

# Veliki podaci u informacijskim ustanovama - primijenjene i kritičke dimenzije

---

Horvat, Sven

Master's thesis / Diplomski rad

2021

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Humanities and Social Sciences / Sveučilište u Zagrebu, Filozofski fakultet**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:131:599508>

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

Download date / Datum preuzimanja: **2023-12-06**



Repository / Repozitorij:

[ODRAZ - open repository of the University of Zagreb Faculty of Humanities and Social Sciences](#)



SVEUČILIŠTE U ZAGREBU  
FILOZOFSKI FAKULTET  
ODSJEK ZA INFORMACIJSKE I KOMUNIKACIJSKE ZNANOSTI  
SMJER BIBLIOTEKARSTVA  
Ak god. 2020./2021.

Sven Horvat

**VELIKI PODACI U INFORMACIJSKIM  
USTANOVAMA - PRIMIJEJENE I KRITIČKE  
DIMENZIJE**

Diplomski rad

Mentor: dr. sc. Sonja Špiranec

Zagreb, rujan 2021.

## **Izjava o akademskoj čestitosti**

Izjavljujem da je ovaj rad rezultat mog vlastitog rada koji se temelji na istraživanjima te objavljenoj i citiranoj literaturi. Izjavljujem da nijedan dio rada nije napisan na nedozvoljen način, odnosno da je prepisan iz necitiranog rada, te da nijedan dio rada ne krši bilo čija autorska prava. Također izjavljujem da nijedan dio rada nije korišten za bilo koji drugi rad u bilo kojoj drugoj visokoškolskoj, znanstvenoj ili obrazovnoj ustanovi.

# Sadržaj

1	Uvod .....	1
2	Veliki podaci .....	4
2.1	Volumen .....	5
2.2	Brzina .....	5
2.3	Raznolikost .....	6
2.4	Integritet .....	6
2.5	Vrijednost .....	6
2.6	Metode, tehnike i tehnologije velikih podataka .....	8
3	Odjek velikih podataka .....	11
3.1	Odjek velikih podatak u informacijskim znanostima i knjižnicama .....	13
4	Knjižnice i podaci .....	16
4.1	Veliki podaci knjižnica .....	16
4.2	Uloga knjižnice u eri velikih podataka .....	17
4.2.1	Skrb o podacima .....	18
4.2.2	Dijeljenje podataka .....	19
4.3	Predispozicije za rad s velikim podacima .....	21
5	Primjena velikih podataka u knjižnicama .....	22
5.1	Rudarenje podataka .....	22
5.2	Vizualizacija podataka .....	22
5.3	Personalizirane informacijske usluge .....	23
5.4	Samo-vrednovanje pomoću velikih podataka .....	25
5.5	Podaci kao resurs .....	27
6	Problemi i izazovi .....	29
6.1	Promjena mentaliteta .....	29
6.2	Znanje i vještine potrebne za rad s velikim podacima .....	30
6.3	Problem budžeta .....	32
6.4	Problemi prouzrokovani dimenzijama velikih podataka .....	33
6.5	Potreba za kritičkom podatkovnom pismenošću .....	34
7	Zaključak .....	36
8	Popis literature .....	39
9	Sažetak .....	43

# 1 Uvod

Svjedoci smo neviđene brzine tehnološkog napretka, napretka i probitaka u raznim znanstvenim područjima, te napretka društva općenito. Svaki takav napredak u pozadini ima znanje, odnosno može se reći da napredak i znanje, koje čovjek ima o svijetu, stoje u recipročnom odnosu. Više znanja uvjetovat će brži napredak, a napredak sa sobom nosi mogućnost otkrivanja novih znanja.

Bilježenje informacija konstanta je u evolucijskom procesu čovjeka (Burke, 1992., prema Oliveira i Cendón, 2020: 155). Od pojave pisma čovjek je počeo bilježiti sve više segmenata svojeg života (Hoy, 2014: 2). Bilježi što radi i na koji to način radi, što u svijetu postoji i s čime se sve susreće. Bilježe se odnosi među ljudima, jednostavne transakcije, misli, dojmovi, te mnoge druge pojave i ideje, sve u svrhu budućeg korištenja takvih zapisa (Hoy, 2014: 2). Dakle, čovjek bilježi informacije i stvara podatke o svojem životu i svijetu u nadi da će mu informacije sadržane u tim zapisima biti korisne u budućim pothvatima. Podaci su svugdje i svaki dio čovjekova života može biti pretvoren u podatak. Sve što ljudi rade, kažu i promatraju stvara podatke. Iako je stvaranje podataka u prošlosti možda bio naporan i dugotrajan proces, danas, s razvojem tehnologije, posebice razvojem računalne tehnologije i pojavom Interneta, stvaranje podataka nikada nije bilo lakše. Razvoj računalne tehnologije rezultirao je neviđenim brzinama stvaranja podataka, tim više, što je gotovo svaki segment čovjekovog života u nekoj mjeri povezan s računalima i Internetom (Hoy, 2014: 2). Zahvaljujući tom razvoju, sposobnosti da se podaci prikupe, analiziraju i obrade znatno su se povećale u zadnjim desetljećima (Hoy, 2014: 2). Digitalizacija čovjekovog svijeta neprestano napreduje i ne pokazuje znakove zaustavljanja (Ball, 2019: 1). Nema sumnje da se čovjek danas nalazi u eri podataka, podataka koje dobivamo iz različitih izvora kao što su: senzorna tehnologija, pametni uređaji, interakcije na društvenim mrežama, web stranice, video zapisi, slike, *online* transakcije itd. (Alotaibi i Abdullah, 2016: 687 – 696, Wu i sur., 2014: 97 – 107, prema Li i Dou, 2017: 4). Podaci različitih formata stvaraju se neviđenim brzinama upravo zato što proces prikupljanja podataka, koji je prije bio naporan i dugotrajan, sada postaje sve više automatiziran i pohrana tih podataka postaje sve jeftinija zahvaljujući razvoju tehnologije.

Ljudi su počeli prepoznavati značaj rastuće količine informacija i podataka do te mjere da se podaci nazivaju „naftom 21. stoljeća“ (Ball, 2019: 2). Podaci se smatraju sirovinom čijom obradom i analizom čovjek dolazi do novih vrijednosti, sirovinom za nove proizvode i usluge (Aaltonen, Aleski i Tempini, 2014: 108, prema Oliphant, 2017: 6). Radna pozicija stručnjaka za podatke prepoznata je kao najprimamljivije zanimanje 21. stoljeća upravo zato što nam oni objašnjavaju kako analiza i primjena podataka može riješiti najsloženije i najhitnije svjetske probleme (Meyer, Schroeder i Linnet, 2013, prema Oliphant, 2017: 2). Može se reći da je u svijetu prisutna određena zaludnost podacima. Stoga, nije čudno, da se tu masovnu količinu podataka nastojalo opisati jednim novim pojmom – veliki podaci.

Veliki podaci predstavljaju ogromnu količinu strukturiranih, polu-strukturiranih i nestrukturiranih podataka (podataka različitih formata), koji se generiraju visokim brzinama i koji se ne mogu pohraniti, analizirati i obraditi tradicionalnim metodama (Oliveira i Cendón, 2020: 155). Njihovom obradom i analizom čovjek dolazi do vrijednih uvida, odnosno otkrivaju se obrasci, trendovi i ponajprije korelacije koje su često vezane uz ljudsko ponašanje i ljudske interakcije (Mayer-Schönberger i Cukier, 2013, 50-72). Oni stvaraju nove oblike vrijednosti, te preoblikuju način na koji ljudi shvaćaju i doživljavaju zbilju i način na koji odlučuju (Mayer-Schönberger i Cukier, 2013., prema Oliphant, 2017: 6). Više podataka ne znači naprosto više, već i drugačije (Anderson, 2008, prema Hoy, 2014: 2). Prikupljanjem, obradom i analizom velikih podataka podupire se i omogućuje otkrivanje informacija i znanja kao i diseminacija i korištenje tog znanja u različitim područjima i u različite svrhe (Weiss i Zgorski, 2012., prema Li i Dou, 2017: 4). Naravno, osim očitih koristi, vrijednosti i novih perspektiva koje prate velike podatke, važno je shvatiti da veliki podaci pred društvo postavljaju nove izazove. Ti izazovi tiču se potrebe za prikladnom i ispravnom obradom ogromnih količina podataka, obrade koja će dovesti do vrijednih zaključaka. Osim tehnološke potpore potrebne za ispravnu obradu, potrebne su i čovjekove vještine i kompetencije.

Prepoznaje se da će otkrivanje, prisvajanje i obrada tih ogromnih količina podataka biti jedan od glavnih čimbenika društvenog napretka (Wang, 2016: 664). Veliki podaci utjecat će na svaku industriju gdje oni mogu pomoći bolje shvatiti potrošača, odnosno korisnika (Ball, 2019: 2). Od 2012. godine gotovo je svaki sektor razvio zanimanje za velike podatke i njihove neviđene potencijale za analitičke probitke

(Wittmann i Reinhalter, 2014: 2). Institucije koje ne uspiju prihvatiti ili odbace velike podatke kao mogućnost zaostajat će za onima koje donose odluke na temelju tih podataka (Bailey, 2015., prema Ball, 2019: 2). Jasno je da veliki podaci i podatkovna analitika mijenjaju čitave industrije kao što su zdravstvo i obrazovanje, te nije iznenađenje da se zbog popularnosti velikih podataka oblikovalo tržište usmjereno na njih (Waller i Fawcett, 2013: 77 – 84, prema Jaiswal, 2020: 4).

Golemim porastom podataka, informacija i znanja u svijetu, informacijske potrebe postale su složenije nego su bile ikad prije, te se podrazumijeva da ih je sada teže zadovoljiti. Prisutnost informacijskih znanosti u tom novom informacijskom kontekstu je nužna, kao i potreba da informacijske znanosti ponovno potvrde svoj položaj u informacijskom scenariju u kojem su prisutni veliki podaci (Oliveira i Cendón, 2020: 156). Ovo ponajprije podrazumijeva odrediti ulogu koju će informacijski, odnosno podatkovni stručnjak imati u kontekstu novih radnih pozicija, za koje bi on trebao biti najosposobljeniji (Oliveira i Cendón, 2020: 156). Skrb o podacima, upravljanje, objavljivanje, vizualizacija, pohrana i očuvanje podataka kao i pružanje podatkovnih usluga su područja stručnosti knjižničara i informacijskih stručnjaka kojima oni mogu značajno doprinijeti (Oliphant, 2017: 2). Prema tome, neizbježan je utjecaj velikih podataka na tradicionalne centre informacijskih usluga – knjižnice. Da bi knjižnice ostale aktualne, te da bi pružale kvalitetne informacijske usluge, odnosno bile u stanju zadovoljiti sve složenije informacijske potrebe svojih korisnika one će trebati prikupljati i obrađivati ogromne količine podataka i pružati nove alate za otkrivanje znanja (Wang, 2016: 664). Također, budući da knjižnice nude sve više podatkovnih usluga, očita je mogućnost i prilika da knjižničari postanu stručnjaci i autoriteti u doba podataka. Dakle, izazovi koji postavljeni pred knjižnice kao informacijske ustanove jesu pružanje adekvatnih informacijskih usluga temeljenih na podacima, njihov razvoj i općenito, način na koji će se suočiti s velikim podacima (Wang, 2016: 664).

Ovaj rad će pokušati odrediti odjek velikih podataka u društvu, posebice u informacijskim institucijama, odnosno knjižnicama, načine na koje se veliki podaci mogu primijeniti u knjižnicama, te izazove koji oni postavljaju knjižnicama

## 2 Veliki podaci

. Kako bi se postigla svrha ovog rada, mora se, prije svega, navesti precizna definicija velikih podataka. Međutim, precizno definiranje podataka nije lak pothvat, a razlog tome jest da se definicije i određenja velikih podataka, od njihove prvobitne konceptualizacije, konstantno mijenjaju, odnosno razvijaju (Kitchin i McArdle, 2015: 471 – 481, prema Oliveira i Cendón, 2020: 156). Ne postoji jedinstveno autorstvo ili jedinstvena definicija za koncept velikih podataka (Oliveira i Cendón, 2020: 156). Ispravne definicije velikih podataka u obzir moraju uzimati uvijek nove načine dolaženja do podataka, rastuće brzine prikupljanja, obrade i analize podataka, formate podataka, odnosno njihovu različitost kao i načine na koje se dolazi do vrijednosti, odnosno koristi od velikih podataka.

Veliki podaci su podaci koji premašuju kapacitete obrade (procesuiranja) konvencionalnih sustava baza podataka, odnosno podaci su preveliki, kreću su previsokim brzinama i ne uklapaju se u strukture tradicionalnih arhitektura baza podataka (Dumbill, 2012, prema Oliveira i Cendón, 2020: 156). Kako bi se došlo do vrijednosti sadržanih u velikim podacima moraju se odabrati alternativni načini obrade (Dumbill, 2012, Hey i Hey, 2006: 515 – 528, prema Oliveira i Cendón, 2020: 156). Zato, nije dovoljno velike podatke svesti na puke podatke s određenim karakteristikama. Koncept velikih podataka obuhvaća i metode, odnosno tehnike kojima se prikupljaju, pohranjuju, distribuiraju, upravljaju i analiziraju veliki skupovi podataka, skupovi za čije upravljanje tradicionalne metode upravljanja podataka nisu prikladne (Wang i sur. 2016: 2). Dakle, koncept velikih podataka obuhvaća i novu generaciju tehnologija, tehnika i metoda stvorenih u svrhu ekonomične ekstrakcije vrijednosti iz velike količine raznolikih podataka time što omogućuju veliku brzinu prikupljanja, otkrivanja i analize (Gantz i Reinsel, 2011: 1 – 12, prema Al-Barashdi i Al-Karousi, 2018: 3). Sjedinjujući prethodne dvije definicije može se reći da su veliki podaci oni podaci koje karakteriziraju velik volumen, visoke brzine stvaranja i prikupljanja, te različitost formata i za čiju su transformaciju u vrijednost potrebne posebne tehnologije i analitičke metode.

Iako postoje mnoge definicije koje velike podatke pokušavaju definirati pomoću različitih pristupa i s različitih gledišta, kao najpopularnija se pokazala ona koja velike



podatke definira na osnovi njihovih karakteristika. Prvu takvu definiciju ponudio je Laney, 2001. godine, opisujući skup karakteristika koje podaci trebaju zadovoljiti, odnosno imati kako bi se ih nazivalo velikim podacima (Laney, 2001., prema Wang i sur. 2016: 1). Te karakteristike, tj. dimenzije velikih podataka nazvane su „tri V-ova“, a označavaju početna slova engleskih riječi za volumen (engl. *volume*), brzinu (engl. *velocity*) i raznolikost (engl. *variety*) (Laney, 2001., prema Wang i sur. 2016: 1). Naknadnim izučavanjem velikih podataka dodane su još dvije dimenzije: integritet podataka (engl. *veracity*), te jedna od najvažnijih dimenzija – vrijednost (engl. *value*), koja označuje koristi velikih podataka (Waqar i Kanwal, 2017: 21). Rad će u nastavku ukratko opisati što svaka od ovih dimenzija označuje, s iznimkom dimenzije vrijednosti, na kojoj se potrebno duže zadržati, budući da se ona pokazuje kao najvažnija karakteristika velikih podataka. Navedenih pet dimenzija nisu sve koje su se pojavile u istraživanjima koncepta velikih podataka i nove konstantno nastaju (Kitchin i McArdle, 2015: 471 – 481, prema Oliveira i Cendón, 2020: 156).

## 2.1 Volumen

Veličina skupova velikih podataka jest neusporedivo veća od veličine uobičajenih podataka. Iako ne postoji točno određena granica koju veličina podataka treba prijeći kako bi ih se nazivalo velikim podacima, često se govori o terabajtima, petabajtima pa čak i eksabajtima podataka (Wang i sur. 2016: 2). Veličina podataka varira i ovisi o disciplinama i industrijama u kojima se stvaraju, prikupljaju i obrađuju ali konstanta, bez obzira na područje, jest da količina podataka raste do te mjere da alati za obradu i prikupljanje pred sobom imaju „pokretnu metu“ (Wang i sur. 2016: 2). Ta količina određuje vrijednost i potencijal prikupljenih i obrađenih podataka (Jharotia, 2016: 2).

## 2.2 Brzina

Brzina, odnosno „velocity“ označava brzinu kojom se stvaraju podaci. Razvojem računalne i senzorne tehnologije, kao i razvojem Interneta podaci se stvaraju neviđenim brzinama i konstantno, a prikupljanje, obrada i analiza tih podataka nastoji se odraditi u vremenu koje se gotovo preklapa s vremenom njihovog nastanka (Alotaibi i Abdullah, 2016: 687 – 696, prema Li i Dou, 2017: 4). Razlog tome jest što su uvidi i informacije na osnovi podataka često potrebni u stvarnom vremenu, odnosno, podaci koji se ne prikupljaju, ne obrađuju i ne analiziraju kada je to

najpotrebnije, gube na vrijednosti (propušta se donošenje ispravnih odluka na temelju podataka, u vremenu kada su te odluke ključne za neki cilj ili svrhu). Dakle, radi se o brzini kojom se stvaraju i procesuiraju podaci kako bi se zadovoljili zahtjevi društvenog razvoja kao i potrebe tržišta.

### 2.3 Raznolikost

Podaci se prikupljaju iz raznih izvora te su prikazani u raznim formatima (Li i Dou, 2017: 6). Za razliku od prijašnjih vremena kada su podaci većinom bili strukturirani, danas su podaci većinom polu-strukturirani ili nestrukturirani, odnosno pojavljuju se u obliku video-zapisa, audio-zapisa, slika, web stranica itd, podataka koje generiraju korisnici itd. Doći do maksimalne vrijednosti sadržane u podacima ujedno znači koristiti sve podatke koji se nude, a to predstavlja priliku i izazov za sve one koji se bave velikim podacima (Jharotia, 2016: 2).

### 2.4 Integritet

S obzirom na to, da je količina podataka do kojih čovjek može doći toliko velika, da se podaci stvaraju na različite načine i na različitim mjestima, te da se stvaraju konstantno i visokim brzinama, važno je na umu imati njihovu kvalitetu. Kontrola kvalitete podataka oduvijek je bila od ključne je važnosti, a to se u doba velikih podataka ne mijenja. Vjerodostojnost izvora podataka varira od izvora do izvora, a kako bi se donosile najbolje odluke u nekom danom kontekstu, te odluke moraju biti temeljene na kvalitetnim podacima (Jharotia, 2016: 3).

### 2.5 Vrijednost

Svaka od navedenih dimenzija podataka do neke mjere daje odgovor na pitanje: „Što su veliki podaci?“ Međutim, ključno pitanje jest: „Zašto koristiti velike podatke?“ Odgovor na ovo pitanje daje dimenzija vrijednosti.

Da bi se dimenzija vrijednosti objasnila potrebno je ukazati na razliku obrade podataka prije pojave velikih podataka i nakon nje. Ono što je obilježavalo prijašnja vremena jest oskudnost podataka i rigoroznost u njihovoj obradi (Mayer-Schönberger i Cukier, 2013, 13). Obično se obrađivao, odnosno, analizirao manji skup podataka – uzorak, na temelju kojeg su se donosile odluke i zaključci (Mayer-Schönberger i

Cukier, 2013, 13). Podaci koji su činili uzorak morali su biti strogo kontrolirani, uređeni i standardizirani, a nakon njihove obrade i analize bilo je moguće otkriti kauzalne odnose među pojavama u pitanju (Mayer-Schönberger i Cukier, 2013, 13). Rigoroznošću i pažljivim odabirom podataka nastojalo se pojavu pogrešaka u zaključivanju, na temelju uzorka, svesti na minimum, no što je količina podataka za obradu bila veća to su greške bile učestalije (Mayer-Schönberger i Cukier, 2013, 13). Također, vrijeme prije velikih podataka obilježavala je teškoća prikupljanja, pohrane i analize podataka što je uvelike ograničavalo mogućnosti ekstrakcije potencijalnih vrijednosti iz tih podataka (Mayer-Schönberger i Cukier, 2013, 100). Pojavom velikih podataka ova se obilježja drastično mijenjaju. U prvom redu, prikupljanje i pohranjivanje podataka postaje jeftine i lakše zbog razvoja tehnologije. Prikupljanje podataka sve je više automatizirano i pasivno, te ne iziskuje truda kao što je prije. Troškovi pohrane tih podataka također postaju sve niži, do te mjere da je lakše opravdati zadržavanje tih podataka negoli je njihove odbacivanje (Mayer-Schönberger i Cukier, 2013, 100). Zbog ovih promjena društvo se približilo novom načinu mišljenja, odnosno zaključivanja – načinu koji se razlikuje od donošenja odluka na temelju uzorka. Nikad prije čovjek nije imao bolju priliku prikupiti i obraditi sve (ili približno sve) podatke o nekoj složenoj pojavi (Mayer-Schönberger i Cukier, 2013, 26). Glavna posljedica (ili prednost) posjedovanja gotovo svih podataka o nekoj pojavi jest ta da se može odstupiti od prethodno navedene rigoroznosti (Mayer-Schönberger i Cukier, 2013, 13). Zahtjevi za točnošću podataka koji su pratili postupak zaključivanja iz uzorka gube na svojoj težini, odnosno može se reći da se na osnovi manjeg broja rigorozno odabranih i uređenih podataka podrazumijeva točnost zaključaka, dok na osnovi velikog broja neuređenih podataka varijabilne kvalitete čovjek može biti zadovoljan uvidom u općeniti smjer kretanja neke pojave (Mayer-Schönberger i Cukier, 2013, 13). Obradom svih (gotovo svih) podataka o nekoj pojavi i konstantnim prikupljanjem novonastalih podataka, promjene vezane uz neku pojavu postaju očite, a to znači da obradom velikih podataka čovjek otkriva obrasce i korelacije koje često dovode do novih i iznenađujućih uvida (Mayer-Schönberger i Cukier, 2013, 14). Iako traženje uzroka (kauzalnosti) u svemu jest način na koji čovjekov um spoznaje svijet oko sebe, neosporno je da takva logika ograničava ono što može vidjeti, odnosno ono što je pred njime. Ovakav način razmišljanja i spoznavanja zbilje veliki podaci trebaju nadopuniti i proširiti korelacijama i obrascima do kojih njihovom obradom dolazimo. Te korelacije nam možda neće reći

zašto se nešto događa ali će svakako upozoriti na to što se događa (Mayer-Schönberger i Cukier, 2013, 14). Nadopunjavajući karakter kauzalnih odnosa i korelacija Mayer i Cukier naglašavaju u djelu „Big Data A Revolution that will transform how we live, work and think“.

Dakle, vrijednosti do kojih dolazimo pomoću velikih podataka, u prvom se redu očituju u obliku uvida i neočekivanih korelacija, koje postaju temelj za djelovanje i odlučivanje, te opisuju ili čak predviđaju neku pojavu ili ponašanje. Međutim, postoje i vrijednosti do kojih dolazimo naknadno, nakon prvobitnog korištenja podataka, odnosno ponovnim korištenjem podataka (Mayer-Schönberger i Cukier, 2013, 107). Za razliku od materijalnih dobara vrijednost podataka se ne iscrpljuje korištenjem, tj. korištenje podataka od strane jedne osobe ne onemogućuje korištenje istih podataka od strane druge osobe (Mayer-Schönberger i Cukier, 2013, 101). Prvobitno korištenje podataka u neku svrhu tek je početak nakon kojeg slijede moguće uporabe u brojne druge svrhe – svrhe koje jedino ograničava ljudska kreativnost i domišljatost. Prema tome, vrijednost podataka mora se razmotriti u kontekstu potencijalnih budućih uporaba, jer ograničiti se na prvobitnu uporabu i ondje stati dovodi do gubitka vrijednosti. Upravo se zato istinska vrijednost podataka uspoređuje sa santom leda, čiji veći dio površnim promatranjem čovjek ne uspijeva vidjeti (Mayer-Schönberger i Cukier, 2013, 103). Velikim podacima čovjek otkriva dio zbilje koji mu dotad nije nužno bio poznat.

## 2.6 Metode, tehnike i tehnologije velikih podataka

Porastom količine i složenosti podataka porasle su i potrebe pohrane i obrade. Dva načina rješavanja problema pohrane i obrade, svojstvena tehnologiji velikih podataka, predstavljaju računanje na oblaku (engl. *cloud computing*) i „cluster“ računanje (engl. *cluster computing*) (Minelli, Chambers i Dhiraj, 2013: 94). Osnovni princip rada obje vrste računanja jest dijeljenje računalnih resursa, odnosno pohranjivanje podataka na oblak i raspodjela zadataka obrade na skup umreženih računala (Minelli, Chambers i Dhiraj, 2013: 94). „Cluster“ računanje podrazumijeva skup računala koja se prilikom rješavanja zadataka ponašaju kao jedinstvena procesna jedinica (Minelli, Chambers i Dhiraj, 2013: 94). U ovoj vrsti računanja jedno računalo postavlja zadatak i distribuira ga na ostale članove „cluster“, te nadgleda rješavanje tog zadatka (Minelli, Chambers i Dhiraj, 2013: 94). Time se zahtjevi obrade postavljani jednoj procesnoj

jedinici razlamaju na prihvatljive dijelove koji se obrađuju istovremeno, čime se kapaciteti i brzine obrade znatno povećavaju (Minelli, Chambers i Dhiraj, 2013: 92). Sustav „Apache Hadoop“ ovaj princip rada primjenjuje na obradu ali i na pohranu podataka. „Hadoop Distributed File System“ podatke razlama i distribuira svojim članovima, odnosno svaki poslužitelj pohranjuje tek mali dio većeg skupa podataka (Minelli, Chambers i Dhiraj, 2013: 63). Također, ukidanjem granica pohrane, fragmente podataka moguće je replicirati u više navrata, čime se rizik gubljenja podataka znatno smanjuje (Minelli, Chambers i Dhiraj, 2013: 63).

U središtu novog informacijskog konteksta nalazi se potreba za analizom velikih podataka (Jaiswal, 2020: 3). Analiza velikih podataka jest proces kojim se obrađuju ogromne količine podataka, kako bi se otkrili skriveni obrasci i nepoznate veze među podacima (Jaiswal, 2020: 3). Tim se procesom informacije razlamaju, te se primjenom kalkulacija, programiranja i metoda mjerenja otkrivaju profitabilne veze među informacijama, odnosno podacima (Jaiswal, 2020: 3). Prikupljeni i obrađeni podaci se tek pomoću analize mogu transformirati u vrijedne uvide (Tableau Software, 2018) S obzirom na složenost i količinu podataka koje treba analizirati, razvile su se mnoge vrste i tehnike analize kao i alati za analize u specifične svrhe.

Vrste analiza zastupljenih u eri velikih podataka mogu se podijeliti u tri kategorije: „batch“ analiza kojom se podaci filtriraju prije nego li budu analizirani, „stream“ analiza kojom se podaci analiziraju u vremenu istovjetnim s njihovim nastankom ili od trenutka prikupljanja, te interaktivna analiza koja omogućuje da korisnici provode analize iz perspektiva vlastitih potreba i ciljeva (Al-Barashdi i Al-Karousi, 2018: 7). Rudarenje podataka, predviđajuća analitika (engl. *predictive analytics*), duboko učenje (engl. *deep learning*) i vizualizacija podataka samo su neke su od tehnika, koje omogućuju pronalaženje obrazaca i uvida sadržanih u odnosima velikih količina podataka (Tableau Software, 2018) Rudarenje podataka uključuje procese razvrstavanja podataka i algoritme za uočavanje anomalija, predviđajuća analitika oslanja se na historijske podatke neke organizacije, kako bi se identificirali budući rizici i prilike, duboko učenje u pozadini ima strojno učenje i umjetnu inteligenciju u svrhu pronalaženja obrazaca u podacima s visokom razinom složenosti i apstraktnosti, a vizualizacija podataka omogućuje korisnicima interakciju s podacima, tj. s procesom analize (Tableau Software, 2018) Svaka od ovih tehnika će u većoj ili manjoj mjeri reflektirati svaku vrstu analize, međutim jedna vrsta analize uvijek

prevladava (Al-Barashdi i Al-Karousi, 2018: 7). Primjerice, u vizualizaciji podataka naglašena je interaktivna analiza, dok je rudarenju podataka svojstvena „batch“ analiza (Al-Barashdi i Al-Karousi, 2018: 7).

### 3 Odjek velikih podataka

Odjek, odnosno utjecaj velikih podataka vidljiv je u komercijalnom i ne-komercijalnom sektoru, ali ono što je bitno za svrhu ovog rada jest utjecaj velikih podataka na informacijske znanosti, LIS (Library and information science) i knjižnice, pa će se, prema tome, taj utjecaj opširnije opisati.

Odjek velikih podataka u komercijalnom svijetu, u prvom je redu vezan uz pronalaženje novih vrijednosti čime neka organizacija može ostati konkurentna – što ona mora biti ako želi opstati u nemilosrdnom kapitalističkom okruženju. Upravo se zato, era velikih podataka uspoređuje s vremenima „zlatne groznice“, što se očituje u hitrosti potraga privatnog sektora za probicima omogućenim velikim podacima (Wittmann i Reinhalter, 2014: 5). Korporacije sve više počinju istraživati koncept velikih podataka i ulažu u razvoj i primjenu tehnologije prikladne za obradu takve vrste podataka (Li i Dou, 2017: 4). Google, Facebook, IBM i Microsoft su primjerice izgradili centre za istraživanje velikih podataka, te surađuju kako bi prikupljali, analizirali i primijenili velike podatke (Zikopoulos i sur., 2012., prema Li i Dou, 2017: 4). Razvojem tehnologije obrada masivnih skupova podataka postaje moguća. Ti se skupovi mogu povezivati u svrhu pronalaženja informacija, a obrasce je moguće pronaći u stvarnom vremenu (Al-Barashdi i Al-Karousi, 2018: 13). Ne samo da organizacije na osnovi promotrenih obrazaca mogu predviđati ishode događaja ili fenomena koji će utjecati na tržište, već im je sada moguće predviđati, često s vrlo visokom vjerojatnošću, kretanje samog tog tržišta. Pomoću obrade velikih skupova podataka one mogu donositi pravovremene odluke u svrhu izbjegavanja rizika i procijeniti koliko je isplativo izložiti se nekom riziku (Al-Barashdi i Al-Karousi, 2018: 13). Također, uvidi u ponašanje korisnika, odnosno potrošača izrazito su vrijedni uvidi za komercijalni sektor, budući da oni mogu, u velikoj mjeri, utjecati na krajnji proizvod, te time osigurati konkurentnost organizacije. Navedeni uvidi i obrasci, na temelju kojih organizacije mogu donositi odluke kako bi osigurale postizanje vrijednosti, opravdavaju zašto komercijalni sektor ulaže golemu financijska sredstva u istraživanje velikih podataka.

Vlade također počinju uviđati važnost velikih podataka, a primjer toga jest „Big Data Research and Development Initiative“ (Inicijativa za istraživanje i razvoj velikih

podataka) Sjedinjenih Američkih Država u koju je uloženo više od 200 milijuna američkih dolara. (Alotaibi i Abdullah, 2016: 687 – 696, prema Li i Dou, 2017: 4). Cilj te inicijative jest unaprijediti tehnologije za rudarenje, prikupljanje, obradu, analizu i primjenu velikih podataka u svrhu ubrzanog otkrivanja informacija i znanja na području prirodnih znanosti, unapređenja državne sigurnosti, te općenito, u korist stanovništva (Weiss i Zgorski, 2012., prema Li i Dou, 2017: 4). Njihova vrijednost očituje se u sposobnosti da prikupljaju masovne količine podataka o svojim državljanima, koje sada mogu učinkovito i jeftino pohranjivati i dati na korištenje raznim institucijama u svrhu boljitka cjelokupne zajednice.

Primjena velikih podataka u zdravstvu također ima zanimljive implikacije o kojima već postoje brojna istraživanja. Razvojem senzorne tehnologije pacijentima postaje moguće svakodnevno motriti svoje tjelesne funkcije i na temelju tih podataka liječnici mogu donositi informirane odluke i predviđanja o njihovom zdravstvenom stanju (Joiner, 2018, 101). Učestalijom primjenom velikih podataka zdravstvo će sve više biti usmjereno na prevenciju.

U obrazovanju se veliki podaci mogu koristiti, primjerice, u svrhu prepoznavanja učenika koji ne postižu zadovoljavajuću uspjeh (Joiner, 2018, 101). Škole bi onda mogle takve učenike smjestiti u odgovarajuće obrazovne programe, a prosvjetnim radnicima bi mogle osigurati odgovarajuće resurse kako bi se postigli traženi ishodi (Joiner, 2018, 101). Također, obradom velikih podataka nastoji se identificirati i predvidjeti kako se učenike može najbolje motivirati za učenje (Joiner, 2018, 101). Iako su primjena i istraživanje velikih podataka od strane obrazovnih institucija rijetki, unapređenje obrazovnog procesa i procesa istraživanja nameću se kao važni ciljevi primjene velikih podataka u knjižnicama.

Navedena su samo neka područja i institucije koje su uvidjele važnost velikih podataka i sukladno tome počele ulagati resurse u istraživanje koncepta i razvoj tehnologije potrebne za primjenu tih podataka. Gdje se god stvaraju podaci u velikim količinama, čijim prikupljanjem i obradom je moguće unaprijediti poslovanje, saznanja, proizvod ili uslugu, ondje će primjena tehnologije velikih podataka biti ključna (Ball, 2019: 2).



### 3.1 Odjek velikih podatak u informacijskim znanostima i knjižnicama

Jedan od glavnih problema s kojima su se informacijske znanosti suočavale jest učiniti znanje pristupačnim. U eri velikih podataka ovaj izazov postaje teži ne samo zbog veće količine znanja u svijetu, već i zbog činjenice da podaci i informacije na kojima je to znanje osnovano često nemaju točno određen format (Oliveira i Cendón, 2020: 155). Novi informacijski kontekst, u kojem su prisutni veliki podaci, predstavlja zahtjev za promjenama unutar informacijskih znanosti (Oliveira i Cendón, 2020: 156). Ova disciplina mora moći odrediti kako će informacijski stručnjak djelovati u novom okruženju, kako će se unaprijediti postojeće informacijske usluge i koje vještine i kompetencije pojedinac, koji radi s velikom podacima, treba imati (Ribeiro, 2014: 96 – 105, prema Oliveira i Cendón, 2020: 156). Neke od glavnih rasprava unutar informacijskih znanosti jesu one vezane uz pristup i upravljanje informacijama i podacima, te pronalazak i odabir informacija – rasprave, koje se pojavom velikih podataka produbljuju. U okruženju velikih podataka potencijalno znanje informacijskih stručnjaka nameće se kao glavni izvor podrške u upravljanju, istraživanju, očuvanju i pohranjivanju sve većih količina podataka, a da bi za to bio osposobljen potrebna su istraživanja o konceptu velikih podataka na temelju kojih će se obrazovanje informacijskih stručnjaka prilagoditi informacijskom okruženju. (Oliveira i Cendón, 2020: 156). Iako je potreba za istraživanjima evidentna, prema istraživanju provedenom u „Big data in Information Science literature“, u kojem se ispitala zastupljenost koncepta velikih podataka u znanstvenim publikacijama baza podataka „LISA“ i „LISTA“, ukazuje da je stvarnost drugačija (Oliveira i Cendón, 2020: 170). Istraživanje je pokazalo da je zastupljenost koncepta velikih podataka (kao glavnog koncepta) u znanstvenim radovima relativno niska –11% od ukupno ispitanih radova se u većoj mjeri tiče tematike velikih podataka, a tek mali dio tih članaka potječe iz discipline informacijskih znanosti i LIS-a (Oliveira i Cendón, 2020: 170).

Bez obzira na mali broj provedenih istraživanja o velikim podacima unutar područja bibliotekarstva, knjižnice će imati važnu ulogu u pohrani dokumenata i upravljanju podataka, a utjecaj velikih podataka na knjižnice je neizbježan (Li i Dou, 2017: 5). Razvojem Interneta i informacijske tehnologije mijenjaju se načini kako pojedinci dolaze do informacija. Mnogi čitatelji preferiraju čitati e-knjige i ne trebaju posjećivati knjižnice u potrazi za relevantnim informacijama, već to čine

pretraživanjem interneta (Zhou, 2016: 66 – 73, prema Li i Dou, 2017: 7). Postoji stav da se Internet pokazao kao glavna konkurencija tradicionalnim knjižnicama u području organizacije znanja, a dokaz tome jest da su se mnoge knjižnice restrukturirale ili zatvorile (Li i Dou, 2017: 7). Prema tome, nije čudno da se u bibliotekarskim krugovima ozbiljno razmatra primjena velikih podataka u svrhu učvršćivanja pozicije knjižnica kao centara za organizaciju znanja (Li i Dou, 2017: 7). Tehnologija velikih podataka omogućuje knjižnicama da analiziraju i rudare velike skupove podataka u svrhu prepoznavanja potreba čitatelja, te da poduzmu mjere u skladu s tim potrebama, odnosno da čitatelja zadrže (Provost i Fawcett, 2013: 51 – 59, Narendra, 2016: 83 – 93, prema Li i Dou, 2017: 8). Promjene informacijskih potreba korisnika predstavljaju neizravan utjecaj velikih podataka na knjižnice – oni sve više koriste velike podatke u istraživanjima, a zadovoljiti složene potrebe postaje sve teže (Ball, 2019: 3). Moći prepoznati kako velike skupove podataka učiniti korisnim, vidljivim i pristupačnim stvaranjem taksonomija i metapodataka jedan je od zahtjeva koje era velikih podataka stavlja pred knjižnice, i pokazuje se kao imperativ za unaprjeđenje informacijskih usluga koje ona nudi (Jharotia, 2016: 5). Popularnost istraživanja na temelju obrade velike količine podataka također je povezana s nastojanjima knjižnica da unaprijede informacijske usluge, budući da su one, itekako uključene u životni ciklus podataka (Krier i Strasser, 2014: 4, prema Wittmann i Reinhalter, 2014: 9). Knjižnice su centri za pohranu i očuvanje podataka nastalih iz studentskih i fakultativnih istraživačkih projekata, a učinkovitim upravljanjem svojih repozitorija podataka one postaju izvor podrške takvim vrstama projekata i istraživanjima općenito (Krier i Strasser, 2014: 10, prema Wittmann i Reinhalter, 2014: 9). Da bi knjižnice zadovoljile potrebe korisnika i istraživanja one moraju prikupljati i analizirati podatke, te ponuditi prikladne alate za otkrivanje znanja što primjena tehnologije velikih podataka omogućuje (Wang, 2016: 664).

Iako su mogućnosti i zahtjevi primjene tehnologije velikih podataka u knjižnicama očigledni, mnogi misle da su trenutni sustavi za upravljanje bazama podataka dovoljni za pohranu i obradu knjižničnih podataka (Wang, 2016: 664). Međutim, u obzir treba uzeti rastuću količinu nestrukturiranih podataka kojima se metode i tehnologije prisutne u knjižnicama, prema definiciji velikih podataka, s vremenom neće moći nositi. Da bi knjižnice opstale, odnosno ostale konkurentne u novom informacijskom okruženju, knjižničari moraju biti sposobni analizirati i prikazati podatke u svrhu

stvaranja znanja, što će ponajprije biti moguće poznavanjem tehnologije velikih podataka. (Jaiswal, 2020: 6). Primjena te tehnologije nije lak pothvat te zahtjeva stručnost knjižničara, čije će vještine u budućnosti sve više nalikovati onima informacijskih i podatkovnih stručnjaka, odnosno znanstvenika (Oliveira i Cendón, 2020: 155).

Primjenu velikih podataka u knjižnici se može shvatiti i kao „dvosjekli mač“ (Kim i Choi, 2016: 303 – 308, prema Li i Dou, 2017: 8). S jedne strane, primjenom tehnologije velikih podataka knjižnice mogu unaprijediti informacijske usluge koje nude u svrhu zadovoljavanja informacijskih potreba korisnika, korisnike zadržati i biti podrška u istraživačkim projektima dok s druge strane trebaju biti svjesne da primjena te tehnologije pred njih postavlja izazove promjene infrastrukture i prilagodbe vještina i kompetencija knjižničara (Li i Dou, 2017: 8). Međutim, u vremenu brze digitalizacije, razvoja Interneta i računalne tehnologije prednosti koje primjena velikih podataka nudi prevelike su da bi ih se zanemarilo. Nije riječ samo o tome da, knjižnica kao informacijska institucija opstane i ostane relevantna, već da iskoristi sve što joj se nudi u svrhu otkrivanja znanja i zadovoljavanja informacijskih potreba svojih korisnika.

## 4 Knjižnice i podaci

Jedna od svrha ovog rada jest prikazati na koje se sve načine veliki podaci mogu integrirati u poslovanje i rad knjižnice. No, u svrhu boljeg razumijevanja primjena, potrebno je odrediti što su knjižnice dosad radile s podacima i kakvi su to podaci, do koje mjere se ti podaci mogu smatrati velikim podacima, koje uloge povezane s podacima knjižnice imaju, te zašto knjižničari imaju predispozicije za rad s velikim podacima.

### 4.1 Veliki podaci knjižnica

Knjižnične zbirke općenito sadrže knjige, e-knjige, online časopise, istraživačke i seminarske radove, te knjižnice sve češće izgrađuju repozitorije (Wang i sur. 2016: 3). Ako se u obzir uzmu statične knjižnične zbirke teško je da će one odgovarati konceptu velikih podataka, budući da se većina knjižnica, odnosno alata koje one koriste mogu nositi s vlastitim zbirkama (Wang i sur. 2016: 2). Međutim, knjižnice prikupljaju velike količine „malih istraživačkih podataka“ koje stvaraju individualni istraživači (Wang i sur. 2016: 2). Ako se te podatke promatra kao cjelinu, oni se približavaju volumenu karakterističnom velikim podacima (Wang i sur. 2016: 2). Iako se čini da su podaci sadržani u knjigama, časopisima i sličnoj građi, te podaci o toj građi dobro organizirani, to nije slučaj s istraživačkim podacima pohranjenim u knjižnicama (Salo, 2010, prema Wang i sur. 2016: 3). Takvi podaci su raspršeni, njih se slabije opisuje, te su formata neprikladnih za dugoročno korištenje, što je posljedica subjektivnih praksa istraživača (Wang i sur. 2016: 3). Podacima istraživanja često nedostaju i standardi, budući da istraživači nerijetko sami određuju koji će parametri biti važni za projekt (istraživanja su individualizirana), te koriste vlastite formate za podatke koje prikupljaju (Wang i sur. 2016: 3). Raznolikost formata i nedostatak standarda predstavlja poteškoću prilikom integracije podataka (Vogel, 2017: 168). S obzirom na dimenziju raznolikosti velikih podataka, istraživački podaci koje pohranjuje knjižnica približavaju se definiciji koncepta velikih podataka i prema tome za njihovu obradu potrebni su prikladni alati i tehnologija. Važno je spomenuti da je posljedica popularnosti istraživanja konstantno generiranje ove vrste podataka, a prema tome potrebe za brзом obradom rastu. Međutim, istraživački podaci nisu jedini podaci koje se može povezati s dimenzijom raznolikosti. Knjižnične često posjeduje

različite tipove podataka kao što su mape, filmovi, slike, audio zapisi i sl. (Wang i sur. 2016: 2). Jedan dio tih podataka je nestrukturiran (Wang i sur. 2016: 2). Iz knjižničnih zbirki se mogu stvoriti sheme podataka tako da se u odnose dovede autore, citiranja, geo-lokacije, predmetne klasifikacije, izdavače i sl. – odnose, čijim se daljnjim povezivanjem s drugim radovima, ljudima, događajima mogu stvoriti grafovi (Wang i sur. 2016: 2). Stvaranje, obrada i davanje na korištenje takvih grafova podrazumijeva tehnike svojstvene velikim podacima, odnosno može ih se nazvati velikim podacima (Teens i Goldner, 2013: 429 – 438, prema Wang i sur. 2016: 2). Knjižnice također, sve češće prikupljaju podatke o korisnicima koje oni stvaraju prilikom korištenja informacijskih usluga knjižnica (Wang, 2016: 665). Pomoću podataka o navikama i preferencijama korisnika knjižnica ih može bolje razumjeti, te svoje usluge prilagoditi njihovim potrebama.

Općenito govoreći, što su skupovi podataka koje knjižnica posjeduje veći, pronalazak vrijednosnih obrazaca analizom tih podataka bit će lakši, pod uvjetom, da je knjižnica spremna koristiti alate i tehnologije koje nude veliki podaci. Primjena tehnologije velikih podataka nameće se kao imperativ ako se u obzir uzmu i raznolikost i brzina stvaranja podataka u knjižnicama.

## 4.2 Uloga knjižnice u eri velikih podataka

Knjižnice prikupljaju, obrađuju i osiguravaju pristup informacijskim resursima u svrhu pronalaženja novih znanja. Ako se podaci mogu smatrati jednom vrstom dokumenata, proizlazi da knjižnica mora prikupljati i obrađivati podatke, te mora nastojati osigurati pristup skupovima podataka koje posjeduje. Uloge knjižnice određuje nastojanje da, u svrhu otkrivanja znanja i zadovoljavanja informacijskih potreba, korisniku pomoću dostupnih alata, metoda i tehnologija, ponudi kvalitetne informacijske usluge u danoj informacijskoj okolini. Zbog porasta popularnosti istraživanja temeljenih na analizi velike količine podataka pokazalo se da knjižnice imaju izvanrednu priliku i sposobnosti biti savjetodavno tijelo i podrška takvim istraživanjima (Al-Barashdi i Al-Karousi, 2018: 2). Znanstvenim disciplinama, vladama, sveučilištima i istraživačkim institucijama knjižnice mogu ponuditi: alate za obradu, analizu i prikupljanje velikih količina podataka i pomoći im pri odabiru prikladnih alata, savjete vezane uz korištenje platforma za upravljanje podacima, savjete vezane uz odabir metodologija istraživanja, savjete uz istraživački proces

općenito, te stručno obrađene i odabrane podatke koji su spremni za daljnju obradu i korištenje (Al-Barashdi i Al-Karousi, 2018: 11). Knjižnice će takve podatke osigurati ako ozbiljno shvate zadatke skrbi o podacima i dijeljenja podataka.

#### 4.2.1 Skrb o podacima

Knjižnična građa prolazi kroz procese analize sadržaja, pohrane i vrednovanja. Budući da podaci jesu jedna vrsta resursa, njih je potrebno učinkovito organizirati, odnosno potrebno je njima upravljati na način koji će omogućiti otkrivanje i inovaciju znanja (Wang, 2016: 665). Prema tome, razložno je zahtijevati da podaci prolaze kroz iste procese kroz koje prolazi ostala knjižnična građa (Wang, 2016: 665). Skrb o podacima (engl. *data curation*) jest vrsta upravljanja koja ima funkcije pohrane, nadzora, odabira, održavanja i čuvanja podataka i podrazumijeva procese potrebne za stvaranje kontroliranih podataka i dodavanje vrijednosti tim podacima (Zipei, 2012: 161, prema Wang, 2016: 665).

Skrb o podacima tiče se ekstrakcije i transformacije podataka (Tiwari, 2018: 2). Taj proces predstavlja aktivno i neprestano upravljanje podacima kroz njihov životni ciklus, uzimajući u obzir korist podataka i interese, koji se njihovom analizom žele postići. On omogućuje otkrivanje podataka kao i njihovo pretraživanje, održavanje njihove kvalitete, dodavanje vrijednosti, te uvelike utječe na buduće korištenje podataka (Tiwari, 2018: 3). Podaci ne opisuju sebe i prema tome, važna funkcija skrbi je opisivanje podataka (Tiwari, 2018: 8). Metapodaci su podaci kojima se opisuje podatke u svrhu otkrivanja resursa i omogućavanja pristupa (Tiwari, 2018: 8). Pomnim opisivanjem, velike skupove podataka moguće je pretraživati, a stvaranje kvalitetnih metapodataka nameće se kao nužnost, kako bi se izbjegao gubitak važnih informacija sadržanih u tim velikim količinama podataka, što u kontekstu konstantno rastućih količina postaje opasnost. Kvalitetni opisi jednako tako podupiru ponovno korištenje podataka, koje odgovara bitnom segmentu dolaska do vrijednosti pomoću njih. Prema Tiwariu, procesi skrbi o podacima uključuju: opisivanje i prikaz informacija (metapodaci se dodjeljuju podacima), planiranje očuvanja, konceptualizaciju (određuje se plan za stvaranje podataka), stvaranje i zaprimanje (stvaranje podataka s odgovarajućim metapodacima i zaprimanje podataka iz drugih izvora), evaluaciju i odabir (odabiru se podaci za dugoročnu skrb i očuvanje), pohranu podataka (podaci se prenose u arhive ili repozitorije, te se pohranjuju u potrebnim formatima), te objavljivanje podataka, odnosno osiguravanje javnog pristupa

podacima (Tiwari, 2018: 6). Skrb o podacima također uključuje povremene procese, koji se obavljaju u vremenu između glavnih procesa (Tiwari, 2018: 6). U povremene procese spadaju: odstranjivanje podataka koji nisu odabrani za dugoročno čuvanje (primjerice, zbog zakonskih razloga) i odstranjivanje podataka koji nisu zadovoljili proces evaluacije (Tiwari, 2018: 6).

Podaci koji su provedeni kroz procese skrbi knjižnicama mogu biti izvor informacija za samo-vrednovanje, no ako se u obzir uzme uloga podrške istraživanjima i kvaliteta podataka na kojima će ta istraživanja biti temeljena, očito je da će kvaliteta rezultata istraživanja ovisiti o kvaliteti podataka, koju procesi skrbi trebaju osigurati. Također, skrb o podacima predstavlja preduvjet za stvaranje kvalitetnih repozitorija podataka i uspješno dijeljenje podataka. Repozitoriji podataka jesu jedna vrsta kontrole podataka koji dopijevaju u knjižnicu, a njihovu kvalitetu knjižnica može unaprjeđivati stvaranjem standarda podataka i nametanjem određenih praksi istraživačkom procesu (Al-Barashdi i Al-Karousi, 2018: 10). Time knjižnice osiguravaju da podaci istraživačkih radova bude prikladni za buduće korištenje i omogućuje se učinkovito dijeljenje podataka.

Kao najpopularniji pristupi skrbi o podacima pokazali su se: „Sheer Curation“, odnosno integracija skrbničkih aktivnosti i izvora podataka, te „Colaboration Spaces“, odnosno kolaborativna skrb strukturiranih i nestrukturiranih podataka (Tiwari, 2018: 6). Popularni alati za skrb o podacima uključuju „OpenRefine“, „Tableau“ i „Data-Driven Documents“ (Tiwari, 2018: 6).

#### 4.2.2 Dijeljenje podataka

Porast količine podataka i brzine kojom se stvaraju otežalo je istraživački proces. Iako istraživači nikada prije na raspolaganju nisu imali tolike količine podataka, oni se često susreću s poteškoćama vezanim uz pristup tim podacima. Jednako tako, podaci na kojima će se neko istraživanje temeljiti često su nestrukturirani, odnosno u formatima neprikladnim za obradu, te postoji rizik da će brzo postati zastarjeli, odnosno da neće biti aktualni. Ovu su neki problemi za čije je rješavanje potrebna visoka razina stručnosti obrade podataka prisutna u knjižnicama. Repozitoriji su sustavi za prikupljanje, objavljivanje, diseminaciju i arhiviranje digitalnog sadržaja i predstavljaju fizičku infrastrukturu koja istraživačima omogućuje dijeljenje i ponovno korištenje rezultata vlastitih i tuđih radova (Jeffrey, 2013, prema Wusu i sur. 2020: 4).

S obzirom na to da podaci, koje neki repozitorij sadrži prolaze kroz procese skrbi i apriorne standardizacije, njihova kvaliteta je zajamčena. Osim što omogućuju dijeljenje aktualnih podataka prikladnih za trenutno i buduće korištenje, repozitoriji kultiviraju otvorenost znanstvene i akademske komunikacije, čime se osigurava daljnja diseminacija znanja (Cullen i Chawner, 2011: 460 – 470, prema Wusu i sur. 2020: 4). Obilježje ere velikih podataka jest raspršenost podataka različitih formata, koji se stvaraju različitim brzinama i prema tome, važno je skupove podataka povezivati u smislene strukturalne cjeline (Al-Barashdi i Al-Karousi, 2018: 10). U skladu s time, mnoge knjižnice suradnjom nastoje stvoriti zajedničke repozitorije, međutim to nije dovoljno (Al-Barashdi i Al-Karousi, 2018: 10). Skupove podataka koje knjižnica posjeduje i daje na korištenje potrebno je povezati s onima koji se nalaze na webu, što predstavlja priliku za primjenu tehnologije velikih podataka u svrhu diseminacije i otkrivanja znanja. Povezivanje podataka nije lak pothvat, budući da, podaci koje se nastoji povezati imaju različite formate i podrazumijevaju različite sintakse i semantike (Vogel, 2017: 168). Oni većinom nisu dovoljno strukturirani, odnosno metapodaci ne otkrivaju dovoljno informacija o sintaksi i semantici (Vogel, 2017: 168). Tehnologijom povezanih podataka (engl. *linked data technology*) nastoji se formalnije kodirati podatke i metapodatke u svrhu povećanja fleksibilnosti i ekspresivnosti podataka (Vogel, 2017: 168). Princip rada ove tehnologije jest okvir za opisivanje resursa („Resource Description Framework“), tj. RDF format koji koristi iskaze za opisivanje dijelova informacija (Vogel, 2017: 168). Iskaz povezuje jedan podatak (subjekt) s drugim (objekt) posredstvom nekog svojstva, odnosno radi se tripartitnom obliku odnosa podataka – subjekt, predikat i objekt (Vogel, 2017: 168). Navedenim načinom opisivanja i standardizacije podaci se transformiraju u oblik koji računalo i čovjek mogu razumjeti (iskazi se mogu interpretirati), čime se znatno povećava razina automatizacije, odnosno integracija podataka postaje lakša (Vogel, 2017: 168).

Dakle, ključna je uloga knjižnice, u eri velikih podataka, da djeluje kao savjetodavno tijelo i bude podrška istraživačima koji se susreću s poteškoćama prikupljanja, obrade i analize velikih količina raznolikih podataka. Tu funkciju knjižnica može uspješno obavljati ako u svoje poslovanje uključi učinkovite mehanizme skrbi o podacima i osmisli inovativne načine dijeljenja podataka. Iako osiguravanje pristupa kvalitetnim



informativnim resursima knjižnicama nije strano, ono se mora prilagoditi zahtjevima vremena i proširiti na dimenziju velikih podataka.

### 4.3 Predispozicije za rad s velikim podacima

Knjižnice i knjižničari imaju predispozicije za rad s velikim podacima, budući da slove kao vrsni korisnici novo-razvijenih tehnologija, a tehnologija velikih podataka ne smije biti iznimka (Huwe, 2014: 17 – 18, prema Hoy, 2014: 7). Zajedno s informativnim stručnjacima, knjižničari su tradicionalno bili predvodnici pristupa informacijama, savjetnici istraživačima i zagovornici privatnosti (Gordon-Murnane, 2012: 34, prema Wittmann i Reinhalter, 2014: 3). Oni su stručnjaci u pretraživanju, odabiru i analizi podataka (Joiner, 2018, 101). S obzirom na to da se knjižnice svakodnevno bave razvojem svojih zbirki, očuvanjem građe i savjetovanjem korisnika, u kontekstu velikih podataka ove je zadatke potrebno konceptualno proširiti na način koji će obuhvaćati podatke. Knjižničari imaju uloge u prikupljanju, razvoju i preobrazbi statističkih jedinica i podataka o korištenju, upravljaju informacijama istraživanja, te u sveučilišne programe uključuju podatkovnu pismenost (Al-Barashdi i Al-Karousi, 2018: 11). Velike količine podataka knjižničari moraju znati transformirati u informacije i znanje, a da bi to učinili moraju razumjeti kako podatke mogu analizirati i prikazati u svrhu stvaranja znanja. (Wang i sur. 2016: 1). Moraju znati kako velike skupove podataka, pomoću odgovarajućih alata mogu učiniti korisnim, vidljivim i pristupačnim stvaranjem taksonomija, shema metapodataka i sustavnih metoda pretraživanja (Joiner, 2018, 102). Također, trebaju proučiti koje bi skupove velikih podataka mogle prikupljati i analizirati, koristeći odgovarajuće alate. Skrb o podacima, upravljanje, objavljivanje, vizualizacija, održavanje i očuvanje podataka, učenje podatkovnoj pismenosti, rudarenje podataka i razvoj infrastruktura za digitalne podatke su područja stručnosti knjižničara kojima oni mogu značajno doprinijeti i po mnogima predstavljaju prirodnu ekstenziju misije zaštite i diseminacije informacija (Oliphant, 2017: 2). No, rad s velikim podacima zahtjeva znanje o statističkim i analitičkim metodama, jednako kao i znanje o konceptu velikih podataka. Posjedovati predispozicije nije isto kao znati koju statističku metodu ili konkretni alat primijeniti u danoj situaciji da bi se postigao željeni cilj. Prema tome, iako knjižničari posjeduju predispozicije za rad s velikim podacima, njima je potrebno kvalitetno obrazovanje i usavršavanje kako bi bili osposobljeni za taj rad.

## 5 Primjena velikih podataka u knjižnicama

Primjena tehnologije velikih podataka može se podijeliti na tri skupine. Prva se skupina odnosi na primjenu velikih podataka u svrhu stvaranja personaliziranih usluga i prilagođavanja korisničkim potrebama. Druga skupina usmjerena je na samu knjižnicu i primjenjuje se radi donošenja odluka na osnovi podataka. Treća skupina odnosi se na usluge koje knjižnica treba pružati i djelatnosti koje treba obavljati ako se podaci shvate kao resursi koje korisnicima daje na korištenje, te je usko povezana s ulogom podrške i savjetovanja naglašenom u ovom radu. Za svaku skupinu rad nastoji ponuditi konkretni primjer knjižnica, koje su u svojem poslovanju primijenile tehnologiju velikih podataka. Rudarenje podataka i vizualizacija podataka su naročito zastupljene tehnike u primjeni velikih podataka zbog čega ih rad ukratko opisuje.

### 5.1 Rudarenje podataka

Rudarenje podataka jest tehnika analize velikih informacijskih repozitorija u svrhu otkrivanja skrivenih i potencijalno vrijednih informacija (Al-Barashdi i Al-Karousi, 2018: 5). Pomoću ove tehnike, analizom ogromnih količina podataka moguće je otkriti skrivene odnose i nepoznate obrasce (Al-Barashdi i Al-Karousi, 2018: 5). To čini identificiranjem anomalija i stvaranjem grupa (engl. *cluster*) podataka (Tableau Software, 2018) Otkrivanjem tih obrazaca i odnosa knjižnice dolaze do vrijednih uvida, koji su temelj za pružanje informacijskih usluga koje će zadovoljiti dinamične informacijske potrebe korisnika, te za povećanje učinkovitosti djelatnosti knjižnica. Rudarenje podataka u sveučilišnim knjižnicama naziva se „biblio-rudarenje“ (engl. *bibliomining*) i koristi se za: praćenje bihevioralnih promjena i trendova u sustavima knjižnica, praćenje performansa zaposlenika, otkrivanje područja nedostataka, predviđanje budućih potreba (Siguenza-Guzman, 2015: 499 – 510, prema Al-Barashdi i Al-Karousi, 2018: 5).

### 5.2 Vizualizacija podataka

Veliki podaci omogućuju čovjeku da donosi zaključke na temelju obrade golemih količina podataka. Iako alati za obradu i analizu podataka obavljaju dio posla, čovjek je, u konačnici, taj koji interpretira podatke. Prema tome, rezultati analiza, i otkrivene korelacije, te podaci općenito trebaju biti prikazani na prikladan način. Procesom

vizualizacije se atributi podataka dovode u vezu s vizualnim svojstvima kao što su pozicija, veličina, oblik i boje što omogućuje spoznavanje na osjetilnoj razini (Minelli, Chambers i Dhiraj, 2013: 113).

Funkcije tradicionalnih tehnika vizualizacije podataka uključuju opisivanje i izvještavanje, međutim, u eri velikih podataka, one uključuju promatranje i otkrivanje, odnosno proširene su na dimenzije vremena i interaktivnosti (Minelli, Chambers i Dhiraj, 2013: 110). Promatranjem se može shvatiti značaj obrazaca koji se pojavljuju kroz određen period vremena, omogućuje se praćenje razvoja varijabli kroz vrijeme, a interakcijom je čovjeku omogućeno da intuitivno shvati odnose podataka i konteksta u kojem su ti podaci nastali (Minelli, Chambers i Dhiraj, 2013: 111). Prikladnom vizualizacijom podaci dobivaju svoju „priču“ (Minelli, Chambers i Dhiraj, 2013: 111). Međutim, vrijeme nije jedina dimenzija u kojoj je moguće prikazati podatke (Minelli, Chambers i Dhiraj, 2013: 113). Podaci se na vizualno prikladan način mogu prikazati u kontekstu prostora, relativnosti, blizine, prioriteta i ograničenja, a vrijednost metoda i alata za vizualizaciju velikih podataka očituje se u mogućnosti kombinacije ovih konteksta, odnosno dimenzija – radi se o višedimenzionalnoj „priči“ (Minelli, Chambers i Dhiraj, 2013: 113).

S obzirom na kontekst velikih podataka, slika čovjeka, koji tablicu Microsoft Excela popunjava sa stotinjak jedinica podataka da bi dobio jednostavan graf, postaje stvar prošlosti (Simon, 2013: 103). Danas su čitave tvrtke specijalizirane za razvoj alata koji će korisnicima pomoći shvatiti svoje podatke, najčešće kroz kreativne i interaktivne načine prikazivanja podataka (Simon, 2013: 103). Neki od najpoznatijih projekata i alata uključuju: „Tableau“, „Data Driven Documents java script library“, „Qlikview“ i „Microstrategy“ (Minelli, Chambers i Dhiraj, 2013: 118).

### 5.3 Personalizirane informacijske usluge

Informacijske potrebe korisnika su složene, heterogene, mijenjaju se ili čak sukobljavaju (Huang i Zhong, 2013: 37 – 43, prema Wang, 2016: 664). Zajednice i grupe korisnika su složene strukture koje knjižnica mora razumjeti, odnosno mora poznavati njihovo ponašanje, navike, interese i preferencije kako bi im bila u stanju ponuditi odgovarajuću informacijsku uslugu i time osigurati bolje korisničko iskustvo (Ball, 2019: 7). Zadovoljiti potrebe korisnika postaje sve veći izazov.

Međutim, u eri velikih podataka ljudsko ponašanje često ostavlja „podatkovni trag“ u informacijskim sustavima (Wang, 2016: 664). Podaci tih tragova se pohranjuju u različite sustave i pomoću tehnologije velikih podataka, automatikom se mogu povezivati kako bi ukazali na obrasce u ponašanju korisnika (Wang, 2016: 665). Korisnički podaci se analiziraju pomoću alata za rudarenje velikih podataka, a rezultati takvih analiza mogu ukazati na načine kako knjižnice mogu potencijalno poboljšati zadovoljstvo korisnika ili čak predvidjeti njihovo ponašanje. Primjenom tehnologije velikih podataka knjižnice imaju mogućnost korisnicima pružiti personalizirane i prilagođene informacijske usluge (Wang, 2016: 665). One se, primjerice, postižu na sljedeći način: primjenom „batch“ analize u tri koraka, odnosno funkcijom neovisnog filtra korisnički podaci se analiziraju kako bi se otkrila individualna pravila ponašanja, funkcijom kooperativnog filtra korisnički podaci se povezuju i analiziraju kako bi se dobio model ponašanja određenih vrsta korisnika, te se funkcijom kontekstualnog filtra analiziraju podaci kako bi se određeno ponašanje povezalos određenim kontekstualnim promjenama (Wang, 2016: 665). Konzistentnim povezivanjem ovih analiza moguće je na dosljedan način informacijske usluge prilagođavati potrebama korisnika. (Wang, 2016: 665). Da bi te analize bile moguće potrebno je da knjižnica na jednako dosljedan način osigura dotok korisničkih podataka, što može, primjerice, učiniti implementacijom jakih funkcija povratnog informiranja u svoje informacijske sustave (Wang, 2016: 665). U ovu svrhu, osmišljavanje načina na koje će se pratiti aktivnosti korisnika također se pokazuje izrazito važnim (Wang i sur. 2016: 5).

Jedan primjer pružanja personaliziranih usluga jest projekt u kojem je sudjelovalo deset knjižnica širom područja SAD-a, „Institute of Museum and Library Service (IMLS) i CIVICTechnologies“ (Ball, 2019: 7). Sudionici su zajedničkim naporima iskoristili mogućnosti tehnologije za analizu velikih podataka u svrhu preciznog definiranja korisničkih potreba (Ball, 2019: 7). Projekt je analizirao javne statistike i podatke na regionalnoj i nacionalnoj razini, te su rezultati tih analiza bili povezani s korisničkim podacima pojedinačnih knjižnica (Ball, 2019: 7). Projekt je rezultirao poboljšanjem iskustva korisnika time što su knjižnice korisnicima ponudile popularne (nove) usluge i programe (Ball, 2019: 7).

„Harvard University Library Analytics Toolkit“ još je jedan primjer ovakve vrste primjene velikih podataka u knjižničnom poslovanju (Joiner, 2018, 103). Razvojem se

ovog skupa alata knjižnicama i korisnicima omogućilo identificirati promjene u zbirkama, korištenju i drugim podacima što je omogućilo knjižničarima da shvate, analiziraju i vizualiziraju obrasce aktivnosti kao što su posudbe, povrati, nedavne akvizicije (Joiner, 2018, 103). Oni su to mogli činiti preko područja više knjižnica (Joiner, 2018, 103).

Primjena tehnologije velikih podataka u svrhu prilagođavanja korisniku i pružanja personaliziranih informacijskih usluga, knjižnici omogućuje da pravovremeno poduzme mjere kako bi se korisnici zadržali, te da shvati potencijalne navike korisnika koje bi mogla privući (Provost i Fawcett, 2013: 51 – 59, Narendra, 2016: 83 – 93, prema Li i Dou, 2017: 7). Također, ovakvom primjenom knjižnice mogu učvrstiti svoj odnos s obrazovnim institucijama, time što mogu pružati predskazujuće podatke povezane s trendovima u obrazovanju (Wittmann i Reinhalter, 2014: 10). Ako se uzme u obzir Internet s kojim su knjižnice sve više u konkurentnom odnosu, personalizirane informacijske usluge od ključne su važnosti za razvoj knjižnica (Sollenberger i sur., 2013: 38, Head i sur., 2014: 61, prema Li i Dou, 2017: 8).

#### 5.4 Samo-vrednovanje pomoću velikih podataka

U eri velikih podataka odlučivanje na osnovi iskustva postaje suvišno (Wang, 2016: 665). Iako ljudski faktor nikada u potpunosti ne može biti zamijenjen, tehnologija velikih podataka omogućuje donošenje odluka na osnovi velikih skupova obrađenih i analiziranih podataka, odnosno u odluke dodaje razinu znanstvenosti (Simon, 2013: 103). Prema tome, vrijednost velikih podataka leži u automatiziranoj i pametnoj potrazi za modelima, pravilima i karakteristikama podataka pomoću kojih se donose pravovremeni i precizni sudovi i odluke (Wang, 2016: 665). Dokazivanje vlastite vrijednosti svakodnevnica je mnogih knjižnica, a tu vrijednost one mogu povećati kvalitetnom organizacijom svojih djelatnosti, troškova i resursa. U pozadini takve organizacije leže odluke utemeljene na činjenicama. Primjenom tehnologije velikih podataka u ovu svrhu, knjižnice mogu postići drastično smanjenje troškova, pametnije alociranje ljudskih resursa, odnosno učinkovitije strukturirane djelatnosti i radne pozicije, te razvoj informacijskih usluga koje nude (Ball, 2019: 6).

Kako bi knjižnice, odnosno djelatnici knjižnice mogli donositi odluke utemeljene na učinkovitoj analizi podataka potrebno im je sljedeće: sposobnost brzog stapanja

strukturiranih, polu-strukturiranih i ne-strukturiranih podataka, snažna sposobnost obrade, odnosno sposobnost pronalaženja znanja i vrijednosti kombinirajući analitičku tehnologiju i metode (primjerice, mrežna analiza, rudarenje podataka, serijska analiza, vizualizacija i „Hadoop“ tehnologija), te sposobnost inovativnog razmišljanja, koja se odnosi na pronalaženje novih modela i informacija sadržanih u velikim podacima (Wang, 2016: 665). Ukratko, za donošenje pametnih odluka potrebni su podaci, tehnologija i kreativnost (Wang, 2016: 665). Vrijedi spomenuti, da se, ovakav način primjene tehnologije velikih podataka često koristi u svrhu dizajniranja baza podataka i unapređenja digitalnih zbirki knjižnica (Wang i sur. 2016: 5). Primjerice, knjižnica bi na osnovi posudbenih transakcija i korisničkog pretraživanja, pomoću rudarenja podataka i tekstualne analize, mogla optimizirati svoje zbirke tako da daju bolje rezultate pretraživanja i preporuke čime bi osiguralo učinkovitije korištenje knjižnične građe (Wang i sur. 2016: 5).

Brže donošenje odluka temeljenih na podacima bio je cilj projekta provedenog u „Brooklyn Public Library“, knjižnici koja se sve teže nosila s razinom raspršenosti svojih podataka (Ball, 2019: 6). Nakon zahtjevnog procesa povezivanja raspršenih podataka knjižnica se poslužila alatom „Tableau“, pomoću kojeg je te podatke uspješno vizualizirala (Tableau Software, 2018, prema Ball, 2019: 6). Postiglo se sljedeće: bolja alokacija ljudskih resursa i to na način da su jedan novi djelatnik i drugačije pozicionirani analitičar preuzeli djelatnosti koje je obavljala skupina od pet vanjskih savjetnika, vrijeme potrebno za proces mjesečnih izvještaja smanjilo se za dva tjedna, razriješili su se računovodstveni problemi vezani uz zakasnine, te je knjižnica počela sastavljati jasne i sveobuhvatne izvještaje koji su na vrijeme bili distribuirani svim djelatnicima (Ball, 2019: 6). Projekt je zaposlenicima omogućio autonomno pronalaženje mogućih smjerova djelovanja (O’Maley Voliva, 2015, prema Ball, 2019: 6).

Knjižnice svakodnevno trebaju dokazivati svoju vrijednost u svrhu ostvarivanja poticaja i suradnji. Da bi ta vrijednost bila veća one, s dostupnim resursima (ljudskim, tehnološkim i financijskim) trebaju ostvariti maksimalnu učinkovitost. Djelatnosti, troškove i upotrebu resursa trebaju organizirati na optimalan način što je moguće isključivo donošenjem pametnih odluka. U eri velikih podataka knjižnice se ne moraju oslanjati na iskustvo u donošenju takvih odluka, već na podatke i alate za njihovu analizu i obradu.

## 5.5 Podaci kao resurs

U eri velikih podataka uviđa se, više nego ikada prije, da su podaci iznimno vrijedan resurs. Međutim, količina, raspršenost i raznolikost tih podataka knjižnicama predstavlja izazov. Uz dosljedno prikupljanje podataka pomoću stvaranja kvalitetnih repozitorija, one korisnicima moraju omogućiti pristup podacima u obliku prikladnom za korištenje. Osim toga, trebaju velike skupove podataka učiniti vidljivim i korisnijim što one mogu postići primjenom tehnologije velikih podataka.

Porastom popularnosti istraživanja osnovanim na obradi velike količine podataka korisnici sve više zahtijevaju sveobuhvatne informacije, a pristup pojedinačnim jedinicama građe u svrhu pronalaska specifične informacije, u ovom kontekstu, postaje stvar prošlosti (Ball, 2019: 4). S obzirom na navedeno podatke je potrebno urediti na način koji će biti prikladan za korištenje, te ih je potrebno povezati.

Vrijednost tehnologije povezivanja podataka jest u njezinoj sposobnosti da nestrukturirane podatke i strukturirane podatke s različitim standardima transformira u strukturirane podatke s istim standardima uz pomoć RDF modela što uvjetuje da računalo može „razumjeti“ podatke (Wang, 2016: 665). Opisivanjem semantičkih odnosa između povezanih podataka, računalo je moguće prepoznati i nositi se s podacima automatizmom (Wang, 2016: 665). Kada se ovo obavi, koristeći tehnologiju povezanih podataka, podatke je moguće integrirati i ponuditi korisniku na vizualno prikladan način (Wang, 2016: 665). Primjerice, ako su potrebne informacije o nekretninama, korisniku se također ponude povezane informacije kao što su informacije o obrazovanju, prometu, financijama, sigurnosti i okolišu (Wang, 2016: 665).

„WorldCat“ jest jedan primjer nastojanja da se podaci povežu u svrhu osiguravanja pristupa. Predstavlja platformu koja omogućuje pretraživanje građe sadržane u mnogobrojnim knjižnicama diljem svijeta i zahvaljujući naporima knjižničara i drugih informacijskih stručnjaka raste svakim danom. U ovom smjeru krenula je i Sveučilišna knjižnica Harvard, koja je 2012. godine objavila metapodatke za dvanaest milijuna jedinica knjižnične građe (Ball, 2019: 4). Iako, zbog autorskih prava, nije mogla objaviti sve podatke, ovaj se projekt opisao kao „veliki podaci za knjige“ (Ball, 2019: 4).

Ako se podatke shvati kao resurse koje će knjižnica davati na korištenje, važno je da ona pruži kontinuiranu podršku u obliku savjetovanja istraživačima. Knjižničari, koji su prakticirali katalogiziranje i stvaranje metapodataka, imaju sposobnost konceptualizacije odnosa među podacima, i prema tome su prikladni za savjetovanje o upravljanju podacima. One također trebaju moći ponuditi prikladne alate korisnicima koji trebaju prikupiti, obrađivati i analizirati velike količine podataka. Podrškom tijekom cijelog istraživačkog procesa olakšava se prikupljanje, organizacija i očuvanje podataka koji će se generirati. Ovime se podupire i promovira dijeljenje podataka jednako kao što se vrši kontrola nad generiranim podacima. Knjižnice moraju pomoći korisnicima shvatiti što sve veliki podaci mogu, a što ne, te kako će najbolje iskoristiti podatke u svrhu postizanja istraživačkih ciljeva.

Sveučilišna knjižnica Michigan primjerice, nudi široki spektar usluga usmjerene na podatke kroz „Research Data Services“ i „Deep Blue Data services“ koje služe kao podrška istraživačima kroz cijeli ciklus istraživanja (Joiner, 2018, 104). „Research Data Services“ je mreža alata koji pokrivaju sljedeća područja: planovi za upravljanje podacima, otkrivanje i pristup, organizacija podataka, metapodaci i dokumentacija, dijeljenje podataka i objavljivanje, očuvanje podataka, vizualizacija podataka (Joiner, 2018, 104). Deep Blue Data services predstavlja ekstenziju sveučilišnog repozitorija i korisnicima nudi platformu specijaliziranu za skupove podataka pomoću koje podatke svojih istraživanja mogu dijeliti s kolegama diljem svijeta (Joiner, 2018, 104).

Još jedan primjer knjižnice koja ozbiljno shvaća ulogu podrške istraživanjima jest Knjižnica Sveučilišta Elmer Holmes Bobst u New Yorku (Joiner, 2018, 103). Ova knjižnica nudi: trening, podršku, stručno savjetovanje tijekom čitavog procesa istraživanja, alate koji podupiru kvantitativna, kvalitativna i geografska istraživanja, te usluge pristupa specijaliziranim paketima softwera za statističku analizu (Joiner, 2018, 103).



## 6 Problemi i izazovi

Ovaj dio rada nastoji prikazati glavne izazove koje je pojava velikih podataka postavila pred knjižnice, te probleme s kojima se knjižnice, koje žele primjenjivati tehnologiju velikih podataka, suočavaju. Era velikih podataka je pred knjižnicu postavila izazove u obliku ogromnih količina neuređenih podataka, koje knjižnica treba znati učinkovito obraditi ako želi ostati relevantna informacijska ustanova. Također, pojava velikih podataka ozbiljno narušava privatnost i sigurnost osobnih podataka zbog čega knjižnice trebaju postati oprezne i kritične u svakom segmentu rukovanja s podacima. Promijenjeni informacijski kontekst i sve složenije informacijske potrebe korisnika uvjetovali su da primjena tehnologije velikih podataka postane nužnost, a ne izbor. Međutim, Mentalitet zahtijevanja apsolutne točnosti i preciznosti u svakom segmentu poslovanja prva je prepreka primjeni tehnologije velikih podataka u knjižnicama. Knjižničari moraju shvatiti da ovakav mentalitet, u eri velikih podataka, opstruira zadovoljavanje informacijskih potreba korisnika. Nedostatak knjižničara s dubokim poznavanjem analitike velikih podataka onemogućuje donošenje odluka temeljenih na velikim količinama podataka. Knjižnicama također, često nedostaju knjižničari koji bi na adekvatan način mogli savjetovati korisnike prilikom istraživanja. Kao još jedan izazov pokazuje se nedostatak financijskih sredstava potrebnih za integraciju tehnologije velikih podataka.

### 6.1 Promjena mentaliteta

Iako se promjena mentaliteta pojedinih knjižnica i knjižničara na prvi pogled ne čini toliko važnom jer ne uključuje financijska sredstva ili promjene infrastrukture, ona je ključna za primjenu tehnologije velikih podataka u knjižnicama. Svaka ljudska djelatnost započinje željom, a neracionalnim odbijanjem primjene velikih podataka svaki sljedeći izazov ili problem postaje nebitan. Jedan razlog zašto neke knjižnice odbijaju primijeniti tehnologiju velikih podataka zasigurno proizlazi iz njihove tendencije da ostanu samodostatne organizacijske jedinice koje se drže podalje od novih tehnologija (Wang i sur. 2016: 4). Također, mnoge knjižnice svoje repozitorije drže podalje od javnosti jer knjižničari misle da tako povećavaju njihovu vrijednost (Wusu i sur. 2020: 4). No, odbijanje primjene velikih podatak ponajprije ima veze s

načinom zaključivanja na osnovi korelacija (Mayer-Schönberger i Cukier, 2013, 12). U nedostatku idealnog uzorka, koji bi na sveobuhvatan način pokrio svaku pojedinu jedinicu nekog skupa entiteta, porast količine entiteta dovodi do nepreciznosti u zaključivanju (Mayer-Schönberger i Cukier, 2013, 13). Zaključivanjem na osnovi korelacija izvedenih iz obrade velikih skupova podataka, razina točnosti možda opada na mikro-razini ali moguće je dobiti vrijedne uvide na makro-razini (Mayer-Schönberger i Cukier, 2013, 13). Iako u nekim djelatnostima knjižnice trebaju inzistirati na što većoj razini točnosti i usmjeriti se na kauzalne odnose izvedene iz analize uzorka, problem nastaje kad zbog takvog mentaliteta tvrdoglavo odbijaju uvidjeti prednosti koje bi primjena tehnologije velikih podataka mogla donijeti. Jedan primjer gdje inzistiranje za točnošću ide na štetu korisnika navodi Ball opisujući klasično katalogiziranje koje se oslanja na ekstremno precizno strukturirane i složene metapodatke i kao rezultat ima zahtjeve za jednakom preciznošću prilikom pretraživanja što korisniku može otežati zadovoljavanje informacijske potrebe (Ball, 2019: 2). Kao dokaz navodi činjenicu da se većina informacija i literature vezane uz prirodne znanosti, tehnologiju i medicinu pretražuje na Google-u, koji je u svoje usluge odavno integrirao tehnologiju velikih podataka, te zaključuje da su se indeksna kultura knjižnica i navike pretraživanja korisnika razvile u suprotnim smjerovima (Ball, 2019: 2). Primijeniti tehnologiju velikih podataka ne znači u potpunosti odbaciti zahtjeve za točnošću, već samo odustati od predanosti prema točnošću (Mayer-Schönberger i Cukier, 2013, 13). S mentalitetom koji zahtijeva apsolutnu točnost u svakoj djelatnosti mnogim knjižnicama nemoguće je uopće zamisliti što bi s primjenom velikih podataka mogle postići, a promjena ovog mentaliteta prvi je, i ključni korak k primjeni (Ball, 2019: 3).

## 6.2 Znanje i vještine potrebne za rad s velikim podacima

Nedostatak djelatnika koji su stručno osposobljeni za rad s velikim količinama podataka onemogućuje knjižnicama kontinuirano prilagođavanje sve složenijim informacijskim potrebama korisnika, te znatno otežava funkcije podrške i savjetovanja u istraživačkom procesu (Jaiswal, 2020: 6). Također, nedostatkom potrebnih vještina dovodi se u pitanje kvaliteta podataka koji se daju na korištenje. U nastavku rada navode se vještine, sposobnosti i znanja koje knjižničari, u eri velikih podataka trebaju posjedovati.

Postepenom pojavom velikih podataka i promjenom informacijskom konteksta društvo je uvidjelo vrijednosti odluka temeljenih na velikim količinama podataka. Međutim, podaci koji ne prođu kroz proces analize nemaju vrijednosti (Kirkwood, 2016: 275 – 284, prema Jaiswal, 2020: 2). Prema tome, u središtu novog informacijskog konteksta nalazi se analitika velikih podataka. Unutarnji procesi i algoritmi ovakve analize izrazito su složeni, a posljedica toga jest da pronalaženje skrivenih obrazaca i veza nalikuje pronalaženju „igle u plastu sijena“ (Al-Barashdi i Al-Karousi, 2018: 13). Ako knjižnica želi primijeniti tehnologiju velikih podataka u poslovanju, složenost analitike velikih podataka predstavlja izazov koji se očituje u nedostatku knjižničara koji posjeduju prikladne statističke vještine, poznavanje domena znanja i osnova računalne znanosti (Wang i sur. 2016: 4). Poznavanje domena znanja pokazuje se bitnim preduvjetom za izvršavanje procesa skrbi o podacima, budući da podaci ovise o domeni jednako kao i modeli skrbi (Tiwari, 2018: 4). Analize svih vrsta podataka, iako imaju potencijalnu vrijednost, nisu učinkovite ako podaci prije nisu prošli proces skrbi (Tiwari, 2018: 4). Da bi knjižničari bili upoznati s mogućnostima inherentnim velikim skupovima podataka, moraju posjedovati statističke i analitičke vještine, te biti sposobni provoditi procese skrbi, a nedostatak takvih vještina izravno će uvjetovati u kojoj mjeri knjižnice mogu velike podatke integrirati u poslovanje (Wang i sur. 2016: 4).

S obzirom na porast istraživanja temeljenih na velikim količinama podataka i ulogu knjižnice da istraživačima služi kao savjetodavno tijelo i bude podrška tijekom čitavog istraživačkog procesa, osim analitičkih vještina, od knjižničara se zahtijevaju vještine stvaranja, obrade, korištenja, pronalaženja, vrednovanja, prikazivanja informacija. Uz ove vještine, u svrhu savjetovanja, trebaju posjedovati i sposobnosti savjetovanja vezane uz očuvanje rezultata istraživanja, znanje o upravljanju istraživačkim podacima, znanje o različitim istraživačkim standardima, znanje o alatima za manipulaciju istraživačkih podataka i znanje o korištenju metapodataka (Wusu i sur. 2020: 6). Posjedovanjem ovih vještina knjižničari, na učinkovit način mogu osigurati kvalitetu istraživanja čime se indirektno potiče otkrivanje i inovacija znanja.

Također, knjižničari trebaju posjedovati vještine, konverzije, redukcije i kompresije podataka (Li i Dou, 2017: 3). U kombinaciji sa sposobnostima analize i obrade

podataka, ove će vještine uvjetovati kvalitetu repozitorija podataka, odnosno kvalitetu podataka koje neka knjižnica daje na korištenje.

Nedostatak potrebnih vještina ukazuje na potrebu promjene u obrazovanju informacijskih stručnjaka, posebice knjižničara. Nezaustavljiv tempo kojim veliki podaci postaju sve relevantniji u svijetu znatno će povećati potrebe tržišta rada za osobama koje posjeduju duboke analitičke vještine i vještine upravljanja podacima (Waqar i Kanwal, 2017: 23). U tu svrhu kratkotrajni tečajevi neće biti dovoljni (Wang i sur. 2016: 4). Obrazovni programi informacijskih znanosti trebaju se prilagoditi novim potrebama tržišta rada i usmjeriti napore na kultiviranje informacijskih stručnjaka s adekvatnim analitičkim i podatkovnim sposobnostima.

### 6.3 Problem budžeta

Koristi primjene tehnologije velikih podataka su očite, no često, implementacija takve tehnologije nije u budžetu knjižnica (Wang i sur. 2016: 4). Uz to, u mnogo slučajeva se ti budžeti smanjuju što mnoge knjižnice navodi da implementaciju tehnologije velikih podataka (koja zahtjeva strukturalne promjene unutar knjižnice) uopće ne stavljaju na dnevni red (Salo, 2010, prema Wang i sur. 2016: 4). Također, budući da je većina knjižnične građe u mnogim knjižnicama diljem svijeta u ne-digitaliziranom obliku, za primjenu velikih podataka potrebna je digitalizacija građe, što je proces koji zahtjeva ljudske i financijske resurse (Wang i sur. 2016: 4). Još jedan, knjižnicama dobro poznati problem jest ubrzano povećavanje količine literature u svijetu (Long i Harvey, 2014: 2 – 9, prema Li i Dou, 2017: 10). U eri velikih podataka problem balansiranja budžeta i potreba čitača postaje samo teži. Osim zahtijevanja poticaja, knjižnicama, s problemom financiranja implementacije velikih podataka, preporučuje se suradnja s drugim knjižnicama u svrhu dijeljenja resursa kao što su tehnologija, informacije i knjižnična infrastruktura (Oladokun, 2014: 24, prema Li i Dou, 2017: 15). S obzirom na to da se primjenom velikih podataka omogućuje otkrivanje i inovacija znanja, odnosno napredak zajednice općenito, problem financiranja trebao bi se rješavati na nacionalnoj i internacionalnoj razini, a ne prepustiti individualnim knjižnicama (Li i Dou, 2017: 10).

## 6.4 Problemi prouzrokovani dimenzijama velikih podataka

U skladu s kontinuiranim porastom količine podataka, informacija i znanja u svijetu informacijske potrebe postaju sve složenije. Taj porast količine podataka toliko je brz da će se knjižnice vrlo vjerojatno, ako ne počnu primjenjivati tehnologiju velikih podataka u poslovanju, naći u situaciji gdje neće moći zadovoljiti informacijske potrebe korisnika ili uopće neće imati korisnika kojima bi ponudile informacijske usluge. Tome ne pomaže ni rastuća popularnost istraživanja temeljenih na obradi velike količine podataka.

Veliki podaci koje knjižnica treba obrađivati kako bi ostala relevantna ogromnih su količina, različitih formata i dolaze iz različitih izvora. Iako možda ne u istim količinama, istraživački podaci također su raznoliki, budući da ih obilježava subjektivnost istraživača (Wang i sur. 2016: 3). Podaci su preveliki, stvaraju se prebrzo i ne odgovaraju regularnim strukturama baza podataka. Posljedica toga jest da se knjižnice sve manje fokusiraju na knjige, časopise i kataloge a sve više podatke kao što su: tekstovi, metapodaci, slike, audio-zapisi, video-zapisi, istraživački podaci itd. Učinkovito indeksirati i prikazati ove podatke knjižnica ne može bez primjene odgovarajuće tehnologije. Transformaciju, skrb, analizu i prikaz ovakvih podataka također je nemoguće postići primjenom zastarjelih tehnologija. Problem je u tome što je većina postojećih sustava za upravljanje podacima i analizu, koje knjižnice koriste namijenjeno za strukturirane podatke (Li i Dou, 2017: 9). Velika količina istraživačkih podataka proteklih deset godina i dalje je analogna i zahtjeva digitalizaciju (Wang i sur. 2016: 4). Uz to, beskorisni i prljavi podaci neizbježni su u velikim skupovima podataka (Long i Harvey, 2014: 2 – 9, prema Li i Dou, 2017: 8).

Da bi se knjižnica nosila s podacima, tj. mogla primijeniti odgovarajuću tehnologiju prvo treba posjedovati centralni repozitorij u kojeg će dospijevati standardizirani podaci, a prikupljat će ih članovi koji podatke još održavaju i provode kroz proces skrbi (Wang i sur. 2016: 3). Postojeće podatke knjižnica mora provesti kroz procese redukcije i kompresije čime se postiže da beskorisni podaci ne ugrožavaju kvalitetu analize i smanjuje njenu zahtjevnost (Li i Dou, 2017: 9). Nakon toga, primjenom računanja na oblaku podatke može analizirati i pohraniti na efikasan način. (Al-Barashdi i Al-Karousi, 2018: 2). Na kraju, knjižnica može primijeniti prikladnu

metodu prikaza podataka na način koji osigurava točnost podataka (Li i Dou, 2017: 8).

## 6.5 Potreba za kritičkom podatkovnom pismenošću

U svijetu se svakodnevno stvaraju podaci koji dolaze iz različitih izvora, upitne su kvalitete i koriste se u različite svrhe. Društveno uvjetovani karakter podataka u eri velikih podataka izrazito je naglašen. Dalton i Thatcher svojim sustavnim pristupom kritici podataka navode sljedeće karakteristike podataka: podaci su smješteni u vremenu i prostoru, tehnologija je inherentno politička, podaci se oblikuju u kontekstu društvenih procesa i inherentni su odnosima moći, podaci nikad nisu sirovi ni objektivni, podaci ne mogu u potpunosti zamijeniti kako čovjek spoznaje – oni su komplementarni spoznaji, podaci se mogu upotrijebiti na društveno progresivan način (Dalton i Thatcher, 2014, prema, Oliphant, 2017: 4). Prema ovim karakteristikama podaci su smješteni u historijski i društveni kontekst što znači da se uz njih uvijek veže neka društvena vrijednost. Schönberger i Cukier navode da veliki podaci mijenjaju načine kako čovjek misli, radi i živi (Mayer-Schönberger i Cukier, 2013, 190). Primjer društvenih i ideoloških konotacija podataka nudi Day opisujući usku povezanost velikih podataka i neoliberalizma u proizvodnji potreba i oblikovanju pojedinca (Day, 2014: 123-144). Značenje ovog jest da podaci sve više postaju faktori proizvodnje (Mayer-Schönberger i Cukier, 2013, 101). S obzirom na spomenuto, čovjek se mora pitati kome se i zašto omogućuje pristup podacima, tko ima koristi od određenog načina diseminacije, zašto se potiče određeni način korištenja podataka, te zašto i kako, zbog podataka, na ovaj ili onaj način, doživljava zbilju. Uključivanje kritičnosti u upravljanju s podacima ključno je ako se pomoću njih treba postići društveni napredak.

Velik podaci ozbiljno narušavaju privatnost i sigurnost. Ljudi, često nesvjesno, stvaraju podatke kojima nikada neće moći pristupiti i koji se koriste u svrhe praćenja ponašanja, lokacije itd (Joiner, 2018, 99). Također, problemi nastaju kada se podaci, koje je pojedinac povjerio nekoj instituciji, dijele sa strankama, kojima isti pojedinac nikada ne bi povjerio podatke. Ponovo korištenje podataka omogućuje da se oni koriste u svrhe za koje pojedinac ne bi dao pristanak ili, čak, na njegovu štetu. Općenito gledajući, problem jest što pojedinci, često podcjenjuju profitno društvo i moć obrade velikih podataka, te misle da je gotovo nemoguće dobiti uvid u njihove

privatne živote bez njihovog pristanka (Minelli, Chambers i Dhiraj, 2013: 151). Bez pažljive prosudbe, pravo na privatnost osobnih podataka samostalno zamjenjuju za trivijalne online usluge, koje pred njih postavljaju prividni izbor davanja pristanka u svrhu korištenja usluge ili odustajanja od usluge uopće (Minelli, Chambers i Dhiraj, 2013: 155). S obzirom na izrazitu teškoću regulacije privatnosti na Internetu, među-organizacijsko profiliranje korisnika integracijom njihovih „podatkovnih tragova“ (preko onih podataka koje korisnici daju na korištenje) realan je problem (Minelli, Chambers i Dhiraj, 2013: 155). Zbog ovakve „erozije“ privatnosti u svijetu se počinje raspravljati o pravu da pojedinac bude zaboravljen (Haga, 2017: 97). Iako takva nastojanja predstavljaju dobar smjer za rješavanje problema privatnosti, uloga prevencije, u prvom redu, pripada pojedincu, koji u svojem ponašanju treba biti oprezniji i, prije svega, svjesniji konotacija velikih podataka. Promjena mentaliteta, u onaj koji će biti primjeren eri velikih podataka, uključuje prisvajanje „zdravog“ skepticizma i kritičnosti prema rukovanju s podacima.

Knjižnice su institucije u koje korisnici imaju povjerenje i one si ne smiju dopustiti izigrati to povjerenje. Kod svojih korisnika trebaju kultivirati kritičnost u upravljanju s podacima u jednakoj mjeri, u kojoj tu kritičnost same trebaju prakticirati. Stoga, bitno je da se u obrazovne programe knjižničara i u informacijske usluge koje knjižnice nude korisnicima uključi podatkovna pismenost s naglašenom dimenzijom kritike (Oliphant, 2017: 1). Na taj će način osigurati da se podaci koriste u skladu s moralnim vrijednostima zajednice i da podupiru društveni napredak, te će očuvati sigurnost i privatnost svojih korisnika.

## 7 Zaključak

Razvojem računalne, telekomunikacijske i senzorne tehnologije, te razvojem Interneta količina podataka, informacija i znanja u svijetu znatno je porasla. U skladu s tim razvojem povećale su se i sposobnosti prikupljanja, analize i obrade podataka. Proces prikupljanja podataka, koji je prije bio naporan i dugotrajan, gotovo je u potpunosti automatiziran, a troškovi pohrane podataka znatno su se smanjili. Može se reći da su tehnološke granice prošlosti, koje su diktirale koliko čovjek može posjedovati podataka, informacija i znanja nestale. Količina podataka u svijetu, brzine kojima se oni stvaraju, raznolikost formata na kojima se oni pojavljuju te vrijednosti do kojih se njihovom obradom može doći rezultirala je stavom da će podaci, odnosno njihova obrada biti glavni čimbenik društvenog i komercijalnog napretka. Ta neprestano rastuća količina podataka dobila je ime – veliki podaci.

Veliki podaci jest pojam, kojim se nastoji obuhvatiti ogromne količine strukturiranih i nestrukturiranih podataka, koji se generiraju izrazito visokim brzinama i koje je nemoguće pohraniti, analizirati i obraditi tradicionalnim metodama. Zbog stalno mijenjajućeg karaktera podataka u svijetu teško je ponuditi jedinstvenu definiciju koncepta velikih podataka, stoga se oni najčešće definiraju pomoću skupa karakteristika, odnosno dimenzija, koje podaci i tehnike trebaju posjedovati ako ih se želi obuhvatiti tim pojmom. Količina podataka nije precizno određena te je dovoljno reći da se tradicionalne metode i alati ne mogu nositi s prikupljanjem, obradom i analizom velikih podataka. Zahvaljujući senzornoj tehnologiji i činjenici da se sve više segmenata čovjekovog života povezuje s računalima i Internetom, podaci se stvaraju izrazito visokim brzinama. Pojavljuju se u različitim oblicima, odnosno uključuju strukturiran, polu-strukturirane i nestrukturirane podatke, te dolaze iz različitih izvora. Najznačajnija karakteristika velikih podataka tiče se implikacija koje oni imaju za pronalaženje vrijednosti. Pomoću učinkovite analize velikih skupova podataka otkrivaju se veze, obrasci i korelacije, te konstantnom obradom i analizom postaju vidljive i promjene. Vrijednost proizlazi iz uvida na temelju kojih institucije i organizacije donose pravovremene i učinkovite odluke. Također, do uvida je moguće doći ponovnim korištenjem podataka u svrhe koje ne trebaju nužno imati veze s prvobitnim korištenjem, budući da se vrijednost podataka, za razliku od materijalnih dobara, ne iscrpljuje korištenjem.



Društvo prepoznaje važnost ovih vrijednosti i prema tome se tehnologija velikih podataka primjenjuje u širokom spektru disciplina, organizacija i institucija, dok za one, u kojima se ne primjenjuju ima ozbiljne implikacije. Jedna od tih institucija jest knjižnica. Knjižnice su oduvijek nastojale osigurati pristup informacijama i znanju. Međutim, u eri velikih podataka ta funkcija je otežana. Zbog prisutnosti ogromnih količina podataka, porasta količine informacija i znanja, informacijske potrebe korisnika knjižnice postaju sve složenije. Također, navike kako ljudi pretražuju informacije se mijenjaju što omogućuje konkurentima kao što je Google, koji na svoje usluge primjenjuje tehnologiju velikih podataka, da preuzmu funkcije knjižnice i time ozbiljno ugroze njihovu poziciju u zajednici. Iako u području LIS-a istraživanja o velikim podacima nisu učestala, implikacije su jasne.

Pojavom velikih podataka i pratećih mogućnosti za prikupljanje, obradu i analizu podataka promijenio se i istraživački proces. Korisnici se teško nose s velikim količinama podataka, te često ne posjeduju potrebna znanja i vještine da rukuju s alatima potrebnim za obradu i analizu tih podataka. Time je ugrožena kvaliteta njihovih istraživanja. U skladu s tim, glavna uloga knjižnice u eri velikih podataka postaje biti savjetodavno tijelo i podrška istraživačkom procesu. Kako bi knjižnica poticala otkrivanje i inovaciju znanja ona mora te ogromne količine podataka učiniti dostupnim i prikladnim za korištenje. To postiže procesima skrbi, dijeljenjem i povezivanjem podataka. Uz to, mora posjedovati temeljita znanja o istraživačkim metodama, standardima i alatima za prikupljanje, obradu i analizu velikih podataka.

Primjena tehnologije velikih podataka u knjižnicama može se podijeliti na tri skupine. Prva se skupina odnosi na primjenu velikih podataka u svrhu stvaranja personaliziranih usluga i prilagođavanja korisničkim potrebama čime knjižnica osigurava svoju relevantnost. Druga skupina usmjerena je na samu knjižnicu i primjenjuje se radi donošenja odluka na osnovi podataka čime knjižnica može vrednovati i pravovremeno korigirati način svog poslovanja. Treća skupina odnosi se na usluge koje knjižnica treba pružati i djelatnosti koje treba obavljati ako se podaci shvate kao resursi koje korisnicima daje na korištenje, te je usko povezana s ulogom podrške i savjetovanja.

Da bi se tehnologija velikih podataka primijenila, knjižničari se moraju suočiti s nedostatkom dovoljno osposobljenih radnika. Potrebni su knjižničari s dubokim

poznavanjem analitike podataka, što znači da se njihovo obrazovanje mora prilagoditi ovim zahtjevima. Također, kao jedan od izazova primjene velikih podataka pokazao se mentalitet ponekih knjižnica i knjižničara koji iz neracionalnih razloga odbijaju tu tehnologiju primijeniti ili pak zahtijevaju apsolutnu točnost u svakom segmentu djelatnosti. Veze, obrasci i korelacije, koje se otkrivaju primjenom tehnologije velikih podataka, često su slučajne i iznenađujuće, a ograničeni mentalitet onemogućuje da se posredstvom njih ostvari vrijednost. Uz mentalitet prihvaćanja mogućnost i prilika koje veliki podaci donose, treba kultivirati kritičnost u svakom segmentu rada s podacima. Budući da su podaci društveno uvjetovani, knjižnice trebaju, u svrhu napretka zajednice i sigurnosti i privatnosti, poticati kritičnost te ju primjenjivati na sebe jednakom mjerom Budžeti kojima knjižnice raspolažu u velikom broju slučajeva predstavlja problem za implementaciju tehnologije velikih podataka. Mnoge knjižnice tako ovise o svojim matičnim institucijama i vlastitim sposobnostima za suradnju. Također, mnogi podaci koje knjižnice već posjeduju, poprimaju karakteristike velikih podataka. Primjenom tehnologije velikih podataka knjižnicama je moguće nositi se s takvim podacima, te iz njih dobiti maksimalnu vrijednost.

Napredak kulture velikih podataka ne pokazuje znakove zaustavljanja. Ako knjižnice ne uspiju kontinuiranim razvojem pratiti ovaj tempo postat će nerelevantne zajednici. Primijeniti tehnologiju velikih podataka je imperativ u brzo-mijenjajućem i rastućem svijetu podataka.

## 8 Popis literature

- Aaltonen, A., Tempini, N. (2014.) "Everything Counts in Large Amounts: A Critical Realist Study on Data-Based Production," *Journal of Information Technology*, sv. 29, 108.
- Al-Barashdi, H., Al-Karousi, R. (2018.) „Big Data in academic libraries: literature review and future research direction“ *Journal of Information Studies & Technology*, sv. 13, 1 – 16.
- Alotaibi, N.M., Abdullah, M.A. (2016) "Big data mining: A classification perspective. In Communication, Management and Information Technology" *Proceedings of the International Conference on Communication, Management and Information Technology*, 687 – 696.
- Anderson, C. (2008.) "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." *WIRED*, (Pristupljeno: 27. lipnja 2008. [http://archive.wired.com/science/discoveries/magazine/16-07/pb\\_theory](http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory)).
- Bailey, J. (2015.), "Top 4 ways to apply big data in higher education," *Information Age*, (Pristupljeno: 6. studenog 2018 <http://www.information-age.com/top-4-ways-apply-big-data-higher-education-123460114/>).
- Ball, R. (2019.) „Big Data and Their Impact on Libraries“ *American Journal of Information Science and Technology*, sv. 3 (1), 1 – 9.
- Burke, P. (Org.) (1992) *A escrita da história: novas perspectiva*, São Paulo: Unesp.
- Cullen, R., Chawner, B. (2011.) "Institutional repositories, open access, and scholarly communication: A study of conflicting paradigms" *Journal of Academic Librarianship*, sv. 37 (6), 460 – 470.
- Dalton, C., Thatcher, J. (2014.) "What Does a Critical Data Studies Look Like, and Why Do We Care? Seven Points for a Critical Approach to 'Big Data'" *Society & Space*, 2014, <http://societyandspace.com/material/commentaries/craig-daltonand-jim-thatcher-what-does-a-critical-data-studies-look-like-and-why-do-we-care-seven-points-for-a-critical-approach-to-big-data/>.
- Day, R. E. (2014.) *Indexing it All*, Cambridge, Massachusetts: The MIT Press, 2014.
- Dumbill, Edd, (2012). „What is dig data“, u: O'Reilly Radar Team (ur.) *Big Data Now: Current Perspectives from O'Reilly Media*. Sebastopol: O'Reilly Media, Inc.
- Gantz, J., Reinsel, D. (2011) „Extracting value from chaos“ *IDC iView*, 1–12.
- Gordon-Murnane, L. (2012.) "Big Data: A Big Opportunity for Librarians," *Online*, sv. 36 (5): 34.
- Haga, Y. (2017.) "Right to be Forgotten: A New Privacy Right in the Era of Internet" u: M. Corrales, M. Fenwick, N. Forgo (ur.), *New Technology, Big Data and the Law*, Singapore: Springer Nature Singapore.
- Hoy, M. B., (2014.) „Big Data: An Introduction for Librarians“ *Medical Reference Services Quarterly*, sv. 33 (3), 1 – 12.
- Head, S. R., Komori, H. K., LaMere, S. A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D.R. and Ordoukhanian, P. (2014) "Library construction for next-

- generation sequencing: overviews and challenges” *Biotechniques*, sv. 56 (2), str. 61.
- Hey, T. (2006.) „E-science nad Its Implications for the Library Community“ *Library Hi Tech*, sv. 24 (4), 1 – 15.
  - Huang, X., Zhong, H. (2013.) “The Construction of Enterprise Competitive Intelligence System Based on Big Data” *Intelligence Magazine*, sv. 3: 37 – 43.
  - Huwe, T.K. (2014.) “Big Data and the Library: A Natural Fit.” *Computers in Libraries*, sv. 34 (2), 17–18.
  - Jaiswal, P., Singh Pratap, A. (2020.) „Big Daa Analytics: A Catalyst for re-energizing LIS Education“ *Library Philosophy and Praticice (e – journal)*, 3867, 1 – 13.
  - Jeffrey, R. S. (2013.) “The Open Science EFramework: Improving Science by Making it Open and Accessible” *A Dissertation presented to the Graduate Faculty of the University of Virginia in Candidacy for the Degree of Doctor of Philosophy 2013*.
  - Jharotia, A. K. (2016.) “Big Data Technology: Big Opportunity for Librarians.” (Pristupljeno: 24. lipnja 2021.  
[https://www.researchgate.net/publication/326972552\\_Big\\_Data\\_Technology\\_Big\\_Opportunity\\_for\\_Librarians](https://www.researchgate.net/publication/326972552_Big_Data_Technology_Big_Opportunity_for_Librarians)).
  - Joiner, I. A. (2018.) *Emerging library technologies*, Cambridge: Chandos Publishing, 2018.
  - Kim, S. and Choi, M.S. (2016), “Study on data center and data librarian role for reuse of research data” *In Knowledge and Smart Technology (KST), 2016 8th International Conference on IEEE*, 303 – 308.
  - Kirkwood, R. J. (2016.) „Collection development or data-driven content curation? An exploratory project in Manchester“ *Library Management*, sv. 37 (4/5), 275 – 284.
  - Kitchin, R. (2015.) „Big data and official statistics: Opportunities, challenges and risks“ *Statistical Journal of the IAOS*, sv. 31, 471 – 481.
  - Krier, L., Strasser, C. A. (2014.) *Data Management for Libraries*, Chicago: ALA TechSource.
  - Laney, D. (2001.) “3-D data management: Controlling data volume, velocity and variety” *META Research Note*.
  - Long, R., Harvey, P. (2014) “Technologies for Libraries” *The ANZTLA EJournal*, sv. (10), 2 – 9.
  - Li, J., Dou, G. (2017.) „Big data application framework and its feasibility analysis in library“ *Information Discovery and Delivery*, sv. 45 (4), 1 – 24, [https://www.researchgate.net/publication/320338751\\_Big\\_data\\_application\\_framework\\_and\\_its\\_feasibility\\_analysis\\_in\\_library](https://www.researchgate.net/publication/320338751_Big_data_application_framework_and_its_feasibility_analysis_in_library).
  - Mayer-Schönberger, V., Cukier, K. (2013.) *Big Data*, New York: Houghton Mifflin, 2013.
  - Meyer, E. T., Schroeder R., Taylor, L. E. M. (2013.) “Big Data: Rewards and Risks for the Social Sciences.” *Oxford Internet Institute*, Call for workshop papers, <http://www.oii.ox.ac.uk/events/?id=557>
  - Minelli, M., Chambers, M., Dhiraj, A. (2013.) “Big Data, Big Analytics”, Haboken: Jon Wiley & Sons, Inc.

- Narendra, A.P. (2016), "Big Data, Data Analyst, and Improving the Competence of Librarian" *Record and Library Journal*, sv. 1 (2), 83 – 93.
- Oladokun, O. (2014) "Library Cooperation and Resource Sharing: Some Lessons from the University of Botswana" *African Journal of Library, Archives and Information Science*, sv. 24 (2).
- Oliphant, T. (2017.) „A Case for Critical Studies in Library and Information Studies“ *Journal of Critical Library and Information Studies*, sv. 1 (1): 1 – 24.
- Oliveira, L. C. F. de, Cendón, B. V. (2020.) „Big Data in Information Science literature“ *Qualitative and Quantitative Methods in Libraries*, sv. 9 (2): 155 – 172.
- O'Maley Voliva, C. (2015.) "Data Visualization for Public Libraries" , *Public Libraries Online*, (<http://publiclibrariesonline.org/2015/04/datavisualization-for-public-libraries/> pristupljeno: 06. studenog 2018).
- Provost, F., Fawcett, T. (2013.) „Data Science and its Relationship to Big Data and Data-Driven Decision Making“ *Big Data*, sv 1 (1), 51 – 59.
- Ribeiro, C. J. S., (2014) „Big data: os novos desafios para o profissional da informação“ *Revista Informação & Tecnologia (ITEC)*, sv. 1 (1), 96 – 105.
- Salo, D., (2010.) "Retooling Libraries for the Data Challenge" *Ariadne*, sv. 64, (<http://www.ariadne.ac.uk/issue64/salo/>)
- Siguenza-Guzman, L., Saquicela, V., Avila-Ordóñez, E., Vandewalle, J., Cattrysse, D. (2015) "Literature review of data mining applications in academic libraries" *The Journal of Academic Librarianship*, sv. 41 (4), 499 – 510.
- Simon, P. (2013) "Too Big to Ignore: The Business Case For Big Data", Hoboken: John Wiley & Sons, Inc.
- Sollenberger, J., Cavanaugh, M.L.S., S.K. and MSIS, C.A.T. (2013) "The value of library and information services in patient care: results of a multisite study" *Journal of the Medical Library Association*, sv. 101 (1), 38.
- Tableau Software (2018.) "Tableau Website", (<https://www.tableau.com/> site pristupljeno: 06. studenog 2018.).
- Teens, M., Goldner, M. (2013.) „Libraries' Role in Curating and Exposing Big Data“ *Future Internet*, sv. 5, 429 – 438.
- Tiwari, A. (2018.) "Data Curation: An opportunity for the libraries" *Conference paper, MANLIBNET 2018: International Conference on Emerging Trends in Librarianship: Role of Libraries in Learning Environment: IIM, Trichy*, 1 – 10.
- Vogel, J. (2017.) "Distributed and Connected Information in the Internet", u: A. J. Schuster (ur.) *Understanding Information: From the Big Bang to Big Data*, Cham: Springer International Publishing.
- Waller, M. A., Fawcett, S. E. (2013). "Data science, predictive analytics, and Big Data: a revolution that will transform supply chain design and management" *Journal of Business Logistics*, sv. 34 (2), 77 – 84.
- Wang, C., Xu, S., Chen, L., Chen, X. (2016.) „Exposing Library Data with Big Data Technology: A Review“, *2016 IEEE/ACIS 15th International Conference on Computer and Information Science*, 1 – 6.
- Wang, D. (2016.) „Information Service in the Big Data Era and Development Strategies for University Libraries“, *Proceedings of the 2nd Annual*

*International Conference on Social Science and Contemporary Humanity Development, Wuhan, China, 664 – 668.*

- Waqar, A., Kanwal, A. (2017.) „Defining big data and measuring its associated trends in the field of information and library management“ *Library Hi Tech News*, sv. 9, 21 – 24.
- Weiss, R., Zgorski, L.J. (2012) “Obama administration unveils “big data” initiative: Announces \$200 million in new R&D investments” *Office of Science and Technology Policy Executive Office of the President*.
- Wittmann, R. J., Reinhalter, L. (2014) „The Library: Big Data's Boomtown“ *The Serials Librarian: From the Printed Page to the Digital Age*, sv. 67 (4), 1 – 14.
- Wu, X., Zhu, X., Wu, G.Q. and Ding, W. (2014), “Data mining with big data”, *IEEE transactions on knowledge and data engineering*, sv. 6 (1), 97 – 107.
- Wusu, O., Lazarus, N., Otun, M., Arekemase, O., Olayiwola, T. (2020.) „Big data and openness: a big issue with librarians“ *Library Philosophy and Practice (e – journal)*, 4279, 1 – 11.
- Zikopoulos, P., Parasuraman, K., Deutsch, T., Giles, J. and Corrigan, D. (2012) „Harness the power of big data The IBM big data platform“ *McGraw Hill Professional*.
- Zipei, T. (2012.) *The Big Data*, Guangxi Normal University Press.
- Zhou, Q. (2016), “eBook Service in Shanghai Library” *International Journal of Librarianship*, sv. 1 (1), str. 66 – 73.

# Veliki podaci u informacijskim ustanovama – primijenjene i kritičke dimenzije

## 9 Sažetak

Rad prikazuje velike podatke u kontekstu knjižnica. To čini objašnjavajući glavne prednosti i izazove primjene tehnologije velikih podataka u knjižnicama, te objašnjavajući implikacije, koje era velikih podataka ima za knjižnice. Kako bi se prednosti, izazovi i implikacije razumjele rad nudi definiciju velikih podataka. Veliki podaci definiraju se pomoću dimenzija veličine, brzine, raznolikosti, integriteta i vrijednosti. Dimenzija vrijednosti se detaljnije obrađuje. Imajući na umu poteškoće s kojima bi se korisnici mogli susresti u eri velikih podataka, knjižnice bi trebale preuzeti ulogu podrške i savjetovanja o istraživačkom procesu i upravljanju podataka općenito. Glavne prednosti primjene tehnologije velikih podataka su personalizirane informacijske usluge, donošenje odluka temeljenih na obradi podataka i mogućnost pomaganja u istraživačkom procesu. Glavni izazovi za primjenu uključuju nedostatak osposobljenih knjižničara, potrebu za promjenom mentaliteta i probleme s budžetom. Zbog sve većih količina podataka varijabilne kvalitete i iz varijabilnih izvora knjižnice će morati naučiti upravljati tim podacima. Kritička podatkovna pismenost također je potrebna. Rad pokazuje da je primjena tehnologije velikih podataka u knjižnicama imperativ ako one žele ostati relevantne.

**Ključne riječi:** veliki podaci, knjižnice, informacijske usluge, istraživački podaci, vrijednost, kritička podatkovna pismenost, informacijski stručnjak, analiza

# **Big Data in Information Institutions – Application and critical dimensions**

## **Abstract**

This paper reviews big data in the context of libraries. It does so by explaining the main benefits and challenges of applying big data technology in libraries and by explaining the main implications of the big data era. To help understand these benefits, challenges and implications a definition of big data is given. Big data is defined by its dimensions of volume, velocity, variety, veracity and value, with special focus on the dimension of value. Keeping in mind, the difficulties library users might face in the era of big data, libraries should take on the role of supporters and consultants regarding the research process and handling data in general. The main benefits of applying big data technology are personalized information services, data-based decision making and being able to assist researchers. The main difficulties include the lack of skilled librarians, a need for change in the mindset and budget issues. Because of the ever increasing amount of data of various quality and from various sources libraries will need to learn how to manage it. Critical data literacy is also required. The paper shows that applying big data technology is an imperative if they wish to stay relevant.

**Key words:** big data, libraries, information service, research data, value, critical data literacy, information professional, analysis