Analiza dostupnosti osobnih podataka nastalih prije i nakon 2000. godine

Humjan, Josip

Master's thesis / Diplomski rad

2020

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, 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:131:183593

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-29**

Filozofski fakultet
Sveučilišta u Zagrebu

Repository / Repozitorij:

ODRAZ - open repository of the University of Zagreb
Faculty of Humanities and Social Sciences

zir.nsk.hr

dabar
DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI

SVEUČILIŠTE U ZAGREBU
FILOZOFSKI FAKULTET
ODSJEK ZA INFORMACIJSKE I KOMUNIKACIJSKE ZNANOSTI
SMJER ARHIVISTIKA
Ak. god. 2019./2020.

Josip Humjan

Analiza dostupnosti osobnih podataka nastalih prije i nakon 2000. godine

Diplomski rad

Mentor: prof. dr. sc. Hrvoje Stančić
Neposredni voditelj: dr. sc. Jozo Ivanović

Zagreb, rujan 2020.

Izjava o akademskoj čestitosti

Izjavljujem i svojim potpisom potvrđujem 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.

Zahvaljujem svojoj obitelji i kolegici Tei Miroslavić na pomoći i podršci prilikom izrade ovoga rada.

Sadržaj

1. Uvod 1
2. Definicija osobnih podataka 3
3. Osobni podaci prije 20. stoljeća 4
4. Percepcija osobnih podataka 8
5. Genealogija i povijesno-demografska istraživanja 9
6. Nove tehnologije u 20. i 21. stoljeću i Web 2.0 15
7. Privatnost 18
8. Zaštita privatnih podataka u 21. stoljeću 21
9. Velike količine podataka 24
10. Povijesni razvoj i karakteristike velikih količina podataka 26
11. Tehnike prikupljanja i obrade velikih količina podataka 35
12. Pozitivni utjecaj velikih količina podataka u poslovanju 37
13. Manjkavosti i zloupotreba velikih količina podataka 38
14. Paradoks privatnosti 44
15. Zaključak 46
16. Literatura 48
Sažetak 52
Summary 53

1. Uvod

Osobni podaci, odnosno pitanje njihove zaštite, u posljednjih nekoliko desetljeća sve više interesira najširu javnost.[1] Mnoge rasprave, članci i konferencije posvećene su toj temi. Osobni podaci najuže su vezani uz privatnost, osobnost i individualnost - vrijednosti koje smatramo temeljima modernog društva. Ipak, dostupnost osobnih podataka i odnos društva prema tim podacima do nedavno bio je temeljito drugačiji. Među najširim slojevima stanovništva koncept privatnosti nije bio raširen u današnjem smislu riječi te se osobnim podacima nije se pristupalo pažljivo. Cilj je ovoga rada prikazati i objasniti promjene u odnosu društva prema osobnim podacima te objasniti zašto je do tih promjena dolazilo. Rad je strukturiran tako da se u prvom, kraćem dijelu, govori o osobnim podacima i odnosu društva prema njima u periodu prije 19. stoljeća. Razvoj pisma i izum tiskarskog stroja ključni su preduvjeti za prikupljanje osobnih podataka. Od 16. stoljeća počinju prva sustavna prikupljanja osobnih podataka. Podaci su tada bili u analognom obliku, nepovezani te je njihova uporaba bila rezervirana isključivo za organizacije koje su te podatke prikupile - državu i Crkvu. Ključan faktor koji je definirao odnos pojedinca i društva prema osobnim podacima bila je činjenica da individualnost kao koncept tada nije afirmiran, što znači da je i percepcija osobnih podataka bila drugačija - svijest o zaštiti osobnih podataka nije bila razvijena. Podaci prikupljeni u tom periodu predaju se državnim arhivskim institucijama sredinom 20. stoljeća te se u velikoj mjeri digitaliziraju na prijelazu s 20. na 21. stoljeće. To otvara nove perspektive za istraživanja tih podataka te se njima počinju još intenzivnije baviti znanstvenici i genealozi amateri. Drugi dio rada odnosi se na suvremeno prikupljanje digitalnih (osobnih) podataka. Do prijelaza s 20. na 21. stoljeće počinje se prikupljati znatno veća količina osobnih podataka u digitalnom obliku. Takvi podaci iznimno su brojni i umreženi. Ovoga puta njima se najviše koriste države i razni poslovni subjekti. Pojavom Weba 2.0 korisnicima interneta omogućuje se da po prvi puta postajnu glavni kreatori internetskog sadržaja, što posredno znači da nikada prije više ljudi nije ostavilo više svojih podataka javno dostupnima. Počinje se upotrebljavati termin velike količine podataka (engl. *big data*). Koncept velikih količina podataka predstavlja kompleksnu temu - postoje benefiti, ali i opasnosti koje oni pružaju. Potencijali analize takvih podataka nedvojbeno su veliki, međutim, do trenutka pojave Weba 2.0 razvila se svijest o potrebi zaštite osobnih podataka. Individualizam postaje jedan od postulata zapadnog svijeta, što tu vrijednost i pravo dovodi u

[1] *Google Ngram Viewer*: "personal data", URL: https://books.google.com/ngrams/graph?content=personal+data&year_start=1800&year_end=2008&corpus=15&smoothing=3&share=&direct_url=t1%3B%2Cpersonal%20data%3B%2Cc0 (25.5.2020.).

konflikt s tendencijama prikupljanja i obrađivanja što većeg broja podataka. U kontekstu tih napetosti posebno je zanimljiv slučaj društvenih mreža. Njihovi korisnici, iako zabrinuti za svoje osobne podatke, ipak na svojim društvenim profilima svojevoljno otkrivaju velik broj osobnih podataka.

Prvi dio rada se uvelike oslanja na nekoliko monografija. *Studije i prilozi iz arhivistike* Igora Karamana korištene su u dijelu o razvoju arhivistike. Djelo *Heraldika, sfragistika, genealogija* Bartola Zmajića korišteno je za dio rada o genealogiji, a za razvoj povijesno-demografskih istraživanja korištena je *Povijesna demografija Hrvatske* Vladimira Stipetića i Nenada Vekarića. Za drugi dio rada iznimno su korisna bila djela Thomasa Davenporta *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities* i Roba Kitchina *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences.* Osim navedenih djela, korisno je bilo i ono Phila Simona *Too Big to Ignore: The Business Case for Big Data.* O negativnim stranama velikih količina podataka opširno se raspravlja u *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* Cathy O'Neil. *Data Thieves in Action: Examining the International Market for Stolen Personal Information* autora Holta, Smirnove i Chua korišteno u dijelu u krađi osobnih podataka. Na kraju, zbornik *Privacy Online: Perspectives on Privacy and Self-Disclosure in the Social Web* korišten je za detaljniju analizu osobnih podataka, odnosno privatnosti u kontekstu društvenih mreža.

2. Definicija osobnih podataka

Za početak, potrebno je definirati osobne podatke – središnji pojam ovoga rada. Za potrebe ovoga rada spomenimo ilustracije radi nekoliko relevantnih definicija i interpretacija osobnih podataka. Najvažnija pravna definicija osobnih podataka u Europskoj uniji i Hrvatskoj je iz Opće uredbe od zaštiti osobnih podataka (engl. *General Data Protection Regulation* - GDPR) donesene u svibnju 2018. godine. Prema članku 4. tog zakona osobni podaci su "...svi podaci koji se odnose na pojedinca čiji je identitet utvrđen ili se može utvrditi („ispitanik"); pojedinac čiji se identitet može utvrditi jest osoba koja se može identificirati izravno ili neizravno, osobito uz pomoć identifikatora kao što su ime, identifikacijski broj, podaci o lokaciji, mrežni identifikator ili uz pomoć jednog ili više čimbenika svojstvenih za fizički, fiziološki, genetski, mentalni, ekonomski, kulturni ili socijalni identitet tog pojedinca".[2] Marija Boban u svom doktoratu *Pravni i kulturološki aspekti sigurnosti i zaštite osobnih podataka* ističe da su osobni podaci i privatnost povezani, ali ne i jednoznačni pojmovi. "...pojam *osobnosti* pokriva područje koje izričito predstavlja osobu i njezin osobni život, a pojam *privatnosti* zahvaća širi krug značajki osobe koje uzete pojedinačno, ne moraju imati neko značenje, ali povezane u postojanu strukturu rezultiraju portretom osobe, za koji najčešće samo ta osoba treba znati".[3] Dakle, privatnost je po autorici širi pojam od osobnih podataka.[4]

Ovaj se rad ne oslanja isključivo na jednu definiciju / interpretaciju. S obzirom da se pokriva relativno dugački period u kojem je došlo do velikih promjena u ne samo kvalitativnom i kvantitativnom razvoju osobnih podataka, već i percepciji pojedinaca odnosno društva prema tim podacima, nije niti moguće osloniti se samo na jednu ili nekoliko definicija. U ovome se radu osobne podatke promatra kroz prizmu odnosa pojedinca odnosno cijelog društva prema tim podacima, tj. s načinom na koji se mijenjala percepcija tih podataka.

[2] Zakon.hr, *Opća uredba o zaštiti osobnih podataka*, URL: https://www.zakon.hr/z/1021/Opća-uredba-o-zaštiti-podataka---Uredba-(EU)-2016-679, (22.4.2020.).
[3] Boban, Marija. "Pravni i kulturološki aspekt sigurnosti i zaštite osobnih podataka". Doktorska disertacija, Sveučilište u Zagrebu, 2012, str. 10-11.
[4] Boban, "Pravni i kulturološki aspekti sigurnosti", n. dj., str. 10.

3. Osobni podaci prije 20. stoljeća

Zapisivanje osobnih podataka staro je gotovo koliko i pismo. Ljudi su uglavnom iz političkih ili ekonomskih razloga zapisivali određene pojedinosti o osobi. Dostignuća vladara, vojskovođa i religijskih figura smatrani su vrijednima zapisivanja zbog njihova doprinosa zajednici. Osobni podaci nisu nužno bili primarni razlog zapisa, već nusprodukt, ali i kao takvi predstavljaju neke od prvih zapisa osobnih podataka. Ekonomski razlozi zapisivanja osobnih podataka vežemo uz Mezopotamiju prije gotovo 3.000 godina pr. Kr., kada su se na glinene pločice zapisivali ugovori pravne odnosno ekonomske naravi. Bili su vezani uglavnom uz trgovinu i poljoprivredu. Primarni razlog nastanka ovakvih zapisa također nije bio zapisivanje osobnih podataka, već trgovinskih ugovora - osobni su podaci samo nužni nusprodukt takvih dokumenata.[5] U narednim tisućljećima došlo je do određenih promjena vezanih uz pismo kao takvo. Središta pismenosti odlaze van Mezopotamije, a nositelji pisane riječi postaju državne zajednice Grčke i Rima, međutim, ključnih promjena u odnosu prema osobnim podacima nema.[6]

Prve veće promjene dolaze tek u 15. stoljeću. Naime, do tada se kao materijal za zapisivanje najčešće koristila pergamena, koja je bila skupa. S obzirom na to, broj zapisa je bio ograničen. Izumom i masovnom proizvodnjom papira koja započinje u tom stoljeću, ostvareni su preduvjeti za masovnije širenje pisane riječi.[7] Druga ključna promjena je izum tiskarskog stroja sredinom 15. stoljeća. Do tada, tekstovi su bili pisani ručno što je značilo da je umnažanje jedne knjige zahtijevalo mukotrpan rad prepisivanja cijelog teksta. Tiskarski je stroj znatno smanjio vrijeme potrebno za dovršenje knjige, što je posljedično značilo nastanak puno više tekstova u kraćem periodu. Prije izuma tiskarskog stroja u Europi je bilo oko 250 tisuća rukopisa,[8] a već oko 1500. godine (50 godina nakon izuma) Europa je bila preplavljena s oko 20 milijuna knjiga. U 16. stoljeću ta je brojka narasla na čak 200 milijuna.[9] Potencijal tiska ubrzo je prepoznat i kao opasnost. Godine 1487. papa donosi bulu kojom se naređuje kontrola svih tiskara.[10] Crkva je sumnjičavo gledala na širenje pisane riječi nad kojom nije imala kontrolu.

[5] Boban, "Pravni i kulturološki aspekti sigurnosti", n. dj., str. 50.
[6] *Britannica*, Genealogy, https://www.britannica.com/topic/genealogy (25.5.2020.).
[7] Stipčević, Aleksandar, Povijest Knjige (Zagreb: Matica hrvatska, 2006), str. 343.
[8] Tuđman, Miroslav, *Informacijsko ratište i informacijska znanost*. Zagreb: Hrvatska sveučilišna naklada, 2008, str. 52.
[9] Tuđman, *Informacijsko ratište*, n. dj., str. 53.
[10] Tuđman, *Informacijsko ratište*, n. dj., str. 58.

Njihove su sumnje očito bile opravdane jer da nije bilo tiska, utjecaj Lutherovih teza iz 1517. godine ne bi bio ni približno tako velik.[11] Jedna od posljedica teza bila je podjela kršćanskog svijeta na što je Katolička crkva odgovorila sazivanjem Tridentskog sabora krajem 16. stoljeća. U kontekstu osobnih podataka, iznimno je važna odredba sabora kojom se propisuje obvezno vođenje matičnih knjiga. Crkva je na taj način, između ostalog, provodila svojevrsni cenzus katolika, čime se htjela dobiti evidencija korpusa katolika naspram sve brojnijih protestanata.[12] Masovno korištenje papira u kombinaciji s navedenim odredbama značilo je do tada najveći kvalitativni, a pogotovo kvantitativni skok u broju zapisanih osobnih podataka. Do tada su osobni podaci bili nužno vezani uz istaknute pojedince ili određene pravne radnje, ali sada, po prvi puta, imamo namjeru jedne od tadašnjih najvećih organizacija da vodi svojevrsne evidencije svojih vjernika, koji su tada činili velik dio cjelokupnog, globalnog stanovništva. To ne znači da su navedene mjere odmah uspješno implementirane od strane cjelokupne crkvene hijerarhije, a ako i jesu, ti zapisi nisu možda sačuvani, međutim, to je bio početak sustavnog vođenja matičnih knjiga, tj. zapisivanja osobnih podataka. Da je 16. stoljeće iznimno važano za širenje tiska i matičnih knjiga dokazuje i gradivo Hrvatskog državnog arhiva, koji u svojoj nadležnost ima cijelu Republiku Hrvatsku. Naime, gradivo koje je pohranjeno u tom arhivu uglavnom je nastalo tijekom ili nakon 16. stoljeća.[13] Slična je situacija i s matičnim knjigama. Tek od 16. stoljeća matične knjige postaju relativno učestale, da bi u narednim stoljećima postajale sve raširenije.[14]

Na prijelazu s 18. na 19. stoljeće, kada tehnike proizvodnje papira postaju sve jeftinije i brže, tisak postaje još raširenij. [15] Osim toga, u tom se periodu snažno razvija državna uprava, koja proizvodi velike količine pisanih dokumenata. Upravo zato 19. stoljeće predstavlja još jedan veliki porast u broju pisanih dokumenata i matičnih knjiga.[16]

Same tehnike proizvodnje papira, tiska te odluke Crkve o vođenju matičnih knjiga važne su za razumijevanje početka zapisivanja i vođenja evidencija osobnih podataka,

---

[11] Tuđman, *Informacijsko ratište*, n. dj., str. 55.
[12] Cik, Nikola, "Crkvene matične knjige kao izvor za povijesna istraživanja, na primjeru župe Đurđevac", *Pro tempore: časopis studenata povijesti*, 10-11 (2016), str. 238.
[13] Kolanović, Josip, *Pregled arhivskih fondova i zbirki Republike Hrvatske* (Zagreb: Hrvatski državni arhiv, 2006): str. XXI.
[14] *FamilySearch*, URL: https://www.familysearch.org/search/catalog/results?count=20&query=%2Bauthor_id%3A845152&page=1&topic0=Birth%2C+Marriage+and+Death&year0=1500, (10.4.2020.).
[15] Stipčević, *Povijest knjige*, n. dj., str. 585.
[16] Kolanović, *Pregled arhivskih fondova i zbirki*, n. dj., str. XXI i *FamilySearch*, URL: https://www.familysearch.org/search/catalog/results?count=20&query=%2Bauthor_id%3A845152&page=1&topic0=Birth%2C+Marriage+and+Death&year0=1500, (10.4.2020.).

međutim, to ne objašnjava kakav je bio odnos pojedinca i društva prema tim podacima. Do 19. stoljeća iskristalizirala su se dva glavna odnosa. Prvi se odnosi na visoko plemstvo koje je temeljem svojeg imetka i plemićkog statusa imalo pristup najvažnijim političkim, vojnim i sudbenim funkcijama u zemlji. Te su funkcije bile namijenjene samo njima.[17] Pojedinci, tj. cijele obitelji, koji su dolazili iz najviših i najprivilegiranijih društvenih slojeva pridavali su veliku važnost rodoslovljima. Obiteljska se povijest istraživala i prikazivala. S obzirom da je društveni, ekonomski i politički položaj takvih obitelji uvelike ovisio o njihovom plemićkom statusu, rodoslovlja su predstavljala svojevrsni legitimitet takvih obitelji.

S druge strane, svi ostali slojevi stanovništva nisu pridavali pažnju svojim rodoslovljima. Razlog tomu je činjenica nisu imali prava poput plemića što je značilo da njihova rodoslovlja nisu imala što legitimirati. Glavni razlog nebrige oko vlastite obiteljske prošlosti ipak je bilo ekstremno siromaštvo u kojem je živjela većina stanovništva. Život je bio težak i kratak.[18] Najveće brige bile su vezane uz egzistencijalna pitanja. U takvoj situaciji brige oko rodoslovlja, a pogotovo njihova izrada mogla je biti percipirana samo kao veliki rizik. Sukladno tomu, sve do druge polovice 20. stoljeća, većina stanovništva nema vlastita rodoslovna stabla. Seljačko je društvo sve do kraja 19. stoljeća izrazito zatvoreno. Gotovo nikada ne odlaze van svoje župe. Na bilo koju vrstu njima nepoznatih ljudi gledaju negativno. To uključuje sve državne službenike, trgovce i razne putnike. Jedina iznimka bili su lokalni župnici. Tradicionalna društva, iako u mnogim aspektima zapravo praznovjerna, smatrala su vjeru iznimno važnom. Prema tome, predstavnik Crkve bio je prihvaćen u zajednici. To je iznimno važno za napomenuti jer su upravo svećenici bili ti koji su stoljećima zapisivali osobne podatke svojih župljana. Bilo tko drugi tko bi pokušao isto naišao bi na pasivni ili čak aktivni otpor lokalne zajednice.

Crkva nije bila jedina organizacija koja je pokušala voditi evidencije "svojeg" stanovništva. Isto je pokušala i svjetovna vlast. Tijekom srednjeg vijeka, a pogotovo novog vijeka u zapadnoj Europi teritorijalne države počinju mijenjati svoj karakter. Postavljaju se državni službenici koji su zaduženi za izvršavanje i kontrolu svih vladarevih naredbi. Država nastoji biti iznad Crkve. Razlog tomu je rat. On je skup, težak za organiziranje te iznimno važan za budućnost vladara odnosno države. S obzirom na to, rat se treba što bolje organizirati, sve mora biti pod nadzorom, a vladareve odluke moraju se efikasno provoditi.

[17] Iveljić, Iskra, *Banska Hrvatska i Vojna krajina od prosvijećenog apsolutizma do 1848. godine*. Zagreb: Leykam Internat., 2010, str. 94.
[18] Iveljić, *Banka Hrvatska*, n. dj., str. 113.

Jedan od ključnih elemenata uspjeha je pravedno i efikasno ubiranje poreza.[19] Da bi se sve navedeno uspješno implementiralo, bilo je potrebno ostvariti efikasan nadzor nad cjelokupnim stanovništvom. Srednjovjekovne države nisu bile u stanju to napraviti. Trebalo je sačekati 18. i 19. stoljeće da bi svjetovna vlast uspješno nadzirala veliku većinu stanovništva. Jedan od razloga zašto je crkvena vlast brže i uspješnije implementirala svoj nadzor od svjetovne vlasti leži u već spomenutoj činjenici da je Crkva imala na raspolaganju svećenike kojima je lokalno stanovništvo vjerovalo, dok svjetovna vlast nije imala sličan ekvivalent. Crkva je vodila matične knjige, a neki od najvažnijih i najreprezentativniji načina svjetovnog nadzora bili su katastri i popisi stanovništva. Prvi sustavni cjelokupni popis stanovništva na hrvatskom povijesnom prostoru napravljen je 1857. godine, dok prve sustavne katastarske izmjere sežu u 18. stoljeće, kada je dovršena tzv. jozefinska izmjera.[20]

[19] Braudel, Fernand, *Civilizacije kroz povijest* (Zagreb: Globus, 1990), str. 284.
[20] Slukan, Mirela, *Povijesna kartografija* (Samobor: Meridijani, 2003), str. 142.

4. Percepcija osobnih podataka

Svijest o zaštiti osobnih podataka u 21. stoljeću je na visokoj razini. Suvremenicima je stoga možda teško razumjeti da su ljudi do prije nekoliko stoljeća imali sasvim drugačiju percepciju osobnih podataka. Već spomenuta briga oko egzistencijalnih pitanja stoljećima je bila primarna briga većine stanovništva. U takvoj situaciji briga oko osobnih podataka nije mogla biti imperativ, međutim, činjenica da se o tim podacima nije brinulo ovisi o još jednom ključnom faktoru. Naime, privatnost kao koncept produkt je prosvjetiteljstva i razvoja kapitalizma u 18. stoljeću. Tek se tada odvaja privatna od javne sfere te pojedinac sebe počinje percipirati više kao individuu, a ne samo kao pripadnika neke zajednice. Primjera radi, spomenimo da riječ "Career" (hrv. *karijera*) u engleskom jeziku postoji od 16. stoljeća, ali značenje koje danas ima (razvoj osobne karijere) dobiva tek na samom početku 19. stoljeća.[21] Prema tome, da bi netko imao svijest o potrebi zaštite svojih osobnih podataka, mora percipirati sebe kao individuu.

[21] Do tada je riječ označavala „smjer/razvoj događaja" (engl. „general course of action or movement") *Online Etimology Dictionary*, „Career", URL: https://www.etymonline.com/word/career#etymonline_v_5378, (11.4.2020.).

5. Genealogija i povijesno-demografska istraživanja

Pun potencijal crkvenih i svjetovnih zapisa osobnih podataka nije mogao biti ostvaren dok god se ti zapisi nalaze u analognom obliku. Razvojem novih tehnologija krajem 20. i u 21. stoljeću to se promijenilo. Sve više zapisa nalazi se u digitalnom obliku što uvelike olakšava povijesno-demografska i genealoška istraživanja - od genealoga amatera do znanstvenika.

Genealogija se pojavljuje u raznim oblicima, od jednostavnih do vrlo kompleksnih. Osnovni pojmovi genealogije koje je definirao Zmajić su:

1. Obitelj (*familia*) kućanska je zajednica kojoj je na čelu otac kao glava obitelji, zatim žena i djeca.
2. Rod (*gens*) jest cjelina pojedinih obitelji istoimene muške linije, sa kćerima, ali bez njihovih potomak
3. Pleme (*stirps, prosapia*) čine rođaci jednog čovjeka s očeve i majčine strane, tj. njegovi krvni i pravni rođaci.

Zmajić razlikuje više vrsta rodoslovlja:

- tabla ascendenata: prikaz u kojem je najstarija poznata osoba na dnu tabele, a njegovi potomci su iznad njega,
- tabla descendenata: prikaz u kojem je najstarija poznata osoba na vrhu tabele, a njegovi potomci su ispod njega,
- tabla konscendenata, tj. tabla plemena odnosno roda; prikaz u kojem su prikazani muški i ženski potomci s tim da se za ženske potomke ne daju informacije o njihovim potomcima,
- tabla svojte: prikaz u kojem su prikazani muški i ženski potomci s tim da i za muške i za ženske potomke se daju informacije o njihovim potomcima.[22]

Za izradu genealogije korisni su: "krsni listovi, zaručnički listovi, ženidbeni listovi, vjenčana pisma, crkveni dispensi za vjenčanje, zapisivanje jutarnjeg dara, oporuke, zadužbinska pisma zadušnice 'altarije', kupoprodajni ugovori, odštetni ugovori, hipotekarni ugovori, založni ugovori, sporazumni ugovori, lenska pisma, dekreti o imenovanju, dekreti u unapređenju, majstorski listovi, doktorske diplome, grbovni listovi (grbovnice), plemićki listovi".[23]

[22] Zmajić, Bartol, *Heraldika, sfragistika, genealogija, veksilologija, rječnik heraldičkog nazivlja*. Zagreb: Golden Marketing, 1996, str. 101.
[23] Zmajić, *Heraldika*, n. dj., str. 105.

Povijest genealogije može se podijeliti u tri faze. Prva je oralna tradicija, dakle ono što se prenosilo s koljena na koljeno, druga faza je zapisivanje informacija relevantnih za genealogiju i treća faza je period od otprilike 1500. godine u zapadnoj Europi. Tada se počinje toliko zapisivati da je ljudima s tog područja danas moguće istražiti svoju obiteljsku povijest do tog perioda. Prva faza odnosi se na početke civilizacije kada nisu postojali pisani dokumenti ili su oni bili vrlo rijetki. S pojavom pisma oralna tradicija mogla se početi zapisivati. To se dogodilo u Grčkoj i Rimu, međutim, te genealogije nisu kao današnje - one su bile svojevrsni nusprodukt napora da se dokaže povezanost određene osobe s nekim od božanstava. Nakon pada Rima slijedi svojevrsna crna rupa povijesti gdje nema toliko zapisa, pa se tako i prekidaju rodoslovne linije. Važno je napomenuti da neki dijelovi svijeta nisu zapisivali podatke koje bi danas mogli smatrati relevantnima za genealogiju. Npr. u Japanu je tek 1947. godine započelo sustavno zapisivanje podataka o stanovništvu. U drugim krajevima zapisi su se počeli voditi tek nakon što su kristijanizirani, kao što je primjer s britanskim otočjem. Zapisi su nakon 1500. godine postali toliko rašireni da od tada postoje podaci o sve većem udjelu cjelokupnog stanovništva.[24] Crkva je bila stvaralac i čuvar takvih zapisa. U takvim uvjetima sustavna znanstvena istraživanja matičnih knjiga je moguća, ali je otežana. Suvremena arhivska obrada tih materijala (što podrazumijeva i čuvanje gradiva od strane državnih institucija) važan je preduvjet sustavnijim istraživanjima. Sve do 18. stoljeća svo gradivo koja se čuva ima neku pravnu važnost, dokazuje vlasništvo nad imovinom ili nešto slično. Stvaratelj je u pravilu taj koji gradivo čuva. Prema tome, oni su odlučivali i koje je gradivo vrijedno čuvanja, a koje nije. To je za povijesnu znanost, genealogiju, kulturu itd. loše, jer je na taj način stvaralac izbacivao svo gradivo koje za njega nema direktnu vrijednost. Od 18. stoljeća javljaju se samostalni arhivi koji čuvaju gradivo više od jednog stvaraoca. Više je razloga zašto se to dogodilo. Staro gradivo nije bilo više potrebno, a htjelo ga se javno prezentirati, pogotovo radi povijesne znanosti koja se počinje razvijati.[25] Uz to, sve razvijenija državna administracija stvara dosta materijala koji se treba predavati arhivima kada više nije potreban. Arhivi na taj način postaju sve važniji za znanost. Gradivo se počelo percipirati kao opće dobro, kao nešto što ima umjetničku vrijednost i što je vrijedno čuvanja.[26]

[24] *Britannica*, Genealogy, https://www.britannica.com/topic/genealogy (25.5.2020.).
[25] Stulli, Bernard, *Arhivistika i arhivska služba: studije i prilozi*. Zagreb: Hrvatski državni arhiv, 1997, str. 350.
[26] Stulli, *Arhivistika*, n. dj., str. 351.

Matične knjige na hrvatskom povijesnom prostoru datiraju uglavnom od 16. i 17. stoljeća.[27] Sustavna genealoška istraživanja datiraju iz 18. stoljeća. U *Ljetopisu Petra Kegljevića* sadrži informacije o rodoslovljima plemstva, a kasnije je nastala genealogija Ohmučevića i nekih drugih obitelji. Sustavnim istraživanjima bavili su se Pavao Ritter Vitezović, Ivan Kukuljević Sakcinski, Vjekoslav Klaić, Ferdo Šišić i drugi.[28] Amaterska genealoška istraživanja u Hrvatskoj i svijetu započinju tek u drugoj polovici 20. stoljeća. Zemlje engleskog govornog područja i one s imigrantskom prošlošću, poput SAD-a, Kanade i Australije, imaju najrazvijenije amaterske genealogije. U ovom kontekstu vrijedi spomenuti jednu od najvećih, ako ne i najveću bazu podataka relevantnih za genealogiju koju drži mormonska zajednica u Salt Lake Cityju.[29]

Crkvene zajednice u Hrvatskoj brinule su o matičnim knjigama do sredine 20. stoljeća kada ta briga prelazi na državne institucije. Zakonom o državnim matičnim knjigama iz 1946. godine određeno je da Crkva može i dalje voditi svoje evidencije u vidu matičnih knjiga, ali da ta dokumentacija više nema javni značaj. Time su stvoreni preduvjeti da se to gradivo počne otpremati u državne arhive, što je i učinjeno.[30] Zakonom iz 1949. godine sve matične knjige starije od 1858. godine moraju se predati arhivima. Prema dostupnim podacima, 1950-ih na području NR Hrvatske državni arhivi čuvali su oko 5.000 crkvenih matičnih knjiga, uglavnom katoličkih, manje pravoslavnih.[31] Odmah po primitku gradiva, arhivi su krenuli u njihovu obradu kako bi što prije bilo dostupno javnosti.[32] Treba napomenuti da Zakon kojim je određeno da se sve matične knjige moraju predati nadležnim ustanovama nije u potpunosti proveden. Naime, postoje podaci da u Vojvodini neke od najstarijih matičnih knjiga nisu predane. Postavlja se pitanje, ako se to dogodilo u Vojvodini, koje još gradivo nije predano?[33] Ubrzo nakon arhivističke obrade djela gradiva osniva se zbirka matičnih knjiga. Desetak znanstvenih radnika je u kratkom roku nakon osnivanja zbirke koristilo te materijale u znanstvene svrhe.[34] Treba napomenuti da matične knjige ne pružaju uvijek točne podatke. Prilikom zapisivanja znale su se događati greške zbog raznih komplikacija ili nemarnosti župnika. Pomoću detaljnih analiza utvrdilo se da u drugoj polovici 19. stoljeća ima relativno velikih grešaka u maticama krštenih. Naime, oko 3 do 4% osoba nije upisano u matice

[27] Karaman, Igor, *Studije i prilozi iz Arhivistike*. Zagreb: Arhiv Hrvatske, 1997, str. 81-82.
[28] *Hrvatska enciklopedija*, „Genealogija“, URL: https://www.enciklopedija.hr/natuknica.aspx?id=21578, (5.4.2020.).
[29] *Britannica*, Genealogy, https://www.britannica.com/topic/genealogy (25.5.2020.).
[30] Karaman, *Studije*, n. dj., str. 80.
[31] Karaman, *Studije*, n. dj., str. 81.
[32] Karaman, *Studije*, n. dj., str. 82.
[33] Karaman, *Studije*, n. dj., str. 83.
[34] Karaman, *Studije*, n. dj., str. 84.

krštenih ili je upisano naknadno. Ukupna točnost o natalitetu u 18. stoljeću je otprilike 95%, a u 19. stoljeću između 94 i 98%.[35]

Velike promjene zadesile su genealoge počevši od samog kraja 20. stoljeća. Naime, tada su nove tehnologije omogućile lakša, kvalitetnija i brža genealoška istraživanja. Velik dio informacija dostupan je online, a nezanemariv dio gradiva je digitaliziran. Unatoč tome, suština genealoških istraživanja nije se promijenila, samo je postala lakšom. Christian je o implementaciji novih tehnologija u genealogiji zapisao: "Iako internet nije promijenio temeljna načela genealoškog istraživanja, promijenio je način na koji se neka od tih istraživanja provode i donio ogromnu razliku u onome što pojedini genealog može lako napraviti"[36].

Zbirka matičnih knjiga Hrvatskog državnog arhiva najveća je takva zbirka u Hrvatskoj. Razlog tome je što Hrvatski državni arhiv ima pod svojom nadležnošću ima cijelu zemlju, što tu zbirku čini reprezentativnom. Promotrimo li dostupne matične knjige u Arhivu, možemo uočiti neke trendove. Iako su najstarije matične knjige s hrvatskog povijesnog prostora iz 15. stoljeća,[37] najstarije matične knjige u Hrvatskom državnom arhivu su iz 16. stoljeća i sve su iz priobalnih područja, često s otoka. Razlog tomu je što su priobalni krajevi imali snažniju crkvenu organizaciju te što su otoci bili relativno zaštićeni od osmanskih prodiranja. Osim toga, uočljivo je da kontinentalna Hrvatska ima sačuvane matične knjige iz tek druge polovice 17. stoljeća. Primarni razlog tomu su osmanska prodiranja, koja su tada prestala biti intenzivna. Od hrvatskih regija u kontekstu sačuvanosti matičnih knjiga najzakinutija je Slavonija. Razlog tomu je što je Slavonija inkorporirana u sastav Habsburške monarhije, tj. unutar službene crkvene hijerarhije, konačno tek 1699. godine.

Tablica 1. Sačuvanost matičnih knjiga po regijama

Regija	**Period**			
	1500.-1549.	1550.-1599.	1600.-1649.	1650.-1699.
Otoci	1 (33%)	12 (55%)	10 (33%)	6 (6%)
Priobalna Hrvatska (bez otoka)	2 (67%)	10 (45%)	20 (67%)	39 (37%)

[35] Stipetić, Vladimir i Nenad Vekarić, *Povijesna demografija Hrvatske* Zagreb: Hrvatska akademija znanosti i imjetnosti, Zavod za povijesne znanosti, 2004, str. 250.
[36] Christian, Peter, *The Genealogist's Internet*. London: Bloomsbury Publishing Plc, 2012, str. 1.
[37] Stipetić i Vekarić, *Povijesna demografija*, n. dj., str. 28.

Kontinentalna Hrvatska	0 (0%)	0 (0%)	0 (0%)	60 (57%)[38]

Druga vrsta istraživanja koja ovisi o osobnim podacima su povijesno demografska istraživanja, koja svoje začetke imaju u 17. stoljeću s djelima Johna Graunta i William Pettyja. Ta istraživanja odnosila su se na uski geografski prostor poput istraživanja matičnih knjiga u Londonu ili Breslavi.[39] Povijesna se demografija zapravo razvila iz dva smjera. Jedan je je *tradicijski smjer* (razvio se iz etnografije), a drugi je *povijesno-ekonomski smjer* (razvio se iz ekonomije). Začetnik tradicijskog smjera je Baltazar Bogišić koji se bavio zadrugama. Godine 1888. osnovan je u JAZU "Odbor za sabiranje spomenika tradicionalne literature" koji je kasnije počeo izdavati "Zbornik za narodni život i običaje južnih Slavena" - tamo su objavljivali autori tradicijskog smjera.[40] Drugi smjer razvio se iz ekonomije iz razloga jer su ekonomiji bili potrebni podaci o stanovništvu. Začetnik tog smjera bio je Fran Vrbanić.[41] Vrbanić je nailazio na velike probleme jer nije postojalo dovoljno njemu korisnih podataka. Kasnije se cijeli niz autora bavio tom temom. Problem je što su svi bili ekonomisti kojima su trebali demografski podaci, ali najčešće samo oni osnovni, bez detaljnijih

[38] Najstarije župe s godinom prvog zapisa: Hvar, 1516.; Labin, 1536.; Bale, 1538.; Rovinj, 1553.; Dobrinj, 1560.; Osor, 1560.; Olib, 1565.; Krk, 1565.; Svetivinčenat, 1568.; Rab, 1569.; Sv. stošija (Zadar), 1569.; Cres, 1571.; Pag, 1577.; Šibenik, 1581.; Pazin, 1582.; Pašman, 1582.; Marina, 1583.; Dubašica, 1585.; Banj, 1587.; Omiš, 1588.; Rijeka, 1590.; Mali Lošinj, 1592.; Novalja, 1596.; Sv. dujam (Split), 1597.; Vrbanj, 1599.; Split stari grad, 1600.; Ugljan, 1601.; Zrenj, 1606.; Žman, 1607.; Lukoran, 1607.; Sukošan, 1608.; Tkon, 1612.; Labinci, 1612.; Sali, 1613.; Silba, 1613.; Vrsar, 1613.; Pula, 1613.; Petrčane, 1614.; Baška, 1616.; Svinišće, 1620.; Podgora, 1621.; Primošten, 1622.; Sutivan, 1628.; Lopud, 1630.; Rogoznica, 1632.; Martinski, 1634.; Sv. Ilija (Zadar), 1637.; Bolnica Dubrovnik, 1637.; Pile, 1639; Boljun, 1640; Materada, 1640; Župa Dubrovnik grad, 1642; Murter, 1643; Medulin, 1648; Trsat, 1648; Brbinj, 1650; Novi Vinodolski, 1650; Sv. Ivan (Zagreb), 1650; Roč, 1651; Vrgada, 1651; Vranjic, 1653; Taborsko, 1653; Sv. Marka (Zagreb), 1653; Vrh, 1654; Zanković, 1656; Poreč, 1657; Grohote, 1658; Lipnik, 1659; Kloštar Ivanić, 1659; Sv. Filip i Jakov, 1659; Mađarevo, 1659; Bistra, 1659; Sv. Trojstvo (Karlovac), 1660; Makarska, 1662; Novigrad, 1662; Vižinada, 1663; Osekovo, 1664; Šterna, 1667; Tučepi, 1667; Granešina, 1667; Klis, 1669; Sisak, 1669; Ivanec, 1669; Šibenik varoš, 1669; Markuševac, 1669; Nedelišće, 1669; Tuhelj, 1670; Sveti Martin pod Okićem, 1670.; Pregrada, 1671.; Vrhovac, 1672.; Gornji Dubovec, 1672.; Kamenica (Ivanec), 1672.; Kastav, 1672.; Brdovec, 1672.; Šumber, 1673.; Hreljin, 1673.; Hum (Buzet), 1673.; Moravče, 1674.; Kostel, 1674.; Prelog, 1674.; Velika Gorica, 1675.; Stupnik, 1675.; Ludbreg, 1676.; Mrežnica, 1677.; Bela, 1677.; Krapinske Toplice, 1677.; Sv. Petar (Zagreb), 1677.; Konjščina, 1678.; Hrašćina, 1678.; Sesvete, 1678.; Vugrovec, 1678.; Bribir, 1678.; Visoko, 1678.; Zmino, 1679.; Začretje, 1679.; Gorica Svetojanska, 1679.; Novaki Motovunski, 1680.; Sveta Nedjelja, 1680.; Petrijanec, 1680.; Varaždinske Toplice, 1680.; Hrnetić, 1680.; Petrovina, 1680.; Podaca, 1681.; Veliko Trgovišće, 1681.; Vinica, 1681.; Donja Voća, 1681.; Rozga, 1683.; Gornja Stubica, 1683.; Gruž (Dubrovnik), 1683.; Bartolovec, 1684.; Maranovići, 1684.; Štrigova, 1685.; Šibenik-Dolac, 1685.; Virovitica, 1686.; Starigrad (Zadar), 1687.; Mala Gorica, 1687.; Vrlika, 1688.; Lobor, 1688.; Knin, 1689.; Križevci, 1689.; Skitača, 1690.; Karlobag, 1691.; Vlašići, 1691.; Omišalj, 1691.; Kostajnica, 1692.; Bokanjac, 1692.; Bakar, 1692.; Grižane, 1692.; Kutjevo, 1692.; Podjezerja, 1693.; Osijek, 1693.; Bednja, 1693.; Turanj, 1694.; Promina, 1694.; Sv. Juraj (kod Senja), 1695.; Maruševac, 1697.; Župa Juraj u trnju, 1698.; Zapuntel, 1699.; Sinj, 1699. i Orašac, 1699. *FamilySearch*, URL: https://www.familysearch.org/search/catalog/results?count=20&query=%2Bauthor_id%3A845152&page=1&topic0=Birth%2C+Marriage+and+Death&year0=1500, (10.4.2020.). zadnja posjeta 28.3.2020.

[39] Stipetić i Vekarić, *Povijesna demografija*, n. dj., str. 10.

[40] Stipetić i Vekarić, *Povijesna demografija*, n. dj., str. 13.

[41] Stipetić i Vekarić, *Povijesna demografija*, n. dj., str. 15.

demografskih analiza.[42] Povijesno demografska istraživanja počinju se značajnije razvijati u drugoj polovici 20. stoljeća.[43] Godine 1972. u JAZU je osnovan "Odbor za demografsku povijest", a svoj doprinos povijesnoj demografiji dali su još Ekonomski fakultet u Zagrebu, Filozofski fakultet u Zagrebu, Hrvatski institut za povijest, Institut za folkloristiku i drugi.[44]

Iako začetke genealogije i povijesno demografskih istraživanja pronalazimo u 19. stoljeću i ranije, njihov razvoj primarno je vezan uz period 20. stoljeća. Uzmemo li u obzir da su se podaci potrebni za takva istraživanja počeli sustavno prikupljati između 16. i 18. stoljeća, možemo uočiti relativno dugo vremensko razdoblje u kojem su podaci bili dostupni, ali se nisu koristili u znanstvene svrhe. Više je razloga tomu, poput nedostatka potrebnog kadra za obradu toliko podataka, skromne mogućnosti kombiniranja podataka iz više različitih izvora i slično. Međutim, ključni razlozi su u tome što, iako je količina prikupljenih podataka u odnosu na prijašnja razdoblja bio iznimno velika, njihova kvantiteta i kvaliteta ipak nije pružala mnogo potencijala. Drugi razlog je što podaci nisu bili fizički dostupni na jednom mjestu te treći činjenica da su bili u analognom, a ne digitalnom obliku. Tek nakon što su se navedeni problemi počeli postepeno rješavati, od sredine 20. stoljeća, navedeni se podaci intenzivnije koriste. Znanstvenici i amateri podatke još lakše, brže i kvalitetnije koriste od prijelaza s 20. na 21. stoljeće, kada se počinju prebacivati u digitalni oblik.

[42] Stipetić i Vekarić, *Povijesna demografija*, n. dj., str. 16.
[43] Stipetić i Vekarić, *Povijesna demografija*, n. dj., str. 9.
[44] Stipetić i Vekarić, *Povijesna demografija*, n. dj., str. 23.

6. Nove tehnologije u 20. i 21. stoljeću i Web 2.0

Jedan od ključnih preduvjeta za masovno prikupljanje podataka u digitalnom obliku je razvoj osobnih računala. Godine 1975. proizvedeno je prvo osobno računalo koje je moglo stati na stol, "Altair 8800".[45] Do 1980-ih osobna su računala bila toliko razvijena da ih je mogao koristiti gotovo bilo tko, od amatera do profesionalaca.[46] Tih je godina New York Times zapisao „Nacionalna računalna industrija ove bi godine trebala dostići prekretnicu: po prvi puta u SAD-u vrijednost prodanih osobnih računala – računala koja su bila gotovo nepoznata do prije samo osam godina – prestići će vrijednost prodanih središnjih računala (engl. mainframe) koja su postavila SAD kao lidera u računalnoj tehnologiji".[47] Broj korisnika osobnih računala je tada još uvijek bio relativno mali, svega nekoliko desetaka tisuća. Međutim, do 2000. godine ta je brojka narasla na preko 100 milijuna, što teoretski otvara potencijal prikupljanja podataka od više od 100 milijuna korisnika.[48] Drugi ključan preduvjet bio je internet. Iako je internet neizostavni faktor naše svakodnevice, njegovi počeci nisu bili toliko uspješni. Naime, ljudi ga nisu znali koristiti. Još 1996. godine Amazonova je podrška dobivala upite o tome kako se koristi traka za pomicanje (engl. *scroll bar*) na njihovoj web stranici.[49] Kreator hiperteksta, Ted Nelson, izjavio je da je internet samo jedna verzija papira. Prvi korisnici interneta nisu u potpunosti iskorištavali njegove potencijale.[50] Sukladno tomu, broj korisnika interneta je dugo bio mali. Tek od sredine 1990-ih je primjetan značaj rast broja korisnika.[51] Situacija se do 2000. godine značajno promijenila u odnosu na samo pet godina ranije. Broj korisnika osobnih računala koji su bili umreženi putem interneta bio je toliko velik da je počeo biti uočljiv trend za koji će se ispostaviti da će utjecati na gotove sve sfere naših života. Naime, nove tehnologije i internet su se do danas toliko razvili da gotovo ne postoji segment ljudskog postojanja na koji nisu utjecali.[52] To ih čini jednim od važnijih faktora globalizacije kao takve.[53]

Razvoj osobnih računala i interneta postao je rapidan. Već na prijelazu stoljeća počeli su se ocrtavati glavni trendovi onoga što će za par godina postati poznato kao Web 2.0.

---

[45] Ryan, Johnny, *A History Of The Internet And The Digital Future*. London: Reaktion Books, 2013, str. 54.
[46] Ryan, *A History Of The Internet*, n. dj., str. 60-61.
[47] Ryan, *A History Of The Internet*, n. dj., str. 61.
[48] Kanellos, Michael, *PCs: more than 1 billion served*, URL: https://www.cnet.com/news/pcs-more-than-1-billion-served/, (5.3.2020.).
[49] Ryan, *A History Of The Internet*, n. dj., str. 127.
[50] Ryan, *A History Of The Internet*, n. dj., str. 137.
[51] *Our World in Data*, Internet, URL: https://ourworldindata.org/internet, (6.9.2020.).
[52] Boban, "Pravni i kulturološki aspekti sigurnosti", n. dj., str. 51.
[53] Boban, "Pravni i kulturološki aspekti sigurnosti", n. dj., str. 49.

Jimmy Wales je 1999. godine odlučio napraviti skup znanja, svojevrsnu enciklopediju, u koju će svatko moći dati svoj doprinos.[54] To nije bio prvi Walesov projekt. "Nupedia" je bila slično zamišljena. Također se radi o enciklopediji u koju je svatko mogao dati svoj doprinos, međutim, članci su morali proći provjeru stručnjaka. To je bilo prekomplicirano pa je objavljeno malo članaka i projekt je u konačnici propao. Enciklopedija u koju svatko, bez uplitanja autoriteta iz znanstvene zajednice, može dati svoj doprinos pokrenuta je 2001. godine pod imenom "Wikipedija".[55] Web stranice u kojima su korisnici glavni ili čak jedini kreatori njena sadržaja do tada su bili nepoznanica. Tzv. Web 1.0 imao je iz perspektive korisnika definiran sadržaj kojem oni nisu mogli doprinijeti. S druge strane, Web 2.0 omogućio je korisnicima da daju svoje doprinose. To je u kontekstu prikupljanja osobnih podataka iznimno važno. Naime, tada se počinju masovno koristiti društvene mreže, koje ovise isključivo o doprinosima svojih korisnika.[56] Prva društvena mreža *Six Degrees* pokrenuta je još 1997. godine. Bila je povezana s e-mailom i nije imala opciju korištenja fotografija. One će ljudima postati dostupne tek za nekoliko godina, kada internet postane brži, a digitalni fotoaparati dostupniji. Ta prva društvena mreža je zatvorena 2001. godine. Uspješnija društvena mreža bila je *Friendster* pokrenuta 2003. godine. Ona je imala opciju korištenja fotografija te je relativno brzo imala velik broj korisnika. Problem je bio što je *Friendster* imao tehničkih problema zbog čega je stranica u konačnici ukinuta.[57] Kasnije, iste 2003. godine, pokrenut je *MySpace*, koji je imao još dramatičniji rast. Razvoj novih tehnologija omogućio je ubrzo i stavljanje videa, što je pak omogućilo pokretanje *YouTubea* 2005. godine.[58] U periodu oko 2004. godine pokrenuto je nekoliko društvenih mreža poput *Facebooka*, *Yahooa* i *Twittera*. S obzirom da su takve vrste stranica postale iznimno popularne, a kreatori njihovih sadržaja su njihovi korisnici, to je značilo da je broj dostupnih osobnih podataka postao iznimno velik. Činjenica da su podaci bili u digitalnom obliku samo je dodatno povećavala njihov potencijal. Nikada u prošlosti veća količina osobnih podataka nije bila dostupna za obradu.[59]

Osim činjenice da su osobna računala postala umrežena putem interneta te da je putem društvenih mreža sve više osobnih podataka dostupno javnosti, još je nekoliko faktora koji utječu na činjenicu da je svakim danom sve više podataka dostupno za obradu. Prvi je da

[54] Ryan, *A History Of The Internet*, n. dj., str. 146.
[55] Ryan, *A History Of The Internet*, n. dj., str. 147.
[56] Ryan, *A History Of The Internet*, n. dj., str. 148.
[57] Ryan, *A History Of The Internet*, n. dj., str. 149.
[58] Ryan, *A History Of The Internet*, n. dj., str. 150.
[59] Simon, Phil, *Too Big to Ignore: The Business Case for Big Data*. Hoboken, NJ : John Wiley & Sons, 2013, str. 34.

je moć računala postala iznimno velika, dok im je u istom trenutku cijena sve niža. Osim toga, razne manje elektroničke naprave postale su prenosive.[60] Sljedeći faktor je činjenica da su nove tehnologije postale gotovo pa sveprisutne te da prožimaju gotovo svaki aspekt naših života. Nikada se više radnji nije pratilo i snimalo.[61] Razvoj tehnologija omogućio je da se sve više podataka može pohraniti, a u kontekstu njihova organiziranja i pretraživanja važno je da se ti podaci danas lakše strukturiraju (ako se već ne prikupljaju u strukturiranom obliku) i indeksiraju.[62] Osim toga, strojno čitanje se iznimno razvilo. Čitanje barkoda ili očitavanje automobilskih registracija danas je jednostavno.[63] Ovo su samo neke od tendencija koje su utjecale i koje će utjecati na daljnji razvoj novih tehnologija, prikupljanja i obrade podataka.

[60] Kitchin, Rob, T*he Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Los Angeles: SAGE Publications Ltd, 2014, str. 81.
[61] Kitchin, *The Data Revolution*, n. dj., str. 83.
[62] Kitchin, *The Data Revolution*, n. dj., str. 84.
[63] Kitchin, *The Data Revolution*, n. dj., str. 85.

7. Privatnost

Koncept privatnosti vezan je uz razvoj individualnosti, koja se počinje afirmirati od 18. stoljeća. Danas je individualizam, a posredno i privatnost, jedan od postulata Zapadne civilizacije.[64] Pravo na privatnost jedno je od najosnovnijih ljudskih prava. S obzirom da su privatnost, individualnost i osobnost, iznimno važni u demokratskim društvima, povreda njih *de facto* je povreda te osobe u cjelini.[65] O pitanju nadzora počinje se ozbiljno raspravljati od 1975. godine kada je Michel Foucault objavio svoje djelo "Surveiller et puni" (hrv. "Nadzor i kazna").[66] Po Foucaultu, postoji nekoliko stupnjeva implementiranja discipline. U određenoj mjeri, navedeni stupnjevi su kronološki poredani. Prvi je mučenje, koje se masovno provodilo do prije nekoliko stoljeća. Cilj mučenja maksimalno fizički mučiti okrivljenika, pritom ga javno poniziti s ciljem nanošenja što više fizičke boli i poniženja prije smrti. Sljedeći stupanj je fizičko kažnjavanje, koje nužno ne dovodi do smrti. Posljednja faza je disciplineriranje. U njemu cilj "normalizirati" okrivljenika. Za Foucaulta panoptikum je simbol disciplineriranja i nadzora u modernom društvu.[67] Panoptikum je građevina u kojoj se u svakom trenutku svaku osobu ili direktno nadzire ili ju se može trenutno početi nadzirati. Ljudi su svjesni da ih se nadzire, samo ne znaju u kojem točno trenutku. S obzirom na to, cijelo se vrijeme ponašaju kao da ih se nadzire. Onaj koji ih nadzire nije im vidljiv. U takvoj situaciji, mučenja, fizička kažnjavanja, ili bilo koja vrsta fizičkog ograničavanja slobode više nije potrebna.[68] Iako je svoj koncept panoptikuma Foucault predstavio prije gotovo pola stoljeća, on je aktualan i u kontekstu nadzora na internetu. S jedne strane postoje oni koji smatraju da je nadzor na internetu u velikoj mjeri neutralan te da čak ima svoje pozitivne strane, poput obrane od raznih prijetnji.[69] S druge strane postoje oni koji smatraju da je nadzor na internetu iznimno negativna pojava. Naime, smatraju da oni koji su do tada bili u poziciji moći, sada putem internetskog nadzora dobivaju moćan alat s kojim mogu još efikasnije provoditi svoje interese.[70]

[64] Boban, "Pravni i kulturološki aspekti sigurnosti", n. dj., str. 11-12.
[65] Boban, "Pravni i kulturološki aspekti sigurnosti", n. dj., str. 14.
[66] Allmer, Thomas, "Critical Internet Surveillance Studies and Economic Surveillance", u *Internet and Surveillance: The Challenges of Web 2.0 and Social Media*, ur. Christian Fuchs, Kees Boersma, Anders Albrechtslund i Marisol Sandoval. New York: Routledge, 2012., str. 125.
[67] Allmer, "Critical Internet Surveillance", n. dj., str. 127.
[68] Allmer, "Critical Internet Surveillance", n. dj., str. 127.
[69] Allmer, "Critical Internet Surveillance", n. dj., str. 128.
[70] Allmer, "Critical Internet Surveillance", n. dj., str. 131-132.

Tri su najvažnije teorije privatnosti. To su one Alana Westina, Irwina Altmana, i Sandre Petronio.[71] Prema Westinovoj teoriji privatnost je definirana od strane pojedinaca ili grupa - oni određuju koliko privatnog će biti predstavljeno odnosno dostupno drugima. Granice privatnosti su dinamične, mijenjaju se ovisno o prilikama. Prema Westinovoj teoriji, četiri su stupa privatnosti: prvi je da je privatnost ne smije biti promatrana/nadgledana od strane drugih. Drugi je da se privatnost odnosi i na grupe, s ciljem ostvarenja bliskih i iskrenih odnosa. Treći je da u javnom prostoru anonimnost pojedinca mora biti osigurana - on ne smije biti nadziran, identificiran i slično. Posljednji je da granice privatnosti budu prepoznate od strane drugih te da drugi te granice poštuju. Prema istom autoru, četiri su razloga zašto je potrebna privatnost. Prvi je da se kroz privatnost pojedinci nastoje zaštiti od manipulacija i iskorištavanja. Drugi je da privatnost služi za emotivnu relaksaciju. Naime, društveni kontakti s vremenom postanu teret pojedincu te mu je potreban svojevrstan odmor - kroz privatnost dobiva upravo to. Treći je da se privatnost iskorištava kao prilika za samoevaluaciju, analiziranje iskustava, konstruiranje/korigiranje vlastiti moralnih i/ili religioznih uvjerenja. Posljednji je da privatnost u kontekstu komunikacija omogućava s jedne strane postavljanje granica (što je pojedinac s kime voljan komunicirati) te s druge strane omogućuje uspostavljanje komunikacijskih kanala putem kojih se razmjenjuju intimnije informacije.[72]

Druga teorija je Altmanova. Njegova je teorija donekle slična Westinovoj. Altman smatra da je pitanje otvorenosti odnosno zatvorenosti oko privatnosti prilično dinamično i definirano internim i eksternim razlozima. Prema Altmanu, pet je svojstava privatnosti. Prvo je da privatnost podrazumijeva dinamičan proces definiranja međuljudskih granica. Drugo je da postoji diskrepancija između željene i stvarne razine privatnosti. Treće je da su granice privatnosti fluidne te da nije uvijek moguće ostvariti željenu razinu privatnosti. Ponekad je ima previše, a ponekada premalo. Četvrto je da su granice privatnosti definirane iz dva pravca. S jedne strane definira ih odnos pojedinca prema drugima te s druge odnos drugih prema pojedincu. Posljednje je da je privatnost prisutna na individualnoj i grupnoj razini.[73]

Petronijeva teorija (Communication Privacy Management - CPM teorija) najvažnija je teorija u kontekstu društvenih mreža. Kao i kod Altmana, granice privatnosti su dinamične i ovise o situaciji. Pet je postavki CPM teorije. Prva je da je "vlasništvo" nad privatnim informacijama definirano činjenicom da li određeni pojedinac informaciju smatra "svojom".

[71] Margulis, Margulis,"Three Theories of Privacy: An Overview," *Privacy Online: Perspectives on Privacy and Self-Disclosure in the Social Web,* ur. Sabina Trepte i Leonard Reinecke. New York: Springer, 2011, str. 9.
[72] Margulis, "Three Theories", n. dj., str. 10.
[73] Margulis, "Three Theories", n. dj., str. 11.

Ukoliko se smatra vlasnikom određene informacije, to je za njega dovoljan kriterij da bi sam sebe smatrao vlasnikom. Druga je da ukoliko osoba smatra određenu informaciju privatnom, da onda ima pravo koristiti tu informaciju kako god želi.[74] Treća je da ljudi postavljaju granice vezane uz njihovu privatnost. Te granice ovise o kulturnom okruženju, trenutnoj situaciji i motivaciji.[75] Četvrto je da jednom kada se privatna informacija podijeli s drugima oni postaju suvlasnici te informacije. Suvlasnici, barem iz perspektive originalnog vlasnika, imaju odgovornost da s tom informacijom postupaju odgovorno. Taj odnos vlasnika i suvlasnika određen je prije dijeljenja informacije, tako da novi suvlasnici znaju svoje odgovornosti. Peta je da ukoliko dođe do kršenja pravila od strane suvlasnika dolazi do redefiniranja odnosa između vlasnika i suvlasnika informacije.[76] Ovo su samo glavna teorijska polazišta o pitanju privatnosti. S obzirom da se radi o kompleksnoj temi, postoji još mnogo različitih perspektiva.[77]

[74] Margulis, "Three Theories", n. dj., str. 12.
[75] Margulis, "Three Theories", n. dj., str. 12-13.
[76] Margulis, "Three Theories", n. dj., str. 13.
[77] Margulis, "Three Theories", n. dj., str. 14.

8. Zaštita privatnih podataka u 21. stoljeću

Država, tj. njene institucije još od 18. stoljeća rade na sakupljanju podataka o svojim građanima. Na taj se način posredno želi, između ostalog, ostvariti određena razina nadzora i kontrole nad stanovništvom.[78] Danas se sve razine i vrste državnih tijela bave sakupljanjem osobnih podataka. To se posebno odnosi na socijalnu skrb i obrazovne ustanove. Država sve više ulaže u bolje tehnologije za obradu podataka, što znači da će država imati dosta velik nadzor nad ljudima, a to neke zabrinjava.[79] Osim države, prikupljanjem podatka bave se i razne druge organizacije, uglavnom velike korporacije. One na taj način prate produktivnost svojih zaposlenika, navike kupaca, analizom podataka mogu smanjiti troškove poslovanja te općenito imaju bolju podlogu za donošenje kvalitetnijih odluka.[80] Korištenje informacija i podataka u poslovanju staro je koliko i kapitalizam, ali raširena uporaba tih resursa je relativno nova pojava. Sve je započelo 1980-ih automatizacijom procesa u kojima su bili strukturirani podaci. S vremenom se sve više pažnje posvećivalo podacima, međutim, to su bili podaci iz interne uporabe, dakle nastali unutar poslovnog subjekta.[81] Ključne promjene počinju se događati sredinom 2000-ih s nastankom Weba 2.0. Tada vrsta i količina podataka na webu raste iznimno brzo, što po prvi puta korporacijama omogućava masovno prikupljanje podataka iz vanjskih izvora. Od tada privatni poslovni subjekti poput Facebooka i Amazona postaju važni faktori na tržištu prikupljanja (osobnih) podataka.[82] Time je broj subjekata koji mogu prikupljati i koristiti osobne podatke narastao, čime su dodatno narasle prijetnje zlouporabe osobnih podataka. Posljedično, porasla je razina svijesti o tome problemu.

Privatnost se štiti na tri razine - pravnoj, etičkoj i tehnološkoj - koje štite privatnost. Pravna zaštita privatnosti gotovo je svugdje prisutna, ali dosta varira od zemlje do zemlje. UN ova deklaracija o ljudskim pravima iz 1948. godine i Europska konvencija o ljudskim pravima iz 1950. godine eksplicitno spominju pravno na privatnost kao pravo. To se odnosilo na osobnu privatnost, privatnost doma i slično. Situacija se drastično promijenila s novim tehnologijama i društvenim mrežama. Druga vrsta zaštite je etička samoregulacija (engl. *ethical self-regulation*).[83] To se npr. odnosi na određeno ne otkrivanje informacija obrazovnim institucijama ili zdravstvenim djelatnicima ili pak u poslovnoj komunikaciji.

[78] Kitchin, *The Data Revolution*, n. dj., str. 114.
[79] Kitchin, *The Data Revolution*, n. dj., str. 115.
[80] Kitchin, *The Data Revolution*, n. dj., str. 117.
[81] Simon, *Too Big to Ignore*, n. dj., str. 33.
[82] Simon, *Too Big to Ignore*, n. dj., str. 34.
[83] Debatin, Berhard, "Ethics, Privacy, and Self-Restraint in Social Networking", *Privacy online,* str. 49.

Čuvanje privatnosti pomoću tehnologija podrazumijeva implementiranje tehnologija u proces zaštite.[84] Kod pristupa privatnim informacijama postoje "sive zone" u kojima je pitanje zaštite privatnosti diskutabilno. Ponekad se trebaju objaviti nečiji osobni podaci radi slučaja koji je od javnog interesa. Tada se trebaju provjeriti dvije stvari. Prvo, može li ste to riješiti na bilo koji drugi način osim preko objavljivanja osobnih podataka te drugo, je li javni interes doista toliko velik da opravdava objavljivanje osobnih podataka. Ovo posljednje je problematično jer se može različito tumačiti. Ukoliko se već mora ići u smjeru objavljivanja onda se mora dogoditi jedna od dvije stvari. Prva je da je javni interes veći od potencijalne štete koju objavljivanje osobnih podataka može napraviti pojedincu i drugo, osoba čiji su osobni podaci u pitanju mora pristati na objavu.[85] Načelno se lako složiti da je moguća situacija da je javni interes iznad privatnog, ali što ako se dogodi slučaj da nema javnog interesa, nego da su u pitanju država i korporacije?[86] Sve veća ugroza osobnih podataka i sve veća kompleksnost pitanja zaštite takvih podataka zahtijevali su bolju pravnu reguliranost tog područja.[87] Prvi ozbiljniji napori da se osobni podaci zaštite na pravnoj razini su već spomenute Deklaracija UN-a i Europska konvencija o ljudskih pravima iz sredine 20. stoljeća. Ipak, u kontekstu Weba 2.0 najvažniju pravnu zaštitu donio je GDPR iz 2016. godine. Radi se o dokumentu koji vrijedi samo u članicama Europske unije, međutim, GDPR je zbog svoje sveobuhvatnosti i kompleksnosti dodatno potaknuo rasprave o potrebi zaštite osobnih podataka. Kalifornija je 2019. godine donijela sličan zakon o zaštiti osobnih podataka upravo po uzoru na zakonodavce u Bruxellesu.[88] GDPR je jedna od najkompleksnijih regulativa koje EU ima i kompanije se žale da je njegova implementacija problematična te da će imati negativnih posljedica.[89] Glavni mu je cilj da pojedinci imaju više kontrole nad time kako i tko koristi njihove osobne podatke, a kompanije da bolje paze na podatke koje pohranjuju. Regulativa se ubrzo nakon implementacije našla na udaru kritika. Naime, neki tvrde da je GDPR podbacio u onome za što je bio namijenjen, tj. da štiti osobne podatke.[90] Osim toga, neke zemlje članice EU-a se optužuju da ne provode dovoljno efikasno

[84] Debatin,"Ethics", n. dj., str. 50.
[85] Debatin,"Ethics", n. dj., str. 52.
[86] Debatin,"Ethics", n. dj., str. 53.
[87] Boban, "Pravni i kulturološki aspekti sigurnosti", n. dj., str. 25.
[88] Swisher, Kara, *Be Paranoid About Privacy,* URL: https://www.nytimes.com/2019/12/24/opinion/location-privacy.html?searchResultPosition=3 (1.4.2020.).
[89] Europe's tough new data-protection law, URL: https://www.economist.com/business/2018/04/05/europes-tough-new-data-protection-law (1.4.2020.).
[90] Hare, Stephanie, *These new rules were meant to protect our privacy. They don't work*, URL: https://www.theguardian.com/commentisfree/2019/nov/10/these-new-rules-were-meant-to-protect-our-privacy-they-dont-work (1.4.2020.).

GDPR i tako *de facto* stvaraju prostor za manipulacije.[91] Ilustrativno je da je do ožujka 2020. godine Njemačka izdala čak 208 kazni temeljem GDPR-a, a Irska niti jednu.[92] Prve primjedbe temeljem GDPR-a odnosile su se na Google i Facebook, a došle su svega 48 minuta nakon uvođenja odredbe. Kazne za kršenje odredbi GDPR-a mogu biti iznimno visoke.[93] Najveće kazne su bile ona dodijeljena Googleu od 50 milijuna EUR, potom TIM-u u iznosu od gotovo 28 milijuna EUR, Austrian Post je dobio kaznu od 18 milijuna EUR, a Deutsche Wohne SE kaznu od gotovo 15 milijuna EUR.[94] Do sada je u ime kršenja GDPR-a naplaćeno preko 153 milijuna EUR. Najveća naplaćena kazna, ona Googleu, obrazložena je time da Google nije bio dovoljno jasan, tj. transparentan oko prikupljanja i korištenja podataka svojih korisnika.[95] Najveće dosuđene, ali za sada još nenaplaćene kazne (pravni postupak nije u potpunosti gotov) su British Airwaysu (BA) u iznosu od gotovo 205 milijuna EUR. BA-u su hakirane stranice i ukradeni osobni podaci putnika.[96] Iznimno veliku kaznu, koja također još nije naplaćena, dobio je lanac hotela Marriott International u iznosu od preko 110 milijuna EUR. U cyber napadu ukradeni su osobni podaci gostiju.[97] U Hrvatskoj je donesena jedna kazna temeljem GDPR-a i to od strane Agencije za zaštitu osobnih podataka. Kaznu je dobila neimenovana banka u Zagrebu, a iznos kazne nije naveden. Kazna je izrečena u ožujku 2020. godine.[98]

[91] Stupp, Catherine, *Brussels Report: Frustated European Regulator Seek Tweaks to GDPR*, URL: https://www.wsj.com/articles/brussels-report-frustrated-european-regulators-seek-tweaks-to-gdpr-11584135669?mod=searchresults&page=1&pos=6 (1.4.2020.).
[92] Stupp, Catherine, *Brussels Report: Frustated European Regulator Seek Tweaks to GDPR*, URL: https://www.wsj.com/articles/brussels-report-frustrated-european-regulators-seek-tweaks-to-gdpr-11584135669?mod=searchresults&page=1&pos=6 (1.4.2020.).
[93] Roose, Kevin, *The Privacy Lawyer Giving Big Techan $8.8 Billion Headache*, URL: https://www.nytimes.com/2018/05/30/technology/gdpr-privacy-lawyer.html?auth=login-google&searchResultPosition=15 (1.4.2020.).
[94] *GDPR Fines Tracker & Statistics*, URL: https://www.privacyaffairs.com/gdpr-fines/, (1.4.2020.).
[95] *5 biggest GDPR fines so far [2020]*, URL: https://dataprivacymanager.net/5-biggest-gdpr-fines-so-far-2020/, (1.4.2020.).
[96] *5 biggest GDPR fines so far [2020]*, URL: https://dataprivacymanager.net/5-biggest-gdpr-fines-so-far-2020/, (1.4.2020.).
[97] *5 biggest GDPR fines so far [2020]*, URL: https://dataprivacymanager.net/5-biggest-gdpr-fines-so-far-2020/, (1.4.2020.).
[98] *Rješenje kojim se izriče upravno novčana kazna zbog odbijanja dostave osobnih podataka*, URL: https://azop.hr/aktualno/detaljnije/rjesenje-kojim-se-izrice-upravno-novcana, (1.4.2020.).

9. Velike količine podataka

Termin velike količine podataka vezuje se uz prikupljanje i obradu digitalnih podataka. Etimologija termina datira iz 1990-ih, ali se tada rijetko koristila. Učestalije se počinje pojavljivati krajem 2000-ih.[99] S obzirom da rad u velikoj mjeri obrađuje velike količine podataka, pojam je potrebno definirati i odrediti. Ne postoji konsenzus što su točno velike količine podataka, međutim, često se spominju tri "V", "Volume" (količina), "Velocity" (brzina, pristupačnost) i "Variety" (raznolikost). Količina se odnosi na činjenicu da su velike količine podataka iznimno brojne. Brzina se odnosi na činjenicu da velika količina podataka nastaje u kratkome roku te da su oni odmah ili vrlo brzo nakon svog nastanka dostupni za analizu. Raznolikost se odnosi na raznolikost podataka. I prije stvaranja koncepta velikih količina podataka postojala su tri "V", međutim, u određenom trenutku mogla su postojati samo dva "V", dok su kod velikih količina podataka sva tri prisutna u istom trenutku.[100] Kao i s nekim drugim pojmovima, poput privatnosti, velike količine podataka su kompleksna tema na koju postoji više različitih pogleda. Sukladno tomu, postoji i više različitih definicija velikih količina podataka. Kenneth Cukier i Viktor Mayer-Schönberger velike količine podataka definiraju kao "faktore koji se mogu iskoristiti samo u velikim obujmima, pomoću kojih se mogu dobiti novi uvidi ili stvoriti neki novi oblici vrijednosti koji pozitivno utječu na tržišta, organizacije, odnose između građana, institucija i slično".[101] Prema McKinsey & Company termin velike količine podataka se odnosi na "skupove podataka čija je veličina tolika da konvencionalni softverski alati nisu u mogućnosti prikupiti i analizirati toliku količinu podataka. U definiciji se namjerno ne spominje koja količina podataka je potreba da se ti podaci smatrali velikim količinama podataka. Pretpostavljamo da će se, kako se tehnologija s vremenom sve više razvija, povećati i veličina baza podataka koja je potreba da bi tu bazu nazvali velikom bazom. Također treba imati na umu da se definicija može razlikovati po sektorima, ovisno o tome koja vrsta baza podataka je uobičajena u određenom sektoru. Uzevši sve to u obzir, za velike količine podataka danas u većini sektora možemo smatrati baze podataka veličine između nekoliko desetaka terabajta i nekoliko petabajta".[102] Tom Davenport za velike količine podataka smatra da iako se radi

[99] Kitchin, *The Data Revolution*, n. dj., str. 67.
[100] Kitchin, *The Data Revolution*, n. dj., str. 68.
[101] Mayer-Schönberger, Viktor i Kenneth Cukier, *Big Data: a Revolution That Will Transform How We Live, Work and Think*. London: John Murray, 2017, str. 6.
[102] Manyika, James, Michael Chui, Brad Brown, Jacques Baghin, Richard Dobbs, Charles Roxburgh i Angela Hung Byers. *Big data: The next frontier for innovation, competition and productivity.* McKinsey Global Institute, 2011. URL:

"nedvojbeno o kvantitativno brojnim podacima, termin velike količine podataka, nije sasvim prikladan. Termin se počeo upotrebljavati za sve podatke koji nisu unutar konvencionalnih okvira obrade podataka. Velike količine podataka odnose se na one količine su prevelike da bi stali u jedan server, previše nestrukturirani da bi bili primjereni za format baze koja strukturira podatke u redove i stupce te podatke koji se neprestano nadopunjuju novim podacima, što znači da statične baze podataka nisu primjerene. Iako veličina velikih količina podataka privlači najviše pozornosti, zapravo je njihov najkompleksniji element - nestrukturiranost.[103] Davenport smatra da je termin velike količine podataka problematičan. Riječ "big" (veliki, brojni), implicira da se nužno radi o mnogobrojnim podacima, što je istina, ali ljudima koji rade s velikim količinama podataka njihova veličina rijetko je problem. Puno je veći problem već spomenut nedostatak strukture podataka. Termin velike količine podataka raširen je jer zvuči moderno. Autor smatra da je termin u suštini pogrešan te da će nestati. Ono što će naravno ostati je fenomen pod kojim danas podrazumijevamo velike količine podataka.[104]

https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_full_report.ashx, 1.

[103] Davenport, Thomas, *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities*. Boston: Harvard Business Review Press, 2014, str. 1.

[104] Davenport, *Big Data at Work*, n. dj., str. 8.

10. Povijesni razvoj i karakteristike velikih količina podataka

Za razumijevanje velikih količina podataka potrebno je razumjeti njegovu osnovnu komponentu - podatke. Više je vrsta podataka i kategorizacija podataka. Jedna od kategorizacija je na kvantitativne i kvalitativne podatke. Kvantitativni podaci su numerički, mogu se odnositi na fizičke atribute poput visine, duljine i sl. ili na neke nefizičke elemente određenog fenomena poput razine obrazovanja, razine pismenosti, razine nejednakosti i slično.[105] Kvalitativni podaci su nenumerički poput fotografija, slika i audio zapisa.

Druga kategorizacija je na strukturirane, polustrukturirane i nestrukturirane podatke. Strukturirani podaci su sređeni, oni su npr. u stupcima i organizirani su određenom logikom koju analitičar može primijeniti prilikom rada na takvim bazama podataka. Polustrukturirani imaju elemente strukturiranih i nestrukturiranih podataka.[106] Nestrukturirani podaci nemaju reda. često su nastali u različitom kontekstu, imaju različite strukture, teško ih je iskoristiti odnosno organizirati.

Podatke je moguće kategorizirati na prikupljene, sekundarne, prolazne i izvedene. Prikupljeni podaci se odnose na podatke koji su dobiveni nekom vrstom promatranja, ankete, skeniranja itd., dakle ti podaci su ciljano prikupljeni s određenom svrhom. Sekundarni podaci su podaci nastali nekom drugom radnjom kojoj nije primarni cilj stvaranje/prikupljanje tih podataka. Prolazni podaci su podaci koji su toliko masivni i nestrukturirani da ih se odluči jednostavno ne iskoristiti. Izvedeni podaci su podaci koji se kombiniraju s drugim podacima, dodatno su analizirani.[107]

Nadalje, podaci se mogu dijeliti i na primarne, sekundarne i tercijarne podatke. Primarni podaci su oni podaci koje određeni subjekt stvori ili prikupi s ciljem da ih upotrebljava. Sekundarni je kad subjekt stvori ili prikupi podatke i da ih nekom drugom da ih koristi (što je nekom primarni podatak, nekom drugom je sekundarni).[108] Tercijarni podaci su oni prikupljeni podaci koji se javno objavljuju, npr. statistički godišnjaci i slično.[109]

Neki podaci su nastali u digitalnom obliku (engl. *born digital*). S njima se lako manipulira, a njihovo stvaranje je često automatizirano ili poluautomatizirano. Njihov

[105] Kitchin, *The Data Revolution*, n. dj., str. 4.
[106] Kitchin, *The Data Revolution*, n. dj., str. 5.
[107] Kitchin, *The Data Revolution*, n. dj., str. 6.
[108] Kitchin, *The Data Revolution*, n. dj., str. 7.
[109] Kitchin, *The Data Revolution*, n. dj., str. 8.

nedostatak je da oni mogu biti lako izgubljeni u odnosu na podatke nastale u analognom obliku. Dok je njihova prednost ta da ih se može odmah i lako analizirati.[110]

Obrada podataka nije nastala s velikim količinama podataka. Sustavne analize većih količina podataka korporacije prakticiraju od sredine 20. stoljeća, samo što su tada terminologija i kontekst drugačiji. Podaci su bili analogni i prikupljali su se interno unutar poslovnog subjekta. Analiza podataka trajno prolazi kroz brojne promjene što mijenja njezino značenje i ulogu. Velike količine podataka u današnjem smislu formiraju se tek u 21. stoljeću. Njihov nastanak i razvoj donekle se oslanja na sada već zaboravljene konvencionalne načine obrade podataka. Tom Davenport sistematizirao je razvoj analize podataka od 1970-ih pa sve do razvoja velikih količina podataka (Tablica 2).

Tablica 2. Značenje termina "analiza podataka"

Termin	**Razdoblje**	**Značenje**
Potpora donošenju odluka	1970.-1985.	Analiza podataka s ciljem podržavanja donošenja odluka
Izvršna podrška	1980.-1990.	Fokus na analizu podataka od strane visokorangiranih rukovoditelja
Internetska analitička obrada	1990.-2000.	Softver za analizu višedimenzionalnih podataka
Poslovna inteligencija	1989.-2005.	Alati za podršku odlukama temeljem podataka s naglaskom na izvještavanje
Analitika	2005.-2010.	Fokus na statističke i matematičke analize kao podrška donošenju odluka
Velike količine podataka	2010.-danas	Iznimno velike, nestrukturirane i fluidne baze podataka

[110] Kitchin, *The Data Revolution*, n. dj., str. 32.

Nije svaka terminološka promjena bila u potpunosti opravdana. Neke promjene u analizi podataka bile su relativno male.[111] Ipak, uvođenje termina velikih količina podataka u potpunosti je opravdan jer je analiza podataka s njima doživjela iznimno velike promjene, a podaci postaju digitalni i nestrukturirani.[112] Ključne razlike između velikih količina podataka i konvencionalnih metoda analize podataka su prikazane u Tablici 3.

Tablica 3. Razlike između velikih količina podataka i konvencionalnih analiza[113]

	Velike količine podataka	**Konvencionalna analitika**
Tip podataka	Nestrukturirani	Strukturirani
Veličina podataka	Od 100 terabajta do više petabajta	Deseci terabajta ili manje
Tok podataka	Konstantan priljev novih podataka	Fiksan obujam podataka
Metode analize	Strojno učenje	Temeljem hipoteza
Primarna svrha	Proizvodi nastali temeljem analize podataka	Potpora internom donošenju odluka

Ovakav pregled razvoja analize podataka ne objašnjava zašto je došlo do razvoja velikih količina podataka, tj. koji su bili preduvjeti za njihov nastanak. Phil Simon navodi nekoliko ključnih razloga. Stalno uključeni potrošač (engl. *the always-on consumer*) odnosi se na činjenicu da su elektroničke naprave (engl. *gadgets*) uvijek "pri ruci" što znači da se uvijek koriste. To je važno jer se putem takvih naprava prikupljaju velike količine podataka.[114] Drugi je sve brži razvoj tehnologija, što posljedično utječe na njihov povećan kapacitet i sve manju cijenu.[115] Ova pojava uočena je prije pola stoljeća, poznata je kao Mooreov zakon, koji tvrdi da se broj tranzistora udvostručuje otprilike svake dvije godine.[116] Nasuprot njemu Kryderov zakon nastao sredinom 2000-ih tvrdi da se gustoća pohrane

[111] Davenport, *Big Data at Work*, n. dj., str. 10.
[112] Davenport, *Big Data at Work*, n. dj., str. 11.
[113] Davenport, *Big Data at Work*, n. dj., str. 4.
[114] Simon, *Too Big to Ignore*, n. dj., str. 13.
[115] Simon, *Too Big to Ignore*, n. dj., str. 14.
[116] *Moores law*, URL: http://www.mooreslaw.org/, (20.4.2020.).

magnetskog pogona udvostručuje svakih 13 mjeseci.[117] Treći je promjena odnosa prema podacima i informacijama. Termin *infonomics* (spoj riječi "informacija" i "ekonomija") nastao je krajem 1990-ih u kontekstu tendencija da se informacije počinju smatrati sve vrjednijim poslovnim resursom. Osim toga, Google je uvelike zaslužan za bolje pretraživanje informacija. Google tražilica doživjela je globalni uspjeh temeljem dostavljanja relevantnih informacija svojim korisnicima.[118] Osim ispunjenja određenih preduvjeta, razvoj velikih količina podataka dodatno su ubrzala dva konkretna događaja. Prvi je bio vezan uz izbore u Sjedinjenim Američkim Državama (SAD). Pred same izbore Nate Silver je pomoću analize velikih količina podataka predvidio da će Barack Obama, tada autsajder, postati predsjednikom. Konvencionalna predizborna istraživanja nisu bila u skladu sa Silverovim predviđanjima, što je njegovo točno predviđanje dodatno istaknulo nakon što se pokazalo da je imao precizniju metodologiju.[119] Drugi događaj je iz 2012. godine kada je *National Weather Service* (NWS) iz SAD-a pomoću velikih količina podataka predvidio da će uragan Sandy pogoditi sjeveroistok SAD-a. Procjena se pokazala točnom, a s obzirom da su upozorenja ozbiljno shvaćena, područje se pravovremeno pripremilo za uragan čime je umanjena šteta od nepogode.[120]

Prema Robu Kitchinu, nekoliko je karakteristika velikih količina podataka. Prva je već spomenuta količina, koja postaje sve veća.[121] Druga je iscrpnost - odnosi se na to koga se sve obuhvaća i koji podaci se prikupljaju. Velike količine podataka, zahvaljujući suvremenim tehnologijama, obuhvaćaju velik broj ljudi i iznimno brojne i raznolike podatke.[122] Treća je rezolucija, što podrazumijeva da su podaci detaljniji i kvalitetniji te da ih se lakše pretražuje.[123] Četvrta je povezanost. Veliki podaci su povezani u odnosu na druge podatke, što ih čini još korisnijima, jer mogu služiti za interpretaciju ili nadopunjavanje drugih podataka. To vrijedi čak i u slučajevima kada primarni cilj nastanka ili obrade nije određivanje povezanosti. Nakon što se sakupi iznimno velika količina podataka u pravilu se onda ti podaci počnu preklapati, tj. povezuju se bez da ih je netko nužno skupljao s ciljem da utvrdi njihovu povezanost.[124] Obama je to svojevrsno umrežavanje (engl. *networking*) velikih količina podataka iskoristio u svojim predsjedničkim kampanjama. Prikupio je sve

[117] PCMag eciklopedija: „Kryders law", URL: https://www.pcmag.com/encyclopedia/term/kryders-law, (20.4.2020.).
[118] Simon, *Too Big to Ignore*, n. dj., str. 18.
[119] Simon, *Too Big to Ignore*, n. dj., str. 20-21.
[120] Simon, *Too Big to Ignore*, n. dj., str. 20.
[121] Kitchin, *The Data Revolution*, n. dj., str. 69.
[122] Kitchin, *The Data Revolution*, n. dj., str. 72.
[123] Kitchin, *The Data Revolution*, n. dj., str. 73.
[124] Kitchin, *The Data Revolution*, n. dj., str. 74.

razne vrste podataka o (potencijalnim) glasačima te ih kombinirao s dostupnim istraživanjima poput popisa stanovništva i slično. Temeljem toga kreirana je kampanja.[125] Peta i šesta karakteristika su već spomenute brzina i raznolikost.[126] Sedma karakteristika je fleksibilnost. Baze podataka koje su konvencionalne često nastaju, i naknadno se popunjavaju, po određenoj formi. To s velikim količinama podataka nije slučaj. Podaci u takvim bazama lako se korigiraju, nadopunjuju i premještaju.[127]

Iako velike količine podataka u velikoj mjeri proizlaze s Weba 2.0 odnosno društvenih mreža, takve podatke moguće je prikupiti i na druge načine. Jedna od prvih popularnih aplikacija koja je analizirala osobne podatke bila je *Nike+ shoes* iz 2006. godine koja je nastala u suradnji Nike+ i iPod-a. Tenisica bila je povezana s iPodom i pratila je određene faktore vezane uz tjelesne aktivnosti. Korisnik je aplikaciju koristio upravo s ciljem da prikuplja podatke koje on, i vlasnici aplikacije mogu koristiti.[128] Puno kompleksnije prikupljanje velikih količina podataka koje nije vezano uz Web 2.0 započinje 2003. godine. Te je godine Oren Etzioni, koji je diplomirao informatiku na Harvardu, rezervirao avionsku kartu par mjeseci prije leta vjerujući da će više uštedjeti što ranije kupi kartu. Na samom letu bio je znatiželjan te je ispitivao suputnike o njihovoj cijeni karte te saznao da unatoč ranijoj kupnji nije ništa uštedio. Odlučio je analizirati velike količine podatke vezane uz avionske karte.[129] Analizirao je prodaju karata i vidio koje su cijene bile istaknute u određenom periodu. Nije mogao saznati zašto su određene cijene prisutne u određenom trenutku, ali je mogao vidjeti tendencije u kojim periodima cijene padaju, a u kojima rastu. Prema tome je mogao odrediti treba li kupiti kartu odmah ili još sačekati. Prvo je analizirao samo period od mjesec dana, a kasnije je počeo analizirati puno više podataka u duljem vremenskom periodu.[130] To se pretvorilo u poduzetnički projekt i velik broj kupaca kupovao je karte temeljem te analize. Kasnije je Etzioni planirao isto napraviti i za hotele i koncerte, ali to nije stigao napraviti jer je njegov projekt kupio Microsoft za 110 milijuna USD.[131] Navedeni primjer ilustrira jednu od karakteristika velikih količina podataka. Naime, velike količine podataka daju odgovore na pitanje što, ali ne i zašto. Odgovor na pitanje zašto može se eventualno dobiti nakon rezultata, ali to onda više nije analiza velikih količina podataka u

[125] Kitchin, *The Data Revolution*, n. dj., str. 75.
[126] Kitchin, *The Data Revolution*, n. dj., str. 77.
[127] Kitchin, *The Data Revolution*, n. dj., str. 77-78.
[128] Davenport, *Big Data at Work*, n. dj., str. 12.
[129] Mayer-Schönberge i Cukier, *Big data*, n. dj., str. 3.
[130] Mayer-Schönberge i Cukier, *Big data*, n. dj., str. 4.
[131] Mayer-Schönberge i Cukier, *Big data*, n. dj., str. 5.

užem smislu.[132] Podaci za analizu velikih količina podataka mogu se dobiti od brokera podataka. To su pravne osobe koje prikupljaju podatke koje onda prodaju, preprodaju ili daju na korištenje uz određenu cijenu. Podatkovni brokeri pojavljuju se masovnije nakon nastanka Weba 2.0.[133]

Baze podataka teško se izrađuju - to je ponekad dugotrajan posao, potrebne su određene tehnologije i stručnjaci. S obzirom da država, tj. njene institucije već stoljećima sustavno rade na prikupljanju podataka o svojim građanima njihove digitalne baze podataka su neke od najvećih i najkompleksnijih te su stoga i iznimno korisne. Privatni sektor je lobirao da se te baze podataka učine javno dostupnima. Na taj način moguće je uštedjeti iznimno mnogo vremena i novaca. Tako je Pokret za otvorene podatke (engl. *the open data movement*) nastojao je vršiti pritisak na vlade s ciljem otvaranja njihovih baza podataka.[134] Krajem 2000-ih vlade su počele popuštati takvim pritiscima. The Guardian je u Ujedinjenoj Kraljevstvu imao kampanju Oslobodite naše podatke (engl. *Free Our Data*). Vlada u SAD-u napravila je web stranicu *data.gov* u kojoj su dostupni podaci od povijesne važnosti (ne sadrže osjetljive podatke). Kasnije su slično učinile i druge države i organizacije poput EU-a, UN-a itd. U tome su pomogle i grupacije koje su lobirale za takve stvari, kao što su Open Knowledge Foundation i Sunlight Foundation. S vremenom se dogodila paradoksalna situacija - počelo se uviđati da otvaranje svih tih podataka nije uvijek dobro, tako da se pojavljuju i kritike. Još jedna otežavajuća okolnost je što se u zadnjih pola stoljeća počelo puno ulagati u pravnu zaštitu osobnih podataka i intelektualnih prava - upravo u trenutku kada se počinje s otvaranjem takvih vrsta podataka.[135]

Otvaranje podataka u vlasništvu države pozitivna je stvar za korporacije, ali one i bez toga imaju veliki potencijal s velikim količinama podataka koje same prikupe. Facebook je 2010. godine pokrenuo kampanju Glasački megafon (engl. *voter megaphone*) u kojoj se ljudima omogućilo da na svoj Facebook profil stavljaju obavijest da su glasali te tako potaknu i druge da učine isto.[136] Poticanje na glasanje se naizgled čini benigno, ali s obzirom da je Facebook korporacija koja ima podatke o velikom dijelu globalne populacije, ta kampanja ima veliki potencijal. Procjene su govorile da je ta kampanja potaknula na glasanje oko 340.000 ljudi, što je velika stvar, pogotovo ako se radi o saveznim državama SAD-a za koje je teško predvidjeti kojem će kandidatu većina glasača dati svoj glas (engl. *swing states*).

[132] Mayer-Schönberge i Cukier, *Big data*, n. dj., str. 7.
[133] Kitchin, *The Data Revolution*, n. dj., str. 42.
[134] Kitchin, *The Data Revolution*, n. dj., str. 48.
[135] Kitchin, *The Data Revolution*, n. dj., str. 49.
[136] O'Neil, Cathy, *Weapons Of Math Destruction*. London: Penguin books, 2017, str. 180.

George Bush je na izborima 2000. godine pobijedio na Floridi za manje od 500-tinjak glasova. Uzevši to u obzir, ako Facebook odluči manipulirati s glasačkim tijelom, mogu rezultate izbora usmjeriti u određenom smjeru.[137] Ključni element kampanje je da su korisnici vidjeli da njihovi prijatelji glasali, što uvelike povećava šansu da će i oni sami glasati.[138] Velika je razlika između jumbo plakata koji poziva na glasanje od toga kada nečiji Facebook prijatelj napiše da je glasao. Ovo posljednje, *de facto*, vrši puno veći pritisak na druge pojedince da učine isto, tj. da glasaju. Ključno za shvatiti je da ničiji prijatelj nije nikoga nagovorio na glasanje *per se,* već da je Facebookov algoritam odredio koji korisnici će vidjeti nečije sudjelovanje u navedenoj kampanji. Ovakav sustav je održiv jer većina ljudi ne sumnja u *Facebook*. Prema istraživanju provedenom na Sveučilištu Illinois 2013. godine čak 63% ispitanika smatra da Facebookovi algoritmi nemaju nikakvog utjecaja na to što oni vide na svojim Facebook profilima. Facebooku to otvara prostor za utjecanje na izborima.[139] Dugo istraživanje provedeno je o emocijama korisnika Facebook mreže. Otkriveno je da ljudima kojima je Facebookov algoritam namjerno dao više pozitivnih novosti i sami počinju stavljati više pozitivnih novosti na Facebook i obrnuto, da oni koji su dobili više negativnih novosti i sami postanu više negativni.[140] Dakle, Facebook može manipulirati ne samo glasanje, nego i osjećaje milijuna ljudi. Slične manipulacije mogu raditi i druge kompanije koji imaju iznimno velik broj korisnika, poput Googlea.[141] Sve to ima masovni potencijal, milijarde korisnika, velik prostor za manipulaciju pri čemu algoritmi nisu javno dostupni.[142]

Iako velike količine podataka za imatelje imaju veliki potencijal, postoje i drugi, otežavajući faktori vezani uz manipulaciju velikog broja podataka. Kod izgradnje bilo koje vrste baze podataka potrebno je paziti na dvije vrste problema – tehnički i politički.[143] Kod izgradnje bilo koje vrste baze podataka treba se voditi računa o npr. načinu prikupljanja podataka, sigurnosnim kopijama, sigurnosti, zaštiti osobnih podataka, hardveru, softveru, autorskim pravima, kvaliteti podataka itd. Baze podataka mogu biti uspješne samo ako imaju dugoročni plan - ako se zna što se treba raditi, kako treba raditi, tko što radi i sl. Drugo pitanje - političko, ne odnosi se na politiku u užem smislu nego na međusobne odnose onih

[137] O'Neil, *Weapons of math destruction*, n. dj., str. 181.
[138] O'Neil, *Weapons of math destruction*, n. dj., str. 182.
[139] O'Neil, *Weapons of math destruction*, n. dj., str. 183.
[140] O'Neil, *Weapons of math destruction*, n. dj., str. 183-184.
[141] O'Neil, *Weapons of math destruction*, n. dj., str. 184.
[142] O'Neil, *Weapons of math destruction*, n. dj., str. 185.
[143] Kitchin, *The Data Revolution*, n. dj., str. 38.

koji stvaraju podatke i onih koji ih koriste. Ti se odnosi moraju urediti, moraju biti unutar zakonskih okvira i moraju biti koordinirani.[144]

Veliki količine podataka kompleksni su za analizu. Poslovni subjekti imaju problema s obradom vlastitih internih strukturiranih podataka, a velike količine podataka, koje su eksterne i nestrukturirane predstavljaju veliki izazov.[145] Više je razloga zašto se kompanije ne nose uspješno s obradom svojih podataka. Jedan dio problema je da rukovodstvo vanjske nestrukturirane podatke ne smatra dovoljno vrijednima da bi se na njima radilo. Drugi problem je nedostatak *know-how*-a i nedostatak kadrova.[146] Iako je jedna od ključnih karakteristika velikih količina podataka upravo njihova količina, nabava podataka često je jeftina ili čak besplatna, dok je korištenje hardvera i softvera relativno jeftino. Najskuplji element velikih količina podataka su stručnjaci, analitičari velikih količina podataka.[147] Oni su ti koji rade sve ono najvažnije s velikim količinama podataka. Pronalaze podatke, strukturiraju podatke, analiziraju ih, interpretiraju rezultate i savjetuju donositelje odluka.[148] Analiza velikih količina podataka nije konvencionalna analiza. Podataka je iznimno mnogo, toliko da ih ljudi sa svojim ograničenim kognitivnim sposobnostima nisu u stanju obuhvatiti i analizirati. Analize se obavljaju posredno.[149] Koliko je analiza velikih količina podataka kompleksna ilustrira činjenica da se na globalnoj razini svega 0,5% podataka analizira. Nisu naravno svi podaci korisni - otprilike četvrtina ima određeni potencijal.[150] Prema anketi koju je napravio *Harvard Business Review* 2013. godine samo 3,5% ispitanika reklo da njihove tvrtke uspješno koriste velike količine podataka.[151] Osim što su vrlo kompleksni za implementaciju, treba znati da velike količine podataka ipak imaju samo ograničen utjecaj na cjelokupno poslovanje. Velike količine podataka povećavaju mogućnost uspjeha, ali nisu neophodni. Mnogo tvrtki je uspješno bez da se oslanjaju na velike količine podataka.[152] Osim toga, treba znati da velike količine podataka nisu primjerene za sve probleme, nego samo za neke. Npr. nije moguće saznati kolika je prosječna dob zaposlenika neke institucije. Za takve se stvari koriste druge metodologije. Mnogo toga se može napraviti s internim podacima koje

[144] Kitchin, *The Data Revolution*, n. dj., str. 40.
[145] Simon, *Too Big to Ignore*, n. dj., str. 41.
[146] Simon, *Too Big to Ignore*, n. dj., str. 45.
[147] Davenport, *Big Data at Work*, n. dj., str. 14.
[148] Davenport, *Big Data at Work*, n. dj., str. 110.
[149] Manyika et. al. *Big data: The next frontier for innovation, competition and productivity.* McKinsey Global Institute, 2011. URL: https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_full_report.ashx, str. 18.
[150] Davenport, *Big Data at Work*, n. dj., str. 2.
[151] Davenport, *Big Data at Work*, n. dj., str. 6.
[152] Simon, *Too Big to Ignore*, n. dj., str. 55.

tvrtka već ima. Štoviše, mnogo toga se može barem otprilike procijeniti ili predvidjeti. Na primjer kolika će biti prodaja sljedeći kvartal ako su tada božićni i novogodišnji blagdani.[153]

[153] Simon, *Too Big to Ignore*, n. dj., str. 57.

11. Tehnike prikupljanja i obrade velikih količina podataka

Prikupljanje, analiziranje i upotreba velikih količina podataka iznimno su heterogene aktivnosti. Kreativnost je karakteristika koja je poželjna kod ljudi koji se bave takvim poslovima. Načini obrade velikih količina podataka mogu biti vrlo različiti, inovativni i međusobno kombinirani. Ovdje se navode samo neke, stvarni broj metoda obrade velikog broja podataka nemoguće je utvrditi. Razvoj tehnologija, koje posredno utječu na podatke, njihovo analiziranje i prikupljanje, suviše je brz da bi konačan popis bio moguć. Treba napomenuti da su neke od navedenih tehnika prikupljanja i metode analiziranja postojale i prije pojave koncepata velikih količina podataka.

Jedna od metoda je "vrijabilnost". Ona služi za otkrivanje veza između dviju varijabli. Npr. u dućanima se pokušava otkriti koji proizvodi se kupuju zajedno. Ako netko kupi čokoladu, da li će onda kupiti i više drugih slatkiša i slično. Druga je "analiza grupacija" (klastera). Radi se o analizi određene heterogene skupine (npr. podataka, kupaca), koje se razvrstava u više skupina tako da svaka skupina ima neke zajedničke karakteristike. Sljedeći je "nabava iz mnoštva" (engl. *crowdsourcing*), a podrazumijeva da veća skupina ljudi preko javnog poziva odradi neki posao za koji bi pojedincu ili manjoj skupini zaposlenika trebalo značajno više vremena. „Integracija podataka" podrazumijeva kombiniranje podataka iz više različitih izvora u jedan korpus podataka koji se onda analizira. "Rudarenje podataka", odnosno ispravnije „dubinska analiza podataka" je dosta interdisciplinarno područje. Cilj joj je otkriti uzorke u velikim bazama podataka.[154] Podaci se mogu prikupljati i pomoću raznih senzora. Praćenje leta aviona pomoću senzora može pomoći prilikom sakupljanja važnih podataka o letu, što, nakon analize, može dovesti do implementiranja mjera koji će buduće letove učiniti ekonomičnijima i bržima.[155]

Neke od tehnika analiziranja velikih količina podataka su "genetski algoritmi" koji su inspirirani prirodnim odabirom u biološkom smislu. Nadahnuti darvinizmom više rješenja se stavlja u obzir i onda se sva rješenja za neki problem kombiniraju, mutiraju, eliminiraju, nadopunjuju i slično, da bi se na kraju dobio najkvalitetniji rezultat. "Strojno učenje" ima za cilj razviti algoritme ili program u koji se ubacuje velik broj podataka koji se obrađuju. Računalo tijekom obrade "uči" i razvija nove metode obrade tih podataka. Postoje i

[154]Manyika et. al. *Big data: The next frontier for innovation, competition and productivity.* McKinsey Global Institute, 2011. URL: https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_full_report.ashx, str. 28.
[155] Simon, *Too Big to Ignore*, n. dj., str. 91.

"neuronske mreže" koje su inspirirane radom ljudskih stanica, neurona u ljudskome mozgu, tj. međusobno povezanih stanica. Tehnika je korisna za pronalaženje nelinearnih uzoraka. U "analizi mreža" cilj je utvrditi sve veze u jednoj mreži podataka. Cilj je utvrditi koje su točno veze među entitetima, kako putuju podaci/informacije, kuda te koliko brzo.[156] Nadalje postoji i "prepoznavanje uzoraka" u kojem se pokušavaju prepoznati razni uzorci među podacima te razlozi zašto se oni pojavljuju. "Prostorna analiza" služi za analiziranje fizičkih karakteristika neke pojave u kontekstu velikih količina podataka. To su primjerice geografija, topografija i slično.[157] Pomoću "simulacija" nastoji se simulirati određena situacija za koju se očekuje da bi se mogla (ponovo) pojaviti. "Analiza u vremenskim intervalima" podrazumijeva praćenje određenih elemenata velikih količina podataka kroz dulji period s ciljem uočavanja trendova. "Vizualizacija" služi samo za vizualni prikaz velikih količina podataka s ciljem lakšeg uočavanja ključnih elemenata određenog problema. Pomoću vizualizacije ljudi, čije kognitivne sposobnosti nisu u stanju obuhvatiti velike količine podataka, mogu lakše dobiti "pregled" velikih količina podataka i donijeti određene zaključke.[158] "A/B testiranje" je postupak kojim se u kontroliranim uvjetima daju dva inputa, A i B, te se promatra kako se reagira na svaki od njih, tj. koje su razlike.[159] "Procesiranje prirodnog jezika" odnosi se na procesiranje ljudskog jezika i njegovog prebacivanja u digitalni oblik koji je čitljiv ljudima i računalima.[160] "Analiza teksta" analizira veće količine strukturiranih ili nestrukturiranih tekstova (poruke e-pošte, recenzije i slično). Konačno "analiza sentimenta" je lingvistička analiza teksta koja analizira generalni ton teksta: dobar, loš, emotivan, neemotivan i slično.[161]

[156] Manyika et. al. *Big data: The next frontier for innovation, competition and productivity.* McKinsey Global Institute, 2011. URL: https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_full_report.ashx, 29.
[157] Manyika et. al. *Big data: The next frontier for innovation, competition and productivity.* McKinsey Global Institute, 2011. URL: https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_full_report.ashx, 30.
[158] Manyika et. al. *Big data: The next frontier for innovation, competition and productivity.* McKinsey Global Institute, 2011. URL: https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_full_report.ashx, 31.
[159] Simon, *Too Big to Ignore*, n. dj., str. 82.
[160] Simon, *Too Big to Ignore*, n. dj., str. 94.
[161] Simon, *Too Big to Ignore*, n. dj., str. 97.

12. Pozitivni utjecaj velikih količina podataka u poslovanju

Velike količine podataka nisu korisne samo za pravne osobe koje ih koriste, već i za krajnje korisnike proizvoda i usluga tih poslovnih subjekata. Treba napomenuti da se interesi i pozitivne strane velikih podataka za krajnje korisnike i poslovne subjekte u određenoj mjeri preklapaju. Ono što je korisno za korisnike u pravilu je korisno i za poslovne subjekte. Jedan od projekata za koji se lako složiti da je vrlo koristan za korisnike (tj. u ovom slučaju opću populaciju) je UN-ov projekt *HunchWorks*. Radi se o projektu koji prati velike količine podataka s interneta s fokusom na društvene mreže. Cilj mu je otkriti povećanje učestalosti pojave termina koji bi mogli alarmirati na opasnost poput suše ili poplave. Ukoliko se npr. u određenom dijelu Afrike počinje intenzivnije spominjati suša, UN to shvaća kao signal na koji trebaju obratiti pažnju. To naravno ne znači da će se elementarna nepogodna nužno dogoditi, već da treba obratiti pažnju. To se naziva „digitalnim dimnim signalima (engl. *digital smoke signals*).[162] Na razini poslovnih subjekata u privatnom sektoru, velike količine podataka pomažu u nekoliko ključnih segmenata. Jedan od njih je smanjenje troškova poslovanja.[163] Drugi vrlo čest razlog iskorištavanja velikih količina podataka je smanjenje vremena potrebnog za određenu radnju. Neke stvari koje su se ranije sakupljale, računale, obrađivale satima ili danima, a sada se mogu obaviti u par sati ili kraće.[164] Velike količine podataka mogu pomoći u uvođenju novih proizvoda i usluga. Analiza velikog broja podataka može ukazati na želje i potrebe tržišta za određenim novim proizvodima i uslugama.[165] Korisnici proizvoda i usluga kompanija koje koriste velike količine podataka imaju direktne i indirektne koristi od implementacije velikih količina podataka u poslovanje tih kompanija. Proizvodi i usluge koje će im se nuditi biti će kvalitetniji, jeftiniji, dostupniji, moguće i personaliziraniji. Iako su pozitivne strane velikih količina podataka za krajnje korisnike kvantitativno malobrojni, njihov je utjecaj na globalno gospodarstvo i društvo dubok i dugotrajan. Omogućuje daljnji razvoj globalne ekonomije na temeljima jeftinijih i kvalitetnijih proizvoda i usluga.

[162] Davenport, *Big Data at Work*, n. dj., str. 17.
[163] Davenport, *Big Data at Work*, n. dj., str. 60.
[164] Davenport, *Big Data at Work*, n. dj., str. 63.
[165] Davenport, *Big Data at Work*, n. dj., str. 65.

13. Manjkavosti i zloupotreba velikih količina podataka

Podaci o pojedincu sve se češće prikupljaju.[166] Tu činjenicu ilustrira, između ostalog, i *Dutch Data Protection Authorty* sa svojom objavom da se prosječni Nizozemac nalazi u čak 250 do 500 baza podataka, utječu na to da je pitanje zaštite osobnih podataka sve važnije u javnom životu.[167] Kao jedna od posljedica takvih tendencija uvedeno je sedam globalnih principa privatnost (engl. *Fair Information Practice Principles*).[168] To su: "obavijest", korisnik je upoznat s činjenicom da se njegovi podaci prikupljaju i za što će biti iskorišteni; "izbor", odnosi se na pravo da pojedinac ima pravo izbora kada će dopustiti, a kada neće prikupljanje vlastitih podataka; "pristanak", pojedinac mora eksplicitno pristati na prikupljanje svojih podataka; "sigurnost", podaci koji su prikupljeni moraju biti sigurno skladišteni; "integritet" se odnosi na to da podaci moraju biti potpuni, a osoba čiji su to bili podaci mora imati pravo pristupa tim podacima te posljednje "odgovornost" se odnosi da na to da je prikupljač podataka obavezan da gore navedene principe provede u praksi.[169]

Praksa je naravno drugačija od proklamiranih principa. O kolikom se nesrazmjeru radi ilustrira Adam Tanner. On je svojedobno 1980-ih bio u posjeti Njemačkoj Demokratskoj Republici (tzv. Istočna Njemačka) – s obzirom da je bio stranac, Stasi, tajna policija, pratila ga je tijekom cijelog njegovog boravka. Tek nakon pada komunizma i otvaranja arhiva Tanner je dobio uvid u vlastiti dosje. Iznenadio se kada je vidio da Stasi zapravo nije saznao ništa osobito intrigantno ili barem donekle zanimljivo. Vratimo li se u sadašnjost možemo vidjeti da određene kompanije mogu saznati puno više o nama. Gdje živimo, što radimo, s kim se družimo, tko su nam susjedi, koji su nam hobiji, telefonski brojevi i slično.[170] Treba naglasiti da Tanner donekle pretjeruje. Društvene mreže mogu prikupiti više podataka, ali one nisu tajna policija. Ipak, navedeni primjer dobro ilustrira potencijal novih tehnologija i uloge društvenih mreža u suvremenom društvu. Jedna od razlika društvenih mreža i suvremenih sigurnosnih agencija je u tome što je rad društvenih mreža ipak slabije reguliran od rada državnih agencija.[171] Koliko se ta činjenica očituje u praksi ilustrira analiza njemačke potrošačke organizacije *Stiftung Warentest* iz 2010. godine na deset društvenih mreža. Samo su dvije imale manje greške vezane uz za zaštitu osobnih podataka. Četiri su imale "očite

[166] Kitchin, *The Data Revolution*, n. dj., str. 169-170.
[167] Kitchin, *The Data Revolution*, n. dj., str. 167.
[168] Kitchin, *The Data Revolution*, n. dj., str. 170.
[169] Kitchin, The *Data Revolution*, n. dj., str. 171.
[170] Tanner, Adam, *What Stays In Vegas: The World Of Personal Data, Lifeblood Of Big Business And The End Of Privacy As We Know It*. New York: PublicAffairs, 2014, str. XII.
[171] Tanner, *What Stays in Vegas*, n. dj., str. XIV.

mane" (u originalu "clear flaws"), a ostale četiri vrlo ozbiljne mane. *LinkedIn*, *Facebook* i *MySpace* bili su u skupini s vrlo ozbiljnim manama. Postoje načini da se korisnik zaštiti, može ograničiti tko gleda njegove objave na društvenim mrežama, ali čak i takve vrste mjera ograničenog su uspjeha.[172] Jedan od ključnih problema vezanih uz privatnost na društvenim mrežama je što pravna zaštita koja postoji vrlo često nije dovoljno razumljiva većini korisnika. Kao što kaže Daniel Solove "(1) ljudi ne čitaju pravila privatnosti; (2) ako ih čitaju, ne razumiju ih; (3) ako ih čitaju i razumiju onda im nedostaje dovoljno drugih informacija da bi mogli donijeti ispravnu odluku; i (4) ako ih čitaju, razumiju i imaju dovoljno drugih informacija onda na njihov izbor utječu razni drugi faktori koji negativno utječu na ispravno donošenje odluke".[173] Sve ovo naravno ne znači da je svako prikupljanje podataka nužno loše. Profesor Melvin Kranzberger s Case Western Reserve University je rekao da "tehnologija nije niti dobra niti loša; niti je neutralna".[174] Osim toga, *Facebook*, *Google* i *LinkedIn* nikoga ne prisiljavaju da koriste njihove usluge odnosno da im daju svoje podatke.[175]

Počeci razvoja korištenja analize podataka i ponuda personaliziranih proizvoda i usluga naizgled nisu problematični. Poslovni subjekti izrađivali su jednostavne klasifikacije stanovništva vezane uz prodaju svojih proizvoda. Mogli su ciljati na određenu skupinu ljudi tako da bi se fokusirali na određenu geografsku lokaciju, npr. imućniji dio grada. Naizgled se radi o *win-win* situaciji, (potencijalni) kupci dobivaju personalizirane oglase, a prodavači bolju prodaju za manji trošak.[176] Situacija se tijekom desetljeća temeljito promijenila. Velike količine podataka i personalizirani oglasi više ne igraju sporednu ulogu u poslovanju, već su često jedan od ključnih elemenata poslovnog uspjeha. Inzistiranje nekih poslovnih ljudi na velikim količinama podataka stoga ne začuđuje. Cathy O'Neil izjavila je da je dok je radila u *start up*-u Intent Media kod njih došao jedan poslovni čovjek ("capitalist" u originalu) raspravljati o budućim ulaganjima. Govorio je o velikim količinama podataka i personaliziranim reklamama u iznimno pozitivnom svjetlu.[177] Smatrao je da je jedini razlog zašto su ljudi nezadovoljni s personaliziranim oglasima taj što ti oglasi nisu dovoljno personalizirani. Smatrao je da će s daljnjim razvojem velikih količina podataka personalizirani oglasi postati još kvalitetniji (tj. još personaliziraniji) što će u konačnici

[172] Kitchin, *The Data Revolution*, n. dj., str. 55.
[173] Kitchin, *The Data Revolution*, n. dj., str. 175.
[174] Simon, *Too Big to Ignore*, n. dj., str. 184.
[175] Simon, *Too Big to Ignore*, n. dj., str. 185.
[176] Kitchin, *The Data Revolution*, n. dj., str. 176.
[177] O'Neil, *Weapons Of Math Destruction*, n. dj., str. 68.

promijeniti stav javnosti prema takvim oglasima.[178] Kao što je već spomenuto u ovom radu, velike količine podataka nedvojbeno imaju pozitivan učinak i na poslovanje i na krajnje potrošače, međutim, postoji i druga strana ovog fenomena.

Velike količine podataka teoretski mogu biti rasistički i seksistički postavljeni, ali čak i kad to nije slučaj, kada rukovodstvo želi stvoriti efikasan sustav on može bit loš. Primjera radi, recimo da osoba koja je vrlo marljiva, ali nema dobro financijsko zaleđe, nema visoku naobrazbu, živi u lošoj gradskoj četvrti želi kredit. Svi gore navedeni faktori mogu biti uzeti u obzir od banke i temeljem toga može dobiti lošiji kredit.[179] Kreditna sposobnost (engl. *credit score*) često se analizira od strane poslodavaca prilikom zapošljavanja ili napredovanja. Međutim, ta praksa je *de facto* diskriminatorna te je stoga u nekim saveznim državama SAD-a zabranjena. Takvim mjerama ljude se zatvara u krug siromaštva. Ako osoba ima loš rejting, manje su šanse da dobije posao ili napredovanje. Određene skupine u SAD-u djelomično su deprivilegirane i zbog analiziranja kreditne sposobnosti. Utvrdilo se da su skupine stanovništva koje su rođene siromašne još siromašnije u odnosu na ostatak stanovništva do svoje 60. godine. To naravno ne znači da se to radi namjerno, ali u konačnici se situacija tako razvije.[180] Događaju se slučajevi da su ljudi odbijani od strane stanodavca prilikom želje da unajme stan zbog dosjea. U konkretnom slučaju koji O'Neil spominje da je riječ o ženi koja je imala dosje zbog sukoba s mužem, ali nikada nije bila osuđena. Uspjela je izbrisati podatke o uhićenju kod državnih institucija, ali problem je što je jedna privatna kompanija u nekom trenutku spremila taj dosje i koristi ga. Stanodavci provjeravaju te podatke i zbog toga su ju odbili. Ne samo da joj je dosje izbrisan nego ona nije kriminalac, samo je uhićena, ali algoritmu je to dovoljno da ju odbije. Da je igrom slučaja ljudski faktor igrao veću ulogu prilikom dodjeljivanja stana, velike su šanse da bi joj stan bio iznajmljen. Osoba je na kraju tužila kompaniju koja je imala njen dosje te je on maknut iz baze podataka, ali pitanje je da li je njen dosje prosljeđivan ili prodan nekim drugim privatnim kompanijama.[181] Ključni je problem kada se donošenje neke odluke temelji samo na analizi velikih količina podataka. Ljudski faktor je nezamjenjiv u donošenju odluka. Osim toga, velike količine podataka nisu imuni na greške. The Federal Trade Commision objavio je da je 2013. godine oko 5% ljudi (10 milijuna ljudi) imalo neku vrste greške kod procjene kredite sposobnosti, radi čega su vjerojatno dobili manje povoljan kredit.[182]

[178] O'Neil, *Weapons of math destruction*, n. dj., str. 69.
[179] O'Neil, *Weapons of math destruction*, n. dj., str. 146.
[180] O'Neil, *Weapons of math destruction*, n. dj., str. 149.
[181] O'Neil, *Weapons of math destruction*, n. dj., str. 151-152.
[182] O'Neil, *Weapons of math destruction*, n. dj., str. 150.

Osim nenamjernih grešaka vezanih uz velike količine podataka, postoji i mogućnost zlonamjerne uporabe velikih količina podataka. Slučaj sa Sveučilištem Corinthian jedan je od takvih primjera. Njihovi su studijski programi prezentirani potencijalnim studentima kvalitetnijima nego što jesu, a njihova cijena bila je previsoka. Jedan online program kojeg su izvodili naplaćivao se gotovo 70.000 USD, dok su neki drugi fakulteti koji su program izvodili konvencionalno takav program naplaćivali manje od 10.000 USD. U svojim kampanjama za regrutaciju, u kojima su se djelomično oslanjali na velike količine podataka, otkriveno je da su ciljani ljudi koji su "izolirani" bili oni "s niskim samopouzdanjem" i slično. Još je gori slučaj s Vatterott Sveučilištem.[183] Oni su svojim regruterima rekli da ciljaju na "ljude na socijalnoj pomoći", "trudnice", "nedavno rastavljene", "s niskim samopouzdanjem", "s niskim primanjima", "koji su nedavno iskusili smrt u obitelji", "koji su bili fizički ili psihički maltretirani", "na rehabilitaciji" i slično.[184] Osobe s takvim karakteristikama i iskustvima posebno su ranjivi na manipulacije. Njihove slabosti nazivaju se "točke boli" (engl. *pain points*).[185] Točke boli ljudi mogu se otkriti putem velikih količina podataka, što je dakle iznimno koristan alat regruterima takvih sveučilišta.[186]

Velike količine podataka mogu biti opasani i iz još jednog razloga. Vrijedni podaci na jednom mjestu meta su *cyber* kriminalaca. Godine 2009. dogodila se najveća uspješna krađa osobnih podataka u SAD-u. Heartland Payment Systems je napadnut. Rečeno je da su hakeri ukrali informacije od oko 130 milijuna kreditnih i debitnih kartica. Takva vrsta napada učestala je i ima veliki potencijal.[187] Razmjeri takvih krađa, kada se dogode, u toliko su veliki da kradljivci ne mogu osobno i na direktan način iskoristiti podatke pa ih zato prodaju.[188] Ukradeni podaci mogu se pokušati prodati na raznim internetskim stranicama i forumima.[189] Čak postoje forumi posvećeni krađi i prodaji osobnih podataka u kojima se otvoreno raspravlja o takvim temama.[190] Nije sasvim jasno kako se točno i za koju cijenu prodaju ukradeni podaci. To vjerojatno ovisi o trenutnom stanju na tržištu i načinu pregovaranja. U monografiji *Data Thieves in Action* analizira se 13 foruma na kojima se obavljaju kupovine i

[183] O'Neil, *Weapons of math destruction*, n. dj., str. 71.
[184] O'Neil, *Weapons of math destruction*, n. dj., str. 72.
[185] O'Neil, *Weapons of math destruction*, n. dj., str. 73.
[186] O'Neil, *Weapons of math destruction*, n. dj., str. 74.
[187] Holt, Thomas, Olga Smirnova and Yi-Ting Chua, *Data Thieves In Action*. New York: Palgrave Macmillan, 2016, str. 2.
[188] Holt, Smirnova i Chua, *Data Thieves in Action*, n. dj., str. 3.
[189] Holt, Smirnova i Chua, *Data Thieves in Action*, n. dj., str. 5.
[190] Holt, Smirnova i Chua, *Data Thieves in Action*, n. dj., str. 9.

prodaje ukradenih podataka i prezentiraju se rezultati. Niti jedna stranica više nije aktivna.[191] Primjer prodaje na forumu je:

"*Vendmaster: Selling CCV: German/ Italy/ USA/ UK/ Netherlands/*
Greetings,
Selling CCV: Germany, Italy, USA, UK, Netherlands
German cc [credit card]: 15USD
Italy CC: 15USD
USA CC: 6USD
UK Without DOB [Cardholder's Date Of Birth]: 14USD
UK with DOB: 30USD
Netherlands: 15USD
Note: We sell all unused/Untouch cc to our Customers! to ensure Quality we sell each cc to each person! And we fully Replace cc in 48 Hours of Delivery TIME
We Accept CC orders only by WU at the moment!
Min order: 350usd
for Orders please e-mail me or add me in your messenger either you can pm me".[192]

Treba napomenuti da se ovdje radi o traženoj cijeni, dok konačna cijena nije poznata..[193] Zanimljivo je da se u analiziranim oglasima navodi otprilike 25% slučajeva da se uplate prihvaćaju i preko *Western Uniona* i *MoneyGrama*. Kod njih se kod slanja treba navesti ime primatelja, a primatelj kad prihvaća novac mora pokazat identifikacijsku iskaznicu.[194] Ponekad postoje čak recenzije prodavača ukradenih podataka.[195] Negativne recenzije korisnika radi prevara relativno su česte.[196] Najviše ukradenih podataka je iz Europe i iz Sjeverne Amerike - ondje ljudi najčešće koriste kartice.[197] Štete od *cyber* kriminala na globalnoj razini vrlo variraju. Procjene su od 226 do 575 milijardi USD godišnje. U SAD-u je 2012. godine bilo štete od *cyber* kriminala u vrijednosti 20,7 milijardi USD. Prave razmjere štete nikada neće biti poznate jer kompanije koje su oštećene rijetko daju detalje o tome

[191] Holt, Smirnova i Chua, *Data Thieves in Action*, n. dj., str. 19.
[192] Holt, Smirnova i Chua, *Data Thieves in Action*, n. dj., str. 21-22.
[193] Holt, Smirnova i Chua, *Data Thieves in Action*, n. dj., str. 25-26.
[194] Holt, Smirnova i Chua, *Data Thieves in Action*, n. dj., str. 30.
[195] Holt, Smirnova i Chua, *Data Thieves in Action*, n. dj., str. 33-34.
[196] Holt, Smirnova i Chua, *Data Thieves in Action*, n. dj., str. 36.
[197] Holt, Smirnova i Chua, *Data Thieves in Action*, n. dj., str. 38.

koliko su ih *cyber* kriminalci točno oštetili i na koji način.[198] Osim toga, štete je teško procijeniti jer postoji direktna i indirektna šteta.[199] S obzirom na druge vrste ilegalnih tržišta, *cyber* prijevare među samim kriminalcima učestalije su jer ne postoje mehanizmi za rješavanje prijevara (osim postavljanja negativnog komentara na forumu). Na dva foruma bio je koncentriran velik broj negativnih komentara, što navodi na zaključak kako se radi o forumu s velikim brojem loših prodavača.[200] *Cyber* kriminal nešto je novo i ne može ga se riješiti konvencionalnim sredstvima. Jedna od predlaganih strategija je da se naprave profili na tim forumima i počne masovno ubacivati ponude koje nisu valjane. Tako će se smanjit povjerenje u te forume što je važno jer takvi forumi se uvelike oslanjaju na povjerenje. Postoje i organiziraniji forumi u kojima bi administratori sigurno izbacili takve lažne ponude. Osim toga, treba više kontrolirati *WebMoney* i *Liberty Reserve* jer preko njih ponekad idu transakcije.[201]

Posebno ranjiva skupina u kontekstu osobnih podataka su stariji građani. Oni sve češće koriste internet i društvene mreže, a s obzirom da su osobe starije životne dobi brojnije u zapadnim zemljama u odnosu na druge dobne skupine, oni su relativno i apsolutno brojniji.[202] Stariji su dosta zabrinuti za svoju privatnost online. *Burst Media* je 2009. godine napravio istraživanje i po rezultatima u dobnoj skupini 18-24 godine 67% ispitanika je zabrinuto za svoju privatnost, dok u dobnoj skupini od 55 godina pa na više gotovo 86% ispitanika je zabrinuto. Iako su natprosječno zabrinuti za svoju privatnost, starije osobe nisu upoznate s metodama i tehnikama zaštite privatnosti te stoga predstavljaju ranjivu skupinu.[203]

[198] Holt, Smirnova i Chua, *Data Thieves in Action*, n. dj., str. 46.
[199] Holt, Smirnova i Chua, *Data Thieves in Action*, n. dj., str. 47.
[200] Holt, Smirnova i Chua, *Data Thieves in Action*, n. dj., str. 135.
[201] Holt, Smirnova i Chua, *Data Thieves in Action*, n. dj., str. 142.
[202] Maaß, Wiebke, "The Elderly and the Internet: How Senior Citizens Deal with Online Privacy", u *Privacy Online*, str. 235.
[203] Maaß, "The Elderly and the Internet", n. dj., str. 240.

14. Paradoks privatnosti

Konvencionalna komunikacija podrazumijeva da se izmijenjene informacije ne dijele s trećim osobama, izuzev slučajeva kasnijeg prepričavanja na koje su direktno ili indirektno pristale obje strane u komunikaciji.[204] Međutim, kada govorimo o društvenim mrežama, situacija je drugačija. Sama narav takvih mreža potiče ljude na širenje informacija.[205] S obzirom da velik broj ljudi svakodnevno komunicira putem društvenih mreža, velik udio komunikacije obilježen je tim fenomenom.[206] Ljudi otkrivaju razne informacije o sebi koje mogu biti kvalitativne i kvantitativne. U jednoj otkrivaju detalje o sebi, vrlo privatne stvari, želje i fantazije, a u drugo otkrivaju količinski puno, ali bez detalja. Širenje informacija o sebi ključno je za stvaranje i održavanje društvenih veza. Po toriji smanjenja nesigurnosti (engl. *ucertainity reduction theory*, URT) što više znamo o nekome to su veće šanse da ćemo se s njime socijalizirati, dok neizvjesnost utječe da osobu manje volimo.[207] Koliko će neka osoba otkriti o sebi ovisi o tome kakav odnos želi uspostaviti s nama. Osim toga, kulturni kontekst također ima utjecaja. Postoje kulture koje više glorificiraju zadovoljenje osnovnih životnih potreba od drugih. To je u ovom kontekstu vezano uz teoriju CPM jer u društvima u kojima se takve stvari glorificiraju ljudi češće objavljuju više privatnih stvari o sebi jer njihova društva cijene glorifikaciju tih stvari pa onda imaju više benefita za objavljivanje. Osim kulture, po teoriji CPM postoji razlika između muškaraca i žena, odnosno odnosu dvaju spola prema pitanju privatnosti. Žene se općenito više brinu o svojoj privatnosti i privatnim podacima, tako da su manje šanse da će objaviti nešto privatno.[208] Dijeljenje informacija o sebi na društvenim mrežama vežemo uz određene rizike, od kojih je krađa podataka samo jedna od njih. Digitalna prenapučenost (eng, *digital crowding*) jedna je od potencijalnih opasnosti. Odnosi se na pojavu kada osoba ima previše kontakata na društvenim mrežama te joj stoga postane preteško kontrolirati privatnost vlastitih informacija na društvenoj mreži. Ono što je bilo namijenjeno uskom krugu prijatelja može postati dostupno širokom krugu ljudi.[209] Osoba kojoj se to dogodi može osjećati da joj je povrijeđena privatnost. Osim toga, takvi korisnici zasipani su sadržajem drugih ljudi koji inače igraju perifernu ulogu u

[204] Boyd, Danah, *It's Complicated: The Social Lives of Networked Teens*. New Haven: Yale University Press, 2014, str. 61.
[205] Boyd, *It's Complicated*, n. dj., str. 62.
[206] Boyd, *It's Complicated*, n. dj., str. 8.
[207] Joinson, Adam, David J. Houghton, Asimina Vasalou i Ben L. Marder, "Digital Crowding: Privacy, Self-Disclosure, and Technology", u *Privacy Online*, str. 36.
[208] Joinson, Houghton, Vasalou i Marder, "Digital Crowding", n. dj., str. 30.
[209] Joinson, Houghton, Vasalou i Marder, "Digital Crowding", n. dj., str. 36.

njihovom društvenom životu pa stoga društvene mreže mogu postati zatrpane nerelevantnim sadržajem.[210]

S obzirom da su opasnosti vezane uz privatnost na društvenim mrežama dosta visoke, postavlja se pitanje zašto su takve mreže toliko popularne. Korisnici društvenih mreža u istraživanjima redovito izjavljuju da im je stalo do zaštite njihovih osobnih podataka na webu, ali u isto vrijeme ne ograničavaju značajno dostupnost svojih osobnih podataka na svojim profilima na društvenim mrežama. Ta diskrepancija naziva se paradoks privatnosti (engl. *privacy paradox*).[211] Neki paradoks privatnosti povezuju s digitalnom/informacijskom nepismenošću. Iako je i to djelomično točno, pitanje dijeljenja privatnih informacija ima duge uzroke. Po teoriji CPM granice privatnosti konstantno se redefiniraju ovisno o okolnostima. Ukoliko netko smatra da dijeljenjem informacije ostvaruje određene benefite i da su ti benefiti veći od potencijalnih rizika, onda će ta informacija biti podijeljena. S obzirom da su društvene mreže stvorene s primarnom svrhom dijeljenja informacija, one za mnoge predstavljaju idealnu platformu za takve stvari.[212] Postoje tri vrste privatnosti - informacijska, društvena i psihološka privatnost. Korisnici društvenih mreža boje se za svoju informacijsku privatnost, ali smatraju da imaju kontrolu nad društvenom i psihološkom privatnosti te da od njih imaju i određene benefite. Informacijska privatnost odnosi se na to koliko su korisnici u mogućnosti kontrolirati koliko drugi šire informacije o njima. Društvena privatnost odnosi se na kontrolu interakcija s drugim ljudima, koliko ih ima, koliko su si bliski ili udaljeni. Psihološka privatnost odnosi se na mogućnost kontrole emocionalnih *inputa* i *outputa*.[213] Glavni razlog masovnog korištenja društvenih mreža leži u činjenici da ljudi smatraju da, iako im je ugrožena informacijska privatnost, imaju kontrolu i benefite od društvene odnosno psihološke privatnosti. Ti benefiti veći su od potencijalnih rizika, što je ključno za razumijevanje masovnog korištenja društvenih mreža.

[210] Joinson, Houghton, Vasalou i Marder, "Digital Crowding", n. dj., str. 41.
[211] Quinn, Kelly, "Why We Share: A Uses And Gratifications Approach To Privacy Regulation In Social Media Use", *Journal Of Broadcasting & Electronic Media* 60, no. 1 (2016), str. 63.
[212] Quinn, "Why we share", n. dj., str. 28.
[213] Trepte, Sabine i Leonard Reinecke, "The Social Web as a Shelter for Privacy and Authentic Living," Privacy Online, 2011, str. 61.

15. Zaključak

Dva su temeljna fenomena središnji dio ovoga rada. S jedne strane, prati se afirmacija individualnosti i privatnost, a s druge strane prikupljanje i korištenje osobnih podataka. Fenomene su definirali (ponekad) vrlo različiti faktori, ali oni su u istom trenutku nerazdvojivo isprepleteni. Prije prosvjetiteljstva pripadnost skupini, vjerskoj, regionalnoj ili nekoj trećoj, smatrano je temeljom nečijeg identiteta. Privatnost u takvim okolnostima nije mogla biti percipirana kao važna tema. Tek tijekom i nakon doba razuma individualnost i privatnost se počinju smatrati relevantnima. Iako su osobni podaci prikupljeni u tom periodu bili, do tada, najopsežniji korpus takvih podataka ikada prikupljen, te baze iz današnje perspektive imale su relativno male potencijale. Činjenica da su bile u analognom obliku te nepovezane ključni su za razumijevanje njihove ograničene uporabe. Koliko digitalni format te arhivska sređenost igraju ulogu u iskoristivosti podataka ilustrira činjenica da danas relativno širok krug ljudi, od znanstvene zajednice do genealoga amatera, koristi i analizira te podatke na način i u opsegu koji ne bi nikada bio moguć prije nekoliko stoljeća. Na samom početku novog tisućljeća već su jasno bile iskristalizirane nove tendencije u odnosu osobnih podataka i privatnosti odnosno u promjenama dostupnosti osobnih podataka. Nove tehnologije omogućile su prikupljanje, pohranjivanje i kompleksno analiziranje iznimno velikog broja podataka. Ključan preduvjet za to bila je "proizvodnja" većeg broja podataka. U tom kontekstu Web 2.0 predstavlja najveći pojedinačni doprinos suvremenom prikupljanju velikih količina podataka. Naime, on je omogućio korisnicima interneta da sami budu kreatori njegovog sadržaja, a ne samo konzumenti. Mogućnost je objeručke prihvaćena te je internet preplavljen novim sadržajima - osobnim podacima. Individualnost, osobnost i privatnost do tada su ne samo afirmirani, nego su postali postulatima zapadne civilizacije. Te dvije tendencije, sve veća potreba za prikupljanjem i analiziranjem osobnih podataka i percipiranje privatnosti ljudskim pravom, počele su dolaziti u sukob. Te su napetosti kompleksne i višedimenzionalne, tako da njihovo analiziranje i konačan sud uvelike ovisi o tome iz koje perspektive je problem sagledan. Nedvojbeno je da prikupljanje i analiziranje velikih količina podataka ima pozitivnih utjecaja na poslovne subjekte i njihove kupce te da je taj utjecaj vidljiv i na globalnoj razini. S druge strane, postoje aspekti velikih količina podataka koji predstavljaju određenu ugrozu privatnosti i javnoj sigurnosti. Takvi problemi variraju sve od "sivih zona" i moralno diskutabilnih praksi, pa sve do krađa podataka. Paradoks privatnosti predstavlja zanimljivu iznimku koja ilustrira da elementi koji donekle izlaze van okvira sveprisutnih globalnih tendencija postoje i da su faktor na kojeg treba

računati. S obzirom da je privatnost jedan od temelja društva u kojem živimo, a korištenje velikih podataka jedan od važnijih elemenata poslovanja, možemo sa sigurnošću tvrditi da će napetosti koje proizlaze iz njihovih neusklađenih ciljeva i interesa nastaviti postojati. Pravna uređenost, kulturni kontekst i slični faktori utjecati će na daljnji razvoj ovog problematičnog odnosa, ali konačnu i apsolutnu pobjedu vjerojatno nećemo vidjeti u narednim desetljećima. Obje strane imaju prečvrste temelje u politici, financijama, ljudskoj naravi i poslovnoj logici da bi mogli u potpunosti nestati.

16. Literatura

5 biggest GDPR fines so far [2020]. URL: https://dataprivacymanager.net/5-biggest-gdpr-fines-so-far-2020/, (1.4.2020.).

Allmer, Thomas. "Critical Internet Surveillance Studies and Economic Surveillance". U *Internet and Surveillance: The Challenges of Web 2.0 and Social Media*, ur. Christian Fuchs, Kees Boersma, Anders Albrechtslund i Marisol Sandoval, 124-143. New York: Routledge, 2012.

Boban Marija. "Pravni i kulturološki aspekt sigurnosti i zaštite osobnih podataka". Doktorska disertacija, Sveučilište u Zagrebu, 2012.

Boyd Danah. It's *Complicated: The Social Lives of Networked Teens*. New Haven: Yale University Press, 2014.

Braudel, Fernand. *Civilizacije Kroz Povijest*. Zagreb: Globus, 1990.

Britannica, "Genealogy". https://www.britannica.com/topic/genealogy (25.5.2020.).

Christian, Peter. *The Genealogist's Internet*. London: Bloomsbury Publishing Plc, 2012.

Cik, Nikola. "Crkvene matične knjige kao izvor za povijesna istraživanja, na primjeru župe Đurđevac*". Pro tempore: časopis studenata povijest* 10-11 (2016): 236-251.

Davenport, Thomas. *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities*. Boston: Harvard Business Review Press, 2014.

Debatin, Berhard. "Ethics, Privacy, and Self-Restraint in Social Networking", U *Privacy Online: Perspectives on Privacy and Self-Disclosure in the Social Web*, ur. Sabina Trepte i Leonard Reinecke, 47-60. New York: Springer, 2011.

Europe's tough new data-protection law. URL: https://www.economist.com/business/2018/04/05/europes-tough-new-data-protection-law (1.4.2020.).

FamilySearch. URL: https://www.familysearch.org/search/catalog/results?count=20&query=%2Bauthor_id%3A845152&page=1&topic0=Birth%2C+Marriage+and+Death&year0=1500, (10.4.2020.).

GDPR Fines Tracker & Statistics. URL: https://www.privacyaffairs.com/gdpr-fines/, (1.4.2020.).

Google Ngram Viewer: Personal data. URL: https://books.google.com/ngrams/graph?content=personal+data&year_start=1800&ye

ar_end=2008&corpus=15&smoothing=3&share=&direct_url=t1%3B%2Cpersonal%20data%3B%2Cc0 (25.5.2020.).

Hare, Stephanie, *These new rules were meant to protect our privacy. They don't work*. URL: https://www.theguardian.com/commentisfree/2019/nov/10/these-new-rules-were-meant-to-protect-our-privacy-they-dont-work (1.4.2020.).

Holt, Thomas J, Olga Smirnova, and Yi-Ting Chua. *Data Thieves In Action*. New York: Palgrave Macmillan, 2016.

Hrvatska enciklopedija, Genealogija. URL: https://www.enciklopedija.hr/natuknica.aspx?id=21578, (5.4.2020.).

Iveljić, Iskra. *Banska Hrvatska i Vojna Krajina od prosvijećenog apsolutizma do 1848. godine*. Zagreb: Leykam Internat., 2010.

Joinson N. Adam, David J. Houghton, Asimina Vasalou i Ben L. Marder, "Digital Crowding: Privacy, Self-Disclosure, and Technology".U *Privacy Online: Perspectives on Privacy and Self-Disclosure in the Social Web*, ur. Sabina Trepte i Leonard Reinecke, 33-46. New York: Springer, 2011.

Kanellos, Michael, *PCs: more than 1 billion served.* URL: https://www.cnet.com/news/pcs-more-than-1-billion-served/, (5.3.2020.).

Kitchin, Rob. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Los Angeles: SAGE Publications Ltd, 2014.

Kolanović, Josip. *Pregled arhivskih fondova i zbirki Republike Hrvatske*. Zagreb: Hrvatski državni arhiv, 2006.

Maaß, Wiebke. " The Elderly and the Internet: How Senior Citizens Deal with Online Privacy". U *Privacy Online: Perspectives on Privacy and Self-Disclosure in the Social Web*, ur. Sabina Trepte i Leonard Reinecke, 235-250. New York: Springer, 2011.

Manyika, James, Michael Chui, Brad Brown, Jacques Baghin, Richard Dobbs, Charles Roxburgh i Angela Hung Byers. *Big data: The next frontier for innovation, competition and productivity.* McKinsey Global Institute, 2011. URL: https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_full_report.ashx

Margulis, Stephen T. "Three Theories of Privacy: An Overview." U *Privacy Online: Perspectives on Privacy and Self-Disclosure in the Social Web*, ur. Sabina Trepte i Leonard Reinecke, 9-17. New York: Springer, 2011.

Mayer-Schönberger Viktor i Kenneth Cukier. *Big Data: a Revolution That Will Transform How We Live, Work and Think*. London: John Murray, 2017.

Moores law. URL: http://www.mooreslaw.org/, (20.4.2020.).

O'Neil, Cathy. *Weapons Of Math Destruction*. London: Penguin books, 2016.

Online Etimology Dictionary, Career. URL: https://www.etymonline.com/word/career#etymonline_v_5378, (11.4.2020.).

Our World in Data, Internet. URL: https://ourworldindata.org/internet, (6.9.2020.).

PCMag eciklopedija: Kryders law. URL: https://www.pcmag.com/encyclopedia/term/kryders-law, (20.4.2020.).

Quinn, Kelly. "Why We Share: A Uses And Gratifications Approach To Privacy Regulation In Social Media Use". *Journal Of Broadcasting & Electronic Media* 60, no. 1 (2016): 61-86.

Rješenje kojim se izriče upravno novčana kazna zbog odbijanja dostave osobnih podataka. URL: https://azop.hr/aktualno/detaljnije/rjesenje-kojim-se-izrice-upravno-novcana, (1.4.2020.).

Roose, Kevin, *The Privacy Lawyer Giving Big Techan $8.8 Billion Headache.* URL: https://www.nytimes.com/2018/05/30/technology/gdpr-privacy-lawyer.html?auth=login-google&searchResultPosition=15 (1.4.2020.).

Ryan, Johnny. *A History Of The Internet And The Digital Future*. London: Reaktion Books, 2013.

Simon, Phil. *Too Big to Ignore: The Business Case for Big Data. Hoboken*, NJ : John Wiley & Sons, 2013.

Slukan, Mirela. *Povijesna Kartografija*. Samobor: Meridijani, 2003.

Stipčević Aleksandar. *Povijest knjige*. Zagreb: Matica hrvatska, 2006.

Stupp, Catherine, *Brussels Report: Frustated European Regulator Seek Tweaks to GDPR.* URL: https://www.wsj.com/articles/brussels-report-frustrated-european-regulators-seek-tweaks-to-gdpr-11584135669?mod=searchresults&page=1&pos=6 (1.4.2020.).

Swisher, Kara, *Be Paranoid About Privacy.* URL: https://www.nytimes.com/2019/12/24/opinion/location-privacy.html?searchResultPosition=3 (1.4.2020.).

Tanner, Adam. *What Stays In Vegas: The World Of Personal Data, Lifeblood Of Big Business And The End Of Privacy As We Know It*. New York: PublicAffairs, 2014.

Trepte, Sabine, and Leonard Reinecke. "The Social Web as a Shelter for Privacy and Authentic Living." U *Privacy Online: Perspectives on Privacy and Self-Disclosure in*

the Social Web, ur. Sabina Trepte i Leonard Reinecke, 61-73. New York: Springer, 2011.

Tuđman, Miroslav. *Informacijsko ratište i informacijska znanost*. Zagreb: Hrvatska Sveučilišna Nakl., 2008.

Zakon.hr, *Opća uredba o zaštiti osobnih podataka*. URL: https://www.zakon.hr/z/1021/Opća-uredba-o-zaštiti-podataka---Uredba-(EU)-2016-679, (22.4.2020.).

Zmajić, Bartol. *Heraldika, sfragistika, genealogija, veksilologija, rječnik heraldičkog nazivlja*. Zagreb: Golden Marketing, 1996.

Analiza dostupnosti osobnih podataka nastalih prije i nakon 2000. godine

Sažetak

U diplomskom radu uspoređuje se dostupnost osobnih podataka nastalih u 19. i 20 stoljeću s onima iz 21. stoljeća. Pod terminom „dostupnost“ podrazumijeva se dostupnost podataka u trenutku pisanja rada. Prvi će dio rada *de facto* odnosi na genealoška istraživanja, ali u manjoj mjeri i na druge vrste istraživanja (npr. imovinsko-pravnih odnosa). Analiziraju se izvori u kojima se ti podaci nalaze, njihova dostupnosti očuvanost, relevantnost i sl. te prikazuju kvalitativne i kvantitativne vrijednosti podataka iz tih izvora. Drugi dio rada obrađuje javno dostupne osobne podatke u kontekstu njihove zaštite (npr. GDPR) i društvenih mreža. Ukazuje se na činjenicu da iako su podaci zaštićeni, pojedinci, često svojevoljno, imaju relativno velik broj podataka o sebi javno dostupan (uglavnom preko društvenih mreža). Cilj rada je ukazati na promjene u odnosu prema osobnim podacima i vrsti javno dostupnih osobnih podataka.

Ključne riječi: privatnost, podaci, genealogija, velike količine podataka, zaštita osobnih podataka

Analysis of the availability of personal data generated before and after the year 2000

Summary

The thesis compares the availability of personal data created in the 19th and 20th centuries with those of the 21st century. "Availability" means the availability of data at the time of writing of this thesis. The first part of the paper *de facto* refers to genealogical research, but also to a lesser extent to other types of research. The sources in which these data are located, their availability, preservation, relevance, etc. are analyzed. Qualitative and quantitative values of data from these sources are shown. The second part of the thesis deals with publicly available personal data in the context of their protection (e.g. GDPR) and social networks. It is pointed out that although data is protected, individuals, often voluntarily, have a relatively large amount of information about themselves publicly available (mostly through social networks). The aim of the paper is to point out changes in the attitude towards personal data and their public availability.

Key words: privacy, data, genealogy, big data, personal data protection