Primena koncepta Big data u ekologiji

Big Data Concept in Ecology

Jelena Banović[1], *Aleksandra Bradić-Martinović*[1]
[1] Institut ekonomskih nauka, Zmaj Jovina 12, Beograd

APSTRAKT
Promene koje donosi razvoj tehnologije doprinele su da naučna istraživanja postanu obimnija i sveobuhvatnija. Podaci dobijeni u okviru istraživanja zahtevaju posebnu analizu, obradu i skladištenje i imaju veliku vrednost za istraživačku zajednicu, kao i društvo u celini. Kako se ekologija kao nauka menja i prihvata nove standarde koje 21. vek postavlja, tako jača i svest o značaju ove vrste naučnih podataka. Globalna ekološka pitanja, kao što su klimatske promene i zaštita ekosistema zahtevaju proučavanja koja kao rezultat donose vrlo vredne i u pojedinim slučajevima obimne podatke, čije prikupljanje često zahteva obimna ulaganja. Iako se ekologija tretira kao „mala nauka“, brojna istraživanja generišu jako velike skupove podataka –*Big Data*, čiju obradu, analizu i skladištenje ne mogu da zadovolje tradicionalni servisi. Ove podatke karakteriše veliki obim, raznovrsnost i mogućnost efikasne obrade. U skladu sa tim, predmet rada je specifična vrsta podataka koja nastaje tokom brojnih istraživanja u ekologiji i njihova važnost za istraživače u ovoj oblasti, a i šire, kao i način njihovog arhiviranja, udruživanja i ponovne upotrebe. Cilj rada je da se ukaže na izuzetan značaj i mogućnost upotrebe *Big Data* u ekologiji, iz perspektive istraživača, finansijera istraživačkih projekata, ali i donosioca javnih politika u oblasti zaštite životne sredine.

Ključne reči: podaci, *Big Data,* arhiviranje, ponovna upotreba podataka

ABSTRACT
The changes brought by the rapid development of technology made scientific research more extensive and comprehensive. Data created in a research require specific analysis, processing and archiving, and have huge value for research community and society. With ecology changing and accepting standards of the 21st century, conscience about the importance of this data is growing. Global ecological agenda, such as dealing with climate changes and vulnerable ecosystem protection demands research which as a result has very valuable, but usually extensive data. Even though ecology is being treated as a „small science“, many research, as a result, has massive datasets – also known as *Big Data*, whose analysis, processing and archiving cannot be satisfied by traditional services. Main characteristics of this data are volumen, variety, velocity, and veracity. This paper presents specific kind of data that can be created during numerous research in ecology, their importance for researchers in the field of ecology and wider, the way of archiving and reusing. The aim of the paper is to present the significance of this data and possibility of using in ecology from the perspective of researchers, funders of research projects and public policy makers.

Key words: data, Big Data, archiving of data, reusing of data

1. UVOD

Početak 21. veka može se okarakterisati kao Doba podataka. Glavni razlog za to je nagli razvoj i masovna upotreba računara, Interneta i tehnologije koja je u stanju da prikuplja informacije iz realnog, fizičkog sveta u kome živimo i da ga konvertuje u digitalni format, odnosno digitalni podatak ili informaciju. Podaci se generišu u svim situacijama koje podrazumevaju *online* aktivnosti, kao na primer upotreba pametnog telefona koji ima ugrađenu aplikaciju za GPS, komunikacija putem društvenih mreža ili aplikacija za čet, kupovina preko *online* sajtova i sl. Sa pravom se može reći da čovečanstvo ostavlja digitalne tragove kroz bilo koju aktivnost koja podrazumeva digitalne transakcije, a to je danas – sve. Povrh svega, broj uređaja koji generišu podatke, takođe raste. Koncept nastavlja da evoluira i kroz brojne pojave koje se mogu okarakterisati kao digitalna transformacija, uključujući veštačku inteligenciju, nauku o podacima i *Internet of Things* (IoT). Širom sveta, industrijske mašine opremljene su senzorima koji sakupljaju i odašilju podatke. Uskoro, automobili sa samo-navođenjem pojaviće se na saobraćajnicama, a princip njihove navigacije zavisiće od brojnih podataka. Fenomen eksponencijalnog rasta količine generisanih i transmitovanih podataka, kao i načini čuvanja i obrade tih podataka naziva se *Big Data*. Prevod termina na srpski jezik nije prihvaćen, već se koristi originalni engleski naziv, a prevod bi, u skladu sa značenjem pojma, mogao biti Obimni podaci.

Promene koje je doneo razvoj tehnologije doprineo je i da naučna istraživanja postanu sveobuhvatnija, a kao posledica toga i obimnija. Podaci dobijeni u okviru jednog istraživanja zahtevaju posebnu analizu, obradu i skladištenje i imaju veliku vrednost za celokupnu istraživačku zajednicu. S obzirom na to da se ekologija kao nauka menja i prihvata nove standarde koje 21. vek donosi, svedočimo i procesu u kome naučna zajednica postaje svesna značaja podataka kao vrednog resursa.

Predmet ovog rada je prikaz primene *Big Data* ekologiji, a cilj rada je da se ukaže na izuzetan značaj ovih podataka u ekologiji. Rad, pored uvoda i zaključka sadrži još tri segmenta. U prvom segmentu određen je pojam i predstavljene karakteristike koncepta *Big Data*, a u drugom delu je objašnjen značaj ovog koncepta za ekologiju, a pre svega za ekološka istraživanja. Arhiviranje i upravljanje velikom količinom podataka predstavljeno je u trećem segmentu rada.

2. BIG DATA – POJAM I KARAKTERISTIKE

Od početka upotrebe, jasno je da se pojam *Big Data* prepliće sa mnogim tehnološkim pitanjima, ali još uvek nije prihvaćena jedinstvena definicija[1]. Najjednostavnije je reći da se pod tim pojmom podrazumevaju velike i kompleksne grupe podataka, kod kojih tradicionalne aplikacije i servisi za obradu nisu primenljivi[2]. Da bi se postavio temelj za definisanje ovog pojma, neophodno je odrediti karakteristike koje ovu vrstu podataka razlikuje od ostalih. Prema Gartner izveštaju iz 2001, *Big Data* imaju tri osnovna obeležja – volumen (veličinu), raznovrsnost i brzinu obrade (takozvani *3v's of Big Data*)[1]. Navedene odrednice nisu postigle koncenzus, jer ova obeležja ne obuhvataju odrednicu tačnosti, odnosno istinitosti informacija koju ti podaci nose. U kasnijim proučavanjima došlo se do zaključka da su, u poređenju sa tradicionalnim podacima, osnovne karakteristike *Big Data* volumen, raznovrsnost, brzina, tačnost i vrednost[3]. Kompanija Intel, koja masovno skladišti *Big Data*, ovaj pojam povezuje sa svima koji generišu preko 300 terabajta podataka nedeljno[1]. Microsoft, sa druge strane, termin *Big Data* koristi da opiše primenu kapaciteta računara na

velike skupove informacija[1]. Pored različitih definicija, sve se suštinski slažu da su glavne odlike *Big Data* sledeće[1]:

- Volumen (veličina) podataka;
- Složenost podataka i
- Mogućnost primene savremenih alatki i tehnologija za njihovu obradu.

Izvor: Šematski prikaz 5 osobenosti *Big Data[4]*

Koncept *Big Data* je značajan zbog mogućnost lake obrade, koja podrazumeva redukciju troškova i vremena rada, uprkos tome što se radi o ogromnoj količini podataka. Infrastruktura *Big Data* obuhvataj čitav životni ciklus podataka i istražuje sve prednosti skladištenja na duži vremenski period[5]. Da bi podaci bili iskorišćeni neophodno je da budu[3]: dostupni u potpunosti, održivi, kompatibilni, da olakšavaju rad drugima i da podržavaju i olakšavaju različite naučne eksperimente.

Usled velikog obima podataka, postavlja se pitanje kako iz njih izvući krajnju vrednost i korist i kako ih što bolje razumeti. Segment u kome se razvijaju alati za analizu *Big Data* je odlučujući za odgovor na prethodno pitanje. Ona podrazumeva **ispitivanje skupova podataka kako bi se otkrila korelacija među podacima, skriveni obrasci, odnos i uticaj strukturiranih i nestrukturiranih podataka u skupu**[5]. Takođe, analizom se smatra svako korišćenje naprednih analitičkih tehnika na velikim skupovima podataka. Najčešće se vrši u sklopu platformi za pohranjivanje *Big Data*, kojih je danas puno. Analizom se dolazi do boljih rezultata, povećava se efikasnost i mogućnost predviđanja potencijalnih poteškoća i prepreka. Takođe, utiče na bolje razumevanje heterogenosti među podacima[6]. Zbog veličine podataka, jedan od glavnih izazova analize je izbegavanje moguće lažne korelaciju i algoritamske nestabilnost podataka. Zbog toga, kvalitetno realizovana analiza *Big Data* ostavlja prostor za mogućnost ponovne upotrebe podataka.

Pored komercijalnih korisnika, nauka je takođe jedan od velikih distributera *Big Data*. Najbolju primer transformacije nauke kroz velike podatke vidimo u primeru CERN-a, koji koristi na hiljade računara u preko 150 centara širom sveta, kako bi skladištili i obrađivali podatke koji nastaju u masovnim istraživanjima[8]. Važnost ovih podataka prepoznaju mnoge zemlje iz Evrope i sveta, pa tako imamo primer iz SAD-a, gde u Beloj kući postoji

inicijativa kojom se izdvaja preko 200.000.000 USD za pribavljanje alata za unapređenje naučnih istraživanja koja kao rezultat imaju *Big Data*[9].

Pored velikog značaja za sve sfere društva, važno je pomenuti da je ovaj tip podataka od velike koristi i za kreatore javnih politika, posebno ukoliko su okrenuti konceptu *Evidence Based Policy*.

3. BIG DATA U EKOLOGIJI

U savremenim uslovima, istraživanja i u oblasti ekologije kreiraju veliku količinu različitih podataka[10]. Obim se eksponencijalno povećava upotrebom linearnih akceleratora, senzornih mreža, satelita, seizmografa i sličnih uređaja[10]. Često se postavljalo pitanje da li istraživanja u ekologiji, kao maloj nauci, mogu da kreiraju *Big Data*. U prethodnim decenijama, ekolozi bi se fokusirali na jednu vrstu životinja ili biljaka, na ograničenom prostoru i praćenje vrste je vršeno u ograničenim vremenskim intervalima. Danas istraživanja podrazumevaju praćenje i posmatranje velikih ekosistema, više različitih organizama na duži vremenski period, u različitim vremenskim uslovima. Savremena tehnologija je toliko napredovala da se u istraživanjima u oblasti ekologije brojni podaci skupljaju automatski i velikom brzinom, pa senzori koji prate kretanja životinja mogu slati informacije o lokacijama u realnom vremenu. Takođe, najnovija klimatska istraživanja se obavljaju uz pomoć uređaja koje određene klimatske varijable šalju u bazu na svakih par minuta. Količina podataka koja se dobije u takvim istraživanjima je često velikog obima i mnogo veća od onoga što jedna istraživačka laboratorija može da obradi, te se s pravom svrstavaju u *Big Data*.

Međutim, pored podataka dobijenih iz vrlo kompleksih istraživanja, *Big Data* u ekologiji mogu nastati i udruživanjem sličnih podataka iz više manjih istraživanja koja su vršila različiti istraživački timovi ili radne grupe. Takođe, nastaju i udruživanjem i kompletiranjem podataka sa senzornih platformi koje su prostorno jako udaljene od lokalne baze[11]. Pojedini stručnjaci smatraju da je osnovni problem u ekologiji nedovoljna transparentnost i razmena podataka, pa se smatra da će povećanje *Big Data* u ekologiji, kome smo trenutno svedoci, raširiti svest o značaju naučnih otkrića, pravi način analiziraju, obrađuju, skladište, i transparentni su i dostupni širokoj naučnoj zajednici za neka buduća istraživanja.

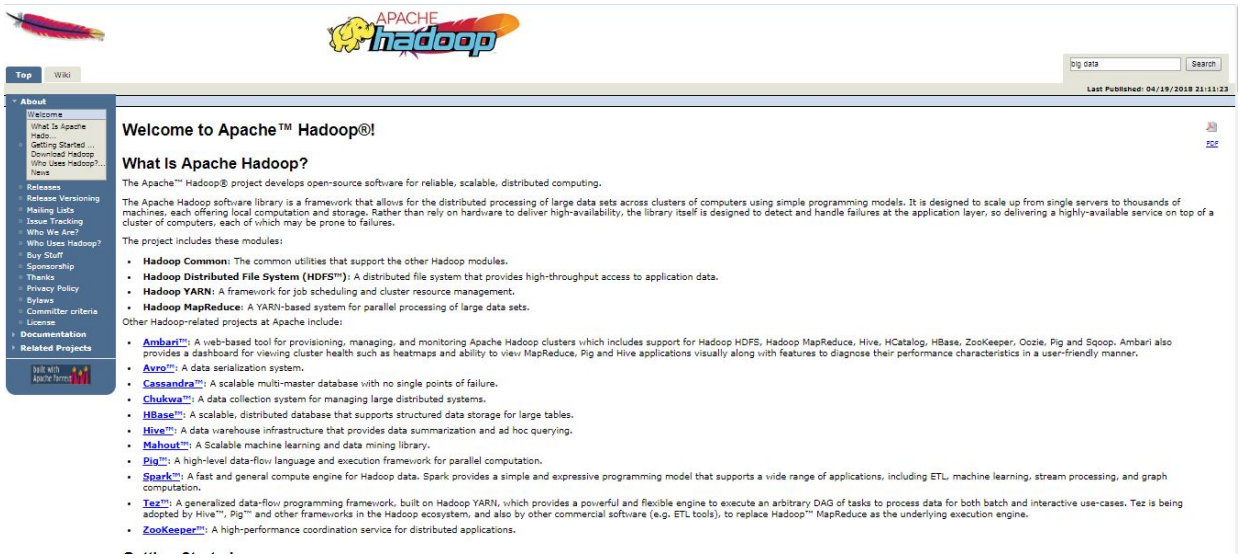
Rad istraživačkih timova je u mnogome olakšan uz ovu vrstu podataka – njihovim izučavanjem, lakše se sagledavaju prednosti, mane i nedostaci projekta na kome se radi, a mogu se predvideti i potencijalne poteškoće, pa se s pravom kaže da istraživačkim timovima postaju glavno oruđe za rad. Njihovo korišćenje i ponovna upotreba skraćuju vreme ekoloških istraživanja, štede novac i pružaju mogućnost za jednostavniji rad u budućim istraživanjima.

Smatra se da *Big Data* u potpunosti menjaju pristup ekološkim istraživanjima jer pružaju brojne pogodnosti. Njihovom analizom i upotrebom, dolazi se do razvijanja novih modela zasnovanih na teorijama koje objašnjavaju određeni ekološki fenomen[12]. Takođe, uspešno se utvrđuju već postavljene teoreme i moguće je predvideti buduće ekološke pojave bez ranijeg posmatranja, koje često može biti dugo i skupo[12]. Pravilnom upotrebom, mogu se pružiti istraživačima dodatne informacije o entitetu koji se proučava u datom trenutku i to sa više različitih aspekata. Bitno je istaći i da dolazi do kvalitetnije saradnje ekologa – ako su okupljeni oko istih ili sličinih podataka, razmenom iskustava i mišljenja se lako rađaju nove ideje. Sve ovo kao krajnji rezultat može dati kvalitetnija istraživanja, smanjenje utrošenog vremena u radu i manje finansijske izdatke.

Ekolog Majk Viling (Mike Willing) je u svojoj knjizi, *Dugoročna ekološka istraživanja: promena prirode naučnika,* objasnio da su *Big Data* u potpunosti promenili naučnike i njihov pristup nauci[13]. On navodi da su podaci okupili ekologe, ali ne samo njih – okupili su i stručnjake koji se bave drugim oblastima, a mogu pomoći u dobijanju odgovora na pitanja koja se tiču, na primer, velikih klimatskih promena[13].

Prilikom praćenja određene životinjske ili biljne vrste, odašiljači postavljeni na terenu mogu slati informacije u određenim vremenskim intervalima. To znači da se za vrlo kratko vreme prima jako veliki broj podataka različite strukture. Veliki nedostatak *Big Data* je to što su podaci koji kontinuirano stižu često vrlo neorganizovani, te mora postojati unapred određen plan za pravilno razvrstavanje prema strukturi. Ako se njihova organizacija i selekcija ne izvrši na vreme, mogu podleći nekvalitetnoj analizi ili pak ostati u potpunosti neiskorišćeni. Zbog toga je neophodno da postoje obučena lica koja će na pravi način upravljati ovom vrstom podataka. Takođe, količina podataka koja stiže u bazu sa terena akumulira se velikom brzinom i postavlja se pitanje njihovog skladištenja. Zbog vrednosti koju ovi podaci nose sa sobom, potreba da se arhiviraju i osposobe za ponovno korišćenje je neminovna.

Tradicionalna skladišta i repozitorijumi nemaju mogućnosti za arhiviranje *Big Data*, te su neophodna skladišta visokih performansi koja će ove podatke uspeti da čuvaju, štite, omoguće pristup i obezbede ponovno korišćenje. Infrastruktura ovih skladišta pruža valjano upravljanje i obradu podataka u realnom vremenu. Jedna od najpopularnijih platformi je *Hadoop*, programski okvir koji podržava procesiranje i skladištenje ekstremno velikih skupova podataka, i inicijalno je postavljen 2011. godine[14]. Dizajniran je za pravilno upravljanje *Big Data*. Mnogi stručnjaci kažu da je upravo *Hadoop* promenio viđenje i rukovanje *Big Data* jer danas zauzima čak 90% tržišta u odnosu na ostale softvere koji se bave arhiviranjem. Pored ovog, tu su i mnoga druga rešenja predviđena za skladištenje i upravljanje podacima, poput *BigTable* i *MapReduce* platforme. Kao najčešći problem, javlja se cena ovih platformi jer brojni istraživački timovi nemaju dovoljno finansijskih sredstava. Gledajući širu sliku, instalacija omogućava kasnije moguće migracije podataka, upravljanje i manipulaciju podacima, što ima više prednosti od nedostataka za instituciju i istraživački tim. Koristeći pohranjene velike podatke, dobijaju se važni odgovori na pitanja postavljena na početku nekog novog istraživanja, što olakšava rad, smanjuje troškove i skraćuje vreme rada.



Izvor: početna strana platforme Hadoop[14]

4. ZAKLJUČAK

Big Data su ušli u sve pore društva – zdravstvo, preduzetništvo, obrazovanje, nauku. Njihova budućnost se teško može predvideti, jer će nove tehnologije doneti napredne alate i platforme koje će znatno unaprediti mogućnosti njihove upotrebe, analize, skladištenja i upravljanja. Istraživanja u oblasti ekologije iz godine u godinu su na sve zavidnijem nivou i očekuje se ekspanzija *Big Data* koji će biti od velike koristi svim naučnicima. Iako su u mnogim zemljama godinama unazad na snazi razne inicijative za upravljanje *Big Data*, u Srbiji to nije slučaj. Konačno, predviđa se da će razvoj procedura za upravljanje *Big Data* pružiti šansu za kreiranje javnih politika.

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

Literatura

[1] Ward, J.S., Barker, A. (2013). Undefined by Data: A Survey of Big Data Defintions
[2] Big Data- What it is and why it matters. https://www.sas.com/en_us/insights/big-data/what-is-big-data.html pristupano: 16.4.2018.
[3] Jin, X, Wah, B.W., Cheng, X., Wang, Y. (2015). Significance and Challenges of Big Data Research
[4] Amazing Big Data in MicroStrategy. https://www2.microstrategy.com/producthelp/10.7/WebUser/WebHelp/Lang_1033/Content/mstr_big_data.htm pristupano: 16.4.2018.
[5] Shin, D.H. Choi, M.J. (2015). Ecological views of big data: Perspectives and issues
[6] What is Big Data Analysis? https://www.quora.com/What-is-big-data-analysis-2 pristupljeno: 21.4.2018.
[7] Fan, J., Han, F., Liu, H. (2013). Challenges of Big Data Analysis.
[8] How is Big Data Used in Practice? 10 Use Cases Everyone Must Read. https://www.bernardmarr.com/default.asp?contentID=1076 pristupano: 22.4.2018.
[9] Big Data is a Big Deal. https://obamawhitehouse.archives.gov/blog/2012/03/29/big-data-big-deal pristupano: 23.4.2018.
[10] Bradić-Martinović, A., Malić, L., Banović, J. (2017). Izazovi kreiranja i ponovne upotrebe podataka u ekologiji.
[11] Soranno, Patricia A., Schimel, David S. (2014) Macrosystems ecology: big data, big ecology.
[12] Xu, M., Cai, H., Liang, S. (2015). Big Data and Industrial Ecology.
[13] How Big Data Changed the Science of Ecology. https://today.uconn.edu/2016/09/big-data-changed-science-ecology/ pristupano: 17.4.2018.
[14] Hadoop Apache. http://hadoop.apache.org/ pristupano: 16.4.2018.