Anonimizacija zapisa u digitalnim arhivima

Lužar, Andreja

SVEUČILIŠTE U ZAGREBU
FILOZOFSKI FAKULTET
ODSJEK ZA INFORMACIJSKE I KOMUNIKACIJSKE ZNANOSTI
SMJER ARHIVISTIKA
Ak. god. 2022./2023.

Andreja Lužar

Anonimizacija zapisa u digitalnim arhivima

Diplomski rad

Mentor: dr. sc. Hrvoje Stančić, red. prof.

Zagreb, rujan 2023.

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.

Sadržaj

1. Uvod

U današnjem digitalnom dobu, zaštita informacija i njihova povjerljivost postali su ključni aspekti u održavanju sigurnosti i integriteta različitih sustava i organizacija. S obzirom na sveprisutnost tehnologije i mogućnost brze razmjene informacija, potreba za osiguravanjem da određene informacije ne dođu u ruke neovlaštenim pojedincima postala je izuzetno važna. Jedan od najranijih primjera zaštite informacija i povjerljivosti je Hipokratova zakletva, u kojoj su se, između ostalog, liječnici zaklinjali na čuvanje povjerljivosti informacija o pacijentima. Tijekom godina se mijenjala, tako da se danas polaže njena suvremena inačica, Ženevska deklaracija, koja i dalje u sebi sadrži element liječničke tajne, odnosno čuvanja tajnosti podataka o pacijentu[1].

Slično tome, u religioznim kontekstima kao mehanizam za čuvanje tajni i povjerljivih informacija služi ispovjedna tajna. Ova tajna omogućuje pojedincima da slobodno dijele svoje najintimnije misli i osjećaje, ali i priznanja bez straha od osude ili izdaje. U pitanju je profesionalna tajna čije kršenje dovodi do posljedica i kazni. U poslovnom svijetu, osim profesionalne tajne, organizacije koriste složene sustave zaštite kako bi osigurale da osjetljive informacije o proizvodima, strategijama i klijentima ostanu zaštićene od neovlaštenog pristupa.

Posljednjih godina bilježi se ogroman porast količine podataka, pa tako i osobnih podataka koji su u digitalnoj sferi dostupni za prikupljanje i analizu. U svijetu u kojemu su pitanja zaštite privatnosti i osobnih podataka puno problematičnija zbog lakše dostupnosti informacija nego iz vremena prije interneta, sve je teže kontrolirati koje su to informacije kojima korisnici imaju pristup. Jedan od većih problema je i što korisnici na različitim web mjestima nesvjesno i ne čitajući Pravila privatnosti i Pravila o korištenju kolačića tih web mjesta pristaju na kolačiće (engl. *cookies*)[2], ne razmišljajući kome i kojim informacijama tim postupkom daju pristup te što točno spada pod podatke koji se obrađuju i u koje svrhe će se oni kasnije koristiti. Iako je cijeli proces prikupljanja i obrade osobnih podataka korištenjem kolačića propisan u zakonodavnom okviru direktivama Europske unije i hrvatskim zakonima, može se postaviti pitanje je li privola za obradu osobnih podataka dana svjesno ili samo zato da korisnik što prije može pristupiti internetskoj stranici koju je otvorio na svom uređaju.

[1] Hrvatska liječnička komora. „Ženevska deklaracija“. Dostupno na: https://www.hlk.hr/EasyEdit/UserFiles/pdf-ovi-za-vijesti-web/2020/lijecnicka-prisega-preambula.pdf

[2] „mrežni kolačić (engl. *cookie*) je mala tekstovna datoteka s informacijama o aktivnostima i postavkama korisnika, koja se na njegov uređaj (računalo, molitel i sl.) pohranjuje prigodom posjeta pojedinoj mrežnoj stranici“ (Hrvatska enciklopedija, mrežno izdanje, LZMK)

Prihvaćanjem kolačića daje se pristanak za analizu podataka koji su spremljeni na osobnom računalu.

Veliki problem koji nastaje kad se osobni podaci koriste u različitim analizama je mogućnost kršenja privatnosti. Osobni podaci često sadrže osjetljive informacije kao što su imena, adrese, identifikacijski brojevi, telefonski brojevi, financijski i zdravstveni podaci, koji se, u slučaju ako dospiju u pogrešne ruke, mogu koristiti za identifikaciju pojedinca i uzrokovanje štete kao što su krađa podataka ili krađa identiteta. Suvremeni svijet donosi probleme novih razmjera pa se tako krađa identiteta pojavljuje i u digitalnom okruženju tako što internetske stranice prikupljaju podatke pa pojedinci koji nisu dovoljno oprezni ili educirani o opasnostima na internetu objavljuju svoje podatke, ponajviše na društvenim mrežama, gdje je moguće otvoriti lažne profile koji se koriste u različite svrhe. Neke od tih svrha mogu biti i uništavanje ugleda pojedinca u čije se ime profili otvaraju. Otvaranje lažnih profila smatra se oblikom *cyberbullyinga* ili virtualnog zlostavljanja. S druge strane, određene su skupine podložne *phishingu* putem elektroničke pošte, odnosno protuzakonitom pokušaju prijevare i prikupljanju povjerljivih osobnih podataka (podataka s kreditnih kartica, lozinki i sl.) u svrhu financijskih prijevara. Kako bi se spriječile opasnosti do kojih dolazi zbog kršenja privatnosti, važno je poduzeti korake za zaštitu osobnih podataka, naročito u slučajevima kada se ti podaci koriste u različitim analizama. Osim što Opća uredba o zaštiti podataka (General Data Protection Regulation, GDPR)[3] Europske unije propisuje da pojedinci moraju biti obaviješteni o obradi svojih osobnih podataka u svrhu analize, postoje načini na koje je moguće spriječiti neovlašteni pristup podacima i zaštititi privatnost pojedinaca čiji se podaci koriste. Jedan od načina je implementacija tehnika za očuvanje privatnosti kao što je anonimizacija, o čemu će biti riječ u ovome radu. Nakon pregleda pravnog okvira zaštite informacija i podataka u Republici Hrvatskoj slijedi teorijski pregled anonimizacije, a nakon njega i praktični pregled programa koji se mogu koristiti za anonimizaciju. Na kraju rada daje se pregled provođenja anonimizacije u hrvatskim državnim arhivima.

[3] Opća uredba o zaštiti podataka. https://eur-lex.europa.eu/legal-content/HR/TXT/?uri=celex%3A32016R0679

2. Zaštita informacija i podataka

Pravni okvir zaštite osobnih podataka važan je kako bi se sačuvala i zaštitila privatnost pojedinaca. On uređuje prikupljanje, obradu, pohranu i upravljanje osobnim podacima, a zakone koji ga sačinjavaju donose države kako bi osigurale da institucije koje rukuju osobnim podacima poštuju privatnost pojedinaca čije podatke obrađuju te da se tako spriječi zloupotreba osobnih podataka.

U članku 8. Povelje Europske unije[4] o temeljnima pravima navodi se sljedeće:

1. „Svatko ima pravo na zaštitu osobnih podataka koji se na njega ili nju odnose.
2. Takvi podaci moraju se obrađivati pošteno, u utvrđene svrhe i na temelju suglasnosti osobe o kojoj je riječ, ili na nekoj drugoj legitimnoj osnovi utvrđenoj zakonom. Svatko ima pravo na pristup prikupljenim podacima koji se na njega ili nju odnose i pravo na njihovo ispravljanje.
3. Poštovanje tih pravila podliježe nadzoru neovisnog tijela.“

S obzirom na to da ista ta Povelja navodi i da se pravo pristupa dokumentima institucija, tijela, ureda i agencija Unije osigurava svim građanima Unije, fizičkim ili pravnim osobama koje imaju boravište ili sjedište u bilo kojoj državi članici (čl. 42), važno je osigurati da su osobni podaci koji su sadržani u tim dokumentima zaštićeni na odgovarajući način. Zbog toga je na razini Europske unije donesena Uredba (EU) 2016/679 Europskog parlamenta i Vijeća od 27. travnja 2016. o zaštiti pojedinaca u vezi s obradom osobnih podataka i o slobodnom kretanju takvih podataka te o stavljanju izvan snage Direktive 95/46/EZ (Opća uredba o zaštiti podataka, *General Data Protection Regulation*, GDPR) koja se primjenjuje u svim državama članicama Europske unije od 25. svibnja 2018. godine. GDPR predstavlja značajan korak prema većoj zaštiti osobnih podataka građana Europske unije. Osim ove uredbe, u Republici Hrvatskoj se reguliranje sigurnosti osobnih podataka osigurava kroz još nekoliko zakona.

[4] Europski parlament, Vijeće Europske unije, Europska komisija. „Povelja Europske unije o temeljnim pravima“. *Službeni list Europske unije* C 202/389 (2016). Dostupno na: https://eur-lex.europa.eu/legal-content/HR/TXT/PDF/?uri=CELEX:12016P/TXT&from=RO

2.1. Obrada osobnih podataka

U članku 37. Ustava Republike Hrvatske[5] stoji da se svakoj osobi „jamči sigurnost i tajnost osobnih podataka“ te da se „bez privole ispitanika, osobni podaci mogu prikupljati, obrađivati i koristiti samo uz uvjete određene zakonom“. Čak i tako prikupljeni podaci smiju se upotrijebiti samo za svrhu koja je utvrđena prilikom njihovog prikupljanja.[6] Nekoliko različitih propisa bavi se zaštitom pojedinaca u sklopu obrade osobnih podataka. Najznačajniji od njih je već spomenuti GDPR koji je zamijenio Direktivu 95/46/EZ Europskog parlamenta i vijeća od 24. listopada 1995. o zaštiti pojedinaca u vezi s obradom osobnih podataka i o slobodnom protoku takvih podataka[7]. Između izdavanja ove dvije važne uredbe o zaštiti podataka prošlo je više od dva desetljeća. Tijekom tog vremena, društvo se značajno promijenio, posebno u pogledu tehnološkog napretka, komunikacije i razvoja interneta. Razvoj pametnih telefona, društvenih medija, interneta, umjetne inteligencije i napredne analitike podataka temeljito je promijenio način na koji ljudi i institucije komuniciraju, dijele informacije i posluju. Digitalna komunikacija postala je središnji dio društvenog života. Elektronička pošta, društvene mreže, digitalne usluge i online trgovine postale su neizostavan dio svakodnevnog života. Svijet je postao „globalno selo“, izraz koji je prvi upotrijebio Marshall McLuhan,[8] jer je danas putem interneta moguće brzo i jednostavno stupiti u kontakt s ljudima iz cijelog svijeta iz udobnosti vlastitog doma. Ovaj tehnološki napredak omogućio je prikupljanje i analizu ogromnih količina podataka, što je postalo ključno između ostalog za razvijanje poslovnih i marketinških strategija. Istovremeno je doveo do potrebe za usklađivanjem i koordinacijom međunarodnih zakona za zaštitu podataka.

Već su ranije postojali propisi koji su služili zaštiti privatnosti i obrade osobnih podataka, pa je tako, recimo, Direktiva 2002/58/EZ Europskog parlamenta i vijeća od 12. srpnja 2002. o obradi osobnih podataka i zaštiti privatnosti u području elektroničkog komuniciranja

[5] Hrvatski sabor. „Ustav Republike Hrvatske“. *Narodne novine* 56/90, 135/97, 8/98, 113/00, 124/00, 28/01, 41/01, 55/01, 76/10, 76/10, 85/10, 05/14 (2014). Dostupno na: https://www.zakon.hr/z/94/Ustav-Republike-Hrvatske

[6] Ustav Republike Hrvatske, n.dj., čl. 37

[7] „Direktiva 95/46/EZ Europskog parlamenta i vijeća od 24. lipnja 1995. o zaštiti pojedinaca u vezi s obradom osobnih podataka i o slobodnom protoku takvih podataka“. *Službeni list Europskih zajednica* L 281/31 (1995). Dostupno na: https://eur-lex.europa.eu/legal-content/HR/TXT/PDF/?uri=CELEX:31995L0046&from=HU

[8] Hrvatska enciklopedija, mrežno izdanje. „McLuhan, Herbert Marshall“. *Leksikografski zavod Miroslav Krleža* (2021). Dostupno na: https://enciklopedija.hr/natuknica.aspx?id=39694

(Direktiva o privatnosti i elektroničkim komunikacijama) implementirana u hrvatski Zakon o elektroničkim komunikacijama.[9]

Direktiva, pa tako i Zakon o elektroničkim komunikacijama u svojoj srži stavljaju naglasak na privolu kao važan aspekt zaštite privatnosti i obrade osobnih podataka u području elektroničkih komunikacija. Zakon o elektroničkim komunikacijama definira privolu kao „slobodno dano i dostavljeno izričito očitovanje volje korisnika usluga ili pretplatnika kojim on izražava svoju suglasnost s obradom njegovih osobnih podataka u određene svrhe“[10].

Izraz *privola* često se koristi u kontekstu zaštite osobnih podataka i općenito se odnosi na izričito davanje odobrenja ili suglasnosti osoba (korisnika usluga ili pretplatnika) za obradu njegovih osobnih podataka u određene svrhe. Obrada osobnih podataka podrazumijeva bilo kakvu aktivnost koja se provodi s osobnim podacima, kao što su prikupljanje, pohrana, analiza, korištenje i dijeljenje tih podataka. Privola je ključan koncept u okviru regulacije zaštite podataka, poput Opće uredbe o zaštiti podataka na području Europske unije ili Zakona o elektroničkim komunikacijama na području Republike Hrvatske jer naglašava važnost informiranosti i slobode pojedinca u određivanju načina na koji se njegovi osobni podaci obrađuju. Tako Opća uredba o zaštiti podataka jasno definira *privolu* kao „svako dobrovoljno, posebno, informirano i nedvosmisleno izražavanje želja ispitanika kojim on izjavom ili jasnom potvrdnom radnjom daje pristanak za obradu osobnih podataka koji se na njega odnose“[11]. Ova suglasnost temelji se na dobrovoljnosti, informiranosti i nedvosmislenosti izraženih želja ispitanika.

Već Direktiva o privatnosti i elektroničkim komunikacijama govori o anonimizaciji kao mehanizmu zaštite osobnih podataka kada podaci više nisu potrebni za prijenos komunikacije ili kad se korisnik ne složi s obradom podataka. Podaci o prometu (koji se odnose na komunikacijske aktivnosti) koji su pohranjeni od strane davatelja usluge trebaju biti obrisani ili anonimizirani kad više nisu potrebni za prijenos komunikacije. To znači da se ti podaci moraju ili potpuno ukloniti iz sustava ili anonimizirati kad prestane potreba za njihovom svrhom. Tako se i podaci o lokaciji (koji nisu podaci o prometu) mogu obraditi samo ako su

[9] Hrvatski sabor. „Zakon o elektroničkim komunikacijama“. *Narodne novine* 76/22(2022). Dostupno na: https://narodne-novine.nn.hr/clanci/sluzbeni/2022_07_76_1116.html

[10] Hrvatski sabor. „Zakon o elektroničkim komunikacijama“. *Narodne novine* 73/08 (2008). Dostupno na: https://narodne-novine.nn.hr/clanci/sluzbeni/2008_06_73_2420.html

[11] Europski parlament, Vijeće Europske unije. „Uredba (EU) 2016/679 Europskog parlamenta i Vijeća od 27. travnja 2016. o zaštiti pojedinaca u vezi s obradom osobnih podataka i o slobodnom kretanju takvih podataka te o stavljanju izvan snage Direktive 95/46/EZ (Opća uredba o zaštiti podataka)“. *Službeni list Europske unije* L 199/1 (2016). Dostupno na: https://eur-lex.europa.eu/legal-content/HR/TXT/?uri=celex%3A32016R0679

prvo anonimizirani ili ako je korisnik dao svoj pristanak, što znači da se anonimizacija mora primijeniti prije obrade tih podataka ili se obrada mora temeljiti na pristanku korisnika.[12]

Zakon o elektroničkim komunikacijama[13] također definira i pojam *povrede osobnih podataka* kao „povredu sigurnosti koja uzrokuje slučajno ili nezakonito uništenje, gubitak, izmjenu, neovlašteno razotkrivanje ili pristup osobnim podacima što se prenose, pohranjuju ili na drugi način obrađuju u vezi s pružanjem javno dostupnih elektroničkih komunikacijskih usluga u Europskoj uniji". Povreda može biti posljedica slučajnih događaja ili nezakonitih aktivnosti te može dovesti do negativnih posljedica i ozbiljno ugroziti privatnost korisnika. U slučaju da dođe do povrede osobnih podataka, operator javno dostupnih elektroničkih komunikacijskih usluga dužan je odmah obavijestiti tijelo nadležno za zaštitu osobnih podataka i korisnika da je došlo do povrede, ako postoji mogućnost da će povreda utjecati štetno na privatnost korisnika ili osobne podatke. Ako se zaključi da je operator na vrijeme primijenio odgovarajuće tehnološke mjere zaštite koje osiguravaju nerazumljivost podataka, korisnika se ne treba obavijestiti.[14]

Dakle, korištenje elektroničkih komunikacijskih mreža za pohranu ili pristup veći pohranjenim podacima u uređajima korisnika dopušteno je samo ako korisnik da svoj pristanak nakon što je jasno i u potpunosti obaviješten o svrsi obrade podataka u skladu s pravilima o zaštiti osobnih podataka. Ovo pravilo ne sprječava tehničku pohranu ili pristup podacima koji su nužni za prijenos komunikacije putem elektroničke komunikacijske mreže.[15]

2.1.1. Opća uredba o zaštiti podataka

U dva desetljeća, koliko je prošlo od izlaska Direktive o zaštiti pojedinaca u vezi s obradom osobnih podataka i o slobodnom protoku takvih podataka do donošenja Opće uredbe o zaštiti podataka, promjene u tehnologiji, društvu i komunikacijama dovele su do potrebe za modernizacijom zakonodavstva o zaštiti podataka kako bi se što bolje odgovorilo na izazove povezane s privatnošću, sigurnošću i upravljanjem osobnim podacima. Opća uredba o zaštiti

[12] Europski parlament, Vijeće Europske unije. „Direktiva 2002/58/EZ Europskog parlamenta i Vijeća od 12. srpnja 2022. o obradi osobnih podataka i zaštiti privatnosti u području elektroničkih komunikacija (Direktiva o privatnosti i elektroničkim komunikacijama)". *Službeni list Europske unije* L 201/37 (2002). Dostupno na: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX%3A32002L0058%3Ahr%3AHTML
[13] Zakon o elektroničkim komunikacijama NN 76/22, n.dj., čl. 5.
[14] Isto, čl. 42.
[15] Isto, čl. 43.

podataka donesena je kako bi se uskladile tehnološke i društvene promjene s regulativama, postavljajući čvrste temelje za zaštitu podataka u digitalnom dobu.

Ovom Uredbom uspostavljena su pravila koja se odnose na zaštitu privatnosti pojedinaca i slobodno kretanje njihovih osobnih podataka unutar Europske unije s ciljem osiguranja zaštite temeljnih prava i sloboda pojedinaca, uključujući prava na zaštitu osobnih podataka, bez ograničavanja slobodnog kretanja tih podataka unutar Europske unije.[16] Odnosi se na automatiziranu i neautomatiziranu obradu osobnih podataka koji su dio sustava pohrane ili su namijenjeni biti njegovim dijelom. Ne uključuje obradu osobnih podataka:

a) „tijekom djelatnosti koja nije obuhvaćena opsegom prava Unije;
b) koju obavljaju države članice kada obavljaju aktivnosti koje su obuhvaćene područjem primjene glave V. poglavlja 2. UEU-a;
c) koju provodi fizička osoba tijekom isključivo osobnih ili kućnih aktivnosti;
d) koju obavljaju nadležna tijela u svrhu sprečavanja, istrage, otkrivanja ili progona kaznenih djela ili izvršavanja kaznenih sankcija, uključujući zaštitu od prijetnji javnoj sigurnosti i njihova sprečavanja.“[17]

Opća uredba o zaštiti podataka usmjerila se na obradu osobnih podataka, pa odmah na početku definira *osobne podatke* kao

> „sve podatke 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“,

a *obradu* kao

> „svaki postupak ili skup postupaka koji se obavljaju na osobnim podacima ili na skupovima osobnih podataka, bilo automatiziranim bilo neautomatiziranim sredstvima kao što su prikupljanje, bilježenje, organizacija, strukturiranje, pohrana, prilagodba ili izmjena, pronalaženje, obavljanje uvida, uporaba, otkrivanje

[16] Opća uredba o zaštiti podataka, n.dj., čl. 1.
[17] Isto, čl. 2

prijenosom, širenjem ili stavljanjem na raspolaganje na drugi način, usklađivanje ili kombiniranje, ograničavanje, brisanje ili uništavanje“.[18]

Prema Uredbi, osnovna načela obrade osobnih podataka su zakonitost, poštenost i transparentnost, ograničavanje svrhe, smanjenje količine podataka, točnost, ograničenje pohrane, cjelovitost, povjerljivost i pouzdanost. Ograničavanje svrhe podrazumijeva prikupljanje osobnih podataka u određene svrhe te se ti podaci smiju obrađivati samo na način koji je u skladu s tim svrhama, a kod ograničenja pohrane da su čuvani u izvornom obliku samo onoliko dugo koliko nalaže svrha za koju se obrađuju.[19]

Obrada osobnih podataka smatra se zakonitom samo ako postoji pravno osnovana podloga koja opravdava takvu obradu. Uredba navodi da se „zabranjuje obrada osobnih podataka koji otkrivaju rasno ili etničko podrijetlo, politička mišljenja, vjerska ili filozofska uvjerenja ili članstvo u sindikatu te obrada genetskih podataka, biometrijskih podataka u svrhu jedinstvene identifikacije pojedinca, podataka koji se odnose na zdravlje ili podataka o spolnom životu ili seksualnoj orijentaciji pojedinca“ (posebne kategorije podataka), osim u slučaju da je te podatke već objavio sam ispitanik ili je dao izričitu privolu za njihovu obradu, u slučaju javnog interesa u području javnog zdravlja, za zaštitu ljudskog života ako ispitanik nije sposoban dati privolu ili u svrhe arhiviranja u javnom interesu radi istraživanja ili u statističke svrhe.[20]

Svaki ispitanik u podacima prikupljenim za obradu ima pravo na ispravak (čl. 16), brisanje ili zaborav (čl. 17), ograničenje obrade (čl. 18), prenosivost podataka (čl. 20) i pravo na prigovor (čl. 21). Kako bi njihovi podaci bili obrađeni na siguran način, voditelj i izvršitelj obrade dužni su provesti tehničke mjere s odgovarajućom razinom sigurnosti uključujući, između ostalog, i tehnike anonimizacije u proces obrade (čl. 32).

Članak 89. Opće uredbe o zaštiti podataka najvažniji je za arhive kao institucije koje u svojem gradivu sadrže osobne podatke. On obrađuje zaštitne mjere i iznimke za obradu osobnih podataka u svrhe arhiviranja u javnom interesu, znanstvenih istraživanja i u statističke svrhe. Ova obrada zahtijeva primjenu odgovarajućih zaštitnih mjera kako bi se osigurala zaštita prava i sloboda pojedinaca. Spomenute mjere uključuju tehničke i organizacijske metode s naglaskom na smanjenje količine podataka, kao primjerice pseudonimizaciju, dok god se svrhe obrade mogu postići u takvim uvjetima.

[18] Isto, čl. 14
[19] Isto, čl. 5.
[20] Isto, čl. 9.

Uvodna odredba 26[21] kaže da bi se načela zaštite podataka trebala primjenjivati „na sve informacije koje se odnose na pojedinca čiji je identitet utvrđen ili se može utvrditi“. S obzirom na to da bi za otkrivanje kome pripadaju osobni podaci koji su pseudonimizirani trebalo uzeti u obzir sva sredstva i čimbenike koji su potrebni za otkrivanje identiteta, zaključeno je da se „načela zaštite podataka stoga 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 tako da se identitet ispitanika ne može ili više ne može utvrditi“. Jednostavnije rečeno, anonimizirani podaci ne smatraju se osobnim podacima te se na njih ne primjenjuje klasična zaštita podataka.

Kako bi se prikazala općenita slika upotrebe Uredbe za funkcioniranje hrvatskih arhiva, Bukvić se dotiče problematike opisa osobnih podataka u arhivskom gradivu u skladu s Uredbom s obzirom na to da standardi za opis arhivskog gradiva ISAD(G) i ISAAR(CPF) nemaju posebne smjernice za opis gradiva ograničene dostupnosti. S obzirom na to da je Uredbom naglašena važnost smanjenja količine podataka, ključni kriterij za odlučivanje o tome kako postupiti s podacima, osobito u kontekstu izrade obavijesnih pomagala i opisa fondova i zbirki, jest procjena osjetljivosti i prirode osobnih podataka. Arhivist treba pronaći ravnotežu između obrade podataka da se zaštiti privatnost pojedinaca i arhivističkog nagona i zakonske obveze da se gradivo od općeg interesa prikupi, opiše i da na korištenje. Kao jedno od rješenja ovog problema Bukvić predlaže izradu dvije verzije obavijesnog pomagala za određeni fond – detaljno pomagalo za interne potrebe arhiva i manje detaljno pomagalo u kojem je gradivo opisano samo na višim razinama za javno korištenje i distribuciju.[22]

2.2. Pravo na pristup informacijama

Ustavom se građanima također jamči i pravo na pristup informacijama[23] koje su u posjedu tijela javnih vlasti. To je pravo ograničeno u nekim slučajevima, koji se propisuju zakonom.[24] Zakon o pravu na pristup informacijama (NN 25/2013)[25] donesen je kako bi se osiguralo da se odredbe o pravu na pristup informacijama ostvarenih Ustavom Republike Hrvatske zaista i ispunjavaju. Tako ovaj Zakon govori da pravo na pristup informacijama „obuhvaća pravo

[21] Isto, (26)
[22] Bukvić, Nenad. „Opća uredba o zaštiti podataka (GDPR) i opis osobnih podataka u arhivskom gradivu: o nekim praktičnim i etičkim aspektima“. *Arhivski vjesnik* 63, br. 1 (2020): 22-23. Dostupno na: https://doi.org/10.36506/av.63.1
[23] Ustav Republike Hrvatske, n.dj., čl. 38
[24] Isto.
[25] https://narodne-novine.nn.hr/clanci/sluzbeni/2013_02_25_403.html

korisnika na traženje i dobivanje informacije kao i obvezu tijela javne vlasti da omogući pristup zatraženoj informaciji, odnosno da objavljuje informaciju neovisno o postavljenom zahtjevu kada takvo objavljivanje proizlazi iz obveze određene zakonom ili drugim propisom“[26]. Zakon o pravu na pristup informacijama odnosi se na informacije koje su u posjedu tijela javnih vlasti, odnosno „svaki podatak koji posjeduje tijelo javne vlasti u obliku dokumenta, zapisa, dosjea, registra, neovisno o načinu na koji je prikazan (napisani, nacrtani, tiskani, snimljeni, magnetni, optički, elektronički ili neki drugi zapis), koji je tijelo izradilo samo ili u suradnji s drugim tijelima ili dobilo od druge osobe, a nastao je u okviru djelokruga ili u vezi s organizacijom i radom tijela javne vlasti“[27].

Iako pravo na pristup informacijama jest temeljno, postoje situacije u kojima se to pravo ograničava radi zaštite drugih vrijednosti kao što su privatnost, nacionalna sigurnost ili zakonom propisana tajnost. To znači da određene informacije mogu biti nedostupne javnosti, jer bi njihovo objavljivanje moglo nanijeti štetu tim vrijednostima. Tako se Zakon o pravu na pristup informacijama ne odnosi na informacije koje su klasificirane ili za koje postoji obveza čuvanja tajnosti (čl. 1.). Takve informacije prošle su cijeli proces klasifikacije te postoji valjani, zakonom određen razlog zašto su klasificirane određenom razinom tajnosti što ih čini nedostupnima za korištenje sve do isteka rokova tajnosti ili deklasifikacije.

Pravo na pristup informacijama može biti ograničeno prema onim informacijama „koje se tiču svih postupaka koje vode nadležna tijela u predistražnim i istražnim radnjama za vrijeme trajanja tih postupaka“. Prema Zakonu o pravu na pristup informacijama[28], tijela javne vlasti mogu ograničiti pristup informacijama:

1. „ako je informacija klasificirana stupnjem tajnosti, sukladno zakonu kojim se uređuje tajnost podataka;
2. ako je informacija poslovna ili profesionalna tajna, sukladno zakonu;
3. ako je informacija porezna tajna, sukladno zakonu;
4. ako je informacija zaštićena zakonom kojim se uređuje područje zaštite osobnih podataka;
5. ako je informacija zaštićena propisima kojima se uređuje pravo intelektualnog vlasništva, osim u slučaju izričitoga pisanog pristanka nositelja prava:

[26] Hrvatski sabor. „Zakon o pravu na pristup informacijama“. *Narodne novine* 25/13 (2013). Dostupno na: https://narodne-novine.nn.hr/clanci/sluzbeni/2013_02_25_403.html
[27] Zakon o pravu na pristup informacijama, 2013, 2015
[28] Zakon o pravu na pristup informacijama, 2013, n.dj., čl. 15

6. ako je pristup informaciji ograničen sukladno međunarodnim ugovorima ili se radi o informaciji nastaloj u postupku sklapanja ili pristupanja međunarodnim ugovorima ili pregovora s drugim državama ili međunarodnim organizacijama, do završetka postupka, ili se radi o informaciji nastaloj u području održavanja diplomatskih odnosa;
7. u ostalim slučajevima utvrđenim zakonom."

Prema istom Zakonu, „tijela javne vlasti mogu ograničiti pristup informaciji ako postoje osnove sumnje da bi njezino objavljivanje:

1. onemogućilo učinkovito, neovisno i nepristrano vođenje sudskog, upravnog ili drugog pravno uređenog postupka, izvršenje sudske odluke ili kazne,
2. onemogućilo rad tijela koja obavljaju upravni nadzor, inspekcijski nadzor, odnosno nadzor zakonitosti".[29]

Tijela javne vlasti također imaju pravo ograničiti pristup informaciji ako je:

1. „informacija u postupku izrade unutar jednog ili među više tijela javne vlasti, a njezino bi objavljivanje prije dovršetka izrade cjelovite i konačne informacije moglo ozbiljno narušiti proces njezine izrade;
2. informacija nastala u postupku usuglašavanja pri donošenju propisa i drugih akata te u razmjeni stavova i mišljenja unutar jednog ili među više tijela javne vlasti, a njezino bi objavljivanje moglo dovesti do pogrešnog tumačenja sadržaja informacije, ugroziti proces donošenja propisa i akata ili slobodu davanja mišljenja i izražavanja stavova".[30]

Informacije koje su podložne ograničenjima pristupa zbog navedenih razloga mogu postati dostupne široj javnosti kada to odobri osoba koja bi mogla pretrpjeti štetu ako bi te informacije bile objavljene. Međutim, ova objava ne može biti odgođena duže od 20 godina od trenutka kada je informacija prvotno nastala, osim ako zakon ili neki drugi propis ne određuje dulji vremenski rok. Sama nedostupnost informacija javnosti prestaje kad prestanu

[29] Isto, čl. 15
[30] Isto, čl. 15

razlozi koji su tijelu javne vlasti omogućili ograničavanje prava na pristup tim informacijama.[31] Ovo pruža fleksibilnost i osigurava da informacije postanu dostupne čim više ne postoji opravdanje za njihovo zadržavanje u tajnosti.

Izmjenom Zakona o pravu na pristup informacijama, čija je najnovija verzija stupila na snagu 2022. godine, uvažena je važnost anonimizacije. Kroz definiciju je pojašnjen proces anonimizacije, kojim se informacija transformira u anonimne podatke koji ne otkrivaju identitet fizičkih osoba te uz pomoć njih nije moguće utvrditi identitet pojedinca.[32] Uvođenje anonimizacije u Zakon čini regulaciji prava na pristup informacijama sveobuhvatnijom i relevantnijom za suvremene izazove povezane s obradom i zaštitom podataka. Anonimizacija omogućava da se vrijedne informacije i podaci koriste za općenite analize, donošenje informiranih odluka i istraživanje, a da pritom ne ugrožavaju privatnost pojedinaca.

2.3. Dostupnost arhivskog gradiva

Dostupnost arhivskog gradiva za korištenje regulira se kroz Zakon o arhivskom gradivu i arhivima. Prema članku 14. Zakona o arhivskom gradivu i arhivima (NN 61/18)[33], imatelji bi javno arhivsko gradivo trebali „predati nadležnom državnom arhivu u roku koji u pravilu nije dulji od 30 godina od njegova nastanka“, a javno arhivsko gradivo u digitalnom obliku u roku ne duljem od 10 godina od njegova nastanka. Javno arhivsko gradivo od javnog je interesa i važno za razumijevanje povijesti i funkcioniranja društva. Imatelji takvog gradiva predaju ga nadležnom državnom arhivu kako bi se relevantni materijali prikupili i sačuvali te bili dostupni povjesničarima, istraživačima i drugima koji su zainteresirani za proučavanje prošlosti i dokazivanje prava. Uz sve veći razvoj tehnologije, informacije u digitalnom obliku postaju sve važnije. Rok od 10 godina, kraći od onog za konvencionalno fizičko gradivo, zadan je kako bi se osigurala dugoročna dostupnost digitalnih materijala, jer digitalni formati mogu biti osjetljivi na tehnološke promjene i zastarijevanje.

Zakon o arhivskom gradivu i arhivima u članku 18. nalaže da javno arhivsko gradivo mora biti dostupno od svog nastanka, osim u slučajevima kad je zakonom određeno drukčije (npr. zbog prisustva osobnih podataka u tom gradivu i slično), a u slučaju ograničenja dostupnosti arhivskoga gradiva to gradivo mogu koristiti samo ovlaštene osobe. Propisi kojima je

[31] Isto, čl. 15
[32] Zakon o pravu na pristup informacijama, 2022, n.dj., čl. 2
[33] Hrvatski sabor. „Zakon o arhivskom gradivu i arhivima“. *Narodne novine* 61/18 (2018). Dostupno na: https://narodne-novine.nn.hr/clanci/sluzbeni/2018_07_61_1265.html (1.4.2023.)

ograničena dostupnost gradiva reći će koje su osobe ovlaštene za pristup. Javno gradivo nastalo do 30. svibnja 1990. godine, prema Zakonu, mora biti dostupno za korištenje bez ograničenja, osim onoga koje sadrži osobne podatke ili vojnu dokumentaciju od značenja za nacionalnu sigurnost Republike Hrvatske.[34]

Rokovi ograničene dostupnosti od 40 godina od nastanka arhivskog gradiva namijenjeni su sljedećem gradivu:

- „koje sadrži klasificirane i druge tajne podatke (…),
- koje nije označeno stupnjem tajnosti sukladno propisima kojima se uređuje tajnost podataka, a sadrži projektnu i tehničku dokumentaciju štićenih i vojnih objekata, industrijskih postrojenja, infrastrukturnih objekata i sl. te podatke o prirodnim i strateškim bogatstvima od značenja za nacionalnu sigurnost Republike Hrvatske, a koje je predano u nadležni arhiv“.[35]

Ovo se gradivo može dati na korištenje i prije isteka roka od 40 godina „ako je na temelju zakona kojima se uređuje pravo na pristup informacijama i zaštita tajnosti podataka utvrđeno da javni interes, koji se ostvaruje dostupnošću podataka, preteže nad interesima koji se štite ili ako su prethodno poduzete mjere koje osiguravaju zaštitu javnih i privatnih interesa zbog kojih je utvrđeno ograničenje dostupnosti odnosno koje onemogućuju uvid u klasificirane podatke“.[36]

2.3.1. Zakon o arhivskom gradivu i arhivima i osobni podaci

Dostupnost javnog arhivskog gradiva u Republici Hrvatskoj koje sadrži osobne podatke regulira Zakon o arhivskom gradivu i arhivima. Prema Zakonu, osobni podaci u javnom gradivu za korištenje su dostupni nakon smrti osobe ili 100 godina od rođenja osobe na koju se podatak odnosi. Ako je teško ili nemoguće utvrditi datume rođenja i smrti osobe na koju se osobni podatak odnosi, javno arhivsko gradivo dostupno je na korištenje 70 godina od dana nastanka tog gradiva[37], što bi u većini slučajeva trebalo biti moguće odrediti, s obzirom na tradiciju datiranja službenih spisa koji nastaju kao podrška poslovanju. U slučaju potrebe za korištenjem arhivskog gradiva koje sadrži osobne podatke prije nego isteknu rokovi od 70 ili 100 godina, nadležni državni arhiv dužan je zaštititi identitet osobe na koju se osobni podaci

[34] Isto, čl. 20
[35] Isto, čl. 18
[36] Isto, čl. 18
[37] Isto, čl. 19

odnose poduzimajući potrebne tehničke mjere, kao što je anonimizacija, kako bi sačuvao tajnost identiteta osobe, dok će korisnik potpisati izjavu da, u slučaju da prema dostupnim podacima ipak može identificirati osobu, neće otkriti njen identitet. Te tehničke mjere izvorno arhivsko gradivo ne smiju oštetiti ili uništiti te mora biti cjelovito dostupno za korištenje nakon isteka roka. Članak 19. ovog Zakona navodi slučajeve u kojima su osobni podaci u javnom arhivskom gradivu dostupni i prije isteka roka:

- „ako korištenje arhivskoga gradiva zahtijeva osoba na koju se gradivo odnosi ili osoba koju ona ovlasti,
- ako je od nastanka namijenjeno javnosti,
- ako je osobni podatak osobe na koju se osobni podatak odnosi već postao dostupan javnosti ili je općepoznat,
- ako na to pristane osoba na koju se osobni podatak odnosi,
- ako je osoba na koju se osobni podatak odnosi već postao dostupan javnosti ili je općepoznat,
- ako je osoba na koju se osobni podatak odnosi sama ili putem druge osobe objavila taj podatak.“[38]

Nisu svi osobni podaci zaštićeni i nedostupni za korištenje sve do smrti osobe na koju se odnose ili nakon isteka rokova od 70 ili 100 godina. Zakon o arhivskom gradivu i arhivima navodi da su „osobni podaci osobe koja je do 30. svibnja 1990. obnašala javne dužnosti ili bila pripadnik ili suradnik službe sigurnosti dostupni bez ograničenja iz stavka 1. [gradivo nastalo do 30. svibnja 1990. dostupno je za korištenje bez ograničenja] u dijelu koji se odnosi na obavljanje te javne dužnosti odnosno službe“[39].

Kako bi se bar malo umanjila šteta koju dostupnost tih podataka može učiniti po ugled osobe na koju se odnose, „svaka osoba čiji se osobni podaci nalaze u javnom arhivskom gradivu iz stavka 1. [gradivo nastalo do 30. svibnja 1990. dostupno je za korištenje bez ograničenja] ima pravo o gradivu koje se na nju odnosi dati pisanu izjavu kojom osporava istinitost ili potpunost svojih osobnih podataka uz navođenje arhivskog fonda i arhivskoga gradiva na koje se izjava odnosi“, a ta se izjava daje na uvid svakom korisniku gradiva koje je navedeno u izjavi zajedno s tim gradivom.[40]

[38] Zakon o arhivskom gradivu i arhivima, n.dj., čl. 19
[39] Isto, čl. 20
[40] Isto, čl. 20

Navedeni mehanizam za zaštitu osobnih podataka u okviru javnog arhivskog gradiva osmišljen je kako bi se uskladio pristup povijesnim i kulturnim materijalima s potrebom zaštite privatnosti i ugleda pojedinaca koji se spominju u toj baštini. On omogućuje pojedincima čije se osobne informacije nalaze u arhivskom gradivu da zatraže zaštitu svojih interesa pružanjem izjave koja osporava istinitost ili potpunost tih podataka, a isto tako omogućuje pojedincima i da reagiraju na informacije te u pisanoj izjavi isprave ili dopune te podatke. Ovo je iznimno važno kako bi se osigurala točnost i ispravnost informacija sadržanih u arhivskom gradivu. Na ovaj se način pokušava stvoriti okruženje u kojem se javno arhivsko gradivo može koristiti u svrhe istraživanja i edukacije, uz istovremeno poštovanje prava i dostojanstva pojedinaca.

2.3.2. Načela dostupnosti arhivskoga gradiva i tehničke smjernice

Načela dostupnosti arhivskoga gradiva[41] (u daljnjem tekstu koristit će se izraz Načela) usvojena su na Općoj godišnjoj skupštini Međunarodnog arhivskog vijeća 2012. godine, a na hrvatski jezik prevedena 2015. godine u izdanju Hrvatskog državnog arhiva. Sastoje se od 10 načela, a svako načelo uz sebe ima i svoje tumačenje. Načela su svjesna da nije sve arhivsko gradivo u cijelosti dostupno za javnost, pa je potrebno postaviti određena ograničenja na gradivo. Nekoliko načela (i njihovih tumačenja) je važno u kontekstu zaštite osobnih podataka:

> 4. Ustanove s arhivskim gradivom skrbe da odluke o ograničenju dostupnosti budu jasne i s navedenim trajanjem, da su utemeljene na pozitivnim propisima, da priznaju pravo na privatnost i poštuju prava vlasnika privatnoga arhivskoga gradiva.

Ograničenja se postavljaju na arhivsko gradivo radi zaštite osobnih podataka, privatnosti, poslovnih tajni, sigurnosti pojedinaca i nacionalne sigurnosti, podataka o istrazi ili primjeni zakona. Vremenski rokovi ovakvih ograničenja moraju biti jasni.

> 5. Arhivsko gradivo dostupno je pod pravednim i jednakim uvjetima za sve.

Svi korisnici moraju imati mogućnosti korištenja gradiva bez diskriminacije, što je njihovo Ustavom zajamčeno pravo. Prema Načelima, postoje različite skupine korisnika, pa tako i one koje imaju pravo pristupa podacima koji su ograničeni (npr. istraživači iz medicine koji

[41] Međunarodno arhivsko vijeće. „Načela dostupnosti arhivskoga gradiva“. Zagreb: *Hrvatski državni arhiv* (2015). Dostupno na: https://www.ica.org/sites/default/files/ICA_Access_Principles__CR.pdf

kroz bolničke zapise skupljaju statističke podatke). Unutar takve, određene skupine korisnika, pravila dostupnosti također se primjenjuju jednako, bez diskriminacije.

> 6. Ustanove s arhivskim gradivom osiguravaju žrtvama teških zločina prema međunarodnom pravu dostupnost arhivskoga gradiva koje pruža dokaz potreban za obranu njihovih ljudskih prava te za dokumentiranje njihova kršenja, čak i ako je to gradivo nedostupno javnosti.

Osobama koje koriste arhivsko gradivo za istraživanje u svrhu zaštite ljudskih prava to je gradivo dostupno čak i ako nije dostupno široj javnosti zbog postojanja osobnih podataka i činjenice da nisu prošli rokovi određeni Zakonom o arhivskom gradivu i arhivima. To je gradivo dokaz kršenja ljudskih prava kroz teške (ratne) zločine kažnjive u međunarodnom pravu, a znajući da arhivi od svojih početaka čuvaju dokumente važne za dokazivanje određenih prava (npr. ugovore), ne čudi što je ovo gradivo, iako većini korisnika nedostupno, ipak dostupno u ovakvim, iznimnim slučajevima.

Kako bi se smanjila opasnost od klevete i objavljivanja lažnih informacija iz arhivskog gradiva, postoji mehanizam kojim prema Zakonu o arhivskom gradivu i arhivima (članak 20.), „svaka osoba čiji se osobni podaci nalaze u javnom arhivskom gradivu [nastalom prije 30. svibnja 1990.] ima pravo o gradivu koje se na nju odnosi dati pisanu izjavu kojom osporava istinitost ili potpunost svojih osobnih podataka“, a ta se pisana izjava daje na uvid korisnicima zajedno s javnim arhivskim gradivom na koje odnosi.[42]

Uz Načela su izdane i Tehničke smjernice za upravljanje arhivskim gradivom[43]. Smjernice su sljedeće:

> A. Informirati javnost o arhivskome gradivu

Javnost bi trebala imati pristup gradivu i ustanovi, odnosno informacijama o gradivu i ustanovi koja ga čuva, te bi se te informacije trebale pružati besplatno. Isto tako bi besplatno dostupna za korištenje trebala biti i obavijesna pomagala.

> B. Razviti politiku dostupnosti

Politika dostupnosti gradi se prema različitim zakonima i propisima, nalozima i sudskim odlukama, politikama i internim pravilima te sporazumima o darovanju.

---

[42] Zakon o arhivskom gradivu i arhivima, n.dj., čl. 20

[43] Međunarodno arhivsko vijeće. „Načela dostupnosti arhivskoga gradiva. Tehničke smjernice za upravljanje arhivskim gradivom ograničene dostupnosti“. Zagreb: *Hrvatski državni arhiv*. (2016). Dostupno na: https://www.ica.org/sites/default/files/tech-guidance_hr.pdf

C. Tijekom preuzimanja gradiva u arhiv postići dogovor o ograničenjima dostupnosti

Predavatelj gradiva zajedno s arhivskom ustanovom dogovara ograničenja dostupnosti gradiva koje predaje arhivu te se prema tim ograničenjima formira opća politika dostupnosti. Ista je situacija i kod donacija gradiva neke organizacije ili osobnih dokumenata. Bitan element je dogovor.

D. Kontrolirati fizički pristup gradivu ograničene dostupnosti

Ovisi o vrsti gradiva i razini sigurnosti, različiti su načini na koje se kontrolira fizički pristup gradivu. Za fizičko gradivo kontrolira se fizički pristup arhivskim spremištima ili se gradivo čuva u zasebnome području (odvojeno područje za fizičko gradivo, računalna zaštita za elektroničko gradivo).

E. Omogućiti zaposlenicima pristup gradivu ograničene dostupnosti u svrhu arhivističke obrade

I gradivo ograničene dostupnosti trebalo bi biti sređeno, opisano i zaštićeno, tako da arhivisti moraju imati pristup arhivskom gradivu čak i ako ono nije dostupno za korištenje. Postoje načini na koje se osigurava sigurnost informacija ograničene dostupnosti, npr. potpisivanjem ugovora o čuvanju tajnosti, arhivisti mogu proći postupak provjere kako bi dobili odobrenje za pristup gradivu ograničene dostupnosti, ustanova može ograničiti broj zaposlenika koji imaju pristup takvom gradivu i slično.

F. Opisati gradivo ograničene dostupnosti

Ova smjernica povezana je s prethodnom, jer arhivistima treba dati pristup gradivu ograničene dostupnosti da bi ga mogli opisati. Opis je važan kako bi i ostali arhivisti i korisnici znali da se u nekom fondu nalazi gradivo kojemu je dostupnost ograničena, ali i razlog ograničenosti. I sama Opća međunarodna norma za opis arhivskoga gradiva (ISAD(G)) u elementu 3.4.1. Uvjeti dostupnosti navodi da je cilj u opisu gradiva ograničene dostupnosti „pružiti obavijesti o pravnom statusu ili drugim propisima koji ograničavaju ili utječu na dostupnosti jedinice opisa“, a pravilo pri opisu „navesti zakonski ili pravni status, ugovor, propis ili politiku što utječu na dostupnosti jedinice opisa. Ako je moguće, naznačiti rok i nadnevak nakon kojeg će gradivo biti dostupno.“[44] Važno je dokumentirati sve

[44] Međunarodno arhivsko vijeće. „Opća međunarodna norma za opis arhivska gradiva ISAD(G)“. Zagreb: *Hrvatski državni arhiv* (2001). Dostupno na: https://www.ica.org/sites/default/files/isad_g_2_edition_hr.pdf

informacije koje mogu pomoći i arhivistima i korisnicima, a treba biti svjestan da će i to gradivo nakon isteka ograničene dostupnosti postati u cijelosti dostupno za korištenje.

G. Odgovoriti na zahtjeve za korištenjem gradiva ograničene dostupnosti

Nakon dolaska upita za gradivom, arhivist provjerava je li gradivo javno dostupno za korištenje ili postoje uvjeti ograničenosti. Činjenica da je nekom gradivu dostupnost ograničena ne znači da mu baš nijedan korisnik nema pristup. Postoje iznimke koje su navedene u politici dostupnosti gradiva ograničene dostupnosti svake ustanove. Jedan od primjera proizlazi iz šestog načela Načela dostupnosti arhivskoga gradiva (ustanove s arhivskim gradivom osiguravaju žrtvama teških zločina prema međunarodnome pravu dostupnost arhivskoga gradiva koje pruža dokaz potreban za obranu njihovih ljudskih prava te za dokumentiranje njihova kršenja, čak i ako je to gradivo nedostupno javnosti). Također se pristup može dati i u svrhe statističkih istraživanja, radi obavljanja službenih dužnosti i slično. Svaki pristup gradivu ograničene dostupnosti se evidentira u ustanovi u kojoj korisnik zahtijeva pristup čuvajući informacije o identitetu korisnika te revizijama. Korisnici ne smiju nikome otkriti koje su to informacije čija je dostupnost ograničena.

H. Donositi odluke o dostupnosti

Arhivisti provode revidiranje odluka o ograničenju dostupnosti kada: „a) potencijalni korisnik zahtijeva korištenje zapisa koji prethodno nisu bili dostupni javnosti ili b) arhivska ustanova utvrdi da je protokom vremena odluka o ograničenju dostupnosti izgubila vrijednosti“.[45] Revidiranje kreće pregledom ograničenja (zakona i propisa, internih politika i slično), a nakon revizije se ili gradivo daje na javno korištenje ili se potvrđuje odluka o ograničavanju dostupnosti gradiva.

I. Primjenjivati ograničenja dostupnosti

Arhivisti iz javno dostupnog gradiva izdvaja predmete, zapise ili njihove dijelove koji sadržavaju informaciju kojoj je dostupnost ograničena. Neke ustanove ulaze u veći rizik neovlaštena otkrivanja informacija jer ne žele ograničiti korištenje gradiva pa korisnicima na potpis daju ugovor o neotkrivanju informacija. Druge ustanove koriste različite fizičke tehnike kako bi redigirali informacije ograničene dostupnosti na preslikama izvornih dokumenata, što je jedan od načina provođenja anonimizacije.

[45] Načela dostupnosti. Tehničke smjernice, n.dj., str. 15

J. Dokumentirati odluke o dostupnosti gradiva

Svaka odluka o dostupnosti mora biti evidentirana, pa tako i rezultati revizija ograničenja dostupnosti. Odluke moraju biti navedene u obavijesnim pomagalima i opisima fondova.

K. Preispitivati odluke o ograničenju dostupnosti slijedom žalbe ili interne procedure

Svaki zahtjev za korištenjem arhivskog gradiva može biti odbijen ako postoje razlozi za takvo postupanje, odnosno ako je gradivo ograničene dostupnosti. Svaki korisnik ima pravo žalbe u slučaju da mu je zahtjev za korištenjem arhivskoga gradiva odbijen.

L. Učiniti dostupnim gradivo kojemu je prethodno bila ograničena dostupnost

Nakon isteka razloga ograničene dostupnosti nekog arhivskog gradiva, što se evidentira u reviziji dostupnosti, ukida se ograničenje dostupnosti i gradivo postaje dostupno za korištenje široj javnosti.

Načela dostupnosti arhivskoga gradiva i njihove tehničke smjernice primjenjuju se u arhivistici kako bi se osiguralo odgovarajuće rukovanje i pristupanje arhivskim materijalima. Arhivsko gradivo često sadrži podatke relevantne za povijest, pravosuđe i istraživanje. Pravo javnosti na pristup ovim informacijama osigurava demokratičnost i omogućava kontekstualizaciju povijesnih događaja. Ova načela pomažu održavanju ravnoteže između zaštite osjetljivih podataka i prava javnosti (korisnika) da pristupi povijesnim informacijama. Ravnoteža je ključna za očuvanje kulturne i povijesne baštine te za postizanje povjerenja javnosti u upravljanje i očuvanje arhivskih resursa.

3. Anonimizacija

Postupak anonimizacije je najjednostavnije definirati kao postupak uklanjanja identifikacijskih informacija iz skupa podataka radi zaštite privatnosti pojedinaca, čime se sprječava ponovna identifikacija pojedinaca iz anonimiziranih podataka. Zakon o izmjenama i dopunama Zakona o pravu na pristup informacijama definira anonimizaciju kao „proces izmjene informacije u anonimne informacije koje ne upućuju na fizičke osobe čiji je identitet utvrđen ili čiji se identitet može utvrditi ili proces kojim se osobni podaci čine anonimnima tako da se ne može utvrditi identitet ispitanika ili se više ne može utvrditi identitet ispitanika“[46]. Anonimizacija je ključna točka zaštite privatnosti jer omogućava korištenje podataka u analizama i istraživanjima bez rizika od lake identifikacije pojedinaca čiji se podaci obrađuju i koriste. Ovaj proces osigurava da se informacije koriste na odgovoran način, čime se balansira između potrebe za podacima i poštovanja privatnosti pojedinaca.

Alati za anonimizaciju razlikuju se po tome koriste li statičku li dinamičku anonimizaciju. U statičkoj anonimizaciji, izdavač (osoba koja će provesti anonimizaciju) anonimizira bazu podataka i potom je objavljuje tako da treće strane mogu jednostavno pristupiti podacima. Kako ne bi došlo do otkrivanja podataka, važno je da se precizno definiraju atributi koje je potrebno anonimizirati zbog zaštite podataka i privatnosti i svrha za koju će se ti podaci koristiti. Ovdje je ključno detaljno razumjeti svrhu korištenja anonimiziranog skupa podataka kako bi se u procesu anonimizacije sačuvala kvaliteta informacija koje su bitne za postizanje svrhe, uz prihvaćanje većih ograničenja za manje bitne podatke. U statičkoj anonimizaciji podaci su često samo pseudonimizirani, pa je moguće izvući zaključke spajanjem tih podataka s drugim dostupnim informacijama. Primjeri alata za statičku anonimizaciju su Amnesia, Open Anonymizer i ARX Anonymization Tool.[47]

Kod dinamičke ili interaktivne anonimizacije, anonimizacija se primjenjuje na rezultate upita kako dolaze u sustav, a ne na cijeli skup podataka na samom početku (kao kod statičke). Osoba zadužena za analizu pristupa bazi podataka putem sučelja i postavlja upit te prima anonimizirane rezultate iz baze podataka koja nije anonimizirana, već se anonimizacija događa automatski tijekom upita. S obzirom na to da se anonimizacija provodi automatski, analitičari ne moraju biti stručnjaci za zaštitu podataka ili anonimizaciju, što mu omogućava

[46] Zakon o izmjenama i dopunama Zakona o pravu na pristup informacijama, 2022, čl. 2
[47] Sartor, Nicolas. „Data Anonymization Software – Diferences Between Static and Interactive Anonymization“. *Aircloak* (2019). Dostupno na: https://aircloak.com/data-anonymisation-software-differences-between-static-and-interactive-anonymisation/

da se potpuno posveti analizi podataka. Dinamičku anonimizaciju koriste alati kao što su Googleov RAPPOR, GUPT, FLEX i PINQ.[48]

3.1. Podatak

Hrvatska enciklopedija Leksikografskog zavoda Miroslav Krleža definira *podatak* kao „poznatu ili pretpostavljenu činjenicu na osnovi koje se oblikuje informacija“[49], a Institut za jezik i jezikoslovlje u svom projektu Struna kao „prikaz obavijesti na formaliziran način, prikladan za komunikaciju, tumačenje, pohranu i obradbu“[50].

Zakon o tajnosti podataka u kontekstu zaštite podataka ide malo dalje, pa podatak definira kao „dokument, odnosno svaki napisani, umnoženi, nacrtani, slikovni, tiskani, snimljeni, fotografirani, magnetni, optički, elektronički ili bilo koji drugi zapis podatka, saznanje, mjera, postupak, predmet, usmeno priopćenje ili informacija, koja s obzirom na svoj sadržaj ima važnost povjerljivosti i cjelovitosti za svoga vlasnika“. Razlikuje dvije vrste podatka: klasificirani (za koji je utvrđen stupanj tajnosti) i neklasificirani podatak (za koji nije utvrđen stupanj tajnosti).[51]

U okviru pravnog okvira zaštite podataka razlikuju se četiri vrste podataka: osobni podaci, osjetljivi podaci (ili posebne kategorije osobnih podataka), anonimni podaci i pseudonimizirani podaci. Osobni podaci su informacije koje se odnose na identificiranu fizičku osobu. To su sve informacije koje omogućuju direktno identificiranje pojedinca, kao što su ime i prezime, adresa, broj osobne iskaznice ili osobni identifikacijski broj. Njihova obrada podliježe Općoj uredbi o zaštiti podataka. Osjetljivi podaci, ili posebne kategorije osobnih podataka su posebno osjetljivi podaci povezani s privatnim aspektima nečijeg života, odnosno osobni podaci koji otkrivaju „rasno ili etničko podrijetlo, politička mišljenja, vjerska ili filozofska uvjerenja ili članstvo u sindikatu, genetski podaci, biometrijski podaci u svrhu jedinstvene identifikacije pojedinca, podaci koji se odnose na zdravlje te podaci o spolnom životu ili seksualnoj orijentaciji pojedinca“[52]. Anonimizirani podaci su podaci koji su obrađeni i promijenjeni tako da više nije moguće identificirati pojedinca na temelju tih

[48] Isto.
[49] Hrvatska enciklopedija, mrežno izdanje. „podatak“. *Leksikografski zavod Miroslav Krleža* (2021). Dostupno na: https://www.enciklopedija.hr/natuknica.aspx?id=48887
[50] Struna. „podatak“, *Institut za hrvatski jezik i jezikoslovlje*. Dostupno na: http://struna.ihjj.hr/naziv/podatak/6110/
[51] Hrvatski sabor. „Zakon o tajnosti podataka“. *Narodne novine* 79/07 (2007). Dostupno na: https://narodne-novine.nn.hr/clanci/sluzbeni/2007_07_79_2483.html
[52] Opća uredba o zaštiti podataka, n.dj., čl. 9

podataka. Upravo zbog toga, anonimizirani podaci ne smatraju se osobnim podacima pa ne podliježu Općoj uredbi o zaštiti podataka te za njihovu daljnju obradu tih podataka nije potrebna privola ispitanika. Pseudonimizirani podaci su oni kod kojih su identifikatori zamijenjeni pseudonimima ili šiframa, čime se smanjuje rizik od identifikacije. Sve četiri vrste podataka igraju ključnu ulogu u regulaciji zaštite podataka kako bi se osiguralo da se informacije koriste odgovorno i da se očuva privatnost pojedinaca.

3.2. Mjerila anonimizacije i modeli privatnosti

3.2.1. *K*-anonimnost

S pojmom *k*-anonimnost opisujemo tehnike sakrivanja identiteta neke osobe u grupi osoba sličnih svojstava. U *k*-anonimnosti *k* predstavlja broj koji pokazuje veličinu skupine, odnosno količinu osoba u skupini. Kao ilustracija, ako se zamisli određeni skup podataka u kojem je određeno da je *k* jednako 100, svojstvo koje povezuje sve osobe u podacima neka bude ime, cilj je spriječiti mogućnost da se iz skupine izdvoji određena osoba grupirajući je s *k*-1 ostalih osoba istog svojstva. Svih 100 osoba u tom skupu podataka imaju isto ime, tako da ne možemo identificirati jednu konkretnu samo na temelju odabranog svojstva jer postoji *k* zapisa koji su jednaki po vrijednosti odabranog svojstva.[53]

3.2.2. *L*-raznolikost

Sama *k*-anonimnost nije dovoljna zaštita jer sve osobe u grupi dijele istu informaciju. Ako prema gornjem primjeru pretražimo određeno ime koje dijele svi članovi skupine, može se prepoznati da su svi dio iste skupine, iako se ne zna točno tko je tko. Da bi se ublažio rizik otkrivanja podataka, tu dolazi proširenje u obliku *l*-raznolikosti. *L*-raznolikost osigurava da će svaki atribut imati najmanje *l* različitih vrijednosti. Što znači da u gornjem primjeru svako svojstvo (kao što su dob i spol) u toj grupi mora imati barem *l* različitih vrijednosti za sva svojstva. Svaki put kada napadač pokuša zaključiti nešto o određenoj osobi koristeći podatke, zbog *l*-raznolikosti će biti nesiguran jer će svaka grupa sadržavati različite vrijednosti za svako svojstvo.[54]

[53] Google. „Privatnost i uvjeti“. Dostupno na: https://policies.google.com/technologies/anonymization?hl=hr
[54] Radna skupina za zaštitu podataka. „Mišljenje 05/2014 o tehnikama anonimizacije“. (2014). Dostupno na: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_hr.pdf , str. 18

3.2.3. *T*-bliskost

Ali ni *l*-raznolikost nije savršena pa je proširena modelom *t*-bliskosti koja pomaže da anonimizirani podaci izgledaju više kao originalni podaci. To znači da se pokušava očuvati raspodjela vrijednosti svojstava iz originalnih podataka u anonimiziranih podacima, ali i osigurati da svaka vrijednost svojstava bude zastupljena dovoljno puta kako bi se zadržala početna raspodjela svojstava. Drugim riječima, ne bi bilo dobro da bude premalo ili previše ljudi s istom vrijednošću svojstva u skupini.[55] *T*-bliskost kombinira očuvanje originalnog izgleda podataka i zaštitu privatnosti kako bi se postigla ravnoteža između ta dva važna aspekta.

3.2.4. Diferencijalna privatnost

Diferencijalna privatnost je matematička definicija privatnosti. Diferencijalno privatan algoritam jamči da se njegovo ponašanje gotovo ne mijenja kada se jedan zapis pridruži ili napusti skup podataka. Gotovo je jednako vjerojatno da će se isti podaci koje će algoritam ispisati iz skupa podataka koji sadrži podatke nekog pojedinca ispisati i iz skupa podataka koje ne sadrži podatke o tom pojedincu.[56]

Diferencijalna privatnost osigurava da čak i ako netko postavi upite o podacima, neće biti moguće jednostavno zaključiti je li određeni zapis prisutan u bazi ili ne. Osnovna ideja iza diferencijalne privatnosti je da prisutnost ili odsutnost bilo kojeg pojedinačnog zapisa u bazi podataka ili skupu podataka ne bi trebala biti primjetna kada se gledaju odgovori koji se vraćaju na upite. Ova ideja štiti privatnost pojedinaca jer čini teškim ili gotovo nemogućim istraživanje pojedinačnih informacija samo na temelju upita i odgovora.[57]

[55] Mišljenje o tehnikama anonimizacije, n. dj., str. 18
[56] Harvard University. „Differential Privacy“. Dostupno na: https://privacytools.seas.harvard.edu/differential-privacy
[57] Domingo-Ferrer, Josep; Sanchez, David; Soria-Comas, Jordi. „Database Anonymization: Privacy Models, Data Utility, and Microaggreegation-based Inter-model Connections“. *Synthesis Lectures on Information Security Privacy and Trust* 8 (1) (2016): 1-136. Dostupno na: https://www.researchgate.net/publication/290229262_Database_Anonymization_Privacy_Models_Data_Utility_and_Microaggregation-based_Inter-model_Connections

3.3. Tehnike anonimizacije

Kako bi se lakše provele tehnike anonimizacije, podaci se često organiziraju u obliku tablica, gdje svaki stupac predstavlja određeno svojstvo ili atribut. Postoje tri kategorije atributa:

- identifikator – služe jedinstvenoj identifikaciji pojedinca (npr. OIB – osobni identifikacijski broj),
- kvazi-identifikator – mogu pomoći u identifikaciji pojedinca u kombinaciji s informacijama iz drugih izvora, i
- osjetljivi atributi – sadrže osjetljive informacije koje ne smiju biti objavljene.[58]

Identifikatori su osnovni identifikatori pojedinca koji najbrže dovode do direktne identifikacije pa se obično prvi uklanjaju iz skupa podataka. To je naprimjer osobni identifikacijski broj, jer svaka osoba ima svoj jedinstveni OIB. Kvazi-identifikatori sami po sebi neće dovesti do direktne identifikacije, ali ih je moguće koristiti u kombinaciji s informacijama iz drugih izvora kako bi se osoba identificirala. Osjetljivi atributi mogu biti podaci kao što su zdravstvene informacije ili bilo koja druga privatna pitanja. Glavni cilj anonimizacije je zaštititi ove podatke od neovlaštenog pristupa i curenja osjetljivih informacija.[59]

3.3.1. Pseudonimizacija

Opća uredba o zaštiti podataka definira pseudonimizaciju kao „obradu osobnih podataka na način da se osobni podaci više ne mogu pripisati određenom ispitaniku bez uporabe dodatnih informacija, pod uvjetom da se takve dodatne informacija drže odvojeno te da podliježu tehničkim i organizacijskim mjerama kako bi se osiguralo da se osobni podaci ne mogu pripisati pojedincu čiji je identitet utvrđen ili se može utvrditi“.[60] Tijekom pseudonimizacije izravni identifikator mijenja se pseudonimom, naprimjer nasumičnim brojem. Tablica s izvornim podacima čuva se na odvojenom mjestu te povezuje identifikatore i pseudonime. Od anonimizacije se razlikuje u tome što se nakon anonimizacije anonimizirani podaci više ne smatraju osobnim podacima te ih nije moguće povezati s određenim pojedincem, dok se nakon pseudonimizacije podaci i dalje smatraju osobnim podacima i moguće ih je povezati s

[58] Bradić-Martinović, Aleksandra; Zdravković, Aleksandar. „Zaštita privatnosti – anonimizacija podataka“. V naučni skup USPON, Beograd (2013): 206-213. Dostupno na: https://www.academia.edu/45437258/Za%C5%A1tita_privatnosti_anonimizacija_podataka_Privacy_protection_data_anonymization , str. 208
[59] Isto, str. 208.
[60] Opća uredba o zaštiti podataka, n.dj., čl. 4

određenim pojedincem. Neke od tehnika pseudonimizacije su tokenizacija, pseudonimizacija šifriranjem i pseudonimizacija funkcijom sažimanja.[61]

3.3.1.1. Tokenizacija

Tokenizacija se često koristi u financijskom sektoru gdje služi za zamjenu identifikacijskih brojeva kartice tokenima (nevezanim vrijednostima koje nisu matematički izvedene iz originalnih podataka).[62] Tijekom tokenizacije se pseudonim (token) generira neovisno o izvornim podacima, pa tako nastaju nasumično generirani brojevi. Da bi se povezali tokeni i izravni identifikatori, potrebna je tablica u kojoj su zapisane njihove veze.[63] Slika 1 prikazuje primjer tokenizacije u kojem se kombinacija spola i datuma rođenja osobe zamjenjuje tokenom. Preko tokena nije moguće doći do izvornih podataka bez tablice koja služi kao ključ.

Spol	Datum rođenja	Predmet	Ocjena
Ž	5.9.1998.	Hrvatski jezik	5
Ž	7.8.1998.	Engleski jezik	2
M	30.10.2000.	Matematika	4
Ž	3.4.2000.	Matematika	3
M	1.9.1999.	Matemaika	5
M	27.3.1999.	Engleski jezik	5
M	6.10.1998.	Biologija	4

→

Token	Predmet	Ocjena
65896	Hrvatski jezik	5
46852	Engleski jezik	2
31687	Matematika	4
15968	Matematika	3
13456	Matemaika	5
57489	Engleski jezik	5
25478	Biologija	4

Token	Spol	Datum rođenja
65896	Ž	5.9.1998.
46852	Ž	7.8.1998.
31687	M	30.10.2000.
15968	Ž	3.4.2000.
13456	M	1.9.1999.
57489	M	27.3.1999.
25478	M	6.10.1998.

Slika 1. Tokenizacija

3.3.1.2. Pseudonimizacija šifriranjem

Slično tokenizaciji, pseudonimizacija šifriranjem temelji se na šifriranju tajnim ključem. Identifikatori se šifriraju čime šifrirana vrijednost postaje pseudonim. Za razliku od tokenizacije, nije potrebna cijela tablica da bi se povezali pseudonimi s identifikatorima, već

[61] CARNET. „Anonimizacija i pseudonimizacija podataka“. Dostupno na: https://www.cert.hr/wp-content/uploads/2018/08/anonimizacija_i_pseudonimizacija_podataka.pdf
[62] Mišljenje o tehnikama anonimizacije, n.dj., str. 21
[63] CARNET. Anonimizacija i pseudonimizacija podataka, n.dj., str. 15

samo tajni ključ.[64] Bez tajnog ključa dešifriranje je nemoguće, a pristup originalnim identifikatorima onemogućen. Slika 2 prikazuje pseudonimizaciju šifriranjem u kojoj su vrijednosti atributa 'spol' i 'datum rođenja' šifrirani tajnim ključem bez kojeg nije moguće dešifrirati podatke.

Spol	Datum rođenja	Predmet	Ocjena
Ž	5.9.1998.	Hrvatski jezik	5
Ž	7.8.1998.	Engleski jezik	2
M	30.10.2000.	Matematika	4
Ž	3.4.2000.	Matematika	3
M	1.9.1999.	Matemaika	5
M	27.3.1999.	Engleski jezik	5
M	6.10.1998.	Biologija	4

→

Šifriran spol i datum rođenja	Predmet	Ocjena
GDsgsDgSDGsdgdGS	Hrvatski jezik	5
DGFaČKfafčaLSJh	Engleski jezik	2
ĆAGhAOighAOPshgAĐO	Matematika	4
čKJABgsćLAJSbćgl	Matematika	3
OAJghćaOSghć	Matemaika	5
ĆAigžaPIghž	Engleski jezik	5
APsgjĆASLjgbćl	Biologija	4

Slika 2. Tokenizacija šifriranjem

3.3.1.3. Pseudonimizacija funkcijom sažimanja

Pseudonimizacija funkcijom sažimanja temelji se na obradi identifikatora funkcijom sažimanja te se kao pseudonim koristi izlaz funkcije sažimanja. Obično se ne koriste posebni podaci za povezivanje pseudonima i identifikatora, kao što su tablice ili tajni ključevi, čija je odsutnost vidljiva na slici 3. U ovom pristupu, za povezivanje identifikatora sa pseudonimima, prvo je potrebno imati početne identifikatore te iz njih generirati pseudonim. Razvijene su razne dodatne varijante pseudonimizacije funkcijom sažimanja koje su sposobnije od izvorne varijante zaštititi podatke od reidentifikacije: pseudonimizacija funkcijom sažimanja s dodanom vrijednosti, s tajnim ključem, uz pohranu tajnog ključa i uz uništavanje tajnog ključa.[65]

Spol	Datum rođenja	Predmet	Ocjena
Ž	5.9.1998.	Hrvatski jezik	5
Ž	7.8.1998.	Engleski jezik	2
M	30.10.2000.	Matematika	4
Ž	3.4.2000.	Matematika	3
M	1.9.1999.	Matemaika	5
M	27.3.1999.	Engleski jezik	5
M	6.10.1998.	Biologija	4

→

Spol i datum rođenja obrađeni funkcijom sažimanja	Predmet	Ocjena
afhadrhddhSGsLDKčdjgaAčakjg	Hrvatski jezik	5
GAFDBSDčkjagčksdjgČagdgdas	Engleski jezik	2
kjhčkasjdhgoHŠSOHGČSDFHSJČ	Matematika	4
pahsphSOHŠIOAHSOHhjčkjčgka	Matematika	3
OAJghćaOSghćpiGčajčkjgbčkJK	Matemaika	5
ĆAigžaPIghžLsjgČKJADGčkksdb	Engleski jezik	5
APsgjĆASLjgbćlfklahsdkgHOAH	Biologija	4

Slika 3. Pseudonimizacija funkcijom sažimanja

[64] Isto, str. 16
[65] Isto, str. 16-18

3.3.2. Generalizacija

Generalizacija ili tehnika poopćavanja temelji se na promijeni vrijednosti atributa kroz njihovo poopćavanje. Tako se, naprimjer, datum rođenja može zamijeniti godinom rođenja (kao što je vidljivo na slici 4), vrijednost plaće može se zamijeniti rasponom vrijednosti u koje taj iznos ulazi, mjesto rođenja (npr. Zagreb) može se zamijeniti državom (npr. Hrvatska) i slično. Reidentifikaciju otežava činjenica da poopćavanjem više osoba dijeli istu vrijednost atributa.[66] Generalizacija omogućuje postizanje *k*-anonimnosti.[67]

Spol	Datum rođenja	Predmet	Ocjena
Ž	5.9.1998.	Hrvatski jezik	5
Ž	7.8.1998.	Engleski jezik	2
M	30.10.2000.	Matematika	4
Ž	3.4.2000.	Matematika	3
M	1.9.1999.	Matemaika	5
M	27.3.1999.	Engleski jezik	5
M	6.10.1998.	Biologija	4

→

Spol	Datum rođenja	Predmet	Ocjena
Ž	1998.	Hrvatski jezik	5
Ž	1998.	Engleski jezik	2
M	2000.	Matematika	4
Ž	2000.	Matematika	3
M	1999.	Matemaika	5
M	1999.	Engleski jezik	5
M	1998.	Biologija	4

Slika 4. Generalizacija

3.3.3. Randomizacija

Randomizacija ili tehnika nasumične izmjene podataka temelji se na promjeni istinitosti podataka tako da podaci i dalje nose korisne informacije, ali se smanjuje veza između podatka i pojedinca na kojeg se odnosi. Neke od tehnika nasumične izmjene podataka su tehnika dodavanja šuma i tehnika permutacije.[68] Ova tehnika održava vrijednost podataka, ali dodaje element slučajnosti radi otežavanja povratnog povezivanja podataka s pojedincima.

3.3.3.1. Dodavanje šuma

Tehnika dodavanja šuma dopušta vrijednostima da i dalje imaju istu distribuciju u skupu, ali individualne vrijednosti više nisu precizne zbog promjene vrijednosti nekog atributa. U to se ubraja dodavanje ili oduzimanje nekoliko dana datumima i slično.[69] Slika 5 prikazuje dodavanje šuma dodavanjem 5 dana na datume rođenja. Tehniku dodavanja matematičkog šuma podacima opisuje diferencijalna privatnost. Da bi se u nekom skupu podataka postigla diferencijalna privatnost, dodaje se šum u taj skup podataka. Zbog dodavanja šuma podaci

[66] Isto, str. 9-10
[67] Google. Privatnost i uvjeti, n.dj.
[68] CARNET. Anonimizacija i pseudonimizacija podataka, n.dj., str.10
[69] Isto, str. 10

mogu izgubiti na koristi koju imaju za analize, ali može se dogoditi i da izlaz određenog algoritma izgleda jednako kad su podaci o nekoj osobi uključeni kao i kad su izostavljeni.[70]

Spol	Datum rođenja	Predmet	Ocjena
Ž	5.9.1998.	Hrvatski jezik	5
Ž	7.8.1998.	Engleski jezik	2
M	30.10.2000.	Matematika	4
Ž	3.4.2000.	Matematika	3
M	1.9.1999.	Matemaika	5
M	27.3.1999.	Engleski jezik	5
M	6.10.1998.	Biologija	4

Spol	Datum rođenja	Predmet	Ocjena
Ž	10.9.1998.	Hrvatski jezik	5
Ž	12.8.1998.	Engleski jezik	2
M	4.11.2000.	Matematika	4
Ž	8.4.2000.	Matematika	3
M	6.9.1999.	Matemaika	5
M	1.4.1999.	Engleski jezik	5
M	11.10.1998.	Biologija	4

Slika 5. Dodavanje šuma

3.3.3.2. Permutacija

Tehnika permutacije zamjenjuje mjesta vrijednostima atributa i tako smanjuje povezanost podataka unutar jednog zapisa. Distribucija u skupu ostaje jednaka jer se nisu promijenili podaci o određenom atributu, već su samo izmiješani.[71] Slika 6 prikazuje primjer permutacije vrijednosti atributa predmeta. Iako su vrijednosti izmiješane, distribucija je ostala jednaka (u obje tablice se predmeti pojavljuju jednak broj puta, npr, vrijednost atributa 'matematika' se pojavljuje tri puta). Kao i dok tehnike dodavanja šuma, moguće je iz skupa podataka izdvojiti podatke o pojedincu, ali s obzirom na izmiješanost atributa, zapisi su manje pouzdani te je moguće izvući zaključke iz skupa podataka.[72]

Spol	Datum rođenja	Predmet	Ocjena
Ž	5.9.1998.	Hrvatski jezik	5
Ž	7.8.1998.	Engleski jezik	2
M	30.10.2000.	Matematika	4
Ž	3.4.2000.	Matematika	3
M	1.9.1999.	Matemaika	5
M	27.3.1999.	Engleski jezik	5
M	6.10.1998.	Biologija	4

Spol	Datum rođenja	Predmet	Ocjena
Ž	5.9.1998.	Matematika	5
Ž	7.8.1998.	Matematika	2
M	30.10.2000.	Engleski jezik	4
Ž	3.4.2000.	Hrvatski jezik	3
M	1.9.1999.	Engleski jezik	5
M	27.3.1999.	Biologija	5
M	6.10.1998.	Matematika	4

Slika 6. Permutacija

[70] Google. Privatnost i uvjeti, n.dj.
[71] CARNET. Anonimizacija i pseudonimizacija podataka, n.dj., str. 11
[72] Mišljenje o tehnikama anonimizacije, n.dj., str. 13-14

3.4. Problemi pri anonimizaciji

Postoji mnogo načina kako se mogu skrivati podaci da bi ostali anonimni, ali nisu svi podjednako pouzdani. Tri su ključna rizika povezana s anonimizacijom: izdvajanje, povezivost i izvođenje zaključaka. Rizik izdvajanja odnosi se na mogućnost da netko može prepoznati nekoliko ili čak sve zapise koji otkrivaju identitet osobe u skupu podataka, odnosno izolirati i identificirati nekoliko ili sve zapise u skupu podataka koji su povezani s određenom osobom. To znači da unatoč pokušajima anonimizacije, netko bi mogao lako otkriti koji su podaci pripadali određenoj osobi. Rizik povezivosti odnosi se na mogućnost da se povežu barem dva zapisa koja se tiču iste osobe. To znači da, iako informacije možda ne otkrivaju identitet pojedinca, kad se povežu s drugim informacijama, identitet bi mogao postati jasan. Povezani zapisi čak se ne moraju nalaziti u istom skupu podataka, mogu biti i dio različitih skupova. Rizik izvođenja zaključaka odnosi se na mogućnost da netko na temelju jednih atributa može zaključiti koja je vrijednost drugih atributa s visokom vjerojatnošću. Tehnike anonimizacije moraju biti dovoljno snažne da spriječe ove rizike, ali istovremeno i dovoljno korisne da podaci zadrže svoju vrijednost za analizu i istraživanje.[73]

[73] Isto, str. 11-12

4. Alati i usluge za anonimizaciju

Upotreba umjetne inteligencije (UI) u anonimizaciji zapisa ima veliki potencijal za zaštitu privatnosti pojedinaca i sigurnosti podataka. Koristi se kako bi se uklonili, zamaglili ili zamijenili identifikacijski podaci iz skupa podataka, čime se osigurava da se ne može jednostavno identificirati pojedinca ili osobu na koju ti podaci upućuju. Svrha umjetne inteligencije je u računalu razviti sposobnost obavljanja zadataka koji zahtijevaju određenu razinu inteligencije, stoga se temelji na neuronskim mrežama koje „uče“ (strojno učenje).[74] Umjetna inteligencija omogućuje računalnim sustavima sposobnost percipiranja svog okruženja, prikupljanje informacija iz okoline, razumijevanje informacija koje primaju, obradu tih informacija te donošenje odluka prema informacijama koje primaju i rješavanje problema. Na taj način ona igra ključnu ulogu u anonimizaciji podataka jer omogućuje tehničkim sustavima da obrade osjetljive informacije na način koji održava privatnost pojedinaca percipirajući i prepoznajući osjetljive podatke.

4.1. Umjetna inteligencija

4.1.1. UAI Anonymizer

Jedan od alata za anonimizaciju uz korištenje umjetne inteligencije je Understand.AI Anonymizer koji koristi umjetnu inteligenciju u svrhu prikupljanja podataka u sklopu prometa autonomnih vozila. Autonomna vozila opremljena senzorima prikupljaju podatke o okolini u kojoj se vozila kreću (npr. informacije o drugim vozilima, pješacima, prometnim znakovima i semaforima), a kako ne bi došlo do prikupljanja osobnih i osjetljivih podataka, UAI Anonymizer automatski uz pomoć umjetne inteligencije prepoznaje i anonimizira prikupljene osjetljive podatke: zamućuje lica i registarske pločice vozila.[75] Iako lica i registarske pločice ostaju anonimizirane, alat i dalje zadržava sve ostale korisne informacije u svrhu analize i obrade, kao što su naprimjer informacije o brzini vozila.

[74] Hrvatska enciklopedija, mrežno izdanje. „umjetna inteligencija“. *Leksikografski zavod Miroslav Krleža* (2021). Dostupno na: https://www.enciklopedija.hr/natuknica.aspx?ID=63150

[75] ASAM. „UAI Anonymizer“. Dostupno na: https://www.asam.net/members/product-directory/detail/uai-anonymizer/

4.1.2. VEIL.AI

Drugi primjer alata koji koristi umjetnu inteligenciju za anonimizaciju je VEIL.AI. VEIL.AI nastao je u Finskoj i koristi se u zdravstvu za anonimizaciju osjetljivih zdravstvenih podataka koji se pretvaraju u sintetičke podatke. Anonimizacija se provodi na svakom izvoru podataka, a u daljnji proces obrade se upućuju samo anonimizirani podaci koji više ne podliježu Općoj uredbi o zaštiti podataka što pojednostavnjuje dijeljenje i prijenos osjetljivih znanstvenih podataka. Umjetna inteligencija u alatu generira sintetičke podatke sa sličnim osobinama kao i stvarni podaci te time ne dolazi do izlaganja stvarnih podataka, a ujedno je očuvana i kvaliteta podataka nakon anonimizacije pa tako zadržavaju svoju korisnost za daljnje analize.[76]

4.2. Alati za anonimizaciju

Postoji nekoliko alata za anonimizaciju, s razvojem tehnologije i umjetne inteligencije taj se broj povećava, a odabir odgovarajućeg trebao bi se temeljiti na osobinama posla koji je potrebno obaviti. Moguće su velike razlike u opsegu i kompleksnosti posla. Različita je i količina informacija koje treba anonimizirati. Također je potrebno i obratiti pozornost na format zapisa koji se anonimizira.

4.2.1. Diffix

Diffix, besplatan alat otvorenog kôda, koristi „snažnu“ anonimizaciju u obradi strukturiranih podataka. U svrhu izrade ovog rada testirana je verzija „aplikacije za obuku“ (engl. *training app*).[77] Nad bazom podataka o vožnjama taksija kroz New York postavljen je upit (unaprijed programiran u uputama za uporabu) koji broji koliko je vožnji u svakom satu prošlo prostorom od metra kvadratnog. Ovaj primjer oslikava korištenje numeričke i tekstualne generalizacije u anonimizaciji. Na Slici 7 prikazano je sučelje aplikacije i dio rezultata upita. Ovim kodom analizirani su i grupirani polasci taksija prema vremenu (engl. *datetime*, *hour*), zemljopisnoj širini (engl. *latitude*) i dužini (engl. *longitude*). Za dohvaćanje podataka iz baze 'jan08' korišteni su upiti u Diffix SQL jeziku. Na traženi upit dobiveno je 2.055 rezultata koji zadovoljavaju uvjete koji su postavljeni upitom. Prvi redak 'SELECT substring

[76] VEIL.AI. „Unlock the power of data“. Dostupno na: https://veil.ai/why-veil-ai/

[77] Open Diffix. „Play with Diffix“. Dostupno na: http://www.open-diffix.org/en/play/

(cast(pickup_datetime) AS text), 1, 13 AS hour' služi za ekstrahiranje prvih 13 znakova iz stupca 'pickup_datetime' (iz baze jan08) te ih pretvara u tekst ('string') i u tablicu upisuje pod stupcem naziva 'hour'. Sljedeća dva retka uz pomoć 'diffix.round_by' zaokružuju širinu i dužinu na dvije decimale (0.01). 'Count(*)' računa broj redaka koji se podudaraju sa zadanim uvjetima u grupiranim podacima.

Diffix SQL

```
    diffix.round_by(pickup_latitude,0.01)
      AS lat,
    diffix.round_by(pickup_longitude,0.01)
      AS lon,
    count(*)
FROM jan08
GROUP BY 1,2,3
ORDER BY 1,2,3
```

Native SQL

```
SELECT substring(cast(pickup_datetime
      AS text),1,13) AS hour,
    diffix.round_by(pickup_latitude,0.01)
      AS lat,
    diffix.round_by(pickup_longitude,0.01)
      AS lon,
    count(*)
FROM jan08
GROUP BY 1,2,3
```

Database: taxi Trust Mode: trusted Run

2055 rows in 7.487 seconds

hour	lat	lon	count
*	None	None	5216
2013-01-07 23	0.0	0.0	36
2013-01-07 23	40.640	-73.790	16
2013-01-07 23	40.640	-73.780	32
2013-01-07 23	40.650	-73.790	45
2013-01-07 23	40.650	-73.780	60
2013-01-07 23	40.690	-73.990	6
2013-01-07 23	40.690	-73.980	8
2013-01-07 23	40.710	-74.010	35
2013-01-07 23	40.710	-74.0	10
2013-01-07 23	40.710	-73.960	9
2013-01-07 23	40.710	-73.950	9
2013-01-07 23	40.720	-74.010	42
2013-01-07 23	40.720	-74.0	65
2013-01-07 23	40.720	-73.990	118
2013-01-07 23	40.720	-73.980	14

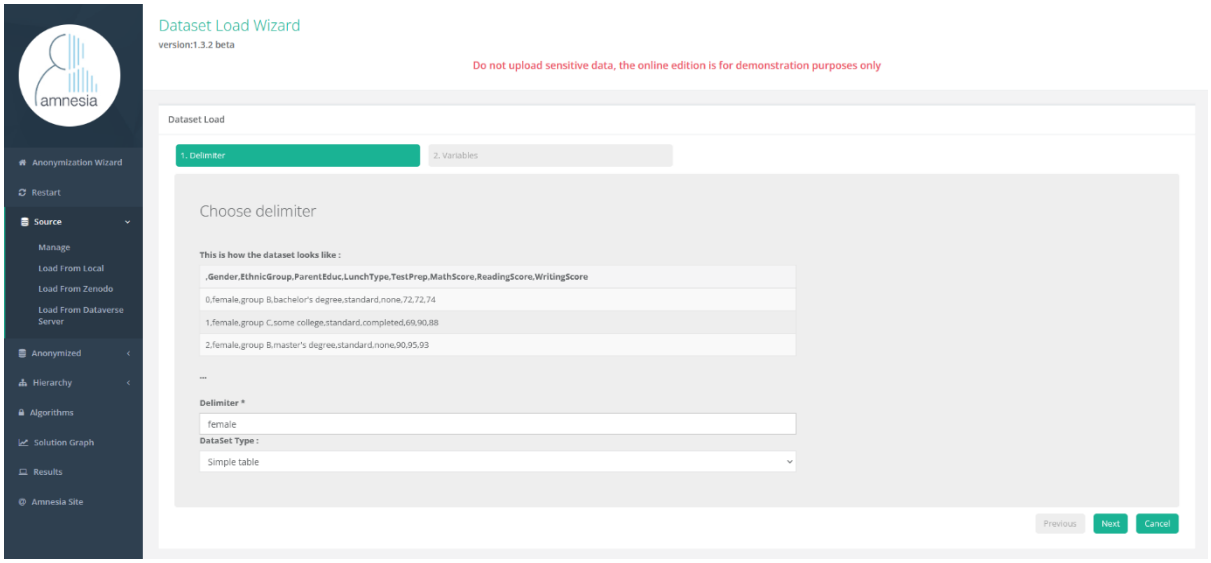
4654 rows in 7.084 seconds

hour	lat	lon	count	abs : rel
2013-01-07 19	40.760	-73.960	1	
2013-01-07 20	40.650	-73.790	1	
2013-01-07 22	40.650	-73.780	1	
2013-01-07 22	40.770	-73.980	1	
2013-01-07 22	40.770	-73.870	1	
2013-01-07 22	40.770	-73.860	1	
2013-01-07 23	0.0	0.0	37	-1 : -2.7%
2013-01-07 23	25.580	-79.120	1	
2013-01-07 23	40.540	-73.890	1	
2013-01-07 23	40.570	-73.990	1	
2013-01-07 23	40.630	-73.950	1	
2013-01-07 23	40.640	-73.790	20	-4 : -20.0%
2013-01-07 23	40.640	-73.780	32	0 : 0.0%
2013-01-07 23	40.650	-73.960	1	
2013-01-07 23	40.650	-73.80	1	
2013-01-07 23	40.650	-73.790	46	-1 : 2.17%

Slika 7. Testiranje Diffixa

4.2.2. Amnesia

Amnesia je alat za anonimizaciju podataka koji uklanja identifikacijske informacije iz podataka koji su mu zadani. Osim što uklanja izravne identifikatore (npr. imena) iz podataka, još i transformira sekundarne identifikatore (poput datuma rođenja) kako bi se smanjila mogućnost povezivanja s drugim izvorima informacija i reidentifikacije. Amnesia podržava *k*-anonimnost i njenu slabiju verziju *km*-anonimnost[78]. Tehnike koje koristi su generalizacija i potiskivanje[79], a hijerarhiju vrijednosti za generalizaciju može definirati korisnik ili je alat sam izrađuje. Alat se može pronaći u dva oblika, kao računalni program ili online servis.[80] U testiranju online oblika, tablica s podacima o rezultatima testova studenata preuzeta je s kaggle-a[81]. Testiranje nije uspjelo jer je program u svim pokušajima javio pogrešku u unutarnjem serveru. Slika 8 prikazuje početnu stranicu koja se pojavi nakon što se na sustav učita baza podataka.



Slika 8. Amnesia

[78] Dok se u *k*-anonimnosti razmatra broj *n* kvazi-identifikatora, *km*-anonimnost ograničava sigurnost protiv napadača koji poznaju samo *m* od tih *n* kvazi-identifikatora. Koristi se za analizu visokodimenzionalnih podataka koji imaju mnogo atributa te osigurava da se, bez obzira na sve te atribute, svaka kombinacija nekoliko odabranih atributa pojavljuje u podacima bar nekoliko puta te tako sprječava da bilo koji mali skup atributa bude jedinstven i lak za identificiranje. (https://amnesia.openaire.eu/Scenarios/AmnesiaKMAnonymityTutorial.pdf)

[79] Potiskivanje uključuje potpuno uklanjanje vrijednosti atributa iz skupa podataka, najčešće irelevantnih atributa koji nisu potrebni za određenu analizu. (https://www.immuta.com/blog/k-anonymity-everything-you-need-to-know-2021-guide/)

[80] OpenAIRE. „Amnesia – Anonymize your data before publishing“. Dostupno na: https://www.openaire.eu/amnesia-guide

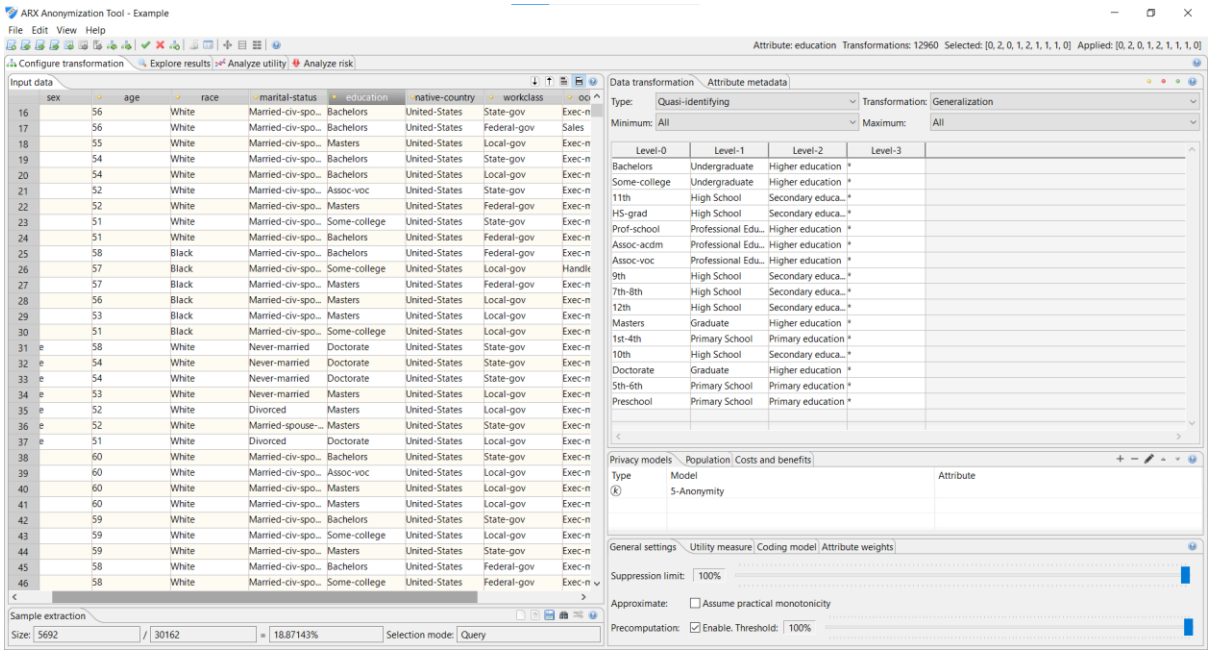
[81] kaggle. „Datasets“. Dostupno na: https://www.kaggle.com/datasets

4.2.3. ARX Data Anonymization Tool

ARX Data Anonymization Tool (u daljnjem tekstu ARX) je višeplatformski alat otvorenog kôda koji podržava nekoliko različitih modela anonimizacije kao što su *k*-anonimnost i njene varijante, *l*-raznolikost i *t*-bliskost te diferencijalna privatnost.[82] Koristi se za anonimizaciju osjetljivih osobnih podataka jer podržava velik izbor modela privatnosti i rizika, metoda za transformaciju podataka i metoda za analizu upotrebljivosti izlaznih podataka. ARX se može koristiti za mnoge oblike anonimizacije, od malih projekata na kojima se treniraju osobe za provođenje anonimizacije do velikih istraživačkih projekata. Sveobuhvatno i intuitivno grafičko korisničko sučelje olakšava rad na velikim skupovima podataka i milijunima zapisa koje alat može podržati.[83]

ARX transformira strukturirane (tablične) skupove osobnih podataka pridržavajući se definiranih modela privatnosti i rizika korištenjem odabranih metoda iz područja anonimizacije podataka. Ti modeli privatnosti i rizika ublažavaju napade na anonimizirane podatke koji mogu dovesti do povreda privatnosti. U ARX-u se mogu ukloniti izravni identifikatori kao što su imena i nazivi iz skupova podataka te provesti daljnja ograničenja na neizravne identifikatore (ključeve), odnosno one atribute koji pojedinca ne identificiraju izravno, ali u kombinaciji s drugim neizravnim identifikatorima mogu tvoriti identifikator koji se može koristiti za daljnja identificiranja.[84] Slika 9 pokazuje primjer kako su osobni podaci generalizirani korištenjem ARX-a. Na lijevoj polovici ekrana prikazana je konačna tablica koja nastaje generalizacijom, a desna polovica ekrana pokazuje koji su podaci zamijenjeni ključevima na tri različite razine.

[82] Sartor, Nicolas. „Top 5 Free Data Anonymization Tools“. *Aircloak* (2019). Dostupno na: https://aircloak.com/top-5-free-data-anonymization-tools/
[83] ARX – Data Anonymization Tool. Dostupno na: https://arx.deidentifier.org/ (26.4.2023.)
[84] Isto.

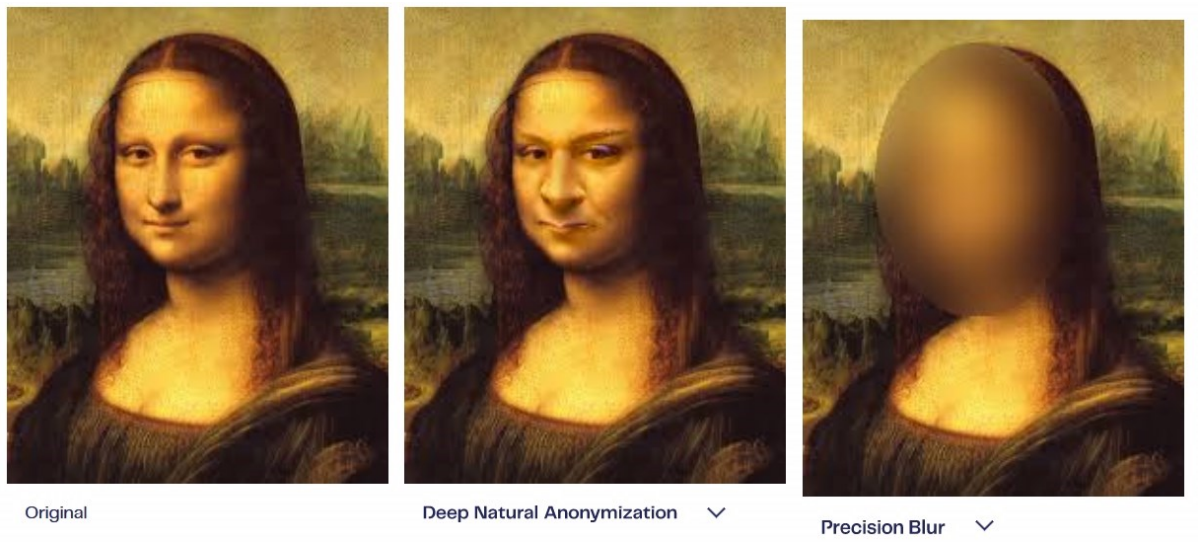


Slika 9. ARX

4.2.4. brighter AI

Brighter AI softver je za anonimizaciju slikovnih i video sadržaja koji uz pomoć umjetne inteligencije anonimizira lica i registracijske tablice. Koristi dva rješenja za skrivanje značajki lica i oznaka na registracijskim tablicama: Precision Blur i Deep Natural Anonymization. Uz pomoć Precision Blura automatski se otkrivaju i zamućuju lica i tablice u videu ili na slici, dok Deep Natural Anonymization automatski otkriva lice i tablice te generira sintetičke podatke kao zamjenu.[85] Softver je jednostavan za korištenje, potrebno je samo učitati fotografiju i odabrati način anonimizacije i fotografija je spremna. Slika 10 prikazuje Deep Natural Anonymization na primjeru ljudskog lica, odnosno portreta Mona Lise.

[85] brighter AI. „Use anonymization to monetize data and be compliant“. Dostupno na: https://brighter.ai/product/#demo

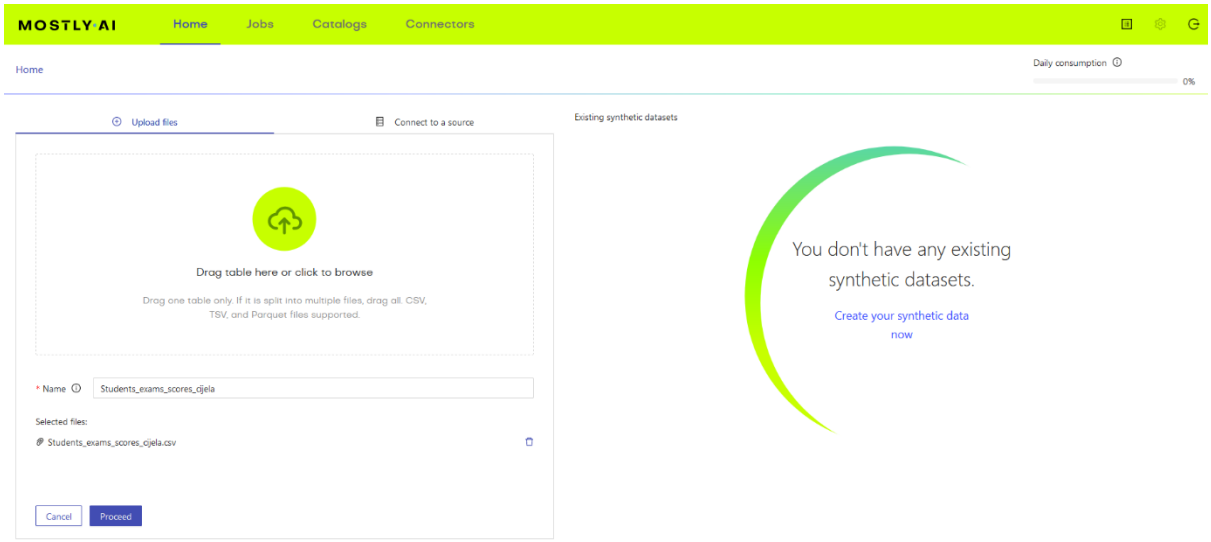


Slika 10. Deep Natural Anonymization i Precision Blur na primjeru Mona Lise

4.2.5. Mostly AI

Mostly AI još je jedan alat za anonimizaciju koji uz pomoć umjetne inteligencije generira sintetičke podatke kako bi se sačuvale informacije i vrijednost skupa podataka za daljnje analize. Generator stvara automatizirana i interaktivna izvješća o svakom skupu podataka koji je generirao iz strukturiranih poslovnih podataka u sintetičke podatke, kojima je lakše manipulirati.[86] Jednostavan je za korištenje, u sam program upisane su upute što i gdje upisati koje se mogu konzultirati u svakom trenutku bez potrebe da se izlazi iz programa. U testiranju programa ponovno je iskorištena baza podataka s kaggle-a o uspjesima studenata na ispitima. Rezultat koji nastaje je nova tablica sa sintetičkim podacima spremna za preuzimanje. Za svaku tablicu mjeri se točnost, odnosno koliko dobro model i novi podaci predstavljaju statistiku izvornih podataka. Što je veći postotak, bolji je i model, i sintetički podaci su točnije generirani. Slika 11 prikazuje početnu stranicu na koju se učitava baza podataka, a nakon toga slijedi samo odabir odredišta i vrste datoteke sa sintetiziranim podacima.

[86] Mostly.AI. „Data anonymization – synthetic data for maximum privacy and utility“. Dostupno na: https://mostly.ai/use-case/data-anonymization-with-synthetic-data



Slika 11. Mostly AI

5. Projekt CabAnon

5.1. LINC

Projekt CabAnon razvio je i vodio tim francuskog LINC-a (fr. *Laboratoire d'Innovation Numérique de a CNIL*) koji pripada pod francuski CNIL (fr. *Commission nationale de l'informatique et des libertés*, engl. *National Commission on Informatics and Liberty*, hrv. Nacionalno povjerenstvo za informatiku i slobode), koji se bavi kontrolom zaštite osobnih podataka u Francuskoj, odnosno osigurava primjenu zakona o privatnosti podataka na prikupljanje, obradu, pohranjivanje i korištenje osobnih podataka.

Cilj projekta CabAnon bio je kroz interaktivnu vizualizaciju podataka procijeniti učinkovitost anonimiziranih skupova podataka i kvantificirati gubitak informacija. Tim iz LINC-a odlučio je osporiti teoriju da su svi anonimizirani skupovi podataka beskorisni, pa su kao primjer stvarnih skupova podataka uzeli podatke o rutama vožnje taksija u New Yorku, podatke koje je javno objavio sam Grad. Podatke su analizirali kako bi kroz konkretan primjer pokazali da postoji ravnoteža između korisnosti i anonimizacije. Sami podaci na početku su izazvali velike probleme jer su podaci o vremenu putovanja i mjestima polaska i dolaska otkrili identitete poznatih putnika.[87] Stručnjaci iz LINC-a odlučili su pokazati da je moguće anonimizirati te podatke i dobiti podatke koji se mogu analizirati i dati konkretne i točne zaključke.

Kako bi ocijenili učinkovitost anonimizacije skupa podataka, a da se istovremeno zadrži njegova učinkovitost, LINC je proveo eksperiment. Uzeli su Uberov[88] proces anonimizacije (umjesto da stvaraju svoj od nule) u koji su uključili koncept k-anonimnosti (s tim da je k=10, kako bi poboljšali vjerojatnost da će doći do očuvanja privatnosti). To bi značilo da su u istraživanje uključili samo putovanja koja su započela ili završila u zonama u kojima je bilo više od 10 putovanja koja su započela ili završila u istom području.[89]

Radili su na 3 skupa podataka: TLC (New York City Taxi & Limousine Commission) skupu podataka o putovanjima koji je javno dostupan na njihovoj službenoj web stranici te dva anonimizirana skupa podataka koja su izvedena iz TLC-a. Ti skupovi podataka, osim

[87] LINC. „CabAnon: exploring and visualising anonymized datasets“. (2017). Dostupno na: https://linc.cnil.fr/cabanon-exploring-and-visualizing-anonymized-datasets

[88] Uber je prijevoznička tvrtka. Putnici pomoću aplikacije naručuju prijevoz u određeno vrijeme na određenu adresu, a cijena prijevoza varira o uvjetima na cesti (npr. gužve), te se prilikom narudžbe može odabrati odgovarajući model prijevoza s predloženom cijenom. Aplikacija prikazuje podatke o vozaču, modelu automobila i broju registracijske pločice kako bi putnik bio siguran da je sjeo u pravi automobil te na kraju vožnje putnik ocjenjuje vozača i vozač putnika.

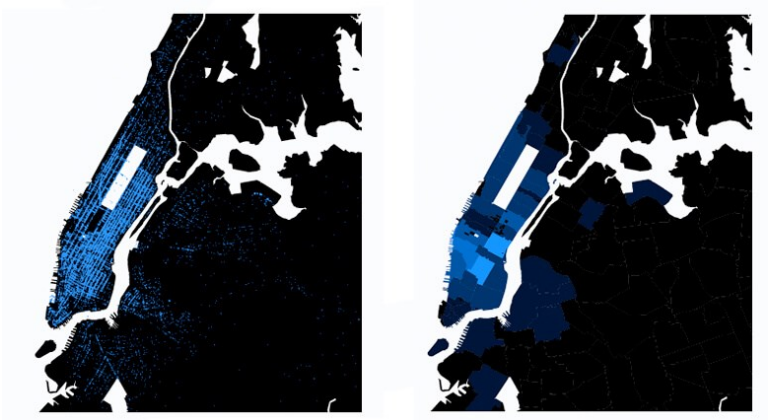
[89] Isto.

geografskih i vremenskih podataka o lokacijama polaska i dolaska, sadrže i podatke o udaljenosti putovanja, vremenskom trajanju putovanja i broju putnika u vozilu.

Projekt CabAnon obuhvaća četiri scenarija koji istražuju primjene za koje bi se mogli upotrijebiti podaci o vožnjama žutih taksija u New Yorku. Neke od primjena koje navode su brže pronalaženje slobodnog taksija ili poboljšanje prometne infrastrukture. Svaki od ta četiri scenarija oslanja se na jedan parametar kvalificiranja gradskog prometa:

1. s obzirom na gustoću prometa, putnik može prepoznati mjesto u blizini gdje će brzo pronaći taksi
2. s obzirom na broj putnika u taksiju, urbanisti mogu smisliti druga rješenja za organiziranje mobilnosti prometa
3. s obzirom na brzinu prometa, putnik može izbjeći prometne gužve, a urbanisti mogu identificirati mjesta na kojima se stvaraju prometne gužve
4. s obzirom na smjer prometa, urbanist može razumjeti kretanje putnika kroz grad i poboljšati sustave javnog prijevoza.[90]

Kako bi, naprimjer, utvrdili je li potrebno uvesti dodatne usluge prijevoza, stručnjaci iz LINC-a upotrijebili su podatke o granularnosti lokacije (poštanski brojevi) i vremena od 30 minuta. Slika 12 prikazuje razliku između neanonimiziranih i anonimiziranih podataka.



Slika 12. CabAnon - neanonimizirani i anonimizirani podaci[91]

[90] Isto.
[91] Le Carnot Télécom & Société numérique. „Les données anonymisées n'ont-elles aucune valeur?“. (2018). Dostupno na: https://www.carnot-tsn.fr/cabanon-teralab-donnees-anonymisees/

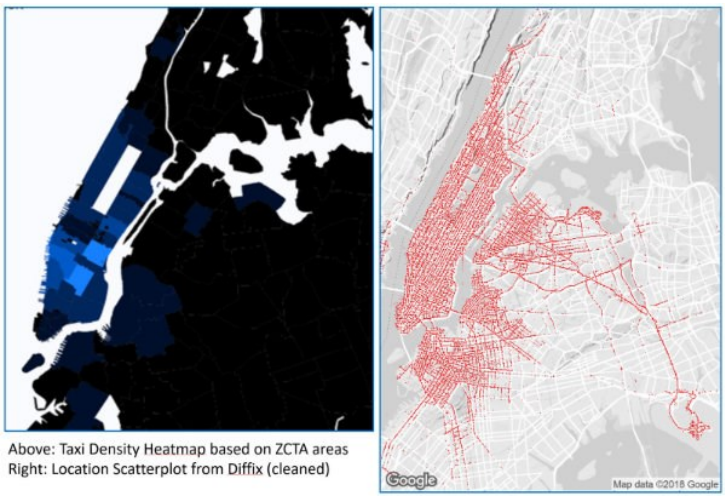
Slika s lijeve strane slike 12 predstavlja putovanja na temelju neanonimiziranih podataka. Slika s desne strane predstavlja interaktivnu kartu dobivenu metodom koropleta kod anonimiziranih podataka, kako bi se olakšalo čitanje skupa podataka. Metoda koropleta omogućuje vizualizaciju podataka tako da se uz pomoć varijable mijenja boja kojom su prikazani podaci. Na primjeru slike 12 vidi se da je gustoća prometa na određenom prometu prikazana različitim nijansama plave boje. Što je nijansa boje tamnija, to je veća gustoća prometa zabilježena na tom području. Ono što se odmah na prvu vidi je da su kod neanonimiziranih podataka označene lokacije polaska i dolaska te putanje taksija, dok se kod anonimiziranih podataka ne može odrediti s kojeg je mjesta neki taksi krenuo i kamo, već su GPS koordinate zamijenili ZCTA (engl. ZIP Code Tabulation Area) kodom (američkom verzijom poštanskog broja).[92] To znači da su položaji odlaska i dolaska taksija grupirani u područjima po nekoliko blokova. Čak ni to može biti nedovoljno da zajamči anonimnost vozača i putnika jer se u određenim satima, recimo noću, događa da samo nekoliko taksija vozi nekim područjem. Stoga je, osim prostorne, uvedena i vremenska degradacija, uzimajući segmente od 5, 25, 30 i 60 minuta, tako da nijedan vremenski segment nema manje od 10 osoba (zbog unaprijed postavljene k-anonimnosti).[93] Na ovaj se način gube pojedinačni podaci, ali kao skup podataka su i dalje iskoristivi, pogotovo, u ovom slučaju, u obliku karte. CNIL predlaže da ovakvi podaci mogu biti javno dostupni u anonimiziranom obliku.[94]

LINC preuzima klasičan pristup anonimizaciji – anonimizaciju su prilagodili određenom slučaju. U ovom slučaju bili su potrebni samo određeni podaci (geografski i vremenski podaci o lokacijama polaska i dolaska taksija, podaci o udaljenosti i trajanju putovanja, podaci o broju putnika u vozilu) pa su ostali podaci, kao naprimjer podaci o uplatama, vozačima i slično bili izbačeni. Da su i ti podaci bili potrebni, LINC bi morao preuzeti drugačiji pristup anonimizaciji.[95]

[92] LINC. „Can anonymised data still be useful?“. (2017). Dostupno na: https://linc.cnil.fr/cabanon-can-anonymised-data-still-be-useful
[93] Carnot. Les données anonymisées n'ont-elles aucune valeur?“
[94] Isto.
[95] Francis, Paul. „Can Anonymized Data Still be Useful? Part Deux“. *Aircloak (2018).* Dostupno na: https://aircloak.com/can-anonymized-data-still-be-useful-part-deux/

5.2. Diffix

Paul Francis iz njemačkog Max Planck Instituta za softverske sustave iste je te podatke o rutama vožnje taksija u New Yorku provukao kroz sustav Diffix, konfigurirajući ga da, od svih dostupnih podataka, zaštiti privatnost taksista. Rezultati su prikazani u web članku *Can Anonymized Data Still be Useful? Part Deux* iz 2018. godine. Točke na slici 13 (desno) prikazuju polaske taksija, ali ne pojedinačne, već unutar okvira veličine približno 10 metara kvadratnih. Diffix je automatski uklonio sva područja koja nisu imala dovoljno različitih taksista (u ovom slučaju procjena je 4).[96]



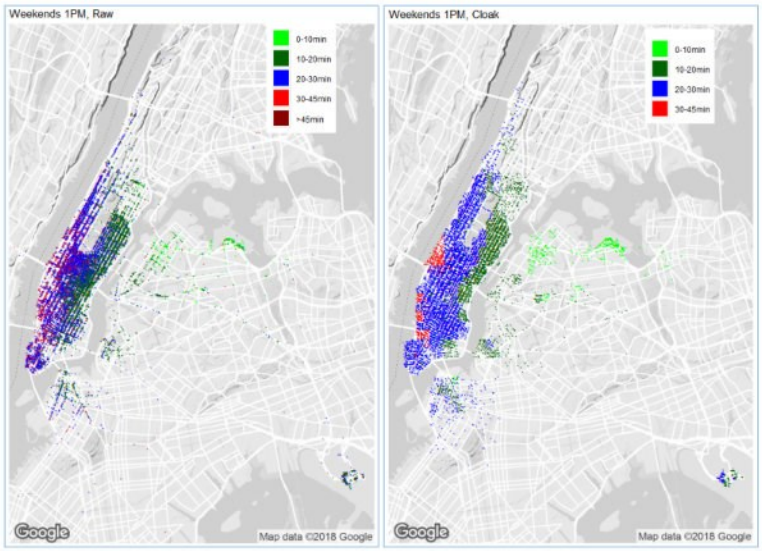
Slika 13. LINC vs Diffix[97]

Slika 13 uspoređuje karte koje su dobivene korištenjem LINC-ovog sustava i Diffixa. Usporedbom se primjećuje da je razina anonimizacije u LINC-ovom slučaju mnogo veća nego kod Diffixa. Ovdje treba uzeti u obzir da Diffix radi s puno više podataka nego što je to bilo u LINC-ovom slučaju jer je ondje odbačen dio podataka koji nije bio potreban za taj konkretan slučaj upotrebe.

Francis je uz pomoć Diffixa proveo tri slučaja upotrebe: očekivano vrijeme do zračne luke LaGuardia (New York), profili vozača i gužva u prometu. Što se tiče prvog slučaja, odabrana su 3 različita vremenska razdoblja kako bi se suprotstavila razdoblja najveće i najmanje gužve u prometu, te kontrastirala razdoblja jutarnje i večernje gužve. Slika 14 prikazuje čiste,

[96] Isto.
[97] Isto.

neobrađene podatke (lijevo) i podatke provučene kroz Diffix (desno, označeno s Cloak) jednog od promatranih razdoblja.[98]

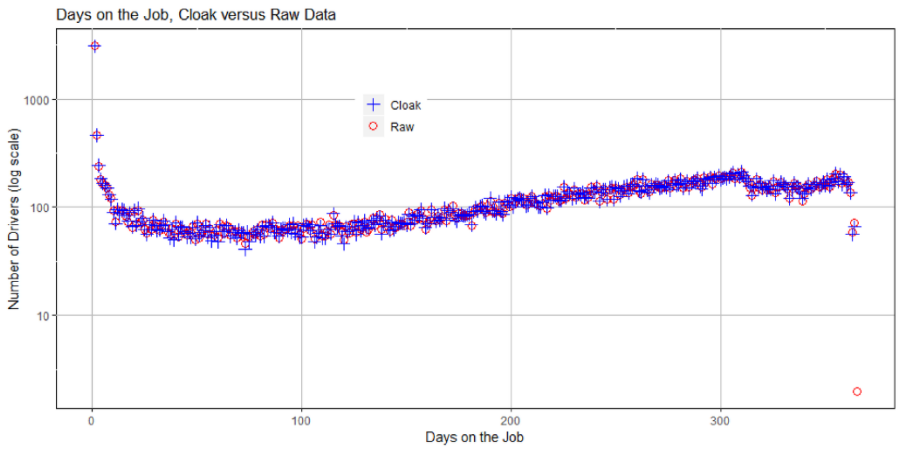


Slika 14. Neobrađeni podaci (lijevo) vs Diffix (desno)

Ne uzimajući u obzir značenje prikazane fotografije, može se zaključiti da iskrivljenje podataka kroz Diffixovu anonimizaciju ne mijenja osnovne zaključke koji se mogu izvući iz vizualizacije (slika 14). Zaključci dobiveni iz neobrađenih i anonimiziranih podataka su otprilike jednaki, što podupire upotrebu anonimizacije u statističke svrhe bez velikih promjena u rezultatima.

Drugi slučaj upotrebe vezan je uz radne profile vozača taksija. Iako su jedini podaci o taksistima u objavljenim podacima bili oni o njihovom identitetu, bez ikakvih drugih informacijama, i iz tih podataka doneseni su određeni zaključci. Kroz Diffix su izvučeni podaci o broju dana u godini koji svaki vozač odradi. Slika 15 prikazuje i anonimizirane (u tablici prikazani kao *Cloak*) i neobrađene (u tablici prikazani kao *Raw)* podatke, a može se zaključiti da su anonimizirani brojevi vrlo točni.[99]

[98] Isto.
[99] Isto.



Slika 15. Diffix - broj dana vožnje[100]

Treći slučaj su podaci o prometnim gužvama. Anonimizirani podaci izvučeni iz podataka Diffixom pokazuju kojom su se brzinom odvijale vožnje između ukrcaja i iskrcaja te su se prema toj brzini izvukli zaključci o količini gužve na cestama (manja brzina, veća gužva i obrnuto).[101] Ovakvi primjeri pokazuju da, ako se podaci ispravno anonimiziraju i s njima se ispravno rukuje, rezultati mogu biti pouzdani i korisni za daljnja postupanja s njima.

Sve ove slike dokazuju da nema velike razlike u upotrebi neobrađenih i anonimiziranih podataka u statističke svrhe. Moguće je imati točnu analitiku uz očuvanje anonimnosti. Ovo je iznimno važno sa stajališta zaštite privatnosti jer ukazuje na mogućnost da se korisne informacije mogu izvući iz anonimiziranih podataka bez potencijalno štetnih posljedica po privatnost pojedinaca. Ovakve statistike vrlo su zanimljive urbanistima i novim taksi prijevoznicima kako bi odredili na kojim područjima i kojim vremenskim razdobljima treba povećati javni promet i/ili promet taksijima kako bi to bilo najprofitabilnije i najkorisnije moguće.

[100] Isto.
[101] Isto.

6. Anonimizacija u hrvatskim arhivima

Kako bi se procijenila situacija u arhivima u Hrvatskoj što se tiče anonimizacije gradiva, kontaktirani su Hrvatski državni arhiv i 17 državnih arhiva koji djeluju na području Republike Hrvatske (državni arhivi u Bjelovaru, Dubrovniku, Gospiću, Karlovcu, Osijeku, Pazinu, Rijeci, Sisku, Slavonskom Brodu, Splitu, Šibeniku, Varaždinu, Virovitici, Vukovaru, Zagrebu, Zadru i Državni arhiv za Međimurje).

U istraživanju su zatraženi odgovori na sljedeća pitanja:

1. Provodite li anonimizaciju?
2. Koje vrste informacija Vaša institucija anonimizira? Na kojim vrstama gradiva?
3. Koje metode koristite? Kako tehnički provodite anonimizaciju analognih, a kako digitalnih sadržaja?
4. Ima li Vaša institucija politike ili smjernice u vezi s korištenjem anonimiziranih podataka u daljnje, istraživačke svrhe?

Prikupljeni podaci pokazuju da se anonimizacija u većini hrvatskih državnih arhiva provodi, a u onima u kojima se ne provodi, to je zato što dosad nije bilo upita koji bi zahtijevali anonimizaciju, ali su je arhivi spremni provesti u slučaju da se pojavi potreba za tim. Anonimiziraju se osobni podaci i oni podaci kojima u trenutku podnošenja zahtjeva za informacijom nije istekao rok dostupnosti za javnost oslanjajući se na zakonsku i drugu arhivsku regulativu. Državni arhivi zasad još uvijek nisu donijeli posebne politike ili smjernice za daljnje korištenje anonimiziranih podataka.

Tehničke mjere koje se poduzimaju moraju osigurati da se izvorno gradivo ne uništi ili ošteti, a nakon isteka roka navedenog u Zakonu o arhivskom gradivu i arhivima (članak 19.) može koristiti u originalu. Državni arhivi u Hrvatskoj koriste većinom slične tehničke mjere anonimizacije analognih i digitalnih sadržaja. Kod analognih sadržaja podaci koji se moraju učiniti nedostupnima se ili prekrivaju papirom pa se kopira traženi dokument, ili se dokument digitalizira pa mu se u programu za uređivanje fotografija zamute ili zacrne ti podaci. Postoji i verzija u kojoj se dokument kopira (izradi se preslika dokumenta) pa se podaci koji se moraju učiniti nedostupnima fizički zacrne (tintom ili crnim flomasterom). U slučaju digitalnih sadržaja podaci koji se moraju učiniti nedostupnima zamute se ili zacrne u programu za uređivanje slike, ili se navode inicijali i poopćavaju podaci. I kod analognih i

kod digitalnih sadržaja važno je da se mjere provode na preslikama, kako bi originalni zapisi ostali u izvornom stanju do isteka svih rokova nedostupnosti, nakon čega se korisnicima na korištenje daje cjeloviti, neanonimizirani dokument u analognoj ili digitalnoj verziji.

Hrvatski državni arhiv kao centralna institucija te središnji i matični arhiv najviše se bavi anonimizacijom, te radi na izradi posebnih smjernica u vezi s korištenjem anonimiziranih podataka. Anonimizaciju provodi na podacima koji se odnose na:

- sudske presude i istražne postupke, počinitelje kaznenih i prekršajnih djela, žrtve kaznenih djela žrtve policijskog progona i nadzora tajnih službi (npr. gradivo pravosudnih tijela, gradivo drugostupanjskih tijela nadležnih za rješavanje molbi za pomilovanja, dosjei zatvorenika, dosjei službi sigurnosti, spisi odvjetničkih kancelarija o zastupanju u sudskim postupcima i dr.),
- radni odnos, mirovinu, invalidnine i slično (npr. dosjei zaposlenika, rješenja o dodjeli mirovina, rješenja o invalidninama, žalbe i predstavke u kojima se navode i osobni podaci i dr.),
- zdravlje, socijalnu, zaštitu, spolni život i spolnu orijentaciju osobe, rasno ili političko podrijetlo, vjerska uvjerenja, politička mišljenja i dr. (npr. medicinska dokumentacija, posvojenja, smještaj u druge obitelji ili u dječje domove, socijalna pomoć, obiteljski odnosi, molbe, žalbe i predstavke u kojima se navode i osobni podaci, članski spisi i dr.),
- rasno ili političko podrijetlo, vjerska uvjerenja, politička mišljenja, spolni život ili spolna orijentacija osobe, zdravlje i slično u gradivu privatne provenijencije (npr. osobni dokumenti, članske iskaznice, privatna korespondencija, osobni dnevnici, bilješke osobne/privatne naravi i dr.),
- druge osjetljive osobne podatke u drugim vrstama gradiva.

Iako još nemaju službene smjernice za anonimizaciju, izrađen je Popis gradiva s podacima za koje je utvrđeno ograničenje dostupnosti sukladno Zakonu o arhivskom gradivu i arhivima (NN 61/18, 98/19). Popis sadrži podatke o arhivskim fondovima i zbirkama ili pojedinim njihovim dijelovima u kojima se nalaze i osobni podaci ograničene dostupnosti. Nakon isteka pojedinih rokova dostupnosti ili preuzimanja novog gradiva, popis se ažurira. Osim ovog popisa, primjenjuju se i Načela dostupnosti arhivskoga gradiva. Tehničke smjernice za upravljanje arhivskim gradivom ograničene dostupnosti u izdanju Međunarodnog arhivskog vijeća i prijevodu Hrvatskog državnog arhiva.

7. Zaključak

U današnjem digitalnom dobu sve veći izazov predstavlja očuvanje privatnosti i sigurnosti osobnih podataka. Iako se anonimizacija kontinuirano razvija i primjenjuje u praksi, postojeći pravni okviri, posebice na području Europske unije s Općom uredbom o zaštiti podataka (GDPR), pokušavaju osigurati konačnost anonimizacije te tvrde da nema načina za povrat izvornih podataka iz anonimiziranih oblika, te se prema tome anonimizirani podaci ne smatraju osobnim podacima i ne podliježu nikakvim pravilima za obradu osobnih podataka. Tehnološki sektor se usmjerava prema razvoju alata za anonimizaciju, pri čemu umjetna inteligencija sve više igra ključnu ulogu. Razvoj anonimizacije i zaštite osobnih podataka dinamičan je proces koji zahtijeva sinergiju pravnih, tehnoloških i praktičnih rješenja. Dok se tehnologija neprestano unaprjeđuje i postaju dostupni sve sofisticiraniji alati, istovremeno je važno prepoznati specifične potrebe različitih vrsta podataka te prilagoditi strategije zaštite sukladno tome.

Važno je istaknuti da, iako se anonimizacija čini obećavajućom, nema apsolutnog jamstva da će podaci ostati potpuno zaštićeni. Neprestani tehnološki napredak može dovesti do razvoja novih metoda deanonimizacije koje bi mogli narušiti trenutne koncepte zaštite osobnih podataka. Stoga je važno održavati kontinuirani dijalog između zakonodavaca, informatičke i tehnološke zajednice te stručnjaka za privatnost kako bi se osigurala pravovremena reakcija na sve izazove koji dolaze s budućim razvojem znanosti, tehnologije i znanja. S obzirom na osjetljivost i kompleksnost ovog problema, potrebno je educirati ljude o važnosti zaštite osobnih podataka i o načinima na koje oni sami mogu doprinijeti očuvanju svoje privatnosti i sigurnosti svojih osobnih podataka. I korisnici, kao i organizacije koje prikupljaju i obrađuju podatke, trebaju biti svjesni rizika i odgovornosti koje dolaze s tim procesima.

8. Literatura

1. „Direktiva 95/46/EZ Europskog parlamenta i vijeća od 24. lipnja 1995. o zaštiti pojedinaca u vezi s obradom osobnih podataka i o slobodnom protoku takvih podataka“. *Službeni list Europskih zajednica* L 281/31 (1995). Dostupno na: https://eur-lex.europa.eu/legal-content/HR/TXT/PDF/?uri=CELEX:31995L0046&from=HU (12.8.2023.)
2. K^m-anonymity. Anonymity Tutorial. *Amnesia*. Dostupno na: https://amnesia.openaire.eu/Scenarios/AmnesiaKMAnonymityTutorial.pdf (16.9.2023.)
3. ARX – Data Anonymization Tool. Dostupno na: https://arx.deidentifier.org/ (26.4.2023.)
4. ASAM. „UAI Anonymizer“. Dostupno na: https://www.asam.net/members/product-directory/detail/uai-anonymizer/ (19.8.2023.)
5. Bradić-Martinović, Aleksandra; Zdravković, Aleksandar. „Zaštita privatnosti – anonimizacija podataka“. V naučni skup USPON, Beograd (2013): 206-213. Dostupno na: https://www.academia.edu/45437258/Za%C5%A1tita_privatnosti_anonimizacija_podataka_Privacy_protection_data_anonymization (16.8.2023.)
6. brighter AI. „Use anonymization to monetize data and be compliant“. Dostupno na: https://brighter.ai/product/#demo (19.8.2023.)
7. Bukvić, Nenad. „Opća uredba o zaštiti podataka (GDPR) i opis osobnih podataka u arhivskom gradivu: o nekim praktičnim i etičkim aspektima“. *Arhivski vjesnik* 63, br. 1 (2020): 9-32. Dostupno na: https://doi.org/10.36506/av.63.1 (10.8.2023.)
8. CARNET. „Anonimizacija i pseudonimizacija podataka“. Dostupno na: https://www.cert.hr/wp-content/uploads/2018/08/anonimizacija_i_pseudonimizacija_podataka.pdf (15.8.2023.)
9. Devane, Heather. „Everything You Need to Know About K-Anonymity“. *Immuta*. (2023) Dostupno na: https://www.immuta.com/blog/k-anonymity-everything-you-need-to-know-2021-guide/ (16.9.2023.)
10. Domingo-Ferrer, Josep; Sanchez, David; Soria-Comas, Jordi. „Database Anonymization: Privacy Models, Data Utility, and Microaggreegation-based Inter-model Connections“. *Synthesis Lectures on Information Security Privacy and Trust* 8 (1) (2016): 1-136. Dostupno na: https://www.researchgate.net/publication/290229262_Database_Anonymization_Privacy

_Models_Data_Utility_and_Microaggregation-based_Inter-model_Connections (18.8.2023.)

11. Europski parlament, Vijeće Europske unije, Europska komisija. „Povelja Europske unije o temeljnim pravima“. *Službeni list Europske unije* C 202/389 (2016). Dostupno na: https://eur-lex.europa.eu/legal-content/HR/TXT/PDF/?uri=CELEX:12016P/TXT&from=RO (26.7.2023.)
12. Europski parlament, Vijeće Europske unije. „Direktiva 2002/58/EZ Europskog parlamenta i Vijeća od 12. srpnja 2022. o obradi osobnih podataka i zaštiti privatnosti u području elektroničkih komunikacija (Direktiva o privatnosti i elektroničkim komunikacijama)“. *Službeni list Europske unije* L 201/37 (2002). Dostupno na: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX%3A32002L0058%3Ahr%3AHTML (12.8.2023.)
13. Europski parlament, Vijeće Europske unije. „Uredba (EU) 2016/679 Europskog parlamenta i Vijeća od 27. travnja 2016. o zaštiti pojedinaca u vezi s obradom osobnih podataka i o slobodnom kretanju takvih podataka te o stavljanju izvan snage Direktive 95/46/EZ (Opća uredba o zaštiti podataka)“. *Službeni list Europske unije* L 199/1 (2016). Dostupno na: https://eur-lex.europa.eu/legal-content/HR/TXT/?uri=celex%3A32016R0679 (25.4.2023.)
14. Francis, Paul. „Can Anonymized Data Still be Useful? Part Deux“. *Aircloak (2018).* Dostupno na: https://aircloak.com/can-anonymized-data-still-be-useful-part-deux/ (25.4.2023.)
15. Google. „Privatnost i uvjeti“. Dostupno na: https://policies.google.com/technologies/anonymization?hl=hr (1.8.2023.)
16. Harvard University. „Differential Privacy“. Dostupno na: https://privacytools.seas.harvard.edu/differential-privacy (18.8.2023.)
17. Hrvatska enciklopedija, mrežno izdanje. „mrežni kolačić“. *Leksikografski zavod Miroslav Krleža*. (2021). Dostupno na: https://enciklopedija.hr/Natuknica.aspx?ID=70676 (15.09.2023.)
18. Hrvatska enciklopedija, mrežno izdanje. „McLuhan, Herbert Marshall“. *Leksikografski zavod Miroslav Krleža* (2021). Dostupno na: https://enciklopedija.hr/natuknica.aspx?id=39694 (15.09.2023.)

19. Hrvatska enciklopedija, mrežno izdanje. „podatak“. *Leksikografski zavod Miroslav Krleža* (2021). Dostupno na: https://www.enciklopedija.hr/natuknica.aspx?id=48887 (17.8.2023.)
20. Hrvatska enciklopedija, mrežno izdanje. „umjetna inteligencija“. *Leksikografski zavod Miroslav Krleža* (2021). Dostupno na: https://www.enciklopedija.hr/natuknica.aspx?ID=63150 (18.8.2023.)
21. Hrvatska liječnička komora. „Ženevska deklaracija“. Dostupno na: https://www.hlk.hr/EasyEdit/UserFiles/pdf-ovi-za-vijesti-web/2020/lijecnicka-prisega-preambula.pdf (15.09.2023.)
22. Hrvatski sabor. „Ustav Republike Hrvatske“. *Narodne novine* 56/90, 135/97, 8/98, 113/00, 124/00, 28/01, 41/01, 55/01, 76/10, 76/10, 85/10, 05/14 (2014). Dostupno na: https://www.zakon.hr/z/94/Ustav-Republike-Hrvatske (25.4.2023.)
23. Hrvatski sabor. „Zakon o arhivskom gradivu i arhivima“. *Narodne novine* 61/18 (2018). Dostupno na: https://narodne-novine.nn.hr/clanci/sluzbeni/2018_07_61_1265.html (1.4.2023.)
24. Hrvatski sabor. „Zakon o elektroničkim komunikacijama“. *Narodne novine* 73/08 (2008). Dostupno na: https://narodne-novine.nn.hr/clanci/sluzbeni/2008_06_73_2420.html (10.8.2023.)
25. Hrvatski sabor. „Zakon o elektroničkim komunikacijama“. *Narodne novine* 76/22 (2022). Dostupno na: https://www.zakon.hr/z/182/Zakon-o-elektroni%C4%8Dkim-komunikacijama (10.8.2023.)
26. Hrvatski sabor. „Zakon o izmjenama i dopunama Zakona o pravu na pristup informacijama“. *Narodne novine* 85/15 (2015). Dostupno na: https://narodne-novine.nn.hr/clanci/sluzbeni/2015_08_85_1649.html (2.4.2023.)
27. Hrvatski sabor. „Zakon o izmjenama i dopunama Zakona o pravu na pristup informacijama“. *Narodne novine* 69/22 (2022). Dostupno na: https://narodne-novine.nn.hr/clanci/sluzbeni/2022_06_69_1025.html (2.4.2023.)
28. Hrvatski sabor. „Zakon o pravu na pristup informacijama“. *Narodne novine* 25/13 (2013). Dostupno na: https://narodne-novine.nn.hr/clanci/sluzbeni/2013_02_25_403.html (2.4.2023.)
29. Hrvatski sabor. „Zakon o tajnosti podataka“. *Narodne novine* 79/07 (2007). Dostupno na: https://narodne-novine.nn.hr/clanci/sluzbeni/2007_07_79_2483.html (2.4.2023.)
30. kaggle. „Datasets“. Dostupno na: https://www.kaggle.com/datasets (19.8.2023.)

31. Le Carnot Télécom & Société numérique. „Les données anonymisées n'ont-elles aucune valeur?“. (2018). Dostupno na: https://www.carnot-tsn.fr/cabanon-teralab-donnees-anonymisees/ (10.8.2023.)
32. LINC. „CabAnon: exploring and visualising anonymized datasets“. (2017). Dostupno na: https://linc.cnil.fr/cabanon-exploring-and-visualizing-anonymized-datasets (16.8.2023.)
33. LINC. „Can anonymised data still be useful?“. (2017). Dostupno na: https://linc.cnil.fr/cabanon-can-anonymised-data-still-be-useful (10.8.2023.)
34. Međunarodno arhivsko vijeće. „Načela dostupnosti arhivskoga gradiva“. Zagreb: *Hrvatski državni arhiv* (2015). Dostupno na: https://www.ica.org/sites/default/files/ICA_Access_Principles__CR.pdf (26.7.2023.)
35. Međunarodno arhivsko vijeće. „Načela dostupnosti arhivskoga gradiva. Tehničke smjernice za upravljanje arhivskim gradivom ograničene dostupnosti“. Zagreb: *Hrvatski državni arhiv*. (2016). Dostupno na: https://www.ica.org/sites/default/files/tech-guidance_hr.pdf (26.7.2023.)
36. Međunarodno arhivsko vijeće. „Opća međunarodna norma za opis arhivska gradiva ISAD(G)“. Zagreb: *Hrvatski državni arhiv* (2001). Dostupno na: https://www.ica.org/sites/default/files/isad_g_2_edition_hr.pdf (15.8.2023.)
37. Mostly.AI. „Data anonymization – synthetic data for maximum privacy and utility“. Dostupno na: https://mostly.ai/use-case/data-anonymization-with-synthetic-data (19.8.2023.)
38. Open Diffix. „Play with Diffix“. Dostupno na: http://www.open-diffix.org/en/play/ (19.8.2023.)
39. OpenAIRE. „Amnesia – Anonymize your data before publishing“. Dostupno na: https://www.openaire.eu/amnesia-guide (19.8.2023.)
40. Radna skupina za zaštitu podataka. „Mišljenje 05/2014 o tehnikama anonimizacije“. (2014). Dostupno na: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_hr.pdf (22.7.2023.)
41. Sartor, Nicolas. „Data Anonymization Software – Diferences Between Static and Interactive Anonymization“. *Aircloak* (2019). Dostupno na: https://aircloak.com/data-anonymisation-software-differences-between-static-and-interactive-anonymisation/ (25.4.2023.)
42. Sartor, Nicolas. „Top 5 Free Data Anonymization Tools“. *Aircloak* (2019). Dostupno na: https://aircloak.com/top-5-free-data-anonymization-tools/ (25.4.2023.)

43. struna. „podatak“, *Institut za hrvatski jezik i jezikoslovlje*. Dostupno na: http://struna.ihjj.hr/naziv/podatak/6110/ (17.8.2023.)
44. VEIL.AI. „Unlock the power of data“. Dostupno na: https://veil.ai/why-veil-ai/ (19.8.2023.)

Popis slika

Anonimizacija zapisa u digitalnim arhivima

Sažetak

Sigurnost osobnih podataka sve je veći izazov čovječanstva kako sve više sfera ljudskog života prelazi u digitalni svijet. Nehotično davanje privole na obradu osobnih podataka na gotovo svakoj internetskoj stranici postalo je svakodnevica. Načini na koje se ti podaci obrađuju podložni su Općoj uredbi o zaštiti podataka (GDPR) koja je donesena sa svrhom da se zaštiti privatnost i dostojanstvo pojedinca kao temeljno i ustavno pravo svake osobe. Arhivske ustanove kao mjesta na kojima se čuvaju dokumenti i dokazi prava i transakcija pune su zapisa koji sadrže osobne podatke. Koji od tih podataka spadaju pod zaštitu, a koji ne određeno je u arhivskom zakonu, a dostupnost arhivskog gradiva s i bez osjetljivih podataka određuju ustanove same za sebe sukladno zakonu i pravilnicima.

Rad se dotiče problematike obrade osobnih podataka i korištenja umjetne inteligencije dok s druge strane daje sliku načina na koji državni arhivi u Hrvatskoj anonimiziraju svoje zapise. U radu se usporedno analizira nekoliko alata za anonimizaciju. Napretkom tehnologije došlo je do razvoja mnogih računalnih sustava za anonimizaciju, a u posljednje vrijeme sve je veća važnost umjetne inteligencije u obradi osobnih podataka, odnosno pripremi podataka za daljnje analize.

Ključne riječi: anonimizacija, zaštita podataka, osobni podaci, GDPR, umjetna inteligencija

Records anonymization in digital archives

Summary

The security of personal data is a growing challenge for humanity as more and more spheres of human life move into the digital world. Involuntarily giving consent to the processing of personal data on almost every internet site has become an everyday practice. The ways in which these data are processed are subject to the General Data Protection Regulation (GDPR), which was adopted with the purpose of protecting the privacy and dignity of an individual as a fundamental and constitutional right of every person. Archival institutions, as places where documents and evidence of rights and transactions are kept, are full of records containing personal data. Which of these data fall under protection and which do not is determined by the archives law, and the availability of archival materials with and without sensitive data is determined by the institutions themselves in accordance with the law and regulations.

The thesis touches on the issues of personal data processing and usage of artificial intelligence, while on the other hand it provides insight into the records anonymization practices in the state archives in Croatia. The comparative analysis of several anonymization tools is given. Advances in technology have led to the development of many computer systems for anonymization, and lately the importance of artificial intelligence in processing personal data, i.e., preparing data for further analysis, is increasing.

Key words: anonymization, data protection, personal data, GDPR, artificial intelligence