

Prejeto / Received: 11. 12. 2021
<https://doi.org/10.54356/MA/2021/LSBO6521>

1.01 Izvirni znanstveni članek

1.01 Scientific Article

AVTOMATIZIRANO PREKRIVANJE Z ARHIVSKIM ZAKONOM VAROVANIH OSEBNIH PODATKOV – ANONIMIZACIJA

Dr. Gregor Jenuš

Arhiv Republike Slovenije, Slovenija
gregor.jenus@gov.si

Dr. Žiga Koncilijs

Arhiv Republike Slovenije, Slovenija
ziga.koncilija@gov.si

Dr. Tatjana Hajtnik

Arhiv Republike Slovenije, Slovenija
tatjana.hajtnik@gov.si

Izvleček:

Arhivsko gradivo je zapuščina naše preteklosti, zato vsebuje tudi takšne podatke, za katere je zakonodajalec določil, da niso dostopni vsakomur. V okviru slovenske javne arhivske mreže so javni arhivi zavezani k anonimizaciji oziroma prekrivanju z arhivskim zakonom varovanih osebnih podatkov. Avtorji v prispevku predstavijo problematiko varstva osebnih podatkov, pomena javne osveščenosti vrednosti le-teh v dobi četrte industrijske revolucije in hibridnega poslovanja. Opozorijo na pridobitve splošne evropske uredbe o varstvu osebnih podatkov in arhivske zakonodaje, ki kot enega od mehanizmov varstva predvidevajo tudi anonimizacijo. Prispevek predstavi tudi praktične vidike anonimizacije arhivskega gradiva v javnih arhivih, le-ta se izvaja na dva načina: fizično na kopijah arhivskega gradiva in elektronsko, slednje javni arhivi izvajajo na izvorno digitalnem ali digitaliziranem arhivskem gradivu s pomočjo namenskega programskega orodja, ki so ga pridobili v okviru projekta vzpostavitve slovenskega javnega elektronskega arhiva.

Ključne besede:

arhivsko gradivo, anonimizacija, varstvo osebnih podatkov, GDPR, arhivska zakonodaja, digitalizacija, arhivi

Abstract:

Automated Processing of Personal Data Protected by Archival Law - Anonymisation

Archival records are a legacy of our past. They also contain such information which the legislator has determined is not accessible to everyone. Within the Slovenian public archival network, public archives are obliged to anonymize or remove personal data that is protected by archival law. The authors present the issue of personal data protection and the importance of the public awareness of the importance of personal data in the age of the fourth industrial revolution and the hybrid documents creation. They present the practical aspects of anonymization of archival records in public archives in two ways; physical on copies of archival records and digital, which public archives perform on born digital or digitized archival records with the help of dedicated software tools acquired within the project of establishing the Slovenian public electronic archive.

Key words:

archival records, anonymization, personal data protection, GDPR, archival legislation, digitization, archives

1. UVOD - ARHIVSKO GRADIVO V DOBI DIGITALNE REVOLUCIJE

Arhivsko gradivo predstavlja vezni člen z našo preteklostjo, saj nam kot zapuščina določenega časa omogoča neposreden vpogled v delovanje družbenih sistemov, institucij, skupin ali posameznikov, njihovo komunikacijo in interakcije. Iz njega je mogoče rekonstruirati splošno ozračje v prelomnih trenutkih naše zgodovine ali družbene procese, ki so k spremembam pripomogli (Jenuš, 2013, 58). Prav zaradi njegove vrednosti kot primarnega vira je arhivsko gradivo med drugim pomembno za humanistične in družboslovne študije, saj nudi neposreden vpogled v našo preteklost. Dokazna vrednost pa mu daje tudi pomen kot pravni vir v upravnih ali sodnih postopkih. Arhivski zapisi so tako v skladu z Zakonom o varstvu dokumentarnega in arhivskega gradiva ter arhivih (ZVDAGA) neprecenljiva zapuščina naše preteklosti, slovenska zakonodaja pa jih vrednoti kot kulturni spomenik in skupno evropsko kulturno dediščino (ZVDAGA, 2014, 2. člen).

Skozi stoletja se je videz arhivskega gradiva spreminjal. V njem se venomer zrcalita čas in okolje, ki sta ga ustvarila. Če so do druge polovice 20. stoletja arhivski zapisi predstavljali predvsem rokopise, tipkopise, risbe, tiske, fotografije, načrte ali kako drugače v fizičnem svetu nastalo gradivo, se je to s postopno informatizacijo naše družbe spreminjalo.¹ Danes smo zaradi široke uporabe informacijsko-komunikacijske tehnologije nenehno soočeni z izzivi hibridne dobe poslovanja tako imenovane četrte industrijske revolucije.² Ta nas ne sooča le z vedno večjimi količinami najrazličnejših oblik arhivskih zapisov – bodisi v digitalnem ali fizičnem svetu, ampak tudi z eksponentno rastjo podatkov (Hajtnik, 2012, 55-70; Asmiyanto, 2019, 2; Koncilijska 2018, 95-96). In čeprav se je videz arhivskega gradiva morda spremenil, pa se njegova zgodovinska, kulturna ali denimo pravna vrednost niso.

¹ Arhivsko gradivo. Dostopno na: <https://www.gov.si teme/arhivsko-gradivo/> (8. 12. 2021).

² Četrta industrijska revolucija ali tudi Industrija 4.0 predstavlja trend v svetovnem gospodarstvu in širše družbeno gibanje, v katerem se vse več delovnih procesov avtomatizira s pomočjo informacijsko-komunikacijskih tehnologij, interneta stvari, računalništva v oblaku in tako imenovanega kognitivnega računalništva. Če se ozremo na prelomne spremembe, ki so jih prinesle predhodne industrijske revolucije, se moramo zavedati, da so spremembe daljnosežne. V primeru prve industrijske revolucije smo bili priča, kako je glavna gonilna sila razvoja – parni stroj – pripomogel k snovanju novih industrijskih obratov, ti pa so povzročili velike migracijske tokove prebivalstva iz ruralnih v urbane ali industrijske centre, kar je zaradi proletarizacije družbe pripeljalo do korenitih socialnih sprememb. V primeru druge industrijske revolucije je razmah industrializacije povzročilo odkritje električne energije, ta je omogočila hitrejšo in učinkovitejšo proizvodnjo, razvoj znanosti in tehnologije, ki so spremenile podobo našega vsakdana in pripeljale do tretje industrijske revolucije. Ta je predstavljala nadgradnjo elektrificirane industrije in vpeljavo novih informacijsko-komunikacijskih tehnologij ter elektronike, ki so povzročile avtomatizacijo proizvodnje. Za tretjo industrijsko revolucijo je bilo značilno, da je vse več dela slonelo na novih tehnologijah, ki so postale vključene vse »pore« naših življenj.

Zahtevnost četrte industrijske revolucije je v dejstvu, da se za dosego ciljev prepleta vse več različnih (informacijsko-komunikacijskih) tehnologij in da je avtomatizacija delovnih procesov dosegla tudi nivo uporabnikov. To pa prinaša izzive za številne deležnike – za gospodarstvo, zakonodajalca (tj. med drugim tudi upravo) in družbo kot celoto. Gospodarstvu prinaša izzive za njihove poslovne modele; ti morajo biti prilagojeni pričakovanjem strank, izboljšavam ali inovacijam produktov in optimizaciji izrabe novih poslovnih okvirjev. Za zakonodajalce prinaša četrta industrijska revolucija izzive v sposobnosti hipnega odzivanja na zahteve družbe in gospodarstva, tehnologije in znanosti. Sposobni morajo biti hitro odzgirati na spremembe, ki jih revolucija prinaša, in zagotoviti pravne okvirje, ki delujejo kot vzpodbujevalec razvoja, hkrati pa ščitijo njegove uporabnike pred morebitnimi zlorabami. In prav ta vidik je pomemben za družbo. Četrta industrijska revolucija je nedvomno pripomogla k spremembi družbe kot celote, naših hotenj in pričakovanj (Asmiyanto, 2019, 16-17). To pa se odraža tudi v arhivskem gradivu, ki danes nastaja – ne le v velikih količinah, ampak tudi v raznovrstnih pojavnih oblikah.

2. ZAVEDANJE POMENA OSEBNIH PODATKOV

Gradivo danes ni zahtevno zgolj zaradi nosilcev ali formatov, v katerih nastaja, ampak je zahtevno tudi zaradi količine in oblik informacij, ki jih nosi v sebi. Digitalna doba je namreč pripeljala do revolucije na področju razmnoževanja podatkov. Danes tako govorimo o problemu „big data“ ali „velikem podatkovju“, le-tega neizbežno spremljajo tudi velike količine osebnih podatkov, ki jih je potrebno zaščititi (Jerome, 2013, 47). Tako imenovano »veliko podatkovje« oz. velike zbirke podatkov dobimo, če svoje podatke združimo z milijoni podatkov drugih uporabnikov.

»Podjetja uporabljajo avtomatizirana orodja za analizo velike količine podatkov milijonov različnih uporabnikov, da prepoznajo vzorce njihovega vedenja. Tako lahko precej natančno napovejo, kaj boste storili glede na to, kaj ste vi in drugi, podobni vam, storili v preteklosti. Zapleteni algoritmi omogočajo, da so posameznemu uporabniku ponujeni prav njemu prilagojeni oglasi in storitve, pa tudi vsebine, ki lahko vplivajo celo na njegove politične odločitve.«³

In tukaj »vse glasneje« stopa v ospredje želja po varovanju osebnih podatkov, ki pa vendarle ni nekaj novega, a v dobi tehnološkega hiperrazvoja, hiperprodukcije podatkov in globalizacije postaja vse pomembnejša. Obseg nezaželenih izmenjav osebnih podatkov se je močno povečal. Tehnologija nam omogoča, da brez večjih naporov delimo osebne podatke in to pogosto nevede oziroma nehote. Vendar pa v tem problemu obstaja dejavnik, na katerega imamo vpliv. Problem ni le v tehnologiji in njenih skoraj neskončnih možnostih; problem so fizične osebe. To smo mi! (Jenuš, 2021, 171). Fizične osebe vse pogostejše dajejo osebne podatke na razpolago tretjim osebam, javno in v svetovnem merilu, ker ne razmišljajo o možnih škodljivih posledicah takšnega ravnanja (Romanski & Nominska, 2020, 5288-5289; Jenuš, 2021, 171).

»Storitve, ki so navidezno brezplačne ali poceni, (uporabniki, op. a.) pogosto plačujemo prav s svojimi podatki. Tržna vrednost teh podatkov je vedno večja, njihovo zbiranje, analiza, obdelava in prodaja pa je velikanski posel, ki je uporabnikom, katerih podatki se znajdejo na trgu, v veliki meri neznan. Uporabniki pogosto nimajo nadzora nad zbranimi podatki in nameni, za katere se ti uporabljajo, pogosto ne vemo niti, katere podatke se o nas zbira in da so med njimi tudi naši osebni podatki. Ne bo torej odveč, če se pri brskanju po spletu ter uporabi aplikacij in družbenih omrežij zavedamo, da smo v svetu, ki ga poganjajo podatki, pogosto naprodaj prav mi. Naši podatki so veliko več kot fotografije, ki jih zavestno naložimo na Facebook ali Instagram, tviti, ki jih delimo, in osebni podatki, ki jih vnesemo pri nastavitvi profilov družbenih omrežij.

Naši podatki so tudi informacije o tem, kje se nahajamo, s kom se družimo, kaj iščemo na brskalnikih, katere članke beremo, kaj kupujemo, kateri telefon imamo in tako naprej. Nemalo uporabnikov brez pomislekov sprejme vse piškotke v svojem brskalniku, saj ne želijo izgubljati časa s prebiranjem pogojev uporabe ali politik zasebnosti, v katerih je zapisano, kaj se bo zgodilo s podatki. Številne aplikacije delujejo tako, da črpajo informacije iz drugih virov, zahtevajo lahko dostop do naših fotografij in videoposnetkov ali imenika, nekatere pa zahtevajo celo več informacij, kot jih potrebujejo za izvajanje svoje storitve – tem se je pametno izogniti.«⁴

³ Ali veste, kaj pomeni »big data« in zakaj so naši podatki tako dragoceni? Dostopno na: <https://www.mipi.si teme/informacijska-pismenost/ali-veste-kaj-pomeni-187big-data171-in-zakaj-so-nasi-podatki-tako-dragoceni> (8. 12. 2021).

⁴ Prav tam.

Digitalna osveščenost in pismenost⁵ se tako torej zdita ključni kompetenci, ki ju mora osvojiti vsak uporabnik, saj med drugim dajeja velik pomen tudi varovanju osebnih podatkov. Pri varstvu osebnih podatkov večina uslužbencev uprave ali gospodarstva v Republiki Sloveniji ali širše v Evropi najprej pomisli na slavno kratico 'GDPR'. Ta predstavlja odgovor na velikokrat malomarno in nespametno ravnanje uporabnikov z lastnimi osebnimi podatki. Evropski parlament in Evropski svet sta razmišljala o vprašanju, kako v Evropski uniji regulirati varstvo osebnih podatkov med svojimi državami članicami v najširšem možnem obsegu. Aprila 2016 je bila sprejeta tako imenovana Splošna uredba Evropske unije o varstvu podatkov (ang. General Data Protection Regulation). GDPR pomeni mejnik pri varstvu osebnih podatkov⁶, vsaj tako je evropski zakonodajalec uredbo predstavil javnosti. Avtorji GDPR navajajo, da je namen te uredbe najti ravnotežje med temeljno pravico fizične osebe do varstva svojih osebnih podatkov in pravnim okvirom, v katerem bi uradni organi lahko izvajali svojo družbeno funkcijo (Jenuš, 2021, 172; GDPR, 2016, uvodna utemeljitev št. 65).

Odziv javnosti na GDPR je bil mešan; kritičen predvsem zaradi določila o zahtevi varstva osebnih podatkov za upravljalce le-teh, brez usmeritev, kako varstvo doseči, ali razmisleka o stroških, ki so zaradi tega nastali (Prelesnik, 2018, 4-5, 8). Pomisleki glede GDPR pa so se pojavili tudi med arhivi. Posebno problematični sta se zdeli tako imenovana »pravica do popravka« in »pravica do pozabe« (GDPR, 2016, 16. in 17. člen). Vendar so bile skrbi neutemeljene. Kot ugotavlja Matevž Košir, je zaščita osebnih podatkov v skladu z GDPR določena le za žive osebe, ne velja pa za umrle. Uredba se tako ne uporablja za osebne podatke umrlih oseb, zato se določila GDPR nanašajo samo na tisto arhivsko gradivo, ki se nanaša na še živeče osebe (Košir, 2020, 61-62; GDPR, 2016, uvodna utemeljitev št. 27).

»Države članice pa lahko določijo pravila za obdelavo osebnih podatkov umrlih oseb. Temelj uredbe je pravica vsake osebe do zaščite osebnih podatkov pri nadaljnji obdelavi. Ključno načelo Splošne uredbe pa je zmanjševanje podatkov. Osebne podatke je treba zbirati in obdelovati le, če je to res potrebno, in jih hraniti, le toliko časa, kolikor je to potrebno, da se dosega namen, za katerega so bili osebni podatki zbrani.« (Košir, 2020, 61-62; GDPR, 2016, uvodna utemeljitev št. 27).

Glede na številna odprta vprašanja je Evropska arhivska skupina (EAG) pripravila smernice za izvajanje splošne uredbe o varstvu podatkov s strani arhivskih služb. Te smernice, ki niso pravni akt in so le nezavezujoč pripomoček, naj bi zagotavljale osnovne informacije in praktične napotke glede izzivov za uporabo GDPR v arhivih, to pa naj bi pripomoglo k dvigu zaupanja javnosti v GDPR (Smernice EAG, 2018).

⁵ Digitalna pismenost je ena izmed ključnih vseživljenjskih kompetenc, ki vključuje varno in kritično uporabo tehnologije informacijske družbe pri delu, v prostem času in pri sporazumevanju. Podpirajo jo osnovna znanja informacijsko-komunikacijske tehnologije: uporaba računalnikov za iskanje, ocenjevanje, shranjevanje, proizvodnjo, predstavitev in izmenjavo informacij ter za sporazumevanje in sodelovanje v skupnih omrežjih po internetu. Poseben vidik digitalne pismenosti je pomen osebnih podatkov oz. tveganje njihove zlorabe.

Digitalna pismenost. Dostopno na: <https://www.portalosv.si/digitalna-pismenost/> (8. 12. 2021).

⁶ 4. člen GDPR v prvem odstavku definira osebne podatke kot katero koli obliko informacije v zvezi z določenim ali določljivim posameznikom. Določljiv posameznik je tisti, ki ga je mogoče neposredno ali posredno določiti, zlasti z navedbo identifikatorja, kot je ime, identifikacijska številka, podatki o lokaciji, spletni identifikator, ali z navedbo enega ali več dejavnikov, ki so značilni za fizično, fiziološko, genetsko, duševno, gospodarsko, kulturno ali družbeno identiteto tega posameznika (GDPR, 2016, prvi odstavek 4. člena). Splošna uredba se uporablja, kadar nekdo obdeluje osebne podatke za svoje gospodarske, poslovne ali poklicne namene. Splošna uredba varuje osebne podatke ne glede na to, katera tehnologija se uporablja za njihovo obdelavo (Košir, 2020, 61-62).

Konec maja 2018 je po dvehletnem prehodnem obdobju GDPR končno začel veljati. Časopisi, strokovni bilteni itd. so bili polni informacij, kaj bo ta mejnik prinesel in kazalo je, da bo petek, 25., postal »črni petek« varstva podatkov. Vendar z uvedbo GDPR ni bilo vse slabo. GDPR je dvignil ozaveščenost slovenske javnosti o pomenu osebnih podatkov, kar ne more biti slabo, in hkrati »prisilil« organe, ki upravljajo s tovrstnimi podatki, k razmisleku o svojih delovnih procesih in podatkovni varnosti (Jenuš, 2021, 173). Pripomogel pa je tudi k povečanju zanimanja za storitve varstva osebnih podatkov. V fokus slovenskih organov javne in državne uprave ter sodišč je postavil tudi slovenske arhive in projekt vzpostavitve slovenskega elektronskega arhiva, saj smo arhivi zaradi izzivov z anonimizacijo iskali rešitve, kako jo čim bolj poenostaviti, in smo torej imeli na tem področju že kar nekaj izkušenj.

3. ZAKONSKA UVEDBA ANONIMIZACIJE

Slovenska zakonodaja se ponaša z razmeroma visokim standardom varstva osebnih podatkov. Neposredne implementacije GDPR v prenovljenem Zakonu o varstvu osebnih podatkov (ZVOP-2) sicer žal, tudi tri leta po objavi predloga le-tega, še nismo dočakali, saj je zakon še vedno v pripravi.⁷ Arhivski svet, ki deluje kot posvetovalno telo ministra za kulturo za strokovno svetovanje na področju arhivske dejavnosti in obravnava ter posreduje mnenja o strokovnih vprašanjih s področja delovanja in razvoja arhivske dejavnosti, je leta 2019 glede takratnega predloga ZVOP-2 zapisal, da predlog zakona v sistem arhiviranja oz. delovanja javnih arhivov ne posega.

»Dovoljena je nadaljnja obdelava osebnih podatkov (opredeli jo ZVDAGA), pravica do pozabe in izbrisa pa se ne nanaša na arhivsko gradivo (se pa nanaša na gradivo pri ustvarjalcu, ki pa na podlagi pisnega navodila arhiva lahko postane arhivsko). ZVDAGA domneva, da je treba do izdaje pisnih strokovnih navodil za odbiranje z vsem gradivom ravnati kot z arhivskim.«⁸

A slovenski arhivi smo – k sreči ali ne – v zadnjih letih poskrbeli za prenovu področne arhivske zakonodaje in vanjo vnesli tudi določila glede varstva osebnih podatkov, ki se nahajajo v arhivskem gradivu.

Podobno kot GDPR leta 2016 oz. 2018 je tudi prenovljeni Zakon o spremembah in dopolnitvah Zakona o varstvu dokumentarnega in arhivskega gradiva ter arhivih (ZVDAGA-A, 2014) leta 2014, še preden je stopil v veljavo, sprožil številne negativne odzive strokovne kot laične javnosti, saj je spremenil pravila dostopa do tistega arhivskega gradiva, ki vsebuje najbolj občutljive kategorije osebnih podatkov. Pri pripravi slovenskega arhivskega zakona se je zakonodajalec zgledoval po tujih modelih varstva arhivskega gradiva – tukaj izstopa zlasti praksa nemškega specialnega arhiva Stasi (Jedlitscha, 2011, 24-30; Jenuš, 2016, 66-74). Zakonodajalec se je tako tudi ob izdatnem sodelovanju uporabnikov arhivskega gradiva odločil za anonimizacijo gradiva, določil je, katere kategorije podatkov se ščitijo, dejansko izvajanje varovanja le-teh pa prepustil slovenski javni arhivski službi.

Zastavljalo se je vprašanje, ali novi zakon z vidika dostopa do arhivskega gradiva uvaja napredek, in je torej arhivsko gradivo zaradi njega bolj dostopno, ali pa uvaja nove restrikcije, ki dostop do njega omejujejo. Predlog uzakonitve anonimizacije, ki v slovenski arhivski praksi sicer ni bila neznanka, saj smo jo v omejeni obliki izvajali že od leta 2011,

⁷ Predlog zakona o varstvu osebnih podatkov (ZVOP-2). Dostopno na: <https://www.iusinfo.si/medijsko-sredisce/v-srediscu/284934> (8. 12. 2021).

⁸ Zapisnik Arhivskega sveta št. 625-4/2018-40 z dne 1. 10. 2019. Dostopno na: <https://www.gov.si/zbirke/delovna-telesa/arhivski-svet/> (8. 12. 2021).

je sprožala burno reakcijo stoke, politike in tudi medijev. Arhivski zakon je (po letu 2011) že drugič šel na referendum. Referendumsko vprašanje na glasovnici se je glasilo: »*Ali ste za to, da se uveljavi zakon o spremembah in dopolnitvah zakona o varstvu dokumentarnega in arhivskega gradiva ter arhivih, ki je sprejel državni zbor na seji 28. januarja 2014?*«. Referendum je zahtevala poslanka iz vrst Slovenske demokratske stranke, Eva Irgl, saj naj v primeru arhivske novele ne bi šlo za strokovno odločitev, ampak za politično odločitev z namenom, da se raziskovalcem onemogoči delo zlasti v povezavi z delovanjem nekdanje Službe državne varnosti. To je bil prvi referendum po sprejetju nove referendumske zakonodaje, po kateri je referendum uspešen samo pod pogojem, če proti sprejetju zakona glasuje večina glasujočih in vsaj 20 odstotkov vseh volilnih upravičencev v Republiki Sloveniji. Vendar, ker prag kvoruma ni bil dosežen, je sprejeti predlog novele ZVDAGA ostal v veljavi.⁹

Direktor Arhiva Republike Slovenije, Bojan Cvelfar, je v odzivu po referendumu ocenil, da naj bi državni arhiv s sprejetjem ZVDAGA delal lažje, saj je po naporni referendumski kampanji končno imel priložnost, da se posveti drugim odprtim težavam, potem ko so se zaposleni v državnem arhivu ves čas »*... od sprejetja arhivske novele do danes ukvarjali z vprašanjem referenduma in uveljavitve novele.*«¹⁰ Vendar je ta ugotovitev direktorja državnega arhiva ostala le delno resnična. Novo uzakonjena anonimizacija je namreč pustila številna vprašanja. Kako jo izvajati? Kaj dejansko pomeni za arhive? Kako anonimizacijo doseči na karseda učinkovit način? Že res, da je Arhiv Republike Slovenije od leta 2011 imel nekaj izkušenj z anonimizacijo, a te so nas naučile predvsem, da jo je bilo skoraj nemogoče izvajati brez ustrezne podpore informacijsko-komunikacijske tehnologije.

Novela ZVDAGA je v drugem odstavku 65. člena spremenila vsebine, ki jih vrednotimo kot osebne podatke. Te je omejila na najbolj občutljive, in sicer: zdravstveno stanje, spolno življenje, podatke o žrtvah kaznivih dejanj zoper spolno nedotakljivost, zakonsko zvezo, družino in otroke, zaščitila je storilce kaznivih dejanj in prekrškov (razen kaznivih dejanj in prekrškov oseb, zoper katere je bil voden postopek zaradi nasprotovanja nekdanjemu enopartijskemu režimu – ti podatki so dostopni), versko prepričanje in etnično pripadnost. Vsi naštetih podatki postanejo dostopni za javno uporabo 75 let po nastanku gradiva ali 10 let po smrti posameznika, na katerega se podatki nanašajo, če je datum smrti znan, če ni z drugimi predpisi drugače določeno (ZVDAGA, 2014, drugi odstavek 65. člen).

Glede anonimizacije je pomembna zakonska pridobitev pomenil četrti odstavek 65. člena ZVDAGA. Ta je anonimizacijo namreč *de jure* uzakonil in določa, da se iz arhivskega gradiva izločijo le tisti dokumenti, ki vsebujejo osebne podatke, s čimer se torej ne zapirajo širše enote gradiva. Kadar posamezni dokumenti vsebujejo podatke, ki so nedostopni, poleg njih pa tudi podatke, do katerih bi bilo mogoče dostopati, so arhivi dolžni za dostop pripraviti anonimizirano kopijo dokumenta, na kateri so varovani

⁹ Zaradi izjemno nizke volilne udeležbe referendum o arhivih ni uspel. Dostopno na: <https://www.24ur.com/novice/slovenija/volivci-odlocajo-o-usodi-arhivske-novele.html> (8. 12. 2021); Na volišča 4. maja. SDS napoveduje ustavno presojo referendumskega odloka. Dostopno na: <https://www.rtvlo.si/slovenija/na-volisca-4-maja-sds-napoveduje-ustavno-presajo-referendumskega-odloka/333524> (8.12.2021); "Referendum je poenotil narod. Vsi se strinjajo, da so zmagali.". Dostopno na: <https://www.rtvlo.si/slovenija/referendum-je-poenotil-narod-vsi-se-strinjajo-da-so-zmagali/338926> (8. 12. 2021).

¹⁰ "Referendum je poenotil narod. Vsi se strinjajo, da so zmagali.". Dostopno na: <https://www.rtvlo.si/slovenija/referendum-je-poenotil-narod-vsi-se-strinjajo-da-so-zmagali/338926> (8. 12. 2021).

podatki prekriti tako, da neposreden ali posreden vpogled ni omogočen (ZVDAGA, 2014, četrti odstavek 65. člen).

4. IZZIVI ANONIMIZACIJE

Če pustimo vprašanje določanja ali vrednotenja osebnih podatkov ob strani, saj bi to vprašanje dejansko zahtevalo lasten interdisciplinarni prispevek, prinaša anonimizacija s seboj povsem tehnične zaplete. Prvi izhajajo iz absolutnega pomanjkanja kadra, ki anonimizacijo izvaja, drugi pa iz 'nosilcev', na katerih se informacije v arhivskem gradivu nahajajo. Kljub dejstvu, da smo v hibridni dobi poslovanja, je anonimizaciji podvrženo zlasti fizično gradivo, mlajše od 75 let; to pa hranimo na papirju, na mikrofilmskih kolutih in v obliki mikrofilmskih žepkov (Jenuš, 2013, 70).

Vsaka od teh oblik 'nosilca' ima svoje zakonitosti in v postopku anonimizacije prinaša svoje izzive. Če posplošimo, se anonimizacija lahko izvaja na dva načina: ročno na fizičnih kopijah dokumentov ali elektronsko na digitalizatih ali izvorno digitalnem gradivu.

Zdi se, da je najlažje anonimizirati gradivo na papirju, vendar je postopek kljub vsemu zapleten. Ko arhivist najde dokument, ki vsebuje varovane kategorije podatkov, izvorno arhivsko gradivo izloči iz celote. Na njegovo mesto priloži list, na katerem uporabnika opozori, da je originalni dokument zaradi vsebovanja občutljivih osebnih podatkov po ZVDAGA ali ZAGOPP¹¹ izločen in nedostopen. Izvirniki arhivskega gradiva

¹¹ *Zakon o arhivskem gradivu, ki vsebuje osebne podatke o zdravljenju pacienta (Uradni list RS, št. 85/16; ZAGOPP) je posledica razprave o pomenu varovanja osebnih podatkov v arhivskem gradivu med referendumsko razpravo leta 2014. Do pomembnosti varovanja ustavno zajamčenih pravic do varstva osebnih podatkov, zasebnega dostojanstva in varstva pravic zasebnosti ter osebnostnih pravic v arhivskem gradivu se je namreč leta 2014 opredelilo tudi Ustavno sodišče Republike Slovenije v odločbi št. U-I-70/12-14, ko je odločilo, da je ZVDAGA v neskladju z 38. in 34. členom Ustave RS v primerih arhiviranja gradiva izvajalcev zdravstvene dejavnosti, ki vsebuje osebne podatke o zdravljenju pacienta. Iz odločbe izhaja, da ZVDAGA v teh primerih ne upošteva v zadostni meri pomena varstva občutljivih osebnih podatkov, ki so vsebovani v zdravstveni dokumentaciji in katerih razkritje lahko pomeni tudi poseg v osebno dostojanstvo pacienta in njegovih bližnjih. Prav tako v zadostni meri ne upošteva zdravnikove molčečnosti kot bistveno predpostavko zaupnosti odnosa med pacientom in zdravnikom, kakor tudi ustavnih jamstev, ki so pacientom zagotovljena zaradi uresničevanja njihove pravice do varstva osebnih podatkov, s posebnim poudarkom na prepovedi uporabe osebnih podatkov v nasprotju z namenom njihovega zbiranja. Zaradi spoštovanja pravic pacientov iz 38. in 34. člena Ustave RS je Ustavno sodišče določilo, da se za arhiviranje zdravstvene dokumentacije, ki vsebuje osebne podatke o zdravljenju pacientov, sprejme in uveljavi drugačna (specialna) zakonska ureditev za gradivo izvajalcev zdravstvene dejavnosti, ki so po ZVDAGA opredeljeni kot javnopravne osebe.*

Odločba Ustavnega sodišča je bila uresničena s sprejetjem ZAGOPP, le-ta določa posebno ureditev varstva arhivskega gradiva, ki vsebuje osebne podatke o zdravljenju pacienta, in tudi pogoje ter postopek, v katerem je izjemoma mogoče pridobiti dostop do tega arhivskega gradiva. Dostop do tega arhivskega gradiva je omogočen le pacientu, na katerega se nanaša, ter osebam, ki imajo pravico do seznanitve z zdravstveno dokumentacijo pacienta po njegovi smrti v skladu z določbami zakona, ki ureja pacientove pravice, in zakona, ki ureja zdravstveno dejavnost. Vendar tudi ZAGOPP za namene znanstvenega raziskovanja omogoča izjemni dostop do arhivskega gradiva, ki vsebuje osebne podatke o zdravljenju pacienta. V skladu s 6. členom ZAGOPP je izjemni dostop mogoč raziskovalcem, registriranim v evidenci izvajalcev raziskovalne in razvojne dejavnosti pri agenciji, pristojni za raziskovalno dejavnost, in sicer: na podlagi pisnega soglasja pacienta, na katerega se gradivo nanaša; po smrti pacienta in če pisnega soglasja pacienta iz prve alineje 3. odstavka 6. člena ni, na podlagi pisnega soglasja zakonitih dedičev pacienta do drugega dednega reda, razen če je pacient to izrecno pisno prepovedal; če gre za zbirke podatkov, iz katerih ni mogoče razbrati identitete posameznega pacienta niti ob nesorazmerno velikih naporih, stroških ali porabi časa (ZAGOPP, 6. člen).

Arhivsko gradivo, ki se nahaja v javnih arhivih, je za namene znanstvenega raziskovanja raziskovalcem, registriranim v evidenci izvajalcev raziskovalne in razvojne dejavnosti pri agenciji, pristojni za raziskovalno dejavnost, dostopno po smrti pacienta tudi, če ni pisnega soglasja pacienta iz prve alineje

so izloženi v zalepljeni kuverti na koncu gradiva. Arhivist nato, v primeru gradiva z varovanimi podatki po ZVDAGA, izdela kopijo dokumenta. Na njej občutljive osebne podatke prekrije s pomočjo črnila, kopijo pa ponovno kopira, saj je informacija pod določenim kotom še vedno berljiva. Če so dokumenti na papirju in je teh dokumentov le nekaj strani, pride v poštev fizično kopiranje dokumentov in prekrivanje osebnih podatkov na kopijah z uporabo ravnala in črnila, kadar pa govorimo o več sto ali celo tisoč strani dokumentov, ali obravnavi analognega gradiva na mikrofilmih ali mikrofiših, fizična (tj. ročna) anonimizacija ne pride v poštev (Jenuš, 2013, 70).

Gradivo na mikrofilmskih kolutih in mikrofilmskih žepkih predstavlja v postopku anonimizacije poseben izziv. Gradivo je problematično zaradi formata, kvalitete zapisa, kakor tudi postopka pregledovanja samega. V teh primerih je potrebno anonimizacijo izvajati elektronsko. Da lahko to dosežemo, je arhivsko gradivo najprej potrebno digitalizirati. In ravno digitalizacija arhivskega gradiva je pravzaprav pozitiven 'stranski' učinek izvajanja anonimizacije, hkrati pa ena od prednostnih nalog slovenskih arhivov, saj omogoča ne le dolgoročno hrambo arhivskega gradiva v digitalni obliki kot ukrep materialne varnosti, ampak večja njegova dostopnost (Jenuš, 2013, 70).

Kot posledica dveh referendumov o arhivskem zakonu v letu 2011 in 2014 se je povečalo zanimanje javnosti zlasti za gradivo nekdanje jugoslovanske tajne politične policije, ki ga hranimo v arhivskem fondu SI AS 1931, Republiški sekretariat za notranje zadeve Socialistične republike Slovenije¹². Povpraševanja je bilo veliko, a zaradi nerazdelanih metodologij varovanja osebnih podatkov in pomanjkanja kadra dohajanje uporabniških zahtevam skoraj ni bilo izvedljivo.¹³

Arhiv Republike Slovenije je zato pričel s pospešeno sistematično digitalizacijo arhivskega gradiva. S pomočjo ponudnikov storitev, ki so certificirani pri državnem arhivu, se je lotil digitalizacije gradiva Republiškega sekretariata za notranje zadeve: 4000 osebnih dosjejev, 45.000 kratkih dosjejev ali tako imenovanih terenskih preverk in 80.000 dosjejev jugoslovanske emigracije. Postopoma so tem zaokroženim celotam dodajali tudi gradiva, vezana na povojne politične procese, pregone posameznih grupacij, letna poročila itd. Fizično gradivo je obsegalo tako dokumente na papirju kot tudi na mikrofilmu in mikrofišu (Jenuš, 2016, 66-74).

Čeprav smo digitalizacijo arhivskega gradiva izvajali v dveh formatih – pdf z vključeno optično prepoznavo znakov (OCR; ang. Optical Character Recognition¹⁴) in slikovni format .jpg, celotnega potenciala digitalizatorov nismo mogli koristiti. Razlogov za

3. odstavka 6. člena ali če pacient tega ni prepovedal in njegovi zakoniti dediči temu ne nasprotujejo, na podlagi obrazložene odločitve arhivske komisije v skladu z zakonom, ki ureja varstvo arhivskega gradiva, h kateri je izdano obrazloženo soglasje komisije, pristojne za medicinsko-etična vprašanja. Komisiji lahko dostop omogočita le, če je utemeljen z vidika javnega interesa na področju napredka znanosti, ki prevladuje nad interesom za nedostopnost, in če bi se lahko z dostopom dosegel predviden znanstveni cilj, ki ga ni mogoče doseči na drug način. Dostop do arhivskega gradiva po 6. členu ZAGOPP pa mora biti ustrezen in po obsegu primeren glede na namen in obseg izvedene raziskave, uporabnik pa gradiva ne sme uporabljati za druge namene (ZAGOPP, 6. člen).

Odločba Ustavnega sodišča v postopku ocene ustavnosti Zakona o varstvu dokumentarnega in arhivskega gradiva ter arhivih št. U-I-70/12-14 z dne 21. 3. 2014. Dostopno na: <https://www.us-rs.si/wp-content/uploads/2020/02/u-i-70-12.pdf> (8. 12. 2021); Tiskovno sporočilo v zvezi z odločbo Ustavnega sodišča št. U-I-70/12-14 z dne 21. 3. 2014. Dostopno na: <https://www.us-rs.si/wp-content/uploads/2020/02/u-i-70-12.-tiskovno.sporocilo.pdf> (8. 12. 2021).

¹² SI AS 1931 Republiški sekretariat za notranje zadeve Socialistične republike Slovenije, 1918-2004 (Fond/zbirka). Dostopno na: <http://arsq.gov.si/Query/detail.aspx?ID=27578> (8. 12. 2021).

¹³ Arhivska novela tudi leto dni po uveljavitvi deležna nasprotujočih ocen. Dostopno na: <https://www.iusinfo.si/medijsko-sredisce/dnevne-novice/147493> (8. 12. 2021).

¹⁴ Kaj je optično prepoznavanje znakov (OCR)? Dostopno na: <https://sl.eyewated.com/kaj-je-opticno-prepoznavanje-znakov-ocr/> (8. 12. 2021).

to je bilo vsekakor več. Anonimizacija se na digitalizatih v formatu pdf ni izvajala, ker OCR zdaleč ni bil zanesljiv, hkrati pa tudi nismo imeli na voljo potrebne programske opreme. V stiski ali iznajdljivosti smo se zatekli k izvajanju anonimizacije na datotekah .jpg, kjer so arhivisti z uporabo programa MS Windows Slikar na digitalizatih prekrili varovane podatke. Čeprav je tovrstna anonimizacija v digitalnem svetu prinašala malenkost več svobode in manevrskega prostora ter arhivistu prihrani nekaj časa zaradi opustitve neprestanega fizičnega kopiranja dokumentov, ima svoje slabosti in je pravzaprav zgolj rešitev v sili. Postopki priprave anonimizirane kopije so vzeli dneve, pogosto tudi tedne; dejstvo, da orodje Slikar ni namenjeno anonimizaciji, pa celotne situacije seveda ni olajšalo.

5. PREMIK NAPREJ – PROJEKT VZPOSTAVITVE SLOVENSKEGA ELEKTRONSKEGA ARHIVA (e-ARH.si)

Premik naprej smo slovenski javni arhivi doživeli leta 2016. Marca istega leta je namreč Vlada Republike Slovenije sprejela Strategijo in izvedbeni načrt razvoja slovenskega elektronskega arhiva za obdobje 2016–2020¹⁵. Med strateške cilje projekta smo, v sklopu nalog namenjenih povečanju dostopnosti arhivskega gradiva, zapisali tudi cilj *»nadgradnje obstoječih ali nakup novih programskih orodij za avtomatizirano razpoznavanje in anonimizacijo z zakonom varovanih podatkov«*. V strategiji smo se zavezali tudi, da bomo rešitev ponudili v novo nastajajočem državnem računalniškem oblaku, da bi lahko služila tudi drugim državnim organom, ki so zavezani k varstvu osebnih podatkov. Naloga je bila zaupana Kompetenčnemu centru 4. Člani delovne skupine smo si nadali ključnik #dostopnost, saj so se področja našega dela dotikala vzpostavitve Virtualne arhivske čitalnice, prilagoditvam arhivskega gradiva in odpravljanju arhitektonskih preprek za ranljive skupine, kot tudi večanju dostopnosti arhivskih dokumentov s pomočjo anonimizacije (Strategija e-ARH.si, 2016).

Strategija je bila odgovor slovenskih javnih arhivov na dobo hibridnega poslovanja, digitalizacije našega vsakdana in potreb javnih arhivov po modernizaciji naših storitev. Bila je popolnoma v duhu časa in slovenske usmerjenosti v digitalizacijo družbe¹⁶. Republika Slovenija se je namreč v strategiji Digitalna Slovenija 2020 – Strategija razvoja informacijske družbe do leta 2020 (2016) izrekla za koriščenje informacijsko-komunikacijske tehnologije tudi na področju kulturne dediščine. Digitalizacija, digitalne vsebine in e-storitve ter javna dostopnost naj bi namreč tudi na področju kulture bile bistvenega pomena za ustvarjanje dodane vrednosti. Pomenile naj bi osnovo za uporabo digitalnih kulturnih vsebin v procesih izobraževanja, usposabljanja, raziskovanja, vseživljenjskega učenja in pridobivanja znanja. Vse to pa naj bi prispevalo k promocije kulturne dediščine in države. Glede na vpetost Slovenije v globalno digitalno skupnost procesi digitalizacije nezadržno spreminjajo tudi slovensko okolje (Hajtnik, 2019, 49-56). *»Digitalne kulturne vsebine so eden najučinkovitejših instrumentov za dvig prepoznavnosti in konkurenčnosti Slovenije v Evropi in svetu.«* Hkrati pa tudi prispevajo k dvigu zaupanja javnosti v delo države in njenih organov (Digitalna Slovenija 2020, 2016; Jenuš, 2018, 136-137).

To je bilo nedvomno eno od vodil, ki nas je vodilo v postopku pridobitve orodja za anonimizacijo. V sklopu nalog, povezanih z anonimizacijo, smo se lotili analize temeljnih

¹⁵ Strategija in izvedbeni načrt razvoja slovenskega elektronskega arhiva za obdobje 2016–2020. Dostopno na: https://www.gov.si/assets/organi-v-sestavi/Arhiv-RS/Projekt-e-ARH-si/PR-material/Strategija_e-ARH_si_2016-2020_1.0.pdf (8. 12. 2021).

¹⁶ Digitalizacija družbe. Dostopno na: <https://www.gov.si/teme/digitalizacija-druzbe/> (8. 12. 2021).

delovnih procesov, ki smo jih arhivisti izvajali do tedaj. Analizirali smo posamezne korake, prednosti in slabosti ter izpostavili želje arhivistov.

Naša želja je bila pridobiti programsko rešitev, ki bi omogočala izrabo optične prepoznavne znakov, predvidevanja in označevanja besed. Program bi moral omogočati uvoz digitalizator v programsko okolje, kjer bi s signirjem označil tiste dele arhivskega gradiva, ki uporabnikom ne smejo biti dostopni. Na tak način bi izvajalec anonimizacije videl, katere dele arhivskega gradiva je označil in zakaj jih je potrebno prekriti – brez potrebe po podvojenemu odpiranju dokumentov. Hkrati pa bi ta postopek omogočil, da bi anonimizacijo lahko izvajalo več oseb. Praviloma se namreč arhivisti pri anonimizaciji ob dilemah – in teh je veliko – posvetujemo. Program bi nato moral omogočiti, da se z izvozom pripravi anonimizirana kopija arhivskega dokumenta, kot tudi njegova interna delovna kopija s signirji.

Prvi korak pri iskanju rešitev je bil, da smo se ozrli po programskih orodjih, s katerimi smo že razpolagali. Eno takšnih je bilo denimo orodje Adobe Acrobat Professional. Ta v omejeni obliki omogoča anonimizacijo gradiva z uporabo OCR, do neke mere je omogočeno tudi ustvarjanje geslovnika (tezavra besed) za avtomatizirano iskanje. Izdelali smo natančna navodila, ki uporabnika vodijo po posameznih korakih. Od paketne pretvorbe digitalizator, vključitve OCR in ureditve prepoznanega besedila do ureditve geslovnika, ki služi za izvajanje anonimizacije. Po ureditvi teh korakov navodila uporabnika vodijo k redigiranju – torej anonimizaciji gradiva v pdf formatu.

Navodila so pomenila začetek, prvi korak k iskanju rešitve, hkrati pa tudi spoznanje, da anonimizacija na tržišču nastopa pod različnimi pojmi: redact, redaction software, redigiranje, programska oprema za redigiranje.

6. OD ANALIZE TRŽIŠČA DO JAVNEGA RAZPISA

Naš naslednji korak je bila analiza tržišča. Kot smo ugotovili v uvodu prispevka, varstvo osebnih podatkov ni nekaj, kar bi zadevalo le arhive. Zahvaljujoč GDPR in implementaciji v evropske nacionalne zakonodaje se je na tržišču pojavilo veliko programskih orodij, namenjenih anonimizaciji, kar predstavlja donosen posel.

Z njo se ukvarjajo tudi slovenski ponudniki. Enega prvih stikov, ki se nam je zdel zelo obetaven, smo vzpostavili z Laboratorijem za umetno inteligenco Inštituta Jožef Štefan. Ta je za anonimizacijo sodb Višjega in Vrhovnega sodišča RS razvil orodje Tacita.

Tacita s pomočjo strojnega učenja pomaga uslužbencem sodišč z avtomatsko detekcijo delov besedila, ki jih je pred javno objavo sodb treba anonimizirati. Vhodni dokumenti analize so bili neanonimizirani dokumenti v open office formatu dot in anonimizirani dokumenti v formatu json, ki jih sodišča objavljajo na svojem spletnem portalu. Šlo je za dejanske anonimizirane dokumente, ki so odražali prakso anonimizacije sodišč. V postopku predprocesiranja je prišlo do ekstrakcije besedila in postopka lematizacije, v kateri se je izdelal glosar/tezaver besed, ki so bile najpogostejše anonimizirane. Tako nastane statistični model, ki na podlagi predvidevanja napove, ali je besede v besedilu treba anonimizirati. Sistem deluje na podlagi besedilnih vzorcev, v različni dolžini izpisa in »IF« stavkov, ki preverjajo, ali so posamezne predpostavke izpolnjene. Program iz glosarja črpa najpogostejše anonimizirane besede ali sklope besed in se na podlagi tega (prim. velikih črk, ki nakazujejo ime, priimek, ulico) odloča, ali je besedo treba prekriti. V času naše raziskave so sistem preizkusili s petkratnim prečnim preverjanjem (ang. 5-fold cross validation), ki je pokazal natančnost v 72,8 odstotkih. Slednje kaže, da program seveda ne zagotavlja popolne gotovosti, a je pomemben korak v prihodnost, saj lajša postopke anonimizacije. Procesiranje ene

sodbe naj bi bilo hitro in naj bi trajalo le eno sekundo. Omogoča tudi učenje sistema in s tem večanje geslovnika. Uporabniški vmesnik omogoča nastavitve občutljivosti besed za anonimizacijo, ki uslužbencu, ki jo opravlja, služi kot orodje, ki anonimizacijo pospeši, nikakor pa uslužbenca ne more nadomestiti. V postopku priprave orodja Tacita se je pokazala tudi težava konsistentnosti prakse pri anonimizaciji, ki kaže na to, da je treba poenotiti prakso anonimizacije, kar bi dodatno izboljšalo sistem prepoznavanja besed. Uporabnost rešitve Tacita za potrebe arhivov se v prvi vrsti ponuja za anonimizacijo arhivskega gradiva sodne provenience, saj je program s tehnikami strojnega učenja prilagodljiv in omogoča prilagajanje geslovnika zakonskim potrebam. Predvsem bi bil uporaben kot orodje za anonimizacijo strukturiranih besedil s predpisano formo – ob predpostavki, da je arhivsko gradivo v ustreznem (besedilnem) formatu. Njegovo uporabnost za potrebe anonimizacije starejšega 'nestrukturiranega' arhivskega gradiva pa se je po preizkusih na praktičnih primerih izkazala za pomanjkljivo in za arhive nezadostno.¹⁷

Rešitev se je zdela na dlani, idealna. No, skoraj. Žal v orodju Tacita nismo našli rešitve. Orodje je briljanten dokaz slovenskega znanja, a žal ni bilo kos arhivskemu gradivu. Po testu na živem arhivskem gradivu smo v sodelovanju z Inštitutom Jožef Štefan ugotovili, da orodje žal v primeru arhivskih dokumentov doseže zgolj nekajodstotno točnost in še v teh primerih je šlo bolj za srečo kot dejansko gotovo določitev besed za anonimizacijo. Težava arhivskega gradiva je, da povečini podatki niso strukturirani, dokumenti nimajo nujno predpisane forme oz. besedišča in je zato strojno učenje zelo zahtevno. Druga težava, ki je še bolj izrazita, pa je kvaliteta. Vsi programi strojnega učenja ali analize besedila so odvisni od prepoznavne znakov. Napake, pomanjkljivosti pomenijo veliko coklo. In OCR starih dokumentov je velikokrat pomanjkljiv.

Nazaj na začetek torej ...

Sledila je analiza orodij, ki jih koristijo sorodne arhivske institucije v tujini, in orodij na trgu. Programske rešitve podjetij, kot so TIS Group & Newton Technologies, IBM Data Capture, Objective redact, Rapid redact, Redact express, so vsa obetala anonimizacijo arhivskega gradiva. Omogočala so nekatere željene funkcionalnosti: uvoz bodisi slikovnih formatov ali pdf z vključenim OCR, in nekatere zahtevne funkcionalnosti. A se je pri večini zataknilo bodisi pri analizi dokumentov, izvozu ali mehanizmihi anonimizacije.¹⁸

Podjetje TIS Group & Newton Technologies je razvilo orodje, ki omogoča, da občutljive podatke za anonimizacijo zazna sistem strojnega učenja K2, nato pa se ti podatki odstranijo iz wordovih dokumentov ali zatemnijo v primeru dokumentov pdf. Omejitve testne različice, ki smo jo imeli na voljo, so bile, da je bil koncept pripravljen samo za datoteke .doc in .txt ter da se je lahko anonimiziralo samo imena in priimke. V kasnejši produkcijski različici naj bi bilo možno pokriti še ostale funkcionalnosti, kot so drugi formati dokumentov, in zaznati ter anonimizirati tudi druga dejstva oz. druge nastavljene parametre.¹⁹

IBM-ovo orodje DataCap je bilo primarno razvito za poslovno dokumentacijo in omogoča zajem, prepoznavo in klasifikacijo poslovnih dokumentov ter ekstrakcijo informacij. Orodje omogoča uvoz .pdf ali slikovnih formatov in transformacijo v obliko za

¹⁷ Interni dokument projekta e-ARH.si: e-ARH.si_KC4.4.1_Poročilo orodja za anonimizacijo TACITA, 20. 3. 2017.

¹⁸ Programsko orodje za avtomatizirano razpoznavanje in zakritje z zakonom varovanih podatkov (anonimizacijo). Dostopno na: <https://nio.gov.si/nio/asset/programsko+orodje+za+avtomatizirano+razpoznavanje+in+zakritje+z+zakonom+varovanih+podatkov> (8. 12. 2021).

¹⁹ Prav tam.

obdelavo, vključno z OCR. Objective Redact je programska rešitev, ki ponuja napredno anonimizacijo dokumentov. Procesira datoteke iz različnih virov, ne da jo uporabnik natisne ali shrani kot pdf. Orodje lahko išče po frazah, regularnih izrazih ali posameznih besedah. Besede se tako anonimizira ireverzibilno iz različnih vrst datotek. Redact Express pa je orodje, ki omogoča visoko nastavljive načine anonimizacije, podporo za različne formate uvoza in izvoza, ireverzibilne oblike anonimizacije ter avtomatizacijo anonimizacije strukturiranih dokumentov.²⁰

V poslovno analizo smo zajeli tudi rešitev za izvajanje anonimizacije, ki jo uporablja Zgodovinski arhiv madžarske službe državne varnosti (ABTL). Dokumentni sistem, v katerem ABTL upravlja s svojim arhivskim gradivom, je prilagojen za potrebe arhivske službe. Njegove glavne funkcionalnosti so obdelava podatkov, obdelava osebnih podatkov, upravljanje z digitalizati, ki se jih v bazo uvozi, in izvajanje/opravljanje anonimizacije. Z vidika naše naloge je najbolj zanimiva funkcionalnost anonimizacija. Delovni tok predvideva, da se vhodni dokument (na papirju ali mikrofišu) najprej digitalizira v formatu .tiff, nato se digitalizat uvozi v bazo, pri čemer se zajamejo identifikacijski in tehnični metapodatki. Ko administrator prične z izvajanjem anonimizacije, se ustvari 'dvojnik' zadeve, ki je označen s podatki izvirnika in shranjen v bazi kot anonimizirana zadeva. Strokovni sodelavec, ki izvaja anonimizacijo z uporabo programskega orodja s funkcionalnostjo Slikarja, ročno označuje dele besedila, ki jih je treba prekriti. Funkcija v bistvu vstavlja kvadratke, ki so transparentni. Tako je besedilo, ki ga prekrivajo, označeno, a vidno bralcu. Ko sodelavec zaključi s pregledom in postopkom anonimizacije, izvozi dokument .pdf, kjer so označeni deli besedila prekriti in uporabniku niso vidni. Za lažjo identifikacijo so dokumenti, ki jih posredujejo uporabniku, označeni z unikatno 'lebdečo' identifikacijsko vrstico (watermark), ki vsebuje signaturo fonda in naziv popisne enote. Posebna funkcionalnost orodja je t. i. združevanje posameznih delov dokumentov (text compiling), ki izhaja iz madžarske zakonodaje. Ta omogoča, da uporabniki vidijo tiste dele dokumentov, ki se nanje nanašajo. Strokovni sodelavec, podobno kot v postopku anonimizacije, s transparentnim kvadratom označi dele dokumentov, ki jih je treba pripraviti za strankino rabo, in pri izvozu namesto funkcije 'prekrij označene dele besedila' vključi 'izvozi izbrani del besedila'. Na tak način nastane nov dokument z izrezi posameznih dokumentov, ki jih stranka dobi v vpogled in ki se nanjo nanašajo. ABTL je v postopkih digitalizacije in anonimizacije poskusno uvedel tudi OCR, čigar raba pa je omejena in je zato v postopkih anonimizacije ne uporabljajo. Težava so rokopisni dokumenti in slabo berljivi tipkopisni dokumenti, ki učinkovitost OCR zmanjšujejo. Še več omejitev pa prinašajo posebne črke madžarske abecede. Sicer se nadejajo, da bi v prihodnosti OCR koristili za potrebe iskanja informacij in podatkov (data mining), a implementacija te funkcionalnosti zahteva dodatne finančne vložke. Programske rešitve, ki jih v postopkih upravljanja z dokumenti, ki vsebujejo osebne podatke, in anonimizacije uporablja ABTL, so zanimive, a prilagojene ne le zakonskim okvirom madžarskih arhivov, ampak tudi madžarskemu jeziku. Sistem temelji na kombinaciji podatkovne baze in dodanih funkcionalnosti, ki omogočajo popis in zajem metapodatkov, kakor tudi anonimizacijo in upravljanje z digitalizati. Anonimizacija se izvaja ročno; avtomatizmov ali funkcionalnosti, pri katerih bi sistem strokovnemu sodelavcu predlagal, katere dele gradiva je treba prekriti, ni. Tudi ni funkcionalnosti, ki bi strokovnemu sodelavcu pri vstavljanju kvadratkov kakorkoli pomagala ali mu olajšala delo. Transparentnost izbranega besedila je dobrodošla, kakor tudi izvoz generiranega

²⁰ Programsko orodje za avtomatizirano razpoznavanje in zakritje z zakonom varovanih podatkov (anonimizacija). Dostopno na: <https://nio.gov.si/nio/asset/programsko+orodje+za+avtomatizirano+razpoznavanje+in+zakritje+z+zakonom+varovanih+podatkov> (8. 12. 2021).

in prekritega dokumenta .pdf po zaključku anonimizacije. Vendar je pri obsežnih dokumentih in zadevah izvoz lahko zelo dolgotrajen.²¹

Po raziskavi tržišča in pridobljenih izkušnjah smo se odločili za izvedbo javnega naročila. Izdelali smo specifikacijo naših pričakovanj in zahtev. Predmet javnega naročila je bila programska rešitev, ki bi omogočala izvajanje anonimizacije, bodisi digitaliziranega ali izvirno digitalnega gradiva najmanj v formatih .pdf, .jpg ali .tiff z uporabo optične prepoznave znakov. Seveda bi morala delovati najmanj v okolju MS Windows in omogočati namestitev strežniške licence ter sočasno uporabo vsaj desetih uporabnikov z delovnimi postajami.

Natančnejše tehnične specifikacije so bile, da mora rešitev omogočati:

- uvoz gradiva posameznega dokumenta ali skupine dokumentov za anonimizacijo (najmanj v formatih .pdf, .jpg oz. .tiff);
- izvedbo avtomatizirane optične prepoznave znakov (najmanj OCR);
- izdelavo in dodajanje iskanih nizov ali besed v bazo (geslovnik), ki so predmet željene anonimizacije;
- možnost uvoza prednastavljenih baz, oziroma kreiranje in urejanje novih (obstajati mora tudi možnost izvoza baz v druga okolja);
- izvedbo avtomatizirane anonimizacije posameznega dokumenta ali skupine dokumentov;
- ročno (s signiranjem) označbo besed, delov besedila, predmeta ali področja gradiva, ki so predmet anonimizacije; označeni deli besedila naj se vnašajo v bazo (geslovnik), ki se stalno veča in uporabnikom s predvidevanjem besed pokaže, kateri deli gradiva so morebiti takšni, da jih je treba anonimizirati;
- uporabniški vmesnik, ki uporabniku v fazi pregledovanja dokumenta oz. besedila omogoča, da predlagane iskalne nize ali besede za anonimizacijo sprejme ali jih zavrne.²²

Na javni razpis so se prijavila tri podjetja. Podjetje TIS d. o. o. je ponudilo TIS Anonimizator, programsko rešitev za iskanje, označevanje in zakrivanje občutljivih podatkov na digitalnih dokumentih. Rešitev je sestavljena iz več programskih sklopov in je zasnovana trinivojsko. Zajem dokumentov se vrši s pomočjo rešitve IBM Datacap, ki zagotavlja optimalno optično branje dokumenta in pripravo podatkov. Operativna logika, kjer se izvaja anonimizacija podatkov, je zasnovana s pomočjo programskega jezika Python. Podatkovni sloj rešitve je zasnovan na tehnologiji QT5, pri čemer rešitev uporablja podatkovno bazo SQLite. Programska rešitev omogoča, da se z izvozom pripravi uporabniška kopija gradiva, v kateri so podatki prekriti tako, da neposreden ali posreden vpogled ni omogočen (kar je tudi v skladu s četrnim odstavkom 65. člena ZVDAGA) in da povrnitev v izvirno stanje dokumenta nikakor ni mogoča. Uporabnik programske rešitve ima v uporabniškem vmesniku možnost izbire videza oznake za anonimizacijo in izvoznega formata. Programska rešitev omogoča izvoz v več formatov hkrati (.jpg oz. .tif in .pdf). Slednji vključuje OCR. Ohranita se obe različici dokumenta, ki sta predmet postopka anonimizacije – tako anonimizirana različica kot tudi delovna

²¹ Programsko orodje za avtomatizirano razpoznavanje in zakritje z zakonom varovanih podatkov (anonimizacija). Dostopno na: <https://nio.gov.si/nio/asset/programsko+orodje+za+avtomatizirano+razpoznavanje+in+zakritje+z+zakonom+varovanih+podatkov> (8. 12. 2021).

²² Specifikacija zahtevanih funkcionalnosti programske rešitve za avtomatizirano razpoznavanje in anonimizacijo z zakonom varovanih podatkov. JN-20-004; št. dokumenta 4301-2/2020/2 z dne 30. 1. 2021.

različica, vključno s pripadajočo bazo iskalnih besed ali besedilnih nizov (geslovnik), v katero se po potrebi lahko še dodajajo dodatne zahteve za prekritje podatkov.

Drugi ponudnik, Mikrografija d. o. o., je ponudil programsko opremo mSCAN, ki ponuja nov in enostaven način skeniranja dokumentov, slik ali fotografij in njihovo indeksacijo. Hkrati obvlada prepoznavo besedil iz formularjev in strukturiranih dokumentov. Odlične rezultate dosega tudi pri prepoznavi besedila iz nestrukturiranih dokumentov, ki danes predstavljajo okoli 90 odstotkov poslovne korespondence. Izvoz dokumentov za nadaljnjo obdelavo je izvedljiv; neposredno v dokumentnem sistemu mDocs, preko FTP, v e-hrambi mSef, preko lokalnega izvoza oz. v morebitne ostale zunanje sisteme. MScan omogoča izvoz v datotekah .tiff, .jpg in .pdf, pri čemer so določene oblike izvoza že primerne za dolgoročno elektronsko hrambo dokumentacije in omogočajo enostavno iskanje po vsebini dokumentov.

Podjetje Akademika d. o. o., ki je bilo glede na kriterije izbrano kot najugodnejši ponudnik, je ponudilo orodje za anonimizacijo oz. OpenText!BravaDesktop in OpenText!BravaBlazon. V okviru sklenjene pogodbe se je ponudnik zavezal, da:

- bo zagotovil dobavo programske rešitve za avtomatizirano razpoznavanje in anonimizacijo z zakonom varovanih podatkov;
- bo izvedel namestitvev programske rešitve na strežniške sisteme (v testnem in produkcijskem okolju) na lokaciji naročnika;
- so dovoljene morebitne spremembe in dopolnitve programske rešitve na podlagi zahtevkov za dopolnitev, v katerih so naknadno specificirane vsebinske in tehnične zahteve;
- bo izvedel predstavitve in šolanje za uporabo programske rešitve v obsegu vsaj dveh delovnih dni za največ 10 uporabnikov;
- bo dobavil neizključno in prenosljivo licenco za uporabo programske rešitve za 10 primerkov;
- bo zagotovil licenčno vzdrževanje programske rešitve (uporabo najnovejših različic licenčne programske opreme, možnost prijave napake v licenčni programski opremi) za dobo treh let po namestitvi programske rešitve v okolje za redno rabo naročnika ter z možnostjo sklenitve kasnejšega licenčnega vzdrževanja;
- bo zagotavljal garancijo za dobavljeno programsko opremo za dobo treh let od končnega prevzema.

7. PROGRAMSKO ORODJE ZA ANONIMIZACIJO VAROVANIH PODATKOV

Programska rešitev je produkt podjetja OpenText²³, ki je vodilno v svetu na področju sistemov za upravljanje z dokumenti oziroma informacijskimi sistemi za upravljanje z dokumenti (angl. Document Management System). Gre za rešitev, ki se jo namesti na računalnik oz. strežnik in velja za neomejeno število uporabnikov (t. i. seat license).

Postopek obdelave oz. anonimizacije se izvede po določenem vrstnem redu. Najprej se izvede uvoz dokumenta, ki je predviden za anonimizacijo (orodje podpira uvoz zelo velikega števila različnih formatov dokumentov – .pdf, .jpg, .png, .tiff itn.). Sledi izvedba »psevdoanonimizacije« (povratna oblika prekrivanja besed) in shranjevanje

²³ OpenText Brava! Dostopno na: <https://www.opentext.com/products-and-solutions/products/enterprise-content-management/opentext-brava> (10. 12. 2021).

anonimizirane datoteke (v formatu .xrl ali .mrk) ter morebitno deljenje datotek drugim anonimizatorjem. Ob koncu sledi zaključek oz. »objava« anonimiziranega dokumenta (nepovratno anonimiziran dokument v željenem formatu).

Funkcionalnosti orodja za anonimizacijo so raznovrstne in so v grobem naslednje:

- »Select« – označimo lahko predhodno anonimizirane dele dokumenta in jih dodatno oz. ponovno obdelamo (brišemo, kopiramo oz. premikamo).
- »Redact Area« – ročno oblikujemo površino, na kateri se v dokumentu nahajajo občutljivi ali zaupni podatki, ki jih želimo anonimizirati.
- »Redact Polygon« – ročno določimo eno ali več poljubnih večkotnih področij anonimizacije, vse ostale funkcionalnosti so podobne funkciji »Redact Area«.
- »Redact Text« – ročno določimo področje anonimizacije v tekstu, vendar brez slik.
- »Allow Area (inside redaction)« – omogoča izrez območja, ki ga želimo znova razkriti oz. izključiti iz obstoječe anonimizacije.
- »Redact Privacy Info« – namenjena je hitremu iskanju/anonimiziranju delno strukturiranih dokumentov (obrazcev) ter anonimizaciji podatkov, kot so: številke socialnega zavarovanja, telefonske številke, e-poštni naslovi, datum rojstva ter ime in priimek.
- »Find&Redact« – najdemo in anonimiziramo več pojavitev iste besede ali besednih zvez. Predpogoj za uporabo funkcionalnosti je izvedena optična prepoznavna znakov v samem dokumentu. Podprta je uporaba iskanja z iskalnimi znaki, nadomestnimi znaki in t. i. »regularnimi izrazi«. Znotraj tega omogoča funkcionalnost »Find Whole Words Only« iskanje celotnih besed oz. besedilnih nizov. Če po besedilu iščemo besedilni niz v točno taki obliki, kot je napisana, izberemo funkcionalnost »Match Case«.
- »Redact Page« – omogoča hitro anonimizacijo celotne strani dokumenta z enim ukazom.²⁴

Bistvena prednost orodja za anonimizacijo je mdr. tudi ta, da je mogoče v obdelan tekst navajati razloge za anonimizacijo. Razlogi za anonimizacijo se shranjujejo v bazo (shranjeno v datoteki »Reasons.ini«), v vmesnem koraku se lahko anonimizirano besedilo in razloge preverja, v končni različici dokumenta pa so nato vsi razlogi prekritja navedeni. Kode so trajno vidne tako takrat, kadar datoteko odprete na računalniku, kot tudi v katerem koli natisnjem dokumentu. Pri preverjanju anonimiziranih dokumentov pred končno objavo lahko uporabnik preverja posamezne razloge oz. po njih tudi išče po določenih filtrih, kar bistveno olajša preverjanje anonimiziranih dokumentov. Pred končno objavo anonimiziranega dokumenta dokument še ročno pregledamo in ga na koncu objavimo (»Finalize Redactions and Publish As«). Ob tem ima upravitelj tudi možnost ročne označitve besed (s signiranjem), dela besedila ter področja arhivskega gradiva, ki je predmet anonimizacije. Uporabniški vmesnik tudi omogoča prikaz, katere dele arhivskega gradiva je uporabnik označil in zakaj jih je prekril – brez potrebe po podvojenemu odpiranju dokumentov. Orodje omogoča sočasno delo večjega števila uporabnikov na istem dokumentu, na daljavo z možnostjo vnašanja komentarjev in uredniško sledjo.

²⁴ Interni dokument projekta e-ARH.si: Brava uporabniška navodila, 10. 12. 2021.

Izvajalec je, skladno s pogodbenimi določili, izdelal ustrezna navodila za namestitve in uporabo OpenText!Brave, navodila za administratorje za OpenText!Blazon, prav tako so bila izdelana video navodila za uporabo in prezentacije ppt z navodili za uporabo orodja za anonimizacijo.²⁵ Izvedenih je bilo več predstavitev in šolanj za uporabo programske rešitve. Aprila 2020 je bila izvedena predstavitev in presoja orodja glede ustreznosti izdelka. Junija 2020 je bila izvedena testna delavnica za deset udeležencev iz vseh arhivov slovenske javne arhivske službe. Oktobra 2020 pa je bilo na on-line delavnici končnim uporabnikom predstavljeno delovanje orodja za anonimizacijo, kakor tudi njegove osnovne funkcionalnosti. Le-to je zajemalo postopke od uvoza datotek do izvedbe različnih tipov anonimizacije ter prikaza izmenjave datoteke med več anonimizatorji.²⁶

Na šestih individualnih delavnicah je bilo izvedeno izobraževanje »superanonimizatorja«, na katerem je bila poleg pregleda osnovnih funkcionalnosti orodja prikazana še uporaba spletnega orodja za pomoč pri kreiranju regularnih izrazov,²⁷ kreiranje skript ter uporaba orodja na primerih. Opredelilo se je tipične primere rabe orodja za anonimizacijo (samostalniške sklanjatlje, spol in število, glagolski čas, pridevniki z izpeljankami) ter posebnosti anonimizacije za arhivsko gradivo v digitalni obliki (pogoste besede in besedne zveze ter opredelitev tipičnih primerov postopka anonimizacije). Naprednega uporabnika se je prav tako usposobilo za uporabo funkcionalnosti programa, namenjenega uporabi šifrantov pri anonimizaciji gradiva, in seznanilo s postopkom kreiranja skript ter njihovo uporabo pri običajni in samodejni anonimizaciji.²⁸

8. ZAKLJUČEK

Projekt pridobitve orodja za anonimizacijo je šel skozi osnovne faze pridobivanja programske rešitve – od izvedbe poslovne analize, v okviru katere smo analizirali obstoječe rešitve, preko izdelave tehnične specifikacije, v kateri smo opredelili ključne funkcionalnosti, do izvedbe javnega naročila. Izbrani izvajalec je nato izvedel namestitve in izobraževalne delavnice ter uvedel orodje v produkcijo.

S pridobitvijo orodja za anonimizacijo (OpenText!Brava in OpenText!Blazon) smo v največji meri zagotovili zahtevanim funkcionalnostim. Torej, uvoz gradiva posameznega dokumenta ali skupine dokumentov za anonimizacijo, izvedbo avtomatizirane optične prepoznavne znakov, izdelavo in dodajanje iskanih nizov ali besed v bazo besed. Obstaja tudi možnost uvoza prednastavljenih baz besed oziroma kreiranje in urejanje novih ter tudi možnost njihovega izvoza.

Na ta način smo dosegli cilje zastavljene naloge v okviru projekta e-ARH.si in pridobili ustrezno rešitev za izvajanje anonimizacije. Na voljo so orodje, natančna navodila z razlago številnih funkcionalnosti in usposobljen kader za uporabo orodja. Z vsem pridobljenim lahko učinkovito zasledujemo in izpolnjujemo zahteve s področja varstva z zakonom varovanih podatkov, ki jih nalaga 65. člen ZVDAGA.

²⁵ Interni dokumenti projekta e-ARH.si: Brava uporabniška navodila, Blazon uporabniška navodila, predstavitev v .ppt.

²⁶ Interni dokumenti projekta e-ARH.si: e-ARH.si: e-ARH.si-KC4.4.4_Zapisnik o izobraževanju ključnih uporabnikov, 14. 10. 2020.

²⁷ Ang. Regular Expressions – gre za zaporedje znakov, ki definirajo iskalne vzorce.

²⁸ Interni dokumenti projekta e-ARH.si: e-ARH.si_KC4.4.4 Zapisnik izobraževanja superanonimizatorja, 13. 1. 2021.

9. VIRI IN LITERATURA

Ali veste, kaj pomeni »big data« in zakaj so naši podatki tako dragoceni? (2021).

Pridobljeno 8. 12. 2021 s spletne strani: <https://www.mipi.si/teme/informacijska-pismenost/ali-veste-kaj-pomeni-187big-data171-in-zakaj-so-nasi-podatki-tako-dragoceni>.

Arhivska novela tudi leto dni po uveljavitvi deležna nasprotujočih ocen. (2015).

Pridobljeno 8. 12. 2021 s spletne strani: <https://www.iusinfo.si/medijsko-sredisce/dnevne-novice/147493>.

Arhivsko gradivo. (s.d.). Pridobljeno 8. 12. 2021 s spletne stran:

<https://www.gov.si/teme/arhivsko-gradivo/>.

Asmiyanto, T. (2019). Metamorphosis archivist in the Industrial Age 4.0: A challenge in the face of digital revolution. *Record and Library Journal*, št. 5 (1), str. 12-21. D3

Perpustakaan, Fakultas Vokasi, Universitas Airlangga.

Blazon uporabniška navodila. Interni dokument.

Brava uporabniška navodila. Interni dokument.

Cvelfar, B. et. al. (2018). Strategija in izvedbeni načrt razvoja slovenskega elektronskega arhiva za obdobje 2016-2020. Pridobljeno 8. 12. 2021 s spletne strani:

https://www.gov.si/assets/organi-v-sestavi/Arhiv-RS/Projekt-e-ARH-si/PR-material/Strategija_e-ARH_si_2016-2020_1.0.pdf.

Digitalizacija družbe. (s.d.) Pridobljeno 8. 12. 2021 s spletne stran:

<https://www.gov.si/teme/digitalizacija-druzbe/>.

Digitalna pismenost. (s.d.). Pridobljeno 8. 12. 2021 s spletne stran:

<https://www.portalosv.si/digitalna-pismenost/>.

Digitalna Slovenija 2020. (2016). Strategija razvoja informacijske družbe do leta 2020 (2016).

Pridobljeno 8. 12. 2021 s spletne strani:

<https://www.gov.si/assets/ministrstva/MJU/DID/Strategija-razvoja-informacijske-druzbe-2020.pdf>.

e-ARH.si_KC4.4.2_Specifikacija zahtevanih funkcionalnosti programske rešitve za avtomatizirano razpoznavanje in anonimizacijo z zakonom varovanih podatkov. JN-20-004; št. dokumenta 4301-2/2020/2 z dne 30. 1. 2021. Interni dokument.

e-ARH.si_KC4.4.4_Zapisnik izobraževanja superanonimizatorja, 13. 1. 2021. Interni dokument.

e-ARH.si_KC4.4.4_Zapisnik o izobraževanju ključnih uporabnikov, 14. 10. 2020. Interni dokument. e-ARH.si_KC4.4.1_Poročilo orodja za anonimizacijo TACITA, 20. 3. 2017. Interni dokument.

GDPR. (2016). Uredba (EU) 2016/679 Evropskega parlamenta in sveta z dne 27. aprila 2016 o varstvu posameznikov pri obdelavi osebnih podatkov in o prostem pretoku takih podatkov ter o razveljavitvi Direktive 95/46/ES. Pridobljeno 8. 12. 2021 s spletne strani: <https://eur-lex.europa.eu/legal-content/SL/TXT/PDF/?uri=CELEX:32016R0679&from=DA>.

Hajtnik, T. (2012). Učinkovito dolgoročno ohranjanje arhivskega gradiva v Arhivu RS. *Knjižnica*, št. 3, 55-70. Ljubljana: Narodna in univerzitetna knjižnica. .

Hajtnik, T. (2019). Digital Age: Time to Transform Public Archive. *Atlanti*, št. 2, 49-56. Trst: International Institute for Archival Science.

Jedlitschka, K. (2011). Arhiv Stasija – preteklost in sedanjost. V: Zupanič, K. (ur.): Primeri različnih praks v slovenskih arhivih. Zbornik referatov: 25 zborovanje Arhivskega društva Slovenije. Ljubljana: Arhivsko društvo Slovenije.

Jenuš, G. (2013). Varstvo občutljivih osebnih podatkov v arhivskem gradivu nekdanje tajne politične policije – med zakonom in prakso. *Arhivi*, št. 1, str. 57-73. Ljubljana: Arhivsko društvo Slovenije.

- Jenuš, G. (2016).** Digitalizacija in anonimizacija arhivskega gradiva, ki vsebuje z arhivskim zakonom varovane kategorije podatkov. V: *Doksis 2016 by Media.doc. Posvetovanje Sodobno upravljanje e-dokumentov*, št. 25, str. 66-74. Ljubljana: Media.doc.
- Jenuš, G. (2021).** The General data protection regulation : the experience of Slovenian archives with the GDPR and the protection of sensitive information, with emphasis on the documents of totalitarian political systems. V: Cseh, G. B. & Kobel, S. (ur.). The application of the GDPR in archives : with special attention on the documents of totalitarian political systems. Budapest: Historical Archives of the Hungarian State Security. 2021, str. 171-182.
- Jenuš, G. et al (2018).** Analiza ustreznosti formatov za dolgoročno hrambo in metapodatkov pri digitalizaciji arhivske kulturne dediščine. *Moderna arhivistika*, št. 1, 135-156. Maribor: Pokrajinski arhiv Maribor.
- Jerome, J. W. (2013).** Buying and selling privacy. Big data's different burdens and benefits. *Stanford Law Review*, št. 66, str. 47-53. Stanford: Stanford Law School.
- Koncilija, Ž. et al (2018):** Virtualizacija čitalniških postopkov: od ideje do koncepta. *Moderna arhivistika*, št. 1, 95-114.
- Košir, M. (2020):** Pomen Splošne uredbe o varstvu podatkov (GDPR) za arhive. V: Semlič Rajh, Z. (ur.): *Arhivi v službi človeka – Človek v službi arhivov*, št. 5, str. 60-71. Maribor : Alma Mater Europaea – ECM.
- Mi. Li. (2014).** "Referendum je poenotil narod. Vsi se strinjajo, da so zmagali.". MMC RTV SLO. Pridobljeno 8. 12. 2021 s spletne strani: <https://www.rtvlo.si/slovenija/referendum-je-poenotil-narod-vsi-se-strinjajo-da-so-zmagali/338926>.
- Midrack, R. L. (s.d.).** Kaj je optično prepoznavanje znakov (OCR)? Pridobljeno 8. 12. 2021 s spletne stran: <https://sl.eyewated.com/kaj-je-opticno-prepoznavanje-znakov-ocr/>.
- Odločba Ustavnega sodišča v postopku ocene ustavnosti Zakona o varstvu dokumentarnega in arhivskega gradiva ter arhivih št. U-I-70/12-14 z dne 21. 3. 2014.** Pridobljeno 8. 12. 2021 s spletne strani: <https://www.us-rs.si/wp-content/uploads/2020/02/u-i-70-12.pdf>.
- OpenText Brava! (s.d.).** Pridobljeno 8. 12. 2021 s spletne stran: <https://www.opentext.com/products-and-solutions/products/enterprise-content-management/opentext-brava>.
- Predlog zakona o varstvu osebnih podatkov (ZVOP-2). (2021).** Pridobljeno 8. 12. 2021 s spletne stran: <https://www.iusinfo.si/medijsko-sredisce/v-srediscu/284934>.
- Prelesnik, Š. (2018):** "GDPR je tu. Poznate vsa pravila?«. *Finance*, št. 96: str. 4–5, 8. Ljubljana: Časnik Finance, časopisno založništvo d.o.o.
- Programsko orodje za avtomatizirano razpoznavanje in zakritje z zakonom varovanih podatkov (anonimizacijo). (s.d.).** Pridobljeno 8. 12. 2021 s spletne stran: <https://nio.gov.si/nio/asset/programsko+orodje+za+avtomatizirano+razpoznavanje+in+za+kritje+z+zakonom+varovanih+podatkov>.
- Romansky, R. P. & Noninska, I. S: (2020):** Challenges of the digital age for privacy and personal data protection. *Mathematical Biosciences and Engineering*, št. 17 (5), str. 5288-5303. Tempe: School of Mathematical and Statistical Sciences, Arizona State University.
- SI AS 1931 Republiški sekretariat za notranje zadeve Socialistične republike Slovenije, 1918-2004 (Fond/zbirka).** Pridobljeno 8. 12. 2021 s spletne stran: <http://arsq.gov.si/Query/detail.aspx?ID=27578>.
- Smernice EAG. (2018).** Smernice Evropske arhivske skupine o izvajanju Splošne uredbe EU o varstvu podatkov v arhivskem sektorju. Pridobljeno 8. 12. 2021 s spletne strani: https://ec.europa.eu/info/sites/default/files/eag_draft_guidelines_1_11_0.pdf.

- STA / N.D./M.R./U.Z. (2014).** »Zaradi izjemno nizke volilne udeležbe referendum o arhivih ni uspel.«. 24ur.com. Pridobljeno 8. 12. 2021 s spletne stran: <https://www.24ur.com/novice/slovenija/volivci-odlocajo-o-usodi-arhivske-novele.html>.
- T. K. B., G. C., A. Č. (2014).** »Na volišča 4. maja. SDS napoveduje ustavno presojo referendumskega odloka.«. MMC RTV SLO. Pridobljeno 8. 12. 2021 s spletne stran: <https://www.rtvlo.si/slovenija/na-volisca-4-maja-sds-napoveduje-ustavno-presajo-referendumskega-odloka/333524>.
- Tiskovno sporočilo v zvezi z odločbo Ustavnega sodišča št. U-I-70/12-14. (2014).** Pridobljeno 8. 12. 2021 s spletne stran: <https://www.us-rs.si/wp-content/uploads/2020/02/u-i-70-12.-.tiskovno.sporocilo.pdf>.
- ZAGOPP. (2016).** Zakon o arhivskem gradivu, ki vsebuje osebne podatke o zdravljenju pacienta. Uradni list RS, št. 85/16. Pridobljeno 8. 12. 2021 s spletne strani: <http://www.pisrs.si/Pis.web/pregledPredpisa?id=ZAKO7541>.
- Zapisnik Arhivskega sveta št. 625-4/2018-40. (2019).** Pridobljeno 8. 12. 2021 s spletne strani: <https://www.gov.si/zbirke/delovna-telesa/arhivski-svet/>.
- ZVDAGA. (2006).** Zakon o varstvu dokumentarnega in arhivskega gradiva ter arhivih. Uradni list RS, št. 30/2006. Pridobljeno 8. 12. 2021 s spletne strani: <http://www.pisrs.si/Pis.web/pregledPredpisa?id=ZAKO4284>.
- ZVDAGA-A. (2014).** Zakon o spremembah in dopolnitvah Zakona o varstvu dokumentarnega in arhivskega gradiva ter arhivih. Uradni list RS, 51/2014. Pridobljeno 8. 12. 2021 s spletne strani: <http://www.pisrs.si/Pis.web/pregledPredpisa?id=ZAKO6681>

SUMMARY

AUTOMATED PROCESING OF PERSONAL DATA PROTECTED BY ARCHIVAL LAW - ANONYMISATION

Gregor Jenuš, Ph. D.

Archives of the Republic of Slovenia,
Ljubljana, Slovenija
gregor.jenus@gov.si

Žiga Koncilija, Ph. D.

Archives of the Republic of Slovenia,
Ljubljana, Slovenija
ziga.koncilija@gov.si

Tatjana Hajtnik, Ph. D.

Archives of the Republic of Slovenia, Ljubljana, Slovenija
tatjana.hajtnik@gov.si

In 2016 the European Parliament and the European Council implemented the General Data Protection Regulation or the infamous GDPR in to the European legislation. Although criticized by many, the regulators actually achieved one major goal; the GDPR raised the awareness of the general public on the value and importance of their personal data. The regulation was a direct response to the era of technological hyper-development and globalization in which it is becoming increasingly important to know what happens with our personal data. The scale of unwanted sharing of it has increased significantly in recent years. Technology allows us to share personal data without much effort and to realize that we do it at all. However, there is a catch to this problem. The problem is not just in the technology and its possibilities; the problem are natural persons. It is us! Natural persons make personal information or data available to third parties publicly and on a global scale, because they do not think about the possible harmful consequences of such doing or simply do not take the time to read though different checkboxes. The consequence is unwanted processing of personal data, which third parties can use to their advantage and exploit it through the so-called "Big Data". Although processing of personal data should be designed to serve humanity, the truth is that it often serves the (financial) interests of third parties. The GDPR therefore marks a milestone in data protection or at least that is the way the European regulation is presented to the public.

However, the desire to protect personal data is nothing new. At least in case of Slovenian archival institutions managing highly sensitive or personal data in archival records is part of our daily tasks. Archival records are a legacy of our past, which also contain such data that the legislator has determined that it is not accessible to everyone. Within the Slovenian public archival network, public archives are therefore obliged to anonymize or remove (blacken out) personal data that is protected by archival law. The anonymization of archival records can be performed in two ways; physical on copies of archival records and digital on born digital or digitized archival records. The digitalization and virtualization of archival services in recent years demanded that the Slovenian public archives acquired a dedicated software tool for anonymization. This task was successfully executed within the project of the establishing of the Slovenian public electronic archive.