



Datu valsts inspekcija

Personas datu dzēšana un anonimizācija

Personas datu apstrādes uzraudzības nodaļas juriskonsults Matīss Liepiņš

67223131

16.03.2018.

Rīga

Personas dati: ir vai nav

- “Personas dati” ir jebkura informācija, kas attiecas uz identificētu vai identificējamu fizisku personu;
- identificējama fiziska persona ir tāda, kuru var tieši vai netieši identificēt, jo īpaši atsaucoties uz identifikatoru, piemēram, minētās personas vārdu, uzvārdu, identifikācijas numuru, atrašanās vietas datiem, tiešsaistes identifikatoru vai vienu vai vairākiem minētajai fiziskajai personai raksturīgiem fiziskās, fizioloģiskās, garīgās, sociālās identitātes faktoriem u.c.;
- Datu aizsardzības principi būtu jāattiecina uz jebkādu informāciju par identificētu vai identificējamu fizisku personu.

Datu dzēšana

Personas datu dzēšana ir informācijas, kas attiecas uz fizisku personu, apstrādes pārtraukšana.

Personas dati var tikt dzēsti dažādos apjomos:

- **Pilnībā.**

Piemēram, tiek neatgriezeniski iznīcināts informācijas nesējs vai likvidēta tajā esošā informācija, kas attiecas uz datu subjektu.

- **Daļēji** – tiek dzēsta noteikta daļa no informācijas. Daļējas dzēšanas gadījumā nav pamatoti paļauties, ka personas identitāte ir noslēpta pilnībā; šāda dzēšana var nodrošināt vien noteikta apjoma drošības garantijas datu subjektam.

Piemēram, tiek izdzēsti tiešie identifikatori – vārds, uzvārds, personas kods, turpinot citas informācijas apstrādi.

Datu dzēšana

- Personas datu dzēšana ir veicama, ja attiecībā uz šiem datiem nepastāv tiesiskais pamats un/vai personas dati vairs nav nepieciešami saistībā ar nolūkiem, kādos tie tika vākti vai citādi apstrādāti.

Piemēram, ja datu subjekts atsauc savu piekrišanu, uzskatāms, ka personas datu apstrādei vairs nav tiesiska pamata.

- Šie nosacījumi attiecas ne vien uz noteiktu datu kopu, piemēram, kartotēku, atsevišķu dokumentu, lietu u.c., bet arī uz konkrētam kopas daļām.

Datu dzēšana

- Attiecībā uz turpmāku apstrādi statistikas un pētniecības nolūkam, ja apkopojamās informācijas raksturs, iespējams, var tieši vai netieši identificēt personu, personas datu aizsardzība būtu nodrošināma ar anonimizācijas metodēm.
- Vērtējot informācijas raksturu uzmanība pievēršama specifiskai informācijai (piem., reta saslimšana) un informācijai, kas ietver vairākus aspektus (piem., persona ir sasniegusi simts gadu vecumu un ir piedalījusies Olimpiskajās spēlēs).

Anonīma informācija

Datu aizsardzības principi nebūtu jāpiemēro anonīmai informācijai, proti, informācijai, kura neattiecas uz:

- identificētu vai identificējamu fizisku personu;
- personas datiem, ko sniedz anonīmi tādā veidā, ka datu subjekts nav vai vairs nav identificējams.

Kas ir anonimizācija?

- Anonimizēta informācija ir tāda, kura nepieļauj identificēt personu, uz kuru attiecās sākotnēji apstrādātā un/vai faktiski attiecināmā informācija.
- Personas datu anonimizācija pēc būtības ir neatgriezeniska – anonimizācija pielīdzināma informācijas pilnīgai dzēšanai.

Kas ir anonimizācija?

Taču..

- Lai noteiktu, vai fiziska persona ir identificējama, būtu jāņem vērā visi līdzekļi, ko pārzinis vai kāda cita persona pamatoti varētu izmantot – piemēram, atsevišķa izdalīšana –, lai tieši vai netieši identificētu fizisku personu.
- Lai pārliecinātos, vai līdzekļus varētu saprātīgi izmantot fiziskas personas identificēšanai, būtu jāņem vērā visi objektīvie faktori, piemēram, identificēšanai nepieciešamās izmaksas un laiks, ņemot vērā apstrādes laikā pieejamo tehnoloģiju un tehnoloģiju attīstību.

- Lai arī pēc būtības anonimizācija ir neatgriezeniska, VDAR neizslēdz hipotētisku iespēju, ka tā var nodrošināt identifikāciju, jo vērtē nepieciešamās izmaksas un laiks, ņemot vērā apstrādes laikā pieejamo tehnoloģiju un tehnoloģiju attīstību. Taču šo norādi nedrīkst interpretēt vieglprātīgi, proti, vērtēt, vai identifikāciju var veikt «vidusmēra» sabiedrības loceklis, konkrētais informācijas saņēmējs, IT speciālists, matemātiķis u.c.
- Ja pārzinis, veicot anonimizāciju, pieļauj jebkādu iespēju, ka informācija var tikt «atšifrēta», ir nepieciešams vērtēt citu vai papildu līdzekļu izmantošanu.

Kas nav anonimizācija?

Anonimizācija nav:

- kriptēšana vai pielīdzināmi līdzekļi;
- personas identifikātoru dzēšana vai izņemšana no informācijas kopuma;
- pseidonimizācija.

- Iegūtā datu kopa bez personas datiem nav uzskatāma par anonimizētu, ja informācijas kopas oriģināls ar personas datiem tiek saglabāts (ir pārziņa vai citas personas rīcībā) vai tiek saglabāts tā, ka tas var tikt salīdzināts ar modificēto datu kopu.

Anonimizācijas metodes

Pastāv šādas datu anonimizācijas metodes:

- Sajaukšana (randomization);
 - Vispārināšana (generalization).
-
- Šo metožu piemērošanas gadījumā personas tiešie identifikatori netiek izmantoti, un tās ir vērstas uz identificēšanas iespējas pēc netiešiem identifikatoriem izslēgšanu.

Sajaukšanas metode

Sajaukšanas metode tiecas «likvidēt» informācijas precizitāti attiecībā uz konkrētu personu, neietekmējot gala informācija funkcionalitāti (mērķa sasniegšanu). Tas tiek panākts, izslēdzot saikni starp informāciju un datu subjektu.

Sajaukšanas tehnikas

Sajaukšanas metode ietver sevī šādas tehnikas:

1. «Noise addition»

Uz personu attiecināmas raksturīgās pazīmes tiek noteiktā pakāpē pārveidotas (samazinātas, palielinātas) jeb tiek pievienoti «trokšņi».

Tā kā informācijas precizitātes labad informācija var netikt pārveidota tādā apmērā, ka tās apstrāde zaudē jēgu, šo tehniku var būt nepieciešams lietot kā papildu līdzekli.

Lai arī ir iespējams ar šādu tehniku apstrādātu informāciju izmantot datu subjekta identificēšanai («izskaitļot»), trešās personas iegūtā informācija par datu subjektu nebūs pilnībā precīza, kā arī pastāv iespēja iegūt viltus pozitīvu informāciju (t.i., par citu subjektu).

2. Permutācijas;

Ar permutācijas tehniku datu subjektu raksturīgās pazīmes datu kopā tiek sajauktas – viena personas noteiktas pazīmes tiek mākslīgi piešķirtas citai.

Šī tehnika ir pielietojama, kad informācijai (vērtībai) datu kopā jābūt precīzai, taču nav būtiskas nozīmes korelācijai starp konkrētām pazīmēm (piem., dzimums pret atalgojuma apmēru).

Nosakot attiecīgās pazīmes, kas tiks apmainītas, ir nepieciešams izvēlēties pazīmes ar būtisku nozīmi vai ar paaugstinātu identificēšanas risku.

Līdzīgi kā «skaņu pievienošanas» tehnikas gadījumā, tā patstāvīgi var nenodrošināt pilnīgu anonimizāciju, taču informācija «atšifrēšanas» gadījumā tās saņēmējs var iegūt neprecīzu un maldinošu informāciju.

Vispārināšanas metode

Šīs metodes pamatā ir uz personu attiecināmo pazīmju vispārināšana, modificējot (paplašinot) noteiktas vērtības - apjomus, mērvienības, mērogus u.c.

Piemērs: Datu kopā norādīta informācija par slimības izplatību reģionā, nevis konkrētās šī reģiona pilsētās.

Vispārināšanas metode

Vispārināšanas metode ietver sevī šādas tehnikas:

1. Apkopošana un K-anonimitāte (aggregation and k-anonymity)

Šīs tehnikas tiecas novērst personas atpazīšanu, sagrupējot šo personu ar noteiktu skaitu citu personu, kas ietilpst paplašinātās vērtības grupā. Proti, noteiktu pazīmju detalizācijas pakāpe tiek modificēta, to samazinot.

Vispārināšanas metode

- Konkrētu pazīmju vērtības tiek vispārinātas tā, ka datu kopā esošo personu noteiktās pazīmes atbilst pārējo citu personu (vai daļas no šīm personām) pazīmēm.
- Piemērs: Veicot statistiku par personām, kas saņem noteiktus valsts pabalstus un saņem noteiktu ārstēšanu, personas tiek sadalītas pa vecuma grupām (nevis norādot precīzu vecumu) un pēc ārstēšanas veidiem (nenorādot konkrētu saslimšanu, bet medicīnas jomu).

Vispārināšanas metode

2. L-diversity

L-diversity ir K-anonimitātes tehnikas paplašinājums, kas vērsts uz to, lai ierobežotu personas konkrētas pazīmes vērtības (piem., vecuma) noteikšanu pēc citu atspoguļoto pazīmju vērtībām, tādējādi pieļaujot identificēšanu. Tas tiek panākts, nodrošinot, ka katrā grupā ar vienlīdzīgām pazīmēm katrai pazīmei ir vismaz viena atšķirīga vērtība.

Galvenais mērķis ir ierobežot gadījumus, kad vienlīdzīgas klases ar zemu pazīmju dažādību (piemēram, 3 kritēriji), tā ka interesents ar iepriekšējām zināšanām par konkrēto datu subjektu, izdarot secinājumus par personu, nevar būt pilnībā drošs par to, vai iegūtā informācija attiecas uz šo personu.

Pseidonimizācija

- Pseidonimizācija ir identitātes maskēšanas process. Šāda procesa mērķis ir iespēja vākt papildu datus, kas attiecas uz to pašu personu, nezinot tās identitāti.
- Tā samazina iespēju savienot informāciju ar personas identitāti, taču tā ir uzskatāma par drošības līdzekli, nevis anonimizācijas metodi.
- Pseidonimizāciju var veikt izsekojamā veidā, lietojot sarakstus ar identitātēm atbilstošajiem pseidonīmiem vai izmantojot pseidonimizācijai divvirzienu kriptogrāfijas algoritmus.
- Pseidonimizācijas procesa efektivitāte ir atkarīga no vairākiem faktoriem (kurā datu apstrādes procesa stadijā to pielieto, cik tā ir droša pret atpakaļ izsekojamību, cik liela ir datu kopa, kurā personas dati ir pseidonimizēti, kāda ir iespēja sasaistīt individuālus ierakstus ar to pašu personu u.tml.).
- Pseidonīmiem būtu jābūt nejauši izvēlētiem un iepriekš nenosakāmiem. Iespējamo pseidonīmu skaitam ir jābūt tik liels, lai to pašu pseidonīmu divas reizes nekad nevarētu izvēlēties. Ja ir nepieciešams augsts drošības līmenis, iespējamo pseidonīmu skaitam ir jābūt vismaz vienādam ar drošas kriptogrāfiskās jaucējfunkcijas vērtību skaitu.

Pseidonimizācija

- Izsekojami pseidonimizēti dati ir uzskatāmi par informāciju, kas attiecas uz netieši identificējamām personām. Patiesībā pseidonīma izmantošana nozīmē, ka pastāv iespēja nonākt atpakaļ pie personas un personas identitāti var atklāt, taču tikai jau iepriekš konkretizētos apstākļos.
- Šādā gadījumā, kaut arī datu aizsardzības noteikumi ir piemērojami, attiecībā uz šādas netieši identificējamās informācijas apstrādi konkrētai personai risks lielākoties būs zems, un tādējādi ir attaisnojama šo noteikumu elastīgāka piemērošana nekā tad, ja tiek apstrādāta informācija par tieši identificējamām personām.

Pārziņa apsvērumi

- Pārzinim ir nepieciešams pirms informācijas apstrādes (apkopošanas u.c.) apzināt mērķus šādai darbībai, kā arī to sasniegšanai nepieciešamo informācijas apjomu.
- No minētajiem apsvērumiem izriet anonimizācijas, pseidonimizācijas un dzēšanas metodes un apjoms.
- Piemērs: Ja nepieciešams apkopot statistiku par sūdzību skaitu attiecībā uz konkrētiem pārkāpumiem, pārzinim nav nepieciešams apkopojumā norādīt sūdzību sadalījumu pēc reģioniem, vecuma u.c.

Pārziņa apsvērumi

- Jāvērtē, vai pastāv citas datu kopas (dataset), kuras, aplūkojot kopsakarā ar apkopoto informāciju, var identificēt personas.
- Pārzinim, tiecoties sasniegt pilnīgu anonimizāciju, pēc iespējas ieteicams papildus domāt «apgriezti».
- Proti, šaubu gadījumā sākotnēji apsvērt nevis, kas no informācijas kopuma ir jādzēš, lai neidentificētu personu, bet, kas tai ir pievienojams, lai vispārēju informāciju varētu attiecināt uz kādu, iespējams, netieši identificējamu personu.

Paldies par uzmanību!

**Datu valsts inspekcija
Blaumaņa iela 11/13-15, Rīga, LV-1011**

**Konsultācijas un informāciju pa tālruni sniedz darba dienās
no plkst. 13.00 līdz 15.00**

Tel.: 67223131

Fakss: 67223556

E-pasts: info@dvi.gov.lv

www.dvi.gov.lv

