20. 5. 2020

Vezměte, prosíme, na vědomí, že text článku odpovídá platné právní úpravě ke dni publikace.

Seriál specifické aspekty ochrany osobních údajů 3/5 Anonymizace a pseudonymizace v ochraně dat a informací

Anonymizace a pseudonymizace osobních údajů, tedy techniky zpracování údajů, zmíněné již v prvním dílu našeho seriálu, mohou z určitého úhlu pohledu vypadat značně abstraktně. V očích laické veřejnosti pak mohou v mnoha případech splývat. Jejich nedůsledné rozlišování či nevhodná aplikace však nicméně může mít konkrétní nepříznivé následky. Cílem tohoto textu je nabídnout srozumitelný výklad těchto dvou pojmů na konkrétním případě, a následně čtenářům přiblížit technické aspekty vybraných anonymizačních a pseudonymizačních technik.

Rozlišování mezi anonymizací a pseudonimazací

Potřebu rozlišování mezi oběma koncepty lze ilustrovat na zpracování biometrických údajů sloužících pro ověření identity vstupu do zabezpečených objektů, případně do softwarových databází apod. Zaměstnavatelé, zejména velké společnosti, totiž mohou mít nezřídka zájem na zavedení opatření, jejichž podstata spočívá v umožnění přístupu na základě užití otisku prstů, rozeznání obličeje či třeba specifických znaků oka. Tím lze totiž snížit některá rizika spojená se zneužitím přístupu, typicky spočívající v hrozbě odcizení klíčů, vstupních karet nebo odhalení hesel. Zásadním problémem však je, že biometrické údaje jsou řazeny mezi tzv. zvláštní kategorie osobních údajů.[1] Pro jejich zpracování je nutné naplnit některou z výjimek ze zákazu tyto údaje zpracovávat, a to dle čl. 9 odst. 2 GDPR. Na první pohled se nabízí výjimka ve smyslu písm. a) uvedeného ustanovení, tedy udělení výslovného souhlasu subjektu údajů s daným zpracováním. Použitelnost této výjimky je však k subjektům údajů v pozici zaměstnanců, tedy slabší strany, značně limitována. Jejich souhlas totiž obecně není považován za svobodný, a tedy platný, a to právě s odkazem na jejich závislé postavení vůči správci v pozici zaměstnavatele.[2] Využitelnost ostatních výjimek pro výše uvedený případ je pak spíše ojedinělá.

Pro řešení nastíněného problému je někdy uvažováno právě o možnosti anonymizace[3] – tj. procesu, kdy jsou osobní údaje nahrazeny údaji anonymními, tedy těmi, které nepodléhají GDPR[4] a ZZOÚ. Při této úvaze však narazíme na skutečnost, že i po převedení otisku prstů např. do číselného kódu nebo do bodové šablony není odstraněno propojení mezi subjektem údajů a danou informací, tudíž se stále jedná o osobní údaje.[5] Právě za účelem jednoznačného určení fyzické osoby je totiž celá aktivita prováděna. V daném případě tak ve skutečnosti nedochází k anonymizaci, ale pouze k tzv. pseudonymizaci, tedy zejména k omezení skupiny osob, která je schopna určité údaje spojit s jejich subjektem. To i proto, že v souvislosti s přístupem na základě biometrických údajů nutně probíhají i jiné zpracovatelské aktivity, než samotné uložení údajů, typicky jejich konverze a porovnávání. Výše uvedené ale neznamená, že by tento krok (tj. pseudonymizace) byl při zpracování biometrických údajů zbytečný. Sice jím nedocílíme možnosti zpracovávat určité údaje zaměstnanců na základě jejich souhlasu, můžeme jej však využít jako adekvátní bezpečnostní opatření při zpracovávání biometrických údajů[6] např. obchodních partnerů, u nichž není faktor svobodné vůle tak problematický. Stejně tak se může jednat o vhodné opatření v případě, že je naplněna jiná z výjimek pro zpracování údajů zvláštní kategorie.

Zjevný problém s možností zpracovávat biometrické údaje zaměstnanců začal aktivně řešit i ÚOOÚ. Ten vznesl připomínku k případné novelizaci současného zákoníku práce, která by (při splnění za dalších podmínek) zaměstnavatelům umožňovala využívat k ochraně výrobních a pracovních prostředků a technologií biometrické údaje identifikující zaměstnance a využívající pouze jejich morfologické znaky. Taková úprava zákoníku práce by byla v souladu s čl. 9 odst. 4 GDPR.[7]

Anonymizaci je vhodné (a někdy i nutné)[8] použít v případech, kdy pro nás to, že se údaje váží ke konkrétní fyzické osobě, není důležité a účelu zpracování tak můžeme docílit i bez uchování daného propojení. Toto je časté např. v sociologických průzkumech, výzkumech ve zdravotnictví a dalších odvětvích, kde pracujeme převážně se statistickými údaji. Pseudonymizaci pak zapojujeme v případech, kdy je pro nás spojitost údaje s danou osobou i nadále zásadní, avšak z bezpečnostních či jiných důvodů a zájmů je třeba omezit okruh osob, které by tuto spojitost mohly snadno dovodit.[9]

Nyní, když jsme si osvětlili využitelnost a základní rozdíly mezi anonymizací a pseudonymizací, můžeme se podívat na vybrané techniky (respektive skupiny technik) pro provádění těchto činností.

Anonymizační techniky

V rámci anonymizace rozlišujeme dvě základní skupiny postupů. Jedná se randomizaci a generalizaci.[10] Při jejich výběru či vzájemné kombinaci je třeba vždy zvolit tu techniku, která nám umožní docílit stavu, kdy nelze rozumně předpokládat použití prostředků k identifikaci fyzické osoby, ke které se údaje vztahují.[11]

Randomizace

Randomizační techniky obecně spočívají v takové úpravě údajů, kdy je narušena spojitost údaje s určitou osobou, ale nedochází ke ztrátě zásadní informační hodnoty daného materiálu.[12] Konkrétní – pro výsledek nepotřebné údaje – bývají nahrazovány údaji náhodnými. Stejný vstup, např. jméno Artur (bude-li v daném případě označitelné jakožto osobní údaj), bude v rámci randomizačních technik typicky převedeno na různé výstupy – kódy, a to třeba i v rámci jednoho dokumentu. Dalším častým úkonem randomizace je provedení dílčích změn, čímž se snižuje organizační hodnota daného systému a opět se tak přispívá k zastření vazby mezi údajem a jeho subjektem.[13] Za specifickou metodu randomizace pak teoreticky můžeme označit dostatečné začernění údajů (bez možnosti srovnání s původním dokumentem a dalšími identifikátory) nebo jejich nahrazení jednotnými znaky jako např. *** nebo xxx.

Generalizace

Generalizační techniky spočívají v zobecnění informací do takové míry, že je výrazně zvýšen počet subjektů údajů, kterým by daný atribut mohl být přiřazen. V rámci deduktivních metod, které se mohou snažit generalizované systémy rozklíčovat, pak má osobě, která se o rozklíčování snaží, vždy zůstat značná nejistota. Generalizační metody jsou tedy vhodné zejména pro velké sady dat, tak aby výsledný data set obsahoval hodnoty společné vždy pro alespoň několik jedinců. Zatímco vztahovat údaj o narození „září 1990“ na všechny osoby v České republice může zajistit dostatečnou anonymitu proto, že je tento údaj společný velkému množství jedinců v rámci dané skupiny, vztahovat tentýž údaj např. ke kolektivu jedné školní třídy již dostatečnou anonymizaci zpravidla představovat nebude.[14]

Pseudonymizační techniky

Pseudonymizačních technik je relativně velké množství. My se zaměříme na ty dvě nejčastější, přičemž za jejich hlavní rozdíl může být s trochou zjednodušení považováno to, že v prvním případě je pseudonymizace obousměrná a z výsledného souboru se dá zpětně vygenerovat soubor původní,

resp. se data dají dešifrovat. V druhém případě je pak funkce jednosměrná.

Šifrování

Šifrování, které je někdy označováno jako samostatná technika (vedle pseudonymizace),[15] je možné provádět jak za pomocí automatických funkcí, kdy je na základě předem nastaveného klíče přiřazována vybraným atributům nová hodnota, pod kterou jsou následně prezentovány, tak i manuálně. V prvním případě není po provedené pseudonymizaci zpravidla potřeba uchovávat zvláštní zdrojový soubor, neboť zdrojem je samotná funkce, jejímž "obrácením", nebo jinou metodou, můžeme původní soubor získat.[16] Při šifrování jména autor za použití stejné šifrovací funkce pak obecně získáváme tentýž výstup. Sofistikovanější metody pak mohou být navázány např. na polohu vstupních dat v rámci systému. I nadále ale platí, že pro tytéž vstupy získáváme tytéž šifrované výstupy. Alternativně je možné osobní údaje nahradit způsobem, že vytvoříme pseudonymizovaný soubor, např. s výsledky přijímacích zkoušek, kde na místo jmen uchazečů budou pouze vygenerované kódy, a tento soubor bude sloužit ke zveřejnění za dodržení zásady minimalizace osobních údajů (bude-li toto s přihlédnutím k dalším aspektům vyhodnoceno jako vhodná metoda pro zveřejnění). Soubor s klíčem, tedy takový soubor, ze kterého bude patrný vztah kódu a jména uchazeče, pak zveřejněn nebude a bude sloužit zejména pro interní potřeby dané školy. Smyslem šifrovaných údajů je však možnost jejich dešifrování ve chvíli, kdy se dostanou do sféry jejich adresáta, který zároveň disponuje možností aplikovat dešifrovací funkci – tedy možnost obousměrné konverze mezi vstupem a výstupem.

Hashování

Hashování spočívá v převedení vstupních dat do výstupního kódu (hashe), přičemž tato aktivita je jednosměrná. To znamená, že by nemělo být možné převést hash zpět do vstupního formátu. Mezi obchodními partnery se pak mohou, typicky pro autentizaci osob, vyměňovat pouze hashe. Zásadní vlastností většiny hashovacích funkcí je, že vytvářejí zásadně odlišné výstupy i pro jen mírně odlišná data. Pro stejná data pak stejná hashovací funkce vytvoří vždy stejný výstup. Pokročilejší hashovací opatření pak mohou zahrnovat např. tzv. "salted and pappered hashe", kdy jsou k datům před jejich zahashováním přidána další související data a teprve potom je spuštěna samotná funkce.[17]

Závěr

Důsledné rozlišování mezi anonymizací a pseudonymizací nám umožní určit nejen rozsah povinností, které se vážou ke zpracování určitých údajů , ale uvědomit si např. i to, že soudní rozhodnutí, byť se o nich mluví jakožto o anonymizovaných, jsou veskrze pseudonymizovaná. Zhodnocení rozdílnosti jednotlivých technik je pak prvním krokem pro výběr vhodné metody (či jejich kombinace) k tomu, aby byl celý proces opravdu úspěšný; ne jen zdánlivý, tedy bez větších potíží rozluštitelný a vratný. Riziko, že užitá technika bude zvrácena, by přitom mělo být stěžejním ukazatelem, zda byl vybrán vhodný postup. Toto je pak pochopitelně třeba revidovat i v průběhu času, neboť zabezpečení osobních údajů (a dodržení podmínek stanovených GDPR a ZZOÚ) po celou dobu jejich zpracování je kontinuální proces.

Radek Buršík

Jan Svoboda

PricewaterhouseCoopers Legal s.r.o., advokátní kancelář
PwC Legal

City Green Court
Hvězdova 1734/2c
140 00 Praha 4

Tel.: +420 251 151 111
Fax: +420 251 156 111

Svobody 91/20
602 00 Brno

Tel.: +420 542 520 111
Fax: +420 542 214 796

e-mail: info@pwc.cz

---

[1] Čl. 9 odst. 1 GDPR.

[2] Srov. ICO. When Is consent appropriate? [online]. [Cit. 18. 5. 2020]. K dispozici >>> zde.

[3] Srov. ÚOOÚ. Změna v hodnocení úrovně právní ochrany biometrických údajů [online]. 2017 [cit. 18. 5. 2020]. K dispozici >>> zde.

[4] Viz rec. 26 preambule GDPR.

[5] Srov. ÚOOÚ. Změna v hodnocení úrovně právní ochrany biometrických údajů [online]. 2017 [cit. 18. 5. 2020]. K dispozici >>> zde.

[6] Srov. např. čl. 25 odst. 1 GDPR.

[7] ÚOOÚ. Připomínka k návrhu zákona, kterým se mění zákon č. 262/2006 Sb., zákoník práce, ve znění pozdějších předpisů, a zákon č. 435/2004 Sb., o zaměstnanosti, ve znění pozdějších předpisů [online]. 2019. 3 s. [cit. 18. 5. 2020]. K dispozici >>> zde.

[8] Srov. např. § 16 odst. 2 ZZOÚ.

[9] Viz např. § 16 odst. 1 písm. f) ZZOÚ.

[10] WP 29. Opinion 05/2014 on Anonymisation Techniques [online]. 2014, s. 12 [cit. 18. 5. 2020]. K dispozici >>> zde.

[11] Viz rec. 26 GDPR.

[12] WP 29. Opinion 05/2014 on Anonymisation Techniques [online]. 2014, s. 3 [cit. 18. 5. 2020]. K dispozici >>> zde.

[13] Data Protection Commission. Guidance on Anonymisation and Pseudonymisation. [online]. 2019, s. 11 [cit. 18. 5. 2020]. K dispozici >>> zde.

[14] Srov. Data Protection Commission. Guidance on Anonymisation and Pseudonymisation [online]. 2019, s. 12 [cit. 18. 5. 2020]. K dispozici >>> zde.

[15] Srov. čl. 6 odst. 4 písm. e) GDPR.

[16] Srov. BURDA, K. Kryptografie okolo nás [online]. 2019, s. 20 a násl [cit. 18. 5. 2020]. K dispozici >>> zde.

[17] Viz FINCK, M. a PALLAS, F. They who must not be identified—distinguishing personal from non-personal data under the GDPR. International Data Privacy Law [online]. 2020, s. 24-28 [cit. 18. 5. 2020]. K dispozici >>> zde.

Další články:

- Nový zákon o veřejných dražbách, aukce a obálkové metody
- Pohled přes hranice – natáčení pornografických klipů jako důvod výpovědi z nájmu bytu
- Nařízení EU o umělé inteligenci a jeho dopady na využití jazykových modelů v advokátní praxi
- Revize zájezdové směrnice: co přináší, co hrozilo a co to znamená pro praxi
- Kupní smlouva o převodu nemovitosti bez uvedení výše kupní ceny
- Druhá „tlačítková novela": povinné tlačítko pro odstoupení od smlouvy
- Souhlas s veřejným užíváním pozemku jako překážka nároku na bezdůvodné obohacení – nález Ústavního soudu sp. zn. I. ÚS 2541/25
- Kupní smlouva bez přesného určení kupní ceny
- Byznys a paragrafy, díl 36.: Doložka o mlčenlivosti
- Detekce podezřelého obchodu v kontextu hazardních her
- AI omnibus