

Jelisaveta Petrović<sup>1</sup>  
Filozofski fakultet  
Univerzitet u Beogradu

Pregledni naučni rad  
UDK 004.65:316.1  
004.6:[316:378  
303.64:316(05)  
Primljen: 02.07.2018



DOI: <https://doi.org/10.2298/SOC1803557P>

## „VELIKI PODACI“ – VELIKI IZAZOV ZA SOCIOLOGIJU?

### Big Data – a Big Deal for Sociology?<sup>2</sup>

*APSTRAKT: U radu se kritički ispituje odnos mainstream sociologije prema fenomenu masovnih podataka („big data“) i primeni analize masovnih podataka u društvenim naukama. Analiza sadržaja tekstova objavljenih u prestižnim sociološkim časopisima rangiranim na Thomson Reuters Impact Factor ssci listi (2000–2017) je pokazala da se temi masovnih podataka izvesna pažnja počinje poklanjati tek u poslednjih nekoliko godina. Ova tema je i dalje zanemarena u glavnim tokovima sociološke misli iako veliku pažnju privlači izvan polja društvenih nauka. Istraživanjem fenomena masovnih podataka dominira kritička perspektiva, dok je primena analize masovnih podataka relativno slabo zastupljena.*

*U zaključku rada se ukazuje na važnost korišćenja masovnih podataka u sociologiji, te na pretnje auto-marginalizacije ukoliko se nastavi sa zanemarivanjem „tektonskih“ promena do kojih, sa razvojem analitike masovnih podataka, dolazi u prostoru kojim su nekada suvereno vladale društvene nauke.*

**KLJUČNE REČI:** masovni podaci, sociologija, društvene nauke, sociološki metod

*ABSTRACT: The paper critically examines the attitude of the mainstream sociology towards the study of big data in social sciences. Content analysis of the scientific papers published in the top-tier sociological journals ranked on the Thomson Reuters Impact Factor ssci list (2000–2017) shows that, in the observed period, the issue of big data was largely neglected. This topic is still rather invisible in the mainstream sociological thought, although it draws a lot of attention outside the academia. The analysis of big data within mainstream sociology is dominated by a critical perspective, while the application of the big data analytics is quite rare.*

---

1 vukelic.jelisaveta@gmail.com

2 Ovaj rad je nastao u okviru projekta „Izazovi nove društvene integracije u Srbiji: koncepti i akteri (ev. broj 179035) koji finansira Ministarstvo prosvete, nauke i tehnološkog razvoja Republike Srbije.

*In the concluding section, the importance of the big data study for sociology is emphasised. Moreover, it is pointed out at the risk of auto-marginalization in case of neglecting the “tectonic” changes induced by the big data analytics in the space once dominated by the social sciences.*

KEY WORDS: Big data, Sociology, Social sciences, Sociological method

*„Napolje sa svakom teorijom ljudskog ponašanja, od lingvistike do sociologije! Zaboravite na taksonomiju, ontologiju i psihologiju. Ko zna zašto ljudi rade to što rade? Poenta je u tome da oni to rade, a mi možemo to da pratimo i merimo sa tačnošću bez presedana. Sa dovoljno podataka, brojevi govore sami za sebe...“ (Anderson, 2008).*

Moglo bi se reći da su masovni podaci („big data“) prvi put privukli pažnju sociologa, upravo onda kada je u popularnom časopisu *Wired* objavljen tekst glavnog urednika Krisa Andersona provokativno naslovljen „Kraj teorije“ („*The End of Theory*“), čiju centralnu ideju predstavlja navedeni odlomak. Odjek ovog teksta bi, u naučnoj zajednici, verovatno bio znatno slabiji da nije reflektovao uverenja koja vladaju tehnološkom i poslovnom sferom.

Godinu dana pre nego što će Anderson samouvereno proglasiti kraj (društvene) teorije, u časopisu *Sociology* je objavljen članak „Nadolazeća kriza empirijske sociologije“ (“*The Coming Crisis of Empirical Sociology*“) dvojice britanskih sociologa – Majka Sevidža i Rodžera Barouza. Tekst je, očekivano, izazvao burne reakcije, a za nas je posebno zanimljiv jer, iz perspektive sociologa, opisuje prvi susret sa masovnim podacima. Majk Sevidž je, kako piše, shvatio da je sociologija upala u krizu 2004. godine na jednom naučnom skupu posvećenom istraživačkim metodama. U toku neformalnog razgovora sa mladim stručnjakom za digitalne tehnologije zaposlenim u jednoj od vodećih britanskih telekomunikacionih kompanija, Sevidž je na veliko iznenađenje saznao da je njegov sagovornik svoje istraživanje društvenih mreža sproveo, potpuno samostalno, na bazi podataka svih razgovora obavljenih putem mobilnog telefona, ukupno nekoliko milijardi poziva. Ovi podaci su bili toliko impresivni, priseća se Sevidž, da su u duboku senku bacili sva druga istraživanja predstavljena na skupu. Pored toga što su nezamislivo veliki spram onoga što jedan društveni naučnik može da prikupi uobičajenim metodama, ovi podaci nisu zahtevali gotovo nikakav dodatni napor da se prikupe, prosto su digitalni nusproizvod svakodnevnih aktivnosti ljudi (Savage & Burrows, 2007).

Iako anegdotskog karaktera, ovi primeri dobro oslikavaju aktuelne trendove u sferi tehnologije, biznisa i nauke te navode da se zapitamo kakva je budućnost sociologije i društvenih nauka uopšte u kontekstu masovnih podataka? Ono čime ćemo se u ovom radu posebno baviti je pitanje: *kako se akademska sociologija postavlja spram izazova koje sa sobom donosi revolucija masovnih podataka?*

Odgovor ćemo potražiti u najuticajnijim sociološkim časopisima i u njima objavljenim radovima koji za temu imaju upravo masovne podatke.

### Masovni podaci<sup>3</sup> („big data“)

U protekle dve decenije, ukupna količina podataka u svetu se uvećavala takvom brzinom da je približno 90% svih podataka koji danas postoje nastalo u poslednjih par godina (Lev-Ram, 2014). Nalazimo se u jeku revolucije podataka, a pojam “big data” je jedan od njenih najprepoznatljivijih simbola.

Iako se pojam masovni podaci prvi put pojavljuje negde potkraj 20. veka, popularnost stiže tek deceniju kasnije. Časopis *Nature* je 2008. godine posvetio poseban broj temi velikih podataka, dok je časopis *Science* to učinio tri godine kasnije. U 2012. godini, masovni podaci su bili jedna od glavnih tema na ekonomskom samitu u Davosu, a Obamina administracija je najavila ulaganje od 200 miliona dolara u projekte masovnih podataka. Uviđajući potencijal ekonomije podataka, čija se vrednost procenjuje na preko 50 milijardi dolara, Gugl, Majkrosoft, Amazon, Fejsbuk i drugi tehnološki giganti su otpočeli velike projekte njihovog skladištenja i obrade (Baruh & Popescu, 2015:1). I u političkoj sferi se pažnja sve više poklanja mogućnostima prediktivne analitike zasnovane na masovnim podacima, pogotovo nakon uvida u njene mogućnosti predviđanja izbornih rezultata. Ujedinjene nacije su 2012. godine izdale prvi izveštaj o ulozi masovnih podataka u javnoj upravi, uzimajući u obzir da se na osnovu analize masovnih podataka ne donose više samo poslovne već i važne političke odluke (Chan et al, 2014:175; Kotler & Armstrong, 2017). Sa nekoliko godina zakašnjenja u odnosu na poslovnu sferu, masovni podaci postaju tema i među društvenim naučnicima. Na godišnjoj konferenciji Američke sociološke asocijacije 2014. godine, sedam sesija je bilo posvećeno masovnim podacima (Liu & Guo, 2016:2). Iste godine je objavljen prvi broj specijalizovanog časopisa *Big Data and Society*<sup>4</sup>.

*Masovni podaci* predstavljaju setove podataka koje odlikuje izuzetno veliki *obim* – reč je često o više petabajta memorijskog prostora. Kako bi se stekao osećaj veličine, recimo da 250,000 DVD filmova zauzima približno jedan petabajt memorije. Da bi se odredilo da li neki set podataka spada u kategoriju velikih podataka, među istraživačima postoji šaljivo neformalno pravilo koje glasi: ako Excel-om ili SPSS-om možete da obradite neki skup podataka ili ukoliko on može da stane na vaš računar, onda nije reč o masovnim podacima (Manovich, 2011; Kitchin, 2014).

Pored ogromnog volumena, masovne podatke odlikuje i izuzetna *raznovrsnost* u pogledu formata (tekst, slike, video zapisi) i nivoa strukturisanosti kao i velika *brzina* kojom se generišu (Laney, 2001). Razmere velikih podataka

3 U radu se kao sinonimi koriste pojmovi masovni podaci i veliki setovi podataka (skraćeno – veliki podaci).

4 Reč je o časopisu otvorenog pristupa koji je fokusiran na interdisciplinarni rad u oblasti društvenih i humanističkih nauka i računarstva. <http://journals.sagepub.com/home/bds> (pristupljeno 15.1.2018. godine).

dobro ilustruje primer Fejsbuka koji na dnevnom nivou procesuiru 5 milijardi različitih jedinica sadržaja (linkovi, slike, komentari) (Chan et al, 2014). Masovne podatke odlikuje i "sekundarni" odnosno "nađeni" karakter (Harford, 2014). Pored toga, kao karakteristike se izdvajaju i iscrpnost, odnosno težnja da se obuhvati čitava populacija, detaljnost, fleksibilnost, varijabilnost, kompleksnost i relacioni karakter. Primeri ovakvih podataka su brojni: elektronski zapisi kupoprodajnih transakcija ili pretrage po internetu; geolokacije zabeležene mobilnim telefonom; podaci različitih senzora itd. Eksperimentalni podaci koji se dobijaju iz Velikog hadronskog sudarača (LHC, Cern) takođe predstavljaju primer masovnih podataka.

Pojedini autori ističu razliku između masovnih podataka i uže kategorije *masovnih društvenih podataka* („social big data“). Dok masovni podaci predstavljaju proizvod kvantifikacije aktivnosti zabeleženih putem senzora ili kroz industrijske i privatne mreže, masovne društvene podatke odlikuje to što predstavljaju proizvod digitalno posredovanih komunikativnih praksi u svakodnevnom životu. Ovi podaci se najčešće kreiraju putem socijalnih medija ili pretragom po internetu (Coté, 2014; Bello-Oraz et al, 2015).

Masovni podaci predstavljaju izazov za standardne načine prikupljanja, skladištenja i obrade podataka. Industrija informacionih tehnologija je vrlo brzo odgovorila na te izazove, razvijajući nova softverska rešenja. Kao osnovno softversko/programsko znanje i radno okruženje koje je neophodno za analizu velikih setova podataka u stručnoj i praktično usmerenoj literaturi se izdvaja „računarstvo u oblaku“ (cloud computing), programski jezik Python (sa specijalizovanim bibliotekama), SQL (programski jezik za manipulisanje bazama podataka), otvoreni softver za statističku analizu R<sup>5</sup>, kao i Hadoop<sup>6</sup> i Spark okruženje (Kitchen, 2014; Chan et al, 2014).

## Izazovi masovnih podataka za društvene nauke

Kao i mnogi drugi socio-tehnološki fenomeni, i masovni podaci pobuđuju kako utopijska tako i distopijska razmišljanja i retoriku. S jedne strane, u masovnim podacima se vidi moćno sredstvo koje će čovečanstvu pomoći da se izbori sa nekim od najvećih problema današnjice kao što su siromaštvo, terorizam ili klimatske promene. S druge strane, na masovne podatke se neretko gleda kao na još jedno sredstvo bogaćenja, masovnog nadzora, ugrožavanja privatnosti, ljudskih prava te rastuće državne i korporativne kontrole (boyd & Crawford, 2012: 663–4).

Kamp entuzijasta za masovne podatke (Anderson, 2008; Mayer-Schoenberger & Cukier, 2013) se nalazi uglavnom izvan akademskih krugova, u sferi biznisa i tehnologije (Strasser, 2012). U najekstremnijem vidu, zagovornici

5 Za razliku od uobičajenih statističkih paketa kao što su SPSS i SAS, R omogućava obradu izuzetno velikih količina podataka kao i primenu naprednijih statističkih tehnika.

6 Hadoop je radni okvir (*framework*) otvorenog koda Apach fondacije koji služi za skladištenje i procesiranje velikih količina podataka. Pored Hadoop-a, druga posebno popularna tehnologija je Apach Spark platforma za izuzetno brzu obradu podataka.

korišćenja masovnih podataka ističu nepotrebnost svake teorije i nauke uopšte, odnosno smatraju da znanje može da bude „rođeno“ iz (dovoljno velike količine) podataka. Kako tvrde, analiza masovnih podataka omogućava da se ustanove korelacije, obrazaci i pravilnosti i dok korelacije ne mogu da kažu zašto se nešto dešava, one mogu da upozore da se nešto dešava (videti šire u: Kitchen, 2014; Mazzochi, 2015:1250–2). Na ovakvim „ateorijskim“ osnovama funkcionišu algoritmi koji se koriste za usmeravanje ponašanja potrošača. Jedan od zastupnika ovog pristupa kaže: „Mi najčešće ne znamo o uzrocima, a vrlo često nam to i nije preterano važno...sve dok algoritam dobro predviđa narednu kupovinu, nije važno kako on to radi“ (Siegel, 2013:90). Pored komercijalnih uspeha (vrtoglavi uspon kompanije Netflix se često navodi kao primer), autori ističu i dostignuća primene analize masovnih podataka u medicini<sup>7</sup>.

Iako se potreba za teorijskim utemeljenjem istraživanja dovodila u pitanje gotovo isključivo izvan akademskih krugova, jedna grupa društvenih naučnika je sklona da temeljno preispita uticaje koje na društvenu teoriju i istraživačku praksu ima digitalna revolucija. Bruno Latour, recimo, zaključuje da podaci nastali kao posledica interakcije korisnika sa digitalnim tehnologijama stvaraju potrebu za ništa manje nego ponovnim promišljanjem celokupne sociološke teorije (Latour et al, 2012). Drugi, pak, smatraju da nove tehnologije mogu pomoći da se razreše neke od najvećih dilema društvenih nauka (Housley et al, 2014). Bek i Puvar blagonaklono gledaju na ono što donosi revolucija masovnih podataka zalažući se za razvoj „žive“ sociologije koja na kreativan, javan ali i kritički način koristi dostupne „žive podatke“, nasuprot „mrtve“ sociologije koja upotrebljava zastarele „zombi“ koncepte (Back, 2012; Back & Puvar, 2012).

Pored isticanja mogućnosti koje se otvaraju u eri masovnih podataka, autori upozoravaju i na pretnju „auto-marginalizacije“ društvenih nauka koja može biti posledica zanemarivanja tehnoloških inovacija, poredeći to sa ignorisanjem mikroskopa u biologiji ili teleskopa u astronomiji (Lohr, 2015: 9). Na istoj liniji razmišljanja su i autori koji ukazuju da bi tehnika anketnog istraživanja kombinovana sa metodama statističkog zaključivanja, jedan od najjačih metodoloških aduta sociologije, trebalo da ustupi svoju povlašćenu poziciju analizi digitalno generisanih podataka (Savage & Burrows, 2007, 2009).

U kontekstu u kome različiti računarski stručnjaci i eksperti za podatke (engl. „data scientist“), zaposleni pri tehnološkim gigantima kao što su Fejsbuk, Gugl ili Amazon, imaju pristup i mogućnost analize ogromne količine društvenih podataka iz digitalnih izvora, tvrdnje sociologa da raspolažu superiornijim znanjem u istraživanju društvenog života postaju, u najmanju ruku, upitne (Rogers 2013: 206; Tinati et al, 2014). Pored konkurencije koja dolazi iz komercijalne sfere, pretnje *mainstream* sociologiji stižu i iz akademskih voda. Računarske društvene nauke (*computational social sciences*) se ubrzano razvijaju i dovode u pitanje kritički pristup karakterističan za društvene nauke preteći da ga „potčine“ kvantitativnoj logici računarskih nauka (Fuchs, 2017:46–7). Pionirka digitalne sociologije Debra Lupton ističe da je budućnost sociologije svetla ako (i samo ako) digitalna sfera bude nešto što će sociologija najozbiljnije uzeti u razmatranje (Lupton, 2015).

7 Kenneth Cukier, TEDSalon Berlin 2014, Big Data is Better Data ([www.ted.com](http://www.ted.com)). Pristupljeno 10.2.2018.

Kao odgovor na preterani optimizam povezan sa masovnim podacima, u poslednjih nekoliko godina su se u okrilju društvenih nauka razvile *kritičke studije podataka*. Iako, načelno, ne odbacuju korišćenje masovnih podataka u društvenim naukama, kritički nastrojeni autori dovode u pitanje teorijske, metodološke i etičke aspekte njihove upotrebe. Očekivano, najoštriju kritiku su izazvali pokušaji ateorijskog zasnivanja istraživanja masovnih podataka. Rob Kiččin ocenjuje da je po sredi javljanje novog empiricizma<sup>8</sup> koga odlikuje pogrešno uverenje da podaci mogu da govore sami za sebe, odnosno da teorija kao takva nije potrebna (Kitchin, 2014:2–3). Umesto toga, ovi autori ističu da podaci nisu „neutralni“ niti nešto što je „tamo negde“ nezavisno od ljudske interpretacije, kao i da nije tačno da je adekvatno tumačenje moguće od strane nekog ko nema specijalizovana znanja iz relevantne naučne oblasti (Andrejevic, 2013; Morozov, 2013; Kitchin, 2014:5; Kitchin & Lauriault, 2014; Lyon & Bauman, 2013). Koliko god da su podaci „veliki“ to ne znači da baze podataka nisu teorijski strukturisane na takav način da određene perspektive omogućavaju, a druge otežavaju (Bowker, 2014). Masovni podaci nikada nisu „sirovi“ već su uvek „skuvani“ (Gitelman, 2013) odnosno preoblikovani od strane društvenih aktera, a neretko tom upotrebom i „iskvareni“ (Boellstorff, 2013) pa samim tim ograničeno upotrebljivi (Michael & Lupton, 2016:107–8). Zapravo, u kontekstu prodora masovnih podataka, zaključuju ovi autori, razvoj društvene teorije nikad nije bio potrebniji (Gonzalez-Bailon, 2013).

Pored toga što su ustale u odbranu društvene teorije, kritičke studije ukazuju i na brojne negativne uticaje masovnih podataka na društvo. Jedna od čestih tema kritičkih studija jeste „velika podela“ (*big-data divide*) (Andrejevic, 2013). Manović (2011) pravi razliku između tri klase ljudi spram masovnih podataka: oni koji ih kreiraju svesno ili nesvesno ostavljajući svoje digitalne tragove, oni koji imaju sredstva da ih prikupljaju i oni koji imaju ekspertizu da ih analiziraju. Sa razvojem analitike masovnih podataka, moć je prešla u ruke onih koji ih poseduju, interpretiraju i na osnovu njih donose odluke (Andrejevic, 2013).

Pitanja nadzora, moći i privatnosti, takođe, zaokupljaju pažnju autora koji se bave kritičkom analizom masovnih podataka (van Dijck, 2014; Arda, 2015; Kitchin, 2014b; Crawford et al, 2014; Gabrys, 2014). Kako bi se razumele novine koje donosi nadzor zasnovan na masovnim podacima, razvijeni su koncepti poput masovnog nadgledanja (*'dataveillance'*) (Best, 2010) i likvidnog nadziranja (Lyon & Bauman, 2013). Na meti kritika su se našli i različiti epistemološki, metodološki i etički aspekti korišćenja masovnih podataka. U pitanje se, tako, dovodi validnost i reprezentativnost podataka prikupljenih putem interneta, transparentnost procesa uzorkovanja kao i narušavanje privatnosti istraživane populacije (Tufekci, 2014; Bone et al, 2016).

---

8 Prvu paradigmu odlikuju eksperimentalne nauke (empiricizam, deskripcija prirodnih fenomena, period pre Renesanse), drugu – teorijske nauke (modeliranje i generalizacije, period pre pronalaska računara), treću – računarske (simulacije kompleksnih fenomena, pre nastanka masovnih podataka) dok su za četvrtu karakteristične *eksplorativne nauke* koje se dovode u vezu sa revolucijom masovnih podataka (Hey et al, 2009 prema Kitchin, 2009:3).

Uz zaoštrene pozicije, u literaturi o masovnim podacima prisutna su i umerena stanovišta koja ističu da je posmatranje masovnih podataka kroz dihotomije teorija vs. podaci, ljudi vs. mašine, dedukcija vs. indukcija, neproaktivno i da maši nešto mnogo važnije a to je da su oba pristupa neophodna i međusobno kompatibilna (Mazzochi, 2015:1254; Ford, 2014; Kelling et al, 2009). Autori ukazuju da „mali“ odnosno dubinski, gusti, etnografski podaci mogu da dopune nalaze dobijene analizom „tankih“, „plitkih“ masovnih podataka, a da je saradnja između društvenih istraživača i računarskih eksperata obostrano korisna (Bone et al, 2016; Halford et al, 2013; Blok & Pedersen, 2014; Boellstorff, 2013; Ford, 2014; Bornakke & Due, 2018).

Bilo da se na masovne podatke gleda sa odobravanjem ili se novine koje donose kritički procenjuju, ono što je gotovo sigurno je da ih sociologija (više) ne može ignorisati. U radu objavljenom sedam godina nakon inicijalnog ukazivanja na nadolazeće pretnje za empirijsku sociologiju, Sevidž i Barouz ostaju pri stavu da je kriza savremene sociologije prisutna, ogromna i da zahteva odgovor, te da je neophodno da se sociolozi pripreme da intervenišu u svetu velikih podataka (Burrows & Savage, 2014:6). Na tom tragu, u radu publikovanom u prvom broju časopisa *Big Data and Society*, Kičin (2014a:10) ističe da postoji neodložna potreba za ispitivanjem implikacija masovnih podataka na društvo uopšte i društvene nauke posebno, a to je zadatak sa kojim su društvene nauke jedva tek započele. Imajući to u vidu, ono čime želimo da se bavimo u nastavku rada jeste pitanje u kojoj meri su se i na koji način masovni podaci našli kao tema u vodećim naučnim časopisima u oblasti društvenih nauka, sa posebnim osvrtom na sociologiju.

## Metod

Osnovna pretpostavka od koje se u ovom radu pošlo jeste da se u glavnim tokovima sociološke misli, za sada, ne poklanja previše pažnje temi masovnih podataka, dok se sociologija masovnih podataka razvija u okvirima specijalizovanih časopisa koji se nalaze na obodu *mainstream* sociologije. Druga pretpostavka je da literaturom dominiraju kritičke pozicije dok je primena analize masovnih podataka srazmerno slabo zastupljena u istraživačkoj praksi.

Kako bismo ispitali u kojoj meri i na koji način je tema masovnih podataka prisutna u *mainstream* akademskoj sociologiji, odlučili smo se da primenimo metod mešovite analize sadržaja na naučne radove koji se bave ovom temom a koji su objavljeni u vrhunskim sociološkim časopisima rangiranim na Thomson Rojters listi (*Thomson Reuters Impact Factor ssci list*). Iako ne predstavljaju potpunu sliku naučnog rada u okviru sociologije, može se reći da ove publikacije jesu jedan od najuticajnijih kanala formalne akademske komunikacije sa širom naučnom javnosti, te da legitimišu intelektualni razvoj i *mejnstrimuju* određene naučne pristupe ukazujući na relativnu važnost određene teme ili polja za društvene nauke<sup>9</sup>.

9 O primeni metode analize sadržaja u izučavanju naučnih publikacija videti detaljnije u: Manić, 2017.

U istraživanju su u obzir uzeti svi tekstovi na engleskom jeziku koji su objavljeni u sociološkim časopisima koji se nalaze na listi Tomson Rojters (2016<sup>10</sup>) u periodu 2000–2017. godine. Dvehiljadita godina je odabrana kao početna tačka s obzirom da se, prema nalazima istraživanja nastanka pojma<sup>11</sup> „big data“, ovaj koncept u akademskim razmatranjima prvi put pojavljuje u periodu između 1998–2000. godine. Selektovani časopisi (tabela 1) imaju težište u sociologiji ali su otvoreni i za teme i pristupe svojstvene drugim društvenim naukama kao što su antropologija, kulturne studije, ekonomija, društvena geografija, mediji, psihologija, semiotika i drugo. Treba napomenuti da su pored časopisa opšte sociološke teorijske i metodološke orijentacije, analizom obuhvaćeni i oni koji se fokusiranije bave odnosom društva i novih informaciono-komunikacionih tehnologija. Kao kriterijum pretrage kroz elektronske baze podataka i odabir tekstova za analizu je korišćen pojam „big data“ kao ključna reč<sup>12</sup>. U obzir su uzeti samo celoviti tekstovi – originalni naučni radovi i pregledni članci, dok su prikazi knjiga, komentari i dr. izostavljeni iz analize. Izdvojeni su sociološki časopisi sa engleskog govornog područja, preciznije iz Sjedinjenih Američkih Država i Velike Britanije. Odabir prostornog okvira je načinjen imajući u vidu kako uticajnost koju časopisi i sama akademska sociologija iz ovih zemalja imaju na svetsku sociologiju, tako i mogućnost analize tekstova na engleskom jeziku<sup>13</sup>. Na osnovu pomenutih kriterijuma („big data“ kao ključna reč, vremenski period 2000–2017; *Thomas Reuters Impact factor ssci list 2016*, engleski jezik) identifikovano je ukupno 77 tekstova objavljenih u 20 naučnih časopisa (u prilogu se nalazi spisak svih analiziranih radova).

---

10 Usled nedostupnosti podataka za 2017. godinu (istraživanje je sprovedeno početkom 2018. godine), vodili smo se pretpostavkom da će se velika većina a verovatno i svi časopisi naći na listi i 2017. godine, a tekstove iz 2017. godine nismo isključili iz analize usled velikog broja relevantnih radova koji je objavljen upravo u toj godini.

11 <https://economics.sas.upenn.edu/pier/working-paper/2012/origins-and-development-term-“big-data”> (pristupljeno 15.1.2018. godine.)

12 Imajući u vidu da su, kako bi podigli šanse da se u pretragama nađe baš njihov rad, pojedini autori skloni da spisku ključnih reči dodaju i zvučne pojmove kojima se u radu tek uzgredno bave, detaljno smo pregledali apstrakte a po potrebi i čitave radove kako bi se odstranili oni koji se ne bave fenomenom masovnih podataka. Na taj način je iz dalje analize eliminisano ukupno 7 tekstova. S druge strane, postoji mogućnost da se izvan analize našao izvestan broj radova koji, iako se bave pitanjem masovnih podataka, to ne ekspliciraju korišćenjem pojma „big data“. Takav je, svakako, slučaj sa ranije pomenutim tekstom Majka Sevidža i Rodžera Barouza (2007). Ipak, verujemo da je ovakvih radova relativno malo te da njihovo izostavljanje značajnije ne utiče na dobijene nalaze.

13 Kako bi se izbegle poteškoće u analizi tekstova koji nisu objavljeni na engleskom jeziku, iz pretrage su isključeni višejezični časopisi.



Tabela 1 Pregled časopisa i broja objavljenih radova sa ključnom reči „big data“

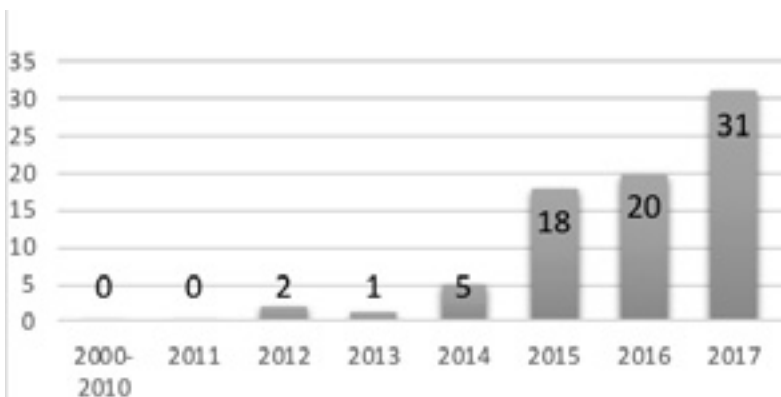
Br.	Naziv časopisa	Zemlja	Impakt faktor (2016)	Broj članaka
	Acta Sociologica	VB	1.225	0
	American Journal of Sociology	SAD	3.088	1
	American Sociological Review	SAD	4.400	1
	Annual Review of Sociology	SAD	5.404	3
	Contemporary Sociology	SAD	1.857	0
	Critical Sociology	SAD	0.973	1
	Current Sociology	VB	1.780	0
	European Journal of Social Theory	VB	1.278	3
	European Sociological Review	VB	2.062	0
	<b>Field Methods</b>	<b>SAD</b>	<b>1.346</b>	0
	Information Society	SAD	1.558	3
	Information, Communication and Society	VB	2.692	14
	Journal of Mixed Methods Research	SAD	1.600	0
	Journal of Sociology	VB	0.841	0
	Media, Culture and Society	VB	1.128	5
	New Media and Society	VB	4.180	7
	Qualitative Inquiry	SAD	1.159	0
	Qualitative Research	VB	2.012	3
	Science, Technology & Human Values	SAD	2.907	1
	Social Epistemology	VB	0.955	1
	Social Research	SAD	0.275	0
	Social Science Computer Review	SAD	2.293	B
	Social Science Quarterly	SAD	0.849	0
	Social Science Research	SAD	1.327	8
	Social Studies of Science	SAD	2.256	0
	Society	SAD	0.500	0
	Sociological Forum	SAD	0.919	0
	Sociological Inquiry	SAD	0.680	0
	Sociological Methodology	SAD	2.667	1
	Sociological Methods and Research	SAD	3.604	0
	Sociological Perspectives	SAD	1.133	0
	Sociological Research Online	VB	0.519	1
	Sociological Spectrum	SAD	0.421	0
	Sociological Theory	SAD	1.545	0
	Sociology	VB	1.963	3
	<b>The Annals of the American Academy of Political and Social Science</b>	<b>SAD</b>	<b>2.118</b>	<b>9</b>
	The British Journal of Sociology	VB	1.963	0
	The Sociological Quarterly	SAD	1.237	0
	The Sociological Review	VB	1.349	3
	Theory, Culture and Society	VB	1.904	3

## Nalazi istraživanja

Pretragom na osnovu postavljenih kriterijuma za vrstu časopisa, izdvojeno je ukupno 40 naučnih publikacija od čega 14 britanskih i 26 američkih. U polovini (20) časopisa je objavljen najmanje jedan rad koji ima ključnu reč „big data“. Od ukupno 77 članaka koji se bave fenomenom masovnih podataka, 34 je objavljeno u američkim, a 43 u britanskim časopisima. Imajući u vidu relativni odnos broja analiziranih časopisa i ukupnog broja radova, može se reći da u britanskim izdanjima postoji nešto veće interesovanje za temu masovnih podataka. Gotovo polovina (37) tekstova je objavljena u sedam časopisa koji su specijalizovani za razmatranje odnosa društva i informaciono-komunikacionih tehnologija, među kojima prednjači *Information, Communication and Society* sa 14 objavljenih radova. U časopisima posvećenim temama opšte sociologije objavljeno je 40 radova. Prema broju objavljenih tekstova u časopisima „opšte“ orijentacije se izdvojio *The Annals of the American Academy of Political and Social Science* sa 9 tekstova, a među publikacijama fokusiranim na metodologiju društvenih nauka *Social Science Research* sa 8 radova.

Ukoliko se u vidu ima da je u posmatranom vremenskom periodu u selektovanoj grupi časopisa objavljeno preko deset hiljada naučnih radova, već se na osnovu samog broja tekstova koji se bave pitanjem masovnih podataka, može zaključiti da ovoj temi nije posvećena značajnija pažnja<sup>14</sup>. Kada se, potom, od ukupnog broja tekstova koji se bave pitanjem masovnih podataka, oduzmu oni objavljeni u specijalizovanim časopisima, sasvim jasno postaje da je u sociološkim debatama opštijeg karaktera tek izuzetno pažnja posvećena pitanju masovnih podataka.

Grafikon 1 Broj objavljenih tekstova koji se bave masovnim podacima prema godini izdanja



14 Okviran broj radova se dobija množenjem broja tekstova objavljenih u jednom izdanju časopisa (koji se u proseku kreće oko 5), broja izdanja u toku godine (računali smo da je prosek 3), broja godina i broja časopisa. Tom računicom, u periodu od 2000–2017. godine, procena je da je u pomenutim časopisima objavljeno približno 11,000 članaka, odnosno da je u posmatranom periodu objavljeno svega 0,7% radova koji se bave temom masovnih podataka.

Pregled raspodele frekvencija tekstova koji se bave pitanjem masovnih podataka prema godini izdanja (grafikon 1) pokazuje da ova tema, iako u krugovima izvan društvenih nauka prisutna već krajem 20. veka, počinje da figurira tek od 2012. godine. Naime, u toku prve dekade 21. veka, u posmatranim časopisima nije objavljen ni jedan naučni članak koji se eksplicitno bavi masovnim podacima iako, verovatno, postoji po neki rad, kao što je ranije pomenuti tekst Majka Sevidža i Rodžera Barouza (2007) koji se, iako ne koristi ovaj pojam, zapravo bavi temom masovnih podataka. Na osnovu podataka prikazanih grafikonom 1, može se zaključiti da tema masovnih podataka počinje da budi veće interesovanje društvenih naučnika tek u poslednje 3–4 godine, sa primetnim porastom broja radova iz godine u godinu. U 2017. godini je objavljeno 40% ukupnog broja tekstova, što navodi na zaključak da će se ovoj temi verovatno više pažnje posvetiti u narednom periodu.

### *Teme*

Radove koji se bave fenomenom masovnih podataka smo svrstali u tri opšte kategorije: 1. tekstovi koji su posvećeni teorijskom razmatranju upotrebe masovnih podataka (38); 2. radovi koji se bave metodološkim aspektima analize masovnih podataka (17); 3. radovi u kojima je primenjena analiza masovnih podataka (22).

#### **1. Teorijsko promišljanje masovnih podataka, njihovog društvenog uticaja i primene u društvenim naukama**

U okviru tekstova koji teorijski promišljaju i preispituju značaj, ulogu i mesto analize masovnih podataka u društvenim naukama, izdvojile su se dve podteme:

- a. Implikacije masovnih podataka na sociološku teoriju i istraživačku praksu (17 članaka)
- b. Društvene implikacije masovnih podataka (21 članak)

##### **a. Implikacije masovnih podataka na društvene nauke**

U ovom segmentu analize, fokus smo stavili na vrednosnu usmerenost tekstova praveći distinkciju između pretežno kritičkih, afirmativnih i tekstova koji zadržavaju neutralan stav prema upotrebi masovnih podataka u društvenim naukama (tabela 2). Pod kritički usmerenim naučnim radovima podrazumevamo one u kojima se ističu rizici i nedostaci korišćenja masovnih podataka u društvenim naukama, dok se pod afirmativnim smatraju oni u kojima se ili zagovara korišćenje masovnih podataka ili odgovara na upućene kritike. Pored ova dva vrednosna usmerenja, izdvojili smo i tekstove koji imaju neutralan odnosno „konstruktivan“ pristup, što znači da u njima autori nastoje da na nepristrasan način utvrde prednosti i mane korišćenja masovnih podataka u sociologiji i da pronađu načine da se uočeni nedostaci prevaziđu.

Tabela 2. Vrednosna usmerenost tekstova prema tipu časopisa

	Afirmativni ton	Kritički ton	Neutralni ton / konstruktivni pristup	Ukupno
Časopisi opšte sociološke orijentacije	2	3	5	10 (59%)
Specijalizovani časopisi	1	3	3	7 (41%)
Ukupno	3 (18%)	6 (35%)	8 (47%)	17 (100%)

Kao što se na osnovu podataka prikazanih u tabeli 2 može videti, nešto je veći udeo radova (59%) koji se bave ispitivanjem implikacija masovnih podataka na sociologiju i druge društvene nauke koji je objavljen u časopisima opšte sociološke orijentacije. Gotovo u polovini (48%) tekstova je zastupljen konstruktivan pristup, u jednoj trećini je naglašen kritički ton, dok je najmanje radova koji imaju pozitivan (afirmativan) prizvuk.

Među *kritički usmerenim tekstovima*, posebno se ističe rad dane bojd<sup>15</sup> i Kejt Kraford „*Critical Questions for Big Data*“ (BOYD & CRAWFORD, 2012<sup>16</sup>) u kome autorke kritički preispituju ulogu i značaj masovnih podataka. Na ovaj rad, za koji se već sad može reći da pripada klasicima digitalne sociologije, poziva se preko polovine (42) svih analiziranih tekstova. Autorke su kroz 6 tvrdnji o masovnim podacima koje iznose u ovom radu, u velikoj meri zacrtale osnovne pravce razmišljanja i debate u sociologiji masovnih podataka:

1. *Masovni podaci menjaju definiciju znanja* odnosno menjaju načine na koji se pristupa procesu istraživanja, pronalaženja informacija te kategorizacije društvene stvarnosti.
2. *Tvrdnje o objektivnosti i tačnosti su pogrešne* – iako može delovati da masovni podaci nude humanističkim naukama „objektivnost“ prirodnih nauka, u realnosti je rad sa masovnim podacima i dalje subjektivan.
3. *Veći podaci nisu uvek i bolji podaci* – veliki setovi podataka su često nepouzdati i nepotpuni, a nedostaci se multiplikuju kada se kombinuje više izvora podataka. Sama veličina seta podataka ne znači i da je on reprezentativan. Prilikom analize masovnih podataka, istraživači moraju biti upoznati sa njihovim potencijalnim pristrasnostima, kao i sa pristrasnošću u njihovoj interpretaciji. Pored toga, masovni podaci mogu dovesti i do situacije da se uočavaju obrasci i pravilnosti tamo gde ih, zapravo, nema.
4. *Izvađeni iz konteksta, podaci gube značenje* – odnosno, nikako se ne može tvrditi da podaci „govore sami za sebe“. Uklapanjem u matematičke modele, podaci bivaju izvađeni iz konteksta, čime gube na sadržaju što dovodi u pitanje mogućnost adekvatne interpretacije.
5. *Samo zato što je dostupno ne znači i da je etički ispravno* – korišćenje masovnih podataka ima ozbiljne etičke implikacije. To što su neki podaci

15 Autorka je promenila lično ime tako da se ime i prezime pišu malim početnim slovima <http://scratchtap.com/danah-boyd-and-capitalization/> pristupljeno 10.1.2017.

16 Velikim slovima su ispisane reference koje se odnose na analizirane naučne radove navedene u Prilogu.

javno dostupni ne znači da su njihovi kreatori svesno pristali na njihovu obradu i korišćenje u marketinške ili bilo koje druge svrhe.

6. *Ograničen pristup masovnim podacima stvara nove digitalne podele* – jedino kompanije u čijem se vlasništvu nalaze društveni mediji imaju potpun pristup podacima. Drugima, koji su u prilici da izdvoje izuzetno visoke svote novca, omogućen je ograničeni pristup. Međutim, velika većina istraživača nema mogućnost da radi na ovim podacima.

Na tekst Boyd i Kraford se nadovezuje nekoliko radova u kojima se dalje razrađuju kritike upućene na račun analitike masovnih podataka (BROCK, 2015; CHAN, 2015; PAPACHARISSI, 2015; SEEVER, 2015; COULDRY, 2014). Kada je u pitanju dekontekstualizacija masovnih podataka, skreće se pažnja da nije problem samo nedostatak konteksta već i šta se pod kontekstom podrazumeva, te da bi i taj aspekt trebalo problematizovati prilikom bavljenja masovnim podacima (SEEVER, 2015). Autori ističu da opasnost ne leži samo u novim digitalnim podelama već i u saradnji sa komercijalnim sektorom koja preta da korporativnu kulturu, koja zanemaruje etičke standarde svojstvene društvenim istraživanjima, uvede u društvene nauke. Masovni podaci predstavljaju fundamentalni izraz savremenog kapitalizma koji nikako ne bi trebalo pozivati u sociologiju (CHAN, 2015; FRADE, 2016). Nik Koldri (2014) nastoji da demistifikuje mit da veliki podaci nude superiornije oblike znanja u odnosu na saznanja koja se mogu dobiti primenom standardnih istraživačkih metoda. Prema njegovom mišljenju, masovni podaci dekontekstualizuju društveno, frakturišu diskurse i dehumanizuju podatke o ljudima. On se zalaže za povratak veberijanskom interpretativnom pristupu kao kontrateži prediktivnoj analitici masovnih podataka.

U manjini analiziranih tekstova prevladava *afirmativni ton*. U ovim radovima se ističe da masovni podaci pružaju mogućnost odgovora na važna pitanja do kojih se ranije nije moglo doći usled ograničenosti standardnih socioloških metoda i tehnika. Mogućnosti se pre svega vide u osvetljavanju jedne od centralnih dilema sociologije – odnos aktera i strukture. Ističe se da masovni podaci mogu da pruže vredne uvide u procese strukturacije, stvaranja kolektivnog iz individualnog, te povratnog uticaja kolektivnog na individualno. Autori kritikuju sociologe što se ne prihvate analize masovnih podataka, smatrajući to propuštenom prilikom. Ipak, uvažavaju postojanje nejednake mogućnosti pristupa masovnim podacima (za koju veruju da se može prevazići povezivanjem sa partnerima iz poslovne sfere), te različitih etičkih ograničenja njihovog korišćenja (HALAVAIS, 2015; HOFFERTH et al, 2017; LIN, 2015).

Na kraju, posebno su interesantni tekstovi u kojima se nude *rešenja* za neke od uočenih problema upotrebe masovnih podataka u sociologiji. Kao način da se „pomire“ tradicionalni i inovativni načini prikupljanja i analize podataka, autori vide *fuziju* (LAZER & RADFORD, 2017), *integrativnu analizu* (HESSE, 2015; HOFFERTH et al, 2017; LAZER & RADFORD, 2017) odnosno razvoj *simfonijskih društvenih nauka* (HALFORD & SAVAGE, 2017). Simfonijski pristup podrazumeva traženje odgovora na ključna društvena pitanja kombinovanjem različitih metoda (metodološki pluralizam) i izvora podataka (mali, etnografski podaci i

veliki podaci). Bez nepotrebne fetišizacije jedne vrste podataka, ključno pitanje jeste kako dati složaj (*assemblage*) podataka može da se iskoristi u pronalaženju odgovora na pitanje koje istraživač postavlja (HALFORD & SAVAGE, 2017:1140).

#### b. Društvene implikacije masovnih podataka

Među radovima u kojima se preispituju društvene implikacije masovnih podataka, najveći broj (14) je posvećen temi *nadzora* (moći i upravljanja) i *privatnosti* u digitalno doba. U analiziranim tekstovima preovlađuje kritički pristup, a kao ključni koncepti se izdvajaju biopolitika, masovni nadzor (*dataveillance*), sistemi nadzora, prediktivna analitika masovnih podataka i društvena moć algoritama, posmatrani u širem kontekstu rasta državne regulacije i korporativne kontrole. Autori ističu da sa usponom masovnih podataka uporedo dolazi do sve većeg prodora različitih tehnika nadzora u svakodnevni život (BRAYNE, 2017:1). Kroz svakodnevne aktivnosti kao što je kupovina namirnica, plaćanje računa ili istraživanje medicinskih simptoma na internetu, potrošači iza sebe ostavljaju digitalne tragove koje kompanije beleže i agregiraju sa podacima drugih korisnika. Sve razvijenija tehnologija omogućava praćenje sve većeg broja korisnika / potrošača na sve efikasniji način, što obezbeđuje i veću mogućnost nadzora i uticaja na ponašanje kroz različite forme personalizovanog marketinga (POPESCU & BARAH, 2013; MARTIN, 2015). U takvom kontekstu, ističe se potreba za ponovnim promišljanjem različitih aspekata privatnosti s obzirom na sve sofisticiranije načine njenog ugrožavanja (MAI, 2016).

Pored tekstova koji se bave pitanjem nadzora i privatnosti, u dva rada autori stavljaju masovne podatke u širi kontekst razmatranja *javne sfere*. Kao problem korišćenja masovnih podataka se izdvaja to što takva praksa može da dovede do uspostavljanja "tiranije većine" nad medijskim sadržajem i javnim debatama, sa tendencijom stvaranja globalne kulture homogene potrošnje i ponašanja. Pokreti koji se zalažu za otvoren pristup podacima, zaštitu kreativnih javnih dobara, podsticanje digitalnog deljenja i razvoj softvera otvorenog pristupa, vide se kao kontrateža moći zasnovanoj na podacima u privatnom vlasništvu (HARPER, 2016).

Jedan manji broj radova (3) je posvećen razmatranju *etičkih aspekata* upotrebe masovnih podataka. Pristupajući iz kritičke perspektive, autori ukazuju na nove etičke probleme koje sa sobom donosi prikupljanje, obrada i korišćenje masovnih podataka. Kao posebno problematičan se navodi proces donošenja odluka zasnovan na algoritmima (ZARSKY, 2016) te se ukazuje na različite zloupotrebe prediktivne analitike i negativne društvene posledice preteranog oslanjanja na statističke podatke i modele (PAULEEN et al, 2016). Autori ističu i to da je jedno od ključnih etičkih pitanja: ko ima pristup masovnim podacima, pod kojim uslovima i za koje potrebe? Za sada, masovni podaci samo doprinose rastu moći i bogatstva onih koji su već moćni i bogati (QUI, 2015).

Na kraju, među analiziranim tekstovima su se našli i oni koji se fokusiraju na posebne teme (ukupno 2) kao što su masovni podaci u kontekstu razvoja muzičke industrije i digitalnog kapitalizma (ARDITI, 2017) te pitanje rekonceptualizacije arhiviranja u kontekstu digitalnih medija (POTZSCH, 2017).

## 2. Metodološka razmatranja upotrebe masovnih podataka u društvenim naukama

Radove koji preispituju različite metodološke aspekte generisanja, prikupljanja i analize velikih podataka (ukupno 17), podelili smo na one u kojima dominira afirmativni ton (3), negativan / kritički stav (6) i neutralan („konstruktivan“) pristup (8). Treba napomenuti da, iako je u trećini radova zastupljen kritički pristup, ni u jednom radu se analiza masovnih podataka ne odbacuje u potpunosti. Radovi sa naglašeno afirmativnim tonom su fokusirani na posebne slučajeve primene analize velikih podataka (npr. mogućnost primene mašinskog učenja u društvenim naukama ili analize članaka u Vikipediji) a ne na opšte aspekte korišćenja ove vrste podataka.

Kao *prednosti* korišćenja masovnih podataka u društvenim istraživanjima u analiziranim tekstovima su se izdvojile: *veličina i obuhvatnost* – odnosno mogućnost realizacije istraživanja na izuzetno velikom uzorku populacije / čitavoj populaciji; *brzina generisanja* usled koje mogu da posluže kao „digitalni dimni signali“ važnih društvenih događaja. Kao kvalitet se vide i relativno *niski troškovi* prikupljanja i obrade u odnosu na tradicionalne metode kao i to što ovi podaci ne predstavljaju iskaze o aktivnostima već *zapise stvarnog delanja*. U jednom broju radova se ukazuje na *nove teme, pristupe i mogućnosti istraživanja* u društvenim naukama koje pružaju masovni podaci. Pored platformi za društveno umrežavanje koje predstavljaju već relativno prihvaćen izvor podataka za analizu društvenog ponašanja, kao relevantni se izdvajaju i različiti administrativni podaci velikog obima, ali i neki tipovi podataka koji do sada nisu bili bliski društvenim istraživačima kao što su masovni genomske podaci ili vizuelni podaci generisani putem *Google street view*. Pored toga, ističu se i prednosti inovativnih istraživačkih metoda i tehnika kao što su mašinsko učenje, napredni statistički modeli, kompjuterski podržana analiza sadržaja i dr.

Kada su u pitanju *nedostaci* korišćenja masovnih podataka, u radovima je najčešće problematizovana *pristrasnost* u analizi i vizuelizaciji masovnih podataka, kao i *greške merenja i uzorkovanja*, upitna *validnost podataka*, *nestandardizovanost analitičkih tehnika* te (ne)mogućnost *generalizacije* zaključaka, pogotovo u slučaju podataka prikupljenih putem *društvenih mreža*. Potom se kao značajna izdvojila tema *neetičnosti* korišćenja podataka o pojedincima koji nisu pristali na beleženje i analizu svog ponašanja. Zatim su istaknuti problemi *dekontekstualizacije* i *simplifikacije* ljudskog ponašanja do kojih dolazi prilikom analize masovnih podataka. Kao značajno se otvorilo i pitanje premeštanje moći od društvenih istraživača ka računarskim stručnjacima i „ekspertima za podatke“. Takođe, u radovima je razmatran problem *nejednake mogućnosti pristupa* i korišćenja velikih setova podataka koji se nalaze u privatnom vlasništvu, a istaknuta je i potreba za ovladavanjem novim tehnikama prikupljanja, menadžmenta i analize podataka.

U tekstovima u kojima se na neutralan način pristupa ispitivanju pretnji i mogućnosti koje sa sobom donosi analiza velikih podataka, primetno je razvijanje *modela* prevazilaženja uočenih nedostataka. Autori su mišljenja da naučna zajednica treba da ustanovi jednako visoke *standarde analize* velikih

podataka kao što je to učinila sa „malim“ podacima. U tom smislu, ističe se da je neophodno *povezati* analizu masovnih podataka sa društvenom statistikom i drugim oblicima znanja razvijenim u okviru društvenih nauka kako bi se unapredili analitički procesi i algoritamske tehnike koji se koriste u analizi velikih podataka.

Kao jedan od načina korekcije nedostataka masovnih podataka, vidi se *kombinovanje istraživanja masovnih podataka i kvalitativnih istraživanja* te kreiranje komplementarnih setova podataka. Pored isticanja važnosti *lokalnog "čitanja"* masovnih podataka sa naglaskom stavljenim na ispitivanje konteksta njihovog nastanka i upotrebe, autori ukazuju i na značaj *participacije* predstavnika onih društvenih grupa čiji se digitalni tragovi prate i analiziraju. *Razvoj administrativnih centara*, koji bi omogućili snižavanje troškova analize masovnih podataka i obezbedili reprodukciju istraživanja, još jedno je od rešenja. Autori analiziranih tekstova, ukazuju i na nužnost *prevazilaženje disciplinarnih granica* te povezivanje društvenih naučnika sa informatičarima i ekspertima za podatke.

### 3. Praktična primena analize masovnih podataka u sociološkim istraživanjima

U 22 članka je primenjena neka od metoda analize masovnih podataka. Prilikom analize ovih tekstova, u obzir su uzeti sledeći elementi: istraživačka tema, akademska afilijacija autora, izvori podataka i njihova ograničenja, primenjeni metodi i korišćeni softveri.

Kao najzastupljenija se pokazala *tema* komunikacije kroz društvene medije (15 radova) u političkom procesu (9 radova) ili u toku kriznih situacija (prirodne katastrofe, masovna ubistva) (6). Četiri rada su posvećena primeni analize masovnih podataka u razmatranju pitanja koja spadaju u domen opšte sociologije: stratifikacija, razvoj društvenih normi, te uticaj digitalnih tehnologija na različite oblike društvenosti. Na kraju, tema tri rada – ispitivanje karakteristika susedstava; pitanje individualne teritorijalnosti i odnosa prema javnom prostoru; testiranje hipoteze „spornih granica“ između različitih susedstava – spada u područje sociologije prostora.

Uzimajući u obzir teme istraživanja zasnovanih na analizi masovnih podataka publikovanih u vrhunskim sociološkim časopisima, može se primetiti podzastupljenost pojedinih oblasti koje se pojavljuju u široj literaturi o masovnim podacima. Kao podzastupljene se izdvajaju prostorne teme koje su, usled mogućnosti koje su otvorili GIS i GPS sistemi, relativno česte u sociologiji masovnih podataka (o čemu svedoče i tri temata časopisa *Big Data & Society*<sup>17</sup> posvećena prostornim aspektima), kao i teme kriminaliteta te kulturoloških aspekata upotrebe masovnih podataka (Kitchin, 2013; Kitchin, 2014b; Bauder, 2016; Chan & Bennett Moses, 2016; Singer, 2016).

*Analiza akademske afilijacije* autora je pokazala da je približno polovina članaka (10) ishod timskog rada naučnika društvenog usmerenja i stručnjaka iz oblasti računarstva. Potom, jedna trećina radova (7) je proizvod rada tima istraživača iz oblasti društvenih nauka, među kojima, po pravilu, barem jedan

17 <http://journals.sagepub.com/page/bds/collections/index> (pristupljeno 2.6.2018.)



a često i više koautora ima diplomu i iz matematike ili statistike. Na kraju, 5 radova je proisteklo iz samostalnog istraživačkog rada naučnika društvenog usmerenja. Imajući u vidu da je, prema podacima *Thomson Reuters*, oko 40% radova iz oblasti društvenih nauka delo jednog autora (King, 2013 prema Metzler et al, 2016:1–2), ovaj poslednji nalaz govori u prilog tome da analiza masovnih podataka predstavlja, za sad, izuzetno veliki izazov za naučnike društvenog usmerenja. Treba napomenuti da se, u naučnim radovima u kojima se kao koautori ne pojavljuju stručnjaci za računarske tehnologije, kao analitički softveri uglavnom koriste platforme koje su relativno jednostavne za korišćenje (Google n-gram ili Twitter i Facebook API) odnosno ne zahtevaju posebno predznanje iz oblasti informatike.

Među *izvorima* velikih setova podataka koji su korišćeni u analizi, izdvajaju se sledeći: 1. analiza društvenih medija (Twitter, Facebook) i veb sajtova (15); 2. analiza administrativnih podataka (4); 3. analiza Google books n-gram (3). Kao izvori administrativnih podataka, autorima su poslužile baze zabeležbi potraživanja od gradskih servisa Bostona i Njujorka (300,000–600,000 u Bostonu, odnosno 4,7 miliona u Njujorku), kao i 15 miliona zabeležbi Federalne izborne komisije SAD o individualnim donacijama (HERVIG, 2016; O'BRIEN et al, 2015; O'BRIEN, 2015, LEGEWIE & SCHAEFFER, 2016). Kada je u pitanju *Google n-gram* korpus knjiga koji je korišćen u tri istraživanja, njegova veličina se kreće između 5 i 8 miliona jedinica (u zavisnosti od kriterijuma selekcije koje su postavili autori). U jednom radu (MUTHY, 2017) je kombinovana analiza *Google n-gram*-a sa analizom 235 miliona tvitova, a u drugom administrativni podaci Federalne izborne komisije i slučajni uzorak od 113 hiljada tvitova (MCKALVEY et al, 2014). Kao osnovni ili pomoćni izvor masovnih podataka, Tviter je korišćen u 10 a Fejsbuk u 3 naučna rada. Pored toga, analiza je primenjena i na podatke prikupljene preko posebnih veb platformi (2).

Kao *ograničenja i nedostake* analize masovnih podataka na koje su nailazili u istraživanju, autori su najčešće isticali varijabilni kvalitet podataka, potrebu čišćenja podataka kombinovanjem posebnih algoritama i naknadne manuelne provere; pristrasnost i iskrivljenost podataka, upitnu reprezentativnost odnosno mogućnost generalizacije te, u pojedinim slučajevima, neophodnost dopune drugim metodama analize primenjene na „malim“ podacima.

U istraživanjima su korišćene raznovrsne analitičke *metode*: kompjuterski podržana analiza sadržaja (13), mrežna analiza (5), socio-prostorna analiza (5) i ekonometrijska analiza (1). U 8 radova je korišćena kombinacija više metoda te povezivanje „malih“ i „velikih“ setova podataka. Autori analiziranih tekstova su naveli sledeća *softverska rešenja / programske jezike* koje su koristila u analizi: Python (3 rada); R (2); SQL (1); Google n-gram (3); Twitter Streaming Application Programming Interface (API) (7), Twitter Firehose (1) i Gardenhose (1); Google developer API (1), Facebook API (2); UCINET softver za mrežnu analizu (1); specijalizovani algoritmi (2); Big data tool Wisdom (1); Wordpress plug-in (1), Admonitor (Java Script) (1). Raznovrsnost metodoloških i tehnoloških rešenja, između ostalog, ukazuje na nepostojanje jedinstvenog standarda te na vrlo dinamični razvoj softvera za obradu masovnih podataka u društvenim naukama.

## Zaključna razmatranja

Iako masovni podaci proizvode tektonska pomeranja u prostoru kojim su nekada suverno vladale društvene nauke, može se reći da je tema masovnih podataka uglavnom zanemarena u okvirima glavnih tokova sociološke misli. Radove koji se bave masovnim podacima objavljene do kraja 2017. godinom odlikuje kritičko preispitivanje prednosti i nedostataka primene analize masovnih podataka u društvenim naukama, uz nastojanje da se razviju modeli koji bi omogućili prevazilaženje uočenih problema. Izuzetno mali broj sociologa se, do sada, upustio u sam proces analize masovnih podataka, a ta pregnuća su uglavnom deo zajedničkog rada multidisciplinarnih timova u koje su, po pravilu, uključeni i stručnjaci iz oblasti računarstva.

Jedno od pitanja koje je otvorila studija masovnih podataka jeste to na koji način bi trebalo pristupiti analizi masovnih podataka imajući u vidu kako standarde rada sa „malim“ podacima tako i teorijska, metodološka i etička ograničenja upotrebe masovnih podataka. Iako proces standardizacije smatramo važnim, mišljenja smo da ranu, eksplorativnu fazu razvoja analize masovnih podataka u sociologiji ne bi trebalo opteretiti strogim pravilima već da bi trebalo omogućiti da se nesmetano razvijaju i proizvode nova kreativna rešenja i inovativne prakse.

Pregled literature koja se bavi masovnim podacima, pokazuje da su ključne dileme savremene sociologije, pre svega pitanje njene krize, zastupljene i u diskusiji o masovnim podacima (Gouldner, 1970; Budon, 1980; Stinchcombe, 1994; Fuller, 2006; Spasić, 2012). Osvrćući se na krizni diskurs prisutan u razmatranju masovnih podataka, rekli bismo da je empirijska sociologija u svojevrsnoj razvojnoj krizi koja, ako se razreši na pravi način, može dovesti do prelaza u jednu novu, inspirativnu fazu. U tom smislu bi sociolozi zainteresovni za masovne podatke trebalo da težište razmatranja pomere sa analize mogućnosti i prepreka na način na koji su to postavile Boyd i Kraford (2012), ka pronalaženju načina na koji bi se, u teorijskom, metodološkom i etičkom pogledu, najbolje moglo pristupiti analizi masovnih podataka, imajući u vidu disciplinarnu i tematske specifičnosti. Iako sociolog može (i treba) kritički da pristupi temi masovnih podataka, opasnost leži u nekritičkom odbacivanju masovnih podataka te povlačenju u sigurnost poznatog sveta društvenih nauka. Na kraju krajeva „sociolozi bi trebalo da se uključe u bavljenje masovnim podacima, jer se to već dešava, sa njima ili bez njih“ (Halford & Savage, 2017:1134).

## Literatura

- Anderson, C. (2008). The end of theory: the data deluge makes the scientific method obsolete. Wired 23 June. [http://www.wired.com/science/discoveries/magazine/16-07/pb\\_theory](http://www.wired.com/science/discoveries/magazine/16-07/pb_theory) (pristupljeno 1.2.2018.)
- Andrejevic, M. (2013). *Infoglut: How Too Much Information Is Changing the Way We Think and Know*. New York, NY: Routledge.
- Ardu, C. (2015). The signature of security: big data, anticipation, surveillance. *Radical Philosophy* 191: 21–8.

- Back, L. (2012). Live sociology: social research and its futures. *The Sociological Review* 60 (S1): 18–39.
- Back, L., Puwar, N. (2012). A manifesto for live methods: provocations and capacities. *The Sociological Review* 60 (S1): 6–17.
- Baruh, L, Popescu, M. (2017). Big data analytics and the limits of privacy self-management. *New Media and Society* 19 (4): 579–596.
- Bauder, M. (2016). Thinking about measuring Auge's non-places with big data. *Big Data & Society* July–December 2016: 1–5.
- Bello-Orgaz, G, Jung, J., Camacho, D. (2015). Social big data: Recent achievements and new challenges. *Information Fusion* <http://dx.doi.org/10.1016/j.inffus.2015.08.005> (pristupljeno 1.6.2018)
- Best, K. (2010). Living in the control society: Surveillance, users and digital screen technologies. *International Journal of Cultural Studies* 13(1): 5–24.
- Blok A, Pedersen MA. (2014). Complementary social science? Quali-quantitative experiments in a Big Data world. *Big Data & Society* 1(2): 1–6.
- Boellstorff, T. (2013). Making big data, in theory. *First Monday* 18(10). <http://firstmonday.org/ojs/index.php/fm/article/view/4869/3750> (pristupljeno 1.2.2018.)
- Bone, J, Emele, C.D, Abdul, A, Coghill, G, Pang, W. (2016). The social sciences and the web: From 'Lurking' to interdisciplinary 'Big Data' research. *Methodological Innovations* 9: 1–14.
- Bornakke, D., Due, B.L. (2018). Big–Thick Blending: A method for mixing analytical insights from big and thick data sources. *Big Data & Society* January–June 2018: 1–16.
- Boudon, R. (1980). *The Crisis in Sociology: Problems of Sociological Epistemology*. London and Basingstoke: Palgrave Macmillan.
- Bowker, G. (2014). The Theory/Data Thing– Commentary. *International Journal of Communication* 8: 1795–1799.
- boyd, d., Crawford, K. (2012). Critical Questions For Big Data. *Information, Communication & Society* 15(5): 662–67.
- Burrows, R., Savage, M. (2014). After the crisis? Big Data and the methodological challenges of empirical sociology. *Big Data & Society*, April–June 2014: 1–6.
- Chan, J, Bennett Moses, L. (2016). Is Big Data challenging criminology?. *Theoretical Criminology* 20(1): 21–39
- Chen, M, Mao, S, Liu, Y. (2014) Big Data: A Survey. *Mobile Netw. Appl.* 19:171–209.
- Coté, M. (2014). Data motility: The materiality of big social data. *Cultural Studies Review* 20(1):121–49.
- Crawford, K, Miltner, K, Gray, M. (2014). Critiquing Big Data: Politics, Ethics, Epistemology – Special Section Introduction. *International Journal of Communication* 8: 1663–1672.
- Ford, H. (2014). Big Data and Small: Collaborations between ethnographers and data scientists. *Big Data & Society* 1(2): 1–3.
- Fuchs, C. (2017). From digital positivism and administrative big data analytics towards critical digital and social media research!. *European Journal of Communication* 32(1) 37–49.

- Fuller, S. (2006). *The New Sociological Imagination*. London: Sage.
- Gabrys, J. (2014). Programming environments: Environmentality and citizen sensing in the smart city. *Environment and Planning D: Society and Space* 32(1): 30–48.
- Gitelman, L. (ed.) (2013). „Raw Data“ Is an Oxymoron. Cambridge: MIT Press.
- González-Bailón, S. (2013). Social Science in the Era of Big Data. *Policy and Internet* 5 (2):147–160.
- Gouldner, A. W. (1970). *The Coming Crisis of Western Sociology*. New York: Basic Books.
- Halford S, Pope C, Weal MJ. (2013). Digital futures? Sociological challenges and opportunities in the emergent semantic web. *Sociology* 47(1): 173–189.
- Harford, T. (2014). Big data: a big mistake?. *Significance* 11:14–19.
- Halford, S, Savage, M. (2017). Speaking Sociologically with Big Data: Symphonic Social Science and the Future for Big Data Research. *Sociology* 51(6): 1132–1148.
- Housley, W. et al. (2014) Big and broad social data and the sociological imagination: A collaborative response. *Big Data & Society* 1: 1–15.
- Kelling S, Hochachka W, Fink D. (2009). Data-intensive Science: A new paradigm for biodiversity studies. *BioScience* 59(7): 613–620.
- Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography* 3(3) 262–267.
- Kitchin, R. (2014a). Big Data, New Epistemologies and Paradigm Shift. *Big Data & Society* 1 (1):1–12.
- Kitchin, R. (2014b). The real-time city? Big data and smart urbanism. *Geo Journal* 79: 1–14.
- Kitchin R, Lauriault T.P. (2014). Towards critical data studies: Charting and unpacking data assemblages and their work. The Programmable City Working Paper 2. Dostupno na: [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id142474112](http://papers.ssrn.com/sol3/papers.cfm?abstract_id142474112) (pristupljeno 1.6.2018.)
- Kotler, P, Armstrong, G. (2018) *Principles of Marketing*, Global edition (17th), London: Pearson.
- Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity and Variety. *META Group Research Note* 6.
- Latour, B., Jensen, P., Venturini, T., Grauwin, S., Boullier, D. (2012) ‘The whole is always smaller than its parts’ – a digital test of Gabriel Tarde’s monads. *British Journal of Sociology* 63 (4): 590–615.
- Lev-Ram, M. (2014). What’s the Next Big Thing in Big Data? Bigger Dana. *Fortune*, June 16, pp. 233–240.
- Lohr, S. (2015). *Data-ism: The Revolution Transforming Decision Making, Consumer Behavior, and Almost Everything Else*. New York: Harper Collins.
- Lupton, D. (2015). *Digital Sociology*. London & New York: Routledge.
- Lyon D and Bauman Z (2013) *Liquid Surveillance: A Conversation*. Oxford: Wiley.
- Manić, Ž. (2017). *Analiza sadržaja u sociologiji*. Beograd: Čigoja štampa i ISI FF.

- Manovich, L. (2011). Trending: the promises and the challenges of big social data. In: M.K. Gold (ed.) *Debates in the Digital Humanities*. Minneapolis, MN: The University of Minnesota Press.
- [http://www.manovich.net/DOCS/Manovich\\_trending\\_paper.pdf](http://www.manovich.net/DOCS/Manovich_trending_paper.pdf) (pristupljeno 1.6.2018.)
- Mayer-Schönberger, V, Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. New York: Houghton Mifflin Harcourt.
- Mazzocchi, F. (2015). Could Big Data be the end of the theory in science? A few remarks on the epistemology of data-driven science. *EMBO Reports* DOI 10.15252/embr.201541001, published online 10 September 2015. (pristupljeno 1.6.2018).
- Metzler, K., Kim, D. A., Allum, N., & Denman, A. (2016). *Who is doing computational social science? Trends in big data research* (White paper). London, UK: SAGE Publishing. doi: 10.4135/wp160926. Dostupno na: <https://us.sagepub.com/sites/default/files/CompSocSci.pdf> (pristupljeno 1.6.2018).
- Michael, M, Lupton, D. (2016). Toward a manifesto for the ‘public understanding of big data’. *Public Understanding of Science* 25(1) 104–116.
- Morozov, E. (2013). *To Save Everything, Click Here: The Folly of Technological Solutionism*. New York, NY: Public Affairs.
- Rogers, R. (2013). *Digital Methods*. Cambridge, MA: The MIT Press.
- Savage, M, Burrows, R. (2007). The Coming Crisis of Empirical Sociology. *Sociology* 41(5): 885–899.
- Savage, M, Burrows, R. (2009). Some Further Reflections on the Coming Crisis of Empirical Sociology. *Sociology* 43(4): 762–772.
- Siegel, E. (2013). *Predictive Analytics*. Hoboken: Wiley.
- Singer, B. (2016). A chronology of tactics: Art tackles Big Data and the environment. *Big Data & Society* July–December 2016: 1–9.
- Spasić, I. (2012). Debate u sociološkoj teoriji na početku 21. veka. *Godišnjak za sociologiju* IX/10–11: 9–32.
- Stinchcombe, A. L. (1994). Disintegrated Discipline and the Future of Sociology. *Sociological Forum* 9 (2): 279–291.
- Strasser, B. (2012). Data-driven sciences: From wonder cabinets to electronic databases. *Studies in History and Philosophy of Biological and Biomedical Sciences* 43: 85–87.
- Tinati R, Halford S, Carr L. (2014). Big data: Methodological challenges and approaches for sociological analysis. *Sociology* 48(4): 663–681.
- Tufekci, Z. (2014). Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, Association for the Advancement of Artificial Intelligence ([www.aaai.org](http://www.aaai.org)) (pristupljeno 1.6.2018).
- van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society* 12(2): 197–208.

**PRILOG:**  
**Spisak analiziranih naučnih radova**

*1. Naučni radovi koji su posvećeni teorijskom razmatranju upotrebe masovnih podataka:*

**a. Implikacije masovnih podataka  
na sociološku teoriju i istraživačku praksu**

- boyd, danah & Kate Crawford. (2012). Critical Questions for Big Data. *Information, Communication & Society* 15(5):662–679.
- Brock, Andre. (2015). Deeper data: a response to boyd and Crawford. *Media, Culture & Society* 37(7):1084–1088.
- Brooker, Phillip, William Dutton, Christian Greiffenhagen. (2017). What would Wittgenstein say about social media?. *Qualitative Research* 17(6): 610–626.
- Chan, Anita. (2015). Big data interfaces and the problem of inclusion. *Media, Culture & Society* 37(7):1078–1083.
- Couldry, Nick. (2014). Inaugural: A Necessary Disenchantment: Myth, Agency and Injustice in a Digital World. *The Sociological Review* 62(4): 880–897.
- Frade, Carlos. (2016). Social Theory and the Politics of Big Data and Method. *Sociology* 50(5): 863–877.
- Halavais, Alexander. (2015). Bigger sociological imaginations: framing big social data theory and methods. *Information, Communication & Society* 18(5):583–594.
- Halford, Susan, Mike Savage. (2017). Speaking Sociologically with Big Data: Symphonic Social Science and the Future for Big Data Research. *Sociology* 51(6): 1132–1148.
- Hesse, Bradford W., Richard P. Moser, William T. Riley. (2015). From Big Data to Knowledge in the Social Sciences. *The ANNALS of the American Academy of Political and Social Science* 659 (1):16–32.
- Hofferth, Sandra L., Emilio F. Moran, Barbara Entwisle, J. Lawrence Aber, Henry E. Brady, Dalton Conley, Susan L. Cutter, Catherine C. Eckel, Darrick Hamilton, Klaus Hubacek. (2016). Introduction: History and Motivation. *The ANNALS of the American Academy of Political and Social Science* 669 (1): 6–17.
- Lazer, David, Jason Radford. (2017). Data ex Machina: Introduction to Big Data. *Annual Review of Sociology* 43:19–39.
- Lin, Jimmy. (2015). On Building Better Mousetraps and Understanding the Human Condition, Reflections on Big Data in the Social Sciences. *The ANNALS of the American Academy of Political and Social Science* 659(1): 33–47.
- Papacharissi, Zizi. (2015). The unbearable lightness of information and the impossible gravitas of knowledge: Big Data and the makings of a digital orality. *Media, Culture and Society* 37 (7): 1095–1100.
- Ruppert, Evelyn, John Law, Mike Savage. (2013). Reassembling Social Science Methods: The Challenge of Digital Devices. *Theory, Culture & Society* 30(4): 22–46.

- Seaver, Nick. (2015). The nice thing about context is that everyone has it. *Media Culture & Society* 37(7):1101–1109.
- Uprichard, Emma. (2012). Being Stuck in (Live) Time: The Sticky Sociological Imagination. *The Sociological Review* 60 (1\_suppl): 124–138.
- de Zúñiga, Homero Gil, Diehl, Trevor. (2015). Citizenship, Social Media, and Big Data Current and Future Research in the Social Sciences. *Social Science Computer Review* 35(1): 3–9.

### b. Društvene implikacije masovnih podataka

- Aradau, Claudia, Tobias Blanke. (2016). Politics of prediction Security and the time/space of governmentality in the age of big data. *European Journal of Social Theory* 20(3):373–391.
- Arditi, David. (2017). Music Everywhere: Setting a Digital Music Trap, *Critical Sociology*. Article first published online: August 31, 2017 (pp. 1–14) <https://doi.org/10.1177/0896920517729192>
- Baruh, Lemi, Popescu, Mihaela. (2017). Big data analytics and the limits of privacy self-management. *New Media and Society* 19 (4): 579–596.
- Beer, David. (2014). The Biopolitics of Biometrics: An interview with Btihaj Ajana. *Theory, Culture & Society* 31(7/8): 329–336.
- Beer, David. (2017a). The Social Power of Algorithms. *Information, Communication & Society* 20(1):1–13.
- Beer, David. (2017b). Envisioning the power of data analytics. *Information, Communication and Society* 21(3): 465–479. |
- Brayne, Sarah. (2017). Big Data Surveillance: The Case of Policing. *American Sociological Review* 82 (5): 977–1008.
- Coll, Sami. (2014). Power, knowledge, and the subjects of privacy: understanding privacy as the ally of surveillance. *Information, Communication & Society* 17(19):1250–1263.
- Dean, Mitchell. (2016). Political acclamation, social media and the public mood. *European Journal of Social Theory* 20 (3): 417–434.
- Dencik, Lina, Arne Hintz, Zoe Carey. (2017). Prediction, pre-emption and limits to dissent: Social media and big data uses for policing protests in the United Kingdom. *New Media and Society* 20(4): 1433–1450.
- Hansen, Hans Krause. (2015). Numerical operations, transparency illusions and the datafication of governance. *European Journal of Social Theory* 18(2): 203–220.
- Harper, Tael. (2017). The big data public and its problems: Big data and the structural transformation of the public sphere. *New Media and Society*, 19(9): 1424–1439.
- Mai, Jens-Erik. (2016). Big data privacy: The datafication of personal information. *The Information Society* 32(3): 192–199.
- Manghani, Sunil. (2017). The Art of Paolo Cirio: Exposing New Myths of Big Data Structures. *Theory, Culture & Society* 34(7–8): 197–214.
- Martin, Kirsten. (2016). Data aggregators, consumer data, and responsibility online: Who is tracking consumers online and should they stop?. *The Information Society* 32(1): 51–63.

- Pauleen, David J, David Rooney, Ali Intezari. (2016): Big data, little wisdom: trouble brewing? Ethical implications for the information systems discipline. *Social Epistemology* 31(4):400–416.
- Pöttsch, Holger. (2017). Archives and identity in the context of social media and algorithmic analytics: Towards an understanding of iArchive and predictive retention. *New Media and Society* 20(1):1–19.
- Qiu, Jack Linchuan. (2015). Reflections on Big Data: 'Just because it is accessible does not make it ethical'. *Media, Culture & Society* 37(7): 1089–1094.
- Yeung, Karen. (2016). 'Hypernudge': Big Data as a mode of regulation by design. *Information, Communication & Society* 20(1):1–19.
- Zarsky, Tal. (2016). The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making. *Science, Technology & Human Values* 41(1): 118–132.

## 2. Naučni radovi koji se bave metodološkim aspektima analize masovnih podataka

- Barratt, Monica J, Alexia Maddox. (2016). Active engagement with stigmatised communities through digital ethnography. *Qualitative Research* 16 (6): 701–719.
- Connelly, Roxanne, Christopher J. Playford, Vernon Gayle, Chris Dibben. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research* 59:1–12.
- Crosas, Mercè, Gary King, James Honaker, Latanya Sweeney. (2015). Automating Open Science for Big Data. *The ANNALS of the American Academy of Political and Social Science* 659 (1): 260–273.
- Evans, James, Pedro Aceves. (2016). Machine Translation: Mining Text for Social Theory. *Annual Review of Sociology* 42:21–50.
- Flyverbom, Mikkel, Anders Koed Madsen, Andreas Rasche. (2017). Big data as governmentality in international development: Digital traces, algorithms, and altered visibilities. *The Information Society* 33(1):35–42.
- Goroff, Daniel, Jules Polonetsky, Omer Tene. (2017). Privacy Protective Research: Facilitating Ethically Responsible Access to Administrative Data. *The ANNALS of the American Academy of Political and Social Science* 675 (1): 46–66.
- Hampton, Keith N. (2017). Studying the Digital: Directions and Challenges for Digital Methods. *Annual Review of Sociology* 43:167–188.
- Hargittai, Eszter. (2015). Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Sites. *The ANNALS of the American Academy of Political and Social Science* 659(1): 63–76.
- Hindman, Matthew. (2015). Building Better Models, Prediction, Replication, and Machine Learning in the Social Sciences. *The ANNALS of the American Academy of Political and Social Science* 659(1): 48–62.
- Liu, Hexuan Liu, Guang Guo. (2016). Opportunities and challenges of big data for the social sciences: The case of genomic data. *Social Science Research* 59:13–22.



- Loukissas, Yanni Alexander. (2017). Taking Big Data apart: local readings of composite media collections. *Information, Communication & Society* 20(5): 651–664.
- Maiers, Claire. (2017). Analytics in action: users and predictive data in the neonatal intensive care unit. *Information, Communication & Society* 20(6): 915–927.
- Mills, Kathy A. (2017). What are the threats and potentials of big data for qualitative research?. *Qualitative Research* Article first published online: November 30, 2017 <https://doi.org/10.1177/1468794117743465>
- Panger, Galen. (2016). Reassessing the Facebook experiment: critical thinking about the validity of Big Data research. *Information, Communication & Society* 19(8):1108–1126.
- Schroeder, Ralph & Linnet Taylor. (2015). Big data and Wikipedia research: social science knowledge across disciplinary divides. *Information, Communication & Society* 18(9): 1039–1056.
- Shapiro, Aaron. (2017). Street-level: Google Street View's abstraction by datafication. *New Media and Society* 20(3) 1201–1219.
- Sharon, Tamar, Zandbergen, Dorien. (2016). From data fetishism to quantifying selves: Self-tracking practices and the other values of data. *New Media & Society* 19(11): 1695–1709.
- Tanweer, Anissa, Brittany Fiore-Gartland, Cecilia Aragon. (2016). Impediment to insight to innovation: understanding data assemblages through the breakdown–repair process. *Information, Communication & Society* 19(6): 736–752.

### 3. Naučni radovi u kojima je primenjena analiza masovnih podataka

- Brandtzaeg, Petter Bae. (2017). Facebook is no “Great equalizer” – A big data approach to gender differences in civic engagement across countries. *Social Science Computer Review* 35(1):103–125.
- Chen, Yunsong, Fei Yan. (2016a). Centuries of Sociology in Millions of Books. *Sociological Review* 64 (4): 872–893.
- Chen, Yunsong, Fei Yan. (2016b). Economic performance and public concerns about social class in twentieth-century books. *Social Science Research* 59: 37–51.
- Fitzhugh, Sean, M., C. Ben Gibson, Emma S. Spiro, Carter T. Butts. (2016). Spatio-temporal filtering techniques for the detection of disaster-related communication. *Social Science Research* 59:137–154.
- Guggenheim, Lauren, S. Mo Jang, Soo Young Bae, W. Russell Neuman. (2015). The Dynamics of Issue Frame Competition in Traditional and Social Media. *The ANNALS of the American Academy of Political and Social Science* 659 (1): 207–224.
- Heerwig, Jennifer A. (2016). Donations and dependence: Individual contributor strategies in house elections. *Social Science Research* 60: 181–198.
- Hilbert, Martin, Javier Vásquez, Daniel Halpern, Sebastián Valenzuela, Eduardo Arriagada. (2017). One Step, Two Step, Network Step? Complementary Perspectives on Communication Flows in Twittered Citizen Protests. *Social Science Computer Review* 35 (4): 444–461.

- Legewie, Joscha, Merlin Schaeffer. (2016) Contested Boundaries: Explaining Where Ethnoracial Diversity Provokes Neighborhood Conflict, *American Journal of Sociology* 122(1) Volume 122 (1): 125–61.
- Liang, Yuhua (Jake), Kerk F Kee. (2016). Developing and validating the A-B-C framework of information diffusion on social media. *New Media and Society* 20(1) 272–292.
- Maireder, Axel, Brian E. Weeks, Homero Gil de Zuniga. (2017). Big Data and Political Social Networks – Introducing Audience Diversity and Communication Connector Bridging Measures in Social Network Theory, *Social Science Computer Review*, 35(1):126–141.
- McKelvey, Karissa Joseph DiGrazia, Fabio Rojas. (2014). Twitter publics: how online political communities signaled electoral outcomes in the 2010 US house election. *Information, Communication & Society* 17(4):436–450.
- Murthy, Dhiraj, Alexander J. Gross. (2017). Social media processes in disasters: Implications of emergent technology use. *Social Science Research* 63: 356–370.
- Murthy, Dhiraj. (2017). Comparative Process-oriented Research Using Social Media and Historical Text. *Sociological Research Online* 22 (4): 3–26.
- O'Brien, Daniel Tumminelli. (2016). Using small data to interpret big data: 311 reports as individual contributions to informal social control in urban neighborhoods. *Social Science Research* 59: 83–96.
- O'Brien, Daniel Tumminelli, Robert J. Sampson, Christopher Winship. (2015). Econometrics in the Age of Big Data – Measuring and Assessing “Broken Windows” Using Large-scale Administrative Records. *Sociological Methodology* 45(1) 101–147.
- Papacharissi, Zizi. (2015). Affective publics and structures of storytelling: sentiment, events and mediality. *Information, Communication & Society* 19(3): 307–324.
- Skeggs Beverly & Simon Yuill. (2016). The methodology of a multi-model project examining how facebook infrastructures social relations. *Information, Communication & Society* 19 (10): 1356–1372.
- Su, Zheng, Tianguang Meng. (2016), Selective responsiveness: Online public demands and government responsiveness in authoritarian China. *Social Science Research* 59: 52–67.
- Tinati, Ramine, Susan Halford, Leslie Carr, Catherine Pope. (2014). Big Data: Methodological Challenges and Approaches for Sociological Analysis. *Sociology* 48(4):663–681.
- Vargo, Chris J., Lei Guo, Michelle A Amazeen. (2017). The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New Media and Society* 20(5): 2028–2049.
- Vargo, Chris J., Toby Hopp. (2017). Socioeconomic Status, Social Capital, and Partisan Polarity as Predictors of Political Incivility on Twitter – A Congressional District-Level Analysis. *Social Science Computer Review* 35(1): 10–32.
- Wells, Chris, Thorson, Kjerstin. (2017). Combining Big Data and Survey Techniques to Model Effects of Political Content Flows in Facebook. *Social Science Computer Review* 35(1):33–52.