PAGAL 29 STRAIPSNĮ ĮKURTA DUOMENŲ APSAUGOS DARBO GRUPĖ

0829/14/LT
WP216

Nuomonė 05/2014 dėl nuasmeninimo metodų

Priimta 2014 m. balandžio 10 d.

Ši darbo grupė įkurta pagal Direktyvos 95/46/EB 29 straipsnį. Ji yra nepriklausomas Europos patariamasis organas duomenų apsaugos ir privatumo klausimams. Grupės užduotys nustatytos Direktyvos 95/46/EB 30 straipsnyje ir Direktyvos 2002/58/EB 15 straipsnyje.

Sekretoriatas yra Europos Komisijos Teisingumo generalinio direktorato C direktorate (Pagrindinės teisės ir ES pilietybė), kuris įsikūręs adresu: B-1049 Brussels, Belgium; kabinetas Nr. MO-59 02/013.

Interneto svetainė: http://ec.europa.eu/justice/data-protection/index_en.htm

ASMENŲ APSAUGOS TVARKANT ASMENS DUOMENIS DARBO GRUPĖ,

įkurta 1995 m. spalio 24 d. Europos Parlamento ir Tarybos direktyva 95/46/EB,

atsižvelgdama į minėtos direktyvos 29 ir 30 straipsnius,

atsižvelgdama į savo darbo tvarkos taisykles,

PRIĖMĖ ŠIĄ NUOMONĘ:

SANTRAUKA

Šioje nuomonėje darbo grupė, paisydama duomenų apsaugos ES teisinio pagrindo nuostatų, nagrinėja esamų nuasmeninimo metodų veiksmingumą bei apribojimus ir pateikia rekomendacijas, kaip taikyti šiuos metodus atsižvelgiant į kiekvienam iš jų būdingą liekamąją asmens tapatybės nustatymo riziką.

Darbo grupė pripažįsta, kad nuasmeninimas gali būti vertingas, ypač kaip atviruųjų duomenų panaudojimo pavienių asmenų ir visos visuomenės reikmėms, kartu mažinant susijusiems asmenims gresiančius pavojus, strategija. Vis dėlto konkrečių atvejų tyrimai ir moksliniai straipsniai parodė, kaip sunku parengti visiškai anoniminį duomenų rinkinį, kartu išsaugant užduočiai atlikti svarbią informaciją.

Vadovaujantis Direktyva 95/46/EB ir kitais susijusiais ES teisės aktais, anonimiškumo pasiekiama asmens duomenis tvarkant taip, kad nebebūtų galima atsekti asmens tapatybės. Todėl duomenų valdytojai, atsižvelgdami į visas priemones, kuriomis koks nors valdytojas arba trečioji šalis „galėtų“ pasinaudoti asmens tapatybei nustatyti, turėtų įvertinti kelis aspektus.

Nuasmeninimas – tai tolesnis asmens duomenų tvarkymas; toks procesas turi atitikti suderinamumo reikalavimą, t. y. turi būti vykdomas atsižvelgiant į teisinį pagrindą ir tolesnio tvarkymo aplinkybes. Be to, nuasmenintiems duomenims netaikomi duomenų apsaugos teisės aktai, tačiau duomenų subjektams vis vien gali būti suteikta teisė į apsaugą pagal kitas nuostatas (pvz., dėl pranešimų konfidencialumo apsaugos).

Šioje nuomonėje aprašyti pagrindiniai nuasmeninimo metodai, t. y. randomizavimas ir apibendrinimas. Pirmiausia šioje nuomonėje aptariami šie konkretūs metodai: iškraipytų duomenų įterpimas, perstatymas, diferencinis privatumas, agregavimas, *k* anonimiškumas, *l* įvairovė ir *t* tankis. Nuomonėje aiškinami šių metodų principai, jų privalumai ir trūkumai, taip pat dažniausios su kiekvieno metodo taikymu susijusios klaidos ir nesėkmės.

Ši nuomonė parengta remiantis kiekvieno metodo patikimumu, kuris grindžiamas šiais trimis kriterijais:

i) ar išlieka galimybė išskirti pavienį asmenį;

ii) ar išlieka galimybė susieti įrašus, susijusius su pavieniu asmeniu;

iii) ar iš informacijos galima gauti išvestinių duomenų apie pavienį asmenį.

Žinant pagrindinius kiekvieno metodo privalumus ir trūkumus, lengviau nuspręsti, kaip atitinkamomis aplinkybėmis parengti tinkamą nuasmeninimo procedūrą.

Siekiant išsiaiškinti kai kuriuos pavojus ir klaidingas nuomones, aptariamas ir duomenų kodavimas pseudonimais, kuris, beje, nėra nuasmeninimo metodas. Tai – tik galimybės duomenų rinkinį susieti su duomenų subjekto pirmine tapatybe sumažinimas, t. y. naudinga saugumo priemonė.

Šioje nuomonėje daroma išvada, kad nuasmeninimo metodais galima užtikrinti privatumą ir parengti veiksmingas nuasmeninimo procedūras, tačiau tik tuo atveju, jeigu šių metodų taikymas tinkamai organizuojamas, t. y., norint pasiekti reikiamą nuasmeninimo lygį ir kartu pateikti tam tikrus naudingus duomenis, turi būti tiksliai nustatytos nuasmeninimo procedūros

prielaidos (aplinkybės) ir tikslas (-ai). Geriausia, jei sprendimai būtų grindžiami remiantis konkrečiais atvejais, galbūt derinant įvairius metodus ir kartu atsižvelgiant į šioje nuomonėje parengtas praktines rekomendacijas.

Pagaliau duomenų valdytojai turėtų atsižvelgti į tai, kad ir nuasmenintas duomenų rinkinys duomenų subjektams vis dar gali kelti liekamąją riziką. Viena vertus, nuasmeninimo ir pakartotinio tapatybės nustatymo srityse išties aktyviai vykdomi moksliniai tyrimai ir reguliariai skelbiama apie naujoves; kita vertus, net ir nuasmeninti duomenys, pvz., statistiniai, gali būti naudojami pavienių asmenų profiliams papildyti, taip sukeliant naujas duomenų apsaugos problemas. Taigi nuasmeninimas neturėtų būti laikomas vienkartine užduotimi, o duomenų valdytojai turėtų reguliariai kaskart iš naujo įvertinti susijusią riziką.

1. Įžanga

Kadangi naudojant įvairius įrenginius, jutiklius ir tinklus sukuriama gausybė duomenų ir atsiranda naujos duomenų rūšys, o duomenų laikymo kaina darosi nereikšminga, visuomenėje stiprėja kartotinio šių duomenų naudojimo interesas ir poreikis. Atvirieji duomenys visuomenei, pavieniams asmenims ir organizacijoms neabejotinai gali būti naudingi, tačiau tik tuo atveju, jeigu bus gerbiamos kiekvieno asmens teisės į asmens duomenų ir privataus gyvenimo apsaugą.

Nuasmeninimas gali būti tinkama naudos išsaugojimo ir rizikos mažinimo strategija. Visiškai nuasmeninus duomenų rinkinį ir panaikinus galimybę nustatyti asmens tapatybę, tokiems duomenims nebetaikomi Europos Sąjungos duomenų apsaugos teisės aktai. Antra vertus, iš konkrečių atvejų tyrimų ir mokslinių straipsnių aiškiai matyti, kad, remiantis gausiu asmens duomenų rinkiniu, parengti visiškai anoniminį duomenų rinkinį ir kartu išsaugoti tiek jame esančios informacijos, kiek reikia užduočiai atlikti, nėra paprasta. Pavyzdžiui, anoniminiu laikomą duomenų rinkinį sujungus su kitu duomenų rinkiniu, gali atsirasti galimybė nustatyti vieno arba daugiau asmenų tapatybę.

Šioje nuomonėje darbo grupė, atsižvelgdama į duomenų apsaugos ES teisinį pagrindą, nagrinėja esamų nuasmeninimo metodų veiksmingumą bei apribojimus ir pateikia rekomendacijas, kaip, rengiant nuasmeninimo procedūrą, apdairiai ir atsakingai taikyti šiuos metodus.

2. Apibrėžtys ir teisinė analizė

2.1. Apibrėžtys atitinkamuose ES teisės aktuose

Kalbant apie nuasmeninimą, Direktyvos 95/46/EB 26 konstatuojamojoje dalyje nurodyta, kad duomenims, kurie paversti anoniminiais, netaikomi duomenų apsaugos teisės aktai:

> *kadangi apsaugos principai turi būti taikomi visai informacijai apie asmenį, kurio tapatybė yra nustatyta arba gali būti nustatyta; kadangi norint nustatyti, ar asmens tapatybė gali būti nustatyta, reikėtų atsižvelgti į visas priemones, kuriomis galėtų pasinaudoti duomenų valdytojas ar bet kuris kitas asmuo minėto asmens tapatybei nustatyti; kadangi apsaugos principai netaikomi duomenims, kurie paversti anoniminiais tokiu būdu, kad duomenų subjekto tapatybė nebegali būti nustatyta; kadangi šiuo tikslu etikos kodeksai, apibrėžti 27 straipsnyje, gali būti naudinga priemonė, nurodant, kaip duomenys galėtų būti paversti anoniminiais ir išlaikyti tokio pavidalo, kad duomenų subjekto tapatybės nebebūtų įmanoma nustatyti;*[1].

Atidžiai skaitant 26 konstatuojamąją dalį, galima susidaryti koncepcinę nuasmeninimo apibrėžtį. 26 konstatuojamojoje dalyje pažymima, jog, norint nuasmeninti kokius nors duomenis, iš jų turi būti pašalinta tiek elementų, kad nebebūtų galima nustatyti duomenų subjekto. Tiksliau sakant, duomenys turi būti tvarkomi taip, kad, taikant „visas priemones,

[1] Be to, reikėtų atkreipti dėmesį, kad tokio pat požiūrio laikomasi ir ES duomenų apsaugos reglamento projekto 23 konstatuojamojoje dalyje, kurioje teigiama, kad „sprendžiant, ar galima nustatyti asmens tapatybę, reikėtų atsižvelgti į visas priemones, kurias asmens tapatybei nustatyti gali naudoti duomenų valdytojas ar bet kuris kitas asmuo“.

kuriomis galėtų pasinaudoti“ duomenų valdytojas arba trečioji šalis asmens tapatybei nustatyti, nebebūtų galima nustatyti fizinio asmens tapatybės. Svarbu tai, kad tvarkymo procesas turi būti nesugrąžinamas. Direktyvoje nepaaiškinta, kaip šis tapatybės duomenų pašalinimo procesas turėtų arba galėtų būti vykdomas[2]. Daugiausia dėmesio joje skiriama rezultatui: duomenys turėtų būti tokie, kad, taikant „visas“ priemones, kuriomis kas nors „galėtų pasinaudoti“, nebūtų įmanoma nustatyti duomenų subjekto tapatybės. Nurodoma, kad galimos nuasmeninimo priemonės turėtų būti nustatytos elgesio kodeksuose, duomenys turėtų būti laikomi tokio pavidalo, kad duomenų subjekto tapatybę „nebebūtų galima“ nustatyti. Taigi direktyvoje neabejotinai keliami labai aukšti reikalavimai.

Direktyvoje dėl privatumo ir elektroninių ryšių (Direktyva 2002/58/EB) sąvokos „nuasmeninimas“ ir „anoniminiai duomenys“ apibūdinamos labai panašiai. 26 konstatuojamojoje dalyje teigiama:

> *Suteikus ryšių rinkodaros paslaugas arba pridėtinės vertės paslaugas, sunaikinami arba padaromi anoniminiais tokioms paslaugoms reikalingi srauto duomenys.*

Atitinkamai 6 straipsnio 1 dalyje nustatyta:

> *Su abonentais ir naudotojais susiję srauto duomenys, kuriuos tvarko ir saugo viešųjų ryšių tinklo ar viešai prieinamų elektroninių ryšių paslaugų teikėjas, turi būti sunaikinti arba pakeisti taip, kad taptų anoniminiai, kai šie duomenys nebėra reikalingi pranešimui perduoti, jeigu nepažeidžiamos šio straipsnio 2, 3 ir 5 dalių ir 15 straipsnio 1 dalies nuostatos.*

Be to, 9 straipsnio 1 dalyje nustatyta:

> *Kai vietos nustatymo duomenys, nesudarantys srauto duomenų ir susiję su viešųjų ryšių tinklų ar viešųjų elektroninių ryšių naudotojais ar abonentais, gali būti tvarkomi, juos galima tvarkyti tik jeigu jie yra pakeisti taip, kad taptų anoniminiai, arba jeigu naudotojai ar abonentai sutinka su tokiu tvarkymu tokia apimtimi ir tiek laiko, kiek yra būtina teikti pridėtinės vertės paslaugai.*

Loginis pagrindas yra toks, kad nuasmeninimo, kaip asmens duomenims taikomo metodo, rezultatas, esant dabartiniam technologijų lygiui, turėtų būti toks pat ilgalaikis kaip ir sunaikinimas, t. y. padarymas, kad nebūtų galima tvarkyti asmens duomenų[3].

2.2. Teisinė analizė

Išanalizavus pagrindiniuose ES duomenų apsaugos teisės aktuose pateikiamas su nuasmeninimu susijusias formuluotes, galima pabrėžti šiuos keturis aspektus:

---

[2] Ši sąvoka išsamiau aprašyta šios nuomonės p. 8.

[3] Čia derėtų prisiminti, kad nuasmeninimas apibrėžtas ir tarptautiniuose standartuose, pvz., ISO 29100: *Process by which personally identifiable information (PII) is irreversibly altered in such a way that a PII principal can no longer be identified directly or indirectly, either by the PII controller alone or in collaboration with any other party* („Procedūra, pagal kurią asmens tapatybės informacija (ATI) nesugrąžinamai pakeičiama taip, kad ATI valdytojas pats vienas arba bendradarbiaudamas su kita šalimi nebegalėtų tiesiogiai arba netiesiogiai nustatyti ATI savininko tapatybės“) (ISO 29100:2011). Atlikto asmens duomenų pakeitimo, susijusio su galimybe tiesiogiai arba netiesiogiai nustatyti asmens tapatybę, negrąžinamumas taip pat labai svarbus ISO aspektas. Šiuo požiūriu esama reikšmingų sąsajų su pagrindiniais Direktyvos 95/46/EB principais ir sąvokomis. Tai pasakytina ir apie apibrėžtis, pateiktas kai kurių šalių (pvz., Italijos, Vokietijos ir Slovėnijos) teisės aktuose, kuriuose daugiausia dėmesio skiriama galimybės nustatyti asmens tapatybę nebuvimui ir vartojama „neproporcingų pastangų“, kuriomis siekiama pakartotinai nustatyti asmens tapatybę, sąvoka (D, SI). Tačiau Prancūzijos duomenų apsaugos įstatyme nustatyta, kad duomenys išlieka asmens duomenimis net ir tuo atveju, kai labai sunku ar beveik neįmanoma pakartotinai nustatyti duomenų subjektą, t. y. šiame įstatyme nėra nuostatos, kurioje būtų minimas „galimumo“ kriterijus.

- nuasmeninimas gali būti asmens duomenų tvarkymo siekiant negrąžinamai panaikinti galimybę nustatyti duomenų subjekto tapatybę rezultatas;

- egzistuoja keletas nuasmeninimo metodų, ES teisės aktuose privalomas jo standartas nenustatytas;

- svarbiais reikėtų laikyti su aplinkybėmis susijusius veiksnius: turi būti atsižvelgiama į „visas“ priemones, kuriomis „galėtų“ pasinaudoti duomenų valdytojas ir kuri nors trečioji šalis asmens tapatybei nustatyti, ypatingą dėmesį skiriant priemonėms, kurios, atsižvelgiant į dabartinį technologijų lygį, pastaruoju metu jau „gali būti“ panaudojamos (dėl padidėjusių skaičiavimo pajėgumų ir prieinamų priemonių);

- nuasmeninimui būdingas rizikos veiksnys: į jį reikėtų atsižvelgti vertinant nuasmeninimo metodikos tinkamumą, be kitų dalykų, atsižvelgiant į galimus tokiu metodu nuasmenintų duomenų panaudojimo būdus; be to, turėtų būti įvertintas tos rizikos dydis ir pasireiškimo tikimybė.

Siekiant pabrėžti pakartotiniam tapatybės nustatymui būdingą liekamąją riziką, susijusią su bet kokia duomenims padaryti anoniminiams skirta technine ir organizacine priemone, šioje nuomonėje pirmenybė teikiama „nuasmeninimo metodo“, o ne „anonimiškumo“ ar „anoniminių duomenų“ sąvokoms.

2.2.1. Nuasmeninimo procedūros teisėtumas

Nuasmeninimas yra asmens duomenims taikomas metodas, kuriuo siekiama panaikinti tapatybės atsekimo galimybę. Todėl pradinė sąlyga yra ta, kad asmens duomenys turėjo būti renkami ir tvarkomi remiantis taikomais teisės aktais dėl duomenų laikymo ir būti tokio pavidalo, kad būtų galima nustatyti asmens tapatybę.

Šiuo požiūriu nuasmeninimo procedūra, kai tokių asmens duomenų tvarkymu siekiama užtikrinti jų nuasmeninimą, yra „tolesnio tvarkymo“ atvejis. Taigi toks tvarkymas turi atitikti suderinamumo kriterijų, kaip nurodyta darbo grupės rekomendacijoje, pateiktoje jos nuomonėje 03/2013 dėl tikslo apribojimo[4].

Iš esmės tai reiškia, kad teisiniu nuasmeninimo pagrindu gali būti bet kuris iš Direktyvos 7 straipsnyje nurodytų motyvų (įskaitant teisėtą duomenų valdytojo interesą), jeigu drauge laikomasi ir 6 straipsnyje pateiktų duomenų kokybės reikalavimų ir deramai atsižvelgiama į konkrečias aplinkybes bei visus darbo grupės nuomonėje dėl tikslo apribojimo nurodytus veiksnius[5].

Kita vertus, reikėtų atkreipti dėmesį ir į Direktyvos 95/46/EB 6 straipsnio 1 dalies e punkto nuostatas (taip pat į Direktyvoje dėl privatumo ir elektroninių ryšių 6 straipsnio 1 dalies ir

[4] Pagal 29 straipsnį įkurtos darbo grupės nuomonė 03/2013 skelbiama šiuo adresu: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

[5] Pirmiausia tai reiškia, kad, atsižvelgiant į visas susijusias aplinkybes, turi būti atliktas esminis vertinimas, ypatingą dėmesį skiriant šiems pagrindiniams veiksniams:
a) duomenų rinkimo tikslų ir jų tolesnio tvarkymo tikslų sąsajai;
b) aplinkybėms, kuriomis buvo renkami asmens duomenys, ir pagrįstiems duomenų subjektų lūkesčiams dėl tolesnio jų naudojimo;
c) asmens duomenų pobūdžiui ir tolesnio jų tvarkymo poveikiui duomenų subjektams;
d) duomenų valdytojo pasirinktoms apsaugos priemonėms, kuriomis siekiama užtikrinti tinkamą tvarkymą ir išvengti nederamo poveikio duomenų subjektams.

9 straipsnio 1 dalies nuostatas); jomis nurodoma, kad asmens duomenys „tokio pavidalo, kad būtų galima nustatyti asmens tapatybę“, turėtų būti laikomi ne ilgiau nei tai yra reikalinga siekiant tikslų, kuriais buvo renkami ar tvarkomi duomenys.

Šia nuostata pabrėžiama, kad asmens duomenys turėtų būti nuasmeninami bent jau standartiniu būdu (laikantis įvairių teisės reikalavimų, pvz., Direktyvoje dėl privatumo ir elektroninių ryšių nurodytų srauto duomenims keliamų reikalavimų). Jeigu duomenų valdytojas pageidauja išlaikyti tokius asmens duomenis ir pasiekus pirminio arba tolesnio tvarkymo tikslus, nuasmeninimo metodai turėtų būti taikomi taip, kad nesugrąžinamai būtų panaikinta tapatybės nustatymo galimybė.

Todėl darbo grupė mano, kad nuasmeninimas, kaip tolesnio asmens duomenų tvarkymo atvejis, gali būti laikomas suderinamu su pirminiais tvarkymo tikslais tik tuo atveju, jeigu nuasmeninimo procedūra suteikia galimybę patikimai parengti nuasmenintą informaciją pagal šio dokumento nuostatas.

Be to, reikėtų pabrėžti, kad nuasmeninimas turėtų būti užtikrinamas laikantis teisinių apribojimų, kuriuos Europos Sąjungos Teisingumo Teismas priminė savo Sprendime *College van burgemeester en wethouders van Rotterdam prieš M.E.E. Rijkeboer* (C-553/07), susijusiame su būtinybe duomenis laikyti tokio pavidalo, kad būtų galima nustatyti asmens tapatybę, pvz., tam, kad duomenų subjektai galėtų pasinaudoti jiems suteiktomis teisėmis susipažinti su informacija. Teisingumo Teismas nusprendė, kad „*Direktyvos [95/46/EB] 12 straipsnio a punktu iš valstybių narių reikalaujama numatyti teisę gauti tiek su dabartim, tiek su praeitim susijusią informaciją apie duomenų gavėjus arba jų kategorijas bei atskleistos informacijos turinį. Valstybės narės turi nustatyti šios informacijos saugojimo trukmę bei atitinkamą teisę gauti ją taip, kad būtų užtikrinta tinkama pusiausvyra tarp, viena vertus, duomenų subjekto intereso apsaugoti savo privatų gyvenimą pasinaudojant, be kita ko, direktyvoje numatytomis įsiterpimo ir reikalavimų iškėlimo priemonėmis ir, kita vertus, naštos, kuri duomenų valdytojui tenka dėl pareigos saugoti šią informaciją.*“

Tai ypač svarbu, jeigu nuasmeninimo klausimu duomenų valdytojas remiasi Direktyvos 95/46/EB 7 straipsnio f punktu: teisėtą duomenų valdytojo interesą visada būtina įvertinti duomenų subjekto teisių ir pagrindinių laisvių požiūriu.

Pavyzdžiui, Nyderlandų duomenų apsaugos institucijos 2012–2013 m. atliktas keturių mobiliojo ryšio operatorių taikomų nuodugnios paketų patikros technologijų tyrimas parodė, kad Direktyvos 95/46/EB 7 straipsnio f punktas yra teisinis pagrindas reikalauti, jog srauto duomenų turinys, surinkus šiuos duomenis, būtų kuo greičiau nuasmeninamas. Išties pabrėžtina, kad Direktyvos dėl privatumo ir elektroninių ryšių 6 straipsnyje nustatyta, jog su abonentais ir naudotojais susiję srauto duomenys, kuriuos tvarko ir saugo viešųjų ryšių tinklo arba viešai prieinamų elektroninių ryšių paslaugų teikėjas, turi būti kuo greičiau sunaikinti arba pakeisti taip, kad taptų anoniminiai. Kadangi šiuo atveju tai leidžiama pagal Direktyvos dėl privatumo ir elektroninių ryšių 6 straipsnį, Direktyvos dėl duomenų apsaugos 7 straipsnis yra tinkamas teisinis pagrindas. Galima pasakyti ir kitaip: jeigu duomenų tvarkymo būdas neleidžiamas pagal Direktyvos dėl privatumo ir elektroninių ryšių 6 straipsnį, Direktyvos dėl duomenų apsaugos 7 straipsnis negali būti teisinis pagrindas.

2.2.2. Galimybė nustatyti asmens tapatybę naudojantis nuasmenintais duomenimis

Nuomonėje 4/2007 dėl asmens duomenų darbo grupė aptarė „asmens duomenų“ sąvoką, daugiausia dėmesio skirdama Direktyvos 95/46/EB 2 straipsnio a punkte pateiktos apibrėžties sudedamosioms dalims, įskaitant žodžius „tapatybė yra nustatyta arba gali būti nustatyta“.

Atsižvelgdama į tai, darbo grupė taip pat padarė tokią išvadą: „Todėl nuasmeninti duomenys yra tokie anoniminiai duomenys, kurie anksčiau buvo susieti su asmeniu, kurio tapatybė galėjo būti nustatyta, bet dabar nebesuteikiantys tokios galimybės.“

Taigi darbo grupė jau išaiškino, kad nustatant, ar nuasmeninimo procedūra yra pakankamai patikima, t. y. ar asmens tapatybės nustatymas tapo „negalimas“, pagal šią direktyvą reikėtų ištirti „priemones, kuriomis [kas nors] galėtų pasinaudoti“. Galimybei nustatyti asmens tapatybę tiesioginės įtakos turi konkrečios nagrinėjamo atvejo aplinkybės ir sąlygos. Šios nuomonės techniniame priede pateikta tinkamiausio metodo pasirinkimo poveikio analizė.

Kaip jau pabrėžta pirmiau, moksliniai tyrimai, priemonės ir skaičiavimo pajėgumai tobulėja. Todėl neįmanoma ir nenaudinga išsamiai surašyti aplinkybes, kuriomis nebegalima nustatyti asmens tapatybės. Vis dėlto derėtų įvertinti kelis veiksnius ir panagrinėti su jais susijusius pavyzdžius.

Pirma, galima teigti, kad duomenų valdytojai daugiausia dėmesio turėtų skirti konkrečioms priemonėms, kurių prireiktų norint atlikti nuasmeninimo metodui priešingą procedūrą, ypač atsižvelgiant į sąnaudas ir praktinę patirtį, reikalingas toms priemonėms įgyvendinti, taip pat įvertinant jų tikėtinumą bei sudėtingumą. Pavyzdžiui, jie turėtų įvertinti nuasmeninimo pastangas ir sąnaudas (būtino laiko ir išteklių požiūriu) atsižvelgdami į vis didesnę nebrangių techninių priemonių, padedančių nustatyti į duomenų rinkinius įtrauktų asmenų tapatybę, pasiūlą, didėjantį kitų duomenų rinkinių (pvz., parengtų pagal atvirųjų duomenų politiką) viešą prieinamumą ir gausius nevisiško nuasmeninimo pavyzdžius, susijusius su atitinkamu neigiamu, o kartais ir nepataisomu poveikiu duomenų subjektams[6]. Reikėtų atkreipti dėmesį į tai, kad asmens tapatybės nustatymo rizika ilgainiui gali didėti, be to, ji priklauso nuo informacinių ir ryšių technologijų pažangos. Todėl teisės aktai, jeigu jie taikomi, turi būti parengti neprimetant konkrečių technologijų ir, pageidautina, atsižvelgiant į informacinių technologijų tobulėjimo galimybių pokyčius[7].

Antra, „priemonės, kurios galėtų būti naudojamos sprendžiant, ar galima nustatyti asmens tapatybę“, yra priemonės, kuriomis gali pasinaudoti „duomenų valdytojas arba kitas asmuo“. Taigi labai svarbu suprasti, kad tuo atveju, jeigu duomenų valdytojas neištrina pirminių duomenų (kuriais remiantis galima nustatyti asmens tapatybę) įvykio lygmeniu ir jeigu duomenų valdytojas perduoda dalį šio duomenų rinkinio (pvz., pašalinęs arba paslėpęs duomenis, kuriais remiantis galima nustatyti asmens tapatybę), iš tokių duomenų sudarytas duomenų rinkinys vis dar laikytinas asmens duomenimis. Tik tuo atveju, kai duomenų valdytojas duomenis agreguoja tokiu lygmeniu, kuriuo nebegalima nustatyti individualių įvykių, iš tokių duomenų sudarytas duomenų rinkinys gali būti laikomas anoniminiu. Pavyzdžiui, jeigu organizacija įvykių lygmeniu renka duomenis apie asmenų judėjimą kelionių metu, įvykių lygmens duomenys apie asmenų keliavimo būdą bet kurios šalies požiūriu vis dar bus laikomi asmens duomenimis, jeigu duomenų valdytojas (arba bet kuri kita šalis) tebeturės galimybę gauti pirminius netvarkytus duomenis, net jeigu iš trečiosioms šalims pateikto rinkinio buvo pašalinti tiesioginiai identifikatoriai. Bet jeigu duomenų valdytojas ištrintų netvarkytus duomenis ir trečiosioms šalims pateiktų tik aukštu lygmeniu

[6] Įdomu tai, kad Europos Parlamentas, neseniai (2013 m. spalio 21 d.) pateikęs Bendrojo duomenų apsaugos reglamento projekto pataisas, 23 konstatuojamojoje dalyje aiškiai nurodė: „Norint įsitikinti, ar priemonės galėtų būti naudojamos asmens tapatybei nustatyti, reikėtų atsižvelgti į visus objektyvius veiksnius, pvz., asmens tapatybei nustatyti reikalingas sąnaudas ir laiką, įvertinant tuo metu esamas duomenų tvarkymo technologijas ir technologijų pažangą.“

[7] Žr. pagal 29 straipsnį įkurtos darbo grupės nuomonę 4/2007, p. 15.

agreguotus statistikos duomenis, pvz., „X kryptimi pirmadieniais važiuoja 160 proc. daugiau keleivių nei antradieniais“, tai būtų laikoma anoniminiais duomenimis.

Taikant veiksmingą nuasmeninimo sprendimą, panaikinama bet kurios šalies galimybė duomenų rinkinyje išskirti konkretų asmenį susiejant du to duomenų rinkinio (arba dviejų atskirų duomenų rinkinių) įrašus ir gauti kokią nors išvestinę informaciją remiantis šiuo duomenų rinkiniu. Todėl apskritai galima teigti, jog norint užtikrinti, kad nebebūtų galima nustatyti duomenų subjekto tapatybės, nepakanka vien tiesiogiai pašalinti identifikavimo elementus. Norint panaikinti galimybę nustatyti asmens tapatybę, dažnai reikės imtis papildomų priemonių, kurios vėlgi priklausys nuo tvarkymo aplinkybių ir tikslų, kuriais numatoma naudoti nuasmenintus duomenis.

PAVYZDYS
Genetinių duomenų profiliai yra vienas iš asmens duomenų pavyzdžių, kai gali kilti asmens tapatybės nustatymo rizika, jeigu dėl unikalaus tam tikrų profilių pobūdžio vienintelis taikomas metodas bus donoro tapatybės panaikinimas. Literatūroje[8] jau įrodyta, kad viešai skelbiamus genetinius išteklius (pvz., genealoginius registrus, nekrologus, paieškos sistemų užklausų rezultatus) sujungus su DNR donorų metaduomenimis (aukojimo laikas, donoro amžius, gyvenamoji vieta), galima nustatyti konkrečių asmenų tapatybę, net jeigu DNR buvo duota neva anonimiškai.

Abi nuasmeninimo metodų grupės – duomenų randomizavimas ir apibendrinimas[9] – turi trūkumų, tačiau tam tikromis aplinkybėmis ir sąlygomis kiekviena iš šių grupių gali būti tinkama pageidaujamam tikslui pasiekti, tuo pat metu nepažeidžiant duomenų subjekto privatumo. Reikėtų aiškiai pasakyti, kad asmens tapatybės nustatymas reiškia ne tik galimybę sužinoti asmens vardą, pavardę ir (arba) adresą, bet ir galimybę nustatyti asmens tapatybę išskyrimo, susiejimo arba išvadų darymo būdu. Be to, duomenų apsaugos teisės aktų taikymo požiūriu nėra svarbu, kokie yra duomenų valdytojo arba gavėjo ketinimai. Jeigu, remiantis duomenimis, galima nustatyti asmens tapatybę, vadinasi, taikytinos duomenų apsaugos taisyklės.

Kai duomenų rinkinį, kuriam buvo pritaikytas koks nors nuasmeninimo metodas (nuasmeninimą atliko ir duomenis paskelbė pirminių duomenų valdytojas) tvarko trečioji šalis, tai ji gali teisėtai daryti neatsižvelgdama į duomenų apsaugos reikalavimus, jeigu ji neturi galimybės (tiesiogiai arba netiesiogiai) nustatyti į pirminį duomenų rinkinį įtrauktų duomenų subjektų tapatybės. Tačiau trečiosios šalys, priimdamos sprendimą, kaip naudoti ir – svarbiausia – derinti tokius nuasmenintus duomenis siekiant savo tikslų, privalo atsižvelgti į pirmiau minėtas aplinkybes ir sąlygas (įskaitant konkrečias pirminių duomenų valdytojo taikomų nuasmeninimo metodų ypatybes), nes dėl susijusių padarinių joms gali būti taikoma skirtinga atsakomybė. Jeigu dėl minėtų veiksnių kyla nepriimtina duomenų subjektų tapatybės nustatymo rizika, tokiu atveju tvarkymui ir vėl bus taikomi duomenų apsaugos teisės aktai.

Pirmiau pateikta informacija tikrai nėra išsami; tai – tik bendros rekomendacijos, kaip įvertinti galimumą nustatyti asmens tapatybę remiantis konkrečiu duomenų rinkiniu, kuriam pagal įvairią esamą metodiką taikomas nuasmeninimas. Visi pirmiau nurodyti veiksniai gali būti laikomi skirtingos rizikos veiksniais, kuriuos duomenų valdytojai turėtų įvertinti nuasmenindami duomenų rinkinius, o trečiosios šalys – naudodamosi nuasmenintais duomenų rinkiniais savo reikmėms.

[8] Žr. John Bohannon, *Genealogy Databases Enable Naming of Anonymous DNA Donors.* Science, Vol. 339, No. 6117 (18 January 2013), p. 262.
[9] Pagrindinės šių dviejų nuasmeninimo metodų ypatybės ir skirtumai aprašyti toliau 3 skyriuje („Techninė analizė“).

2.2.3. Nuasmenintų duomenų naudojimo rizika

Svarstydami nuasmeninimo metodų taikymo galimybes, duomenų valdytojai turi atsižvelgti į šiuos rizikos veiksnius:

- dažnai klaidingai manoma, kad pseudonimais užkoduoti duomenys ir nuasmeninti duomenys yra lygiaverčiai. „Techninės analizės“ skyriuje bus paaiškinta, kad pseudonimais užkoduotų duomenų negalima prilyginti nuasmenintai informacijai, nes, naudojantis tokiais duomenimis, išlieka galimybė išskirti pavienį duomenų subjektą ir jį susieti su įvairiais duomenų rinkiniais. Suteikiant pseudonimą, veikiausiai bus įmanoma nustatyti asmens tapatybę, todėl tokiems duomenims taikoma teisinė duomenų apsaugos sistema. Tai ypač aktualu mokslinių, statistinių arba istorinių tyrimų atveju[10].

PAVYZDYS
Tipinis netinkamo supratimo apie pseudonimų naudojimą atvejis – gerai žinomas AOL (*America On Line*) incidentas. 2006 m. viešai buvo paskelbta duomenų bazė su 20 mln. reikšminių paieškos žodžių, kuriuos trijų mėnesių laikotarpiu pavartojo daugiau kaip 650 000 naudotojų. Vienintelė privatumo apsaugos priemonė joje buvo AOL naudotojo identifikatoriaus pakeitimas skaitiniu požymiu. Tai suteikė galimybę viešai nustatyti kai kurių naudotojų tapatybę ir jų buvimo vietą. Paieškos sistemai pateikiant pseudonimais užkoduotas užklausas, ypač jei jos susiejamos su kitais požymiais, pvz., IP adresais arba kitais kliento konfigūracijos parametrais, atsiranda labai didelė asmens tapatybės nustatymo galimybė.

- kita klaida – manyti, kad jeigu duomenys buvo tinkamai nuasmeninti (buvo įvykdytos visos pirmiau nurodytos sąlygos ir kriterijai, taigi duomenims netaikoma Direktyva dėl duomenų apsaugos), asmenims nebetaikomos jokios apsaugos priemonės; pirmiausia taip manyti klaidinga dėl to, kad šiems duomenims gali būti taikomi kiti teisės aktai. Pavyzdžiui, Direktyvos dėl privatumo ir elektroninių ryšių 5 straipsnio 3 dalyje neleidžiama saugoti informaciją arba suteikti galimybę naudotis galiniame įrenginyje saugoma bet kurios rūšies informacija (įskaitant ne asmens duomenų informaciją) negavus abonento arba naudotojo sutikimo, nes tai yra vienas iš principų, sudarančių platesnį pranešimų konfidencialumo principą;

- trečia klaida susijusi su poveikio, kurį tam tikromis aplinkybėmis pavieniams asmenims gali padaryti tinkamai nuasmeninti duomenys, nepaisymu, ypač profiliavimo atveju. Privatus asmens gyvenimas saugomas pagal Europos žmogaus teisių konvencijos 8 straipsnį ir ES pagrindinių teisių chartijos 7 straipsnį; todėl, net jei tokio tipo duomenims būtų nebetaikomi duomenų apsaugos teisės aktai, dėl trečiosioms šalims pateiktų nuasmenintų duomenų rinkinių naudojimo gali nukentėti privatumas. Jeigu nuasmeninta informacija (dažnai susieta su kitais duomenimis) naudojamasi priimant sprendimus, darančius poveikį (nors ir netiesioginį) pavieniams asmenims, ji turi būti tvarkoma labai apdairiai. Kaip jau nurodyta šioje nuomonėje ir kaip išaiškino darbo grupė, pirmiausia nuomonėje dėl „tikslo apribojimo“ sąvokos (nuomonė 03/2013)[11], teisėti duomenų subjektų lūkesčiai dėl tolesnio jų duomenų tvarkymo turėtų būti įvertinti atsižvelgiant į susijusias aplinkybes, pvz., duomenų subjektų ir duomenų valdytojų tarpusavio ryšio pobūdį, taikomus teisinius įpareigojimus ir tvarkymo operacijų skaidrumą.

[10] Taip pat žr. pagal 29 straipsnį įkurtos darbo grupės nuomonę 4/2007, p. 18–20.
[11] Skelbiama http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf.

3. Techninė analizė, metodų patikimumas ir būdingos klaidos

Yra įvairių praktinių nuasmeninimo būdų ir metodų, jų patikimumas skirtingas. Šiame skyriuje aptarsime pagrindinius aspektus, į kuriuos turėtų atsižvelgti šiuos metodus taikantys duomenų valdytojai. Pirmiausia jie turi įvertinti garantijas, kurias galima suteikti konkrečiu metodu, atsižvelgiant į esamą technologijų lygį ir įvertinant tris labai svarbius nuasmeninimo požiūriu rizikos veiksnius:

- *išskyrimo galimybę*, t. y. galimybę išskirti kai kuriuos arba visus įrašus, pagal kuriuos būtų galima nustatyti į duomenų rinkinį įtraukto asmens tapatybę;
- *susiejimo galimybę,* t. y. galimybę susieti bent du įrašus, susijusius su tuo pačiu duomenų subjektu arba ta pačia duomenų subjektų grupe (toje pačioje duomenų bazėje arba dviejose skirtingose duomenų bazėse). Jeigu išpuolio vykdytojas gali nustatyti (pvz., atlikdamas koreliavimo analizę), kad du įrašai priskirti tai pačiai asmenų grupei, tačiau negali iš tos grupės išskirti pavienių asmenų, tai šiuo metodu apsaugoma nuo išskyrimo, bet neužtikrinama apsauga nuo susiejimo;
- *išvados padarymo galimybę,* t. y. galimybę dedukcijos būdu gana tikėtinai nustatyti požymio vertę remiantis kitų požymių rinkinio vertėmis.

Taigi sprendimu, padedančiu apsisaugoti nuo šių trijų rizikos veiksnių, būtų patikimai panaikinta galimybė iš naujo nustatyti asmens tapatybę labiausiai tikėtinomis duomenų valdytojų arba kurios nors trečiosios šalies pasitelktinomis priemonėmis. Šiuo klausimu darbo grupė pabrėžia, kad asmens tapatybės nustatymo galimybės panaikinimo ir nuasmeninimo metodai yra šiuo metu atliekamų mokslinių tyrimų objektas ir kad tokie tyrimai visuomet parodydavo, jog nėra metodo, kuris neturėtų trūkumų. Plačiąja prasme yra du skirtingi nuasmeninimo būdai: pirmasis grindžiamas ***randomizavimo***, antrasis – ***apibendrinimo*** principu. Šioje nuomonėje aptariami ir kiti principai, pvz., *pseudonimų suteikimo, diferencinio privatumo, l įvairovės, t tankio*.

Toliau paaiškinsime šiame nuomonės skyriuje vartojamas sąvokas. Duomenų rinkinys būna sudarytas iš skirtingų įrašų, susijusių su pavieniais asmenimis (duomenų subjektais). Kiekvienas įrašas susietas su vienu duomenų subjektu ir yra sudarytas iš kiekvienam požymiui (pvz., metai) priskirtų verčių, dar vadinamų įvesties elementais (pvz., 2013). Duomenų rinkinys – tai įrašų rinkinys, kurį taip pat galima pateikti lentelės (arba lentelių rinkinio) ar diagramos su pastabomis ir (arba) svorinėmis vertėmis, kaip tai dabar vis dažniau daroma, pavidalu. Nuomonėje pateikti pavyzdžiai bus susiję su lentelėmis, bet jie tinka ir kitokiems grafinio įrašų vaizdavimo būdams. Požymių, susijusių su duomenų subjektu arba duomenų subjektų grupe, deriniai gali būti vadinami kvaziidentifikatoriais. Kartais duomenų rinkinyje gali būti daug įrašų, susijusių su tuo pačiu asmeniu. Išpuolio vykdytojas – tai trečioji šalis (t. y. ne duomenų valdytojas ir ne duomenų tvarkytojas), netyčia arba tyčia mėginanti gauti pirminius duomenis.

3.1. Randomizavimas

Randomizavimas – tai metodų, kuriais keičiamas duomenų tikrumas siekiant panaikinti aiškią duomenų ir asmens sąsają, grupė. Kai duomenys yra ganėtinai nekonkretūs, jų nebegalima susieti su konkrečiu asmeniu. Taikant randomizavimą, atskirų įrašų savitumas nemažėja, nes kiekvienas įrašas vis viena bus išvedamas pagal atskirą duomenų subjektą, tačiau šiuo būdu užtikrinama apsauga nuo išvestinės informacijos gavimo išpuolių ir (arba) rizikos ir jį, siekiant suteikti didesnę privatumo garantiją, galima derinti su apibendrinimu. Norint

užtikrinti, kad, remiantis įrašu, nebūtų galima nustatyti pavienio asmens tapatybės, gali prireikti papildomų metodų.

3.1.1. Iškraipytų duomenų įterpimas

Iškraipytų duomenų įterpimo metodas pirmiausia naudingas tada, kai požymiai gali turėti reikšmingą neigiamą poveikį asmenims. Šio metodo esmė – į duomenų rinkinį įtrauktų požymių pakeitimas sumažinant jų tikslumą, tačiau išsaugant bendrą pasiskirstymą. Tvarkydamas duomenų rinkinį, stebėtojas manys, kad vertės yra tikslios, bet tai bus teisinga tik iš dalies. Pavyzdžiui, jeigu asmens ūgis iš pradžių buvo išmatuotas centimetrų tikslumu, nuasmenintame duomenų rinkinyje ūgis gali būti nurodomas tik ±10 cm tikslumu. Jeigu šis metodas bus taikomas veiksmingai, trečioji šalis negalės nustatyti asmens tapatybės ir neturėtų galėti ištaisyti duomenis arba kaip nors kitaip nustatyti, kaip duomenys buvo pakeisti.

Iškraipytų duomenų įterpimą paprastai reikia derinti su kitais nuasmeninimo metodais, pvz., su akivaizdžių požymių ir kvaziindentifikatorių pašalinimu. Iškraipymo laipsnis turėtų priklausyti nuo to, kokio lygio informacija yra reikalinga, ir nuo apsaugotų požymių atskleidimo daromo poveikio asmenų privatumui.

3.1.1.1. Garantijos

- *Išskyrimo galimybė:* galimybė išskirti vieno asmens įrašą išlieka (galbūt nenustatant jo tapatybės), tačiau įrašų patikimumas bus mažesnis.
- *Susiejimo galimybė:* galimybė susieti to paties asmens įrašus išlieka, tačiau įrašų patikimumas bus mažesnis, taigi gali būti, kad tikras įrašas bus susietas su dirbtinai įterptu įrašu (t. y. su iškraipytais duomenimis). Kartais dėl neteisingo jo priskyrimo duomenų subjektui gali kilti didelė ar netgi didesnė rizika, nei tai būtų teisingo priskyrimo atveju.
- *Išvados padarymo galimybė:* išvestinių duomenų gavimo išpuoliai yra galimi, tačiau sėkmės tikimybė bus mažesnė, gali būti padarytos klaidingos teigiamos (arba neigiamos) išvados.

3.1.1.2. Dažnos klaidos

- Netinkamų iškraipytų duomenų įterpimas: jeigu iškraipyti duomenys prasmės požiūriu nėra įtikinami (t. y. iškraipymas pernelyg didelis ir neatitinka rinkiniui būdingų požymių logikos), išpuolio vykdytojas, turintis galimybę prisijungti prie duomenų bazės, galės atrinkti iškraipytus duomenis, o kartais – atkurti trūkstamus įvesties elementus. Be to, jeigu duomenų rinkiniui būdingas didelis retumas[12], gali išlikti galimybė iškraipytus duomenų elementus susieti su išorės šaltiniu.
- Manymas, kad iškraipytų duomenų įterpimas yra pakankama priemonė: iškraipytų duomenų įterpimas yra papildoma priemonė, apsunkinanti išpuolio vykdytojo pastangas gauti asmens duomenis. Jeigu iškraipytų duomenų kiekis nėra didesnis už duomenų rinkinyje esančios informacijos kiekį, nederėtų manyti, kad iškraipytų duomenų įterpimas yra vienintelis taikytinas nuasmeninimo sprendimas.

[12] Ši sąvoka išsamiau aprašyta priede, žr. p. 30.

3.1.1.3. Netinkamas iškraipytų duomenų įterpimo metodo taikymas

Naudojantis vaizdo įrašų paslaugų teikėjo *Netflix* naudotojų duomenų baze, buvo atliktas gerai žinomas pakartotinio tapatybės nustatymo eksperimentas. Tyrinėtojai išanalizavo šios duomenų bazės, apimančios daugiau kaip 100 mln. įvertinimų, kuriuos beveik 500 000 naudotojų 1–5 balų skalėje skyrė daugiau kaip 18 000 filmų, geometrines savybes; šiuos duomenis įmonė paskelbė pirmiau juos nuasmeninusi pagal vidaus privatumo politiką – pašalinusi visą klientų tapatybės informaciją, palikdama tik įvertinimus ir datas. Buvo įterpti iškraipyti duomenys – įvertinimai šiek tiek padidinti arba sumažinti.

Nepaisant to, nustatyta, kad atrankos kriterijumi pasirinkus aštuonis įvertinimus ir datas su 14 dienų paklaida, būtų galima nustatyti 99 proc. naudotojų tapatybę, o, supaprastinus atrankos kriterijų (du įvertinimai ir 3 dienų paklaida), vis dar būtų galima nustatyti 68 proc. naudotojų tapatybę[13].

3.1.2. Perstatymas

Taikant šį metodą, lentelėje esančių požymių vertės sukeičiamos vietomis taip, kad kai kurios iš jų būtų dirbtinai susietos su kitais duomenų subjektais. Tai naudinga, kai svarbu išsaugoti tikslų kiekvieno į duomenų rinkinį įtraukto požymio pasiskirstymą.

Perstatymas gali būti laikomas savita iškraipytų duomenų įterpimo rūšimi. Pagal klasikinį iškraipytų duomenų įterpimo metodą požymiai pakeičiami pasirenkant atsitiktines vertes. Dėsningai iškraipytų duomenų įterpimas gali būti sunkiai įvykdomas uždavinys, o nedaug pakeičiant požymių vertes gali būti neužtikrintas pakankamas privatumas. Perstatymo metodas – tai alternatyva, kurią taikant duomenų rinkinio vertės pakeičiamos tiesiog sumaišant vietomis skirtingų įrašų vertes. Tokiu sukeitimu užtikrinama, kad verčių intervalas ir paskirstymas išliktų tokie patys, o verčių ir asmenų koreliacijos pasikeistų. Jeigu dviem arba daugiau požymių būdingas loginis tarpusavio ryšys arba statistinė koreliacija ir atliekamas nepriklausomas jų perstatymas, toks ryšys sunaikinamas. Todėl gali būti svarbu susijusių požymių rinkinio perstatymą atlikti taip, kad nebūtų pažeistas loginis tarpusavio ryšys, nes kitaip išpuolio vykdytojas galėtų nustatyti sukeistus požymius ir atlikti atvirkštinį perstatymą.

Tarkime, medicinos duomenų rinkinyje yra požymių poaibis „hospitalizavimo priežastys, simptomai, atsakingas skyrius“; dažniausiai tarp šių verčių bus stiprus loginis ryšys, todėl, atlikus tik vienos iš šių verčių perstatymą, ją bus galima nustatyti ir gal netgi atlikti atvirkštinį perstatymą.

Panašiai kaip ir iškraipytų duomenų įterpimo atveju, vien perstatymo pritaikymas gali neužtikrinti nuasmeninimo, todėl jis visada turėtų būti derinamas su akivaizdžių požymių ir (arba) kvaziidentifikatorių pašalinimu.

3.1.2.1. Garantijos

- *Išskyrimo galimybė:* kaip ir iškraipytų duomenų įterpimo atveju, galimybė išskirti su konkrečiu asmeniu susijusius įrašus išlieka, tačiau šių įrašų patikimumas bus mažesnis.

[13] Narayanan, A., Shmatikov, V. (2008, May). *Robust de-anonymization of large sparse datasets*. In: *2008 IEEE Symposium on Security and Privacy (SP 2008),* (p. 111–125).

- *Susiejimo galimybė:* perstatymas, darydamas įtaką požymiams ir kvaziindikatoriams, gali trukdyti teisingai susieti požymius tame pačiame duomenų rinkinyje ir su kitais duomenų rinkiniais, tačiau vis viena išlieka galimybė atlikti neteisingą susiejimo operaciją, nes tikrasis įrašas gali būti susietas su kitu duomenų subjektu.
- *Išvados padarymo galimybė:* išvestinių duomenų gavimo, remiantis duomenų rinkiniu, galimybė išlieka, ypač jei požymiai yra koreliaciniai arba susiję stipriais loginiais tarpusavio ryšiais; antra vertus, išpuolio vykdytojas, nežinodamas, kurių požymių perstatymas buvo atliktas, turi turėti omenyje, kad toks išvestinių duomenų gavimas yra pagrįstas klaidinga hipoteze, todėl išlieka tik tikimybinė galimybė.

3.1.2.2. Dažnos klaidos

- Netinkamo požymio pasirinkimas: neslaptų arba nerizikingų požymių perstatymas nebus labai naudinga asmens duomenų apsaugos požiūriu. Išties, jeigu slapti ir (arba) rizikingi požymiai išliktų susieti su pirminiu požymiu, išpuolio vykdytojas vis dar galėtų sužinoti slaptą informaciją apie asmenis.
- Atsitiktinis požymių perstatymas: jeigu tarp dviejų požymių yra tvirtas koreliacinis ryšys, atsitiktinis požymių perstatymas nesuteiks didelių garantijų. Ši dažnai daroma klaida pavaizduota 1 lentelėje.
- Manymas, kad perstatymas yra pakankama priemonė: panašiai kaip ir iškraipytų duomenų įterpimo atveju, vien perstatymo atlikimas neužtikrina anonimiškumo, todėl perstatymas visada turėtų būti derinamas su kitais metodais, pvz., akivaizdžių požymių pašalinimu.

3.1.2.3. Netinkamas perstatymo metodo taikymas

Iš toliau pateikiamo pavyzdžio matyti, kad atsitiktinis požymių perstatymas suteikia menkų privatumo garantijų, kai įvairūs požymiai yra susieti loginiais ryšiais. Remiantis tariamai nuasmenintais duomenimis, nesunku nustatyti kiekvieno asmens pajamas, atsižvelgiant į pareigas (ir gimimo metus). Pavyzdžiui, tiesiog peržiūrėjus duomenis, galima teigti, jog labai tikėtina, kad lentelėje nurodytas generalinis direktorius yra gimęs 1957 m. ir jo atlygis yra didžiausias, o bedarbis gimęs 1964 m. ir jo pajamos yra mažiausios.

Gimimo metai	Lytis	Pareigos	Pajamos (atliktas perstatymas)
1957	V	Inžinierius	70 000
1957	V	Generalinis direktorius	5 000
1957	V	Bedarbis	43 000
1964	V	Inžinierius	100 000
1964	V	Vadovas	45 000

1 lentelė. Neveiksmingo tarpusavyje koreliuojančių požymių nuasmeninimo perstatymo būdu pavyzdys

3.1.3. Diferencinis privatumas

Diferencinis privatumas[14] priskiriamas randomizavimo metodų grupei, tačiau jis pagrįstas kitokiu principu: iškraipytų duomenų įterpimas taikytinas prieš paskelbiant duomenų rinkinį,

[14] Dwork, C. (2006). *Differential privacy*. In: *Automata, languages and programming* (p. 1–12). Springer, Berlin Heidelberg.

o diferencinio privatumo metodas gali būti taikomas, kai duomenų valdytojas parengia nuasmenintus duomenų rodinius, išsaugodamas pirminių duomenų kopiją. Tokie nuasmeninti rodiniai paprastai parengiami naudojant užklausų poaibį, skirtą tam tikrai trečiajai šaliai. Į šį poaibį vėliau sąmoningai įtraukiami atsitiktiniai iškraipyti duomenys. Taikydamas diferencinio privatumo metodą, duomenų valdytojas sužino, kiek iškraipytų duomenų jis turėtų įterpti ir kokiu pavidalu, kad užtikrintų reikiamas privatumo garantijas[15]. Šiuo atveju labai svarbu nuolat stebėti (ne rečiau kaip kiekvienos naujos užklausos atveju), ar neatsirado galimybė nustatyti asmens tapatybę pasinaudojant užklausos rezultatų aibe. Be to, derėtų paaiškinti, kad diferencinio privatumo metodu pirminiai duomenys nepakeičiami, o kol jie išlieka, duomenų valdytojas, atsižvelgdamas į visas galimas pasitelktinas priemones, asmens tapatybę gali nustatyti pasinaudodamas diferencinio privatumo užklausų rezultatais. Šie rezultatai taip pat turėtų būti laikomi asmens duomenimis.

Vienas iš diferenciniu privatumu pagrįsto metodo privalumų yra tas, kad duomenų rinkiniai įgaliotosioms trečiosioms šalims teikiami pagal konkrečias užklausas, o ne paskelbiant visą duomenų rinkinį. Kad būtų lengviau atlikti auditą, duomenų valdytojas gali išsaugoti visų užklausų ir prašymų sąrašą, taip užtikrindamas, kad trečiosios šalys negautų duomenų, su kuriais jos neturi teisės susipažinti. Be to, norint geriau apsaugoti privatumą, užklausai gali būti taikomi nuasmeninimo, pvz., iškraipytų duomenų įterpimo arba pakeitimo, metodai. Tyrinėtojams dar nepavyko sukurti gero interaktyvaus užklausų ir jų rezultatų teikimo mechanizmo, kurį naudojant būtų galima ir gana tiksliai (t. y. kuo mažiau iškraipant duomenis) atsakyti į visas užklausas, ir apsaugoti privatumą.

Siekiant apriboti išvados padarymo ir susiejimo išpuolių galimybę, būtina sekti subjektų teikiamas užklausas ir stebėti apie duomenų subjektus gautą informaciją; todėl diferencinio privatumo metodu valdomos duomenų bazės neturėtų būti prieinamos viešoms paieškos sistemoms, kuriose nėra užklausas teikiančių subjektų sekimo galimybės.

3.1.3.1. Garantijos

- *Išskyrimo galimybė:* jeigu vieninteliai statistiniai duomenys yra užklausos rezultatai ir jeigu tinkamai parenkamos rinkiniui taikytinos taisyklės, neturėtų būti įmanoma atsakymus panaudoti asmens išskyrimo tikslais.
- *Susiejimo galimybė:* pasitelkus daug užklausų, gali būti įmanoma susieti dviejų atsakymų elementus, susijusius su konkrečiu asmeniu.
- *Išvados padarymo galimybė:* remiantis daugybės užklausų informacija, galima gauti išvestinę informaciją apie asmenis arba jų grupes.

3.1.3.2. Dažnos klaidos

- Nepakankamo iškraipytų duomenų kiekio įterpimas: norint išvengti susiejimo su bendrosiomis žiniomis, būtina pasistengti pateikti kuo mažiau įrodymų, ar konkretus duomenų subjektas arba jų grupė yra įtraukti į duomenų rinkinį. Duomenų apsaugos požiūriu sunkiausia užduotis – parengti pakankamą iškraipytų duomenų kiekį, pridėtiną prie teisingų atsakymų, kad būtų apsaugotas asmens privatumas kartu nesumažinant teikiamų atsakymų naudingumo.

3.1.3.3. Netinkamas diferencinio privatumo metodo taikymas

[15] Plg. Felten, E. (2012) *Protecting privacy by adding noise*. Skelbiama adresu https://techatftc.wordpress.com/2012/06/21/protecting-privacy-by-adding-noise/.

Atskiras kiekvienos užklausos vertinimas: derinant užklausų rezultatus, gali būti įmanoma išsiaiškinti informaciją, kuri turėjo būti įslaptinta. Jeigu nebūtų saugoma užklausų istorija, išpuolio vykdytojas diferencinio privatumo metodu valdomai duomenų bazei galėtų parengti daug užklausų, pagal kurias gaunamos imties dydis būtų nuosekliai mažinamas tol, kol determinavimo būdu arba su gana didele tikimybe išryškėtų vieno duomenų subjekto arba jų grupės konkretus pobūdis. Be to, reikėtų vengti klaidingai manyti, kad duomenys prieigą turinčiai trečiajai šaliai yra anoniminiai, jeigu duomenų valdytojas, atsižvelgdamas į visas galimas pasitelktinas priemones ir naudodamasis pirmine duomenų baze, vis dar gali nustatyti duomenų subjekto tapatybę.

3.2. Apibendrinimas

Apibendrinimas yra antroji nuasmeninimo metodų grupė. Pagal šį principą duomenų subjektų požymiai apibendrinami arba, kitaip tariant, susilpninami, kiek pakeičiant atitinkamą mastelį arba dydžio eilę (pvz., informaciją pateikiant ne miesto, o regiono mastu, mėnesio, o ne savaitės apimtimi). Nors apibendrinimas, siekiant panaikinti išskyrimo galimybę, ir gali būti veiksmingas, ne visais atvejais šiuo principu užtikrinamas tinkamas nuasmeninimas; pirmiausia, taikant šį principą, būtina pasitelkti specialius sudėtingus kiekybinius metodus, kuriais būtų panaikinta susiejimo ir išvados padarymo galimybė.

3.2.1. Agregavimas ir *k* anonimiškumas

Agregavimo ir *k* anonimiškumo metodais siekiama panaikinti galimybę išskirti duomenų subjektus, juos grupuojant kartu su ne mažiau kaip *k* kitų asmenų. Šiuo tikslu požymių vertės apibendrinamos tokiu mastu, kad kiekvienam asmeniui būtų priskirta tokia pat vertė. Pavyzdžiui, vietovės mastelį pastambinus nuo miesto iki šalies, bus įtraukta daugiau duomenų subjektų. Pavienių asmenų gimimo datos gali būti apibendrintos datų intervalais arba sugrupuotos pagal mėnesius arba metus. Kitus skaitinius požymius (pvz., darbo užmokestį, svorį, ūgį, vaisto dozę) galima apibendrinti verčių intervalais (pvz., darbo užmokestis nuo 20 000 iki 30 000 EUR). Šie metodai gali būti taikomi tada, kai dėl požymių tikslių verčių koreliacijos gali susidaryti kvaziindikatoriai.

3.2.1.1. Garantijos

- *Išskyrimo galimybė:* kadangi vienodi požymiai dabar būdingi *k* naudotojų, turėtų būti nebeįmanoma iš *k* naudotojų grupės išskirti vieną asmenį.
- *Susiejimo galimybė:* nors susiejimo galimybė ir nėra didelė, išlieka galimybė susieti įrašus pagal *k* naudotojų grupes. Tikimybė šioje grupėje, kad du įrašai atitinka tuos pačius pseudoindentifikatorius, lygi 1/*k* (ši tikimybė gali būti gerokai didesnė už tikimybę, kad šie įvesties elementai nesusiejami).
- *Išvados padarymo galimybė:* pagrindinis *k* anonimiškumo metodo trūkumas yra tas, kad juo nepanaikinama kokio nors išvestinių duomenų gavimo išpuolio galimybė. Išties, jeigu visi *k* asmenų priklauso tai pačiai grupei, tuomet, jeigu žinoma, kuriai grupei asmuo priklauso, nesunku gauti šios savybės vertę.

3.2.1.2. Dažnos klaidos

- Kvaziindentifikatorių trūkumas: labai svarbus *k* anonimiškumo parametras yra *k* riba. Kuo didesnė *k* vertė, tuo tvirtesnės privatumo garantijos. Dažnai daroma klaida – dirbtinis *k* vertės padidinimas sumažinant nagrinėjamą kvaziidentifikatorių aibę. Sumažinus kvaziidentifikatorių skaičių, dėl kitų požymių suteikiamos asmens

tapatybės nustatymo galimybės (ypač jeigu kai kurie iš šių požymių yra slapti arba jiems būdinga labai didelė entropija, kaip tai būna kai kurių retų požymių atveju) tampa lengviau sudaryti *k* naudotojų grupes. Daroma didelė klaida, jeigu, pasirenkant požymį, kuris bus apibendrinamas, atsižvelgiama ne į visus kvaziindentifikatorius; jeigu pagal kuriuos nors požymius *k* asmenų grupėje galima išskirti pavienį asmenį, vadinasi, apibendrinimu neužtikrinta kai kurių asmenų apsauga (žr. 2 lentelėje pateiktą pavyzdį).

- Maža *k* vertė: pasirinkus mažą *k* vertę, taip pat kyla problemų. Kai *k* vertė yra per maža, kiekvieno grupei priklausančio asmens svorinė vertė yra pernelyg didelė, todėl išpuoliai siekiant išvestinių duomenų bus sėkmingesni. Pavyzdžiui, jeigu $k = 2$, tuomet tikimybė, kad dviem asmenims būdinga ta pati savybė, bus didesnė nei tuo atveju, kai $k > 10$.
- Nevienodos svorinės vertės asmenų grupavimas: grupuojant asmenis, kurių požymių pasiskirstymas yra nevienodas, taip pat gali kilti problemų. Asmens įrašo poveikis duomenų rinkiniui bus nevienodas: kai kurie asmenys reprezentuos didelę įvesties elementų dalį, kitų poveikis išliks gana mažas. Todėl svarbu užtikrinti, jog *k* vertė būtų pakankamai didelė, kad su jokiais asmenimis susiję elementai grupėje nesudarytų pernelyg reikšmingos dalies.

3.1.3.3. Netinkamas *k* anonimiškumo metodo taikymas

Pagrindinis *k* anonimiškumo metodo trūkumas yra tas, kad šiuo metodu nepanaikinama išpuolių siekiant gauti išvestinių duomenų galimybė. Iš toliau pateikto pavyzdžio matyti, kad tuo atveju, kai išpuolio vykdytojas žino, jog tam tikras į duomenų rinkinį įtrauktas asmuo yra gimęs 1964 m., jis žinos ir tai, kad asmenį buvo ištikęs širdies smūgis. Be to, žinodami, jog šis duomenų rinkinys gautas iš Prancūzijos organizacijos, galėsime daryti išvadą, kad kiekvienas asmuo yra Paryžiaus gyventojas, nes pirmieji trys Paryžiaus pašto kodų skaitmenys yra 750*.

Gimimo metai	Lytis	Pašto kodas	Diagnozė
1957	V	750*	Širdies smūgis
1957	V	750*	Cholesterolis
1957	V	750*	Cholesterolis
1964	V	750*	Širdies smūgis
1964	V	750*	Širdies smūgis

2 lentelė. Netinkamai parengto nuasmeninimo taikant *k* anonimiškumo metodą pavyzdys

3.2.2. *l* įvairovė ir *t tankis*

l įvairovės metodu išplečiamas *k* anonimiškumo metodas, siekiant užtikrinti, kad nebebūtų galima rengti determinavimo būdu pagrįstų išpuolių, pasirūpinant, kad kiekvienoje lygiavertiškumo klasėje kiekvienam požymiui būtų priskirta ne mažiau kaip *l* skirtingų verčių.

Vienas iš pagrindinių siektinų tikslų – riboti lygiavertiškumo klasių, kurioms būtų būdingas menkas požymių kintamumas, susidarymą, kad bendrųjų žinių apie tam tikrą duomenų subjektą turinčiam išpuolio vykdytojui visada liktų didelių abejonių dėl savo išvadų.

l įvairovės metodas naudingas norint apsaugoti duomenis nuo išpuolių siekiant gauti išvestinių duomenų, kai požymių vertės yra gerai pasiskirsčiusios. Tačiau reikia pabrėžti, kad šiuo metodu negalima panaikinti informacijos nutekinimo galimybės, jeigu požymiai skaidinyje pasiskirstę netolygiai arba priklauso mažam verčių arba reikšminių verčių intervalui. Todėl *l* įvairovės metodas neapsaugo nuo tikimybinio išvadų darymo išpuolių.

t tankio metodas yra patobulintas *l* įvairovės metodas, nes juo siekiama sudaryti lygiavertiškumo klases, kurioms būtų būdingas panašus į pirminį požymių pasiskirstymas lentelėje. Šis metodas naudingas tada, kai svarbu, kad duomenys būtų kuo panašesni į pirminius; todėl lygiavertiškumo klasei taikomas papildomas apribojimas, pagal kurį kiekvienoje lygiavertiškumo klasėje turėtų būti ne tik mažiau kaip *l* skirtingų verčių, bet ir kiekviena vertė turi būti pateikta tiek kartų, kiek reikalinga tam, kad būtų atkurtas pirminis kiekvieno požymio pasiskirstymas.

3.2.2.1. Garantijos

- *Išskyrimo galimybė:* kaip ir *k* anonimiškumo atveju, *l* įvairovės ir *t* tankio metodais užtikrinama, kad duomenų bazėje nebūtų galima išskirti su pavieniu asmeniu susijusių įrašų.
- *Susiejimo galimybė: l* įvairovės ir *t* tankio metodai nėra tobulesni už *k* anonimiškumo metodą susiejimo galimybės nebuvimo požiūriu. Problema tokia pat, kaip ir bet kurios grupės atveju: tikimybė, kad vienodi įrašai priklauso tam pačiam duomenų subjektui yra didesnė už santykį $1/N$ (čia N – duomenų subjektų skaičius duomenų bazėje).
- *Išvados padarymo galimybė:* pagrindinis *l* įvairovės ir *t* tankio metodų privalumas *k* anonimiškumo metodo atžvilgiu yra tas, kad, naudojantis duomenų bazėmis, kurioms buvo pritaikyti šie metodai, nebegalima parengti išvestinių duomenų gavimo išpuolių, kurie duotų 100 proc. patikimus rezultatus.

3.2.2.2. Dažnos klaidos

- Įslaptinto požymio verčių apsauga šias vertes sumaišant su kitais įslaptintais požymiais: norint suteikti privatumo garantiją, nepakanka grupėje turėti dvi požymio vertes. Įslaptintų verčių pasiskirstymas kiekvienoje grupėje turėtų atitikti tų verčių pasiskirstymą visoje aibėje arba bent jau turėtų būti vienodas visoje grupėje.

3.2.2.3. Netinkamas *l* įvairovės metodo taikymas

Toliau pateiktoje lentelėje *l* įvairovės metodas pritaikytas požymiui „diagnozė“; vis dėlto žinant, kad į lentelę yra įtrauktas 1964 m. gimęs asmuo, vis dar galima, esant labai didelei tikimybei, teigti, kad jį buvo ištikęs širdies smūgis.

Gimimo metai	Lytis	Pašto kodas	Diagnozė
1957	V	750*	**Širdies smūgis**
1957	V	750*	**Cholesterolis**
1957	V	750*	**Cholesterolis**
1957	V	750*	**Cholesterolis**
1964	V	750*	**Širdies smūgis**
1964	V	750*	**Širdies smūgis**
1964	V	750*	**Širdies smūgis**
1964	V	750*	**Cholesterolis**
1964	V	750*	**Širdies smūgis**
1964	V	750*	**Širdies smūgis**
1964	V	750*	**Širdies smūgis**
1964	V	750*	**Širdies smūgis**
1964	V	750*	**Širdies smūgis**
1964	V	750*	**Širdies smūgis**
1964	V	750*	**Širdies smūgis**

3 lentelė. *I* įvairovės metodu sudaryta lentelė, kurioje požymio „diagnozė“ vertės pasiskirsčiusios netolygiai

Pavardė	Gimimo metai	Lytis
Smith	1964	V
Rossi	1964	V
Dupont	1964	V
Jansen	1964	V
Garcia	1964	V

4 lentelė. Žinodamas, kad šie asmenys įtraukti į 3 lentelę, išpuolio vykdytojas galėtų numanyti, kad juos buvo ištikęs širdies smūgis

4. Pseudonimų suteikimas

Pseudonimų suteikimas – tai metodas, pagal kurį vienas požymis (paprastai – unikalus) įraše pakeičiamas kitu. Todėl išlieka galimybė netiesiogiai nustatyti fizinio asmens tapatybę; taigi vien pseudonimų suteikimas neužtikrina duomenų rinkinio anonimiškumo. Vis dėlto šis metodas vis tiek aptariamas šioje nuomonėje, nes su jo taikymu susiję dažni nesusipratimai ir klaidos.

Taikant pseudonimų suteikimo metodą, sumažinama galimybė duomenų rinkinį susieti su pirmine duomenų subjekto tapatybe; taigi šis metodas yra naudinga saugumo priemonė, bet tai nėra nuasmeninimo metodas.

Pseudonimų suteikimo rezultatas gali nepriklausyti nuo pirminės vertės (pvz., jeigu tai atsitiktinis duomenų valdytojo sugeneruotas skaičius arba duomenų subjekto pasirinkta pavardė) arba gali būti sukuriamas naudojantis požymio arba jų grupės pirminėmis vertėmis, pvz., taikant maišos funkciją arba šifravimo sistemą.

Toliau aprašyti dažniausiai taikomi pseudonimų suteikimo metodai.

- Šifravimas naudojant slaptą raktą: šiuo atveju raktą turintis asmuo gali nesunkiai atkurti kiekvieno duomenų subjekto tapatybę dešifravęs duomenų rinkinį, nes asmens duomenys, nors ir užšifruoti, tebėra duomenų rinkinyje. Jeigu buvo pritaikyta pažangi šifravimo sistema, dešifravimas galimas tik žinant raktą.
- Maišos funkcija: tai – funkcija, kuri iš bet kokio dydžio įvesties duomenų (tai gali būti vienas požymis arba požymių rinkinys) parengia nustatyto dydžio išvesties duomenis ir kurios negalima atlikti priešinga kryptimi; tai reiškia, kad nebelieka pakartotinio tapatybės nustatymo rizikos, būdingos šifravimui. Tačiau, jeigu yra žinomas maišos funkcijos įvesties verčių intervalas, šioms vertėms galima pakartotinai pritaikyti maišos funkciją ir taip gauti teisingą tam tikro įrašo vertę. Pavyzdžiui, jeigu duomenų rinkiniui buvo pritaikytas pseudonimų suteikimo metodas, pagrįstas nacionalinių asmens tapatybės kodų maišos funkcija, tuomet šiuos kodus galima nustatyti tiesiog pritaikant maišos funkciją visoms galimoms įvesties vertėms ir rezultatą palyginant su duomenų rinkinyje esančiomis vertėmis. Maišos funkcijos paprastai yra skirtos palyginti greitam skaičiavimui ir nėra atsparios jėgos metodo išpuoliams[16]. Kad būtų galima masiškai atkurti didelį verčių, kurioms buvo pritaikyta maišos funkcija, rinkinį, taip pat gali būti parengiamos iš anksto apskaičiuotų verčių lentelės.

  Taikant „druskos“ naudojimu pagrįstą maišos funkciją (prie požymio, kuriam taikoma maišos funkcija, pridedama atsitiktinė vertė, vadinama „druska“), galima sumažinti įvesties vertės nustatymo tikimybę, tačiau pagrįstomis priemonėmis vis vien gali būti įmanoma apskaičiuoti pirminę požymio vertę, paslėptą sudėtingesnės maišos funkcijos (su „druskos“ elementu) rezultatu[17].
- Saugomo rakto naudojimu pagrįsta maišos funkcija: tai tam tikra maišos funkcija, kai naudojamas papildomas įvesties elementas – slaptas raktas (ši funkcija nuo „druskos“ naudojimu pagrįstos funkcijos skiriasi tuo, kad „druska“ paprastai nėra slaptas elementas). Duomenų valdytojas, naudodamas slaptą raktą, šią funkciją gali

[16] Tokiais išpuoliais peržiūrimos visos tikėtinos įvesties vertės, taip siekiant sudaryti atitikties lenteles.

[17] Pirmiausia tai pasakytina apie atvejį, kai žinomas požymio pobūdis (vardas, socialinio draudimo numeris, gimimo data ir t. t.). Norint papildyti skaičiavimo sąlygą, galima remtis rakto nustatymo maišos funkcija, pagal kurią apskaičiuotajai vertei maišos funkcija taikoma kelis kartus, naudojant nedidelį „druskos“ kiekį.

pakartotinai pritaikyti požymiui, tačiau išpuolio vykdytojui tampa gerokai sunkiau pakartoti šią funkciją nežinant rakto, nes mėgintinų variantų skaičius yra per didelis, kad būtų įmanoma tai padaryti.

- Determinavimu pagrįstas šifravimas arba panaikinamo rakto naudojimu pagrįsta maišos funkcija: pagal šį metodą kiekvienam duomenų rinkinyje esančiam požymiui kaip pseudonimas gali būti parenkamas atsitiktinis skaičius, o tada panaikinama atitikties lentelė. Pasitelkus tokį sprendimą, galima[18] sumažinti galimybę duomenų rinkinyje esančius asmens duomenis susieti su kitame duomenų rinkinyje, kuriame naudojamas kitoks pseudonimas, esančiais duomenimis apie tą patį asmenį. Skaičiavimo požiūriu išpuolio vykdytojui, pasitelkusiam net ir pažangų algoritmą, būtų sunku iššifruoti arba pakartoti funkciją, nes, neturint rakto, reikėtų išmėginti kiekvieną galimą raktą.
- Pakaitinių simbolių naudojimas: šis metodas paprastai taikomas (nors gali būti taikomas ir kitur) finansų sektoriuje, siekiant kortelių atpažinimo numerius (angl. ID) pakeisti vertėmis, kurios išpuolio vykdytojui būtų ne tokios naudingos. Šis metodas sukurtas remiantis pirmiau aptartais metodais ir paprastai grindžiamas vienakrypčių šifravimo priemonių taikymu arba eilės numerio ar atsitiktine tvarka sugeneruoto numerio, kuris nėra matematiškai gaunamas iš pirminių duomenų, priskyrimu pasitelkiant indeksavimo funkciją.

4.1. Garantijos

- *Išskyrimo galimybė:* galimybė išskirti pavienio asmens įrašus išlieka, nes asmens tapatybę vis dar galima nustatyti pagal unikalų požymį, sukurtą pritaikius pseudonimų suteikimo funkciją (t. y. pseudonimu užkoduotą požymį).
- *Susiejimo galimybė:* išlieka galimybė nesunkiai susieti įrašus, sąsajai su tuo pačiu asmeniu nustatyti panaudojant ta patį pseudonimu užkoduotą požymį. Net jeigu tam pačiam duomenų subjektui būtų priskirti skirtingi pseudonimais užkoduoti požymiai, vis vien išliktų galimybė atlikti susiejimą remiantis kitais požymiais. Tik tuo atveju, jeigu duomenų subjekto tapatybei nustatyti negali būti naudojamasi jokiu kitu į duomenų rinkinį įtrauktu požymiu ir jeigu buvo panaikinta kiekviena pirminio požymio ir pseudonimu užkoduoto požymio sąsaja (taip pat ištrinant pirminius duomenis), nebeliks jokių akivaizdžių kryžminių sąsajų tarp duomenų rinkinių, kuriuose naudojami skirtingi pseudonimais užkoduoti požymiai.
- *Išvados padarymo galimybė:* remiantis vienu duomenų rinkiniu arba susiejant skirtingus duomenų rinkinius, kuriuose naudojamas toks pat pseudonimu užkoduotas asmens požymis arba kuriuose naudojami pseudonimai yra savaime aiškūs ir tinkamai nepaslepia pirminės duomenų subjekto tapatybės, galima vykdyti tikrosios duomenų subjekto tapatybės nustatymo išpuolius siekiant gauti išvestinių duomenų.

4.2. Dažnos klaidos

- Manymas, kad pseudonimais užkoduotas duomenų rinkinys yra nuasmenintas: duomenų valdytojai dažnai mano, kad vieno arba daugiau požymių pašalinimas arba pakeitimas yra pakankama duomenų nuasmeninimo priemonė. Daugybė pavyzdžių rodo, kad taip nėra; vien tik pakeitus identifikatorių nepanaikinama galimybė nustatyti duomenų subjekto tapatybę, jeigu duomenų rinkinyje lieka kvaziindentifikatorių arba jeigu asmens tapatybę vis dar galima nustatyti pasinaudojant kitų požymių vertėmis.

[18] Tai priklausytų nuo kitų duomenų rinkinyje esančių požymių ir nuo to, ar ištrinti pirminiai duomenys.

Nustatyti asmens tapatybę pseudonimais užkoduotame duomenų rinkinyje gali būti taip pat lengva, kaip ir pasinaudojant pirminiais duomenimis. Kad duomenų rinkinį būtų galima laikyti nuasmenintu, reikėtų imtis papildomų priemonių, įskaitant požymių pašalinimą arba apibendrinimą arba pirminių duomenų ištrynimą ar bent jau jų agregavimą aukštesniu lygmeniu.

- Su pseudonimų suteikimu, kuriuo siekiama sumažinti susiejimo galimybę, susijusios dažnos klaidos:

    - to paties rakto naudojimas skirtingose duomenų bazėse: skirtingų duomenų bazių susiejimo galimybės pašalinimas labai priklauso nuo to, ar naudojamas raktu pagrįstas algoritmas ir ar pavienis asmuo įvairiomis aplinkybėmis atitinka skirtingus pseudonimais užkoduotus požymius. Taigi, norint sumažinti susiejimo galimybę, svarbu vengti skirtingose duomenų bazėse naudoti tokį patį raktą;
    - skirtingų (kaitaliojamų) raktų suteikimas skirtingiems naudotojams: gali būti patrauklu skirtingoms naudotojų grupėms suteikti skirtingus raktus ir keisti raktą, kai jis panaudojamas tam tikrą kartų skaičių (pvz., naudoti tą patį raktą dešimčiai įvesties elementų, susijusių su tuo pačiu naudotoju, įrašyti). Tačiau, jeigu ši operacija bus parengta netinkamai, gali susidaryti tam tikri šablonai, dėl kurių iš dalies sumažės tikėtina nauda. Pavyzdžiui, kaitaliojant raktą pagal tam tikriems asmenims taikomas specialias taisykles, gali būti palengvinta su tais asmenimis susijusių įvesties elementų susiejimo galimybė. Be to, periodinis pseudonimais užkoduotų duomenų dingimas, kai tik atsiranda nauji duomenys, gali būti ženklas, kad abu įrašai susiję su tuo pačiu fiziniu asmeniu;
    - rakto laikymas: jeigu slaptas raktas bus saugomas kartu su pseudonimais užkoduotais duomenimis ir duomenys bus nutekinti, išpuolio vykdytojui gali būti lengva susieti pseudonimais užkoduotus duomenis su jų pirminiu požymiu. Tas pats pasakytina apie atvejį, kai raktas laikomas atskirai nuo duomenų, bet nesaugiai.

4.3. Pseudonimų naudojimo trūkumai

- *Sveikatos priežiūra*

1. Vardas ir pavardė, adresas, gimimo data	2. Specialios pašalpos mokėjimo laikotarpis	3. Kūno masės indeksas	6. Tyrimo grupės numeris
[redacted]	< 2 metai	15	QA5FRD4
	> 5 metai	14	2B48HFG
	< 2 metai	16	RC3URPQ
	> 5 metai	18	SD289K9
	< 2 metai	20	5E1FL7Q

5 lentelė. Lengvai atkuriamos tapatybės, kai pritaikytas pseudonimų suteikimo maišos metodas (vardas ir pavardė, adresas, gimimo data), pavyzdys

Šis duomenų rinkinys buvo sudarytas siekiant ištirti asmens svorio ir specialių pašalpų gavimo sąryšį. Pirminiame duomenų rinkinyje buvo nurodytas asmens vardas ir pavardė, adresas ir gimimo data, tačiau šie duomenys buvo ištrinti. Tyrimo grupės numeris sudarytas ištrintiems duomenims pritaikius maišos funkciją. Nors vardas ir pavardė, adresas ir gimimo

data buvo ištrinti iš lentelės, jeigu, be pritaikytos maišos funkcijos, būtų žinomas duomenų subjekto vardas ir pavardė, adresas ir gimimo data, būtų nesunku apskaičiuoti tyrimo grupių numerius.

- *Socialiniai tinklai*

Įrodyta[19], kad įslaptintą konkrečių asmenų informaciją galima gauti iš socialinių tinklų diagramų, nepaisant to, kad tokiems duomenims tariamai taikomi nuasmeninimo metodai. Socialinio tinklo paslaugos teikėjas laikėsi klaidingos nuomonės, kad toks duomenų nuasmeninimas yra patikimas būdas neleisti nustatyti asmens tapatybės po to, kai šie duomenys bus perduoti kitoms įmonėms rinkodaros arba reklamos reikmėms. Tikruosius vardus ir pavardes paslaugos teikėjas pakeitė pravardėmis, tačiau to aiškiai nepakako naudotojų profiliams nuasmeninti, nes pavienių asmenų tarpusavio ryšiai yra unikalūs ir jais galima pasinaudoti kaip identifikatoriumi.

- *Vietovės*

Masačiusetso technologijos instituto (MIT)[20] mokslininkai neseniai išanalizavo pseudonimais užkoduotą duomenų rinkinį, kuriame buvo per 15 mėnesių laikotarpį užregistruotos 1,5 mln. žmonių judėjimo 100 km spinduliu erdvės ir laiko koordinatės. Jie įrodė, kad pagal keturias buvimo vietas būtų galima išskirti 95 proc. asmenų, o pagal dvi – daugiau kaip 50 proc. duomenų subjektų (žinant, kad viena iš šių vietų veikiausiai yra namai arba darbo vieta); taigi privatumo apsaugos galimybės būtų labai menkos, net jeigu asmenų tapatybė būtų užkoduota pseudonimais, tikruosius jų požymius <...> pakeičiant kitomis žymėmis.

5. Išvados ir rekomendacijos

5.1. Išvados

Tapatybės duomenų panaikinimo ir nuasmeninimo metodai yra intensyvių mokslinių tyrimų objektas. Šiame dokumente nuosekliai parodyta, kad kiekvienas metodas turi privalumų ir trūkumų. Dažniausiai neįmanoma pateikti būtinųjų rekomendacijų dėl pasitelktinų parametrų, nes kiekvienas duomenų rinkinys turėtų būti nagrinėjamas atsižvelgiant į konkretų atvejį.

Daugeliu atvejų ir nuasmenintas duomenų rinkinys duomenų subjektams gali kelti liekamąją riziką. Išties, net jeigu nebelieka galimybės iš įrašo išgauti tikslius asmens duomenis, vis tiek įmanoma surankioti informaciją apie tą asmenį pasinaudojant kitais esamais informacijos šaltiniais (viešais arba neviešais). Reikėtų pabrėžti, kad netinkamai atliktas nuasmeninimas duomenų subjektams daro ne tik tiesioginį poveikį (nemalonumai, laiko gaišimas ir kontrolės praradimo jausmas, kylantis dėl įtraukimo į grupę nepranešus arba negavus išankstinio sutikimo), gali būti ir kitokių netinkamo nuasmeninimo netiesioginio poveikio padarinių, susijusių su tuo, kad koks nors išpuolio vykdytojas, remdamasis sutvarkytais nuasmenintais duomenimis, duomenų subjektą per klaidą pasirinko savo objektu, ypač jeigu tas išpuolio vykdytojas turi neteisėtų ketinimų. Todėl darbo grupė pabrėžia, kad nuasmeninimo metodais galima suteikti privatumo garantijų, tačiau tik tuo atveju, jeigu šių metodų taikymas tinkamai organizuojamas, t. y., norint pasiekti reikiamą nuasmeninimo lygį ir kartu parengti naudingus

[19] Narayanan, A., V. Shmatikov, V. *De-anonymizing social networks.* In: 30th IEEE Symposium on Security and Privacy, 2009.
[20] de Montjoye, Y.-A., Hidalgo, C., Verleysen, M. and Blondel, V. *Unique in the Crowd: The privacy bounds of human mobility*, Nature, No. 1376, 2013.

duomenis, turi būti aiškiai nustatytos nuasmeninimo procedūros prielaidos (aplinkybės) ir tikslas (-ai).

5.2. Rekomendacijos

- Kai kuriems nuasmeninimo metodams būdingi tam tikri apribojimai. Duomenų valdytojai, norėdami pagal tam tikrą metodą parengti nuasmeninimo procedūrą, iš pradžių turi rimtai įvertinti šiuos apribojimus. Jie privalo atsižvelgti į siekiamus nuasmeninimo tikslus, pvz., apsaugoti asmenų privatumą skelbiant duomenų rinkinį arba sudarant galimybę iš duomenų rinkinio gauti tam tikrą informaciją.
- Visi šiame dokumente aprašyti metodai nevisiškai atitinka veiksmingo nuasmeninimo kriterijus (t. y. asmens išskyrimo galimybės nebuvimo; su asmeniu susijusių įrašų susiejimo galimybės nebuvimo; su asmeniu susijusių išvestinių duomenų gavimo galimybės nebuvimo). Antra vertus, taikant tam tikrą metodą, kai kuriuos iš šių rizikos veiksnių galima visiškai arba iš dalies pašalinti, todėl būtina kruopščiai parengti tinkamą konkretaus metodo taikymo konkrečioje situacijoje procedūrą ir derinti šiuos metodus tarpusavyje, kad būtų gautas patikimas rezultatas.

Toliau pateiktoje lentelėje trijų pagrindinių reikalavimų požiūriu apžvelgiami aptariamų metodų privalumai ir trūkumai:

	Ar išlieka išskyrimo rizika?	**Ar išlieka susiejimo rizika?**	**Ar išlieka išvados padarymo rizika?**
Pseudonimų suteikimas	Taip	Taip	Taip
Iškraipytų duomenų įterpimas	Taip	Ne (laikantis tam tikrų sąlygų)	Ne (laikantis tam tikrų sąlygų)
Pakeitimas	Taip	Taip	Ne (laikantis tam tikrų sąlygų)
Agregavimas arba *k* anonimiškumas	Ne	Taip	Taip
l įvairovė	Ne	Taip	Ne (laikantis tam tikrų sąlygų)
Diferencinis privatumas	Ne (laikantis tam tikrų sąlygų)	Ne (laikantis tam tikrų sąlygų)	Ne (laikantis tam tikrų sąlygų)
Maiša ar pakaitinių simbolių naudojimas	Taip	Taip	Ne (laikantis tam tikrų sąlygų)

6 lentelė. Aptariamų metodų privalumai ir trūkumai

- Būtų geriausia, jei sprendimai būtų priimami atsižvelgiant į kiekvieną konkretų atvejį. Radus sprendimą (t. y. nustačius išsamią nuasmeninimo procedūrą), atitinkantį šiuos tris kriterijus, būtų patikimai panaikinta galimybė nustatyti asmens tapatybę labiausiai tikėtinomis priemonėmis, kurias galėtų pasitelkti duomenų valdytojas arba kokia nors trečioji šalis.
- Kai pasiūlymas neatitinka kurio nors iš šių kriterijų, turėtų būti atliekamas išsamus asmens tapatybės nustatymo rizikos vertinimas. Jeigu pagal nacionalinius teisės aktus reikalaujama, kad tam tikra institucija įvertintų arba leistų taikyti nuasmeninimo procedūrą, šis vertinimas turėtų būti pateiktas atitinkamai valdžios institucijai.

Norint sumažinti asmens tapatybės nustatymo riziką, reikėtų atsižvelgti į toliau aprašytą gerąją patirtį.

Geroji nuasmeninimo patirtis

Bendrieji patarimai

- Nesivadovaukite principu „paskelbk ir pamiršk“. Atsižvelgdami į liekamąją asmens tapatybės nustatymo riziką, duomenų valdytojai turėtų:
    - 1) reguliariai nustatyti naują riziką ir atlikti pakartotinius liekamosios rizikos vertinimus;
    - 2) vertinti, ar nustatytos rizikos valdymo priemonės yra pakankamos, ir imtis atitinkamų taisomųjų veiksmų;
    - 3) stebėti ir valdyti riziką.
- Vertindami liekamąją riziką, atsižvelkite į galimybę nustatyti asmens tapatybę remiantis nenuasmeninta duomenų rinkinio dalimi (jeigu tokia yra), ypač jeigu ši dalis būtų sujungta su nuasmeninta dalimi, ir į galimą požymių (pvz., geografinės vietovės ir turto duomenų) koreliaciją.

Su aplinkybėmis susiję veiksniai

- Turėtų būti aiškiai išdėstyti duomenų rinkinio nuasmeninimo tikslai, nes jie yra labai svarbūs vertinant asmens tapatybės nustatymo riziką.
- Šiuo tikslu taip pat turėtų būti atsižvelgta į visus svarbius su aplinkybėmis susijusius veiksnius, pvz., pirminių duomenų pobūdį, taikomas kontrolės priemones (įskaitant saugumo priemones, kuriomis ribojama galimybė naudotis duomenų rinkiniais), imties dydį (kiekybines charakteristikas), viešų informacijos šaltinių (kuriais galėtų remtis gavėjai) buvimą, numatomą duomenų teikimą trečiosioms šalims (ribotas, neribotas, pvz., internetu, ir t. t.).
- Turėtų būti aptarti galimi išpuolio vykdytojai, atsižvelgiant į duomenų patrauklumą tiksliniams išpuoliams (šiuo požiūriu taip pat labai svarbu įvertinti informacijos slaptumą ir duomenų pobūdį).

Techniniai aspektai

- Duomenų valdytojai turėtų nurodyti taikomą nuasmeninimo metodą arba jų rinkinį, ypač jeigu nuasmenintą duomenų rinkinį jie ketina paskelbti.
- Akivaizdūs (pvz., reti) požymiai ir (arba) kvaziidentifikatoriai turėtų būti pašalinti iš duomenų rinkinio.
- Jeigu taikomas iškraipytų duomenų įterpimo metodas (atliekant randomizavimą), įrašams taikytinas iškraipymo lygis turėtų būti nustatomas atsižvelgiant į požymio vertę (t. y. neturėtų būti įterpiami pernelyg iškraipyti duomenys), požymių, kuriuos reikia apsaugoti, poveikį duomenų subjektams ir (arba) duomenų rinkinio retumą.
- Jeigu remiamasi diferencinio privatumo metodu (atliekant randomizavimą), turėtų būti atsižvelgiama į būtinybę stebėti užklausas ir nustatyti privatumą pažeidžiančias užklausas, įvertinus jų kaupiamąjį poveikį.
- Jeigu taikomi apibendrinimo principu pagrįsti metodai, labai svarbu, kad duomenų valdytojas neapsiribotų vienu apibendrinimo kriterijumi, net jeigu jis būtų taikomas tam pačiam požymiui; kitaip tariant, turėtų būti parenkami įvairūs masteliai arba įvairūs laiko

intervalai. Parenkant taikytiną kriterijų, turi būti remiamasi požymių verčių paskirstymu nagrinėjamoje aibėje. Ne visi skirstiniai yra vienodai tinkami apibendrinti, t. y. apibendrinant negalima visada remtis tuo pačiu metodu. Reikėtų užtikrinti kintamumą lygiavertiškumo klasėse; pvz., atsižvelgiant į pirmiau minėtus su aplinkybėmis susijusius veiksnius (imties dydį ir t. t.), turėtų būti parinkta tam tikra ribinė vertė, ir, jeigu ji nėra pasiekta, atitinkama imtis turėtų būti atmesta (arba turėtų būti nustatytas kitoks apibendrinimo kriterijus).

PRIEDAS
Nuasmeninimo metodų pagrindai

A.1. Įžanga

Europos Sąjungoje anonimiškumas suprantamas nevienodai: vienose šalyse jis suprantamas kaip skaičiavimų anonimiškumas (t. y. net ir duomenų valdytojui, bendradarbiaujančiam su bet kuria šalimi, skaičiavimo požiūriu turėtų būti sunku tiesiogiai arba netiesiogiai nustatyti kurio nors duomenų subjekto tapatybę), kitose – kaip visiškas anonimiškumas (t. y. net ir duomenų valdytojui, bendradarbiaujančiam su bet kuria šalimi, turėtų būti neįmanoma tiesiogiai arba netiesiogiai nustatyti kurio nors duomenų subjekto tapatybę). Nepaisant to, abiem atvejais nuasmeninimas reiškia procesą, per kurį duomenys tampa anoniminiai. Šie požiūriai skiriasi nustatant priimtiną galimybės atsekti tapatybę rizikos lygį.

Nuasmeninti duomenys gali būti numatyti naudoti įvairiais būdais: socialiniuose tyrimuose, statistinėje analizėje ir naujų paslaugų ir (arba) produktų kūrimo strategijose. Kartais net ir tokia bendrosios paskirties veikla tam tikriems duomenų subjektams gali daryti poveikį, dėl kurio tariamas sutvarkytų duomenų anonimiškumas netenka prasmės. Galima pateikti daug su tuo susijusių pavyzdžių, pradedant tikslinės rinkodaros iniciatyvomis ir baigiant viešųjų priemonių įgyvendinimu remiantis naudotojų savybėmis, elgsena arba judėjimo pobūdžiu[21].

Deja, neskaitant bendro pobūdžio nuostatų, nėra tokių ištobulintų matavimo priemonių, kuriomis būtų galima iš anksto įvertinti laiką arba pastangas, reikalingus norint pakartotinai nustatyti asmens tapatybę remiantis sutvarkytais duomenimis, arba kaip nors kitaip parinkti tinkamiausią taikytiną procedūrą, kai pageidaujama sumažinti paskelbto duomenų rinkinio susiejimo su nustatytu duomenų subjektų rinkiniu galimybę.

Nuasmeninimo technika (taip ši veikla kartais vadinama mokslinėje literatūroje[22]) – nauja mokslo šaka, dar tik žengianti pirmuosius savo žingsnius, tačiau yra daugybė praktinių būdų, kuriais galima sumažinti su duomenų rinkiniais susijusių asmenų tapatybės nustatymo galimybę; vis dėlto reikėtų nevienareikšmiškai pasakyti, kad daugeliu šių praktinių būdų nepanaikinama galimybė sutvarkytus duomenis susieti su duomenų subjektais. Įrodyta, kad tam tikromis aplinkybėmis galima labai sėkmingai nustatyti su duomenų rinkiniais, kurie laikomi anoniminiais, susijusių asmenų tapatybę, o tam tikromis sąlygomis gali būti padarytos klaidingos teigiamos išvados.

Plačiąja prasme skirtini du būdai: vienas iš jų pagrįstas požymių apibendrinimu, antras – randomizavimu. Išnagrinėję šių praktinių metodų ypatumus ir subtilybes, geriau suprasime duomenų identifikavimo galimybes ir pačią asmens duomenų sąvoką.

A.2. Nuasmeninimas taikant randomizavimo principą

Vienas iš nuasmeninimo būdų – tikrųjų verčių pakeitimas siekiant panaikinti galimybę nuasmenintus duomenis susieti su pirminėmis vertėmis. Tai galima padaryti taikant įvairius metodus, pradedant iškraipytų duomenų įterpimu ir baigiant duomenų sukeitimu (perstatymas). Reikia pabrėžti, kad požymio pašalinimas yra kraštutinis šio požymio randomizavimo būdas (tokiu atveju požymiui priskiriamos tik iškraipytos vertės).

[21] Pavyzdys – *TomTom* atvejis Nyderlanduose (žr. pavyzdį, aprašytą 2.2.3 skirsnyje).
[22] Jun Gu, Yuexian Chen, Junning Fu, Huanchun Peng, Xiaojun Ye, *Synthesizing: Art of Anonymization*, Database and Expert Systems Applications Lecture Notes in Computer Science. Springer, Vol. 6261, 2010, p. 385–399.

Tam tikromis aplinkybėmis tvarkymo bendrasis tikslas – ne paskelbti randomizuotą duomenų rinkinį, o suteikti galimybę gauti duomenis pasitelkiant užklausas. Rizika duomenų subjektui šiuo atveju kyla dėl tikimybės, kad išpuolio vykdytojas galės gauti informaciją be duomenų valdytojo žinios atlikęs daug skirtingų užklausų. Siekiant garantuoti su atitinkamu duomenų rinkiniu susijusių asmenų privatumą, neturėtų būti galima padaryti išvadą, kad duomenų subjektas įtrauktas į duomenų rinkinį, taip panaikinant bet kokią susiejimo su bendrąja informacija, kurią išpuolio vykdytojas gali turėti, galimybę.

Į užklausų rezultatus įterpiant iškraipytus duomenis, galima dar labiau sumažinti pakartotinio tapatybės nustatymo riziką. Šis metodas, literatūroje dar vadinamas diferenciniu privatumu[23], nuo pirmiau aprašytųjų skiriasi tuo, kad šiuo metodu duomenis skelbiantiems asmenims suteikiama didesnė, jei lyginsime su viešu paskelbimu, galimybė kontroliuoti duomenų prieigą. Iškraipyti duomenys įterpiami siekiant dviejų pagrindinių tikslų: pirma, norint apsaugoti į duomenų rinkinį įtrauktų duomenų subjektų privatumą, antra, norint išsaugoti paskelbtos informacijos naudingumą. Pabrėžtina, kad duomenų iškraipymo lygis turi būti proporcingas užklausų skaičiui (kuo daugiau pateikiama užklausų apie asmenis siekiant gauti labai tikslius duomenis, tuo didesnė asmens tapatybės nustatymo tikimybė). Šiais laikais, norint sėkmingai pritaikyti randomizavimą, kiekvieną atvejį būtina vertinti atskirai, nes nėra tokio metodo, pagal kurį būtų taikoma labai paprasta metodika: yra daugybė pavyzdžių, kaip buvo nutekinta informacija apie duomenų subjekto požymius (nesvarbu, ar šis subjektas buvo įtrauktas į duomenų rinkinį, ar ne), nors duomenų valdytojas ir randomizavo duomenų rinkinį.

Gali būti naudinga aptarti konkrečius pavyzdžius ir išsiaiškinti galimus netinkamo randomizavimo, kaip nuasmeninimo būdo, taikymo atvejus. Pavyzdžiui, kai suteikiama interaktyvi prieiga, dėl užklausų, kurios laikomos nepažeidžiančiomis privatumo, gali kilti rizika duomenų subjektams. Iš tiesų, jeigu išpuolio vykdytojas žino, kad asmenų pogrupis S įtrauktas į duomenų rinkinį, kuriame yra informacija apie A požymio paplitimą P aibėje, pateikus paprastą užklausą, sudarytą iš dviejų klausimų – „Kokiam skaičiui asmenų P aibėje būdingas A požymis?“ ir „Kokiam skaičiui asmenų P aibėje, išskyrus tuos, kurie priklauso S pogrupiui, būdingas A požymis?“, galima nustatyti (apskaičiuojant skirtumą) S pogrupiui priklausančių asmenų, kuriems tikrai būdingas A požymis, skaičių (determinavimo arba tikimybinės išvesties būdu). Bet kuriuo atveju gali būti rimtai pažeistas S pogrupiui priklausančių asmenų privatumas, ypač pagal A požymio pobūdį.

Be to, jeigu duomenų subjektas nėra įtrauktas į duomenų rinkinį, tačiau yra žinoma jo sąsaja su duomenų rinkinyje esančiais duomenimis, galima manyti, kad, paskelbus duomenų rinkinį, gali kilti rizika duomenų subjekto privatumui. Pavyzdžiui, jeigu žinoma, kad „tikslinio subjekto A požymio vertė nuo aibės vidurkio skiriasi dydžiu X“, išpuolio vykdytojas, duomenų bazės valdytojo paprašęs atlikti paprastą privatumo nepažeidžiančią operaciją – pateikti A požymio vertės vidurkį, gali tiksliai nustatyti su konkrečiu duomenų subjektu susijusius asmens duomenis.

Duomenų rinkinio tikrųjų verčių nedidelio iškraipymo operacija turėtų būti tinkamai parengta. Siekiant užtikrinti privatumą, turi būti įterpti pakankamai iškraipyti duomenys, tačiau iškraipymo lygis turėtų būtų nedidelis, kad būtų išsaugotas duomenų naudingumas. Pavyzdžiui, jeigu duomenų subjektų, kuriems būdingas savitas požymis, yra labai mažai arba požymis yra labai slaptas, gal būtų geriau nurodyti intervalą arba suformuluoti bendro pobūdžio sakinį, pvz., „reti atvejai, gali būti net nulis atvejų“, o ne pateikti tikrą jų skaičių. Taip, net jeigu iš anksto būtų žinomas netikslaus verčių atskleidimo mechanizmas, vis tiek

[23] Cynthia Dwork, *Differential Privacy*, International Colloquium on Automata, Languages and Programming (ICALP), 2006, p. 1–12.

būtų išsaugotas duomenų subjekto privatumas, nes išliktų tam tikras netikrumo laipsnis. Naudingumo požiūriu, jeigu duomenų iškraipymo procedūra parengiama tinkamai, rezultatai vis dar gali būti naudojami statistikos arba sprendimų priėmimo tikslais.

Be to, turi būti apsvarstytos duomenų bazės randomizavimo ir diferencinio privatumo metodu pagrįstos prieigos galimybės. Pirma, tinkamas iškraipymo lygis gali gerokai skirtis atsižvelgiant į aplinkybes (užklausos tipą, į duomenų bazę įtrauktos aibės dydį, požymio pobūdį ir jam būdingą tapatybės nustatymo galimybę); taigi visiems atvejams tinkamo sprendimo numatyti negalima. Negana to, laikui bėgant, aplinkybės gali keistis, todėl interaktyvus mechanizmas atitinkamai turėtų būti keičiamas. Norint nustatyti reikiamą duomenų iškraipymo lygį, būtina stebėti su bet kurio interaktyvaus mechanizmo taikymu susijusią kaupiamąją riziką duomenų subjektų privatumui. Be to, į duomenų prieigos mechanizmą turėtų būti įtraukti įspėjimai, kad pasiekta riba, kurią peržengus bus neleistinai pažeistas privatumas, ir kad, pateikus dar vieną užklausą, duomenų subjektams gali kilti tam tikra rizika; šie įspėjimai padėtų duomenų valdytojui nustatyti reikiamą iškraipymo lygį, kuris kaskart turėtų būti taikomas tikriesiems asmens duomenims.

Kita vertus, taip pat turėtume apsvarstyti atvejį, kai požymių vertės ištrinamos (arba pakeičiamos). Dažnas sprendimas, taikomas kai kurioms netipinėms požymių vertėms, yra su netipiniais asmenimis susijusio duomenų rinkinio arba netipinių verčių ištrynimas. Pastaruoju atveju svarbu įsitikinti, kad pats vertės nebuvimas nepadėtų nustatyti duomenų subjekto tapatybės.

Dabar aptarkime randomizavimą požymių pakeitimo metodu. Labai klaidinga manyti, kad nuasmeninimas yra tas pat, kas šifravimas arba kodavimas naudojant raktą. Ši klaidinga nuomonė grindžiama dviem prielaidomis: a) kadangi šifravimas taikomas kai kuriems duomenų bazės įrašo požymiams (pvz., vardui, pavardei, adresui, gimimo datai) arba šie požymiai pakeičiami iš pažiūros randomizuota eilute, sudaroma taikant kodavimo naudojant raktą operaciją, pvz., rakto naudojimu pagrįstos maišos funkciją, tai įrašas yra nuasmenintas; b) nuasmeninimas bus veiksmingesnis, jeigu raktas bus tinkamo ilgio ir šifravimo algoritmas bus pažangus. Ši klaidinga nuomonė labai paplitusi tarp duomenų valdytojų, todėl ją reikėtų panagrinėti plačiau, nes ji taip pat susijusi su pseudonimų suteikimu ir tariamai mažesne šio metodo rizika.

Pirmiausia šių metodų tikslai yra visiškai skirtingi: šifravimas, kaip saugumo priemonė, yra skirtas ryšių linijos tarp nustatytų šalių (žmonių, įtaisų ar programinės arba aparatinės įrangos dalių) slaptumui užtikrinti, siekiant išvengti slapto pasiklausymo arba nepageidaujamo informacijos atskleidimo. Rakto naudojimu pagrįstas kodavimas – tai reikšminis duomenų pakeitimas naudojant slaptą raktą. Kita vertus, nuasmeninimo tikslas – panaikinti galimybę nustatyti asmens tapatybę, neleidžiant slapta susieti požymių su duomenų subjektu.

Nei šifravimas, nei rakto naudojimu pagrįstas kodavimas nėra tinkami duomenų subjekto tapatybės nustatymo galimybei panaikinti: kadangi bent vienas asmuo, t. y. duomenų valdytojas, tebeturi pirminius duomenis, išlieka galimybė gauti arba nustatyti pirminius duomenis. Atliekant tik reikšminį asmens duomenų pakeitimą, kaip tai daroma taikant rakto naudojimu pagrįstą kodavimą, nepanaikinama galimybė atkurti pirminę duomenų struktūrą pritaikius atvirkštinį algoritmą, surengus jėgos išpuolius (konkretus būdas priklauso nuo sistemų pobūdžio) arba nutekinus duomenis. Pažangiais šifravimo metodais galima užtikrinti, kad duomenys būtų geriau apsaugoti, t. y. taptų nesuprantami asmenims, nenaudojantiems iššifravimo rakto, tačiau duomenys nebūtinai tampa nuasmeninti. Kol yra pirminių duomenų raktas (net jeigu jį turi patikima trečioji šalis, pagal sutartį privalanti teikti saugią rakto deponavimo paslaugą), galimybė nustatyti duomenų subjekto tapatybę nepanaikinama.

Būtų klaidinga manyti, kad šifravimo mechanizmo patikimumas yra svarbiausias duomenų rinkinio nuasmeninimo lygio matas, nes bendras šifravimo mechanizmo arba maišos funkcijos saugumas taip pat priklauso nuo daugelio kitų techninių ir organizacinių veiksnių. Literatūroje aprašoma daug sėkmingų išpuolių, kai buvo visiškai išvengta algoritmo kliūties – pasinaudota raktų laikymo trūkumais (pvz., taikomas ne toks saugus numatytasis režimas) arba kitais su žmogumi susijusiais veiksniais (pvz., nesunkiai atspėjamais raktų atkūrimo slaptažodžiais). Galiausiai ši šifravimo sistema, grindžiama tam tikro ilgio raktu, yra skirta konfidencialumui užtikrinti tam tikrą laiką (apie 2020 m. turės būti pakeistas daugelio dabartinių raktų dydis), o nuasmeninimo procedūrai laiko apribojimai neturėtų būti taikomi.

Dabar, remiantis įvairiais per pastaruosius kelerius metus randomizavimo principu atlikto netinkamo nuasmeninimo pavyzdžiais, vertėtų panagrinėti požymio randomizavimo (arba sukeitimo ir pašalinimo) trūkumus bei nesėkmių taikant šį metodą priežastis.

Gerai žinomas netinkamai nuasmeninto duomenų rinkinio paskelbimo atvejis, susijęs su *Netflix Prize*[24]. Išnagrinėjus tipinį įrašą duomenų bazėje, kurioje buvo randomizuota keletas su duomenų subjektu susijusių požymių, matyti, kad kiekvieną įrašą vis dar galima padalyti į du smulkesnius įrašus, t. y. {randomizuoti požymiai, aiškūs požymiai}, šiuo atveju aiškiu požymiu gali būti bet koks tariamai ne asmens duomenų derinys. Konkreti išvada, kurią galima padaryti remiantis *Netflix Prize* duomenų rinkiniu, gauta pastebėjus, kad kiekvieną įrašą galima pažymėti kaip tašką daugiamatėje erdvėje, kurioje kiekvienas aiškus požymis yra atskira koordinatė. Pagal šį principą bet koks duomenų rinkinys gali būti laikomas taškų aibe tokioje daugiamatėje erdvėje, kuriai gali būti būdingas labai mažas tankis, o tai reiškia, kad taškai gali būti labai nutolę vienas nuo kito. Išties jie gali būti taip toli vienas nuo kito, kad, erdvę padalijus į plačias sritis, kiekvienoje liks tik po vieną įrašą. Net ir įterpus iškraipytų duomenų, įrašai nepriartėja vienas prie kito tiek, kad patektų į tą pačią daugiamatę sritį. Pavyzdžiui, atliekant su *Netflix* susijusį eksperimentą, įrašai, kurie apėmė tik aštuonis filmų įverčius, kuriuos skiria 14 dienų, buvo pakankamai unikalūs. Į įverčius ir datas įterpus iškraipytų duomenų, sričių sutapimo nepastebėta. Kitaip tariant, pasirinkus tik aštuonis įvertintus filmus, nustatytas visoje duomenų bazėje tarp jokių dviejų duomenų subjektų nesikartojantis pareikštų įvertinimų bruožas. Remdamiesi šia geometrine išvada, tyrinėtojai tariamai anoniminį *Netflix* duomenų rinkinį palygino su kita vieša filmų vertinimo duomenų baze (IMDb) ir nustatė naudotojus, kurie tokiais pat laikotarpiais įvertino tuos pačius filmus. Kadangi daugumą naudotojų buvo galima unikaliai susieti tarpusavyje, iš IMDb duomenų bazės gautą pagalbinę informaciją buvo galima importuoti į paskelbtą *Netflix* duomenų rinkinį ir taip visus tariamai nuasmenintus įrašus papildyti tapatybės duomenimis.

Svarbu pažymėti, kad tai – bendra savybė: likusi bet kokios randomizuotos duomenų bazės dalis išlaiko labai didelę galimybę pagal ją nustatyti asmens tapatybę; ši galimybė priklauso nuo liekamųjų požymių derinių retumo. Šį svarbų aspektą duomenų valdytojai, pasirinkdami randomizavimą, kaip reikiamo nuasmeninimo lygio pasiekimo būdą, visada turėtų turėti omenyje.

Daugelis tokio pobūdžio pakartotinio tapatybės atkūrimo eksperimentų buvo atliekama panašiu principu – dvi duomenų bazes pavaizduojant viename poerdvyje. Tai labai veiksminga pakartotinio tapatybės nustatymo metodika, pastaruoju metu įvairiais būdais taikyta skirtingose srityse. Pavyzdžiui, asmens tapatybės nustatymo eksperimentas buvo

[24] Arvind Narayanan, Vitaly Shmatikov, *Robust De-anonymization of Large Sparse Datasets*. In: IEEE Symposium on Security and Privacy, 2008, p. 111–125.

atliktas socialinio tinklo[25] atžvilgiu – pasitelkta naudotojų socialinių ryšių diagrama, kuri buvo užkoduota pseudonimais naudojant žymes. Šiuo atveju asmens tapatybei nustatyti naudoti požymiai buvo kiekvieno naudotojo kontaktinių asmenų sąrašas, nes įrodyta, jog tikimybė, kad dviejų asmenų kontaktinių asmenų sąrašai bus vienodi, yra labai maža. Remiantis šia intuityvia prielaida, nustatyta, kad reikia gauti tinkle paslėptą topologinį kodą, t. y. labai nedidelio skaičiaus mazgų vidaus saitų dalinę diagramą, ir, identifikavus šį potinklį, didelėje viso socialinio tinklo dalyje būtų galima nustatyti asmenų tapatybę. Tiesiog, norint pateikti kelis su panašaus išpuolio veiksmingumu susijusius skaičius, buvo įrodyta, kad, naudojant mažiau kaip 10 mazgų (iš jų gali būti sudaryta milijonas skirtingų potinklio sąrankų, iš kurių kiekviena gali būti topologinis kodas), socialinis tinklas, sudarytas iš daugiau kaip keturių milijonų pseudonimais užkoduotų mazgų ir 70 mln. saitų, gali tapti neatsparus pakartotinio tapatybės atkūrimo išpuoliams, taigi gali būti pažeistas labai gausaus susijungimų skaičiaus privatumas. Reikėtų pabrėžti, kad toks pakartotinio tapatybės nustatymo būdas nėra pritaikytas konkrečiam socialinių tinklų kontekstui, jis yra gana bendro pobūdžio, jį galima pritaikyti kitoms duomenų bazėms, kuriose nurodyti naudotojų tarpusavio ryšiai (pvz., telefono adresatai, el. laiškai, susitikimo vietos ir t. t.).

Kitas tariamai anoniminio įrašo atpažinimo būdas pagrįstas stilistine rašymo analize (stilometrija)[26]. Jau parengta keletas algoritmų, skirtų iš analizuojamo teksto atrinkti metrikos duomenis, be kitų dalykų, nustatant tam tikro žodžio vartojimo dažnumą, tam tikrų gramatinių modelių paplitimą ir skyrybos ženklų pobūdį. Visomis šiomis ypatybėmis galima remtis tariamai anoniminį tekstą susiejant su nustatyto autoriaus rašymo stiliumi. Tyrinėtojai tinklaraščiuose nustatė daugiau kaip 100 000 rašymo stilių, ir dabar beveik 80 proc. tikslumu galima automatiškai nustatyti paskelbto teksto autoriaus tapatybę; manoma, kad šio metodo tikslumas, pasitelkiant ir kitus požymius, pvz., vietą arba kitus teksto metaduomenis, toliau didės.

Mokslininkų bendruomenė ir pramonės atstovai turėtų daugiau dėmesio skirti galimybei nustatyti asmens tapatybę remiantis įrašo reikšminėmis savybėmis (t. y. likusia nerandomizuota įrašo dalimi). Neseniai atlikti pakartotinio DNR donorų tapatybės nustatymo bandymai (2013 m.)[27] rodo, kad nuo gerai žinomo AOL incidento (2006 m.), kai buvo viešai paskelbta duomenų bazė su 20 mln. reikšminių paieškos žodžių, kuriuos trijų mėnesių laikotarpiu vartojo daugiau kaip 650 000 naudotojų, į priekį pasistūmėta labai nedaug. Pasinaudojant šia duomenų baze, buvo nustatyta kai kurių AOL naudotojų tapatybė ir buvimo vieta.

Kita duomenų, kurie retai kada nuasmeninami tik pašalinant duomenų subjektų tapatybės informaciją arba iš dalies užšifruojant kai kuriuos požymius, grupė yra buvimo vietos duomenys. Žmonių judėjimo pobūdis gali būti pakankamai išskirtinis, kad, pasinaudojus reikšmine vietos duomenų dalimi (vietomis, kuriose duomenų subjektas buvo tam tikru

[25] Backstrom, L., Dwork, C., Kleinberg, J. M. *Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography*. Proceedings of the 16th International Conference on World Wide Web WWW'07, 2007, p.181–190.

[26] http://33bits.org/2012/02/20/is-writing-style-sufficient-to-deanonymize-material-posted-online/

[27] Genetiniai duomenys yra labai svarbus įslaptintų duomenų, kuriems gali grėsti tapatybės atkūrimo rizika, jeigu vienintelis jų nuasmeninimo būdas bus donorų tapatybės duomenų pašalinimas, pavyzdys. Žr. pirmiau 2.2.2 skirsnyje pateiktą pavyzdį. Taip pat žr. John Bohannon, *Genealogy Databases Enable Naming of Anonymous DNA Donors*, Science, Vol. 339, No. 6117 (18 January 2013), p. 262.

konkrečiu metu), netgi nežinant kitų požymių, būtų galima sužinoti daug duomenų subjekto savybių[28]. Tai daug kartų įrodyta atlikus reprezentacinius mokslinius tyrimus[29].

Šiuo požiūriu svarbu įspėti dėl pavojų, susijusių su pseudonimų naudojimu siekiant užtikrinti tinkamą duomenų subjektų apsaugą nuo tapatybės duomenų arba požymių nutekėjimo. Jeigu pseudonimų suteikimas grindžiamas tuo, kad tapatybės duomenys pakeičiami kitu unikaliu kodu, būtų naivu manyti, kad tai yra patikimas galimybės atkurti asmens tapatybę panaikinimo būdas, nes juo neatsižvelgiama į asmens tapatybės nustatymo metodų sudėtingumą ir įvairias galimas jų taikymo aplinkybes.

A.3. Nuasmeninimas apibendrinimo principu

Požymių apibendrinimu pagrįstą metodą gali padėti paaiškinti paprastas pavyzdys.

Tarkime, duomenų valdytojas nusprendžia paskelbti paprastą lentelę, kurioje pateikiama trejopa informacija (trys požymiai): asmens tapatybės numeris, unikalus kiekvieno įrašo atžvilgiu, vietovės kodas, kuriuo duomenų subjektas susiejamas su jo gyvenamąja vieta, ir savybės kodas, rodantis duomenų subjektui būdingą savybę; be to, tarkime, kad šioms savybėms gali būti priskirtos dvi skirtingos vertės, bendrai nurodomos kaip {P1, P2}:

Asmens tapatybės numeris	**Vietovės kodas**	**Savybė**
1	Roma	P1
2	Madridas	P1
3	Londonas	P2
4	Paryžius	P1
5	Barselona	P1
6	Milanas	P2
7	Niujorkas	P2
8	Berlynas	P1

A.1 lentelė. Duomenų subjektų imtis, įtraukiant vietovės ir savybių (P1 arba P2) požymius

Jeigu kuris nors išpuolio vykdytojas iš anksto žino, kad į lentelę yra įtrauktas tam tikras duomenų subjektas (tikslinis subjektas), gyvenantis Milane, tai, peržiūrėjęs lentelę, jis gali sužinoti, kad to subjekto tapatybės numeris yra 6, nes tik šis numeris susietas su tokiu vietovės kodu ir šiam subjektui būdinga P2 savybė.

Šiuo labai paprastu pavyzdžiu parodytos pagrindinės bet kurios tapatybės nustatymo procedūros, taikomos tariamai nuasmenintam duomenų rinkiniui, grandys. Pavyzdžiui, išpuolio vykdytojas (atsitiktinai arba sąmoningai) gavo bendrųjų žinių apie kai kuriuos arba visus į duomenų rinkinį įtrauktus duomenų subjektus. Išpuolio vykdytojas siekia šias

[28] Kai kurios šalys šį klausimą reglamentuoja teisės aktais. Pavyzdžiui, Prancūzijoje skelbiami statistikos duomenys apie buvimo vietą nuasmeninami apibendrinimo ir perstatymo būdais. Taigi INSEE skelbiami statistikos duomenys yra apibendrinti visus duomenis agregavus į 40 000 kvadratinių metrų zonas. Toks duomenų rinkinio mastelis yra pakankamas, kad būtų išsaugotas duomenų naudingumas, o, taikant perstatymą, panaikinama galimybė vykdyti išpuolius retų duomenų srityse. Apskritai šios duomenų grupės agregavimas ir perstatymas gerai apsaugo nuo išvestinių duomenų gavimo ir nuasmeninimo panaikinimo išpuolių (http://www.insee.fr/en/).

[29] de Montjoye, Y.-A., Hidalgo, C.A., Verleysen, M. & Blondel, V.D. *Unique in the Crowd: The privacy bounds of human mobility*. Nature (2013), 3, 1376.

bendrąsias žinias susieti su paskelbtame duomenų rinkinyje esančia informacija, kad susidarytų aiškesnį vaizdą apie tų duomenų subjektų savybes.

Siekdamas užtikrinti, kad duomenų susiejimas su bet kokiomis bendrosiomis žiniomis būtų kuo neveiksmingesnis arba sudėtingesnis, duomenų valdytojas galėtų daugiau dėmesio skirti vietos identifikatoriui, miestą, kuriame gyvena duomenų subjektai, pakeisdamas didesnio ploto vietove, pvz., šalimi. Tuomet lentelė atrodytų taip:

Asmens tapatybės numeris	**Vietovės kodas**	**Savybė**
1	Italija	P1
2	Ispanija	P1
3	Jungtinė Karalystė	P2
4	Prancūzija	P1
5	Ispanija	P1
6	Italija	P2
7	JAV	P2
8	Vokietija	P1

A.2 lentelė. A.1 lentelės duomenų apibendrinimas pilietybės lygmeniu

Atlikus tokį naują duomenų agregavimą, išpuolio vykdytojas iš turimų bendrųjų žinių apie atpažintą duomenų subjektą (pvz., „subjektas gyvena Romoje ir yra įtrauktas į lentelę“) negalėtų daryti tikslios išvados dėl subjektui būdingos savybės, nes lentelėje yra du italai ir jiems būdingos skirtingos savybės – atitinkamai P1 ir P2. Tikimybė, kad išpuolio vykdytojas padarys teisingą išvadą dėl tikslinio subjekto savybės, tebėra 50 proc. Šis paprastas pavyzdys rodo apibendrinimo naudą nuasmeninimo srityje. Iš tikrųjų, nors toks apibendrinimas gali būti naudingas siekiant perpus sumažinti tikimybę, kad bus teisingai nustatyta tikslinio subjekto iš Italijos tapatybė, jis nėra veiksmingas, kai tiksliniai subjektai yra iš kitų vietovių (pvz., JAV).

Be to, išpuolio vykdytojas vis dar gali sužinoti informaciją apie tikslinį subjektą iš Ispanijos. Jeigu bendrosios žinios yra „tikslinis subjektas gyvena Madride ir yra įtrauktas į lentelę“ arba „tikslinis subjektas gyvena Barselonoje ir yra įtrauktas į lentelę“, išpuolio vykdytojas gali daryti 100 proc. patikimą išvadą, kad tiksliniam subjektui būdinga P1 savybė. Taigi apibendrinimo būdu užtikrinamas nevienodas į duomenų rinkinį įtrauktų subjektų privatumo lygis arba jie pasižymi skirtingu atsparumu išpuoliams siekiant gauti išvestinių duomenų.

Remiantis šia logika, gali kilti išvada, kad platesnis apibendrinimas galėtų padėti išvengti bet kokio susiejimo, pvz., jeigu būtų apibendrinama žemynų lygiu. Tuomet lentelė atrodytų taip:

Asmens tapatybės numeris	Vietovės kodas	Savybė
1	Europa	P1
2	Europa	P1
3	Europa	P2
4	Europa	P1
5	Europa	P1
6	Europa	P2
7	Šiaurės Amerika	P2
8	Europa	P1

A.3 lentelė. A.1 lentelės duomenų apibendrinimas žemynų lygmeniu

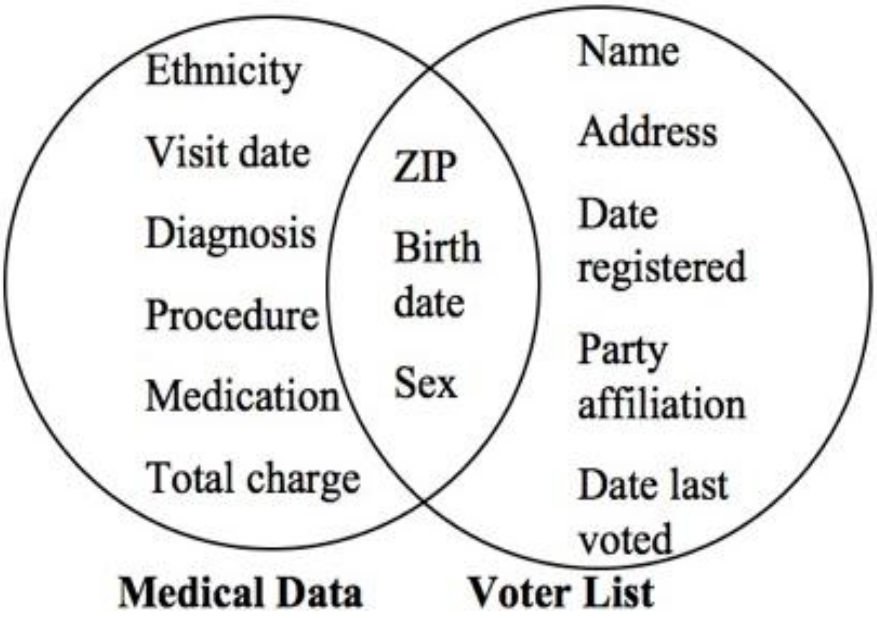
Pasitelkus tokį agregavimo būdą, visi į lentelę įtraukti duomenų subjektai, išskyrus gyvenantį JAV, būtų apsaugoti nuo susiejimo ir tapatybės nustatymo išpuolių, o, naudojantis tokia bendrąja informacija kaip „tikslinis subjektas gyvena Madride ir yra įtrauktas į lentelę“ arba „tikslinis subjektas gyvena Milane ir yra įtrauktas į lentelę“, konkrečiam duomenų subjektui savybę būtų galima priskirti tik vadovaujantis tam tikra tikimybe (P1 – su 71,4 proc. tikimybe, P2 – su 28,6 proc. tikimybe), o tiesioginis susiejimas nebūtų galimas. Be to, atliekant tokį didesnio mastelio apibendrinimą, akivaizdžiai prarandama daug informacijos: remiantis tokia lentele, neįmanoma nustatyti galimo savybių ir vietovės sąryšio, t. y. ar kurioje nors vietovėje gyvenantiems asmenims labiau būdinga kuri nors iš dviejų savybių, nes tokia lentelė suteikia tik galimybę nustatyti vadinamąjį ribinį pasiskirstymą, t. y. neabejotiną tikimybę, kad P1 arba P2 savybė pasireikš visoje aibėje (mūsų pavyzdyje šios tikimybės atitinkamai yra 62,5 proc. ir 37,5 proc.) ir kiekviename žemyne (kaip jau minėta, atitinkamai 71,4 proc. ir 28,6 proc. – Europoje, 100 proc. ir 0 proc. – Šiaurės Amerikoje).

Be to, šis pavyzdys rodo, kad praktinis duomenų naudingumas priklauso nuo taikomo apibendrinimo būdo. Jau siūloma keletas parengtų metodinių priemonių, skirtų iš anksto (t. y. prieš paskelbiant duomenų rinkinį) nustatyti, koks požymių apibendrinimo lygis būtų tinkamiausias siekiant sumažinti į lentelę įtrauktų duomenų subjektų tapatybės nustatymo riziką, pernelyg nesumažinant skelbiamų duomenų naudingumo.

k anonimiškumas

Vienas iš metodų, kuriuo požymių apibendrinimo principu panaikinama galimybė vykdyti susiejimo išpuolius, yra *k* anonimiškumas. Šis praktinis metodas parengtas remiantis pakartotinio tapatybės nustatymo eksperimentu, atliktu pačioje XX a. pabaigoje, kai privati JAV sveikatos priežiūros įmonė viešai paskelbė neva nuasmenintą duomenų rinkinį. Nuasmeninimas atliktas pašalinus subjektų vardus ir pavardes, bet duomenų rinkinyje buvo palikti su sveikata susiję duomenys ir kiti požymiai, pvz., pašto (gyvenamosios vietos) kodas, lytis ir visa gimimo data. Tie patys trys elementai {pašto kodas, lytis, visa gimimo data} yra įtraukti ir į kitus viešuosius registrus (pvz., rinkėjų sąrašą), todėl akademinis tyrinėtojas, naudodamasis šiais duomenimis, konkrečių duomenų subjektų tapatybę galėjo susieti su paskelbto duomenų rinkinio požymiais. Išpuolio vykdytojas (tyrinėtojas) galėjo turėti tokių bendrųjų žinių: „Žinau, kad į rinkėjų sąrašą įtrauktas duomenų subjektas, kuriam priskirti trys konkretūs elementai {pašto kodas, lytis, visa gimimo data}, yra unikalus. Paskelbtame

duomenų rinkinyje yra šių trijų požymių įrašas.“ Atlikus empirinį tyrimą, nustatyta[30], kad absoliučią daugumą (daugiau kaip 80 proc.) duomenų subjektų, įtrauktų į viešąjį registrą, kuriuo buvo naudojamasi per šį eksperimentą, buvo galima vienareikšmiškai susieti su konkrečiu trijų elementų rinkiniu, todėl buvo galima nustatyti ir asmenų tapatybę. Taigi šiuo atveju duomenys buvo nuasmeninti netinkamai.



A.1 paveikslas. Pakartotinis asmens tapatybės nustatymas susiejant duomenis

Buvo teigiama, kad duomenų valdytojai, norėdami sumažinti panašių susiejimo išpuolių veiksmingumą, pirmiausia turėtų peržiūrėti duomenų rinkinį ir sugrupuoti tuos požymius, kuriais galėtų pasinaudoti išpuolio vykdytojas, kad paskelbtą lentelę susietų su kitu, papildomu, šaltiniu; kiekvienoje grupėje turėtų būti ne mažiau kaip *k* vienodų derinių, sudarytų iš apibendrintų požymių (t. y. kiekviena grupė turėtų būti požymių lygiavertiškumo klasė). Duomenų rinkiniai turėtų būti skelbiami tik padalyti į tokias vienodas grupes. Apibendrintini požymiai literatūroje vadinami kvaziidentifikatoriais, nes, tiksliai žinant šiuos požymius, būtų galima tiesiogiai nustatyti duomenų subjektų tapatybę.

Daugybė tapatybės nustatymo eksperimentų parodė, kad netinkamai *k* anonimiškumo metodu parengtos lentelės yra neatsparios išpuoliams. Pavyzdžiui, taip gali būti, kai kiti lygiavertiškumo klasei priskirti požymiai yra vienodi (kaip tai yra lygiavertiškumo klasės, kuriai priskirti duomenų subjektai iš Ispanijos, atveju, nurodytu A.2 lentelėje pateiktame pavyzdyje) arba jų pasiskirstymas labai netolygus dėl akivaizdaus tam tikro požymio vyravimo arba kitų priežasčių, kai lygiavertiškumo klasei priskirtų įrašų yra labai mažai; taigi abiem atvejais galimas tikimybinis išvados darymas. Arba kai lygiavertiškumo klasėms priklausantys aiškūs požymiai reikšminiu požiūriu skiriasi nedaug (pvz., kiekybiniai šių požymių matai gali būti iš esmės skirtingi, tačiau skaitinės vertės – labai panašios, arba jie gali priklausyti reikšminiu požiūriu panašių požymių grupei, pvz., tai pačiai kredito rizikos grupei arba tai pačiai ligų grupei), iš duomenų rinkinio vis dar gali būti nutekinta daug informacijos apie duomenų subjektus, kurią bus galima panaudoti per susiejimo išpuolius[31].

[30] Sweeney, L. *Weaving Technology and Policy Together to Maintain Confidentiality*. Journal of Law, Medicine & Ethics, 25, nos. 2&3 (1997), p. 98–110.

[31] Reikia pabrėžti, kad sąryšius taip pat galima nustatyti duomenų įrašus sugrupavus pagal požymius. Jeigu duomenų valdytojas žino sąryšių, kuriuos nori patikrinti, pobūdį, jis gali pasirinkti tinkamiausius požymius. Pavyzdžiui, „Pew“ tyrimų centro rezultatai nėra pažeidžiami išpuolių siekiant gauti tikslių išvestinių duomenų

Šiuo požiūriu svarbu atkreipti dėmesį, kad visais atvejais, kai duomenys yra išsklaidyti (pvz., geografinėje vietovėje tam tikra savybė būdinga tik keletui asmenų) ir kai, atliekant pirminį duomenų agregavimą, jų negalima sugrupuoti taip, kad skirtingos savybės kartotųsi pakankamai dažnai (pvz., geografinėje vietovėje keletas savybių gali kartotis retai), norint pasiekti reikiamą nuasmeninimo lygį, gali prireikti atlikti papildomą agregavimą.

l įvairovė

Remiantis pirmiau aptartais teiginiais, ilgainiui buvo pasiūlyti įvairūs *k* anonimiškumo metodai ir sudaryti rengimo kriterijai, skirti apibendrinimo principu pagrįstiems praktiniams nuasmeninimo būdams patobulinti ir sumažinti susiejimo išpuolių riziką. Šie metodai pagrįsti duomenų rinkinių tikimybinėmis savybėmis. Kalbant konkrečiau, įtrauktas papildomas apribojimas, pagal kurį kiekvienas lygiavertiškumo klasei priskiriamas požymis turi pasikartoti ne mažiau kaip *l* kartų, kad išpuolio vykdytojui visada liktų didelių abejonių dėl požymių, net jeigu jis turėtų bendrųjų žinių apie tam tikrą duomenų subjektą. Tai reiškia, kad duomenų rinkinyje (arba skaidinyje) pasirinkta savybė turėtų pasikartoti ne mažiau nei tam tikrą būtiną skaičių: taip galima sumažinti pakartotinio tapatybės atkūrimo riziką. Toks yra *l* įvairovės metodu pagrįsto praktinio nuasmeninimo būdo tikslas. Šio praktinio būdo pavyzdys pateiktas A.4 lentelėje (pirminiai duomenys) ir A.5 lentelėje (tvarkymo rezultatas). Akivaizdu, kad A.4 lentelėje, tinkamai parinkus vietovės kodo ir asmenų amžiaus nurodymo būdą, atlikus požymių apibendrinimą, gerokai padidės neaiškumas dėl kiekvieno tyrime dalyvavusio duomenų subjekto tikrųjų požymių. Pavyzdžiui, net jeigu išpuolio vykdytojas žinos, kad duomenų subjektas priklauso pirmajai lygiavertiškumo klasei, jis negalės tvirčiau nustatyti, kuri savybė – X, Y ar Z – būdinga asmeniui, jeigu tai klasei (ir visoms kitoms lygiavertiškumo klasėms) bus priskirtas bent vienas tokių savybių įrašas.

Eilės numeris	Vietovės kodas	Amžius	Savybė
1	111	38	X
2	122	39	X
3	122	31	Y
4	111	33	Y
5	231	60	Z
6	231	65	X
7	233	57	Y
8	233	59	Y
9	111	41	Z
10	111	47	Z
11	122	46	Z
12	122	45	Z

A.4 lentelė. Asmenų, sugrupuotų pagal vietovę, amžių ir tris savybes – X, Y ir Z, sąrašas

požiūriu ir kartu vis dar yra labai tinkami demografinių aspektų ir interesų sąryšiams nustatyti (http://www.pewinternet.org/Reports/2013/Anonymity-online.aspx).

Eilės numeris	Vietovės kodas	Amžius	Savybė
1	11*	< 50	X
4	11*	< 50	Y
9	11*	< 50	Z
10	11*	< 50	Z
5	23*	> 50	Z
6	23*	> 50	X
7	23*	> 50	Y
8	23*	> 50	Y
2	12*	< 50	X
3	12*	< 50	Y
11	12*	< 50	Z
12	12*	< 50	Z

A.5 lentelė. Taikant *l* įvairovės metodą sutvarkytos A.4 lentelės pavyzdys

t tankis

Ypatingam atvejui, kai požymiai skaidinyje pasiskirstę netolygiai arba kai požymių verčių ar semantinių reikšmių intervalas yra nedidelis, yra skirtas metodas, vadinamas *t* tankio metodu. Tai patobulintas nuasmeninimo pagal apibendrinimo principą metodas, kurį taikant duomenys sutvarkomi taip, kad būtų sudarytos lygiavertiškumo klasės, kuo labiau atspindinčios pirminį požymių pasiskirstymą pirminiame duomenų rinkinyje. Šiuo tikslu taikoma dviejų veiksmų procedūra, kuri iš esmės aprašyta toliau. Pirminę duomenų bazę sudarančioje A.6 lentelėje pateikti tikslūs duomenų subjektų duomenys, sugrupuoti pagal vietovę, amžių, darbo užmokestį ir dvi reikšminiu požiūriu panašias savybes – atitinkamai (X1, X2, X3) ir (Y1, Y2, Y3) (pvz., tai gali būti panašios kredito rizikos klasės, panašios ligos). Visų pirma lentelei pritaikomas *l* įvairovės metodas, kai *l* = 1 (A.7 lentelė), įrašus sugrupuojant į reikšminiu požiūriu panašias lygiavertiškumo klases ir ne iki galo pasiekiant reikiamą nuasmeninimo lygį; tada lentelė sutvarkoma siekiant kiekviename skaidinyje užtikrinti *t* tankį (A.8 lentelė) ir didesnį kintamumą. Atlikus antrą veiksmą, į kiekvieną lygiavertiškumo klasę įtraukiami įrašai iš abiejų savybių grupių. Verta paminėti, kad vietovės kodo ir amžiaus duomenų masteliai įvairiais procedūros etapais yra skirtingi: tai reiškia, kad, norint pasiekti reikiamą nuasmeninimo lygį, kiekvienam požymiui gali prireikti taikyti skirtingus apibendrinimo kriterijus, o tam savo ruožtu duomenų valdytojai turės specialiai pasirengti ir prisiimti atitinkamą skaičiavimo naštą.

Eilės numeris	Vietovės kodas	Amžius	Darbo užmokestis	Savybė
1	1127	29	30 000	X1
2	1112	22	32 000	X2
3	1128	27	35 000	X3
4	1215	43	50 000	X2
5	1219	52	120 000	Y1
6	1216	47	60 000	Y2
7	1115	30	55 000	Y2
8	1123	36	100 000	Y3
9	1117	32	110 000	X3

A.6 lentelė. Asmenų, sugrupuotų pagal vietovę, amžių, darbo užmokestį ir dvi savybių grupes, sąrašas

Eilės numeris	Vietovės kodas	Amžius	Darbo užmokestis	Savybė
1	11**	2*	30 000	X1
2	11**	2*	32 000	X2
3	11**	2*	35 000	X3
4	121*	> 40	50 000	X2
5	121*	> 40	120 000	Y1
6	121*	> 40	60 000	Y2
7	11**	3*	55 000	Y2
8	11**	3*	100 000	Y3
9	11**	3*	110 000	X3

A.7 lentelė. Taikant *l* įvairovės metodą sutvarkyta A.6 lentelė

Eilės numeris	Vietovės kodas	Amžius	Darbo užmokestis	Savybė
1	112*	< 40	30 000	X1
3	112*	< 40	35 000	X3
8	112*	< 40	100 000	Y3
4	121*	> 40	50 000	X2
5	121*	> 40	120 000	Y1
6	121*	> 40	60 000	Y2
2	111*	< 40	32 000	X2
7	111*	< 40	55 000	Y2
9	111*	< 40	110 000	X3

A.8 lentelė. Taikant *t* tankio metodą sutvarkyta A.6 lentelė

Reikia aiškiai pasakyti, kad kartais duomenų subjektų požymių apibendrinimo tikslas taikant tokius mokslinius būdus gali būti pasiekiamas ne visų įrašų, o tik nedidelio jų skaičiaus atžvilgiu. Taikant gerosios patirties metodus, turėtų būti užtikrinta, kad kiekvienai lygiavertiškumo klasei būtų priskirta daug asmenų ir kad nebeliktų galimybės vykdyti išvados padarymo išpuolių. Bet kuriuo atveju, taikydami šį metodą, duomenų valdytojai privalo nuodugniai įvertinti turimus duomenis ir atlikti kombinatorinį įvairių alternatyvų vertinimą (pvz., įvertinti įvairius intervalų dydžius, įvairius vietovių arba amžiaus duomenų mastelius ir t. t.). Kitaip tariant, nuasmeninimas apibendrinimo principu negali būti negrabaus duomenų valdytojų pirmojo mėginimo analitines įrašo požymių vertes pakeisti intervalais rezultatas, nes būtina taikyti tikslesnius kiekybinius metodus, pvz., įvertinti požymių entropiją kiekviename skaidinyje arba pirminio požymių pasiskirstymo ir pasiskirstymo kiekvienoje lygiavertiškumo klasėje skirtumus.