CYBERNETICA

Isikuandmete kaitse delikaatsetes registrites

Jan Willemson, Arne Ansper, Monika Oit

Versioon: 1.0

Sisukord

1 Sissejuhatus

Eestis on arvukalt riiklikke andmekogusid, milles sisaldub palju informatsiooni riigi elanike kohta. Koos erinevate ametkondade asjaajamise muutumisega üha elektroonilisemaks on järjest rohkemaid neist andmekogudest võimalik kasutada elektrooniliselt. Arvutipõhine andmehaldus võimaldab elanikele ja ettevõtetele pakkuda ka täiesti uuelaadseid teenuseid ning muuta olemasolevaid teenuseid käideldavamaks. Heaks näiteks selles valdkonnas on Eesti riiklik andmekogude ristkasutuse süsteem X-tee.[1]

Samas käivad suurema mugavusega kahjuks kaasas ka suuremad ohud. Digitaalinformatsiooni kaitsmise meetodid on võrdlemisi uued ning kasutajale raskemini hoomatavad kui tavalised füüsilised turvameetmed. Samuti on arvutis failide kujul hoitavate andmete kopeerimine võimalik kiiresti, ilma kvaliteeti kaotamata ning jälgi jätmata. Varastatud elektrooniliste andmekogude töötlemine (näiteks erinevate andmekogude üheks suuremaks kokkupanemine) käib väga efektiivselt ja nõnda saab erinevaid registreid kuritahtlikult kombineerides väga paljude inimeste kohta panna kokku küllalt täieliku profiili. Viimases sisaldub aga peaaegu kõigi meie kohta midagi delikaatset, mida sobivas olukorras meie vastu ära kasutada võib.

Andmete konfidentsiaalsusega peaagu sama tundlik on ka andmekogude tervikluse tagamise probleem. Ekslik sattumine mõnesse tervisevaldkonna infosüsteemi võib inimesele jätta külge märgi haigusest, mida tal tegelikult pole; sinna kord põhjusega sattunule on tema kohta talletatud andmete õigsus aga tihti elulise tähtsusega.

Käesolevas analüüsis käsitleme delikaatseid isikuandmeid sisalvadate registrite turvaprobleeme ning analüüsime mõnesid väljapakutud lahendusi. Arvestades, et üheks kõige delikaatsemaks inimest puudutavaks infoliigiks on teave tema tervise kohta, keerleb kogu järgnev paljuski just tervishoiualaste infosüsteemide turvalisuse ümber.

2004. aastal võeti vastu Eesti Vabariigi Tervise infosüsteemi arengukava aastateks 2005–2008 [1]. Selle arengukava 4. peatükk käsitleb erinevate tervisevaldkonna elektrooniliste registrite sisseviimisega seotud probleeme ja seda nii õiguslikust, organisatsioonilisest kui infotehnoloogilisest aspektist. Üheks raskemaks õiguslikuks probleemiks osutub asjaolu, et mitmed meditsiinilise teabe andmebaasid sisaldavad väga delikaatseid andmeid, isegi sel määral, et juba vastavasse baasi kuulumise fakti tuleks salastada (arengukava lisast 2 leiame näiteks HIVi andmekogu, narkomaaniaravi registri jt). Põhilise meetmena nähakse arengukava peatükis 4.3.3 ette nn kodeerimiskeskus(t)e loomine, mille ülesandeks on isikut identifitseerivate andmete asendamine

[1] `http://x-tee.riik.ee/`

koodidega nõnda, et vajadusel oleks võimalikud ka tagasiteisendused.

Kodeerimiskeskuse idee on arengukavas esitatud valmismeetmena, kuid samas ilma analüüsita, milliseid riske ta täpselt maandama peaks. Seetõttu ongi käesoleva uuringu esimene eesmärk koostada delikaatseid isikuandmeid sisaldavate registrite turvalisuse analüüsi jaoks sobiv metoodika ning teha selle valguses kindlaks, milliste ohtude vastu kodeerimiskeskus tegelikult aitab.

Üheks esimestest arengukavas [1] mainitud andmekogudest, mida Sotsiaalministeeriumi terviseosakonnas reaalselt looma hakatakse, osutus narkomaaniaravi register/andmekogu (edaspidi narkoregister), mida on kirjeldatud seaduses [3] ja määruse eelnõus [2]. Loomulikult on Eestis ka teisi ametkondi, kes delikaatseid isikuandmeid sisaldavaid andmebaase haldavad ning kasutavad. Käesoleva uurimuse teiseks eesmärgiks on kaardistada niisuguste registrite haldamise praktika ja uurida, kas kodeerimiskeskus võimaldaks neis kuidagi turvalisust tõsta; kaardistuse tulemused võtab kokku peatükk 2. Eraldi alapeatükina 2.3 käsitleme Sotsiaalministeeriumi tööhõiveosakonnas kavandatavat tööpoliitika statistika infosüsteemi, mida küll veel ei eksisteeri, kuid mille juures Sotsiaalministeerium on samuti avaldanud soovi kodeerimiskeskust kasutada.

Käesoleva uurimuse kolmas eesmärk on anda ülevaade sellest, miks ja kuidas erinevaid delikaatseid registreid peetakse. Peatükis 3 käsitleme registripõhiste uuringute põhimõtteid ja võrdleme välisriikide kogemusi Eesti hetkepraktikaga.

Peatükis 4 vaatleme registrite turvaanalüüsiks sobivat metoodikat ja peatükis 5 tutvume lähemalt kodeerimiskeskuse lahendusega. Saadud tulemusi rakendame peatükis 6 narkoregistri näitel. Peatükis 7 teeme kokkuvõtteid ning anname soovitused delikaatseid isikuandmeid sisaldavate registrite turvamiseks.

Muuhulgas uurime kogu analüüsis läbivalt, milliseid meetmeid Eestis hetkel kehtiv seadusandlus registrite turvamiseks kasutada käsib/lubab. Paraku osutub, et mitmeski sättes pole meie seadused ja määrused andmeturbe seisukohast optimaalsed. Sellest johtuvalt võib järelduste seast leida kahte liiki soovitusi – ühed, mis teevad ettepanekuid olukorra parandamiseks läbi seadusemuudatuste, ja teised, mis üritavad kehtivas õigusruumis parimat lahendust leida. Arusaadavalt ei pruugi need soovitused omavahel alati kooskõlas olla.

2 Ülevaade olemasolevatest delikaatsetest registritest

Eestis on arvukalt erinevate ametkondade poolt peetavaid registreid ja paljud neist sisaldavad andmeid, mida võib suuremal või vähemal määral tundlikeks pidada. Isikuandmete kaitse seadus [6] loetleb eraelulistena näiteks isiku perekonnaelu ja sotsiaalse staatuse kohta käivaid andmeid, delikaatsetena aga poliitilisi vaateid, terviseseisundit, kriminaalmenetluse infot jm. Kuna kõiki registreid polnud uuringuks eraldatud ajaliimidi piires kaugeltki võimalik käsitleda, tegime valiku ametkondlikul printsiibil ning intervjueerisime kolme riigiasutust, kus delikaatsete isikuandmetega kõige rohkem kokku puututakse. Valimisse jäid ning käesolevas peatükis leiavad käsitlemist Justiitsministeerium, Siseministeeriumi Politseiamet ja Sotsiaalministeeriumi tööhõiveosakond. Sotsiaalministeeriumi terviseosakonna probleeme vaatleme narkoregistri näitel eraldi 6. peatükis.

2.1 Justiitsministeerium

Justiitsministeeriumi haldusalas on mitu infosüsteemi ja registrit, näiteks:

- kinnipeeturegister (ametliku nimega *riiklik kinnipeetavate, arestialuste ja vahistatute register*) ehk VANGIS, kuhu kantakse kõik, kes on kordki sattunud arestimajja või vanglasse (v.a. kaineri kliendid ja 48ks tunniks kinni peetud);
- kriminaalhoolduse infosüsteem ehk KHIS, kuhu kantakse kõik, kes on kordki sattunud kriminaalhooldamise alla;
- kohtulahendite register ehk KOLA, mida küll alates 1.01.2006 enam ei täiendata, sest selle asemele tuleb Kohtuinfosüsteem ehk KIS;
- riskihindamise süsteem, mida hetkel alles kavandatakse, hakkab sisaldama täiendavat üsna mitteformaalset infot inimese ohtlikkuse kohta (kas ta võib tänaval kellelegi kallale tungida vms);
- kriminaalmenetluse register.

Mitmetes neist registritest on ka delikaatseid isikuandmeid, näiteks viimane elukoht enne kinnipidamist, usutunnistus jm, vt nt VANGISe andmekoosseisu määrusest [4].

Erinevalt Sotsiaalministeeriumi vajadusest kasutada registreid põhiliselt statistika saamiseks ja poliitika kujundamiseks, on Justiitsministeeriumi andmebaasid praktiliselt kõik operatiivkasutuses. Isikustatud andmetele pääseb oma igapäevaste tööülesannete tõttu ligi võrdlemisi lai isikute ring.

Näiteks VANGISel on suurusjärku 500-1000 kasutajat (sotsiaaltöötajad, julgeolekutöötajad, vanglaametnikud) ja kõik neist näevad kõigi vangide nimesid. Samas on realiseeritud rollipõhised õigused ja vangide muid andmeid näevad töötajad vastavalt oma töökohustustest tulenevale vajadusele (meditsiinitöötajad terviseinfot jne).

KHISil on umbes 200 kriminaalhooldajast kasutajat, igaühel keskmiselt 40 hoolealust. Ka KHIS on organiseeritud rollipõhiselt, iga hooldaja näeb ainult nende isikute andmeid, kellega ta vahetult töötab.[2]

VANGISel ja KHISil on ka ristkasutus, nt selleks, et kriminaalhooldajad saaksid vaadata, millist tööd kinnipeetavatega vanglas tehtud on. Isikute arv, kes mõlema andmebaasi kõiki kirjeid näevad, on alla 10 (jällegi rollipõhiselt) ning ristpäringuid saab esitada füüsiliselt turvatud arvutist.

Rõhuv enamus Justiitsministeeriumi andmebaaside kasutusjuhte on ministeeriumisisesed. Ametkonnast välja võib andmeid viia ainult Kaitsepolitsei, ametkonnavälistest baasidest tuuakse infot sisse ainult rahvastikuregistrist ja Politseiameti käes olevast karistusregistrist.

Justiitsministeeriumi registrid kasutavad küll kinnipeetute identifitseerimiseks isikukoodist sõltumatuid koode, kuid seda eeskätt põhjusel, et mitte kõik kinnipeetud pole Eesti Vabariigi seaduslikud residendid (mõnedel pole isegi mitte ühegi riigi passi) ja kuidagi tuleb ju inimesele viidata. Lisaks on Justiitsministeeriumi töö spetsiifika selline, et ametnikele peavad andmed olema nähtavad isikustatud kujul. Sellest johtuvalt oleks kodeerimiskeskuse laadse lahenduse sisseviimine Justiitsministeeriumi haldusalas olevatesse registritesse kunstlik ja takistaks tööd rohkem kui saadav privaatsusevõit väärt oleks. Hetkel on andmeturberiskid maandatud rollipõhiste pääsuõiguste ning füüsiliste ja organisatsiooniliste meetmete rakendamisega ning need meetmed on pälvinud ka Andmekaitse Inspektsiooni heakskiidu.

2.2 Politseiamet

Politseiameti hallata on kaks põhilist registrit – karistusregister ja Politsei infosüsteem (PolIS).

[2]Siinkohal on kohane märkida, et pea kõik Justiitsministeeriumi registrid on paberkandjal dubleeritud ja vahel erinevad elektroonilised pääsuõigused füüsilistest. Näiteks KHISi pabertoimikud on jaotatud piirkonniti ja nt üks Tartu kriminaalhooldaja pääseb põhimõtteliselt ligi kõigi Tartu piirkonna 500 hooldatava andmetele.

Karistusregistris, nagu nimigi ütleb, hoitakse andmeid kõigi karistuste kohta. Kuna põhiseaduse järgi on kohtu otsused avalikud, on tegelikult avalik ka osa karistusregistri infot. Samas sisaldub seal ka karistusi, mis pole määratud kohtus ja neile lai avalikkus väga ligi ei peaks pääsema. Reaalselt on karistusregister ühendatud X-teega ja näiteks saab sealt infot küsida Majandus- ja kommunikatsiooniministeerium (MKM). Millised füüsilised isikud konkreetselt läbi MKMi andmeid kätte saavad, sellest Politseiametil täpne ülevaade puudub, aga nende isikute arvu hindasid Politseiameti töötajad tuhandetesse.

Karistusregistri ajaloo jooksul on olemas üks juhtum, kus Politseiametist läks Justiitsministeeriumisse välja oluline osa andmebaasist, kust isikuandmed olid küll eemaldatud, jättes alles vaid karistuste info. Andmed anti välja täie teadmisega, et karistuste iseloomu ja aja järgi on võimalik isikuid tagasi identifitseerida ning see oli kaalutletud risk. Politseiamet ei välista sarnaseid andmete väljastamisi ka edaspidi, kuid nende üle otsustatakse igal konkreetsel juhul eraldi.

PolIS sisaldab endas kriminaalmenetluste andmeid, mis on delikaatsed, seega PolIS-t ei saa lugeda nii avalikuks andmekoguks kui karistusregistrit. PolIS-esse võib inimene sattuda ka poolkogemata, näiteks pannakse sinna kirja kuriteost teatanud inimeste andmed (juhul kui teataja on nõus neid teatamisel avalikustama). PolIS-t kasutab oma töös üle poole politseis ametis olevatest inimestest, suurusjärku 2000 . . . 3000 inimest.

PolIS on seotud teiste andmebaasidega, näiteks võtab ta andmeid Autoregistri Keskusest ja annab andmeid Justiitsministeeriumi kriminaalmenetluse registrisse. PolIS-est väljastatakse ka statistikat (nt ajakirjandusele). Selle valmistab ette volitatud Politseiameti töötaja, ajakirjanikke PolIS-e kallale ei lasta.

Kuna Politseiameti registrid on operatiivkasutuses ja paljudel töötajatel on vaja isikustatud andmeid, siis mingi isikukoodide kodeerimine ennast ei õigusta. Pigem on politseil andmete varjamisega vastupidine probleem – tuleb kuidagi identifitseerida inimesi, kelle isikukood ei ole teada või kellel polegi mingit mõistlikku isikukoodi. Sellises olukorras osutub tegelikult väga kasulikuks see, kui inimest saab tuvastada teisaste andmete kaudu (auto, vanus, rahvus jms).

Isikustatud andmete ligipääsetavus laiale kasutajate ringile toob loomulikult kaasa probleeme – mitte kõik politseiametnikud pole väga ausad. On esinenud ka infolekkeid, millega tegeletakse *post factum* sisejuurdluse korras. Reeglina lõppeb asi distsiplinaarkaristuse või vallandamisega, aga juba on olemas ka kriminaalasja kaasus.

2.3 Sotsiaalministeeriumi tööhõiveosakond

Sotsiaalministeeriumil on rida tööhõive ja sotsiaalprobleemidega seotud allasutusi, kellest kõigil on olemas oma operatiivandmebaasid. Nende asutuste hulka kuuluvad

- Tööturuamet,
- Töötukassa,
- Sotsiaalkindlustusamet ja
- Tööinspektsioon.

Riigi tööhõivepoliitika analüüsiks, parandamiseks ja paremaks kujundamiseks on Sotsiaalministeeriumis valminud plaan luua tööpoliitika statistika infosüsteem. Sisuliselt on tegu nelja ülalloetletud allasutuse registrite linkimisega[3], mille väljundina oodatakse statistilisi aruandeid.

Võrreldes vaadeldavate registrite turvataset näiteks politsei infosüsteemiga võib öelda, et tööhõiveinfo on oluliselt vähem tundlik kui näiteks teave kriminaalmenetluste kohta. Teisest küljest peab lingitud tööpoliitika statistika infosüsteemile ligi pääsema vaid küllalt piiratud isikute ring. Siit võib järeldada, et kui PolISe kaitsmiseks piisab rollipõhistest pääsuõigustest ja vastavast õiguste kontrollimise infrastruktuurist, siis peaks need tagama piisava turvalisuse ka tööpoliitika statistika infosüsteemi jaoks.

Sotsiaalministeeriumi tööhõivespetsialistid soovivad soovivad aga rakendada rangemaid turvameetmeid ja kustutada oma operatiivregistritest isikukoodid, asendades need kodeerimiskeskuse abil asenduskoodidega (vt peatükk 5). See soov tundub samas põhjendamatu, sest intervjuudes vastavate asjatundjatega ei selgunud ühtki ohustsenaariumi, mille vastu nii keeruka meetme abil võidelda loodetakse. Ainsa ründena mainis spetsialist Siiri Otsmann ebatervet huvi, mida analüütikus isikute identiteetide nägemine esile võib kutsuda. Selle riski saab aga maandada ilma igasuguse koderimiskeskuseta, keelates infosüsteemi tasemel analüütiku rollil nimesid ja isikukoode näha. Seega oleks otstarbekas ehitada tööpoliitika statistika infosüsteem üles rollipõhisena sarnaselt Justiitsministeeriumi ja Politseiameti ülalkirjeldatud infosüsteemidele.

[3]Linkimine ei ole standardne andmekogude käitlemise alane mõiste, aga seda kasutatakse praktikas (sh Sotsiaalministeeriumis) küllalt sageli. Linkimise all peetakse silmas kahest või enamast andmebaasist samade isikute kohta käivate andmete kokkuviimist ning ühendandmebaasi moodustamist. X-tee mõistes on tegu komplekspäringuga.

2.4 Kokkuvõte

Justiitsministeeriumis, Politseiametis ja Sotsiaalministeeriumi tööhõiveosakonnas tehtud intervjuude tulemuste põhjal võib öelda, et neis ametkondades ei ole registreid, millele isikukoodide varjamine midagi juurde annaks. Osade andmekogude puhul välistab selle nende operatiiviseloom, teiste puhul on olemasolevad riskid võrdlemisi väikesed ning maandatavad infosüsteemi pääsuõiguste süsteemi rollipõhise ülesehitusega.

3 Registripõhised uuringud

Terviseandmeid kogutakse statistilise ja teadusuuringute eesmärgil praktiliselt igas riigis. Kuna käesoleva ülevaate ressursid on piiratud, piirdume siinkohal põhiliselt meile kultuuritraditsiooniliselt lähedaste Põhjamaade kogemuse kirjeldusega. Delikaatse terviseinfo kasutamise praktikat nii Põhjamaades kui ka kaugemal on oma artiklites [12, 13, 14] põhjalikult käsitlenud Tervise Arengu Instituudi professor Mati Rahu. Lisainfot terviseuuringute kohta Soomes ja Rootsis võib leida vastavalt artiklist [15] ning ülevaatebrošüürist [16]. Järgnev tugineb märgatavas osas Mati Rahu eelpoolviidatud artiklitele.

Infot kogutakse reeglina mingi eesmärgiga ja terviseinfo kogumise eesmärk on analüüsida rahvastiku tervise hetkeolukorda ning trende, ennustada tervishoiukulutuste vajadust, hinnata senikasutatud meetmete efektiivsust jne. On selge, et mida kõrgema kvaliteediga andmestikku kasutatakse analüüside alusena, seda tõhusamaks kujunevad epidemioloogiliste uuringute tulemusena planeeritud tervishoiuteenused.

Laias laastus võib epidemioloogilised uuringud läbiviimise metoodika järgi jagada kahte klassi.

1. Igalt uuringus osalevalt inimeselt küsitakse eraldi nõusolekut tema andmete kogumiseks ja kasutamiseks. Registreid kas ei peeta üldse või on nende andmestik teadaolevalt puudulik.

2. Uuringuteks kasutatakse eelkogutud registreid, kuhu andmeandjatel (arstidel) on kohustus raporteerida kõik juhtumid; inimese nõusolekut registrisse kuulumiseks või tema kohta käivate andmete kasutamiseks ei küsita.

Arvestades terviseandmete delikaatsust tundub esimesel pilgul, et inimeste privaatsuse tagamisel saavad kõne alla tulla ainult esimese klassi meetodid. Samas on nendel meetoditel rida puudusi.

- Kõigi inimeste käest eraldi nõusoleku küsimine nõuab palju ressursse, seda eriti juhul kui juba andmekogu olemasolu loetakse nii suureks turvariskiks, et seda kusagil ei peetagi. Sel juhul on näiteks mingist haigusest ülevaate saamiseks sisuliselt ainuke võimalus teha ühekordne uuring ja proovida kuidagi leida kõik selle haiguse all kannatajad.

- Isegi kui mingi seisuga õnnestus kokku panna enamvähem adekvaatne ülevaade rahvastiku hetketervisest, ei uuendata esimest tüüpi lähenemise korral seda ülevaadet uute andmete lisandumisel. See viib andmestiku kiirele vananemisele ja siis, kui vastavaid andmeid järgmine kord vaja läheb, tuleb nad uuesti koguda.

- Kui oma andmete esitamine epidemioloogiliste uuringute jaoks on vabatahtlik, puudub meil igasugune võimalus uuringutulemuste põhjal reaalset olukorda ligilähedaseltki hinnata. Esimest tüüpi uuringuid viiakse kulude kokkuhoidmiseks sageli läbi postiküsitluste vormis ja sel juhul on potentsiaalsel andmesubjektil eriti lihtne küsimustikku ignoreerida.

- Olukorras, kus inimene peaks iga uuringu jaoks eraldi nõusoleku andma, on väga raske tagada uuringute järjepidevust. Inimene võib mitmekordsetest küsitlustest tüdineda, otsustada uuringust loobuda just siis, kui tema juhtum muutub huvitavaks (nt kui ta nakatub HIV-sse) jne.

Vaatamata loetletud puudustele on sääraseid uuringuid maailmas tehtud küllalt palju. Näiteks Suurbritannias värvati inimesi toidupoodides, haiglates ja mitmetes ühingutes, et kätte saada ligi 11000 taimetoitlast, keda hiljem 17 aastat jälgiti. Sarnaseid aktsioone on ette võetud ka Saksamaal jt suurriikides.

Väikestes riikides (sh kõigis Põhjamaades) on aga valdavalt kasutuses teist liiki, registripõhine terviseuuringute metoodika. Põhjus on lihtne – uuringud, kus iga kord tuleb andmesubjekti käest eraldi luba küsida, on väga kallid. Samuti tuleb arvesse võtta, et olukorras, kus piiratud ressursside tingimustes pole võimalik saavutada esindavat valimit, on ka need vähesed vahendid kulutatud asjatult.

Loomulikult tähendab süsteem, mille korral patsiendi terviseandmed edastatakse peale arsti külastust automaatselt kesksesse registrisse, potentsiaalselt inimese privaatsusele suuremat riski. Samas on seda riski võimalik analüüsida ja hallata. Mitmed ühiskonnad on just seda teed läinud ning seal käsitletakse privaatsuse kadu kaalutletud kompromissina, mis lõppkokkuvõttes annab läbi täpsemate uuringutulemuste ja paranenud ravimeetodite kasu ka uuringualustele inimestele endile.

Loomulikult tuleb tekkivate riskide maandamiseks terviseandmetele ligipääs täpselt reguleerida. Selles osas on Euroopa Liidus infoturbelahenduste aluseks direktiiv 95/46/EC [9], mis sätestab delikaatsete andmete kogumise, hoidmise ja töötlemise põhimõtted. Direktiiv võimaldab muuhulgas delikaatsete isikuandmete kasutamist teadustöö eesmärkidel ilma andmesubjekte eraldi teavitamata. Nii ütleb direktiiv järgmist:

> ***Artikkel 7***
>
> *Liikmesriigid sätestavad, et isikuandmeid võib töödelda ainult juhul, kui:*
>
> *. . .*
>
> *d) töötlemine on vajalik andmesubjekti eluliste huvide kaitsmiseks või*
>
> *e) töötlemine on vajalik üldiste huvidega seotud ülesande täitmiseks või sellise avaliku võimu teostamiseks, mis on tehtud ülesandeks volitatud töötlejale või andmeid saavale kolmandale isikule, või*
>
> *. . .*
>
> ***Artikkel 8*** *Andmete eriliikide töötlemine*
>
> *. . .*
>
> *3. Lõiget 1 ei kohaldata, kui andmete töötlemine on vajalik ennetava meditsiini, meditsiinilise diagnoosi, meditsiinilise abi või ravi võimaldamise või tervishoiuteenuste juhtimise jaoks ja kui kõnealuseid andmeid töötleb tervishoiutöötaja, kelle puhul kehtib siseriiklikus õiguses või pädevate siseriiklike ametiasutuste kehtestatud eeskirjades sätestatud ametisaladuse hoidmise kohustus, või mõni teine isik, kelle suhtes kehtib samaväärne saladuse hoidmise kohustus.*
>
> *4. Võttes arvesse sobivaid tagatisi, võivad liikmesriigid põhjustel, mis on seotud märkimisväärse avaliku huviga, siseriikliku õiguse või järelevalveasutuse otsusega kehtestada täiendavaid erandeid lisaks lõikega 2 ettenähtud eranditele.*

Lisaks neile sätetele on direktiivi preambulas teadusuuringuid eraldi mainitud, öeldes, et isikuandmete töötlemist ajaloo, statistika või teadusega seotud eesmärkidel ei peeta andmete kogumisega vastuolus olevaks ning et rahvatervishoid ja sotsiaalkaitse kuuluvad oluliste avalike huvide hulka:

> *29) isikuandmete täiendavat töötlemist ajaloo, statistika või teadusega seotud eesmärkidel ei peeta tavaliselt andmete kogumise esialgse eesmärgiga vastuolus olevaks tingimusel, et liikmesriigid kannavad hoolt vajalike tagatiste eest; nimetatud tagatistega tuleb eelkõige kindlustada, et andmeid ei kasutataks ühegi konkreetse isikuga seotud meetmete või otsuste toetamiseks;*
>
> ...
>
> *34) liikmesriikidel peab olema lubatud kalduda kõrvale delikaatsete andmete töötlemise keelust ka siis, kui seda õigustavad olulised üldiste huvidega seotud põhjused sellistes valdkondades nagu tervishoid ja sotsiaalkaitse (eriti selleks, et tagada hüvitisenõuete rahuldamise kord ja teenuste kvaliteet ning tasuvus tervisekindlustuse puhul), teadusuuringud ja riiklik statistika; liikmesriikidel on siiski kohustus pakkuda konkreetseid ja sobivaid tagatisi, et kaitsta üksikisikute põhiõigusi ja eraelu puutumatust;*

Selle direktiivi alusel on valminud ka näiteks Taani, Rootsi, Soome ja Eesti vastavad seadusandlikud aktid.

Vaadeldavas kontekstis on Põhjamaade ja Eesti seadusandluse vahel aga üks oluline erinevus – Põhjamaades on privaatsuse rikkumine teadustöö eesmärgil selgesõnaliselt lubatud, Eestis aga mitte. Nii näiteks ütleb Taani isikuandmete töötlemise seadus [10] §32 (4), et sama seaduse §31 nõuet andmesubjekti informeerimise kohta tema andmete kasutamisest ei tule täita teadusuuringute korral.[4]

Rootsi isikuandmete seaduse [11] paragrahv 9 sätestab direktiivi 95/46/EC [9] vaimus, et vaatamata sellele, mis eesmärgil mingeid isikuandmeid korjati, tohib neid lisaks kasutada ajalooliste, statistiliste ja teaduslike uurimuste läbiviimiseks.[5]

Eestis kehtiv isikuandmete kaitse seadus [6] §14 (3) ütleb samal ajal vaid:

[4] *Section 31 (1) shall not apply where data are processed solely for scientific purposes or are kept in personal form for a period which does not exceed the period necessary for the sole purpose of creating statistics.*

Ingliskeelset tõlget vt `http://www.datatilsynet.dk/include/show.article.asp?art_id=443`

[5] *The controller of personal data shall ensure that*

> ...
>
> *d) personal data is not processed for any purpose that is incompatible with that for which the information is collected,*
>
> ...
>
> *i) personal data is not kept for a longer period than that as is necessary having regard to the purpose of the processing.*
>
> ...

> *(3) Delikaatsete ja eraeluliste isikuandmete töötlemine on ilma andmesubjekti nõusolekuta lubatud:*
>
> *1) seaduse või välislepinguga ettenähtud ülesande täitmiseks;*
>
> *2) andmesubjekti või muu isiku elu, tervise ja vabaduse kaitseks.*

Näeme, et võrreldes direktiiviga 95/46/EC on Eestis avalikes huvides uuringute tegemine vaikimisi keelatud, kui mõni seadus ei sätesta teisiti. Iga uuringu jaoks omaette seadust ilmselt vastu võtma ei hakata, reaalne oleks täiendada kõigi andmekogude aluseks olevaid seadusi eraldi või ainult isikuandmete kaitse seadust. Viimane võimalus oleks loomulikult lihtsaim.

Heaks näiteks konkreetsest kasust, mida niisuguse täienduse korral oleks võimalik saada, on Taanis läbi viidud uuring mobiiltelefoniomanike vähiriski kohta. Selles uuringus lingiti isikutepõhiselt kokku vähiregister ja mobiiltelefoniomanike andmebaas ning jõuti järeldusele, et mobiiltelefoni kasutamine isegi pika perioodi vältel ei tõsta vähki haigestumise riski.

Eestis on sarnaste uuringute legaalsus praeguses õiguslikus olukorras kahtluse all. Näiteks puudub meil 2000. aastast alates üldse usaldusväärne vähistatistika. Põhjuseks on siin riikliku statistika seadus [7], mille §9 (2) ütleb:

> *. . . Andmesubjekti identifitseerimist võimaldavaid andmeid võib riikliku statistilise vaatluse korraldaja erandkorras edastada ja avaldada ainult andmesubjekti kirjalikul nõusolekul. Andmesubjekti identifitseerimist võimaldavaid andmeid võib Vabariigi Valitsuse poolt kehtestatud korras edastada teadusuuringuteks ka andmesubjekti nõusolekuta. . . .*

Seega isegi kui mingile asutusele teadustöö eesmärgil andmeid edastatakse, ei tohi see asutus nende andmete põhjal oma registreid uuendada, sest see pole teadustöö. Nii võib vähiregistri volitatud töötleja Tervise Arengu Instituut küll küsida surmaregistrist postmortaalselt avastatud vähijuhtude infot, kuid ei tohi selle alusel vähiregistrit täiendada. Kuna vähijuhtumitest avastatakse aga ca 5% alles pärast inimese surma, siis on oluline osa vähistatistikast registrist lihtsalt puudu.

> *However, as regards the first paragraph, d), the processing of personal data for historical, statistic or scientific purposes shall not be regarded as incompatible with the purposes for which the information was collected.*
>
> *Personal data may be kept for historical, statistic or scientific purposes for a longer time than that stated in the first paragraph i). However, personal data may not in such cases be kept for a longer period than is necessary for these purposes.*

Ingliskeelset tõlget vt `http://www.datainspektionen.se/pdf/ovrigt/pul-eng.pdf`

Praeguses õiguslikus olukorras oleks ainus lahendus edastada surmajärgselt avastatud vähijuhtumite kohta teave nii surmaregistrisse kui ka vähiregistrisse. Pikemas perspektiivis tuleb aga sarnaste juhtude vältimiseks viia täiendus sisse ka riikliku statistika seadusesse ning lubada ühe registri alusel teist täiendada.

Kokkuvõtteks võime öelda, et kõrgetasemeliste uuringute aluseks on kvaliteetsed isikustatud andmekogud. Teiste maade praktika näitab, et need ühiskonnad, kus tekkivad turvariskid on aktsepteeritud ning sobilike meetmetega maandatud, lõikavad saadavatest uuringutulemustest suuremat tulu, kui kaasnev risk potentsiaalset kahju toob.

4 Registrite turvaanalüüsi metoodika

Ühest küljest pärineb kogu kodeerimiskeskuse ülesandepüstitus algselt narkoregistri juhtumist, teisest küljest aga on see register kõigist vaadeldavatest niisugune, mille puhul andmete delikaatsuse nõuded kõige selgemalt esile tulevad. Seepärast kasutame järgnevas metoodika kirjelduses just narkoregistri näidet.

Narkootiliste ja psühhotroopsete ainete ning nende lähteainete seaduse [3] §11^1 lõige 5 ütleb narkoregistri kohta nõnda:

> *Andmekogu peetakse kujul, mis ei võimalda registrisse kantud isikut tuvastada.*

Paraku on niisugust sätet seadusesse kergem kirjutada kui teda täita. Probleem seisneb selles, et isiku identifitseerimiseks on olemas ka teisi viise peale nime ja isikukoodi. Vastavalt eelnõule [2] tahetakse andmekogus statistiliste uuringute huvides säilitada väga erinevat teavet, sh inimese sugu, sünniaega (kuu täpsusega), emakeelt, rahvust, tegevusala, haridustaset, elukohta[6], infot ravi, pöördumiste ja teatiste esitamise kohta. Arvestades Eesti väikest rahvaarvu võivad need andmed kombineerituna väga sageli viia isiku võrdlemisi kerge tuvastamiseni kas üheselt või vähemalt suure kindlusega.

Vaatleme näiteks olukorda, mil ründaja saab ligipääsu andmekogule, kust on kustutatud nimed ja isikukoodid, kuid alles on jäetud inimeste vanus (kuu täpsusega), rahvus ja elukoht (valla/linna täpsusega) ning et nende väljade mingite konkreetse väärtustega kirjeid on narkoregistris k tükki. Kui ründaja suudab nüüd välja selgitada, et selles vallas/linnas elab antud vanusega antud rahvusest isikuidkokku n tükki, siis võib ta igaühe kohta neist väita,

[6]Vastavalt Sotsiaalministeeriumi spetsialisti Kristo Klesmenti sõnadele planeeritakse elukohta säilitada haldusüksuse, st linna/valla täpsusega.

et see inimene kuulub vaadeldavasse registrisse tõenäosusega $\frac{k}{n}$. Kui kehtib $k = n$ (st kõik vastava vanuserühma vastavast rahvusest vastavas kohas elavad inimesed kuuluvad narkoregistrisse), siis pole isikukoodide ja nimede kustutamisel andmete varjamise seisukohast praktiliselt mingit efekti.

Loomulikult tuleb arvestada asjaoluga, et väärtuse n leidmine ei pruugi ründajale olla niisama lihtne, kuid selleks võib olla erinevaid võimalusi. Saamaks teada, kui palju mingis haldusüksuses kindla vanusega kindlast rahvusest inimesi elab, võib ründaja proovida omandada ebaseaduslikku ligipääsu rahvastikuregistrile aga ta võib ka kohapeale sõita ja üritada kohalikke inimesi küsitledes vastavat infot kokku korjata.

Siit näeme, et üks oluline parameeter registrite turbe korral on potentsiaalse ründe hind, mis ei tohiks ületada ründe tulemusena saadavat kasu. Igast registrist on võimalik kogu seal sisalduv info kätte saada (näiteks andmekogu volitatud töötlejat ära ostes), küsimus on ainult selles, kui palju peab ründaja investeerima, kui täpse info ta ründe tagajärjel omandab ja kui suur on selle info turuväärtus. Praktikas pole ründe hinda ja info väärtust tihti sugugi lihtne leida, mistõttu me käesolevas analüüsis nende hindamisel põhinevaid meetodeid ei kasuta. Küll aga osutub meile allpool väärtuslikuks ka juba pelgalt see eeldus, et ründest saadakse mingit (rahalist) kasu.

Lisaks on selge, et kõik inimesed võib andmetele ligi pääsemise lihtsuse, kättesaadavate andmete detailsuse ja ebasihipärase kasutamise avastamise keerukuse alusel erinevatesse klassidesse jagada (nt registri volitatud töötleja usaldatud isikud, ajalehest üldstatistikat lugevad inimesed jne). Teatud olukordades võib info kättesaamiseks olla vaja ka eri klassidesse kuuluvate isikute koostööd.

Seega tuleb niisuguseid paragrahve nagu seaduse [3] ülaltsiteeritud paragrahvi kindlasti täpsustada. Vastavate seaduste või määrustega tuleb määrata vähemalt

- isikute klassid, kes omavad registrile mingil tasemel ligipääsu (iga inimene kuulub potentsiaalselt mingisse klassi);
- millise täpsusega mingisse klassi kuulujad andmeid näha võivad;
- vajalikud koostööstsenaariumid, mille alusel ründajad info omandamiseks koopereeruma peavad.

5 Kodeerimiskeskus

Nagu peatükis 4 viidatud, keelab narkootiliste ja psühhotroopsete ainete ning nende lähteainete seadus [3] narkoregistri pidamise kujul, mis võimaldab

registrisse kantud isikuid tuvastada. Kui seda sätet interpreteerida nii, et üksi osapool ei tohi ka ükskõik kellega koopereerudes saada ühegi registrisse kuuluva inimese isikut teada 0-st suurema tõenäosusega, siis pole narkoregistrit üldse võimalik pidada. On olemas kaks kasutusjuhtu, mis nõuavad registrisse kuulujate kuidagimoodi identifitseerimist – andmete uuendamine/parandamine ja linkimine teiste registritega. See tähendab, et kusagil tuleb (kasvõi mitme osapoole vahel hajutatult) hoida vahendeid registrikirjete ja füüsiliste isikute kokkuviimiseks. Ükskõik milliseid turvameetmeid me ka ei rakendaks, alati jääb kusagil alles isik või isikute rühm, kes suudab(vad) kõik registrisse kuulujad kindlaks teha.

Vestluses Sotsiaalministeeriumi spetsialistidega ei selgunud, millist realistlikku interpretatsiooni ülaltoodud sättele tegelikult silmas peeti. Küll aga jäi kõlama seisukoht, et kui registrit pidada isikustatud kujul, siis on volitatud töötlejal (näiteks narkoregistri puhul Tervise Arengu Instituudil) väga lihtne registrissekuulujate privaatsust rikkuda, mistõttu see risk tuleks kindlasti maandada.

Volitatud töötleja poolse ründe riski maandamiseks on arengukavas [1] ja määruse eelnõus [2] pakutud välja idee eemaldada andmete hulgast otseselt isikule viitavad andmed (st nimi ja isikukood). Selleks aga, et jääks alles võimalus registrit uuendada ning teiste registritega linkida, pannakse isikukoodi asemele teatav asenduskood. Asenduskoodiks võib olla näiteks isikukoodi krüptogramm, aga põhimõtteliselt ka suvaline teine identifikaator, mille vastavust isikule volitatud töötleja ei tea.

Andmete uuendamiseks ja teiste registritega linkimiseks peab mingi osapool seda vastavust mingil kujul ikkagi teadma. Niisuguseks osapooleks on arengukavas [1] ja eelnõus [2] pakutud välja nn *kodeerimiskeskus*. Tuleb tõdeda, et kuigi see idee on käesoleva kirjutamise hetkeks juba üle aasta vana, ei valitsenud ka Sotsiaalministeeriumi spetsialistide seas täielikku üksmeelt selles osas, kuidas kodeerimiskeskus(ed) täpselt töötama peaks, mis ülesandeid ta/nad täidab/täidavad ning kas neid peaks olema üks või mitu. Järgnevas analüüsime erinevaid variante, mis Sotsiaalministeeriumis läbi viidud intervjuude käigus kõlama jäid.

Kodeerimiskeskus on narkoregistri näitel eelnõus [2] ja tehnilises spetsifikatsioonis [17] esitatud visiooni järgi sisseostetav teenus, mis toimib üldjoontes nii, et keskus saab narkomaani isikukoodi, krüptib selle ära ja saadab TAI andmebaasi. Selleks, et krüptitud isikukoodi teiste andmetega kokku viia, kasutatakse üht arsti infosüsteemi poolt genereeritud juhuslikku arvu (sisuliselt ühekordset sessioonivõtit), mis liigub kaasa nii andmetega kui ka krüptitud isikukoodiga.

Probleemi niisugune lahendus toob aga sisse uue turvariski – nõnda saab kodeerimiskeskuse pidaja ligipääsu andmetele, mida narkoregistri varjestami-

sega just varjata soovitakse, nimelt narkomaanide isikukoodidele. Isegi kui kodeerimiskeskuse pidaja tarkvara on auditeeritud ja kõik logid kustutatakse, jätab isikukoodide kandmine läbi keskuse alles piisavalt ründestsenaariume, mida kõiki ei ole võimalik välistada. Kui pikemas perspektiivis tahta hakata ka teiste delikaatsete andmekogude isikuandmeid samas keskuses kodeerima, omandab selle nõrkuse ründamine järjest suurema väärtuse.

Ääremärkuse korras tuleb mainida, et pakutud kodeerimiskeskuse lahendus [2, 17] on ka eelöeldust sõltumata ebaoptimaalne ning seega muudatusteta realiseerimiseks mittekohane. Andmeandja, kodeerimiskeskuse ja registripidaja vahelist protokolli on võimalik üles ehitada nii, et krüptimata isikukoodid kodeerimiskeskust ei läbi. Üks võimalus sellise protokolli realiseerimiseks on toodud lisas A.

Samas maandame me viidatud protokollitäiendusega ainult ühe ja tegelikult sugugi mitte kõige tõsisema riski. Teiste tekkivate riskide kirjeldamiseks peame kodeerimiskeskus(t)e infrastruktuuri ja töökorda veidi lähemalt vaatlema. Üldiselt on kodeerimiseks kaks põhimõtteliselt erinevat võimalust, mida mõlemat on Sotsiaalministeeriumis kaalutud, kuid lõplikku otsust ühe või teise kasuks pole käesoleva kirjutamise hetkeks veel langetatud.

Esimene võimalus on võtta kõigis delikaatsetes registrites kasutusele üks universaalne asenduskood. See lahendus oleks lihtne realiseerida – tuleks luua ainult üks kodeerimiskeskus ja erinevate registrite linkimiseks poleks vaja ühegi välise osapoole abi. Samas mida rohkemates registrites selle koodiga infot seotakse, seda kergemaks muutub ründajale inimese identifitseerimine ja lõpuks muutub asenduskood sisuliselt isikukoodi aliaseks, kaotades kogu kodeerimise mõtte. Samuti tähendaks olukord, kus kõigi registrite koodide tagasiteisenduseks vajalik info koondatakse ühte keskusesse, suuremat turvariski, sest ründajal piisab kõigi delikaatsete registrite avamiseks pelgalt selle keskuse kompromiteerimisest.

Teine võimalus on seada ühele isikukoodile vastavusse palju erinevaid asenduskoode, piirjuhul lausa iga registri jaoks eraldi. Niisugune lahendus raskendab ründajal inimese profiili koostamist, kui ta erinevate registrite asenduskoodide vahelist vastavust ei tea. Samuti on sel juhul võimalik riski erinevate kodeerimiskeskuste vahel hajutada. Teisest küljest peab erinevate registrite linkimiseks mingi osapool kusagil ikkagi suutma erinevaid asenduskoode kokku viia. Kuna kõiki potentsiaalseid linkimisstsenaariume ei ole võimalik ette näha, siis tuleb luua kas üks suur keskus, kes valdab kõiki vastavusi (ja millega oleme sisuliselt tagasi esimese võimaluse juures) või näha ette protseduurid vastavalt linkimisvajadusele asenduskoodide jagamiseks. Viimane variant tähendab paljude, piirjuhul lausa ühekordsete kodeerimiskeskuste loomist. Samas läheb ka ühekordsete asenduskoodivastavuste määramiseks ikkagi vaja teavet selle kohta, millised koodid vastavusse seada saab. Kok-

kuvõttes ei anna vaadeldav võimalus universaalse asenduskoodi lahenduse ees mingit võitu, kui me vaatleme ohuna kodeerimiskeskuse kompromiteerimise rünnet.

Võimalikke ründeid on aga veel. Nimelt võib ründaja ära kasutada peatükis 4 tehtud tähelepanekut ja üritada isikuid teiseste tunnuste järgi (kasvõi mingi tõenäosusega) tuvastada.[7] Võime endale ette kujutada kuritegelikku "tagasiteisendusteenust", mille pakkuja hangib endale (võibolla mitte päris legaalse) ligipääsu kodeeritud registrile ning koostab selle põhjal tagasiteisendustabeli, kus on iga asenduskoodi kohta kirjas rida lahtisi isikukoode koos tõenäosustega, kui tõenäoliselt mingi neist isikukoodidest sellele asenduskoodile vastab.

Kui kõigis registrites kasutatakse sama universaalset asenduskoodi, saab teenusepakkuja oma andmeid aja jooksul teistele registritele ligi pääsedes järjest täpsustada, kuni lõpuks on tema käsutuses kogu kodeerimiskeskuse saladus. Kui eri registrites on erinevad koodid, siis peab teenusepakkuja muidugi rohkem tööd tegema, tema töö tulemuseks on iga registri kohta üks tabel ja see pole nii täpne. Samas avaneb teenusepakkujal linkimise korral võimalus tuvastada mingi asenduskoodide paari vastavus ning kasutada mõlema registri teiseseid andmeid sellele paarile vastava füüsilise isiku väljaselgitamiseks. Kui mingit liiki registreid teistega ei lingitagi, jääb tagasiteisendusteenuse pakkuja töö tulemus halvemaks, kuid siiski piisavalt väärtuslikuks, et erinevad osapooled (vt jaotis 6.2) selle vastu huvi võiksid tunda. Kuna kogutud informatsiooni võib müüa mitu korda, on kirjeldatud rünne sellevõrra majanduslikult kasulikum ja järelikult ka ohtlikum.

Intervjuu käigus Sotsiaalministeeriumis selgus, et paljude erinevate asenduskoodide puhul pole linkimise kontseptsioon veel väga selgelt läbi mõeldud. Ministeeriumi spetsialistid pakkusid välja, et kõiki neid asenduskoode võiks ikkagi hallata üks kodeerimiskeskus. Kuna asenduskoodide vaheliste vastavuste liigutamine kodeerimiskeskuse turvatud keskkonnast välja oleks turvarisk (kergendades näiteks tagasiteisendusteenuse rünnet), siis pakkusid spetsialistid välja idee teha ka linkimine kodeerimiskeskuse sees. Näiteks kui on vaja linkida kodeeritud registrid A ja B, siis võiks seda teha nii:

- registrist A saadetakse kodeerimiskeskusesse registrisse kuulujate asenduskoodid,
- kodeerimiskeskus asendab need koodid niisugustega, mis on arusaadavad registrile B ning edastab need B-le,

[7] Kodeerimiskeskuse algne motivatsioon on läbi isikukoodi kodeerimise kahandada registri turvanõudeid. Seega võime eeldada, et ründajal on kodeeritud registrile kergem ligi pääseda kui selle kodeerimata kujule. Järelikult võime kodeerimiskeskuse turvaanalüüsis eeldada, et ründaja näeb andmekogusse kantud inimeste teiseseid andmeid.

- register B tagastab kodeerimiskeskusele nende isikute andmed, kes tema registrisse kuuluvad,
- kodeerimiskeskus teisendab nendes andmetes asenduskoodid tagasi A-le mõistetavale kujule ning tagastab andmed registrile A.

Niisuguse lahenduse korral omandab kodeerimiskeskus aga pikas perspektiivis lõpuks kõigi lingitud delikaatsete registrite sisu. Pannes juurde asjaolu, et kodeerimiskeskus teab ka kõigi asenduskoodide vastavusi (seega sisuliselt isikukoodi aliast), tekib kodeerimiskeskusesse see superandmebaas, mille moodustamist me kogu kõnealuse infrastruktuuri loomisega vältida üritasime. Ülaltoodud lahendus ei ole kuidagi turvalisem sellest juhust, kui kodeerimiskeskuses hoitaks lihtsalt kõiki registreid isikustatud kujul.

Käesoleva peatüki kokkuvõtteks võib öelda, et kodeerimiskeskuse kui tehnilise lahenduse väljatöötamisel on jäetud tähelepanuta registrite põhiline kasutusjuhtum – linkimine. Kuna kõiki tulevasi vajalikke linkimisi pole praegu võimalik ennustada (ja väga tõenäoliselt pole hetkel isegi olemas kõiki neid registreid, millega kunagi praeguseid linkida tuleb), siis tuleb linkimiseks sisuliselt isikud ikkagi tuvastada. Seega on kõik seadusesätted, mis nõuavad registrite absoluutset isikustamatust, mõttetud, sest välistavad registrite sihipärase kasutamise. Kodeerimiskeskus omab otsese tuvastamise riski vastu küll teatavat efekti, kuid ei likvideeri kuidagi kaudset tuvastamist, samuti hääbub tema efekt aja vältel. Üritades likvideerida volitatud töötleja poolset rünnet võime me saavutada hoopis olukorra, kus kodeerimiskeskusesse tekib superandmebaas kõigist riigi delikaatsetest registritest.

6 Narkoregister

6.1 Rollid

Narkoregistri ülesandepüstituses [3, 2] on kirjeldatud järgmised rollid.

- Arst – psühhiaatri litsentsi omav meditsiinitöötaja või tema poolt volitatud isik (õde), kes tegeleb narkomaani ravi ja andmete töötlemise ning edastamisega narkoregistrisse.
- Tervise Arengu Instituut (TAI) – narkoregistri pidaja ja töötleja. TAI vahendab päringuid narkoregistrisse ja omab võimalust neid blokeerida, kui päringud võiksid kahjustada registrisse kuuluvate isikute privaatsust. Praktiliselt pääseb registrile ligi vaid 1-2 TAI usaldatud töötajat, kellel lasub kogu vastutus andmelekete korral.

- Avalikkus – kõik ülejäänud isikud, kellel on ligipääs vaid narkoregistrist tehtud üldistele statistilistele kokkuvõtetele. Siia rolli kuuluvad teadlased, kes uurivad narkomaania ja selle raviga seotud trende, aga ka kõik teised kodanikud, kellel on õigus pärida registrist infot vastavalt avaliku teabe seadusele [8]. Avalikkus ei tohi saada teada ühegi konkreetse inimese kuulumisest narkoregistrisse ei otseste ega kaudsete päringute abil.

6.2 Ohustsenaariumid ja nende analüüs

Järgnevalt vaatleme ohustsenaariume narkoregistri näitel; analoogilised riskid kerkivad esile ka teiste kõrge delikaatsusastmega registrite (nt HIV register) korral. Eeskätt huvitab meid, mil määral maandab neid riske kodeerimiskeskuse rakendamine. Nagu nägime peatükis 4, ei pruugi kustutatud nimede ning kodeeritud isikukoodidega registrist olla võimalik inimesi üheselt identifitseerida, kuid teatava lisatööga saab väita, et inimene kuulub registrisse mingi tõenäosusega $\frac{k}{n}$. Vaatleme, millised ründed on sellises vormis väite korral veel aktuaalsed.

Narkoregistri turvameetmete väljatöötamisel tuleb Sotsiaalministeeriumi spetsialisti Kaja Kuivjõgi andmetel lähtuda järgmistest ohustsenaariumitest.

1. Ründaja tahab kompromiteerivaid andmeid kätte saada selleks, et nende abil inimest või tema lähedasi santažeerida või manipuleerida.

2. Ründajaks on pank, kindlustusselts või mõni muu finantsasutus, kes otsustab andmete põhjal inimese kuulumise mõnesse riskirühma, et siis tema makseid suurendada.

3. Ründajaks on uuriv ajakirjanik, kes on saanud mõne tuntud isiku kohta vihje ning tahab kontrollida, kas see isik tõepoolest kuulub sellesse registrisse.

Vaatleme neid stsenaariume eraldi, alustades santaažist. On selge, et paraku pole enamik narkomaane rahaliselt kuigi heal järjel, mistõttu nende santažeerimine ei anna eriti palju tulu. Samas võib narkoregistris olla ka tuntud isikuid (poliitikuid, ärimehi jt) ning väga tõenäoliselt ongi. Nende santažeerimine annab kindlasti kasu, aga potentsiaalse tulu suurust ei osanud Sotsiaalministeeriumi spetsialistid hinnata. Ekstreemse näitena tõid nad Konstantin Pätsi, keda Nõukogude Liit suutis omal ajal poliitiliselt mõjutada tänu sellele, et riigivanem oli kunagi Venemaal hämaravõitu kullaäri ajanud. Sel juhtumil oli santaaži hinnaks Eesti Vabariigi iseseisvus. Samas on selge, et toodud

näide on selgelt äärmuslik ning reaalsem on näiteks ministri mõjutamise hinnaks pidada tema ametiajal saadavate hüvede (palk, auto, preemiad jne) kogusummat.

Kas santažeerimine on võimalik, kui me teame poliitiku kuulumist narkoregistrisse mitte 100%, vaid mingi tõenäosusega $\frac{k}{n}$? Jah, loomulikult, sest poliitik ju ei tea, mida ründaja täpselt teab. Seega võib ründaja mingi piisavalt suure tõenäosuse korral poliitikut ikkagi ähvardada. Kui 100%-lise kindluse korral oleks tema ootetulu x, siis nüüd on see $\frac{k}{n} \cdot x$. Seega ei kaota kodeerimiskeskuse sisseviimine santaažirünnet, vaid kahandab ainult tema tulukust. Eeldusel, et poliitikute käest saadavad summad võivad olla väga suured, jäävad need ründed endiselt alles.

Teine, pankade või klindlustusseltside oht on aga näide sellisest ründest, kus ründajal polegi põhimõtteliselt oluline registris inimesi sajaprotsendiliselt identifitseerida. Teadmist, et konkreetne finantsteenust sooviv isik on tõenäosusega $\frac{k}{n}$ narkomaan, saavad nii pangad kui kindlustusseltsid oma riskiarvutustes täiesti kasutada. See omakorda tähendab, et kodeerimiskeskuse mõte kahandada ründajate motivatsiooni isikute peitmise läbi ei ole selle stsenaariumi puhul kuigi tõhus.

Kolmandat liiki ründaja, uuriva ajakirjaniku jaoks on tõepoolest oluline teada, et poliitik kuulub narkomaanide hulka; tõendid kindlusega alla 100% ei ole piisavad mõjusa artikli kirjutamiseks ning probleemide korral kohtus võitmiseks. Samas kui registri andmekoosseis on selline, et $n = k$ (vt peatükk 4), siis on ajakirjanikul võimalik ka kodeeritud andmekogu pealt 100%-lisi väiteid esitada.

Kokkuvõttes võime öelda, et kolmest Sotsiaalministeeriumi spetsialistide poolt välja toodud ohustsenaariumist kahe puhul aitab kodeerimiskeskus vähendada ründe tulusust, kuid mitte piisavalt selleks, et ründeid kaotada. Kolmanda stsenaariumi puhul on rünne endiselt võimalik, kuid vajab spetsiifilist andmekooseisu ja niisuguse ründe saab välistada mõnesid andmevälju kustutades. Kuna registrite kooseis aga pidevalt muutub, tähendab see vajadust registrit pidevalt monitoorida.

6.3 Narkoregistri täiendavaid kasutusvõimalusi

Narkoregistri kodeerimiskeskuses on ette nähtud teatud piiratud tagasiteisenduse võimalus, aga seda ainult andmete korrigeerimiseks. Seda, et üks arst pääseks ligi teise arsti juures käinud narkomaani kirjetele, ei ole planeeritud. Samas on narkomaania ravi võrdlemisi spetsiifiline, haigetele manustatakse järjest vähenevaid annuseid metadooni, mis on ka sisuliselt narkootikum. Narkomaanid kasutavad hetkel ära seda, et arstide vahel info ei liigu ja käivad ühe arsti juurest teise juurde oma ravi "alustamas", et saada kätte uus suur

annus.

Üks üleriigilise registri täiendav kasutusstsenaarium võiks olla, et arstid saavad näha oma uute patsientide ajalugu – see aitaks senisest palju tõhusamalt narkomaaniaga võidelda. Tõsi, niisugune lahendus eeldaks narkoregistri muutmist sisuliselt operatiivandmebaasiks ja tõstaks turvariske. Samas on psühhiaatritel nagunii narkomaanide ravimise õigus ja ametist tulenev delikaatsuskohustus, mistõttu võib osutuda, et registri niisugusest rakendusest saadav tulu kaalub võetava riski üles. Seda peab näitama täpsem analüüs.

7 Kokkuvõtted ja soovitused

Infotehnoloogia areng pakub meile palju uusi mugavaid võimalusi, kuid põhjustab kahjuks ka mõndagi ebameeldivat. Üheks selliseks kahetahuliseks näiteks on elektrooniliste registrite pidamine, mis ühest küljest võimaldab teha laiahaardelisi uuringuid kiiremini kui ei kunagi varem, kuid teisest küljest lubab registrites sisalduvat delikaatset infot inimeste vastu ka märksa tõhusamalt ära kasutada.

Paraku tuleb registris sisalduvaid andmeid registri ülesannete täitmiseks väga sageli konkreetsete isikutega siduda. Isegi juhul, kui andmekogu pidamise ainus eesmärk on tugevasti umbisikustatud statistika, läheb viiteid andmesubjektidele vaja vähemalt andmete uuendamisel ja teiste registritega linkimisel. Sellest johtuvalt peab alati kusagil eksisteerima (kasvõi hajutatud) osapool, kes suudab registrisse kuulujaid kindlaks teha. Küsimus seisneb ainult selles, kui kergesti ründaja andmeid sellelt osapoolelt kätte saab ning kui palju kahju ta seejärel teha võib.

Kui delikaatseid isikuandmeid sisaldavat registrit pidada isikustatult, on esimeseks kahtlusaluseks volitatud töötleja, kes suudab andmetele ligi pääseda jälgi jätmata. Selle riski maandamiseks, võimaldades samas registritel siiski oma ülesandeid täita, on Sotsiaalministeeriumi eestvedamisel töötatud välja ühe infotehnoloogilise lahenduse, nn kodeerimiskeskuse kontseptsioon. Käesolev analüüs käsitleski seda kontseptsiooni laiemas kontekstis ning püüdis välja selgitada, kui palju kodeerimiskeskuse rakendamine süsteemi reaalset turvalisust tõstab.

Narkoregistri juhtumi näitel võisime kõigepealt veenduda, et kuigi isikukoodi kodeerimine vähendab isiku otsese tuvastamise ohtu, pole suurema osa rünnete jaoks isiku otsest tuvastamist tegelikult vajagi ja sellest tuvastamistäpsusest, mida registri valdaja lahtiseks jäetud andmete abil saavutada suudab, piisab ründamiseks täiesti.

Teiseks nägime, et kuna kodeerimiskeskus üritab lahti siduda andmeid,

millede vahelised seosed füüsilises maailmas alles jäävad (st inimese isikut ja tema erinevaid tunnuseid), siis on kodeerimiskeskuse efekt ajaliselt piiratud – mida enam registrit kasutatakse ja teiste registritega seotakse, seda avalikumaks ja kindlamaks muutuvad asenduskoodide ja isikute vahelised vastavused. Pikas perspektiivis ei varja kodeerimine ründaja eest enam midagi.

Kolmandaks viivad kõik seni välja pakutud kodeerimiskeskuse lahendused sisuliselt superandmebaasi tekkimisele. Kui koondada koodide ja isikute vahelised vastavused ühte keskusesse (mida kõigi potentsiaalsete linkimiste võimaldamiseks teha tuleks), siis peaks selle keskuse turvamiseks kasutama nii tugevaid füüsilisi ja organisatsioonilisi meetmeid, et sama hästi võiks seal hoida isikustatud registreid.

Kokkuvõtlikult võib öelda, et volitatud töötleja poolsete rünnete ohtu polegi praktiliselt võimalik maandada, selle asemel toob kodeerimiskeskuse lahendus aga sisse hoopis täiendavaid turvariske.

Tegelikult algab probleem volitatud töötleja kui riskiallika hindamisest. Kogu analüüsi läbinud narkoregistri näite puhul on selleks töötlejaks TAI, kus uuringuid teevad arstiteadlased. Need inimesed on ühest küljest oma erialale pühendunud arstid, keda seovad nii eetikakoodeks kui ka vastavad vanded. Teisest küljest on tegu teadlastega, kes tegutsevad antud valdkonnas juba aastakümneid, neid tuntakse väga hästi rahvusvahelistes teadusringkondades ja kohalikes riigiasutustes. Ka nemad ei pääse delikaatsetele isikuandmetele ligi sugugi niisama, vaid range korra alusel, kirjutades kõigepealt üksikasjaliku ja põhjendatud uurimisplaani, mis peab saama eetikakomisjoni heakskiidu. Alles seejärel võimaldatakse neil oma uuringute läbiviimiseks andmetele ligi pääseda ja seda täpselt sel määral, mis antud uuringu jaoks vaja on. Niisugune kord kehtib Eestis juba aastaid ja siiamaani pole esinenud ühtki juhtumit, kus delikaatsed tervisealased isikuandmed oleksid lekkinud teadurite kaudu.[8] Seega võib senist turvapraktikat TAI (ning teiste volitatud terviseandmete töötlejate) andmekäitluses pidada adekvaatseks ja isikute varjestamiseks kodeerimise läbi puudub praktikas põhjendatud volitatud töötleja riskist lähtuv vajadus.

Kokkuvõttes näeme, et registriandmete legaalsed kasutajad on ründajatega võrreldes oluliselt halvemas seisus. Selleks, et registritele toetudes teha head ning viia läbi uuringuid, millest sisuliselt tulu tõuseks, peavad kasutatavad andmed olema täpsed ja isikustatavad. Samas selleks, et halba teha ning registritesse kantud isikuid delikaatsete info abil rünnata, piisab ka ebatäpsetest andmetest. Niisiis põhjustab kodeerimiskeskuse laadne lahendus kahju

[8] Küll on kaotatud paberkandjal dokumente madalama taseme meditsiinitöötajate hooletuse tõttu.

eeskätt legaalsetele kasutajatele, likvideerimata samas ühtki tõsisemat rünnet.

Ülalöeldu põhjal anname delikaatseid isikuandmeid sisaldavate registrite käitlemiseks järgmised soovitused.

- Isikuandmete kaitse seadust [6], riikliku statistika seadust [7] (ja vajadusel teisi õigusakte) tuleb täiendada nii, et teadusuuringuteks (sh epidemioloogilisteks) vajalikke registreid võiks isikustatult pidada ning uuringulistel eesmärkidel linkida ilma andmesubjektide nõusolekuta. Eraldi tuleb sätestada nende registrite turvamiseks vajalikud füüsilised ja organisatsioonilised meetmed.

- Registrite linkimiseks tuleb luua nii seadusandlik kui tehnoloogiline raamistik.

- Iga delikaatseid isikuandmeid sisaldava registri kohta tuleb koostada loetelu registriga töötavate isikute rollidest, määrata rollipõhised pääsuõigused ning täiendada infosüsteeme nende õiguste alusel.

Kui mingitel (poliitilistel, ärilistel vms) põhjustel otsustatakse kodeerimiskeskus(ed) siiski rajada, siis tuleks arvestada järgmiste soovitustega.

- Kodeerimiskeskuse ülesehitus peab vältima superandmebaasi tekkimist ja üldse andmebaaside delikaatsete osade lahtist edastamist läbi keskuse. Muuhulgas käib see isikukoodide kohta, mille kodeerimiskeskuse eest peitmise üks võimalus on toodud lisas A.

- Kui tahetakse kasutada üht universaalset isikukoodi asenduskoodi, siis tuleb hinnata aega, mille jooksul koodide vastavused piisavalt suure tõenäosusega avalikustuvad ning analüüsida, kas kodeerimiskeskuse investeering ennast selle ajaga ära tasub.

- Kui tahetakse kasutada paljusid erinevaid asenduskoode, tuleb põhjalikult läbi töötada registrite linkimise metoodika, sest praeguseks välja pakutud lahendus viib superandmebaasi tekkele. Võibolla on võimalik kasutada protokolli, mis sarnaneb spetsifikatsioonis [17] kirjeldatud süsteemile lisa A täiendusega, aga seda peab näitama edasine analüüs.

- Selleks, et varjestatud isikukoodidega registri korral vähendada isikute teiseste andmete põhjal identifitseerimise ohtu, tuleb hoolikalt analüüsida registri andmekooseisu ja leida, kui suure täpsusega isikud selles andmekogus kindlakstehtavad on. Iga niisuguse registri korral tuleb määrata lävi, millest suuremat täpsust peetakse ohtlikuks. Seejuures

võib osade registrite puhul oluliseks osutuda *keskmine*, osade puhul aga *minimaalne* täpsus. Kuna registrite andmekooseis võib ajas muutuda, tuleb määratud läve mitteületamist pidevalt jälgida ning ületamisel vastavad meetmed tarvitusele võtta (nt mõnesid välju kustutada).

- Selleks, et kodeerimiskeskus täidaks oma ülesannet vähendada registri töötleja poolsete rünnete riski, pole teda otstarbekas siduda ühegi toimiva registriga (sh rahvastiku- või pereregistriga).

Viited

[1] Kaja Kuivjõgi, Jaak Parre, Tarmo Tammaru, Tiina Vares, *Tervise infosüsteemi arengukava 2005–2008*, Eesti Vabariigi Sotsiaalministeerium

[2] *Narkomaaniaravi andmekogu asutamise ja andmekogu pidamise põhimäärus*, eelnõu

[3] *Narkootiliste ja psühhotroopsete ainete ning nende lähteainete seadus,* vastu võetud 11.06.1997. a seadusega (RT I 1997, 52, 834), jõustunud 1.11.1997

[4] *Riikliku kinnipeetavate, arestialuste ja vahistatute registri asutamine ja registri pidamise põhimäärus.* Vastu võetud 11.07.2004 (RT I 2004, 57, 409), jõustunud 19.07.2004

[5] *Andmekogude seadus.* Vastu võetud 12.03.1997. (RT I 1997, 28, 423), jõustunud 19.04.1997

[6] *Isikuandmete kaitse seadus.* Vastu võetud 12.02.2003 (RT 2003, 26, 158), praegune redaktsioon kehtib alates 1.05.2004

[7] *Riikliku statistika seadus.* Vastu võetud 11.06.1997 (RT I 1997, 51, 822), jõustunud 17. 07. 1997

[8] *Avaliku teabe seadus.* Vastu võetud 15.11.2000 (RT I 2000, 92, 597), jõustunud 1. 01. 2001

[9] *Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.* Official Journal of the European Communities of 23 November 1995 No L. 281 p. 31

[10] *Lov om behandling af personoplysninger (Act on Processing of Personal Data*). Taani, Act No. 429, vastu võetud 31.05.2000

[11] *Personuppgiftslag (Personal Data Act)* (1998:204). Rootsi, jõustunud 24.10.1998

[12] Mati Rahu, *Epidemioloogilised uuringud ja indiviidi privaatsus*. HIPPOKRATES, märts 2003 (42), lk 188–196

[13] Mati Rahu, *Põhjamaad ja Eesti – isikustatud andmete töötlemise kaks poolust.* Sotsiaalministeeriumi rahvusvaheline konverents. Eesti Arst 2005, 84 (1), lk 50–52

[14] Mati Rahu, Hans Storm, Teadustöö, registrid, inimesed ja andmekaitse. HIPPOKRATES, märts 2004 (52), lk 188–191

[15] Mika Gissler, Jari Haukka, *Soome tervise- ja sotsiaalregistrid ning epidemioloogilised uuringud*. Eesti Arst 2004, 83 (8), lk 543–552

[16] *A Finger on the Pulse. Monitoring public health and social conditions in Sweden 1992–2002*. Centre for Epidemiology, National Board of Health and Welfare, Stockholm, Rootsi, november 2003

[17] *Isikukoodi kodeerimise süsteem*, tehniline spetsifikatsioon, Sotsiaalministeerium 2005

A Kodeerimiskeskuse protokolli täiendus

Nagu mainitud peatükis 5, puudub otsene vajadus kodeerimiskeskusele lahtisi isikukoode saata. On võimalik luua protokoll, kus arst saadab isikukoodi kodeerimiskeskusele krüptitult ja mille siis kodeerimiskeskus omakorda üle krüptib. Selleks, et TAI suudaks eristada korduvaid isikukoode, peaks arsti krüptogramm olema niisugune, mida TAI suudab eemaldada nõnda, et kodeerimiskeskuse krüptogramm alles jääb. Seda on võimalik saavutada näiteks järgmise protokolliga.

- Süsteemis lepitakse kokku suur algarv p, näitlikkuse huvides olgu meil praegu $p = 19$.
- Arst ja TAI lepivad kokku sellised a ja b, et $a \cdot b \equiv 1 \bmod p-1$, näiteks $a = 5$ ja $b = 11$; siis tõepoolest $5 \cdot 11 = 55 = 3 \cdot 18 + 1$.
- Kodeerimiskeskus valib samuti omale arvu k, mis on ühistegurita arvuga $p - 1$; näiteks $k = 7$.
- Isikukoodi $x = 10$ saatmiseks arvutab arst

  $$x^a \bmod p = 10^5 \bmod 19 = 3$$

  ja saadab selle kodeerimiskeskusele.
- Kodeerimiskeskus arvutab $(x^a)^k = x^{ak} \bmod p$

  $$(x^a)^k = x^{ak} \bmod p = 3^7 \bmod 19 = 2187 \bmod 19 = 2$$

  ning edastab TAIle.
- TAI arvutab $(x^{ak})^b = x^k \bmod p$.

  $$(x^{ak})^b = x^k \bmod p = 2^{11} \bmod 19 = 2048 \bmod 19 = 15.$$

  See väärtus jääb siis TAI jaoks isiku kodeeritud tunnuseks.

Nii saab TAI leppida iga arstiga kokku erineva eksponentide paari ja samas jäävad isikukoodid tema eest varjatuks (sisuliselt on isikukoodid kodeeritud kodeerimiskeskuse võtmega). Tuleb aga meeles pidada, et toodud on väga lihtne krüptosüsteem, millel võib olla ootamatuid nõrkusi mõnes kasutusmallis. Sellised nõrkused toob esile põhjalikum analüüs.