CARNET

Anonimizacija i pseudonimizacija podataka

CERT.hr-PUBDOC-2018-8-367

CERT.hr

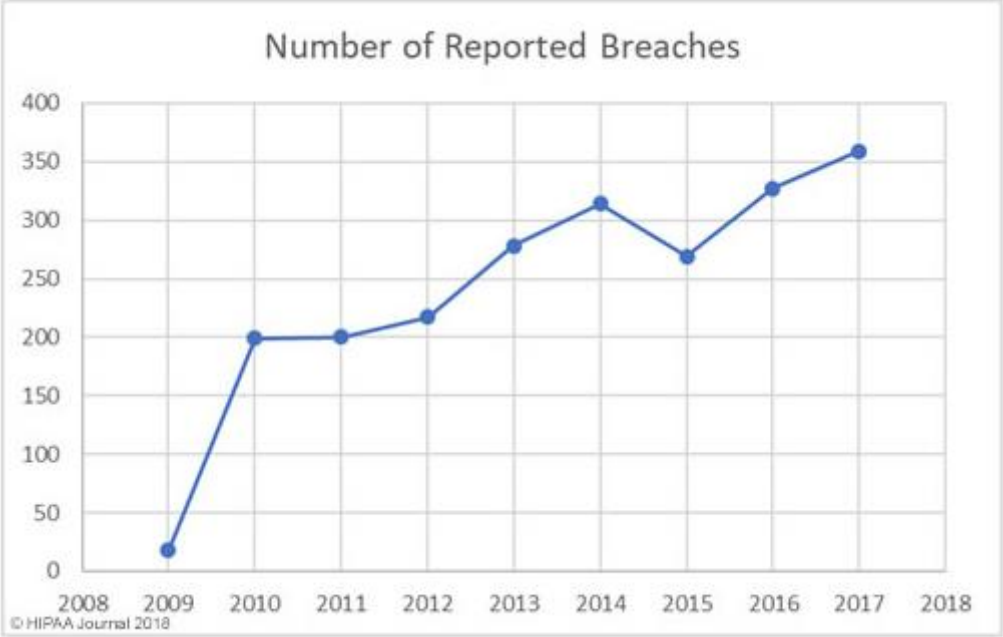
Sadržaj

Ovaj dokument izradio je Laboratorij za sustave i signale Zavoda za elektroničke sustave i obradbu informacija Fakulteta elektrotehnike i računarstva Sveučilišta u Zagrebu.

1 Uvod

Povećanim korištenjem tehnologije, tvrtke i pojedinci postaju sve ovisniji o digitalnim podacima pohranjenim na lokalnim računalima, u poslovnim bazama podataka ili poslužiteljima u oblaku. Zbog takvog načina pohrane, sprječavanje neovlaštenog pristupa podacima svodi se na zaštitu mreža i računala na kojima su podaci pohranjeni.

U ovakvom okruženju, učestalost i posljedice neovlaštenih pristupa podacima najlakše je objasniti brojčano: u periodu od 2009. do 2017. godine, u SAD-u je kompromitirano preko 176 milijuna zapisa zdravstvenih podataka – taj broj odgovara otprilike polovici populacije SAD-a [1]. Incidenti neovlaštenih pristupa podacima, tj. točnije, **povrede sigurnosti podataka, postaju sve učestalije**. Slika 1 prikazuje porast broja prijavljenih povreda sigurnosti zdravstvenih podataka u SAD-u kroz posljednjih osam godina.



Slika 1 – broj prijavljenih povreda sigurnosti zdravstvenih podataka u SAD-u [1]

Povreda sigurnosti podataka (eng. *data breach*) je, općenito, neovlašteni pristup povjerljivim podacima. Ti podaci mogu biti:

- intelektualno vlasništvo tvrtke,
- poslovne tajne,
- podaci o kreditnim karticama,
- zdravstveni podaci pacijenata,
- i slično.

Žrtva povrede sigurnosti podataka u svakom je slučaju **organizacija** čiji su podaci kompromitirani. No kada su ukradeni osobni podaci, tada su žrtve i **sve osobe** čiji su podaci

izloženi neovlaštenom pristupu. U takvim slučajevima pričamo o **povredi sigurnosti osobnih podataka** (eng. *personal data breach*).

Primjerice, informatička tvrtka *FotoOblak* svojim korisnicima prodaje uslugu pohrane fotografija na njihove poslužitelje „u oblaku“. Povreda sigurnosti podataka tvrtke *FotoOblak* može biti **krađa izvornog koda** njihovog softvera za prijenos i pohranu podataka. Žrtva je u tom slučaju sama **tvrtka**, te će posljedica biti da će sada konkurenti imati pristup njihovoj tehnologiji i u konačnici će se smanjiti zarada tvrtke.

No kada bi u istoj tvrtki bile ukradene **fotografije i ostali podaci njihovih klijenata**, tada bi to bila povreda sigurnosti **osobnih** podataka. U tom slučaju žrtva nije samo tvrtka, već i **svi njeni klijenti** čije su fotografije i ostali podaci ukradeni. Kao posljedica, neće se samo tvrtki smanjiti zarada, već će neki od njenih klijenata biti i žrtve krađe identiteta, ucjene i slično.

Očigledno je da posljedice povrede sigurnosti osobnih podataka mogu biti i katastrofalne. No unatoč brojnim rizicima za pojedince, tvrtke često **nemaju poticaje** ulagati u zaštitu njihovih osobnih podataka. Kako bi se to ispravilo, diljem svijeta uvedeni su zakoni koji reguliraju kako tvrtke smiju obrađivati i moraju štititi osobne podatke.

U Europskoj Uniji, donesena je **Opća uredba o zaštiti podataka**, skraćeno **OUZP** (eng. *General Data Protection Regulation*, skraćeno ***GDPR***).

Kako bi se smanjio rizik od povrede sigurnosti osobnih podataka ili umanjile njene posljedice, postoji niz mjera zaštite koje je moguće primijeniti. Te mjere uključuju:

- **fizičku zaštitu** infrastrukture – primjerice zaključavanje i ograničavanje pristupa sobi s poslužiteljima,
- **zaštitu** infrastrukture **od kibernetičkih napada** – korištenje vatrozida (eng. *firewall*), redovito ažuriranje softvera, korištenje sustava za nadzor mreže, upotreba antivirusnih programa i slično,
- **šifriranje podataka** prilikom pohrane kako bi podaci ukradeni bez ključa za dešifriranje bili beskorisni napadaču,
- **edukacija** te **osvještavanje** zaposlenika o sigurnosnim rizicima.

Navedenim mjerama zaštite također pripadaju i **anonimizacija** te **pseudonimizacija podataka**. To su zaštitne mjere, osmišljene kako bi **minimizirale štetu** od povrede sigurnosti **osobnih** podataka. Kao dodatna prednost, anonimizacija i pseudonimizacija podataka pomažu u poštivanju regulativa poput Opće uredbe o zaštiti podataka.

2 Anonimizacija i pseudonimizacija podataka

U ovom poglavlju bit će općenito objašnjeni postupci anonimizacije i pseudonimizacije podataka, kako oni štite podatke, kako pomažu u poštivanju Opće uredbe o zaštiti podataka te koje su neke od konkretnih tehnika koje se koriste u tim postupcima.

2.1 Anonimizacija podataka

2.1.1 Što je anonimizacija podataka?

Anonimizacija podataka je postupak obrade **osobnih** podataka kojim se **nepovratno** sprječava identifikacija pojedinca iz obrađenih podataka [2]. Uzmimo za primjer skup podataka o posjetima bolnicama u kojemu se nalaze zapisi s:

- imenom i prezimenom pacijenta,
- adresom pacijenta,
- razlogom njegovog posjeta bolnici
- te imenom bolnice.

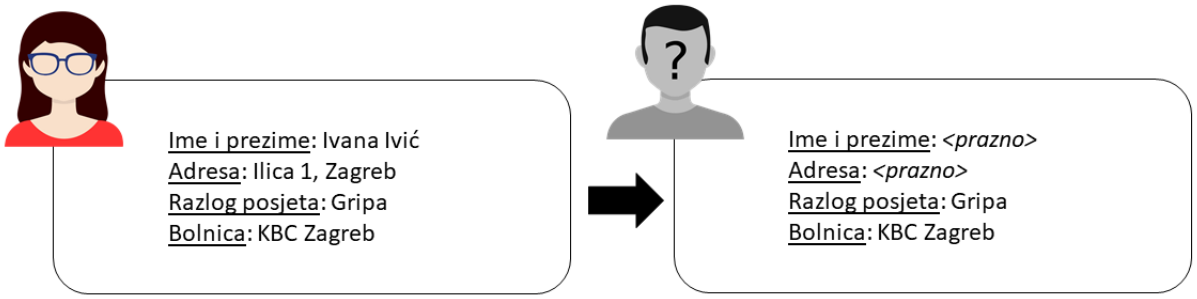
Jednostavan primjer anonimizacije tog skupa podataka bio bi uklanjanje podataka o imenu, prezimenu i adresi pacijenta iz svakog zapisa. Primjerice, kada bi u jednom zapisu pisalo da je

- pacijentica **Ivana Ivić**
- koja živi na **Ilici 1**
- zbog gripe
- bila u bolnici KBC Zagreb,

nakon anonimizacije, u tom zapisu pisalo bi samo da je:

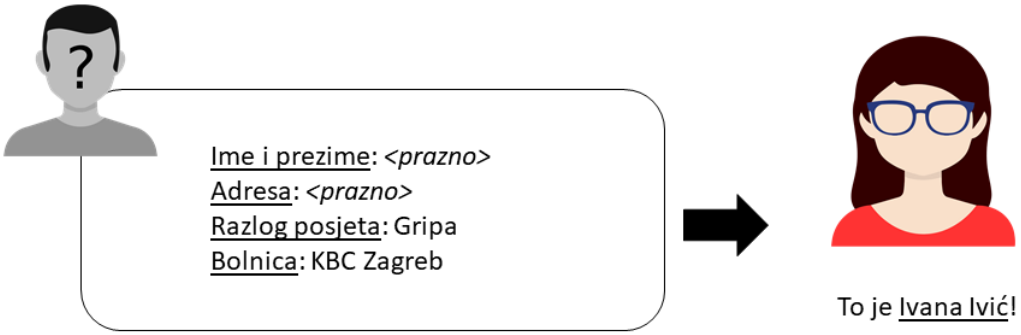
- neki (**anonimni**) pacijent
- zbog gripe
- bio u bolnici KBC Zagreb.

Ovaj primjer ilustriran je na Slika 2. Ovo je dobar uvodni primjer anonimizacije, no bitno je imati na umu (te će kasnije biti i detaljnije pojašnjeno) kako ovakvo uklanjanje elemenata nije uvijek dovoljno za ispravnu anonimizaciju podataka.



Slika 2 – jednostavan primjer anonimizacije podataka

Po definiciji, nakon ispravne anonimizacije **ne smije** biti moguće povezati anonimizirane podatke s određenom osobom. Koristeći prethodni primjer – ne smije biti moguće nekako otkriti da je upravo Ivana Ivić bila pacijent koji je posjetio bolnicu KBC Zagreb zbog gripe, kao što je prikazano na Slika 3. Drugim riječima, ne smije biti moguće **reidentificirati** pojedinca iz anonimiziranih podataka.



Slika 3 – povezivanje podataka s određenom osobom

2.1.2 Zašto anonimizirati podatke?

Prethodno je objašnjeno kakve mogu biti posljedice povrede sigurnosti **osobnih** podataka. No, što ako su kompromitirani **anonimizirani** podaci?

Na primjeru podataka o bolničkim posjetima, nakon anonimizacije više neće biti moguće otkriti zdravstveno stanje pojedinih osoba, primjerice: *Ivana Ivić često boluje od gripe*. Bit će moguće samo saznati općenite, statističke podatke o posjetima bolnicama, primjerice: *10% posjeta bolnici KBC Zagreb je zbog gripe*. Upravo na ovaj način, anonimizacijom je **znatno smanjena šteta** od potencijalne povrede sigurnosti podataka. Zato je vrijednost anonimizacije podataka prepoznata i u OUZP-u – prema njemu, ispravno anonimizirani osobni podaci više se **ne smatraju** osobnim podacima, kao što navodi uvodna izjava 26. [3]:

„ (…) Načela zaštite podataka stoga se ne bi trebala primjenjivati na anonimne informacije, odnosno informacije koje se ne odnose na pojedinca čiji je identitet utvrđen ili se može utvrditi ili na osobne podatke koji su učinjeni anonimnima na način da se identitet ispitanika ne može ili više ne može utvrditi. Ova se Uredba stoga ne odnosi na obradu takvih anonimnih informacija, među ostalim za statističke ili istraživačke svrhe.“

Jednostavnije – obveze OUZP-a uopće se **ne primjenjuju na anonimizirane podatke**. No, izrazito je bitno napomenuti da, dokle god izvorni podaci nisu izbrisani, anonimizirani podaci se **ne smatraju anonimnima** iz perspektive OUZP-a [2]. Drugim riječima, što se OUZP-a tiče, podaci postaju anonimni tek u trenutku kada više ne postoji njihov original [2]. Na primjeru podataka o bolničkim posjetima, dokle god **postoje** izvorni podaci koji sadrže i ime, prezime te adresu pacijenta, „anonimizirani" podaci koji sadrže samo razlog posjeta i ime bolnice se i dalje, po OUZP-u, smatraju **osobnim podacima**. U skladu s time, sve zaštite OUZP-a se i dalje primjenjuju na njih.

2.1.3 Kako se reidentificiraju pojedinci?

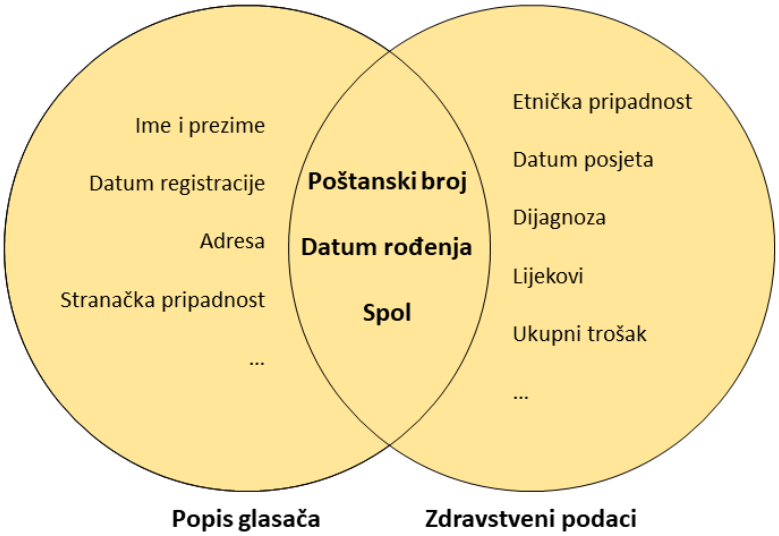
Za razumijevanje tehnika anonimizacije ključno je prvo osnovno razumjeti kako je moguće reidentificirati pojedince iz naizgled anonimiziranih podataka. Dobar primjer reidentifikacije je demonstracija istraživačice Latanye Sweeney u kojoj je ona otkrila **koji zapisi** iz naizgled **anonimizirane** baze zdravstvenih podataka pripadaju tadašnjem guverneru Massachusettsa, Williamu Weldu. Ona je uspješno reidentificirala guvernerove zapise povezivanjem zapisa iz dva različita izvora podataka. Prvi izvor bila je javna baza zdravstvenih podataka koja je sadržavala sljedeće informacije o pacijentima i njihovim posjetima bolnicama:

- poštanski broj mjesta stanovanja,
- datum rođenja,
- spol,
- etnička pripadnost,
- datum posjeta,
- dijagnoza,
- lijekovi,
- ukupni trošak liječenja,
- ...

Drugi izvor bio je javni popis registriranih glasača koji je sadržavao sljedeće informacije o glasačima:

- poštanski broj mjesta stanovanja,
- datum rođenja,
- spol,
- ime i prezime,
- adresa,
- stranačka pripadnost,
- datum zadnjeg glasanja,
- ...

Ključno je bilo to što su oba skupa podataka sadržavali podatke o **poštanskom broju**, **datumu rođenja** i **spolu** pojedinca. Istraživačica je pokazala da kombinacija ta tri podatka često može identificirati pojedinca. Pomoću kombinacije poštanskog broja, datuma rođenja i spola guvernera, ona je povezala zapise iz oba izvora te tako otkrila koji zapisi u zdravstvenoj bazi podatka pripadaju guverneru. Drugim riječima, istraživačica je **reidentificirala** guvernera iz **naizgled anonimne** baze zdravstvenih podataka [4]. Ovaj primjer prikazan je dijagramom na Slika 4.



Slika 4 – povezivanje zapisa iz dva odvojena izvora preko zajedničkih podataka

2.1.4 Tehnike anonimizacije

Kao što pokazuje prošli primjer, za ispravnu anonimizaciju nije dovoljno samo ukloniti podatke koji izravno identificiraju pojedince. Efektivan postupak anonimizacije trebao bi štititi od sljedećih ključnih rizika [2]:

- **izdvajanje** – mogućnost izoliranja nekoliko ili svih zapisa pojedinca u skupu podataka
- **povezivost** – mogućnost povezivanja dva ili više zapisa o istoj osobi ili skupini osoba
- **izvođenje zaključaka** – mogućnost zaključivanja vrijednosti jednog atributa iz vrijednosti skupa ostalih atributa

U svakom slučaju, prvi korak u postupku anonimizacije je uklanjanje elemenata koji izravno utvrđuju identitet pojedinca. Primjerice, potrebno je ukloniti:

- ime i prezime,
- fizičku adresu, IP adresu, adresu elektroničke pošte,
- fotografiju pojedinca,
- broj telefona,
- itd.

Takvi elementi zovu se **izravni identifikatori.** U primjeru podataka o bolničkim posjetima prikazanima na Slika 5, izravni identifikatori su ime i prezime te adresa pacijenta.

Ime i prezime	Adresa	Spol	Datum rođenja	Razlog posjeta	Bolnica
Ivana Ivić	Ilica 1	Ž	5.9.1980.	Gripa	KBC Zagreb
Jakov Marić	Žuta obala 3	M	30.1.1988.	Operacija oka	KBC Split
Ema Novak	Maksimir 76	Ž	23.4.2007.	Lom čeljusti	KBC Zagreb
Marko Jurić	Šegrtova 4	M	17.8.1987.	Ubod ose	KBC Pula
Lucija Perić	Ribarska 51	Ž	6.12.1995.	Trudovi	KBC Osijek
Luka Matić	Vinogradi 25	M	12.5.1972.	Povraćanje	KBC Zadar
Ana Babić	Splitska 89	Ž	9.4.1969.	Srčani udar	OB Koprivnica

Ime i prezime	Adresa	Spol	Datum rođenja	Razlog posjeta	Bolnica
[redacted]	[redacted]	Ž	5.9.1980.	Gripa	KBC Zagreb
		M	30.1.1988.	Operacija oka	KBC Split
		Ž	23.4.2007.	Lom čeljusti	KBC Zagreb
		M	17.8.1987.	Ubod ose	KBC Pula
		Ž	6.12.1995.	Trudovi	KBC Osijek
		M	12.5.1972.	Povraćanje	KBC Zadar
		Ž	9.4.1969.	Srčani udar	OB Koprivnica

Slika 5 – uklanjanje izravnih identifikatora

Uklanjanje izravnih identifikatora je nužno, ali često ne i dovoljno za potpunu anonimizaciju podataka. Daljnje tehnike anonimizacije mogu se grupirati u [2]:

- tehnike **poopćavanja** (eng. *generalization*)
- i tehnike **nasumične izmjene podataka** (eng. *randomization*).

2.1.4.1 Tehnike poopćavanja (eng. *generalization*)

Kao što i ime kaže, tehnike poopćavanja temelje se na **poopćavanju vrijednosti** određenih atributa. Na primjeru podataka o bolničkim posjetima, **datum rođenja** pacijenta može se zamijeniti **godinom rođenja** pacijenta, kao što je prikazano na Slika 6.

Ime i prezime	Adresa	Spol	Datum rođenja	Razlog posjeta	Bolnica
[redacted]	[redacted]	Ž	5.9.1980.	Gripa	KBC Zagreb
		M	30.1.1988.	Operacija oka	KBC Split
		Ž	23.4.2007.	Lom čeljusti	KBC Zagreb
		M	17.8.1987.	Ubod ose	KBC Pula
		Ž	6.12.1995.	Trudovi	KBC Osijek
		M	12.5.1972.	Povraćanje	KBC Zadar
		Ž	9.4.1969.	Srčani udar	OB Koprivnica

Ime i prezime	Adresa	Spol	Godina rođenja	Razlog posjeta	Bolnica
[redacted]	[redacted]	Ž	1980.	Gripa	KBC Zagreb
		M	1988.	Operacija oka	KBC Split
		Ž	2007.	Lom čeljusti	KBC Zagreb
		M	1987.	Ubod ose	KBC Pula
		Ž	1995.	Trudovi	KBC Osijek
		M	1972.	Povraćanje	KBC Zadar
		Ž	1969.	Srčani udar	OB Koprivnica

Slika 6 – poopćavanje datuma rođenja u godinu rođenja

U drugim skupovima podataka, mogli bi primjerice:

- grad stanovanja (npr. Rijeka) zamijeniti regijom (npr. Primorsko-goranska županija),
- vrijednost plaće (npr. 5623 kn) zamijeniti rasponom vrijednosti (npr. 5000-6000 kn),
- i slično.

Nakon poopćavanja, istu vrijednost atributa dijeli veći broj osoba što može **znatno otežati reidentifikaciju**. U ovoj grupi tehnika, značajnu ulogu igra stupanj poopćavanja. Općenitije vrijednosti mogu pružiti dodatnu zaštitu, no za legitimne svrhe, one su manje korisne od specifičnijih vrijednosti jer pružaju manje informacija. Primjerice, kada bi se datum rođenja poopćio na stoljeće rođenja, to bi dodatno otežalo reidentifikaciju, no atribut bi tada izgubio smisao jer više ne pruža gotovo nikakvu informaciju.

2.1.4.2 Tehnike nasumične izmjene podataka (eng. *randomization*)

Tehnikama nasumične izmjene podataka **mijenja se istinitost** podataka tako da se **oslabi veza** između podataka i pojedinca, no da podaci i dalje sadrže korisne informacije [2]. Primjeri tehnika nasumične izmjene podataka su:

- tehnika dodavanja šuma (eng. *noise addition*)
- i tehnika permutacije (eng. *permutation*).

Tehnika dodavanja šuma mijenja vrijednosti nekog atributa tako da individualne vrijednosti više nisu precizne, no sveukupno, vrijednosti i dalje imaju istu distribuciju. Na primjeru prethodnih podataka o bolničkim posjetima, moguće je datumu rođenja pacijenta dodati ili oduzeti nekoliko dana kao što je prikazano na Slika 7.

Ime i prezime	Adresa	Spol	Datum rođenja	Razlog posjeta	Bolnica
[redacted]	[redacted]	Ž	5.9.1980.	Gripa	KBC Zagreb
		M	30.1.1988.	Operacija oka	KBC Split
		Ž	23.4.2007.	Lom čeljusti	KBC Zagreb
		M	17.8.1987.	Ubod ose	KBC Pula
		Ž	6.12.1995.	Trudovi	KBC Osijek
		M	12.5.1972.	Povraćanje	KBC Zadar
		Ž	9.4.1969.	Srčani udar	OB Koprivnica

Ime i prezime	Adresa	Spol	Datum rođenja	Razlog posjeta	Bolnica
[redacted]	[redacted]	Ž	11.9.1980.	Gripa	KBC Zagreb
		M	4.2.1988.	Operacija oka	KBC Split
		Ž	17.4.2007.	Lom čeljusti	KBC Zagreb
		M	19.8.1987.	Ubod ose	KBC Pula
		Ž	6.12.1995.	Trudovi	KBC Osijek
		M	11.5.1972.	Povraćanje	KBC Zadar
		Ž	6.4.1969.	Srčani udar	OB Koprivnica

Slika 7 – dodavanje šuma izmjenom datuma rođenja

Tehnika permutacije permutira vrijednosti atributa (stupca u tablici) i time **smanjuje povezanost** podataka **unutar** jednog **zapisa**, no **zadržava** sveukupnu **distribuciju** permutiranog atributa. Na primjeru podataka o posjetima bolnicama, moguće je permutirati stupac s imenom posjećene bolnice i time smanjiti vezu između bolnice i ostatka zapisa, kao što je prikazano na Slika 8.

Ime i prezime	Adresa	Spol	Datum rođenja	Razlog posjeta	Bolnica
[redacted]	[redacted]	Ž	5.9.1980.	Gripa	KBC Zagreb
		M	30.1.1988.	Operacija oka	KBC Split
		Ž	23.4.2007.	Lom čeljusti	KBC Zagreb
		M	17.8.1987.	Ubod ose	KBC Pula
		Ž	6.12.1995.	Trudovi	KBC Osijek
		M	12.5.1972.	Povraćanje	KBC Zadar
		Ž	9.4.1969.	Srčani udar	OB Koprivnica

Ime i prezime	Adresa	Spol	Datum rođenja	Razlog posjeta	Bolnica
[redacted]	[redacted]	Ž	5.9.1980.	Gripa	KBC Zadar
		M	30.1.1988.	Operacija oka	OB Koprivnica
		Ž	23.4.2007.	Lom čeljusti	KBC Pula
		M	17.8.1987.	Ubod ose	KBC Zagreb
		Ž	6.12.1995.	Trudovi	KBC Osijek
		M	12.5.1972.	Povraćanje	KBC Zagreb
		Ž	9.4.1969.	Srčani udar	KBC Split

Slika 8 – permutacija imena bolnice u skupu podataka o bolničkim posjetima

Iako se na prvi pogled čini da permutacija u potpunosti uništava vezu između atributa i ostatka zapisa, to često nije točno. Primjerice, nakon permutacije na Slika 8, u drugom zapisu piše kako je operacija oka obavljena u Općoj bolnici Koprivnica. No, pretpostavimo da je to složena operacija koju je u Hrvatskoj moguće samo obaviti u KBC-u Split. U tom slučaju, čak i nakon ovakve permutacije, moguće je zaključiti kako posjet iz drugog zapisa na slici nije obavljen u Općoj bolnici Koprivnica, već u KBC-u Split.

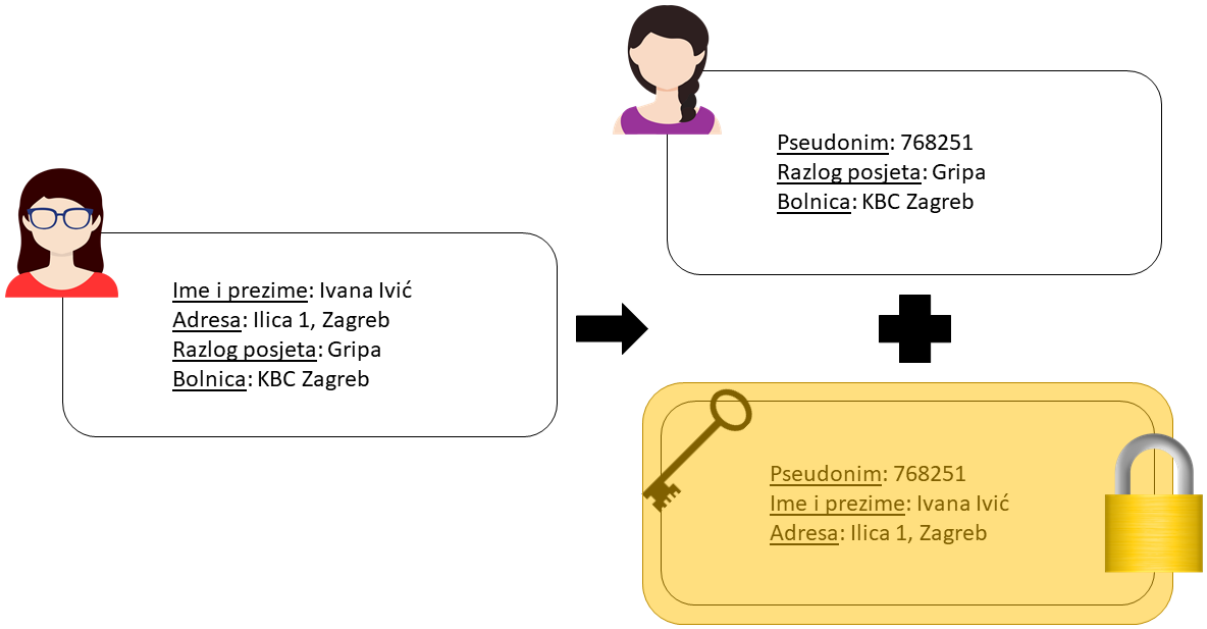
2.2 Pseudonimizacija podataka

2.2.1 Što je pseudonimizacija podataka?

Pojednostavljeno, pseudonimizacija je postupak obrade podataka u kojemu se:

- izravni identifikatori zamijene tzv. **pseudonimom,**
- dok se na **odvojenom** (i **sigurnijem**) mjestu drži tablica, ili općenitije, podaci, koji povezuju pseudonime i identifikatore.

Primjerice, pseudonimizacijom podataka o posjetima bolnicama bi ime, prezime i adresa pacijenta bili zamijenjeni nasumičnim brojem. Tako bi primjerice „Ivana Ivić, Ilica 1“ bilo zamijenjeno brojem 768251 (nasumični broj). Odvojeno od tih zapisa bila bi čuvana tablica koja povezuje pseudonime (npr. „768251“) i izravne identifikatore (npr. „Ivana Ivić, Ilica 1“). Ovaj primjer ilustriran je na Slika 9.



Slika 9 – primjer pseudonimizacije podataka na jednom zapisu

Ključno je razumjeti da pseudonimizacija nije istovjetna anonimizaciji. Nakon pseudonimizacije podataka je i dalje moguće povezati pseudonimizirane podatke s određenim pojedincem. Zato se pseudonimizirani podaci i dalje smatraju **osobnim podacima** te za njih vrijede zaštite iz OUZP-a.

2.2.2 Zašto pseudonimizirati podatke?

Za razliku od anonimizacije, nije očigledno kako pseudonimizacija pomaže u zaštiti podataka. Nakon pseudonimizacije podataka moguće je:

- držati tablicu s **identifikatorima** (odnosno ekvivalentne podatke) na **sigurnome**,
    - ili ih čak **uništiti**,

- te u redovitoj obradi koristiti **samo manje osjetljive podatke** gdje su pojedinci **pod pseudonimom**.

Tako se pseudonimizacijom minimizira rizik na dva načina:

- pošto se tablica s identifikatorima drži odvojeno (na sigurnome) ili je čak uništena, **manja je vjerojatnost** da će ona biti **kompromitirana**,
- a ako tablica gdje su pojedinci pod pseudonimom i bude kompromitirana, **šteta je znatno manja** od alternative u kojoj su kompromitirani identifikatori, a ne samo pseudonimi.

Pseudonimizacija je u OUZP-u izričito navedena kao korisna mjera zaštite koja pomaže u poštivanju uredbe na više načina. Tvrtka *Privacy Analytics* izradila je dokument u kojemu je, između ostalog, detaljno analizirano na koji način pseudonimizacija olakšava poštivanje OUZP-a [5]. Ukratko, načini na koje pseudonimizacija pomaže u poštivanju OUZP-a mogu se svrstati u tri kategorije:

1. Pseudonimizacija i **zaštita podataka**
2. Pseudonimizacija i **prava građana**
3. Pseudonimizacija i **zakonitost obrade podataka**

Po pitanju **zaštite podataka**, OUZP:

- zahtijeva korištenje **mjera sigurnosti** obrade podataka,
- uvodi obvezu tzv. **tehničke zaštite podataka** (eng. *data protection by design*)
- te uvodi obvezu **izvještavanja** nadzornog tijela i žrtava o povredama sigurnosti osobnih podataka **ovisno o riziku** kojega povreda predstavlja.

Pseudonimizacija je u OUZP-u **izričito navedena** (npr. Članak 25.(1), Članak 32.(1)(a)) kao mjera koja pomaže i u osiguravanju obrade podataka i u postizanju tehničke zaštite podataka [3]. Uz to, pseudonimizacija može **smanjiti rizik** povrede sigurnosti osobnih podataka i time potencijalno **ukloniti potrebu izvještavanja** nadzornog tijela i žrtava [5].

OUZP uvodi niz **prava** za **građane**, primjerice:

- pravo na pristup osobnim podacima,
- pravo na ispravak osobnih podataka,
- pravo na brisanje osobnih podataka („pravo na zaborav”).

No OUZP također navodi da (Članak 12(2)) [3] [5]:

- ako voditelj obrade **dokaže da nije u mogućnosti utvrditi identitet** osobe iz podataka koje posjeduje
- onda joj **ne mora**, tj. ne može, **pružiti** navedena **prava**.

Što se tiče zakonitosti obrade podataka, OUZP u pravilu zahtijeva **izričitu privolu** od osobe za obradu njenih osobnih podataka u **određenu svrhu**. No u nekim slučajevima, dopuštena je i obrada tih osobnih podataka **u drugu svrhu** ako je ta svrha u skladu sa svrhom u koju su osobni podaci prvotno prikupljeni. Kako bi se utvrdilo je li druga svrha u skladu s izvornom svrhom, OUZP izričito navodi da se **uzima u obzir** i „*postojanje odgovarajućih zaštitnih mjera, koje mogu uključivati enkripciju ili pseudonimizaciju*“.

2.2.3 Tehnike pseudonimizacije

Tehnike pseudonimizacije mogu se podijeliti, ovisno o tome kako generiraju pseudonime, na:

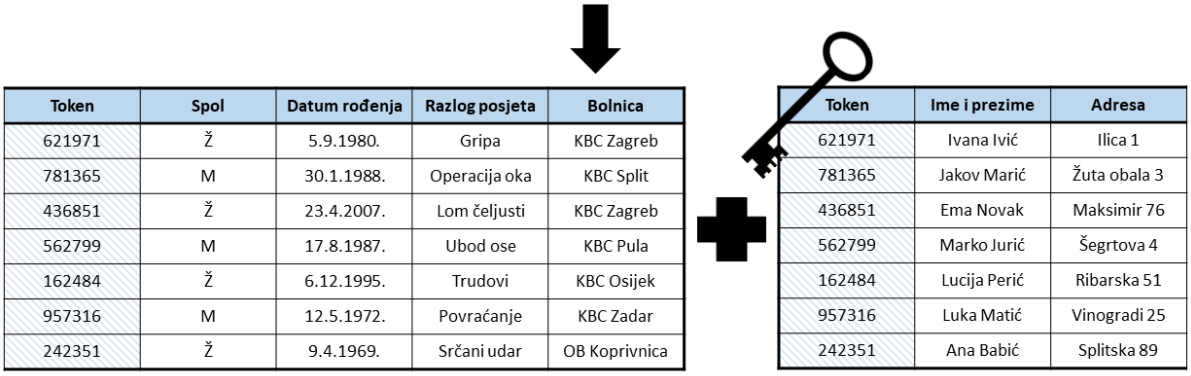
- tehnike u kojima je pseudonim **neovisan** o izvornim podacima
  - npr. tokenizacija
- tehnike u kojima se pseudonim **generira iz izvornih podataka**
  - npr. pseudonimizacija šifriranjem
  - ili pseudonimizacija funkcijom sažimanja (eng. *hash function*)

2.2.3.1 Tokenizacija

Tokenizacija je tehnika pseudonimizacije u kojoj se pseudonim, u ovoj tehnici zvan **token**, generira neovisno o izvornim podacima. Primjerice, pseudonimi mogu biti nasumično generirani brojevi. Tako bi „Ivana Ivić na adresi Ilica 1“ moglo biti zamijenjeno tokenom (pseudonimom) 621971. Primjer pseudonimizacije tokenizacijom ilustriran je na Slika 10.

Kada se podaci ispravno pseudonimiziraju tokenizacijom, **nije moguće** iz pseudonima doći do izravnih identifikatora bez korištenja tablice u kojoj su zapisane njihove veze.

Ime i prezime	Adresa	Spol	Datum rođenja	Razlog posjeta	Bolnica
Ivana Ivić	Ilica 1	Ž	5.9.1980.	Gripa	KBC Zagreb
Jakov Marić	Žuta obala 3	M	30.1.1988.	Operacija oka	KBC Split
Ema Novak	Maksimir 76	Ž	23.4.2007.	Lom čeljusti	KBC Zagreb
Marko Jurić	Šegrtova 4	M	17.8.1987.	Ubod ose	KBC Pula
Lucija Perić	Ribarska 51	Ž	6.12.1995.	Trudovi	KBC Osijek
Luka Matić	Vinogradi 25	M	12.5.1972.	Povraćanje	KBC Zadar
Ana Babić	Splitska 89	Ž	9.4.1969.	Srčani udar	OB Koprivnica



Token	Spol	Datum rođenja	Razlog posjeta	Bolnica
621971	Ž	5.9.1980.	Gripa	KBC Zagreb
781365	M	30.1.1988.	Operacija oka	KBC Split
436851	Ž	23.4.2007.	Lom čeljusti	KBC Zagreb
562799	M	17.8.1987.	Ubod ose	KBC Pula
162484	Ž	6.12.1995.	Trudovi	KBC Osijek
957316	M	12.5.1972.	Povraćanje	KBC Zadar
242351	Ž	9.4.1969.	Srčani udar	OB Koprivnica

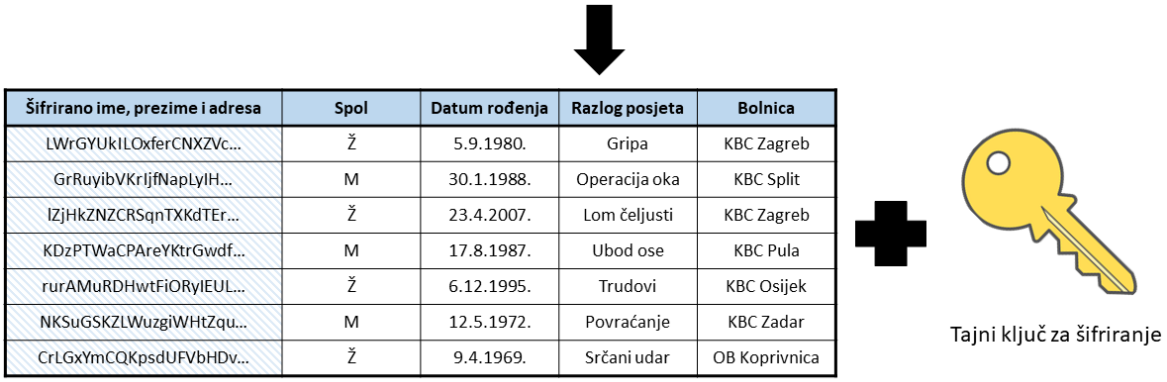
Token	Ime i prezime	Adresa
621971	Ivana Ivić	Ilica 1
781365	Jakov Marić	Žuta obala 3
436851	Ema Novak	Maksimir 76
562799	Marko Jurić	Šegrtova 4
162484	Lucija Perić	Ribarska 51
957316	Luka Matić	Vinogradi 25
242351	Ana Babić	Splitska 89

Slika 10 – primjer pseudonimizacije tokenizacijom

2.2.3.2 Pseudonimizacija šifriranjem

Pseudonimizacija šifriranjem je tehnika pseudonimizacije u kojoj se identifikatori **šifriraju** tajnim ključem te ta **šifrirana vrijednost** zapravo postaje **pseudonim**. U ovoj tehnici, „Ivana Ivić na adresi Ilica 1“ bilo bi šifrirano te bi se šifrirana vrijednost „LWrGYUkILOxfe…“ koristila kao pseudonim. Primjer pseudonimizacije šifriranjem prikazan je na Slika 11.

Ime i prezime	Adresa	Spol	Datum rođenja	Razlog posjeta	Bolnica
Ivana Ivić	Ilica 1	Ž	5.9.1980.	Gripa	KBC Zagreb
Jakov Marić	Žuta obala 3	M	30.1.1988.	Operacija oka	KBC Split
Ema Novak	Maksimir 76	Ž	23.4.2007.	Lom čeljusti	KBC Zagreb
Marko Jurić	Šegrtova 4	M	17.8.1987.	Ubod ose	KBC Pula
Lucija Perić	Ribarska 51	Ž	6.12.1995.	Trudovi	KBC Osijek
Luka Matić	Vinogradi 25	M	12.5.1972.	Povraćanje	KBC Zadar
Ana Babić	Splitska 89	Ž	9.4.1969.	Srčani udar	OB Koprivnica



Šifrirano ime, prezime i adresa	Spol	Datum rođenja	Razlog posjeta	Bolnica
LWrGYUkILOxferCNXZVc...	Ž	5.9.1980.	Gripa	KBC Zagreb
GrRuyibVKrIjfNapLyIH...	M	30.1.1988.	Operacija oka	KBC Split
lZjHkZNZCRSqnTXKdTEr...	Ž	23.4.2007.	Lom čeljusti	KBC Zagreb
KDzPTWaCPAreYKtrGwdf...	M	17.8.1987.	Ubod ose	KBC Pula
rurAMuRDHwtFiORyIEUL...	Ž	6.12.1995.	Trudovi	KBC Osijek
NKSuGSKZLWuzgiWHtZqu...	M	12.5.1972.	Povraćanje	KBC Zadar
CrLGxYmCQKpsdUFVbHDv...	Ž	9.4.1969.	Srčani udar	OB Koprivnica

Slika 11 – primjer pseudonimizacije šifriranjem

U ovoj tehnici **ne postoji odvojena tablica** koja povezuje pseudonime s identifikatorima, već se odvojeno drži samo **tajni ključ** kojim su identifikatori šifrirani. Ako se koristi siguran algoritam šifriranja i tajni ključ, onda **nije moguće** iz pseudonima doći do izravnih identifikatora bez znanja tajnog ključa.

2.2.3.3 Pseudonimizacija funkcijom sažimanja (eng. *hash function*)

U pseudonimizaciji funkcijom sažimanja (eng. *hash function*) identifikatori se obrađuju funkcijom sažimanja te se **izlaz funkcije** koristi kao **pseudonim**. U ovoj tehnici, „Ivana Ivić na adresi Ilica 1“ bilo bi obrađeno funkcijom sažimanja te bi se izlazna vrijednost „nmrTqvciBbkfb…“ koristila kao pseudonim. Primjer pseudonimizacije funkcijom sažimanja prikazan je na Slika 12.

Ime i prezime	Adresa	Spol	Datum rođenja	Razlog posjeta	Bolnica
Ivana Ivić	Ilica 1	Ž	5.9.1980.	Gripa	KBC Zagreb
Jakov Marić	Žuta obala 3	M	30.1.1988.	Operacija oka	KBC Split
Ema Novak	Maksimir 76	Ž	23.4.2007.	Lom čeljusti	KBC Zagreb
Marko Jurić	Šegrtova 4	M	17.8.1987.	Ubod ose	KBC Pula
Lucija Perić	Ribarska 51	Ž	6.12.1995.	Trudovi	KBC Osijek
Luka Matić	Vinogradi 25	M	12.5.1972.	Povraćanje	KBC Zadar
Ana Babić	Splitska 89	Ž	9.4.1969.	Srčani udar	OB Koprivnica

Ime, prezime i adresa obrađeni funkcijom sažimanja	Spol	Datum rođenja	Razlog posjeta	Bolnica
nmrTqvciBbkfbDrORxqyDVcHlLRFmTIFGV...	Ž	5.9.1980.	Gripa	KBC Zagreb
ovHSYCZjgCyMmriqvlfVlnVSujsLTtwkucxj...	M	30.1.1988.	Operacija oka	KBC Split
GKdovDfCnNhxKjzgpWjshzDnFbykJvmARM...	Ž	23.4.2007.	Lom čeljusti	KBC Zagreb
mSkAmSQzLPbmalbltvNCngSNskxeVOnsBa...	M	17.8.1987.	Ubod ose	KBC Pula
zrTxpiSjrAhqxzjbOTLVuCEnMgViDlOrNkJm...	Ž	6.12.1995.	Trudovi	KBC Osijek
POqWEZXtEdwrBHjsimXcnBjEfXiAdhGqgoQ...	M	12.5.1972.	Povraćanje	KBC Zadar
SbawIPYYFQjSfnCxCDCcEGLzeCbrZxrxVZfZ...	Ž	9.4.1969.	Srčani udar	OB Koprivnica

Slika 12 – primjer pseudonimizacije funkcijom sažimanja

U pseudonimizaciji funkcijom sažimanja se u pravilu **ne koriste nikakvi odvojeni podaci** za povezivanje pseudonima i identifikatora, kao što su prethodno bili tablica odnosno tajni ključ. U ovoj tehnici je za povezivanje identifikatora s pseudonimom potrebno prvo znati identifikatore te zatim ponovno iz njih generirati pseudonim.

Ideja pseudonimizacije funkcijom sažimanja je da, bez prethodnog znanja identifikatora, iz pseudonima **ne bude moguće** doći do identifikatora. No, u praksi, to često **ne vrijedi** – uz poznatu funkciju sažimanja, napadači mogu uzastopnim pogađanjem (eng. *brute-force*) i sličnim tehnikama doći do identifikatora iza nekih od pseudonima. Zato su osmišljene razne varijante pseudonimizacije funkcijom sažimanja koje imaju snažnija svojstva zaštite od reidentifikacije. Te varijante uključuju:

1. Pseudonimizaciju funkcijom sažimanja **s dodanom vrijednosti** (eng. *salted hash function*)

   U ovoj se varijanti prilikom generiranja svakog pseudonima ulazu funkcije sažimanja **dodaje nasumična vrijednost**, tzv. „sol“ (eng. *salt*). Dodana nasumična vrijednost čini postupak generiranja pseudonima **jedinstvenim** za svaki zapis. To u konačnici pruža **snažniju zaštitu** od korištenja funkcije sažimanja bez dodane vrijednosti.

2. Pseudonimizaciju funkcijom sažimanja **s tajnim ključem** (eng. *keyed hash function*)

   U ovom slučaju se koristi posebna funkcija sažimanja koja kao parametar prima i **tajni ključ**. Korištenjem takve funkcije, napadač **bez znanja tajnog ključa ne**

može nikako **povezati** pseudonime i identifikatore. No i tu postoji više varijanta:

a) varijanta uz **pohranu** tajnog ključa

U ovoj se varijanti ključ pohranjuje na odvojeno, sigurno mjesto te ova varijanta ima slična svojstva zaštite kao i pseudonimizacija šifriranjem.

b) varijanta uz **uništavanje** tajnog ključa

U ovoj varijanti tajni ključ se uništava čime se i **uništava** bilo kakva **veza** pseudonima i identifikatora. Upravo zato, u ovoj varijanti su pseudonimi slični tokenima, no sada **nema** tablice niti bilo kojeg drugog **načina povezivanja** pseudonima i identifikatora.

3 Zaključak

U današnjem svijetu, povrede sigurnosti osobnih podataka sve su češće. Anonimizacija i pseudonimizacija mjere su zaštite podataka koje prilikom takvih povreda mogu **značajno smanjiti štetu**. Upravo zato, one su izričito navedene u OUZP-u kao mjere koje olakšavaju zadovoljavanje raznih obveza iz uredbe.

Intuitivno, lako je pomisliti kako je provođenje **ispravne** anonimizacije i pseudonimizacije prilično jednostavno. No u stvarnosti, to je izrazito **težak problem**. Tehnike anonimizacije i pseudonimizacije aktivno su područje istraživanja te se redovito dolazi do novih otkrića koja **otkrivaju mane postojećih tehnika**. Podaci koje danas smatramo anonimiziranima možda će u budućnosti, pomoću novih tehnika, **biti moguće povezati s pojedincima**.

I dalje nije u potpunosti jasno postižu li današnje tehnike anonimizacije i pseudonimizacije očekivani rezultat te kako će se one nositi s budućim otkrićima. No u konačnici, neosporno je da one otežavaju reidentifikaciju i time smanjuju potencijalnu štetu prilikom povrede osobnih podataka, te je zato i njihova korist **prepoznata u regulativi**.

4 Literatura

[1] HIPAA Jurnal, »Healthcare Data Breach Statistics,« 2018. [Mrežno]. Available: https://www.hipaajournal.com/healthcare-data-breach-statistics/.

[2] Članak 29. Radna skupina za zaštitu podataka, »Mišljenje 05/2014 o tehnikama anonimizacije,« 10. travnja 2014. [Mrežno]. Available: http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_hr.pdf.

[3] Službeni list Europske unije, »Uredba (EU) 2016/679 Europskog parlamenta i Vijeća,« 4. svibnja 2016. [Mrežno]. Available: https://eur-lex.europa.eu/legal-content/HR/TXT/HTML/?uri=CELEX:32016R0679.

[4] L. Sweeney, »Simple Demographics Often Identify People Uniquely,« 2000. [Mrežno]. Available: https://dataprivacylab.org/projects/identifiability/paper1.pdf.

[5] M. Hintze i K. El Emam, »Comparing the Benefits of Pseudonymization and Anonymization Under the GDPR,« 17. kolovoza 2017. [Mrežno]. Available: https://iapp.org/media/pdf/resource_center/PA_WP2-Anonymous-pseudonymous-comparison.pdf.

[6] M. Rouse, »Data Breach,« TechTarget, prosinac 2017. [Mrežno]. Available: https://searchsecurity.techtarget.com/definition/data-breach.

[7] Europska komisija, »Što su osobni podaci?,« [Mrežno]. Available: https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_hr.

[8] Europska komisija, »Reforma EU-a o zaštiti podataka: bolja prava na zaštitu podataka za građane Europe,« svibanj 2016. [Mrežno]. Available: https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-citizens_hr.pdf.

[9] Europska komisija, »Reforma EU-a o zaštiti podataka: bolji propisi za europske tvrtke,« European Commission, svibanj 2018. [Mrežno]. Available: https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-business_hr.pdf.

[10] N. Lord, »The History of Data Breaches,« Data Insider, 6. travnja 2016. [Mrežno]. Available: https://digitalguardian.com/blog/history-data-breaches.