

D2 – Metóda pre anonymizáciu digitálnych stôp



Projekt Automatizácia digitálnej forenznej analýzy a odpovede na incident (ADFIR) financovaný Európskou úniou – Next GenerationEU prostredníctvom Plánu obnovy a odolnosti Slovenskej republiky pod číslom projektu č. 09I05-03-V02-00079.

OBSAH

1	Popis projektu	3
2	Úvod	3
3	Úloha KPB1.2.....	4
3.1	Inventarizácia a klasifikácia údajov	5
3.1.1	Klasifikácia údajov ako osobné alebo neosobné údaje	5
3.1.2	Iné druhy chránených informácií.....	8
3.2	Súlad s právnymi požiadavkami na anonymizáciu osobných údajov.....	9
3.3	Výber anonymizačných metód	11
3.3.1	Randomizácia (randomization)	11
3.3.2	Zovšeobecnenie (generalization).....	12
3.4	Kontrola výsledku	14
3.5	Využitie anonymizovaných digitálnych stop pre budúcu analýzu.....	14
4	Prehľad	16

1 Popis projektu

Projekt **Automatizácia digitálnej forenznej analýzy a odpovede na incident** (ďalej len „ADFIR“) je financovaný **Európskou úniou – Next GenerationEU prostredníctvom Plánu obnovy a odolnosti Slovenskej republiky** pod číslom projektu č. 09-I05-03-V02-00079. Tento projekt sa zaoberá jednou z kľúčových výziev v oblasti kybernetickej bezpečnosti a informačnej bezpečnosti – ako spracovať obrovské množstvo digitálnych dôkazov, ktoré vznikajú počas incidentov kybernetickej bezpečnosti alebo forenznych vyšetrovaní. V súčasnosti je tento proces veľmi náročný z hľadiska ľudských zdrojov a času. Automatizácia pomocou metód strojového učenia môže preto výrazne **zlepšiť kvalitu digitálnej forenznej analýzy** a skrátiť čas potrebný na jej vykonanie. Celkovo to umožňuje bezpečnostným tímom efektívnejšie reagovať na kybernetické hrozby. Hlavné prínosy tohto projektu sú:

- **Rýchlejšie riešenie incidentov v oblasti kybernetickej bezpečnosti.** Projekt ADFIR zavádza automatizované prístupy k zberu, spracovaniu a analýze digitálnych stôp. Vďaka tomu môžu bezpečnostné tímy rýchlejšie identifikovať príčiny incidentov a prijať účinné opatrenia na ich riešenie.
- **Zníženie pracovnej záťaže forenznych analytikov.** Rutinné a časovo náročné úlohy spojené so spracovaním digitálnych stôp budú nahradené automatizovanými metódami. To umožní analytikom sústrediť sa na zložitejšie prípady a strategické rozhodovanie.
- **Vyššia kvalita a konzistentnosť výstupov.** Použitie jednotných metodík a nástrojov zaručuje, že spracované digitálne stopy budú presnejšie, konzistentnejšie a ľahšie overiteľné. To výrazne znižuje riziko chýb spôsobených ľudskými faktormi.
- **Možné využitie v trestnom konaní.** Výstupy projektu budú vyvinuté v súlade s právnymi požiadavkami a normami, čo umožní, aby digitálne stopy boli akceptované ako relevantné dôkazy pre vyšetrovanie a súdne konania.

2 Úvod

Cieľom pracovného balíka KPB1 je analyzovať a **reagovať na právne otázky súvisiace s navrhovaným projektom**. Tento pracovný balík definuje minimálne štandardy a právne záruky potrebné na zabezpečenie uplatňovania právneho rámca ochrany osobných údajov pri spracovaní digitálnych dôkazov a prípustnosti získaných výstupov automatizovanej forenznej analýzy ako dôkazov v trestnom konaní, pričom zohľadňuje aspekty etické, právne, ako aj otázky legitímnosti a proporcionality týkajúce sa extrakcie digitálnych dôkazov strojovým učením. Predmetom skúmania sú aj vlastnosti digitálnych dôkazov (integrita, autenticita, prípustnosť a relevantnosť).

Keďže projekt implementuje hybridný prístup kombinujúci dáta zo simulovaných útokov v kontrolovanom prostredí s anonymizovanými artefaktmi z reálnej praxe vo fáze implementácie, je nevyhnutné zabezpečiť zodpovedajúcu právnu analýzu anonymizácie dát.

Výstupom je návrh metódy anonymizácie digitálnych stôp, ktorý identifikuje potrebné kroky na zabezpečenie súladu s právnymi predpismi. Cieľom je poskytnúť súbor nástrojov, ktorý umožňuje rozlišovať medzi osobnými a neosobnými údajmi a ďalej, ak sú osobné údaje identifikované, zabezpečiť ich spracovanie a anonymizáciu v súlade s platnými predpismi.

3 Úloha KPB1.2

Úloha KPB1.2 – Analýza podmienok pre anonymizáciu digitálnych stôp a zahrnutie údajov týkajúcich sa bezpečnostných incidentov do datasetu. Zohľadnenie právnych požiadaviek na vytváranie datasetu a podmienok pre zahrnutie údajov do datasetu (UPJS). V rámci tejto úlohy skúmame platné predpisy o ochrane osobných údajov a ich uplatnenie v súvislosti s preskúvaním a spracovaním digitálnych stôp v rámci projektu.

Digitálne stopy môžu zahŕňať širokú škálu rôznych informácií s dôkaznou hodnotou uložených alebo prenášaných v digitálnej forme.¹ Konkrétne informácie obsiahnuté v digitálnych stopách môžu byť chránené právnou ochranou podľa rôznych právnych predpisov. Príkladom takýchto informácií sú napríklad údaje, ktoré možno klasifikovať ako osobné údaje podľa aplikovateľnej právnej úpravy, ktorá je obsiahnutá predovšetkým v Nariadení Európskeho parlamentu a Rady (EÚ) 2016/679 z 27. apríla 2016 o ochrane fyzických osôb pri spracúvaní osobných údajov a o voľnom pohybe takýchto údajov, ktorým sa zrušuje smernica 95/46/ES (všeobecné nariadenie o ochrane údajov) (ďalej len „GDPR“), ktoré má všeobecnú pôsobnosť, je záväzné v celom rozsahu a priamo uplatniteľné vo všetkých členských štátoch. V národnom kontexte možno zodpovedajúcu úpravu nájsť v zákone č. 18/2018 Z. z. o ochrane osobných údajov v znení neskorších predpisov.

Ak sa majú digitálne stopy skúmať a spracovávať na iné účely, ako je účel trestného konania, na ktorý ich pôvodne získali príslušné orgány alebo iné subjekty², musia sa riadiť platnými predpismi. V nasledujúcich kapitolách načrtávame potrebné kroky, ktoré je potrebné podniknúť, aby sa umožnilo ďalšie používanie digitálnych dôkazov prostredníctvom ich anonymizácie. V súhrne ide o:

¹ Scientific Working Group on Digital Evidence (SWGDE). *Digital Evidence: Standards and Principles*. In *Forensic Science Communications*. 2000, No. 2. Available: <https://archives.fbi.gov/archives/about-us/lab/forensic-sciencecommunications/fsc/april2000/swgde.htm#Introduction>

² Ako konštatuje Dražová, „V rámci svojej pôsobnosti sa so zaistovaním dát ako potenciálnych elektronických dôkazov bežne potykajú aj rôzne subjekty mimo trestného konania, napríklad národné alebo vládne jednotky pre riešenie kybernetických bezpečnostných incidentov, orgány verejnej moci na účely civilného či správneho konania, tímy vykonávajúce digitálne vyšetrovanie pôsobiace v komerčnej sfére a podobne.“ Dražová, P. *Electronic evidence in criminal proceedings*. Dissertation thesis, 2021. Available: <https://opac.crzp.sk/?fn=detailBiblioFormChildA52T&sid=4BD9478A3A7038F60D223DAC8853&seo=CRZP-detail-kniha>

1. inventarizáciu údajov a klasifikáciu údajov ako osobných alebo neosobných údajov,
2. zabezpečenie súladu so zákonnými podmienkami pre anonymizáciu osobných údajov,
3. výber vhodných metód anonymizácie údajov,
4. implementáciu vybraných metód anonymizácie údajov,
5. kontrolu výstupu.

3.1 Inventarizácia a klasifikácia údajov

Prvým nevyhnutným krokom je **vytvorenie inventára údajov**. V praxi to zahŕňa určenie všetkých spracovávaných informácií a vytvorenie zoznamu takýchto informácií, napr. vo forme tabuľky.

Nasledujúcim krokom je preskúmanie týchto informácií s cieľom určiť ich právnu povahu, konkrétne či sa na tieto informácie vzťahuje špecifická právna ochrana. Vytvorená tabuľka informácií by mala obsahovať dosiahnuté závery a určené kategórie informácií. Vo väčšine prípadov budú kategórie informácií zahŕňať rozlíšenie medzi osobnými a neosobnými údajmi, ale možno použiť aj iné kategórie (napr. citlivé údaje, osobné identifikačné informácie (tzv. personally identifiable information), dôverné informácie, obchodné tajomstvo) v závislosti od aplikovateľných predpisov.

Tabuľka 1 Príklad inventarizácie a klasifikácie údajov

INFORMÁCIA	KLASIFIKÁCIA INFORMÁCIE
meno	osobný údaj
emailová adresa	osobný údaj / dôverná informácia
názov súboru	osobný údaj
IP adresa	osobný údaj
časová pečiatka	osobný údaj
zoznam zákazníkov	obchodné tajomstvo

3.1.1 Klasifikácia údajov ako osobné alebo neosobné údaje

Podľa článku 4 ods. 1 GDPR, **osobné údaje** sú „akékoľvek informácie týkajúce sa identifikovanej alebo identifikovateľnej fyzickej osoby (ďalej len „dotknutá osoba“); identifikovateľná fyzická osoba je osoba, ktorú možno identifikovať priamo alebo nepriamo, najmä odkazom na identifikátor, ako je meno, identifikačné číslo, lokalizačné údaje, online identifikátor, alebo odkazom na jeden či viaceré prvky, ktoré sú špecifické pre fyzickú, fyziologickú, genetickú, mentálnu, ekonomickú, kultúrnu alebo sociálnu identitu tejto fyzickej osoby.“

Článok 9 GDPR rozlišuje **osobitné kategórie osobných údajov**, ktoré zahŕňajú údaje, ktoré odhaľujú rasový alebo etnický pôvod, politické názory, náboženské alebo filozofické

presvedčenie alebo členstvo v odborových organizáciách, ako aj genetické údaje, biometrické údaje na individuálnu identifikáciu fyzickej osoby, údaje týkajúce sa zdravia alebo údaje týkajúce sa sexuálneho života alebo sexuálnej orientácie fyzickej osoby. Spracúvanie osobitných kategórií osobných údajov je, v princípe, **zakázané**, ak nie je daný niektorý z právnych základov spracúvania vymedzený v článku 9 ods. 2 GDPR.

Klasifikácia informácií ako osobných údajov je kľúčová pre určenie uplatniteľnosti právnej úpravy ochrany osobných údajov vo vzťahu k predmetným informáciám. Ak konkrétne informácie nemožno klasifikovať ako osobné údaje, ich spracovanie nebude podliehať podmienkam a obmedzeniam definovaným v príslušnej právnej úprave ochrany osobných údajov.

Pri určovaní, či možno informácie klasifikovať ako osobné údaje, **sa musia zvážiť všetky prostriedky**, ktoré prevádzkovateľ (fyzická alebo právnická osoba, orgán verejnej moci, agentúra alebo iný subjekt, ktorý sám alebo spoločne s inými určí účely a prostriedky spracúvania osobných údajov) alebo iná osoba s primeranou pravdepodobnosťou použije, aby sa dospelo k definitívnemu záveru o povahe daných informácií. Údaje môžu byť identifikovateľné, napr. ak obsahujú informácie umožňujúce priamu identifikáciu (meno, audio alebo zvukové nahrávky, e-mailová adresa, údaje o polohe atď., tzn. ak umožňujú identifikáciu konkrétnej fyzickej osoby (v prípade prítomnosti jedinečných údajových bodov alebo vzorcov správania, ktoré sa môžu vzťahovať iba na jednu osobu)), ak je možné odvodiť informácie o jednotlivcovi na základe vytvoreného súboru údajov alebo prepojiť záznamy, ktoré sa týkajú jednotlivca (kombináciou atribútov v súbore údajov alebo prepojením viacerých súborov údajov), alebo ak je možná deidentifikácia (pseudonymizácia). Vo všeobecnosti, ak sa informácie zhromažďujú od fyzických osôb, možno predpokladať, že budú predstavovať osobné údaje.

V kontexte právneho rámca ochrany osobných údajov sa informácie, ktoré nespádajú pod širokú definíciu osobných údajov, považujú za **neosobné údaje**, ktoré zahŕňajú:

- a) údaje, ktoré sa pôvodne netýkali identifikovanej alebo identifikovateľnej fyzickej osoby, ako napríklad údaje o poveternostných podmienkach generované senzormi nainštalovanými na veterných turbínach alebo údaje o potrebách údržby priemyselných strojov, a
- b) údaje, ktoré boli pôvodne osobnými údajmi, ale neskôr boli anonymizované (anonymné údaje). Riadne anonymizované údaje nemožno priradiť ku konkrétnej osobe, a to ani s použitím dodatočných údajov, a preto nie sú osobnými údajmi.³

Podľa Recitálu 26 GDPR, „*Zásady ochrany údajov by sa preto nemali uplatňovať na anonymné informácie, konkrétne na informácie, ktoré sa nevzťahujú na identifikovanú alebo*

³ Communication from the Commission to the European parliament and the Council. Guidance on the Regulation on a framework for the free flow of non-personal data in the European Union. COM(2019) 250 final.

identifikovateľnú fyzickú osobu, ani na osobné údaje, ktoré sa stali anonymnými takým spôsobom, že dotknutá osoba nie je alebo už nie je identifikovateľná. Toto nariadenie sa preto netýka spracúvania takýchto anonymných informácií **vrátane spracúvania na štatistické účely alebo účely výskumu.**⁴ Toto ustanovenie poskytuje **konceptnú definíciu anonymizácie.** Vyžaduje, aby boli údaje, ktoré sú predmetom anonymizácie, **zbavené dostatočného množstva prvkov vylučujúcich identifikáciu dotknutej osoby.** Dôležitým faktorom je, že spracovanie musí byť **nezvratné.** GDPR neobjasňuje, ako by sa takýto proces deidentifikácie mal alebo mohol vykonať, ale **zameriava sa na výsledok, ktorý sa musí dosiahnuť, a to nemožnosť opätovnej identifikácie dotknutej osoby.**

Anonymizácia je definovaná aj v medzinárodných normách, ako napríklad v norme ISO 29100 ako „proces, pri ktorom sa osobné identifikačné údaje (tzv. personally identifiable information alebo PII) nezvratne zmenia takým spôsobom, že poskytovateľa PII už nemožno priamo ani nepriamo identifikovať, či už samotným prevádzkovateľom PII alebo v spolupráci s akoukoľvek inou stranou“ (ISO 29100:2011). Kľúčovou je aj pre ISO nezvratnosť zmeny, ktorou osobné údaje prešli, aby sa umožnila priama alebo nepriama identifikácia.

Posúdenie, či sú údaje riadne anonymizované, a teda či sa na ne v skutočnosti nevzťahuje právna úprava o ochrane osobných údajov, **závisí od špecifických a jedinečných okolností každého jednotlivého prípadu.** Niekoľko príkladov opätovnej identifikácie súborov údajov, ktoré boli pôvodne anonymizované, ukázalo, že takéto hodnotenie môže byť náročné.⁵ Vytvorenie skutočne anonymného súboru údajov z bohatej sady osobných údajov pri zachovaní čo najväčšieho množstva podkladových informácií potrebných na danú úlohu nie je jednoduchá úloha. Napríklad súbor údajov považovaný za anonymný možno kombinovať s iným súborom údajov takým spôsobom, že je možné identifikovať jednu alebo viaceré fyzické osoby.⁶ Súčasťou procesu anonymizácie je vždy **riziko zlyhania anonymizácie;** mala by sa preto zvážiť závažnosť tohto rizika a pravdepodobnosť jeho výskytu.

⁴ V súčasnosti GDPR neobsahuje právnu definíciu pojmu „vedecký výskum“. Právne nezáväznú definíciu možno nájsť v Recitáli 159 GDPR, podľa ktorého „Na účely tohto nariadenia by sa spracúvanie osobných údajov na účely vedeckého výskumu malo vykladať široko, aby zahŕňalo napríklad technický rozvoj a demonštračné činnosti, základný výskum, aplikovaný výskum a výskum financovaný zo súkromných zdrojov.“ Podľa dôvodovej správy k Protokolu, ktorým sa mení Dohovor o ochrane jednotlivcov pri automatizovanom spracovaní osobných údajov, sa „**štatistické účely**“ vzťahujú na vypracovanie štatistických prieskumov alebo na tvorbu štatistických, agregovaných výsledkov; štatistika sa zameriava na analýzu a charakterizáciu hromadných alebo kolektívnych javov v danej populácii. Okrem toho, spracovanie údajov na „**účely vedeckého výskumu**“ má za cieľ poskytnúť výskumníkom informácie prispievajúce k pochopeniu javov v rôznych vedeckých oblastiach (epidemiológia, psychológia, ekonómia, sociológia, lingvistika, politológia, kriminológia atď.) s cieľom stanoviť trvalé princípy, zákony správania alebo vzorce kauzality, ktoré presahujú všetkých jednotlivcov, na ktorých sa vzťahujú.

⁵ Pozri The study on future data flows conducted for the European Parliament's ITRE Committee by Blackman, C., Forge, S.: Data Flows — Future Scenarios: In-Depth Analysis for the ITRE Committee, 2017, p. 22, Box 2. Available online

⁶ ARTICLE 29 DATA PROTECTION WORKING PARTY. Opinion 05/2014 on Anonymisation Techniques. Adopted on 10 April 2014. 0829/14/EN WP216.

Na určenie toho, či je fyzická osoba identifikovateľná, by sa mali brať do úvahy všetky prostriedky, pri ktorých existuje primeraná pravdepodobnosť, že ich prevádzkovateľ alebo akákoľvek iná osoba využije, napríklad osobitným výberom, na priamu alebo nepriamu identifikáciu fyzickej osoby. Ak možno neosobné údaje akýmkoľvek spôsobom spojiť s jednotlivcom, na základe čoho ho možno priamo alebo nepriamo identifikovať, je potrebné tieto údaje klasifikovať ako osobné údaje.

V tejto súvislosti je dôležité zreflektovať aj pojem pseudonymizácia. Podľa článku 4 ods. 5 GDPR sa pseudonymizáciou rozumie „*spracúvanie osobných údajov takým spôsobom, aby osobné údaje už nebolo možné priradiť konkrétnej dotknutej osobe bez použitia dodatočných informácií, pokiaľ sa takéto dodatočné informácie uchovávajú oddelene a vzťahujú sa na ne technické a organizačné opatrenia s cieľom zabezpečiť, aby osobné údaje neboli priradené identifikovanej alebo identifikovateľnej fyzickej osobe*“. Pseudonymizované údaje, ktoré by bolo možné priradiť k fyzickej osobe s použitím dodatočných informácií, sa považujú za informácie o identifikovateľnej fyzickej osobe, a preto ide o osobné údaje.⁷ Zjednodušene povedané, pseudonymizované údaje sú stále osobnými údajmi a vzťahuje sa na ne príslušná právna úprava ochrany osobných údajov.

3.1.2 Iné druhy chránených informácií

Osobné údaje nie sú jediným druhom informácií, ktoré sú chránené platnými právnymi predpismi. Osobitná ochrana sa poskytuje aj:

1. **dôverným informáciám**, ktoré vo všeobecnosti zahŕňajú akékoľvek informácie určené za dôverné príslušnými subjektmi, predovšetkým zmluvnými stranami v rámci ich zmluvných vzťahov;
2. **obchodným tajomstvám**, ktoré chránia informácie, ktoré spĺňajú všetky nasledujúce požiadavky:
 - a) sú tajné v tom zmysle, že nie sú ako celok alebo pri presnom usporiadaní a spojení ich častí všeobecne známe alebo bežne prístupné osobám v kruhoch, ktoré sa dotknutým druhom informácií bežne zaoberajú;
 - b) majú obchodnú hodnotu, pretože sú tajné;
 - c) osoba, ktorá s týmito informáciami oprávnene nakladá, podnikla za daných okolností primerané kroky, aby ich udržala v tajnosti.⁸

Informácie môžu byť tiež chránené v rámci právnej úpravy duševného vlastníctva, napr. neregistrované ochranné známky. Uplatniteľnosť takejto špecifickej právnej ochrany sa musí posudzovať individuálne a zohľadniť v procese inventarizácie a klasifikácie údajov.

⁷ European Data Protection Board. Guidelines 01/2025 on Pseudonymisation. Adopted on 16 January 2025. P. 3.

⁸ Čl. 2 ods. 1 Smernice Európskeho parlamentu a Rady (EÚ) 2016/943 z 8. júna 2016 o ochrane nesprístupneného know-how a obchodných informácií (obchodného tajomstva). Ú. v. EÚ L 157, 15.6.2016, s. 1 – 18.

3.2 Súlad s právnymi požiadavkami na anonymizáciu osobných údajov

V prípade určenia potreby anonymizácie skúmaných údajov, nasledujúcim nevyhnutným krokom je zabezpečenie, že proces anonymizácie je uskutočnený v súlade s aplikovateľnou právnou úpravou ochrany osobných údajov.

Anonymizácia osobných údajov predstavuje formu spracúvania osobných údajov. Podľa článku 4 ods. 2 GDPR sa spracúvaním rozumie „*operácia alebo súbor operácií s osobnými údajmi alebo súbormi osobných údajov, napríklad získavanie, zaznamenávanie, usporadúvanie, štruktúrovanie, uchovávanie, prepracúvanie alebo zmena, vyhľadávanie, prehliadanie, využívanie, poskytovanie prenosom, šírením alebo poskytovaním iným spôsobom, preskupovanie alebo kombinovanie, obmedzenie, vymazanie alebo likvidácia, bez ohľadu na to, či sa vykonávajú automatizovanými alebo neautomatizovanými prostriedkami*“. Akékoľvek spracúvanie osobných údajov musí byť vykonávané **v súlade so zásadami spracúvania osobných údajov** vymedzených v čl. 5 GDPR.

Jednou z najdôležitejších zásad spracúvania osobných údajov je **zákonnosť spracúvania**. Osobné údaje sú spracúvané zákonne iba v prípade a v rozsahu, v akom je daný niektorý z právnych základov spracúvania vymedzený v článku 6 ods. 1 GDPR. Toto ustanovenie konkrétne vymedzuje 6 právnych základov spracúvania, ktoré zahŕňajú:

- a) dotknutá osoba vyjadrila **súhlas so spracúvaním** svojich osobných údajov na jeden alebo viaceré konkrétne účely,
- b) spracúvanie je nevyhnutné na **plnenie zmluvy**, ktorej zmluvnou stranou je dotknutá osoba, alebo aby sa na základe žiadosti dotknutej osoby vykonali opatrenia pred uzatvorením zmluvy,
- c) spracúvanie je nevyhnutné na **splnenie zákonnej povinnosti** prevádzkovateľa,
- d) spracúvanie je nevyhnutné, aby sa **ochránili životne dôležité záujmy** dotknutej osoby alebo inej fyzickej osoby,
- e) spracúvanie je nevyhnutné na **splnenie úlohy reaizovanej vo verejnom záujme alebo pri výkone verejnej moci zverenej prevádzkovateľovi**,
- f) spracúvanie je nevyhnutné **na účely oprávnených záujmov**, ktoré sleduje prevádzkovateľ alebo tretia strana, s výnimkou prípadov, keď nad takýmito záujmami prevažujú záujmy alebo základné práva a slobody dotknutej osoby, ktoré si vyžadujú ochranu osobných údajov, najmä ak je dotknutou osobu dieťa.

Spracúvanie osobných údajov musí byť **spravodlivé a transparentné**. Dotknuté osoby by si mali byť vedomé spracovania svojich osobných údajov a rozsahu takéhoto spracovania. Dotknuté osoby by mali byť informované o rizikách, pravidlách, zárukách a právach súvisiacich so spracovaním osobných údajov a o tom, ako si môžu uplatniť svoje práva v súvislosti s takýmto spracovaním.⁹

⁹ Recitál 39 GDPR.

Zásada obmedzenia účelu vyžaduje, aby osobné údaje boli získavané na konkrétne určené, výslovne uvedené a legitímne účely, pričom nesmú byť ďalej spracúvané spôsobom, ktorý s týmito účelmi nie je zlučiteľný; ďalšie spracúvanie na účely archivácie vo verejnom záujme, **na účely vedeckého alebo historického výskumu či štatistické účely** sa v súlade s článkom 89 ods. 1 GDPR nepovažuje za nezlučiteľné s pôvodnými účelmi. Konkrétne účely, na ktoré sa osobné údaje spracúvajú, by mali byť výslovne uvedené a legitímne a stanovené v čase získavania osobných údajov.¹⁰ Osobné údaje by sa mali spracúvať len vtedy, ak účel spracúvania nebolo možné za primeraných podmienok dosiahnuť inými prostriedkami.¹¹

Zásada minimalizácie údajov vyžaduje, aby osobné údaje boli primerané, relevantné a obmedzené na rozsah, ktorý je nevyhnutný vzhľadom na účely, na ktoré sa spracúvajú

Zásada správnosti vyžaduje správnosť a aktualizáciu osobných údajov v prípade potreby; v prípade ak sú osobné údaje nesprávne z hľadiska účelov, na ktoré sa spracúvajú, je potrebné prijať všetky potrebné opatrenia na to, aby sa dosiahol bezodkladný výmaz alebo oprava takýchto údajov.

Zásada minimalizácie uchovávaní vyžaduje, aby osobné údaje boli uchovávané vo forme, ktorá umožňuje identifikáciu dotknutých osôb najviac dovtedy, kým je to potrebné na účely, na ktoré sa osobné údaje spracúvajú. Osobné údaje sa môžu uchovávať dlhšie, pokiaľ sa budú spracúvať výlučne na účely archivácie vo verejnom záujme, na účely vedeckého alebo historického výskumu či na štatistické účely v súlade s článkom 89 ods. 1 GDPR za predpokladu prijatia primeraných technických a organizačných opatrení vyžadovaných týmto nariadením na ochranu práv a slobôd dotknutých osôb. Súlad s touto zásadou predpokladá obmedzenie obdobia, počas ktorého sa osobné údaje uchováujú, na nevyhnutný rozsah.¹²

Zásada integrity a dôvernosti ustanovuje, že osobné údaje musia byť spracúvané spôsobom, ktorý zaručuje primeranú bezpečnosť osobných údajov, vrátane ochrany pred neoprávneným alebo nezákonným spracúvaním a náhodnou stratou, zničením alebo poškodením, a to prostredníctvom primeraných technických alebo organizačných opatrení.

Zodpovedná osoba, ktorá musí zabezpečiť súlad s týmito zásadami, je prevádzkovateľ. Prevádzkovateľ musí vedieť tento súlad preukázať (**zásada zodpovednosti**).

¹⁰ Ibidem.

¹¹ Ibidem.

¹² Recitál 39 GDPR.

3.3 Výber anonymizačných metód

Ďalším krokom je výber metód anonymizácie, ktoré budú implementované. Anonymizácia údajov by sa mala vždy vykonávať s použitím **najnovších anonymizačných techník**, ktoré odrážajú vývoj v oblasti technológií a analýzy údajov, a ktoré môžu umožniť konverziu anonymizovaných údajov na osobné údaje. V tejto kapitole popisujeme rôzne metódy anonymizácie osobných údajov relevantné v skúmanom kontexte anonymizácie digitálnych stôp.

Osobné údaje je možné anonymizovať pomocou rôznych anonymizačných metód alebo techník. Prevádzkovateľ musí zvážiť hlavné riziká spojené s procesom anonymizácie, a to:

1. **vyčlenenie (singling-out)** – možnosť izolovať niektoré alebo všetky záznamy, ktoré identifikujú jednotlivca v súbore údajov;
2. **prepojitelnosť (linkability)** – schopnosť prepojiť aspoň dva záznamy týkajúce sa tej istej dotknutej osoby alebo skupiny dotknutých osôb (buď v tej istej databáze, alebo v dvoch rôznych databázach);
3. **inferencia (inference)** – možnosť odvodiť s vysokou pravdepodobnosťou hodnotu atribútu z hodnôt súboru iných atribútov.¹³

3.3.1 Randomizácia (randomization)

Randomizácia predstavuje skupinu techník, ktoré **menia pravdivosť údajov** s cieľom odstrániť silné prepojenie medzi údajmi a jednotlivcom. Ak sú údaje dostatočne neisté, nemožno ich už priradiť ku konkrétnej osobe¹⁴, a preto ich nemožno klasifikovať ako osobné údaje, ktorých spracovanie podlieha platným predpisom o ochrane osobných údajov.

Samotná randomizácia nezníži jedinečnosť každého záznamu, pretože každý záznam bude stále odvodený od jednej dotknutej osoby, ale môže chrániť pred inferenčnými útokmi/rizikami a možno ju kombinovať s technikami zovšeobecnenia, aby sa poskytli silnejšie záruky ochrany súkromia. Na zabezpečenie toho, aby záznam nemohol identifikovať jednu osobu, môžu byť potrebné ďalšie techniky.¹⁵

3.3.1.1 Pridanie šumu (noise addition)

Pridanie šumu spočíva v úprave atribútov v súbore údajov tak, aby boli menej presné, pričom sa zachová celkové rozloženie. Táto metóda mení atribúty, zavádza do súboru údajov náhodnosť, zabezpečuje, že skúmané údaje nemožno prepojiť s jednotlivcami, a zároveň

¹³ ARTICLE 29 DATA PROTECTION WORKING PARTY. Opinion 05/2014 on Anonymisation Techniques. Adopted on 10 April 2014. 0829/14/EN WP216.

¹⁴ ARTICLE 29 DATA PROTECTION WORKING PARTY. Opinion 05/2014 on Anonymisation Techniques. Adopted on 10 April 2014. 0829/14/EN WP216.

¹⁵ Ibidem

zachováva agregované vzorce relevantné pre budúci výskum. Túto metódu bude často potrebné kombinovať s inými technikami anonymizácie, ako je odstránenie zjavných atribútov a kvázi identifikátorov. Úroveň šumu by mala závisieť od nevyhnutnosti požadovanej úrovne informácií a vplyvu na súkromie jednotlivcov v dôsledku zverejnenia chránených atribútov.¹⁶

3.3.1.2 Permutácia

Permutácia preskupuje hodnoty atribútov v tabuľke tak, že niektoré z nich sú umelo prepojené s rôznymi dotknutými osobami, čím sa prerušia priame prepojenia medzi jednotlivcom a jeho osobnými údajmi. Napríklad v tabuľke obsahujúcej mená zamestnancov a ich platy by táto metóda premiešala zodpovedajúce platy medzi všetkými jednotlivcami bez zmeny výšky platu. Takáto výmena zabezpečí, že rozsah a rozloženie hodnôt zostanú rovnaké, ale korelácie medzi hodnotami a jednotlivcami sa nezmenia.¹⁷

3.3.1.3 Differential privacy

Metóda differential privacy využíva odlišný prístup: hoci v skutočnosti dochádza k vkladaniu šumu už pred zverejnením súboru údajov, táto metóda sa môže použiť, keď prevádzkovateľ generuje anonymizované pohľady na súbor údajov a zároveň si ponecháva kópiu pôvodných údajov. Metóda informuje prevádzkovateľa o tom, koľko šumu musí pridať a v akej forme, aby získal potrebné záruky súkromia.¹⁸

3.3.2 Zovšeobecnenie (generalization)

Druhá skupina anonymizačných techník pozostáva zo zovšeobecnenia alebo riadenia atribútov týkajúcich sa dotknutých osôb modifikovaním príslušnej mierky alebo rádu veľkosti.¹⁹

3.3.2.1 Agregácia and K-anonymita

Techniky agregácie a K-anonymity sa zameriavajú na zabránenie izolácie dotknutej osoby zoskupením s najmenej k ďalšími osobami. Na dosiahnutie tohto cieľa sa hodnoty atribútov zovšeobecnia do takej miery, že každá osoba zdieľa rovnakú hodnotu.²⁰

¹⁶ Ibidem

¹⁷ Ibidem

¹⁸ Cf. Ed Felten (2012) Protecting privacy by adding noise. URL: <https://techatftc.wordpress.com/2012/06/21/protecting-privacy-by-adding-noise/>.

¹⁹ ARTICLE 29 DATA PROTECTION WORKING PARTY. Opinion 05/2014 on Anonymisation Techniques. Adopted on 10 April 2014. 0829/14/EN WP216.

²⁰ Ibidem

3.3.2.2 L-diverzita

L-diverzita rozširuje k-anonymitu tak, aby sa zabezpečilo, že deterministické inferenčné útoky už nebudú možné tým, že sa zabezpečí, že v každej triede ekvivalencie má každý atribút aspoň l rôznych hodnôt. Jedným zo základných cieľov, ktoré treba dosiahnuť, je obmedziť výskyt tried ekvivalencie so slabou variabilitou atribútov, aby útočník so základnými znalosťami o konkrétnej dotknutej osobe vždy zostal so značnou neistotou. Je užitočné chrániť údaje pred inferenčnými útokmi, keď sú hodnoty atribútov dobre rozložené. Úniku informácií sa nedá zabrániť, ak sú atribúty v rámci oddielu nerovnomerne rozložené alebo patria do malého rozsahu hodnôt alebo sémantických významov. Nakoniec, l-diverzita je vystavená pravdepodobnostným inferenčným útokom.²¹

3.3.2.3 T-closeness

T-closeness je zdokonalením l-diverzity v tom zmysle, že sa zameriava na vytvorenie ekvivalentných tried, ktoré sa podobajú počiatočnému rozloženiu atribútov v tabuľke. Táto technika je užitočná, keď je dôležité udržať údaje čo najbližšie k pôvodným; na tento účel sa na triedu ekvivalencie kladie ďalšie obmedzenie, a to, že v každej triede ekvivalencie by malo existovať nielen aspoň l rôznych hodnôt, ale aj to, že každá hodnota je zastúpená toľkokrát, koľkokrát je potrebné na zrkadlenie počiatočného rozloženia každého atribútu.²²

Výber konkrétnych metód anonymizácie údajov alebo ich kombinácie, ktoré sa majú v praxi implementovať, musí určiť konkrétny prevádzkovateľ, pričom musí zohľadniť:

- povahu dotknutých údajov a ich citlivosť – rôzne kategórie údajov vyžadujú rôzne úrovne ochrany,
- rozsah súboru údajov (dataset),
- zamýšľaný účel budúceho použitia údajov a požadovanú presnosť sledovaného výsledku,
- riziko opätovnej identifikácie, napr. dostupnosť externých súborov údajov, ktoré by sa mohli prepojiť, alebo iných doplnkových informácií,
- technickú uskutočniteľnosť, škálovateľnosť a udržiavateľnosť,
- kontextovú integritu a ďalšie relevantné faktory.

Vzhľadom na široký rozsah možných súborov údajov a techník anonymizácie, ako aj na budúci vývoj techník (opätovnej) identifikácie, si akékoľvek posudzovanie bude vyžadovať analýzu s prihliadnutím na okolnosti konkrétneho prípadu.

²¹ Ibidem

²² Ibidem

3.4 Kontrola výsledku

Po výbere a implementácii techník anonymizácie údajov je posledným nevyhnutným krokom zabezpečenie dosiahnutia požadovaného výsledku, a to úspešnosti anonymizácie údajov. Preto je potrebné skontrolovať kvalitu výstupu vytvoreného aplikáciou zvolených techník anonymizácie údajov.

V nadväznosti na najnovší vývoj v oblasti hodnotenia úspešnosti anonymizácie navrhol Európsky výbor pre ochranu údajov tri kritériá, ktoré je v tejto súvislosti potrebné zohľadniť, a to konkrétne: (1) vylúčenie izolácie záznamu, (2) vylúčenie prepojenia a (3) vylúčenie vyvodzovania záverov. Kritérium vylúčenia izolácie záznamu je splnené, ak údaje neobsahujú jedinečnú kombináciu hodnôt atribútov, ktorá sa vzťahuje na jednu konkrétnu fyzickú osobu. Kritérium vylúčenia prepojenia je splnené, ak údaje neobsahujú záznam o fyzickej osobe, ktorý by bolo možné prepojiť s iným záznamom, ktorý (a) sa takisto vzťahuje (s istotou alebo vysokou pravdepodobnosťou) na tú istú fyzickú osobu a (b) pochádza z iného súboru údajov. Kritérium vylúčenia vyvodzovania záverov je splnené, ak z daných údajov nemožno vyvodiť žiadny konkrétny a významný záver. Vyvodené závery môžu vychádzať z daných údajov a/alebo z dodatočných informácií vrátane informácií o procese anonymizácie. Konkrétne:

- Vyvodený záver je konkrétny, ak sa vyvodená informácia vzťahuje na jednu identifikovanú alebo identifikovateľnú fyzickú osobu (t. j. ide o osobné údaje); a
- Vyvodený záver je konkrétny a zároveň významný, ak spracúvanie vyvodenej informácie môže mať vplyv na práva a záujmy dotknutej osoby, pričom tento záver vychádza z daných údajov a nebolo by možné ho získať na základe všeobecných poznatkov alebo údajov o populácii ako celku.²³

Ak sú splnené všetky tri kritériá, informácie možno považovať za anonymné.

3.5 Využitie anonymizovaných digitálnych stop pre budúcu analýzu

Anonymizácia digitálnych dôkazov si môže zachovať svoju užitočnosť pre budúcu automatizovanú analýzu, ak je navrhnutá s ohľadom na zachovanie štruktúry údajov, a nie výlučne na jednoduché vymazanie určitých údajov. Namiesto odstraňovania alebo náhodného maskovania identifikačných polí by sa mali používať rôzne techniky, ktoré zachovávajú štatistické vzory, vzťahy a formáty, na ktorých sa algoritmy spoliehajú. Udržiavanie konzistentnej schémy mapovania – kde je tá istá entita reálneho sveta vždy nahradená rovnakým anonymizovaným tokenom – zabezpečuje, že modely rozpoznávania vzorov, analýzy prepojení a strojového učenia dokážu stále detegovať korelácie medzi záznamami bez

²³ European Data Protection Board. Guidelines 02/2026 on Anonymisation. Adopted on 07 July 2026 for public consultation.

toho, aby odhalili pôvodné identity. Je tiež dôležité zdokumentovať metodiku anonymizácie a uchovávať metadáta o typoch údajov, jednotkách a vzťahoch, aby automatizované nástroje mohli správne interpretovať transformovaný súbor údajov. Nakoniec by sa anonymizácia mala overiť pomocou špecifických analytických techník určených na budúce použitie (napr. klastrovanie, identifikácia anomálií, asociačné pravidlá atď.), aby sa potvrdilo, že základné vlastnosti a odchýlky v údajoch nie sú neúmyselne sploštené alebo zničené, čím sa dosiahne rovnováha medzi ochranou súkromia a analytickou užitočnosťou.

4 Prehľad



METÓDA PRE ANONYMIZÁCIU DIGITÁLNYCH STÔP



INVENTARIZÁCIA ÚDAJOV

- určenie všetkých spracovávaných informácií
- vytvorenie zoznamu takýchto informácií, napr. vo forme tabuľky



KLASIFIKÁCIA ÚDAJOV

- preskúmanie spracovaných informácií a určenie ich právnej povahy, najmä či sa na takéto informácie vzťahuje osobitná právna ochrana
- tabuľka informácií by mala obsahovať dosiahnuté závery a určené kategórie informácií.

OSOBNÉ V. NEOSOBNÉ ÚDAJE

- čl. 4 ods. 1 GDPR definuje osobné údaje ako „akékoľvek informácie týkajúce sa identifikovanej alebo identifikovateľnej fyzickej osoby (ďalej len „dotknutá osoba“); identifikovateľná fyzická osoba je osoba, ktorú možno identifikovať priamo alebo nepriamo, najmä odkazom na identifikátor, ako je meno, identifikačné číslo, lokalizačné údaje, online identifikátor, alebo odkazom na jeden či viaceré prvky, ktoré sú špecifické pre fyzickú, fyziologickú, genetickú, mentálnu, ekonomickú, kultúrnu alebo sociálnu identitu tejto fyzickej osoby.“
- **neosobné údaje**
 - a. údaje, ktoré sa pôvodne netýkali identifikovanej alebo identifikovateľnej fyzickej osoby
 - b. údaje, ktoré boli pôvodne osobnými údajmi, ale neskôr boli anonymizované (anonymné údaje)



SÚLAD S PRÁVNOU ÚPRAVOU

- ak sa zistí, že skúmané údaje musia byť anonymizované, tento proces musí byť v súlade s platnými predpismi
- anonymizácia predstavuje **formu spracovania osobných údajov**
- akékoľvek spracovanie osobných údajov podlieha zásadám týkajúcim sa spracovania osobných údajov uvedeným v článku 5 GDPR



VÝBER ANONYMIZAČNÝCH METÓD

- anonymizácia údajov by sa mala vždy vykonávať s použitím najnovších anonymizačných techník, ktoré odrážajú vývoj v oblasti technológií a analýzy údajov



KONTROLA VÝSLEDKU

- kontrola úspešnosti anonymizácie údajov

