Zaštita privatnosti – anonimizacija podataka

Privacy protection – data anonymization

Aleksandra Bradić-Martinović, Aleksandar Zdravković,

Institut ekonomskih nauka, Beograd

Ovaj rad je sastavni deo projekata 179001, 179015 i III47009 koje finansira Ministarstvo obrazovanja, nauke i tehnološkog razvoja Republike Srbije.

Rezime: **Eksponencijalni rast digitalnih podataka u prethodnih dvadeset godina uzrokovao je povećanje mogućnosti ugrožavanja privatnosti pojedinca. Privatnost može biti ugrožena neovlašćenim pristupom i zloupotrebom u poslovnim informacionim sistemima, javnim institucijama, društvenim mrežama, ali i pristupom servisima u kojima se nalaze podaci koji su rezultat istraživanja u društvenim naukama. Širenjem *Open Access* inicijative, koja omogućuje otvoren pristup podacima, problem je dobio novu dimenziju.**

U ovom radu predstavili smo postupak anonimizacije podataka, kao jedan od metoda koji istovremeno omogućuje zaštitu privatnosti uz nesmetano obavljanje istraživačkog procesa. U radu smo precizno definisali fazu istraživanja u kojoj je neophodno primeniti postupak anonimizacije, a zatim smo opisali osnovne dimenzije prirode podataka, od kojih zavisi izbor odgovarajuće tehnike. Centralni deo rada je posvećen najvažnijim tehnikama, posmatranim sa stanovišta ciljeva anonimizacije, *k*-anonimnost, *l*-raznolikost i *t*-bliskost.

Ključne reči: **baze podataka, anonimizacija, privatnost, digitalne arhive podataka, istraživanje**

Abstract: **The exponential growth of digital data over the past twenty years has caused an increase possibility of invasion on individual privacy. Privacy can be compromised by unauthorized access and misuse of business information systems, public institutions, social networks, and access to services in which the data resulting from research in the social sciences. The expansion of *Open Access* initiative, which allows open access to the data, the problem is a new dimension.**

In this paper we present the process of data anonymization, a method that simultaneously provides privacy to the smooth operation of the research process. We also define in a research phase in which it is necessary to apply the anonymization process, and then we describe the main dimensions of the nature of data, the choice of which depends on proper technique. The central part is devoted to the most important techniques, looking at the goals of anonymity, k-anonymity, l-diversity and t-closeness.

Index terms: **databases, anonymization, privacy, digital data archive, research**

1. UVOD

Doba digitalizacije omogućilo je relativno brzo, jednostavno i jeftino prikupljanje i čuvanje podataka o fizičkim i pravnim licima. Javne ustanove, zdravstvene institucije, finansijske organizacije i poslovni sistemi poseduju ogromne količine individualnih informacija koje se čuvaju u odgovarajućim bazama podataka. Mogućnosti upotrebe ovih informacija su praktično neograničene, ali se istovremeno mogu posmatrati dobri i loši aspekti ovog fenomena. Sa jedne strane, mora se uočiti veliki potencijal koji ove informacije pružaju za poboljšanje uslova života, kroz kreiranje novih i unapređenje postojećih usluga, kao i mogućnosti spoznaje navika, želja i preferencija korisnika i potrošača. Sa druge strane,

Aleksandra Bradić-Martinović, Institut ekonomskih nauka, Beograd, Zmaj Jovina 12, 11000 Beograd, Srbija, (abmartinovic@ien.bg.ac.rs)

Aleksandar Zdravković, Institut ekonomskih nauka, Beograd, Zmaj Jovina 12, 11000 Beograd, Srbija, (aleksandar.zdravkovic@ien.bg.ac.rs)

međutim ove informacije u velikoj meri zadiru u privatnost pojedinaca ili osetljive aspekte poslovanja preduzeća, a njihovo objavljivanje može uzrokovati niz vrlo loših posledica [3, str.1]. U većini demokratskih zemalja zaštita privatnih i ličnih podataka garantovana je zakonom [9, str.1].

Jedan od najpoznatijih primera ekstrakcije i zloupotrebe informacije iz baze digitalnih podataka je slučaj koji se desio kada je američka kompanija AOL (*America On-line*) objavila 2006. godine detaljne logove pretraživanja servisa AOL, za veliki broj svojih korisnika. Objavljivanje ovih informacija bilo je namenjeno u svrhe istraživanja, ali se ispostavilo da javno objavljivanje podataka podrazumeva mogućnost da svi imaju pristup podacima, a ne samo ograničena akademska zajednica. AOL nije vršio nikakvu obradu podataka, niti primenio neku od naprednijih tehnika anonimizacije (korisnici su samo imali numeričke identifikacione brojeve umesto imena), čime je otvorio potencijalnu mogućnost za otkrivanje identiteta korisnika od strane javnosti. Prvi slučaj otkrivanja identiteta desio se u slučaju Telme Arnold, sa ID #4417749, šezdesetdvogodišnja udovica iz Luburna, država Džordžija. Pronalaženje gospođe Arnold je dobijeno na osnovu ukrštanja nekoliko obeležja: specifične bolesti, godina i loše navike psa. Posledica ovog propusta dobila je veliki negativni odjek u javnosti, a protiv kompanije AOL je podneta tužba.

Komercijalno prikupljanje i objavljivanje podataka je samo jedna dimenzija, a pored nje postoje i podaci o ličnosti koje prikupljaju administrativne institucije i naučno-istraživačke institucije koje se bave istraživanjima u društvenim naukama (sociologija, psihologija i slično) ili prirodnim naukama, kao što je medicina. Obelodanjivanje ovih podataka, bez prethodne anonimizacije može u velikoj meri da ugrozi privatnost ispitanika.

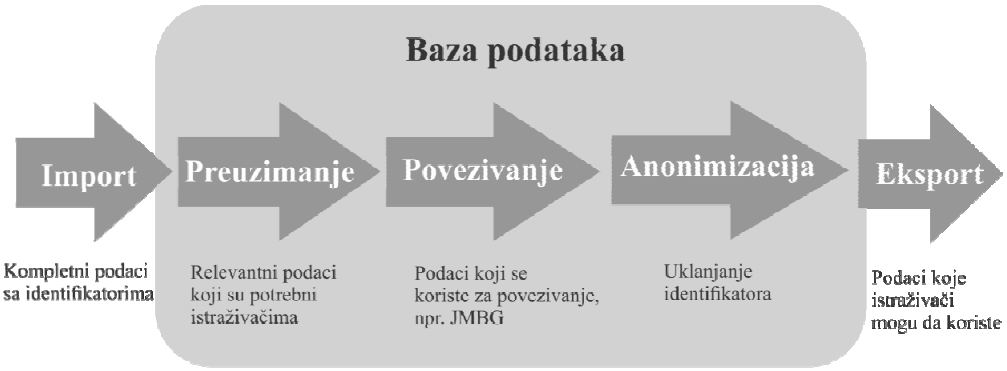
U kontekstu čuvanja podataka posebno mora voditi računa o raspoloživosti podataka u digitalnom obliku, koji se jednostavno i brzo mogu distribuirati i umnožavati. Da bi se izbegle situacije navedene u prethodnom primeru potrebno je pažljivo anonimizirati podatke pre postupka diseminacije. Osnovni cilj anonimizacije je očuvanje privatnosti fizičkog ili pravnog lica. Prema proceni Baker i ostalih [9, str. 2] pitanje anonimizacije je pre svega pravne prirode, ali je direktno povezano sa mogućnostima i primenom informaciono-komunikacione tehnologije u privatnom i poslovnom životu. Osim toga, pitanju je veoma osetljivo i sa stanovišta etike.

U Republici Srbiji ova tema postaje posebno aktuelna nakon 2008. godine, kada je usvojen Zakon o zaštiti podataka o ličnosti, koji ima za cilj da u vezi sa obradom podataka o ličnosti, svakom fizičkom licu obezbedi ostvarivanje i zaštitu prava na privatnosti i ostalih prava i sloboda.

2. POJAM ANONIMIZACIJE

Anonimizacija je postupak trajnog i potpunog uklanjanja identifikacije iz podataka pretvaranjem ličnih informacija u agregatne podatke. Anonimizirani podaci ne mogu više biti povezani sa individualnim podacima na bilo koji način. Drugim rečima, kada se jednom ukloni lični identifikacioni element iz podataka, ti elementi se ne mogu ponovo povezati sa individualnim podacima ili sa konkretnom osobom.

Prikupljanje i diseminacija podataka, uz uključen proces anonimizacije, bez obzira na oblast u kojoj je primenjeno, može se predstaviti šemom na grafiku 1. U bazu podataka, arhiv ili skladište podataka ulaze kompletni podaci koji sadrže identifikatore. Istraživači (korisnici) imaju potrebu za određenim, relevantnim podacima, ali oni ne smeju prilaziti i koristiti podatke koji su povezani sa identifikatorima, pa je pre eksporta podataka potrebno izvršiti proces anonimizacije. Tek nakon toga je omogućeno da se sačuva privatnost lica na koja se podaci odnose.



Grafik 1:Prikupljanje i diseminacija podataka uz proces anonimizacije, prilagođeno [4, str. 1141]

Pre no što pristupimo objašnjenju tehnika kojima je moguće postići anonimizaciju podataka potrebno je pojasniti osnovne odrednice vezane za prikupljene podatke, koji se obično čuvaju u bazama podataka, predstavljenim u obliku tabela. Svaka kolona tabele, koja se definiše kao atribut predstavlja semantičku kategoriju informacije. Za lakšu analizu pravi se podela atributa na sledeće tri kategorije [7, str. 53]: *Identifikatori* su atributi koji mogu jedinstveno identifikovati pojedinca. U Srbiji najbolji primer ovakvog identifikatora je Jedinstveni matični broj građana (JMBG) za fizička lica ili Poreski identifikacioni broj (PIB) za pravna lica. Ova vrsta podataka se prva eliminiše iz seta podataka, jer su oni najlakši i siguran put ka narušavanju privatnosti. *Kvazi-identifikatori,* odnosno *ključni atributi* su oni koji u kombinaciji sa informacijama iz drugog izvora mogu da omoguće identifikaciju. Oni ne moraju da pruže identifikaciju svih osoba u bazi, ali mogu da ukažu na neku osobu ukoliko ona ima specifičan spoj ovih, ključnih atributa. Za fizička lica to su kombinacija atributa npr. datum, mesec i godina rođenja, pol, opština na kojoj žive, a za pravna lica to mogu biti tržišno učešće i grad u kome su registrovani. *Osetljivi atributi* su oni koji sadrže poverljive informacije, npr. bolest pacijenta i koje nikako ne smeju biti objavljene.

Prema istraživanjima neprofitne organizacije Educause [2] bez obzira na to što je pojam anonimizacije relativno lako definisati, to nije slučaj kada pokušamo da posmatramo podatke koji su anonimizirani. U većini slučajeva institucije se suočavaju sa izazovom i rizikom prilikom odabira odgovarajuće procesa koji bi im obezbedio pouzdanu anonimizaciju. Jedan od problema sa kojim se suočavaju je nedostatak precizne definicije pojma informacije koja omogućuje da se izvrši identifikacija (*Personally Identifiable Information - PII*) i informacije pomoću koje nije moguće izvršiti identifikaciju (*Non-PII*). Uzrok tome je složenost problema koji je nemoguće svesti na prost spisak ovih informacija (elemenata), zbog toga što informacije menjaju značenje u zavisnosti od konteksta u kome se nalaze, kao i raspoloživosti ostalih podataka sa kojima su povezani. Zbog ovog problema postupak anonimizacije predstavlja svojevrstan izazov za one koji se bave čuvanjem i publikovanjem setova podataka.

3. TEHNIKE ANONIMIZACIJE

Do sada je razvijen određen broj, manje ili više uspešnih tehnika anonimizacija podataka. Da bi se utvrdila odgovarajuća tehnika kojom će biti sprovedena anonimizacija podataka, potrebno je izvršiti obuhvat više dimenzija koje se odnose na ovaj proces [3, str. 53]

Priroda podataka

Postoji razlika u primeni tehnike anonimizacije u zavisnosti od toga da li su podaci:

a) *tabelarni podaci*, koji predstavljaju informacije o entitetima (npr. fizičkim ili pravnim licima), njihovim kvazi-identifikatorima (npr. godine, pol, poštanski broj) ili njihovim osetljivim podacima (npr. visina plate, bolesti);
b) *stavke seta podataka*, kojima su predstavljene transakcije (ili „tržišna korpa“), odnosno koji povezuju, na primer određene osobe i proizvode koje je ta osoba kupila u transakciji i
c) grafički podaci, kojima su predstavljeni osetljivi odnosi između entiteta (npr. ljudi koji su članovi iste društvene mreže).

Pristupi anonimizaciji

Moguće tehnike anonimizacije podrazumevaju:

a) *Uklanjanje (direktnog) identifikatora*, npr. imena ili adrese sa ciljem da se isključi mogućnost povezivanja konkretnih lica uključenih u bazu sa raspoloživim podacima. Uklanjanje identifikatora je najjednostavnija tehnika anonimizacije, a primer je predstavljen tabelom 1. Problem ove tehnike je u tome što ona često nije dovoljna da obezbedi privatnost lica čiji su podaci raspoloživi.

TABELA 1

Primer uklanjanja direktnog identifikatora

Sirovi podaci

R.br.	Ime	Poštanski broj	Pol	Godine	Hobi
1	Jelena Petrović	11000	Ženski	38	Filmovi
2	Milena Janković	15000	Ženski	32	Filmovi
3	Dragan Todorović	12350	Muški	25	Skejt bord
4	Slavica Pešić	14000	Ženski	48	Heklanje
5	Antonije Dišić	11050	Muški	62	Pecanje
6	Petar Ljubičić	11070	Muški	28	Fitnes

Anonimizirani podaci

R.br.	Ime	Poštanski broj	Pol	Godine	Hobi
1	UKLONJEN IDENTIFIKATOR	11000	Ženski	38	Filmovi
2		15000	Ženski	32	Filmovi
3		12350	Muški	25	Skejt bord
4		14000	Ženski	48	Heklanje
5		11050	Muški	62	Pecanje
6		11070	Muški	28	Fitnes

U prethodnom primeru, bez obzira na to što je uklonjen identifikator, verovatno ne bi bilo teško zaključiti koji dvadesetpetogodišnji mladić u malom mestu intenzivno vozi skejt bord. Zbog toga se pored ove tehnike primenjuju i dopunske, sa ciljem smanjenja verovatnoće otkrivanja identiteta u slučajevima izuzetaka, što svakako nije muškarac od 28 godina, koji se bavi fitnesom na Novom Beogradu.

b) *Agregiranje ili smanjenje preciznosti informacija ili promenljivih*. Ova tehnika se najčešće sprovodi tako što se podaci grupišu, npr. zamena datuma rođenja starosnim grupama. Primer je predstavljen tabelom 2.

TABELA 2

Agregiranje kvazi- identifikatora

Anonimizirani podaci

R.br.	Ime	Poštanski broj	Pol	Godine	Hobi
1	UKLONJEN IDENTIFIKATOR	11000	Ženski	31-40	Filmovi
2		15000	Ženski	31-40	Filmovi
3		12350	Muški	21-30	Skejt bord
4		14000	Ženski	41-50	Heklanje
5		11050	Muški	61-70	Pecanje
6		11070	Muški	21-30	Fitnes

c) *Peturbacija (uvođenje šuma) podataka za jednog ili više kvazi-identifikatora*. Kolone sa određenim obeležjima se generalizuju pomoću odgovarajuće distribucije, tako da više ne postoji veza između podataka koja bi omogućila otkrivanje identiteta [5, str. 10]. Ova tehnika spada u statističke tehnike. Primer je dat u tabeli 3.

TABELA 3

Mešanje podataka jednog ili više kvazi- identifikatora

Anonimizirani podaci					
R.br.	Ime	Poštanski broj	Pol	Godine	Hobi
1	UKLONJEN IDENTIFIKATOR	11000	Ženski	sredji	Filmovi
2		15000	Ženski	sredji	Filmovi
3		12350	Muški	mlađi	Skejt bord
4		14000	Ženski	sredji	Heklanje
5		11050	Muški	stariji	Pecanje
6		11070	Muški	mlađi	Fitnes

d) *Permutacija podataka jednog ili više kvazi-identifikatora.* Ova tehnika se sprovodi tako što se u kolonama sa određenim obeležjima zamene vrednosti (npr. redosled), kao što je predstavljeno tabelom 4. Istraživač i dalje može da istražuje pojavu, jer poseduje podatke o frekvencijama, ali više nije u stanju da otkrije identitet osobe koja obuhvaćena istraživanjem.

TABELA 4

Permutacija podataka jednog ili više kvazi- identifikatora

Anonimizirani podaci					
R.br.	Ime	Poštanski broj	Pol	Godine	Hobi
1	UKLONJEN IDENTIFIKATOR	11000	Ženski	25	Filmovi
2		15000	Ženski	28	Filmovi
3		12350	Muški	32	Fitnes
4		14000	Ženski	38	Heklanje
5		11050	Muški	48	Pecanje
6		11070	Muški	62	Skejt bord

Ciljevi anonimizacije

Osim navedenih kriterijuma koji utiču na izbor odgovarajuće tehnike anonimizacije, različiti ciljevi zbog kojih postoji potreba da se obezbedi privatnost, takođe imaju uticaj na izbor, sve dok podaci ne postignu odgovarajuće osobine. Tu spadaju sledeće tehnike:

a) *K-anonimnost*. Tehniku *k-anonimity* predložili su Samarati i Sweeny [8] da bi se sprečila mogućnost povezivanja kvazi-identifikatora, a sa ciljem da se zaštiti privatnost lica uključenih u bazu podataka. Smatra se da set podataka zadovoljava uslov k-anonimnosti ukoliko se svaki zapis u tabeli razlikuje od najmanje k-1 ostalih zapisa, po pitanju svake kombinacije atributa kvazi-identifikatora. Prema tome, za svaku kombinaciju vrednosti kvazi-identifikatora u k-anonimiziranoj tabeli, postoji bar još *k* zapisa koji dele te vrednosti. Jednostavno rečeno, svaki kvazi-identifikator se mora pojaviti u identičnim kombinacijama u tabeli u najmanje *k* zapisa. Time je obezbeđeno da pojedinačni zapis ne može biti identifikovan metodom ukrštanja vrednosti obeležja (*cross tabluation*). Takva tabela naziva se k-anonimizirana tabela.

Primer: U narednoj tabeli prikazana je medicinska evidencija fiktivne bolnice u Njujorku. Tabela na sadrži direktne identifikatore poput imena, prezimena i broja socijalnog osiguranja. Raspoloživa obeležja podeljena su u dve grupe: 1) osetljivi (zdravstveno stanje) i 2) ne-osetljivi (poštanski broj, godine i zemlja porekla). Obeležja koja su označena kao osetljiva nikako ne smeju biti povezana sa bilo kojom jedinstvenom pojedinačnom vrednošću. Drugim rečima, ne smeju da omoguće otkrivanje identiteta pojedinačnog bolesnika.

TABELA 5

Primer k-anonimizacije

Sirovi podaci

	Ne-osetljivi podaci			Osetljivi podaci
R.br.	Poštanski broj	Godine	Zemlja porekla	Stanje
1	13053	28	Rusija	Srčano oboljenje
2	13068	29	Amerika	Srčano oboljenje
3	13068	21	Japan	Virusna infekcija
4	13053	23	Amerika	Virusna infekcija
5	14853	50	Indija	Psihološki poremećaj
6	14853	55	Rusija	Srčano oboljenje
7	14850	47	Amerika	Virusna infekcija
8	14850	49	Amerika	Virusna infekcija
9	13053	31	Amerika	Psihološki poremećaj
10	13053	37	Indija	Psihološki poremećaj
11	13068	36	Japan	Psihološki poremećaj
12	13068	35	Amerika	Psihološki poremećaj

Anonimizirani podaci

	Ne-osetljivi podaci			Osetljivi podaci
R.br.	Poštanski broj	Godine	Nacionalnost	Stanje
1	130**	<30	*	Srčano oboljenje
2	130**	<30	*	Srčano oboljenje
3	130**	<30	*	Virusna infekcija
4	130**	<30	*	Virusna infekcija
5	148**	≥40	*	Psihološki poremećaj
6	148**	≥40	*	Srčano oboljenje
7	148**	≥40	*	Virusna infekcija
8	148**	≥40	*	Virusna infekcija
9	130**	3*	*	Psihološki poremećaj
10	130**	3*	*	Psihološki poremećaj
11	130**	3*	*	Psihološki poremećaj
12	130**	3*	*	Psihološki poremećaj

Postupak anonimizacije ove tabele podrazumevao je: a) potpuno uklanjanje obeležja „Zemlja porekla“, b) zamena jedne ili više vrednosti u polju „Poštanski broj“ zvezdicom, tako da vrednost 1485* može biti bilo koji broj između 14850 i 14859 i c) druga cifra u obeležju „Godine“ zamenjena je, takođe, zvezdicom, tako da je vrednost 3* u rangu od 30 do 39 godina.

Postignut je cilj postupka anonimizacije u tabeli, jer se jedna ista kombinacija obeležja pojavljuje više puta (najmanje tri puta). U ovom slučaju u pitanju je tzv. 4-anonimizirana tabela, jer najmanje 4 zapisa dele isti kvazi-identifikator.

b) L-raznovrsnost (l-diversity). K-anonimnost često nije dovoljno dobra tehnika, zato što je uprkos prisustvu više identičnih kombinacija obeležja, ipak moguće otkriti identitet pojedinca. Ovo se javlja kao nedostatak raznolikosti kombinacije obeležja. Ukoliko, npr. imamo 10 različitih muškaraca, starih 54 godine, koji žive u malom mestu i oni imaju HIV, onda Piter može biti izložen sumnji/diskriminaciji ukoliko živi u malom mestu i ima 54 godine. Zbog toga je potrebno uvesti raznovrsnost u grupu formiranu na osnovu jednog obeležja. L-raznovrsnost znači da grupa sadrži *l* različitih vrednosti.

TABELA 6

Primer l-raznovrsnosti

Anonimizirani podaci

	Ne-osetljivi podaci			Osetljivi podaci
R.br.	Poštanski broj	Godine	Nacionalnost	Stanje
1	1305*	≤40	*	Srčano oboljenje
2	1305*	≤40	*	Srčano oboljenje
3	1305*	≤40	*	Virusna infekcija
4	1305*	≤40	*	Virusna infekcija
5	1485*	>40	*	Psihološki poremećaj
6	1485*	>40	*	Srčano oboljenje
7	1485*	>40	*	Virusna infekcija
8	1485*	>40	*	Virusna infekcija
9	1306*	≤40	*	Psihološki poremećaj
10	1306*	≤40	*	Psihološki poremećaj
11	1306*	≤40	*	Psihološki poremećaj
12	1306*	≤40	*	Psihološki poremećaj

U tabeli 6 prikazan je primer tabele na kojoj je primenjena tehnika 3-raznolikosti. U poređenju sa k-anonimiziranom tabelom 5, Bob ne može biti identifikovan ukoliko je samo Amerikanac, 31 godine starosti, koji živi na teritoriji poštanskog broja 13053 i ima psihološki poremećaj.

Postoji nekoliko vrsta l-raznolikosti: probablisistička, entropijska, rekurzivna i više-atributivna.

c) *T-bliskost (t-closeness)*. Ova tehnika predstavlja unapređenje prethodno opisanih tehnika i kreirana je sa ciljem da otkloni postojeće nedostatke.
Privatnost se može izmeriti kao informaciona dobit istraživača (posmatrača). Pre no što dobije podatke na uvid istraživač već ima određena uverenja / verovanja o pojedinačnim vrednostima osetljivih obeležja (a' priori verovanja). Nakon uvida u podatke istraživač formira nova uverenja / verovanja (a' posteriori verovanja). Razlika koja postoji između a' priori i a' posteriori verovanja je informaciona dobit. Prema novim istraživanjima [5] informaciona dobit se može podeliti u dve celine: ona koja se ostvaruje na osnovu podataka o celoj populaciji i ona koja se ostvaruje na osnovu podataka o specifičnom pojedincu. Za ekvivalente skupove (klase) podataka kažemo da imaju t-bliskost ukoliko razlika između distribucija osetljivih obeležja u skupu i distribucije obeležja u celoj tabeli nije iznad utvrđenog praga t. Za tabelu kažemo da ima t-bliskost ukoliko svi njeni ekvivalentni skupovi imaju t-bliskost. Poenta tehnike je da se prilagodi distribucija pojedinačnih skupova distribuciji koja postoji u celoj populaciji.

Uprkos razvoju i primeni raznovrsnih tehnika kojima se vrši anonimizacija podataka u bazama, postoji niz metoda kojima se ugrožava privatnost. Veoma je važno istaći da se ni jedna od prethodno navedenih metoda ne može primenjivati automatski. Najbolje je da osoba koja je prikupila podatke sa određenim ciljem ima i uvid u rizik po pitanju ugrožavanja privatnosti koji može nastati ukoliko se izvrši diseminacija podataka na neodgovarajući način.

U praksi Arhiva podataka Velike Britanije (UK Data Archive – UKDA) uvek navode primer istraživanja u kome su direktori verskih škola iznosili mišljenje o načinu vođenja lokalne samouprave. Bez obzira na to da je izvršena anonimizacija podataka i uklonjeni svi direktni identifikatori, promakla je greška, jer se ispostavilo da je ugrožena privatnost izvesne žene koja je bila jedini direktor verske škole u tom okrugu, koja je ujedno ženskog pola. Problem je nastao zbog toga što je iznela vrlo oštre kritike, uz uverenje istraživača da niko neće moći da otkrije njen identitet. Ovaj primer pokazuje koliko je potrebno detaljno poznavati materiju na koju se podaci odnose, kao i specifičnost pojedinih kombinacija obeležja. To je često veoma teško obezbediti.

4. ZAKLJUČAK

Osetljivost privatnih podataka posebno je postala aktuelna pojavom digitalizacije i globalnog umrežavanja, jer je dostupnost podataka u savremenom svetu jednostavna i masovna. Podatke prikupljaju i distribuiraju, pre svega, kompanije, administracija i naučno-istraživački centri. Međutim, svi navedeni moraju strogo voditi računa o privatnosti ličnosti, jer su do sada zabeleženi slučajevi drastičnog ugrožavanja privatnosti od strane pojedinih kompanija, čime se krše i zakoni koji su uvedeni u velikom broju zemalja.

Jedan od načina da se privatnost ličnosti zaštiti, a da se podaci ipak distribuiraju, je primena metoda anonimizacije podataka. U ovom radu prikazali smo najvažnije metode koji se koriste za anonimizaciju podataka u razvijenim sistemima za čuvanje i distribuciju podataka, kao i njihove prednosti i nedostatke. Ukazali smo da sam proces i odabir konkretnog metoda anonimizacije zavisi od tipa podataka, od prostog uklanjanja direktnog identifikatora u slučaju manje osetljivih podataka, do primene neke od kompleksnijih metoda anonimizacije na indirektnim (kvazi) identifikatorima u slučaju kada je rizik otkrivanja identiteta visok. Ova oblast se stalno razvija, tako da se pojavljuju novi metodi kojima je moguće sve uspešnije zaštititi privatnost pojedinaca čiji su podaci raspoloživi i dostupni u bazi podataka.

Republika Srbija uspešno prati IT trendove u raznim oblastima poslovanja, administracije i nauke, tako da tema anonimizacije postaje vrlo aktuelna.

LITERATURA

[1.] Machanavajjhala, J. Gehrke, D. Kifer, (2007), L-diversity: Privacy beyond k-anonymity, ACM Transactions on Knowledge Discovery from Data (TKDD), Volume 1, Issue 1, Article no. 3

[2.] Educause, Guidelines for Data De-Identification, http://www.educause.edu/ero, poslednja poseta oktobar 2013.

[3.] G. Cormode, D. Srivastava, (2010), Anonymized Data: Generation, Model, Usage, Book of Proceedings International Council for Open and Distance Education, SIGMOD '09

[4.] G. Haddow, A. Bruce, S. Sathanandam, J. Wyatt, (2011), Nothing is really safe: a focus group study on the processes of anonymizing and sharing of health data for research purposes, International Journal of Public Health Poliy and Health Services Research

[5.] L. Liu, M. Kantarcioglu, B. Thuraisingham, (2008), The applicability of the peturbation based privacy preserving data mining for real-world data, Data & Knowledge Engineering 65, 5-21

[6.] L. Ninghui, L. Tiancheng, V. Suresh, (2007), t-closeness: Privacy Beyond k-Anonymity and l-Diversity, Data Engineering, ICDE 2007. IEEE 23rd International Conference on Data Enginering

[7.] M. Stanek, A. Jurišić, M. Babić, (2008), Anonimizacija podatkovnih baz, Bilt-ekon organ inform zdrav 2009, (25)2:53-59

[8.] P. Samarati, L. Sweeney, (1998), Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression, Technical report, CMU, SRI, 1998.

[9.] S. Baker, K. Kuilwijk, W. Chang, D. Mah, (2003), Anonimization, Data-Matching and Privacy: A Case Study, Steptoe & Johnson LLP, Attorneys at Law