje meteoroloških stanica, satelitski podaci);
- *Eksperimentalni*: podaci dobijeni u kontrolisanom ili delimično kontrolisanom okruženju, koji se mogu reprodukovati uprkos visokim troškovima (eksperimenti u staklenim baštama, hemijske analize);
- *Simulacija*: podaci dobijeni iz modela (klimatsko modeliranje);
- *Izvedeni*: podaci koji nisu prikupljeni direktno, već su generisani iz drugih podataka (populacija biomase koja se izračunava na osnovu gustine stanovništva i podataka o prosečnoj veličini tela) i
- *Meta podaci*: podaci o podacima (opis svakog podatka).

Ključni izazov koji se danas postavlja pred istraživače je potreba da rade sa različitim izvorima podataka, pri čemu je postalo vrlo uobičajeno da se za potrebe projekata integriše bilo koja kombinacija podataka u okviru jedinstvene metodologije i analize. To se u još većoj meri odnosi i na multidisciplinarnu

linarne projekte. Kako vremenom istraživanja postaju sve više kolaborativna i interdisciplinarna, ovo pitanje će sve više dobijati na značaju. Osim navedenog, uvidom u ekološku informatiku [7] suočavamo se i sa tri glavna tehnološka izazova, kada su podaci u pitanju: disperzija, heterogenost i njihovo poreklo.

**Disperzija:** Ekosistemi i staništa variraju širom sveta, a podaci se prikupljaju na hiljadama lokacija. Uprkos tome što je velika količina podataka predstavljena relativno malim brojem skupova podataka i što njima obično upravljaju veliki istraživački projekti, instituti i agencije, većina ekoloških podataka je teško dostupna, jer se nalazi rasuta među nekoliko desetina hiljada nezavisnih istraživača.

**Heterogenost:** raznolikost podataka stvara izazove zbog širine tema koje ekolozi izučavaju, kao i zbog raznovrsnih eksperimentalnih protokola koji se koriste od strane nezavisnih istraživača.

**Poreklo i istorija:** su neophodni kada se u ekološkim istraživanjima, pojave zanimljivi rezultati pojave kao posledica složenih, mukotrpnih procesa agregacije, modeliranja i analize.



Grafik 1 - Životni ciklus podataka,  
Izvor: Corty, et al. (2011), str. 15

Na grafiku 1. predstavljen je životni ciklus podataka. Kao što se vidi na grafiku, podaci često mogu imati mnogo duži vek trajanja u odnosu na dužinu projekta u okviru koga su prikupljeni. Veliki broj projekata se fokusira samo na određene segmente životnog ciklusa podataka, kao što je primarno prikupljanje ili generisanje podataka, zatim njihova analiza i publikovanje rezultata (diseminacija), a mnogo manji broj projekata obezbeđuje mogućnost ponovne upotrebe tih podataka od strane drugih naučnika. Navedena praksa skraćuje

životni vek podataka, čime se generišu veliki oportunitetni troškovi.

Koncept životnog ciklusa podataka dobija posebnu dimenziju u kontekstu pokreta Otvorene Nauke (*Open Science*). Otvorena Nauka predstavlja novi pristup naučnom procesu zasnovan na kooperativnom radu i novim načinima difuzije znanja pomoću digitalne tehnologije i novih kolaborativnih alata. Ideja obuhvata sistemsku promenu načina na koje nauka sprovodi istraživanja, odnosno pomeranje od standardne prakse objavljivanja konačnih rezultata istraživanja u naučnim publikacijama ka deljenju i upotrebi svih dostupnih znanja u ranijim fazama procesa istraživanja. Otvorena Nauka je u nauci isto što je koncept Veb 2.0 bio za društvene i ekonomskih transakcije: omogućio je krajnjim korisnicima da budu generatori ideja, odnosno i usluga i na taj način je omogućio stvaranje novih radnih modela, novih socijalnih odnosa i doveo do novog *modus operandi* nauke [5].

## 2. DELJENJE PODATAKA U EKOLOGIJI

Deljenje podataka je trend koji se u poslednjih dvadesetak godina širi u svim oblastima nauke i istraživanja. Feinberg (1994) je istakao koji je značaj deljenja i arhiviranja podataka:

- Promoviše nova istraživanja i omogućuje testiranje novih ili alternativnih metoda. Brojni su primeri podataka koji se koriste na način na koji prvobitni istraživači nisu predvideli.
  - Poboljšavaju se metode prikupljanja podataka i njihovo merenje kroz razmatranje većeg broja istraživača. Javno dostupni podaci pružaju mogućnost postizanja konsenzusa o ovim metodama.
  - Redukuju se troškovi time što se izbegava dupliranje prilikom procesa prikupljanja. Pojedini standardni skupovi podataka, kao što su *General Social Survey* ili *National Election Studies*, rezultirale su hiljadama naučnih radova za koje autori nisu morali da odvoje vreme i novac za prikupljanje podataka. Ova prednost se suštinski svodi na to da je nepotrebno dva puta prikupljati iste podatke.
  - Obezbeđuje jako dobar materijal u nastavnom procesu. Sekundarni podaci su izuzetno vredni za nastavnike i studente, koji imaju pristup kvalitetnim podacima, kao model za svoj rad.
- Uprkos svim navedenim prednostima koje deljenje podataka donosi, veliki broj naučnika nije spreman da deli svoje podatke na ovaj način. Navode se različiti razlozi, od kojih su najčešći [4]:
- Želja da se potencijal podataka u potpunosti iskoristi u istraživanju i objavljivanju publikacija (radova u časopisima i na konferencijama). Ovaj problem se prevazilazi tako što je većina javno finansiranih projekata tako koncipirana

da omogućuje istraživačima koji su prikupili podatke da ih izvesno vreme samostalno koriste, a da ih tek nakon određenog vremena dele. Osim toga, istraživali bi uvek trebalo da imaju na umu da svoje podatke poznaju bolje od onih koji se prvi put sreću sa njima, što im pruža veliku prednost. Ukoliko i pored navedenog imaju bojazan, imaju mogućnost da u određenom roku postave vremenski embargo na svoje podatke.

- Stav da podaci nisu zanimljivi za druge istraživače, koji je najčešće pogrešan, jer se uporede studije mogu realizovati samo ukoliko imamo podatke iz prethodnih istraživanja.
- Nemaju vremena, ni sredstava da pripreme podatke za arhiviranje i deljenje. Važno je planirati upravljanje još u toku životnog ciklusa istraživačkih podataka. Upravljanje podacima idealno bi trebalo da bude sastavni deo istraživačke prakse, čime bi bilo redukovano vreme i troškovi, ali i poboljšan kvalitet podataka koje i sami istraživači koriste.

Pored toga navode se i razlozi objektivne prirode [1]:

- Skupovi podataka sadrže poverljive informacije o ugroženim vrstama;
- Skupovi podataka sadrže lične podatke čijim bi se deljenjem ugrozio zakon o poverljivosti ličnih podataka (koju većina zemalja poseduje) i

- Ukoliko treće lice ima vlasništvo nad tim podacima, npr. finansijer istraživanja.

Navedeni problemi rešavaju se pažljivim sprovođenjem procedura anonimizacije i pregovorima sa formalnim vlasnikom podataka u kojima se utvrđuju uslovi deljenja.

Kao posledica ovih razloga, uz tri glavna tehnološka izazova na žalost, samo je mali deo ekoloških podataka javno dostupan i vidljiv, a i oni koji jesu dostupni često nisu praktično upotrebljivi u svrhu novog istraživanja. Na osnovu iskustva [7] arhiviranje podataka u ekologiji procenjuje se na 1% u odnosu na ukupno prikupljene podatke. Umesto da obezbede direktan pristup podacima, istraživači dele tumačenja obrađenih podataka kroz objavljivanje prezentacija i publikacija. Da bi se situacija promenila potrebno je raditi na promeni kulture deljenja i na rešavanju tehnoloških i socioloških izazova. Otvoreni Podaci (*Open Data*) će poboljšati i ubrzati naučna dostignuća pružajući veću transparentnost i reproduktivnost rezultata.

Jedan od vrlo dobrih primera pozitivne prakse u ovoj oblasti je *Department of Ecology, State of Washington* koji deli podatke dobijene u procesu monitoringa životne sredine, kroz portal DRYAD ([www. http://datadryad.org/](http://datadryad.org/)).

Slika 1 - Portal Dryad, Izvor: [www. http://datadryad.org/](http://datadryad.org/)

## ZAKLJUČAK

Uprkos tome što je otpor u procesu deljenja podataka izražen, baze ekoloških podataka nastavljaju da rastu u obimu, širini i kompleksnosti. Razvija se i viši nivo opisa podataka, odnosno meta podaci i informacije izvedene iz naknadne obrade i analize podataka, takozvane "meta-informacije", kao veoma važni prateći elementi arhiviranih po-

dataka, jer pomažu boljem razumevanju i lakšem pretraživanju istih.

Istraživači bi trebali da budu svesni da i najmanji skupovi podataka mogu da doprinesu razvijanju ključnih znanja za rešavanje velikih problema, a ti podaci često već postoje, jer su proizvedeni od strane drugih timova, koji ne žele da ih dele. Očekivanja su da je ova oblast u začetku razvoja, kada

je u pitanju masovno arhiviranje i deljenje podataka u ekologiji, ali je vrlo važno ohrabrivati kulturu saradnje među istraživačima, koji bi u budućnosti bili spremniji da učine svoje podatke transparentnim i korisnim široj naučnoj zajednici i ostalim stejkholderima, poput kreatora javnih politika.

## LITERATURA

- [1] BES. (2014). A Guide to Data Management in Ecology and Evolution. London: British Ecological Society.
- [2] Borgman, C.L., Wallis, J.C., Enyedy, N. (2006). Little Science Confronts the Data Deluge: Habitat Ecology, Embedded Sensor Networks, and Digital Libraries. International Journal on Digital Libraries, Volume 7, Issue 1, str. 17–30.
- [3] Corti L., Van den Eynden V., Bishop L., Morgan-Brett B. (2011). Managing and sharing data. Colchester, Essex: UK Data Archive.
- [4] Corti, L., Van den Eynden, V., Bishop, L., Woollard, M. (2014) Managing and Sharing Research Data: a Guide to Good Practice. Colchester, Essex: UK Data Archive and SAGE Publish.
- [5] EC. (2016). Open Innovation, Open Science, Open to the World - A Vision for Europe. European Commission: Directorate-General for Research and Innovation. Brussels.
- [6] Feinberg, S.E. (1994). Sharing Statistical Data in the Biomedical and Health Sciences - Ethical, Institutional, Legal and Professional Dimensions, Annual Review of Public Health, 15, Palo Alto, CA, Annual Reviews, Inc. str. 1-18.
- [7] Reichman, O. J., Jones, M.B., Schildhauer, M.P. (2011). Challenges and Opportunities of Open Data in Ecology. Science, Volume 331. str. 703-705.

## IZVOD

### IZAZOVI KREIRANJA I PONOVI NE UPOTREBE EKOLOŠKIH PODATAKA

*Svedoci smo promena koje tehnološki napredak izaziva u globalnom okruženju, a posebnu dimenziju čini revolucija u oblasti proizvodnje i upravljanja podacima. U ovom radu postavljen je fokus na izazove upravljanja podacima koji predstavljaju proizvod istraživanja u oblasti ekologije i potrebu prilagođavanja povećanom obimu, raznovrsnosti i sve većoj brzini protoka podataka i informacija. Za razliku od takozvanih „velikih nauka“ poput fizike i astronomije, koje u okviru gigantskih projekata generišu velike količine podataka (Big Data), „male nauke“ poput ekologije u najvećem broju slučajeva realizuju svoja istraživanja i generišu podatke u okviru većeg broja manjih projekata. Uprkos tome, ovi podaci zahtevaju značajna sredstva za prikupljanje i veoma često imaju veliku vrednost za istraživačku zajednicu. Osnovni izazov sa kojim se istraživači u ekološkim istraživanjima suočavaju su nedostatak odgovarajućih alatki i infrastrukture za upravljanje sve većom količinom podataka. Drugi izazov je odgovor na sve intenzivniji trend deljenja podataka među istraživačima u istoj oblasti istraživanja. Potreba za deljenjem zasniva se na koristi njihove ponovne upotrebe, a sam proces arhiviranja i čuvanja zahteva dodatnu infrastrukturu. Čitav proces je povezan sa nizom izazova, kao što su nepostojanje kulture i nedostatak podsticaja za deljenje, nedostatak znanja i sl. U radu je prikazan značaj i stanje u ovoj oblasti, uz predstavljanje ograničenja i izazova.*

**Ključne reči:** podaci, ekologija, digitalni arhivi, deljenje, Otvorena nauka, Otvoreni podaci.

## ABSTRACT

### THE CHALLENGES OF CREATION AND REUSE OF ECOLOGICAL DATA

*We are witnessing changes that technological progress is caused in the global environment, with special dimension of the revolution in production and data management. In this paper, the focus is placed on challenges in management of data that are the product of research in the field of ecology and the need to adapt to the increased volume, diversity and increasing flow of data and information. Unlike the so-called "big science" such as physics and astronomy, which, in the context of gigantic projects generate large amounts of data (Big Data), "small science" such as ecology, in most cases realize their research and generate data in the context of a larger number of smaller projects. Nevertheless, these data require significant resources to collect and often have great value for the research community. The main challenge that researchers face in ecological research are the lack of appropriate tools and infrastructure to manage the increasing amount of data. Another challenge is the respond to intensifying trend of data sharing among researchers in the same field of research. The need for sharing is based on the benefits of reuse, but the process of archiving and storage requires additional infrastructure. The whole process is associated with a number of challenges, such as lack of culture and the lack of incentives for sharing, lack of knowledge etc. The paper presents the importance of this field and the current situation, together with the representation of restrictions and challenges.*

**Keywords:** data, ecology, digital archives, sharing, Open Science, Open Data.