JAANIKA PUUSALU

SUURANDMED: OLEMUS JA KASUTAMISE KITSASKOHAD

SUURANDMED: OLEMUS JA KASUTAMISE KITSASKOHAD

JAANIKA PUUSALU

SISEKAITSEAKADEEMIA
ESTONIAN ACADEMY OF SECURITY SCIENCES

Esikaane foto: data.path Ryoji.Ikeda - 3
Makett: Jan Garshnek
Keeletoimetaja: Victoria Parmas

ISBN 978-9985-67-323-2 (pdf)

www.sisekaitse.ee/kirjastus

SISUKORD

SISSEJUHATUS 4

1. LÄHTEPUNKT: INFOTEHNOLOOGIA LEVIK JA UUED ANDMED 6
    1.1. Maailmatrendid infotehnoloogias: andmed kui uus nafta 6
    1.2. Maailmatrendid kohalikus olustikus: Eesti tehnoloogiline tulevikuvisioon 9
2. SUURANDMED: TEOREETILINE KÄSITLUS 12
    2.1. Suurandmete mõiste 12
    2.2. Suurandmete olemus 13
3. SUURANDMED EESTI KONTEKSTIS 20
    3.1. Digitaliseeritud andmebaasid ja andmekogud 20
    3.2. Interneti (kasutatavate rakenduste) kasutamisel tekkivad andmed 22
    3.3. Seadmed, mis automaatselt andmeid toodavad ja nutistu seadmed 24
    3.4. Digitaalsed suurandmed Eestis 25
4. ANDMETE KASUTAMISE TULEVIK EESTIS 27
    4.1. Kratistrateegia 27
    4.2. Uus tehnoloogia ja tehisintellekti süsteemid korrakaitses 28
    4.3. Suurandmete kasutuselevõtu kitsaskohad 30

KOKKUVÕTE 38
KASUTATUD KIRJANDUS 42

LISA 46

SISSEJUHATUS

Eestil on tehnoloogiliselt arenenud riigi maine, kus digitaalsed avaliku sektori lahendused on laialdaselt töösse võetud ning koostoimivad. Kohalikke lahendusi tutvustatakse ja turustatakse üle maailma[1] ning ka uute IT-rakenduste ja lahenduste leidmisel on eestlased oma initsiatiivi tõestanud.[2] Kuigi Eesti IT-lahendused on tihti omanäolised, on edasine suund uute lahenduste, ka avaliku sektori teenuste laheduste puhul maailmatrendidega sarnane. Üha enam keskendutakse suurandmete (ingl *big data*) töötlemisele ning tehisintellekti (TI) või tehisintellekti süsteemide[3] (ingl *artificial intelligence* ehk *AI*) arendamisega kaasnevatele võimalustele.

Digitaliseerimisega kaasnenud hüppeline andmemaht on katsumus, aga samas pakub uusi võimalusi kogu maailmale – esimest korda on pea kõike toimuvat võimalik kogutud andmete abil analüüsida. Andmeanalüüsist lähtuvalt on võimalus masinaid toimima õpetada, teenuseid täiustada ja tegevust ning kasutust optimeerida. Kuigi andmeanalüüsil on suur potentsiaal nii toimivate kui ka uute tehnoloogiliste lahenduste väljatöötamisel, on andmete kogumine ja koostoime, ligipääs ja kasutamine seotud hulga ülesannete ja probleemidega. Digitaalsed andmed tekivad samal ajal nii avaliku- kui ka erasektori pakutavate rahvusvaheliste ja kohalike teenuste kasutamisest; need on nii sõnalised ja numbrilised kui ka pildis ja helis; neid on nii avalikult ligipääsetavaid kui ka ainult kindla teenusega seotuid ning turvatuid ja privaatseid; need kustuvad kohe või talletatakse pikaks ajaks.

Et suurandmete potentsiaali rakendada ja nende abil teenuseid ning tooteid arendada, on vaja täpsemalt mõista, mida mõiste suurandmed tähendab ehk missuguseid andmeid loetakse suurandmeteks ning millised on nende kasutamisega seotud katsumused.

Selle uurimuse eesmärk on selgitada lähemalt suurandmete olemust ning kaardistada, missugused on võimalikud kitsaskohad suurandmete analüüsil tuginevate rakenduste arendamiseks ja kasutuselevõtuks digitaliseeritud ning juba koostoimivaid avaliku- ja erasektori teenuseid omavas Eestis. Ülesannete kaardistamisel keskendub raport avaliku sektori teenuste pakkumiseks vajalike suurandmete rakendamisele.

Mitu suurandmete kogumise ja kasutamisega kaasnevat probleemi, millele raportis tähelepanu juhitakse, sh andmete kasutuse õiguslikud probleemid, on muutunud eriti aktuaalseks ja ajakriitiliseks just COVID-19 pandeemia valguses, mil arendatakse rahvatervise huvides ülemaailmselt inimeste jälgimist ning terviseandmete analüüsimist või-

[1] Näiteks X-tee lahenduse kasutuselevõtt WHO-s (Marand ja Seppel, 2020).

[2] Näiteks Eestist alguse saanud *Hack the Crisis* häkaton (ERR uudised, 2020).

[3] Üha enam käsitletakse tehisintellekti kui tehnoloogilist lahendust, mis on süsteemi osa – see võimaldab tööprotsessi mõne etapi automatiseerimist tehisintellekti abil, nt suurandmete analüüsiks kasutatav iseõppiv algoritm, mis toetab teenust x. Tehisintellekti kui täisautonoomset masinat käsitletakse pigem tulevikuvõimalusena. Nii räägitakse rohkem tehisintellekti süsteemist kui tehisintellektist.

maldavaid rakendusi.[4] Tekkinud probleemid toovad välja kitsaskohad, nagu reeglite puudumise ja toimivate praktikate ähmasuse ning era- ja riigi teenuste sõltumise üksteisest. Allolev selgitab ilmnevaid kitsaskohti suurandmete olemust avades.

Esmalt selgitab uurimus suurandmete potentsiaali ilmnemist maailma ja Eesti tehnoloogia arengu valguses. Teiseks selgitab uurimus suurandmete mitmekülgset olemust ning sellest tulenevat kasutamise ja koostoime keerukust. Kolmandaks avatakse raportis, milliseid andmeid saab Eesti kontekstis lugeda suurandmeteks ning miks on need teiste riikide suurandmetega võrreldes mõneti erinevad. Viimases alapeatükis avatakse kitsaskohti, mis võivad suurandmete rakendamisel ja teenuste arendamisel Eesti jaoks murekohaks saada.

Raport toob välja, et vaatamata potentsiaalile on suurandmed väikesel riigil, millel on limiteeritud ressursid ise teenuseid ja lahendusi arendada, suur katsumus: samal ajal tuleb hakkama saada nii juba arendatud teenuste heaperemeheliku haldamise kui ka uute teenuste ning tehnoloogiliste potentsiaalide arendamise ja rakendamisega. Need kaks ülesannet ei pruugi aga ühtida ega osutuda samal ajal teostatavaks.

[4] Vaata näiteks O'Neill, 2020; O'Neill, Ryan-Moseley, & Johnson, 2020.

1. LÄHTEPUNKT: INFOTEHNOLOOGIA LEVIK JA UUED ANDMED

1.1. MAAILMATRENDID INFOTEHNOLOOGIAS: ANDMED KUI UUS NAFTA

Uue infotehnoloogia laialdane levik[5] ja üha kättesaadavam internet[6], mis info- ja kommunikatsioonitehnoloogia (edaspidi IKT) rakendusi toetab ja IKT seadmeid sidestab, on olulised arengud kogu inimkonnale, sest on muutnud nii suhtlus- kui ka organisatsioonipraktikaid. Uus tehnoloogia lubab näiteks pakutavaid teenuseid kiiremini ning soodsalt vahendada ja tarbida, sest võimaldab tihti saada teenust füüsiliselt kohale minemata ja ilma vahendava inimeseta: muuta teenindus ja suhtlus virtuaalseks; asendada otsene käsk ja kontroll kaugjuhtimisega; usaldada bürokraatia arvutiprogrammile. Just pandeemiaga kaasnevas üleilmses vajaduses osaleda ühiskondlikes protsessides distantsilt on senine tehnoloogiline areng proovile pandud ning see on end õigustada suutnud. Nimelt on paljud vajalikud toimingud, k.a õppetegevus, uute tehnoloogiliste lahenduste abil tõesti (jätkusuutlikult) läbi viidavad.

Uue tehnoloogia kasutuselevõtul rõhutatakse aga ka selle positiivset ühiskondlikku mõju. Digitaalsete lahenduste puhul eeldatakse näiteks paremat ligipääsu (riigi)teenustele; õiglasemat kohtlemist ja sotisaalsete erisuste kadumist virtuaalseid teenuseid kasutades ning internetis suheldes.[7] Lisaks viidatakse ka riikide tasavägisemale arengule, sest digitaalsete teenuste pakkumiseks ja väljatöötamiseks on üldjuhul vaja vähem ressurssi kui füüsiliste taristute arendamiseks; ausamale konkurentsile, sest firmadel on võimalus olla võrdväärselt nähtav; suuremale sõna- ja informatsioonivabadusele, sest inimestel on vabadus valida infokanalite vahel.[8]

Digitaalsete protsesside laiaulatusliku kasutuselevõtuga on vajalikuks osutunud ja suurenenud ka juba olemasolevate andmete digitaliseerimine. Samuti kaasneb digitaalsete teenuste kasutamisega nüüd suur hulk uusi andmeid. Ainuüksi aastal 2018 tootis inimkond

5 Tänu nutitelefonide soodsale hinnale ning pakutavatele võimalustele võib mõnes arenevas riigis esimeseks ja peamiseks interneti kasutamise vahendiks olla nutitelefon, mitte arvuti (Blank & Dutton, 2014).

6 31. detsembri 2010 seisuga on maailmas 4 574 150 134 interneti kasutajat (Internet World Statistics, 2020).

7 *Eesti Tehisintellekti kasutuselevõtu ekspertrühma aruanne* viitab, et õiglane kratt võimaldab rohkem vaba aega, meelerahu ja tagab õiglasema ühiskonna (Majandus- ja Kommunikatsiooniministeerium ja Riigikantselei, 2019).

8 1990ndaid ja 2000ndate algust võib lugeda Interneti-entusiasmi hiilgeaegadeks, mis on tänu uue tehnoloogia keerukale toimimisele, k.a ebademokraatliku ärimudeli kasutamisele, tänaseks veidi taandunud. Kuid üldine arusaam tehnoloogia arengu vajalikkusest ja sellega kaasnevast ühiskondlikust kasust on jätkuv. Nt Euroopa Liit plaanib teha jõupingutusi tehisintellekti süsteemide kasutuselevõtuks (Sõltumatu kõrgetasemeline tehisintellekti eksperdirühm; Sidevõrkude, sisu ja tehnoloogia peadirektoraat (Euroopa Komisjon), 2019).

rohkem (digitaalseid) andmeid kui kogu eelneva kasutusaja jooksul kokku (Turk ja Pild, 2019, lk 44). Üheskoos moodustavad need eriilmelised andmed nn suurandmed (ingl *big data*), mis oma mahu tõttu eeldavad uusi talletus- ja analüüsimeetodeid. Just suurandmetele on üldise digitaliseerimise valguses pööratud eraldi tähelepanu ning omistatud suur majanduslik ja ühiskondlik potentsiaal.[9] Tekkivaid digitaalseid andmeid ning ligipääsu nendele peetakse uueks (lokaalse ja globaalse) võimu määrajaks ja sümboliks.[10]

Üks andmete tähtsuse kuvandit kinnistada aidanud ilming on näiliselt tasuta eksisteerivad veebirakendused, eriti sotsiaalmeedia rakendused, mis kasutajate andmeprofiile ning teenuse kasutusmustrit hinnates on arendanud teenused kasutajatele meelepäraseks ning seeläbi saanud võimaluse manipuleerida kasutaja nähtava informatsiooni ja kasutajakogemusega.[11] Selliste laialdaselt kasutatavate, üldjuhul vähemalt osaliselt tasuta teenuste puhul[12] omistatakse tekkivate suurandmete omamisele nii otsene majanduslik kasu kui ka kaudne ühiskondlikke muutusi kaasa toov võim. Majanduslik kasumlikkus tuleb sellest, et teenust pakkuvad ettevõtted on kasutajate andmeprofiilide näol suurandmeid omades ning teenust vastavalt andmeanalüüsile kohandades ning sobivatele sihtgruppidele reklaami või informatsiooni vahendades muutnud end vajalikuks teenuseid pakkuvatele ettevõtetele. Kaudne suurandmete omamise ja kasutamise tulem on aga andmete omaniku mõjuvõim ühel üha suuremat tähtsust omaval infoväljal.[13] Võimalus infovooge kohandada tähendab kaude võimalust ühiskondlikesse protsessidesse sekkuda. Nii on suurandmete abil infovooge juhtivate firmade näol tekkinud kas ajakirjandusele omistatavale võimule uus väljendusvorm või lisandunud hoopis uus mõjusfäär.

Just IT laialdane levik ning interneti vahendusel toimuv suhtlus on võimaldanud andmeanalüüsi tehnikate ja analüüsitavate andmete mitmekesistumise. Kui enne laialdast digitaliseerimist olid lihtsasti kodeeritavateks ja analüüsitavateks andmeteks arvandmed ning kodeeritud tekstid ja tekstikorpused, siis digitaliseerimisega on erinevate materjalide kogumise ja võrdlemise võimalused laienenud. Kuigi võimalused on pea piiramatud, on erinevate digitaalsete andmete sisu analüüsi rakendamine, nagu suurandmete lähem vaatlus allpool välja toob, märkimisväärselt erinev. Et mõne andmeliigi, nt video, analüüs nõuab nii keerukama tehnoloogia olemasolu kui ka analüüsiks suuremat ajalist ressurssi, on peamised analüüsitavad andmed ikkagi arvuliselt tuvastatavad ja/või kodeeritavad andmed, k.a metaandmed. Seega, teksti ja numbriliste andmete omajatel on eelis.

Interneti rakendusi pakkuvad firmad on tänu oskuslikule suurandmete kasutamisele unikaalsel positsioonil, sest näiteks personaalset suhtlust ja selle sisu ei olnud seni võimalik nii detailselt analüüsida. Sellisest uuest võimekusest tingitult on ka eelmainitud majanduslik ja ühiskondlik võimupositsioon teenusepakkujatele mõeldav. Siiski, suurandmete analüüsi rakendamispotentsiaali nähakse ka juba toimivate teenuste ja rakenduste kaasajastamises ning kiiremaks muutmises. Loomulikult koondab kiirus enda alla ka opti-

---

9 Näiteks EL-i eetikasuunistes tehisintellekti arendamiseks kirjutatakse: „Usaldusväärne tehisintellekt võib parandada nii üksikisiku head käekäiku kui ka kollektiivset heaolu, luues jõukust ja väärtust ning suurendades rikkust“ (Sõltumatu kõrgetasemeline tehisintellekti eksperdirühm; sidevõrkude, sisu ja tehnoloogia peadirektoraat (Euroopa Komisjon), 2019, lk 10).

10 Kurbalija (2017) peab suurandmeid lausa uueks naftaks.

11 Näiteks märtsis 2020 oli Facebooki väärtus 54 miljardit dollarit, Sina Weibo väärtus oli 2020. aasta mais 19,6 miljardit dollarit ja Twitteril 2019. aasta novembris 22,5 miljardit dollarit. Vaata: www.macrotrends.net.

12 Näiteks muusika kuulamiseks kasutatav rakendus Spotify pakub võimalust nii tasuta muusikat kuulata kui ka Premium teenuse eest maksta. Esimese puhul on kasutaja sunnitud kuulama muusikale lisaks ka reklaamklippe, tasuline teenus annab võimaluse aga kasutajakogemust häirivast reklaamist hoiduda.

13 Näiteks Facebookist ja Twitterist leitava informatsiooni olulisusest ühiskondlike vaadete kujunemisel on palju räägitud nii 2017 Suurbritannia Euroopa Liidu referendumi kui ka 2017 Ameerika Ühendriikide presidendi valimiste kontekstis.

maalsuse, täpsuse, laiema haarde ja paindlikkuse. Nii on suurandmete analüüsimisel ka eelmainitust erinevaid, nii otseseid kui ka kaudsemaid võimalusi ühiskondlikke protsesse mõjutada. Suurandmete potentsiaal seisneb juba toimivate teenuste ja protsesside puhul selles, et lisaks uute tekkivate andmete analüüsile, mis teenust muuta või kohendada võimaldab, võib ka juba olemasolevate andmete digitaliseerimine ja/või uut moodi kasutus anda tähendusrikkaid tulemusi. Nimelt annab juba olemasolevate andmete lisamine ja analüüs võimaluse tuvastada ka uusi näitajaid, mis muutuvad ilmseks ja/või märkimisväärseks ainult juhul, kui hinnatud on väga suurt hulka andmeid. Pealegi võivad erinevaid andmestikke koos kasutades ilmneda suhteseosed, mida pole vaid ühte andmestikku kasutades ja uurides võimalik näha. Nii on suurandmete analüüsiga võimalik toiminguid juhtida ning muuta, tuginedes reaalajas (juurde) tekkivale informatsioonile ning suhtemustritele, mitte toetudes vaid toimunul põhinevatele prognoosidele.

Suurandmete ilmnemise ning kaasnevate analüüsipraktikate näol on seeläbi tegemist paradigma muutusega protsesside planeerimises ja juhtimises, kus andmed kinnitavad hüpoteesi või prognoosi paikapidavust (isegi kohe) (Kitchin, 2014, p. 4). Ka ühiskondlikke protsesse või avalike teenuste toimimist saab sobivate tark- ja riistvara lahenduste olemasolul hinnata nüüd reaalajas. Varem oli nende hindamiseks vaja koguda statistikat ning analüüsiprotsessiks vaja seega aega. Selline paradigma muutus toob kaasa arengud ja praktikate teisenemise kõikides valdkondades, kus andmeanalüüsil tuginev informatsioon saab muuta protsessi täpsust ja eesmärki. Tänu andmehulga suurenemisele ning andmete võrdlemisele on näiteks välja arvestatav inimese käitumise tõenäosus (Mehozay & Fisher, 2019, p. 536). Sellist analüüsi on sarnaselt võimalik rakendada nii tarbimiskäitumist kui ka illegaalset tegevust hinnates. Nii eeldatakse, et digitaalsete andmete analüüsile toetuvad ning nende kasutamiseks arendatavad tehnoloogilised lahendused, k.a tehisintellekti süsteemid[14] (ingl *artificial intelligence* ehk *AI*), saavad oluliseks osaks riiklikes teenustes, sh rakendatakse neid turvalisema elukeskkonna tagamiseks näiteks kuritegude tuvastamiseks[15], aga ka ennetus- ja tõkestustöös.[16]

Et uudsed tehnoloogilised lahendused tähendavad nii teadusinnovatsiooni kui ka majandusarengut, on tehisintellekti süsteemide arendamise näol suurandmete analüüsi rakendamisest saanud osa uue tehnoloogilise võimekuse ja majandusliku ülemvõimu kinnistamise protsessist, mille nimel riigid pingutavad. Ka Euroopa Liit institutsioonina on seadnud sellisesse uudsesse tehnoloogiasse investeerimise endale prioriteediks ning on seda erinevate meetmetega toetamas (Euroopa Komisjon, 2018). Et ootused tehisintellekti süsteemide kiirele arengule on suured, on juba pandud suur uurimisressurss autonoomsete süsteemide agentsuse küsimuste lahendamiseks, nt iseliikuvate autode tehtavate valikute moraalsed ja õiguslikud alused.

Tulevikuvisioonist hoolimata on tehisintellekti süsteemide ehk krattide[17] puhul täna aga tegemist ennekõike nn kitsaste tehisintellekti süsteemidega (ingl *narrow AI*), st nad on

---

14 *Eetikasuunised usaldusväärse tehisintellekti arendamiseks* toovad välja, et tehisintellekti süsteeme on erinevaid: tegemist ei pea olema ainult väga kõrgelt arenenud süsteemiga (nt autonoomsete masinatega), vaid siia alla kuulub ka suurandmeid analüüsiv (iseõppiv) algoritm. (Sõltumatu kõrgetasemeline tehisintellekti eksperdirühm; sidevõrkude, sisu ja tehnoloogia peadirektoraat (Euroopa Komisjon), 2019).

15 Näiteks *lisa järgmise dokumendi juurde: Komisjoni teatis Euroopa parlamendile, Euroopa Ülemkogule, Nõukogule, Euroopa majandus- ja sotsiaalkomiteele ning regioonide komiteele. Tehisintellekti käsitlev kooskõlastatud kava* nimetab tehisintellekti tehnoloogiate arengu üheks uueks võimaluseks kuritegude (näiteks rahapesu ja maksupettuste) tõhusama avastamise (Euroopa Komisjon, 2018, lk 3).

16 Näiteks liiklusvoo hindamine ning vastavalt sellele tehtavad liikluskorralduse muutused on saanud võimalikuks tänu tehnoloogia arengule.

17 Eestindatud versioon sõnast tehisintellekt on kratt (Majandus- ja kommunikatsiooniministeerium, 2019).

rakendatavad kindla ja/või määratletud ülesande lahendamiseks.[18] Kitsaste tehisintellekti süsteemide kasutusvaldkonnad on näiteks soovituste genereerimine (k.a Google'i otsingu vaste, mis on suhestatud eelnevate otsingutega), tegevuse täpsuse tõstmine või prognoosimine,[19] aga ka näiteks näotuvastustarkvara (k.a näotuvastus, mis nutitelefoni lukustab või turvakaamera videolt inimesi tuvastab). Just selliseid suurandmete analüüsi kasutamisel ja täiustamisel põhinevaid süsteeme tahetakse edaspidi arendada ning nii Eestis kui ka EL-is rakendada. (Siin võimegi mõelda näotuvastusprogrammist või viisist, kuidas maksupettusi tuvastada.)

Tulevikult oodatav lai tehisintellekt on vastukaaluks kitsale „rakendus või süsteem, mis on võimeline lahendama inimesele omaselt mis tahes ülesandeid" (Majandus- ja Kommunikatsiooniministeerium ja Riigikantselei, 2019, lk 15). Need on Sophia-sarnased (Hanson Robotics Ltd., 2020) autonoomsed ning füüsiliselt oma keskkonda tajuvad, tihti välimuselt inimest meenutavad masinad. Selliste masinate üldkasutatavaks muutumine nõuab aga nii tark- kui ka riistvara suurt tehnoloogilist arengut ning õigusruumi küsimustega tegelevast uurimusest hoolimata arendatakse täna alles esimesi prototüüpe.

1.2. MAAILMATRENDID KOHALIKUS OLUSTIKUS: EESTI TEHNOLOOGILINE TULEVIKUVISIOON

Eesti on sarnaselt üldisele globaalsele trendile ennustanud juba aastaid IKT positiivset mõju ühiskonna kõikidele tahkudele ja liikmetele.

> *Ettevõtetel lubab IKT kasutuselevõtt optimeerida nii tänaseid äriprotsesse kui ka luua sootuks uusi ja innovaatilisi tooteid ja teenuseid. Üksikisikule annab IKT juurdepääsu lõpututele infovaradele […], mis avardavad märkimisväärselt võimalusi enda arendamiseks ja oma heaolu suurendamiseks.[20] Avalike teenuste korraldamine ja osutamine IKT-lahendustega võimaldab suunata maksumaksja raha valitsusasutuste haldustegevustelt sisuliste ülesannete lahendamisele, luues ühtlasi eeldused suhtlemise lihtsustamiseks.*
>
> *[…] IKT saab olla oluline tööriist majanduskasvu suurendamisel ja inimeste elukvaliteedi tõstmisel (Majandus- ja Kommunikatsiooniministeerium, 2013, lk 4).*

2013. aastal kinnitatud infoühiskonna arengukava vaimus on Eesti tehnoloogilist arengut ja IKT vahendite kasutuselevõttu saatnud mõtteviis, et Eestis on muutuste elluviimine võimalik; uued ja seni laialt levimata lahendused saab kasutusele võetud.[21] Riiklikke digitaalseid teenuseid võimaldava tarkvara areng on olnud kiire. Eesti kuvand maailmas on tänu uudsete digitaalsete laheduste pakkumisele – näiteks e-residentsus ehk võimalus saada digitaalse e-Eesti elanikuks, elektroonilised valimised, andmesaatkond ja X-tee abil koos toimivad digitaalsed avaliku- ja erasektori teenused – kui üks uuendusmeelsemaid

---

18 Tavakõnes olev diskussioon algoritmide abil andmete analüüsist on just see, mida siin silmas peetakse.

19 Majandus- ja Kommunikatsiooniministeerium, 2019. a.

20 *Eesti infoühiskonna arengukava 2020* (Majandus- ja Kommunikatsiooniministeerium, 2013, lk 9) toob samas välja, et eestlased kasutavad Internetti ennekõike suhtluseks, meelelahutuseks ja informatsiooni otsimiseks, mitte aga enesetäiendamiseks. See lubab eestlasi pidada nii passiivseteks kui ka sisukriitilisteks Interneti teenuste tarbijateks.

21 Vaata Vaarik, 2015.

ja ainulaadsemaid riike.[22] Just taristu aeglast arendamist peetakse takistuseks, miks digitaalsed teenused Eestis veelgi suuremat kasutust ei ole leidnud.[23]

Hoolimata omanäolisest tehnoloogilisest arengust ning vaatamata koostoimivatele andmekogudele sarnanevad Eesti tulevikuvisioon, nt projekt #Bürokratt[24], või tehnoloogilised lahendused korrakaitses, nt kitsaste tehisintellekti süsteemide kasutuselevõtt ennetus- ja tõkestustöös, maailmas domineerivale (tehnoloogia) diskursusele ning suurandmete töötlemisele ja tehisintellekti süsteemidele pandud ootustega. Ka Eestis tuvastatud tehnoloogia kasutusest tingitud ees ootavad probleemid, nt privaatsust suurendavate tehnoloogiate kasutamise kasv ning tulevikuohud, nt küborgid ja kvantarvutid[25], on ka maailmas sarnaselt kajastust leidnud.[26]

Globaalselt tuvastatud arengusuunad ning välja toodud probleemid tõukuvad domineerivatest diskussioonidest teaduses ja ekspertgruppide tööst. Rahvusvaheline teadusmaastik, mis tehnoloogia mõju ja uusi ilminguid analüüsib, aga ei suuda ja ka ei saa sammu pidada tehnoloogia arengu ning sellest tõukuvate ühiskondlike probleemidega. Probleemsed nähtused ilmnevad tihti alles tehnoloogia kasutuselevõtul ja laiahaardelisel kasutusel, nt internetipanga pettused on võimalikud vaid siis, kui internetipank laialdaselt kasutusel on; sotsiaalmeedia illegaalsete tehingute vahendajana võetakse kasutusele, kui kasutajaid on palju jne. Samuti ei too globaalsed ilmingud alati välja kindlate riikide tehnoloogiakasutuse eripärast tingitud kitsaskohti.

Loomulikult tuleb igal riigil teada maailmatrende ning võimalikke riske ja arenguid oma tegevuses ning tulevikuvisioonis arvesse võtta. Nagu Euroopa Liidu tehisintellekti kava viitab, on paljud tehnoloogilised arengud ning nende mõjud riikide ülesed (sõltumatu kõrgetasemeline tehisintellekti eksperdirühm; sidevõrkude, sisu ja tehnoloogia peadirektoraat (Euroopa Komisjon), 2019). Näiteks küberkuritegevuse haare on globaalne ning selle tuvastatud arengusuunad mõjutavad kõiki riike, st kuritegelikud organisatsioonid ning kuriteod on riigipiiride ülesed, kuriteo ohvriks võib langeda igaüks. Lähtudes kohalikus praktikas ja tulevikuvisioonis aga ainult sellest, mis maailmas toimumas, ning tehnoloogiatest, mis laialdaselt kasutusel, tähendab, et Eesti jaoks võib kohaliku süsteemi ladusaks ja turvaliseks toimimiseks ning uute rakenduste kasutuselevõtuks oluline õigel ajal märkamata ja tegemata jääda.

Näiteks ID-kaart ning sellega tehtavad toimingud on Eestis igapäevased ja nende toimimine oluline. 2017. aasta kevadel Eestis välja kuulutatud hädaolukorra seaduses on elektrooniline isikutuvastamine ja digitaalne allkirjastamine loetud Eesti elanikele elutähtsate teenuste hulka.[27] Samas ID-kaardi tarkvara probleemid ning sellega kaasnenud turvarisk, mis 2017. aasta sügisel sundisid riiki suurt hulka ID-kaartide sertifikaate peatama ja neid kehtetuks kuulutama või omanikku uuendusi tegema, tähendas, et mingi aja vältel ei olnud kõigi ID-kaardi (kui kohustusliku isikut tõendava dokumendi) omanikel

---

22 Siiski ei tohi pidada Eestit ainulaadseks; mitmed Aasia riigid on oma tehnoloogiliste lahendustega oluliselt innovatiivsemad.

23 2013. aastal kinnitatud arengukavas eeldati, et aastal 2020 on igal pool valguskaabel ja ülikiire internet. Tegelikult on sellega üpris keerulised lood.

24 #Bürokratt võimaldab „tulevikus inimesel mistahes seadmest ja virtuaalse assistendi kaudu saada ühe suhtlussessiooniga kõik vajalik korda aetud“ (Sikkut, Velsberg ja Vaher, 2020, lk 2). See tähendab, virtuaalne assistent suhtleb erinevate institutsioonide tehisintellektisüsteemide ehk krattidega – tehisintellekti süsteemide koostoime; inimene saab pelgalt kõnekeelse suhtlusega avalikke teenuseid kasutada (*Ibid*).

25 Näiteks Siseministeerium, 2020.

26 Näiteks Bundeskriminalamt, 2020; Majandus- ja Kommunikatsiooniministeerium, 2019.

27 Peatükk 5 § 36. Elutähtsate teenuste loetelu ja nende toimepidevust korraldavad asutused (hädaolukorra seadus, 2017).

võimalik e-teenuseid tarbida. Sertifikaatide peatamine pärssis digiriigi kodanike ja elanike ligipääsu teenustele.

Aastal 2017 tuvastatud turvarisk ei olnud küll Eesti-spetsiifiline, vaid tingitud Eesti ID-kaardis kasutatavaid mikrokiipe tootva firma veast kiipide krüpteerimistarkvaras. Nii olid sarnaste turvariskidega ka sama firma tehnoloogiat kasutavate firmade tooted ning neil põhinevad rakendused, nt arvuti emaplaatide salasõnade salvestised, pangakaardid ja autentimislokid (ingl *authentication token*) (Khandelwal, 2017; e-estonia, 2018). Kui turvariski sisaldavate teenuste puhul on aga üldjuhul tegemist kindla firma teenuse või tootega, mille kasutajaskond on küll üleilmne, kuid piirdub kindla tootega ja nii võib olla turvariskiga vaid osa inimese informatsioonist, siis Eesti ID-kaardi puhul tähendas see otsest turvariski nii riigile kui ka elanikkonnale. Selline suuremahuline – võimalik, et riiki halvava võimega – turvaaugu ilmsikstulemine näitab e-riigi haavatavust, riikide sõltuvust välistest teenusepakkujatest ning tõstatab küsimuse ka riigi usaldusväärsusest.

Võimalike riist- ja tarkvaraprobleemide tuvastamise ning riskide maandamisega tegelevad aga ainult inimesed ja teenusepakkujad, kes on otseselt asjaga seotud või asjast huvitatud. Isegi sellisel juhul nagu krüpteerimiskoodi viga ei saa aga eeldada, et laiem maailma digikogukond oleks just Eesti ID-kaardist teadlik ning selle võimalikest turvariskidest räägitaks laialdaselt; et riikide kasutatavate lahenduste turvariskidele pööratakse erilist tähelepanu. Vastupidi, riikliku tähtsusega teenuste turvalisuse tagamine on üks valdkond, kuhu tuleb riigil endal panustada – olukorda analüüsida ja seirata – et ennetada võimalikke probleeme ja tagada teenuse säilimine.

Wolfgang Drechsler toob oma 2018. aasta artiklis „Pathfinder: e-Estonia as the β-version" välja, et kuigi Eesti on uute tehnoloogiate kasutuselevõtus innukas, siis nende lahendustega kaasnevad riskid ei ole tihti läbi mõeldud. Nimelt ei lähtuta tehnoloogia kasutuselevõtul mitte niivõrd selle tehnoloogilise protsessi eripäradest, kuivõrd võimalusest juba olemasolevat protsessi tehnoloogia abil juhtima asuda.[28] Digitaliseerimine ja suurandmete kasutus ei ole siin erand. Ka Eestil on soov kasutada tekkinud andmete potentsiaali ära ja realiseerida maailmas mainitud võimalusi (Majandus- ja Kommunikatsiooniministeerium ja Riigikantselei, 2019, lk 67). Uute tehnoloogiliste võimaluste, k.a suurandmete analüüsi ning omavahel veelgi paremini ühilduvate ja toimivate andmete, kasutamine ja selleks tehisintellekti süsteemide arendamine ning rakendamine on visioonina kirjas[29]; kaude soosivad sellist ideed ka sisejulgeoleku arendamist plaanivad dokumendid.[30] Uutest tehnoloogilistest lahendustest räägitakse aga kui innovatsioonist, mitte võimalikest proovikividest või muutustest, mida tuleks lahenduste kasutuselevõtuks arendada. Nii on Eesti kitsaste tehisintellektisüsteemide kasutuselevõtu plaanid tiivustatud globaalsest suurandmete potentsiaalist, hindamata olemasolevate suurandmete olemust ja sellest tulenevaid katsumusi.

---

28 ingl *tech-driven rather than tech based* (Dechsler, 2018, lk 16).

29 Näiteks: Siseministeerium, 2019; Justiitsministeerium, 2019b.

30 Suurandmete võimaldatava andmete analüüsiga kaasneb ka teatav hirm, sest andmete kogumist samastatakse jälitustegevuse ning autoritaarse kontrolliga inimtegevuse üle. Seda hirmu saab maandada vaid selgitades suurandmete olemust ning sellest tulenevalt ka vägagi piiratud kasutusviise.

2. SUURANDMED: TEOREETILINE KÄSITLUS

Mõistmaks, mida digitaliseerimisega kaasnevad andmed võimaldavad, on vaja avada suurandmete mõiste. Kuigi suurandmetele on iseloomulik nende maht, on nende tekkeviis ja sellest tingitult ka ligipääsu- ja kasutusvõimalused erinevad. Suurandmete analüüsi rakendamise potentsiaali hinnates tuleb andmete eripärast teadlik olla ning sellele tähelepanu pöörata. Olles hetkel erasektoris kasumlik ei tähenda see, et sarnane andmete kasutus oleks tingimata võimalik ka avaliku sektori teenustes. Samas avaliku sektori hallatavate andmete maht kasvab pidevalt ning uued analüüsimeetodid võivad drastiliselt muuta seda, kuidas riigi hallatavaid tegevusi planeeritakse ja pakutakse.

2.1. SUURANDMETE MÕISTE

Suurandmed [vahel kasutatakse ka *massandmed*] moodustavad kõik digitaliseerimisega tekkinud andmed, mille analüüsimiseks on nende mahu tõttu vaja kasutada uusi meetodeid. Kuigi mõiste *suurandmed* viitaks justkui andmete homogeensele massile, iseloomustavad neid peale suure mahu[31] aga just eriilmelisus, andmete üldise struktuuri puudumine ja andmehulga pidev muutumine (Drewer & Miladinova, 2017, p. 299).

Suurandmeid iseloomustavateks tunnusteks loetakse peale andmemahu ja struktuurituse eriilmelisuse veel nende tekkimise või andmebaasi osaks saamise kiirust: digitaalsete toimingute puhul muutub tegevus andmeteks (peaaegu) reaalajas. Samuti on need oma ulatuselt ammendamatud, nt valim võib olla kogu elanikkond; suurandmed on oma olemuselt väga tihedad, kuid vajadusel kergesti eristatavad; nad on üksteisega suhtes ning nad on paindlikud ehk võivad muutuda, nt neile võib lisanduda omadusi, või nad võivad drastiliselt oma mahtu suurendada (Kitchin, 2014, p. 2).

Suurandmete eriilmelisuse tõttu ja vastavalt uurimisküsimusele või kasutuskontekstile võib neid liigitada mitmeti; ühene metodoloogia suurandmete rühmitamiseks puudub.

Andmete tekkeviiside järgi saab suurandmed jagada järgmiselt (Kitchin, 2014):

- otsesed andmed ehk andmed, mille kogumine on eesmärgistatud, nt andmebaaside sissekanded;
- automaatsed andmed ehk andmed, mis tekivad automaatselt mingisuguse toimimisprotsessi käigus, nt süvaandmed digiteenuse kasutamise kohta;

[31] Aastal 2019 kogunes ainuüksi Facebookist iga päev 4 betabitti uusi andmeid (Smith, 2019).

- vabatahtlikud andmed ehk andmed, mille tekkimine on seotud vabatahtlikult teenuse tarbimisega, nt interneti uudistekommentaariumi postitus.

Andmete loojate või tekkepõhjuste järgi seevastu võib andmekogud jagada järgmiselt (Zwitter, 2014):

- loomulike agentide (ingl *natural actors*) tegevusest tekkinud andmekogud, nt inimene, kes sotisaalmeediat kasutab;
- tehisagentide (ingl *artificial actors*) tegevusest tekkinud andmekogud, nt 'küpsised' (ingl *cookies*), mis kasutaja internetitegevust jälgivad;
- füüsiliste nähtuste (ingl *physical phenomenon*) tagajärjel tekkinud andmekogud, nt aktiivsusmonitori andmed inimese südametöö kohta, ilmastikuandmed jne.

Samuti võib suurandmeid jagada kogumise ja talletamise viisi järgi struktureeritud ja struktureerimata andmeteks. Struktureeritud andmed on andmed, mille jaoks on välja töötatud süsteem, nt kodanike register – kõik uued kogutud andmed asetuvad sel juhul olemasolevasse süsteemi. Struktureerimata andmed aga kogutakse ja talletatakse nende tekkimisele vastaval kujul, nt uudiste portaali kommentaarid või videokaamera salvestised. Struktureerimata andmete potentsiaal seisneb võimaluses kogutud andmeid liigitada analüüsiks mitmeti. See eeldab aga kogutud andmete sisu analüüsi: lisatööd ja keerukama või intelligentsema tehnoloogia kasutust.

Lühidalt, suurandmed on kõigest koondnimetaja digitaalsetele andmetele, mille tekkimise ning kogumise on võimaldanud uue tehnoloogia kasutuselevõtt. Tekkeviisi, andmete looja või andmete korrastatuse põhjal andmeid eristada pole ainsad võimalused – andmete kasutamise vajadusest oleneb ka viis, kuidas neid täpselt määratleda. Seega, uusi tehnoloogilisi lahendusi plaanides on oluline täpsemalt teada, missuguseid andmeid saab ja on plaanis kasutada; pelgalt suurandmetele viitamisest ei piisa, et avada plaanitava protsessi toimimist.

2.2. SUURANDMETE OLEMUS

Suurandmete analüüsi kasutuselevõtu võimaluste hindamiseks ning kitsaskohtade mõistmiseks (just) avaliku sektori teenistuses on oluline esmalt kaardistada, missugustel erinevatel viisidel elanikud – ühe digitaalseid teenuseid pakkuva ja võimaldava riigi kontekstis – andmeid loovad ja toodavad. Samuti tuleb tähelepanu pöörata sellele, kes andmeid haldab ja talletab. Era- ja avalikul sektoril on erinevad andmete kogumise põhjused ja põhimõtted. Lisaks ei piirdu digitaalsete teenuste globaalset haaret arvesse võttes kasutajaskond alati riigi füüsilise piiriga, mistõttu on mõne teenuse või rakenduse puhul ühe riigi kodanike andmed talletatud neist füüsiliselt väga kaugele, hoopis kolmandasse riiki. Nagu Eesti e-residentsus tõestab, võib ka riigi digitaalne piir olla laiem füüsilisest – riigi virtuaalruumis teevad toiminguid kodanikud, kelle kodanikuks saamise ja olemise huvi ongi kõigest virtuaalsete toimingutega seotud; füüsiliselt nad riigis viibima ei pea ja kohalikke teenuseid ei tarbi.

Kuigi andmeid tekib riigi kontekstis palju, ei saa ligipääsu kogutud andmetele võtta isegi riigi institutsioonide seisukohalt iseenesest mõistetavalt. Teenuse pakkujal on nii kohustus kui ka vastutus kogutud andmetega seaduslikult/heaperemehelikult ümber käia, mis võib riigi huvidega vastuollu sattuda.

Olgugi et andmeid koguvad nii riik kui ka erafirmad, ei ole riigil võimalust erafirma andmetele (mõjuva põhjuseta) ligipääsu saada ning samuti ei saa riik firmasid kohustada neile andmetele ligipääsetavust tagama. Andmed saab riik küll kohtu otsusega välja nõuda[32], sarnaselt IKT võimaldavale jälitustegevusele, kuid kui ka firmal puudub andmetele ligipääs, siis tähendab tagauste loomine eetilist dilemmat.

Üks lähiaja enimkajastatud juhtumeid, mis kõnealust dilemmat näitlikustab, on seotud tehnoloogiakompanii Apple ja USA julgeolekuteenistuse FBI vastasseisuga. Nimelt keeldus Apple arendamast FBI-le tarkvara, mis võimaldaks krüpteeritud telefoni andmetele ligipääsu. Sellise nõudmise ajendiks oli FBI vajadus vaadata telefoni talletatud andmeid, mille omanikuks oli USAs 2015. aastal terrorirünnaku toimepanemises kahtlustatav. Apple põhjendas oma otsust vajadusega kaitsta oma kliente võimalike edasiste turvariskide, andmete lekkimise ning asjatu/õigustamata jälitustegevuse eest. Lisaks minetab firma oma usaldusväärsuse, andes riigile ligipääsu teenustele, mille turvalisust ja konfidentsiaalsust kliendid hindavad. FBI leidis küll kolmanda osapoole, kes teadete kohaselt FBI-d telefoni avamisel abistas, kuid Apple pole seni krüpteeritud andmetele ligipääsu pidanud võimaldama (Veistennson, 2020).

Vastupidise näitena on aga mitu riiki, k.a Euroopa Liit, nimetanud Hiina tehnoloogiakompanii Huawei tooted ja teenused ebaturvaliseks ning seda just võimalike kolmandate osapoolte, ka väidetava Hiina RV luureteenistuse, andmetele ligipääsetavuse tõttu (NIS COOPERATION GROUP, 2019). Huawei seadmeid ning nende vahendatavaid teenuseid tarbides seab teenuse kasutaja ohtu nii iseenda (andmed) kui ka oma kontaktid; riigi institutsioonide puhul on aga ohus kogu infrastruktuur ja/või riigi kodanike andmed. Valik teenust tarbida ning end võimalikku ohtu seada on aga riigil/inimesel endal.

Nagu suurandmete olemuse lähem vaatlus välja toob ja millele ka Huawei juhtum viitab, ei ole kõigil digitaalsetel andmetel aga riigipiire ning paljude populaarsete teenuste ning seadmete arendajaks on rahvusvahelised suurfirmad; andmed on talletatud kolmandates riikides asuvates serverites või on krüpteeritud. Nii võivad riigi või riikide ühenduse seadused ja volitused jääda liiga lokaalseks, et andmetele ligipääsu isegi taotleda.

Et mõista täpsemalt, kui eriilmelised on tekkivad suurandmed isegi ühe (digitaalselt arenenud) riigi kontekstis, jagan digitaliseerimisprotsessi ja uue tehnoloogia kasutust arvesse võttes riigi suurandmed moodustavad andmed mõtteliselt järgmiselt:

- digitaliseeritud andmebaasid ja nende kasutamisest tekkivad metaandmed (ehk andmed, mis ei salvesta mitte otsest tegevust, vaid selle koordinaate jne[33]);
- internetti toetavate rakenduste kasutamisel tekkivad andmed ning kasutamist salvestavad metaandmed;
- digitaalsed seadmed ja asjade interneti ehk nutistu[34] (ingl *Internet of Things ehk IoT*) tehnoloogia, mis andmeid ja metaandmeid toodab.

---

[32] Siiski, andmetele ligipääsuks on tuvastatud konkreetsed mõjuvad põhjused. Nt Euroopa andmepoliitikat reguleeriv GDPR toob andmetele ligipääsemiseks välja terrorismikahtluse.

[33] Metaandmed on järgmised: (a) kirjeldavad: faili nimi, autor, tekkimise aeg; (b) õigused: nt autoriõigusi puudutav informatsioon; (c) tehnilised: andmed, mis kirjeldavad faili tehnilisi detaile; (d) talletamisega seotud andmed: nt andme asetsemine andmepuus; (e) markeerimise informatsioon: nt informatsioon, mis viitab navigeerimisele ja kooskasutusele (Chapple, 2020).

[34] Nutistu ehk IoT esemed on kõik „nutikad" esemed, mis saavad oma võimekuse võrku ühendamisest, k.a nutitelefonid.

2.2.1. Digitaliseeritud andmebaasid

Esimest liiki andmed, mis moodustavad tinglikult riigi suurandmed, on andmekogud ja andmebaasid, mis on seotud teenustega või süsteemidega ning on tänu tehnoloogia arengule muudetud digitaalseks. Sellised andmekogud ja andmebaasid on struktureeritud ehk määratud on kriteeriumid ja protsessid, võimalik et ka seadusandlus, mille alusel andmeid kogutakse ja talletatakse (Tam & Kim, 2018, p. 580). Sellised struktureeritud andmebaasid on tihti kas retrospektiivselt digitaliseeritud või sisaldavad retrospektiivselt digitaliseeritud andmeid. Järelikult, andmete digitaliseerimisel võib olla tehtud teadlik valik, mida „vajalikuks ja talletamisväärseks" pidada. Digitaliseeritud andmebaase esineb nii era- kui ka avalikus sektrotis. Näiteks rahvastikuregister või panga klientide andmestik, raamatukogu andmebaas või eestikeelsete taimenimede andmebaas.

Need andmebaasid ei ole üldjuhul tehnoloogilise arengu ja digitaliseerimise otsene tulemus, vaid tänu uutele tehnoloogilistele lahendustele on andmed muutunud kõigest digitaalseks ja nii hõlpsasti kasutatavaks, vajadusel analüüsitavaks. Digitaliseerimise puhul peetakse suureks riskiks just andmete sisestamisel ilmnevaid puudujääke, k.a inimlikke vigu (Fernandes, *et al.*, 2014, p. 432; Griffard, 2019, p. 53); grupeerimise metodoloogia õigsust (Lee & Estivill-Castro, 2011, p. 364); seda, et andmete kogumisel on ebaloomulikud piirid, ja nii edasi. Samas, tegemist on andmebaasidega, mille puhul – vähemalt, kui tegemist on otseselt inimesega seotud andmetega – on inimene selles olemisest või selle eksisteerimisest teadlik, sest andmete kogumine on otsene. See ei tähenda küll, et kõik andmebaasid oleksid vabatahtlikud, kuid inimesel on selge, milline on andmete sinna kandumise põhjus. Andmebaasi haldaja on samuti üldiselt kergesti tuvastatav, sest teenus, mida ta pakub, on otse või kaudselt andmete kogumise ja kasutamise või hiljem andmebaasiks korrastatud andmetega seotud. Digitaalsete lahendustega kaasnevad sellistele andmebaasidele aga ka metaandmed, mis andmebaaside kasutuse talletavad. Nende metaandmete tekkimisest võib ei pruugi kasutaja olla teadlik või võib teada vähe. Kuid need annavad pigem võimaluse luua andmebaasile ning teenuse kasutajale lisaväärtust täiustatud teenuse näol.

2.2.2. Interneti (rakenduste) kasutamisel tekkivad andmed

Teine liik andmeid, mis suurandmed moodustab, on uute infotehnoloogiliste vahendite tekkimise ja kasutamise tulemus ehk andmed, mis tekivad interneti kasutamisel. Maailmas oli 2019. aasta detsembri seisuga 4,574,150,134 interneti kasutajat – kõigi nende kohta on mingisugune digitaalne jalajälg olemas (Internet World Statistics, 2020).

Interneti kasutamisest tekkinud andmed võib oma olemuse järgi jagada kaheks.

1) Ühe hulga moodustavad andmed, mis tekivad internetti ja selle rakendusi kasutades. Nende andmete tellimisest on inimene põhimõtteliselt teadlik. Need on vabatahtlikud andmed, sest kuigi paljud teenused pakuvad interneti rakenduse kasutamise võimalusi, on nende kasutamine siiski vabatahtlik. Just selliste andmete kasutus on tänu internetipoodidele ja suhtlusrakendustele suurandmete potentsiaali kõikides eluvaldkondades päevakorda tõstnud. Loomulikult on digitaalses ühiskonnas keerukas interneti kasutamist vältida, kuid see pole võimatu. Viisid, kuidas interneti kasutusel andmed tekivad, on erinevad.

    a. Interneti rakenduse abil (riikliku) teenuse kasutamise puhul on andmete tekkimine loomulik. Nt maksuamet või Amazon.com – otsesed andmed, mis selle teenuse kasutamisest tekivad, suunatakse olemasolevatesse andmebaasidesse. Kuna paljud uued teenused, nt Amazon.com, on vaid internetipõhi-

sed, siis võib andmete tekkimine kui paratamatu osa teenusest jääda inimestel märkamata. Siiski on selline andmete kogumine teenust tarbides arusaadav ja mõistetav, mõnel juhul on see ka kinnituseks salvestatud, nt e-poe ostu puhul.

b. Interneti rakenduste abil suhtlemise jne puhul tekivad seevastu andmed, mille puhul ei pruugi kasutaja ise oma tegevust pidada selliseks, mille tulemusena analüüsitavad ja analüüsimist väärt andmed tekiks. Sellist liiki andmetega on tegu näiteks internetivestluste, kommentaaride, saadetud piltide või „meeldib" nupuvajutuse näol. Teenusepakkujal on võimalus aga ka need andmed kokku koguda ja hiljem, võimalusel ja vajadusel, neid analüüsida.[35] Just teenuse sisu kujundades võtavad paljud firmad, nt Google või Facebook, sellise andmeanalüüsi tulemuse aluseks.

2) Teise hulga andmeid moodustavad metaandmed ehk andmeid kirjeldav informatsioon, mis tekib internetti kasutavaid rakendusi kasutades, kuid mille tekkimisest ei ole inimene põhimõtteliselt teadlik (Yar 2009: 143). Sellised süvaandmed on kursori liimumine või külastatud lehekülgede arv ja olemus, sõidujagamisteenus Bolti sõiduteekond või Spotify kuulamiste arv. Sellised automatiseeritud andmed võivad olla anonüümsed (või neid kogutakse anonüümselt), aga tänu internetitoimingutega kaasnevale informatsioonile on need andmed isikustatavad. Siiski, üha rohkem metaandmeid on juba tekkides isikustatud teenuse lisaprodukt või -väärtus, nt Bolti sõidu(jagamis)teenust saab kasutada ainult kasutajanime ja krediitkaarti omades, mistõttu on sõiduteekonna info isikuga seotud.[36] Seega on teenusepakkujal inimeste harjumusi lihtsam kaardistada ning seda teenust pakkudes arvesse võtta.

Digitaliseeritud teenuste interneti rakenduste kasutamise puhul tekivad otseselt struktureeritud andmed: see on teenustega kaasnevate digitaliseeritud andmebaaside edasine rakendus. Teiste interneti rakenduste kasutamise puhul, nagu ka metaandmete puhul, on tegemist aga struktureerimata andmetega ehk nende andmete maht ja olemus ei ole enne nende tekkimist kindlaks määratud. Kuna internetiteenused on tihti riikide ülesed, siis oleneb kogutavate andmete hulk kasutajabaasist. Globaalsete teenuste puhul on andmete analüüs ning nende tulemusel pakutavad teenused tavaliselt teenusepakkuja otsene teenimisvõimalus; andmeanalüüs aitab muuta teenuse või toote kasutajale meelepärasemaks ehk seda tarbitakse rohkem. Erafirmade kogutavatele andmetele, eriti veel globaalsete firmade andmetele, riigil otsest ligipääsu ei ole – kliendibaasi- ja ärihuvide kaitsmiseks on igal teenusepakkujal õigus andmeid talletada viisil, et kolmandad osapooled neile põhjuseta ligi ei pääseks. See suunab paljud interneti rakendustest kogutavad andmed erasektori hallata ja kasutusse. Riiklike teenuste planeerimisel võib oletada küll erafirmade käsutuses olevatele andmetele ligipääsu saades nende kasu. Samas, riigi- ja erasektori teenuste olemus erineb suuresti ning andmed ja analüüsimise meetodid on erinevad, mistõttu võivad paljud erasektori kogutud andmed osutuda riigile kasutuks.

Oluline on ka mõista, et kuigi inimene on interneti rakendusi kasutades saanud tahes-tahtmatult üheks nendest, kes analüüsitavaid andmeid loob, ei ole ta nende andmete tekkimise protsessist tegelikult teadlik. Ta ei tea, missuguseid andmeid kogutakse ja kuidas neid edasi kasutatakse. Interneti rakendusi kasutades kogutavad andmed on isikuand-

---

35 Loomuliku keele töötlus (ingl *natural language processing)* – arvuti ja inimese vaheline suhtlus; arvuti mõistab inimese kõnet ja kirja. Tänu tarkvara arengule on üha keerukamate tekstide analüüs arvutite abil võimalik.

36 Mõistmaks nende tekkivate andmete mahtu – Eesti Statistikaameti 2019. aasta oktoobris avaldatud andmete kohaselt broneeris veebilehe või mobiilirakenduse kaudu sõiduteenuse (nt Bolt või Yandex) iga kolmas ja majutusteenuse (nt AirBnb.com või Booking.com) iga neljas internetikasutaja (Statistikaamet, 2019).

med ning neid tuleb ka vastavalt käidelda, mistõttu esitatakse teenuse kasutamise alguses potentsiaalsele kasutajale teade kasutajatingimustega (*ingl Terms and Conditions ehk T&C*), kus isikuandmete kogumisest ja käitlemisest teavitatakse. Kuigi nende tingimustega peab teenuse tarbimiseks nõustuma, ei selgita need edasise analüüsi kohta midagi konkreetset.[37]

2.2.3. Seadmed, mis (automatiseeritult) andmeid toodavad

Kolmas liik andmeid, mis suurandmed moodustab, on pärit erinevat liiki seadmetelt, mis automaatselt informatsiooni koguvad või toodavad, saadavad ja talletavad või talletama suunavad: need on võrku ühendatud seadmed. Sellised seadmed on kas (i) traditsioonilised salvestusseadmed, mis on saanud tänu digitaalsetele teenustele uue võimekuse, või (ii) seadmed, mille teke digitaalsete teenuste arenguga on võimalikuks saanud. Seadmed võivad oma olemuselt olla nii passiivsed, nt neid skaneeritakse, kui ka aktiivsed ehk nad saadavad ise informatsiooni välja.

Need andmed on üldiselt struktureeritud, sest seadmetel on kindel ülesanne, mida nad täidavad. Siiski, nende andmete ühtimine ja ühildamine teiste andmekogudega võib olla keeruline. Näiteks videoid on keerulisem analüüsiks kasutada kui numbrilisi või tekstist koosnevaid andmeid, sest video puhul on vaja sisu tehnoloogiale äratuntavaks muuta tehnoloogiat õpetades. Teksti või numbrite puhul on sobilike vastete leidmine seevastu lihtsam, sest näidis sellest, mida süsteem otsib, on üheselt mõistetav.

Automaatselt andmeid tootvad seadmed, mis on uue võimekuse saanud, tähendavad selliseid seadmeid, mis on juba traditsiooniliselt andmeid kogunud ja talletanud, kuid mis saavad nüüd interneti vahendusel ka teiste seadmete või andmebaasidega ühenduda. Näiteks turva- või kiiruskaamerad ei tooda ja ei salvesta täna ainult andmeid, vaid saavad neid ka teistesse masinatesse või seadmetesse vahendada.

Digitaliseerimisega on automaatselt andmeid tootvate seadmete andmemaht suurenenud, sest võimalus on koguda erinevaid andmeid. Juba need tehnoloogiad panustavad nüüd suurandmete tekkimisse. Kuid lisaks juba eksisteerivatele andmekogujatele on tänu uutele võrgulahendustele suurandmeid tootmas ka samalaadsed, kuid uued seadmed – nutistu seadmed.[38] Need on võrku ühendatud seadmed, mis on saanud võimalikuks tänu digitaliseerimisele ehk nende olemus sõltub digitaalsetest lahendustest. Need on üldjuhul seadmed, mis oma andmed teistele seadmetele ja andmebaasidesse saadavad. Kogutud andmeid on hiljem võimalik analüüsida.

Seadme haldaja (ja selle, kelle jaoks ta andmeid kogub) järgi on vajalik seadmed jagada kolme gruppi.

1. Riiklikud seadmed – seadmed, mis koguvad ja talletavad riiklikku informatsiooni. Näiteks ilmajaamad, kiiruskaamerad jne.
2. Erasektorite seadmed – seadmed, mis koguvad ja talletavad spetsiaalsele firmale vajalikku informatsiooni. Näiteks turvafirmade kaamerad, telekommunikatsioonifirmad jne.
3. Eraisikute kasutuses olevad seadmed – seadmed, mis on eraisikute kasutuses ning on kas isikustatud (nt aktiivsusmonitoride rakendused eeldavad kasutaja-

[37] Näiteks RIA lehekülg talletab mõned metaandmed 10 aastaks (Riigi Infosüsteemi Amet, 2019b).

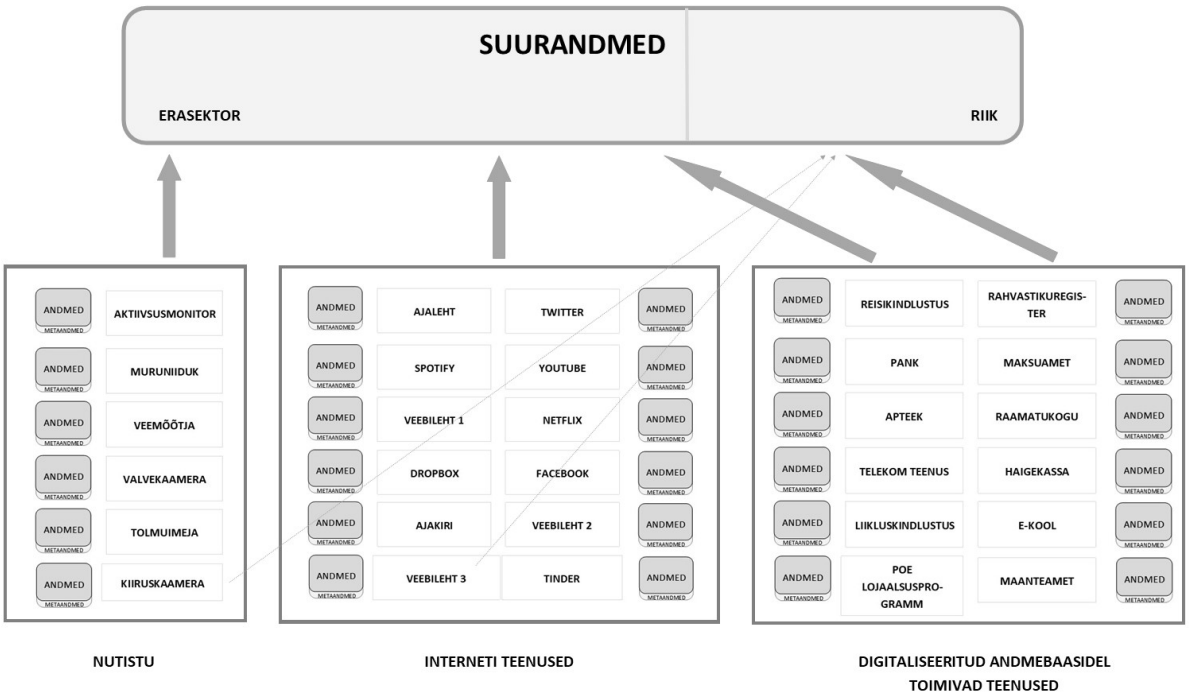
[38] Nutistu ehk asjade internet ehk IoT – aastal 2020 on prognooside kohaselt ligikaudu 31 miljonit seadet, mis on kõik asjade internetti ühendatud; aastaks 2025 on prognoositud nutistu seadmete arvu kahekordistumist (De Sualles, 2019).

nime/e-posti omamist ja süsteemi sisse logimist) või isikustamata (nt lihtsamad robotmuruniidukid ja robottolmuimejad saadavad/koguvad küll liikumisandmeid, kuid ei eelda kasutajanime olemasolu jne.)

Need otsesed andmed, mida seade kogub, on seadme haldaja ja/või andmete koguja käsutuses, nt kiiruskaamerate informatsiooniga tegeleb PPA, „nutika" mõõdikuga Elering. Eraisikute käsutuses olevad seadmed on aga samuti seotud tarkvaraga, mis kas riiklik või erateenus. Seega, ka eraisikute kasutatavate seadmete informatsioon võib saada analüüsi osaks, vastavalt teenuse haldaja soovile. Sarnaselt kuuluvad seadmete kasutamise süvaandmed teenusepakkujale ning võimalused nende analüüsiks on olemas. Kitsaskohad nutistu ja teiste interneti vahendusel tekkivate andmetega on sarnased – tänu nende mahule on lihtne näha peale andmete otsese kasutuse ka analüüsi potentsiaali, kuid ligipääs andmetele on piiratud. Nii peab iga teenusepakkuja piirduma andmetega, mis on tema käsutuses.

Oluline on ka välja tuua, et vaatamata nutistu üha laiemale levikule, nt targad majad või nutikad tänavad, ei tähenda see veel iseenesest suurt läbimurret tarbimise optimeerimises ning aja kokkuhoidu. Kalaranna tänav – Eesti ainus nutikas tänav – kogub näiteks statistikat jalakäijate arvu kohta tunnis (Eliko, 2020). Inimeste liikumise jälgimine ei tohiks muuta aga tänava haldamist või korralisi hooldustöid ehk tänav peab alati ja kõigile jalakäijatele võimalikult ohutu olema: nt kõnnitee lumest puhas ning libisemisoht ennetatud, tänavavalgustus toimimas, prügi ära viskamine võimalik ja nii edasi. Ühesõnaga, pelgalt seadmete võrgustamine ning omavaheline suhtlus, digitaalsete andmete kogumine ja omamine, ei muuda alati veel teenuse kvaliteeti/olemust, kuigi tehnoloogia kasutuselevõtu propageerimiseks just sellist eelist rõhutatakse (Kitching ja Dodge, 2019). Küll aga tähendavad kogutavad digitaalsed andmed erafirmadele tehnoloogia vahendusel ligipääsu järjekordsele hulgale andmetele.

Nutistu ehk koos toimivate tehnoloogiliste lahenduste voorus on sarnane X-tee abil koos toimivate andmebaaside omale: protsesse on võimalus optimeerida, sest tehnoloogia



Joonis 1: Suurandmed (autori koostatud)

suudab ilma inimese vahenduseta vajaliku toimingu ära teha, nt muru on niidetud, tuba soe, elektrinäidud saadetud. Samas kaasneb sellise koostoimimisega aga ka oluline risk – kõikidel tehnoloogilistel lahendustel on omad nõrkused; võib-olla turvavead või -augud (mis alles koos toimimisel ilmnevad). Seega, koos toimima ühendatud tehnoloogia on sellevõrra haavatavam, et süsteemi mõnd teist osa rünnates jõutakse ka selle spetsiifilise seadmeni.

Kokkuvõttes, kuigi suurandmetest räägitakse kui homogeensest andmekogust, on tegemist väga eriilmeliste ja erinevatest allikatest tekkivate andmetega, mis kohest koostoimet ei võimalda.

Suurandmeid koos hinnates on tekkivate andmete maht suur ning nende kasutusviisid- ja võimalused väga erinevad. Samas ei tähenda andmete teke aga ka kohest kasumlikkust või optimeerimise võimalust – teenuse pakkujal on õigus ja võimalus otsustada, milliseid andmeid kasutatakse. Uute andmete analüüsist on olnud võimalik kasu saada peamiselt interneti vahendusel tekkivatest andmetest, sest need mitte ei täienda juba olevat, vaid annavad ka täiesti uut informatsiooni. Pelgalt andmete maht aga ei muuda neid veel vajalikuks – vaja on välja töötada nii analüüsi kui ka kogumise, kasutamise ja talletamise viisid.

3. SUURANDMED EESTI KONTEKSTIS

Lähtudes suurandmete liigitamisest kolmeks, saab nüüd vaadelda, missugune on Eesti suurandmete olemus. Alljärgnev selgitab, kuidas uudsete digitaalsete lahenduste kasutusele võtjana on Eesti suurandmete eripäraks isikustatud andmete suur osakaal nii era- kui ka avalikus sektoris. Need andmed tekivad erinevate andmebaaside ja teenuste koostoimest ning on hiljem kasutatavad. kasutatuava. Lisaks on IKT-teenuste ja nutistu vahendite kasutamine sarnaselt globaalsele trendile kasvamas nii Eesti era- kui ka riigiteenustes. Eesti koostoimivatest lahendustest tulenevalt on sarnaselt andmebaasidega, aga ka nutistust ja interneti vahendusel toimivate teenuste kasutamisest, juba tekkimas andmeid, mis on isikustatud ning põhimõtteliselt koostoimivad.

3.1. DIGITALISEERITUD ANDMEBAASID JA ANDMEKOGUD

> *„Eestis on üks maailma eesrindlikumaid rahvastikuregistreid, mis võimaldab riigi e-teenuste korraldamist ja tagab isikuandmete kasutamise kogu riigis ühtsetel alustel.“ (Siseministeerium, 2016, lk 105)*

Suur roll e-Eesti digitaalses eduloos ei ole mitte ainult ID-kaardil, mis tehnilisi lahendusi võimaldab ja erinevate teenuste koostoimes suurt tähtsust omab, vaid ühtlustatud viisil isikute tuvastamisel – Eesti isikukoodil. Isikukoodi saab pidada „digitaalseks nimeks“ (Kotka, 2014), mille alusel erinevaid teenuseid kasutades, lepinguid sõlmides, kohustusi võttes, vastutades jne isiku tuvastamine toimub. Isikukood on ainukordne ning olemasoleva koodi korral moodustatakse rahvastikuregistri seaduse alusel uus isikukood vaid lapsendaja soovi või isiku sünniaja/soo muutmise või parandamise korral.[39]

Selline üks kindel number ja viis, millega Eestis kodanikku või residenti tuvastada ning mis kõikides riiklikes andmebaasides on sama – mille alusel andmed on erinevates andmebaasides tuvastatavad – on Euroopas üsna ainulaadne lahendus. See on laialdaselt kasutusel olevast nimest, sünniajast ja -kohast (vahel ka vanemate nimede loetlemisest) sõltuvast isikutuvastusest täpsem (näiteks nimepildi erinevusi võõrtähtede transkribeerimisel või inimlikud vead nimede sisestamisel saab nii üldjuhul ennetada), kiirem, digitaalselt lihtsam ning võimaldab isiku tuvastamist erinevaid teenuseid tarbides. Nii on isikukood, mis juba enne digitaalseid lahendusi võimaldas erinevates andmebaasides olevaid andmeid ühe kriteeriumi alusel ühendada, kesksel kohal, et arendatud digitaalsed lahendused koos toimida saaks.

[39] Isikukoodide moodustamise ja andmise kord (2017) § 11. Uue isikukoodi moodustamine. Punkt 1.

Eestis, nimelt, on nii paljud avaliku sektori kui ka erasektori teenused seotud isikukoodi ja/või isikutunnistuse abil isiku tuvastamisega. Et isikukood on iseenesest digitaalne jälg, siis jätab ka kõikide selliste teenuste kasutamine isikustatud digitaalse jälje: lisab isikustatud andmebaasi ja suurandmete mahtu. St erinevate teenuste andmebaasid ei ole eraldiseisvad, vaid ka koos toimivad. Kusjuures, iga isiku andmed on tänu isikustamisele kergesti eristatavad.

Näiteks retseptiravimite ostmisel kontrollitakse vajaliku arsti väljastatud retsepti olemasolu ID-kaardi või isikukoodi alusel, kus süsteem kontrollib nii arsti sisestatut kui ka ravimi soodustuse olemasolu. Ravimi väljakirjutamisest ja ka väljaostmisest jääb digitaalne (isikustatud) jälg. Selline identifitseerimine käib ühe süsteemi sees. Ka mitte-elutähtsate riiklike teenuste puhul, nagu raamatukogu lugejapilet, toimub inimese tuvastamine ID-kaardi alusel. See on samuti süsteemisisene tuvastamine. Nii tekib isikustatud andmebaase Eestis pea iga riiklikku teenust kasutades.

ID-kaardi abil isikustamine on sarnase isikutuvastuse võimaldanud ka erasektoris, kus samuti toimuvad identifitseerimised ühe süsteemi sees. Mitte kõikide erasektori pakutavate teenuste puhul ei ole tegemist aga elutähtsate teenustega – nii kohviku lojaalsusprogrammiga liitumiseks kui ka toidukaupluse hinnasoodustuse saamiseks saab ID-kaardi abil end tuvastada.[40] Nii tekib ka erasektoris hulk isikustatud andmeid.

Et aga ka erinevaid teenuseid oleks võimalik isikukoodi või ID-kaardi abil digitaalselt teostada (oma identiteeti kinnitada) ja digitaliseeritud andmebaasidest saadavad kasu täielikult realiseerida, on EV-s kasutusel ka andmevahetuskiht X-tee, mille kaudu erinevad sellega liitunud teenusepakkujad ja (digitaliseeritud) andmebaasid omavahel informatsiooni jagada ja koos toimida saavad. Selle rakenduse abil saab üks ametiasutus või teenusepakkuja lihtsalt ligi teise andmebaasi isikustatud andmetele, et isikut tuvastada, informatsiooni kontrollida, informatsiooni oma süsteemi lisada jne.

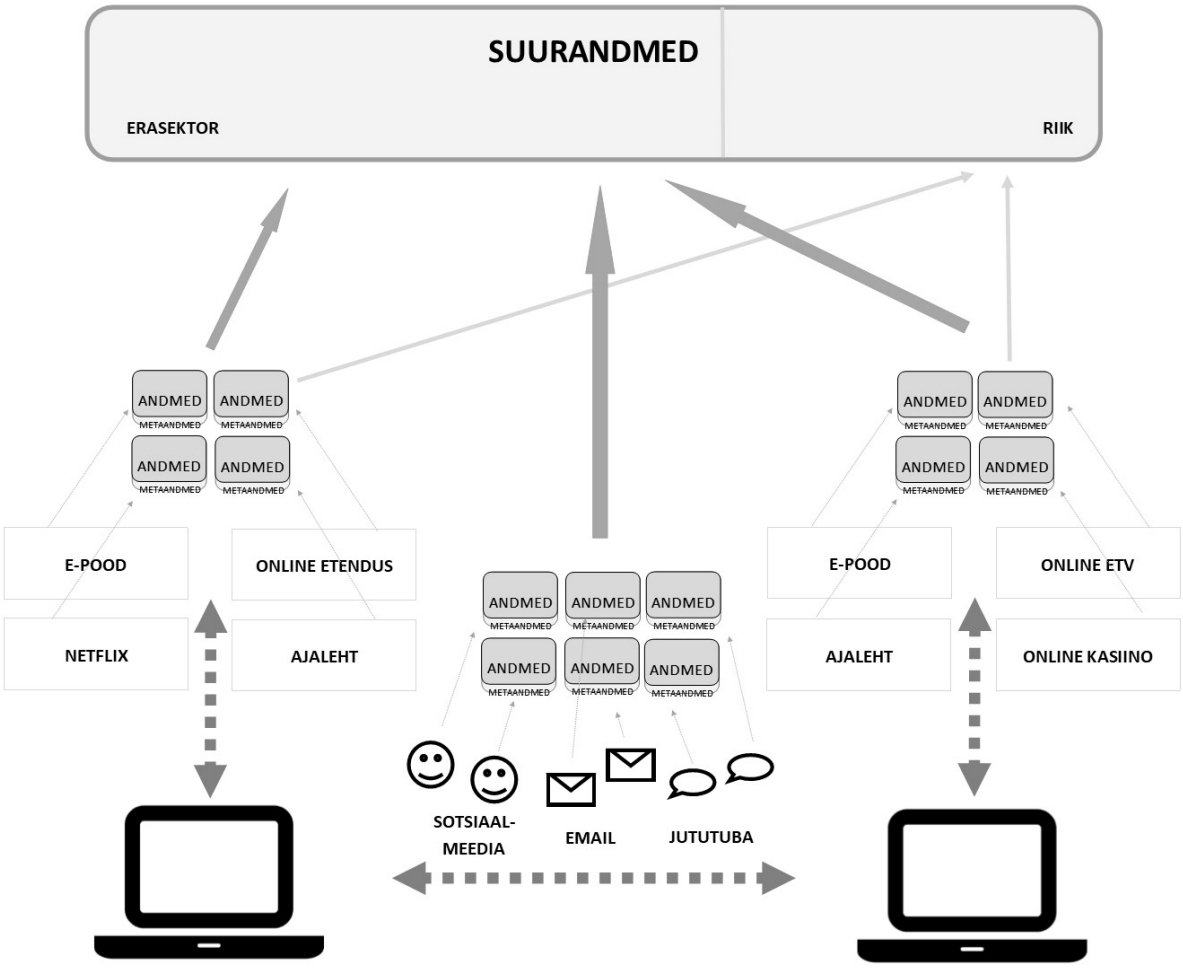
Digitaalsete teenuste kättesaadavust on Eestis hõlbustanud ka erinevate rakenduste kasutuselevõtt, mis võiks väiksemal (Smart-ID) või suuremal määral (mobiil-ID) vajadusel asendada ID-kaardiga toimuvat isiku tuvastamist. Nende kahe lisandunud autentimisviisi kasutusmugavus tähendab, et digitaalsetel teenustel on veelgi rohkem kasutajaid ja omavahel seotud isikustatud digitaalsete andmete hulk on veelgi kasvamas.

Paljud isikustatud digitaalseid andmeid vajavad ning digitaalset isiku tuvastamist nõudvad teenused on elutähtsad. Andmebaaside omavaheline koostalitusvõime on hädavajalik. „Infoühiskonna Arengukava 2020“ arvab, et „avaliku sektori sees on koosvõime suures osas saavutatud ning peamiseks kitsaskohaks on õiguslik, organisatsiooniline ja semantiline koosvõime“ (Majandus- ja Kommunikatsiooniministeerium, 2018, lk 9). Isegi kui selline suutlikkus on saavutatud, siis X-tee toimib interneti vahendusel: teenused, mis selle abil isiku tuvastamisest sõltuvad, sõltuvad X-tee toimimisest. Nii on Eesti koos toimivate digitaalsete lahenduste jätkusuutlikkus siiski haavatav ning isikustatud andmetele ligipääs võib olulisel hetkel olla pärsitud.

[40] Isikustatud andmed võivad olla ka kõigest kõrvalprodukt. Kuigi teenuste kasutamine ja näiteks lojaalsuskaardid eeldavad, vahel isegi nõuavad, isikustamist, ei ole nende teenuste eesmärk siiski mitte metaandmete kogumine, vaid hoopis tihedas konkurentsis ellujäämine ehk oma turupositsiooni/klientuuri kinnistamine (McCullagh, 2004, p. 27).

3.2. INTERNETI (KASUTATAVATE RAKENDUSTE) KASUTAMISEL TEKKIVAD ANDMED

Teine suur hulk andmeid, mis samuti riigi ja selle elanike ning kodanike suurandmetesse panustab, on andmed, mis tekivad internetti ja selle vahendusel toimivaid teenuseid kasutades. Andmeid toodab nii internetis ajalehe lugemine kui ka teatrietenduse vaatamine; sõpradega kirjade ja emotikonide vahetamine kui ka internetipoes ostlemine.



Joonis 2: Suurandmete teke internetis toimivaid teenuseid kasutades (autori koostatud)

Interneti andmed on EU GDPR ehk *Euroopa Liidu isikuandmete kaitse üldmääruse* alusel,[41] millega on ka Eesti seadusandlus kohandatud, samuti isikuandmed. Nende kogumiseks ja töötlemiseks on vaja inimese nõusolekut; nende talletamisel on kindlad reeglid. Andmete omanikul (ingl *data subject*) on õigus saada informatsiooni, milliseid andmeid on tema kohta kogutud. Kogutud andmeid ei tohi kasutada pelgalt automatiseeritud otsuste tegemisel, sh profiilianalüüsil, mis võivad kaasa tuua õiguslikke tagajärgi või muud märkimisväärset mõju (Artikkel 22). Kuid, kui inimene on andnud selleks selgesõnalise nõusoleku ehk talle kuvatavates T&C-s „OK" klikkinud, on teatud töötlemine lubatud. Pealegi, nagu paragrahv 4 märgib, isikuandmete kaitse ei ole absoluutne, vaid seda tuleb kaaluda „vastavalt selle ülesandele ühiskonnas tasakaalustada põhiõigustega vastavalt proportsionaalsuse põhimõttele". Nii on interneti andmed lihtsamini analüüsitavad

[41] EU GDPR nime all tuntakse Euroopa Parlamendi ja Nõukogu määrust (EL) 2016/679, 27. aprill 2016, füüsiliste isikute kaitse kohta isikuandmete töötlemisel ja selliste andmete vaba liikumise ning direktiivi 95/46/EÜ kehtetuks tunnistamise kohta (isikuandmete kaitse üldmäärus) (EMPs kohaldatav tekst).

kui andmebaaside andmed, mille puhul on töötlemise nõusolekut vaja eraldi taotleda või töötlemisest inimesele teada anda.

Riigil on internetis toimuvat tegevust märkivate andmete kättesaamine keeruline ja seda isegi siis, kui teenuse kasutamiseks ja kasutaja autentimiseks kasutatakse isikustamist, millega andmebaaside puhul on tegu, sest suurt hulka andmetest koguvad, haldavad ja ka analüüsivad erafirmad. Teenusepakkujad on tihti rahvusvahelised, mistõttu on ühe konkreetse riigi seadusandluse rakendamine väga keeruline. Nii oma ärimudelist lähtuvalt kui ka kasutajatele konfidentsiaalse teenuse pakkumiseks on nad loonud viise, kuidas andmetele ligipääs on võimatu.

Krüpteerimine, mida kasutatakse ka e-riigi digiteenuste puhul, tähendab, et kolmandatel osapooltel pole andmetele/tekstile/sisule mingit ligipääsu – ligipääsu tagavad ainult dekrüpteerimiseks vajalikud koodid (Yar, 2009, p. 149). (Nt digiallkirjastatud dokumendile on ligipääs vaid inimestel, kellele isikukoodi alusel on dekrüpteerimine võimalikuks tehtud, sh digiallkirjastatud lepingute ja avalduste sisuga tutvumine.)

Ka mitmed sotsiaalmeedia lahendused edastavad kasutajate sõnumid krüpteeritult. Näiteks populaarse rakenduste WhatsApp puhul on ainult sõnumi adressaadil, st kindla telefoninumbriga seotud identiteedil võimalus sõnumi sisuga tutvuda ehk kasutusel on lõppkasutajast lõppkasutajani krüpteerimine (ingl *end-to-end encryption*). Selline krüpteerimistehnoloogia võimaldab kasutajatele suuremat turvalisust sõnumi sisu privaatseks jäämise kohta ja väiksemat riski küberkuritegevuse/häkkimise näol. Teenusepakkuja edastatud sõnumite sisusse ei sekku/seda ei hinda ning teenusepakkuja hallatavale informatsioonile, sh sõnumite metaandmetele, ligipääsuks on riiklikel institutsioonidel seadusandlusega lubatud/limiteeritud võimalused (WhatsApp, 2020). Kuigi teenus on arendatud kasutaja ja andmete turvalisuse tagamiseks (ning suur osa vahetatavast informatsioonist on igati legaalne, aga isiklik), on teenuste turule toomisega antud ka samal ajal organiseeritud kuritegevusele võimalus legaalseid meetodeid kasutades oma tegevust varjata.[42] Nimelt ei ole korrakaitseorganitel ilma loata võimalust suhtlust seirata – seiramise põhjuse annab aga ainult sisust teadlik olemine ja sellele korrakaitseorganitel esmane ligipääs puudub. Lisaks on keeruline ka teenusepakkuja ja riikide institutsioonide vaheline suhe, mida raskendab olukord, mil teenusepakkuja kohustused on ühes riigis, aga andmeid küsivad institutsioonid teises riigis.

Veelgi keerulisem on aga anonüümsuse tagava veebibrauseri (Tor) kasutamine ning sealse tegevuse jälgimine. Tor veebibrauseri kasutamine, sarnaselt krüpteeritud sõnumiteenustega, ei ole iseenesest illegaalne ning arvestades teiste veebibrauserite andmekogumise tehnikaid, on see ka mõistetav. Kuid peale neutraalse/rahuliku veebikogemuse võimaldab anonüümsus märkamatut ligipääsu ka interneti kõige varjatumatesse osadesse – tumeveebi – kus suur osa illegaalseid tehinguid aset leiab. Loomulikult ei kasuta mitte kõik anonüümsust suurendavaid teenuseid illegaalseks tegevuseks, kuid anonüümsuse pakkumine on ka selle võimaldanud. (Riigitu internetiruum teeb aga riiklike seaduste internetile rakendamise keeruliseks.)

Kui suur on eestlaste interneti teenuste tarbimisest tulenev digitaalne jalajälg ja mis on selle sisu, on ainult aimatav. Eestis on 2019. aasta oktoobris avaldatud Statistikaameti 16-74-aastaste seas läbi viidud rahvastiku uuringu kohaselt internetiühendusega 90% leibkondadest (Statistikaamet, 2020).[43] Internetiühenduseta leibkondadest ei pea valdav hulk seda lihtsalt vajalikuks ja huvitavaks. Leibkondade osatähtsus, kus teenus ei ole kät-

[42] Vaata näiteks: National Crime Agency, 2019; Bundeskriminalamt, 2020; Riigi Infosüsteemide Amet, 2019 a.

[43] Vaata ka: andmebaas.stat.ee/Index.aspx?lang=et&DataSetCode=IT20

tesaadav teenuse ja seadmete maksumuse tõttu, on seevastu vähenemas (Statistikaamet, 2019).

Suurandmete tekkimise seisukohalt on oluline näitaja, et 16-44-aastastest kasutas internetti iga päev või peaaegu iga päev 98% elanikkonnast, 65-74-aastaste seas aga 75%. Lisaks on Statistikaameti andmetel üha suurenenud mobiilse interneti kasutamine suhtlemiseks ja teenuste tarbimiseks:

> *16–24-aastaste seas kasutas mobiilset internetti 98,6%[44] ja 65–74-aastaste hulgas 42,5% tarbijatest. Üheksa inimest kümnest kasutas internetti e-kirjade saatmiseks ja internetipanga teenuste kasutamiseks, 72% kasutajatest kuulas muusikat ning kasutas sotsiaalvõrgustikke. Kõige enam on suurenenud interneti kaudu telefoniga rääkimine (Statistikaamet, 2019).*

Freedom House-i 2019. aasta raporti kohaselt on Eesti internetivabaduselt maailmas esirinnas – riik on kehtestanud väga vähe piiranguid internetis kättesaadava sisu kohta: seadusandlusega on reglementeeritud hasartmänguteenuse pakkujate tingimused. Sotsiaalmeedia kasutamine on populaarne ja kättesaadav ning üldiselt on nii internetis sõnavabaduse kasutamine kui ka uudiste neutraalsus tagatud (Freedom House, 2020).

Lühidalt, interneti vahendusel tekkivate andmete puhul on riigil ainult kaudselt võimalik hinnata andmete kogumahtu ja sisu; erafirmade pakutavate interneti teenuste puhul on võimalus, et andmed on avalikud, nt avalikud Facebooki postitused, kuid samas on ka andmeid, mille millele on ligipääs (peaaegu) võimatu. Nii on hetkel internet küll näiliselt ammendamatu andmete ressurss, mis suurandmete analüüsi potentsiaali tiivustab, kuid nende väärtus on väga muutlik. Sellistele andmetele toetudes ei ole riiklikult võimalus toimivaid rakendusi arendada, ilma et suur ressurss läheks jätkusuutlike kasutus- ja analüüsimeetodite väljatöötamisele.

3.3. SEADMED, MIS AUTOMAATSELT ANDMEID TOODAVAD JA NUTISTU SEADMED

Suurandmetesse panustavad ka seadmed, mis oma andmed digitaliseerivad ja need digitaalsel kujul edasi saadavad – need on seadmed, mis on eraldiseisvad, kuid suhtlevad omavahel interneti vahendusel. Sellised seadmed võivad olla passiivset laadi ehk kõigest skanneritavad, aga võivad ka aktiivselt ise informatsiooni välja saata. Ka selle andmete loomise viisi puhul on tuvastatav andmete kasv, sest kasvanud on teenuste pakkumine ja nõudlus ning ka selliste seadmete arv, mida saab võrku ühendada/mis digitaalset jalajälge toodavad, nt targad termomeetrid, targad mõõdikud jne. Nt Telia avas võrgu, Levira haldab mitut erinevat viisi[45], kuidas seadmeid ühendada jne. Samas ei tohi suurandmete hulka kasvatavate seadmete hulka piirata kõigest uute „nutikate" seadmetega, vaid andmestusse/andmebaasidesse toodavad sisu ka seadmed, mille puhul on võrku ühendamine neile uue võimekuse andnud. Näiteks „targad majad" ja „tark tänav" võimaldavad reaalajas toimuvast andmelist ülevaadet saada.

Isikuga seotuse alusel on seadmete tehtavaid andmeid võimalik jagada järgmiselt:

1. Isikustatud andmed – andmed, mis on pärit nutikatelt seadmetelt, mida kasutades on oluline isiku tuvastamine. Nt targad termomeetrid ja aktiivsusmonitorid

---

[44] Freedom House-i 2019. aasta raport toob välja 145% kasvu telefonide kasutamisel võrreldes aastaga 2017. (Freedom House, 2020).

[45] Vaata: Levira, 2020.

on isikustatud (kui mitte nimeliselt, siis kasutajakonto, telefoninumbri jne abil). (Tihti on need isikustamised seotud teiste internetiteenustega ja jällegi võib tuvastamine toimuda nt isikukoodi kasutades.)

2. Vajadusel isikustatavad andmed – andmed, mille puhul on vajadusel võimalik informatsioon ühendada teiste andmetega, nt liikluskaamera andmed (auto registreerimise number jne). Selliseid koostoimivaid nutistu teenuseid Eestis veel palju kasutusel ei ole.
3. Automaatsed andmed – andmed, mille analüüs on ühe kindla asjaga seotud, nt ilmastikuandmed või turvakaamerad. Näiteks üha enam paigaldatakse Eestis avaliku ruumi jälgimiseks ja valvamiseks ka kaameraid, mis salvestavad andmeid automaatselt. Samas ei ole aga Eesti avalikus ruumis olevate valvekaamerate puhul, erinevalt Venemaast või Singapurist, tegemist sellistega, mis automaatselt näotuvastustehnoloogiate abil isikuid tuvastavad. Selliseid andmeid on võimalik isikustada vaid lisarakendusi välja töötades ja appi võttes.

Nutistu andmete mahtu on sarnaselt interneti vahendusel kasutatavate rakenduste kaudu tekkivate andmetega keeruline hinnata. Nutistu seadmete müük Eestis on hoogustumas. Erinevad seadmed moodustavad väga erinevaid andmeid; paljusid neist ei talletata pikaajaliselt või need pole oluliselt kasutatavad hilisemal analüüsil. Vaid väike osa tekkivatest andmetest on aga riigile kättesaadavad.

3.4. DIGITAALSED SUURANDMED EESTIS

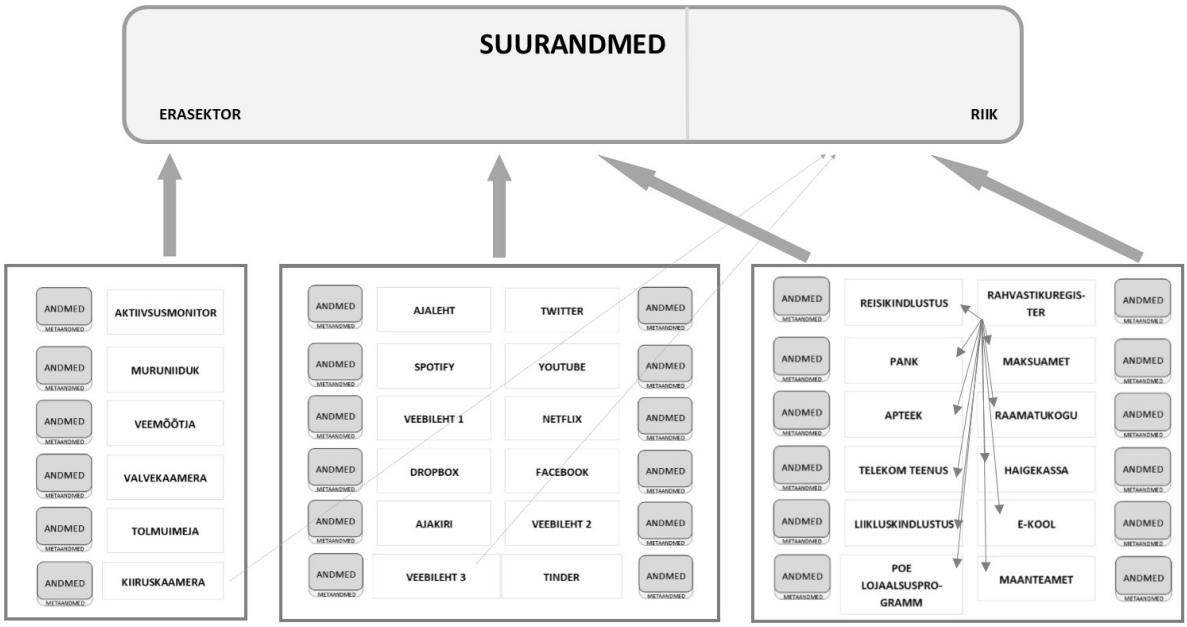
Digitaalsete andmete maht erinevate andmebaaside näol on Eestis suur. Ainuüksi 30. jaanuari 2020 seisuga on kasutusel 1 345 724 aktiivset ID-kaarti. Ühtekokku on digitaalset isikutuvastust tehtud ID-kaartidega 944 131 563 ning digiallkirjastatud dokumente 879 782 012 korral.[46]

Tänu Eesti isikukoodi abil elanike ja kodanike isikute tuvastamise süsteemile ja arenenud digitaalsetele lahendustele, kus isikukoodiga isiku tuvastamine on kesksel kohal, on Eestil hallata väga suur hulk isikustatud digitaalseid andmeid, mis on a) inimesele elutähtsad; b) riigi toimimisele elutähtsad (nt elanikeregister; sõidukiregister). Neile andmetele lisanduvad aga ka nende elutähtsate andmete ja andmebaaside metaandmed ehk andmed näiteks viisist, ajast ja kohast, kus isikustatud andmed on tekkinud; kus ja missuguse meetodiga on süsteemi sisse logitud.

Digitaalsetel lahendustel toimiva riigina on Eesti riigiteenuste toimimiseks, hädavajalike andmete olemasolu ja kättesaadavuse tagamiseks hajutanud riski ning loonud „andmesaatkonna“ – riigi toimimiseks vajalikud andmed on talletatud teise riigi, Luxemburgi territooriumil asuvatesse serveritesse. Sarnaselt riigi teistele saatkondadele laienevad ka andmesaatkonnale kehtestatud diplomaatilised kokkulepped (Siersbutowski, 2019). Niisiis, turvalisus või turvatunne isikustatud andmete olemasolu ja jätkusuutlikkuse kohta on riigil olemas. Samas on aga nii talletatud andmed ainult väike osa suurandmete koguhulgast.

Tänu Eesti digilahendustele on ka eraettevõtetel hallata suurel hulgal digitaalseid andmeid, mis on isikustatud. See tähendab, et vajadusel – kooskõlas andme haldaja ja seadusandlusega – saaks need erasektori isikustatud andmed ühendada riiklike isikustatud andmetega. Nii on inimesed andmestikus kergesti tuvastatavad, st võimalus on nende informatsiooni lihtsalt kontrollida ning ristkasutada.

[46] Reaalajas info: www.id.ee



Joonis 3. Suurandmed Eesti kontekstis (autori koostatud)

Internet ja nutistu on seevastu oma hüppeliselt kasvanud andmemahuga proovikiviks andmete ligipääsetavuse, mahu ja vajalikkuse osas. Kui riik tahab esmavajalikke isikustatud andmeid hallata ja ristkasutada, siis mitte kõik interneti ja nutistu andmed ei ühildu olemasolevasse süsteemi ilma lisarakendusi välja töötamata. Need on tihti struktureerimata ning andmete sisu nõuab edasiseks kasutuseks (keerukamat) sisuanalüüsi. Lisaks on interneti ja nutistu andmetele ligipääs üpris keerukas, sest erafirma halduses olevatele isikustatud andmetele ligipääsemiseks peab riigil olema mõjuv põhjus.[47] Pealegi, alati pole tegemist teenuse pakkujate puhul mitte kohalike ettevõtetega, vaid globaalsete teenusepakkujatega, keda on riiklikke seadusi järgima sundida keeruline.

Lühidalt, arvestades kogu andmete hulka ehk inimese digitaalset jalajälge, on teoreetiliselt kõiki neid erinevates sektorites/teenustes olevaid andmeid koos analüüsides võimalik inimese tegevusest, kuigi selline tegevus on Euroopa Liidu isikuandmete kaitse üldmääruses keelatud, väga ülevaatlik profiil saada. Vaadates suurandmeid moodustavate andmete olemuse erinevust, on aga tähtis mõista, et andmed on killustunud ning kõikide erinevate liikide koostoime nõuab uusi tehnoloogilisi lahendusi, mis suudaks olulise ebaolulisest eristada ja vajalikke andmeid isikustada.

47 Nt riigi põhiseadusliku aluse ohustamine või terrorism.

4. ANDMETE KASUTAMISE TULEVIK EESTIS

4.1. KRATISTRATEEGIA

Tänu juba praegu Eestis toimivatele digitaliseeritud lahendustele, tehnoloogiateadlikkusele ning üldisele tehnoloogia entusiasmile või (vähemalt) vähesele tehnoloogia kartusele (Drechsler, 2018) on Eesti tulevikuvisioon (kitsaid) tehisintellekti süsteeme elik kratte avalike sektori teenustes üha rohkem rakendada. „Eesti positsiooni krattide rakendamisel teevad võimalikuks […] avaliku sektori kõrge digitaliseeritus, Eesti elanike avatus uutele tehnoloogiatele ja inimeste valmidus neid kasutusele võtta“ (Majandus- ja Kommunikatsiooniministeerium ja Riigikantselei, 2019, lk 34). Eesti täpsemat plaani tehisintellekti krattide rakendamiseks on selgitatud 2019. aasta mais esitatud ekspertrühma aruandes „Eesti tehisintellekti kasutuselevõtu eksperdirühma aruanne“.

Aruanne annab mõista, et riigile on pandud suur ootus ja vastutus rahastada teadust ja välja arendada ka baasteenuste ja infrastruktuur, et tehisintellekti potentsiaali saaks Eestis optimaalselt kasutada. Ootused arendatavate tehisintellekti süsteemide rakendamisele on tänu seni õnnestunud digitaalsete lahenduste arendamisele suured.

Ühest küljest on ootused praktilised ning sarnased maailmas sõnastatud ootustega. Tehisintellekti süsteemide kasutuselevõtu planeerimine erinevates sektorites on hädavajalik, et hakkama saada tööealise elanikkonna vähenemisega ning optimeerida nii tootmist kui ka teenuseid. See võimaldaks kasumlikkuse tõusu nii eraettevõtluses kui ka avalikus sektoris.

Teisalt aga on ootused seotud Eesti kinnitamisega (kinnistamisega) tehnoloogiliselt innovatiivse ja arenenud riigina. Nagu aruanne märgib: „Eestil on hea võimalus saada muule maailmale eeskujuks ja katselavaks – kohaks, kus kratte pannakse inimesekeskselt nii valitsusaparaadis kui ka ettevõtluses inimese heaks tööle“ (Majandus- ja Kommunikatsiooniministeerium ja Riigikantselei, 2019, lk 5). Kuigi selline ambitsioon ei ole iseenesest halb, võib liigne keskendumine sellele aga tuua kaasa arendustegevuse, kus rõhk ei ole mitte vajalike vaid muljetavaldavate lahenduste väljatöötamisel. See võib viia situatsioonini, kus tehnoloogia rakendamine ei ole läbimõeldud – sarnalaselt digitaliseerimisele keskendutakse pigem võimalusele kui vajalikkusele tehisintellekti süsteemi rakendada. Nii peavad nii teenuse tellijad kui ka arendajad pöörama suurt tähelepanu, et tehisintellekti rakendused oleksid vajaduspõhised.

Kahe keskse ootuse valguses toob aruanne välja esmased vajakajäämised, mis on krattide arendamisega ilmnenud.

Esiteks, hoolimata visioonist, on vajaliku teadustegevuse rahastamine Eestis puudulik. Seega on ka krattide kasutuselevõtuks vajalikke teadmisi ning spetsialiste vajaka. Pealegi,

uutele süsteemidele lisaks on vaja spetsialiste ka juba toimivate ja suurenevate süsteemide haldamiseks ning arendamiseks. Nii näib, et visiooni on võimalus teostada ainult juhul, kui hinnatakse ümber, millesse on vaja panustada, et toimuks oodatud tehnoloogiline areng.

Teiseks, tehisintellekti süsteemi toimimiseks ei piisa kõigest suurandmete olemasolust ning nende kättesaadavusest. Nii olemasolevate andmete maht kui ka koostoime X-tee abil on hea eeldus, kuid ka sel juhul on kindlasti vaja olemasolevate andmete sünteesimine, et moodustada vajalikke andmestikke/andmebaase. Pealegi ei ole ühest kinnitust, et hetkel toimiv X-tee on krattide jaoks sobilik. Seega on vaja välja töötada viis, kuidas uued tehnoloogilised arendused juba olemasolevat toetaks, mitte ei muudaks seda kasutuskõlbmatuks.

Kolmandaks, visioon näeb ette, et avalik sektor oleks tellija ja kasutajana krattide kasutuselevõtu teerajajaks. Kuigi esmased krati rakenduste kasutajad peaksid olema avaliku sektori asutused, eeldatakse, et kratilahendusi arendaksid erafirmad ehk avaliku sektori asutus on kõigest tellija ja rakendaja rollis. Kuna tehnoloogia on alles arendamisjärgus, siis avaliku sektori esindajatel puudub adekvaatne ülevaade sellest, kuidas saaks süsteeme rakendada; missuguseid teenuseid oleks soovitav ja mõttekas tehisintellekti süsteemide abil toimima panna (Majandus- ja Kommunikatsiooniministeerium ja Riigikantselei, 2019, lk 37). Samal ajal on keeruline ka erafirmadel initsiatiivi võtta ning võimalikke tehnoloogilisi lahendusi välja pakkuda, sest neil puudub ülevaade, missugused andmed on olemas ning kasutamiseks sobilikud (Majandus- ja Kommunikatsiooniministeerium ja Riigikantselei, 2019, lk 42). Pealgi on koostoimivate krati teenuste puhul hädavajalik inimese nõusolek, et kolmandad osapooled saaksid tema informatsiooni kasutada. Ka see süsteem tuleb läbi mõelda. Hetkel võib krattide planeerimine takerduda aga läbipaistmatuse taha.

4.2. UUS TEHNOLOOGIA JA TEHISINTELLEKTI SÜSTEEMID KORRAKAITSES

Üks riigi haldusala, mida uute tehnoloogiliste lahenduste kasutuselevõtt puudutab ning kus ka ise kasutuselevõttu mainitud on ning mida siinkohal näitena tuua, on korrakaitse.[48]

Üldine tehnoloogiliste võimaluste areng tähendab õigus- ja korrakaitseorganitele nii ülesannete kui ka võimaluste laienemist. Kui uue tehnoloogia võimaldatud ja vahendatavatele kuritegudele reageerimist mainitakse süsteemselt (nt küberkuritegevus leiab kajastamist nii KAPO aastaraamatutes kui ka riigi strateegiadokumentides), siis ühiskonnas plaanitavate uute tehnoloogiliste lahendustega kaasnevatest võimalikest muutustest sisejulgeoleku kontekstis räägitakse pigem kaude. Näiteks tehisintellekti süsteemide rakendamisega kaasnevaid võimalikke uusi väljakutseid, nii korrakaitse teenistuse osana kui ka korrakaitse süsteemile, kaardistatud ei ole.

Loomulikult, interneti laialdasest levikust on juba kasvanud välja omaette kuriteoliik – küberkuritegevus –, mille tõkestamine ja tabamine eeldab eraldiseisva võimekuse ning üksuste arendamist. Just küberkuritegevusega võitlemine ning virtuaalses ruumis toimuva hindamine – nii küberkuritegevus, mis kasvab, kui ka virtuaalsete keskkondade jälgimine – ongi seatud prioriteetideks. Interneti abil (veebis) toimuva jälgimist õigusrikkumiste ärahoidmiseks mainitakse „Kriminaalpoliitika põhialused aastani 2030“ eelnõu

[48] Euroopa-üleselt politsei ja piirivalve andmete edastamiseks, vahetamiseks ja kontrollimiseks on võrgustikke arendatud ja täiustatud aastaid, kuid *Entry-Exit Systemi*, *European Travel Authorisation Systemi* või *European Criminal Records of 3rd Contry Nationalsi* puhul on tegu kõigest koos toimivate andmebaasidega, mitte olemasolevate andmete analüüsimisega tehisintellekti süsteemide abil.

punkti 2 lõikes 6, et keskenduda tuleb organiseeritud kuritegevust ning põhiseaduslikku korda ohustavate tegevuste, k.a terrorismi ja vägivaldse äärmusluse, ennetamisele. Sellist riiklikku süsteemset tegevust pärsivad aga juba välja toodud probleemid – veebiruum on killustunud ning ligipääs pole kõigile andmetele vaba.

Samas on Eestil suurandmete olemust hinnates potentsiaal andmeid kasutada olemas ka laiemalt, st toimingute osana ja/või lihtsustajana, sest nii kvalitatiivselt kui ka kvantitatiivselt on andmete olemus muutunud. Kvantitatiivselt on isikuandmete kogus märkimisväärselt suurenenud; kvalitatiivselt on aga viisid andmete kogumiseks, informatsiooni töötlemiseks ja analüüsi võimalusteks, k.a kuritegude väljaselgitamiseks, täiustunud (Lõhmus, 2016, lk 698). Pealegi on avalike meetmete abil korjatud andmed jälitustoimingutes alati eelistatud varjatud toimingutele (Laaring, 2015, lk 174) ning mõne inimese puhul võimaldab internet ligipääsu suurele hulgale isikustatud avalikule informatsioonile. Selline lähenemine sobib eesmärgiga, et julgeolekuasutused piiravad oma ülesande täitmisel isikute põhiõigusi võimalikult vähe (Heldna, 2016, lk 720).

Andmeanalüüsi võimalike rakendamisviiside kohta, mis maailmas üha enam kasutamist leiavad, on seisukoht siiski äraootav – vajadus ja rakendamisvõimalus ei ühti. „Kriminaalpoliitika põhialused aastani 2030“ kaasnevas seletuskirjas on välja toodud, et tehnoloogial ja andmete analüüsil on potentsiaal, kuid „Tehisintellekt on Eestis veel kuriteoennetuses ja kuritegude lahendamises läbi mõtlemata ressurss, samas kui mujal seda juba kasutatakse […]“ (Justiitsministeerium, 2019a, lk 14). Samas märgitakse, et hetkel pole andmesisestusmeetodid kuritegevuse seesuguseks ennetamiseks piisavalt usaldusväärsed ja arenenud. Ka Eesti prokuratuur on juba käsitlenud uute digitaalsete lahenduste kasutuselevõttu, kus arvuti roll on andmete analüüs ning inimesele jääks otsustusprotsess, kui järgmise kümnendi võimalust optimeerida menetlustoiminguid (Kivistik, 2018).[49] Siiski, lahendused ei ole veel läbi töötatud.

„Kriminaalpoliitika põhialused aastani 2030“ eelnõus on punkti 1 lõikes 2 kirjas, et uudsete tehnoloogiliste võimaluste/lahenduste rakendamine analüüsiks on oluline.

> *Kriminaalpoliitika roll on kujundada õiguskuulekat ühiskonda ja selleks vajalikke väärtusi. Kriminaalpoliitika tugineb teadmistele ja analüüsile, kasutades selleks andmeid ja tehnoloogiat. Kriminaalpoliitika arvestab tehnoloogiast ja globaalsetest suundumustest johtuvaid tulevikuriske ja -võimalusi.*

Punkti 2 lõikes 4 rõhutatakse tehnoloogia kasutuse ja digitaliseerimise tähtsust kriminaalpoliitikas ning uudsete oskuste õpetamist ning arendamist.

> *Süüteomenetlus muudetakse digitaalseks,[50] personaalseks ja asjatut bürokraatiat vältivaks. Õiguskaitsetöötaja ettevalmistus annab sellised teadmised ja oskused, mis soodustavad tehnoloogia kasutamist […].*

Kuigi antud dokumentides ei ole kindlaid plaane tehisintellekti süsteemide väljatöötamiseks välja toodud ja valdkondi, kus neid rakendada, mainitud, on eeldatud, et tulevikus on toimingud digitaalsed ning suuremahulisi andmeid analüüsitakse. „Siseturvalisuse arengukava 2020–2030“ kavand rõhutab sarnaselt teistele, et info- ja digiajastul on uudsete

---

49 Samal teemal: Parmas „Esimene Stuudio“, 4.02.2020.

50 Eelnõu juurde käivas seletuskirjas avatakse digitaalsuse mõistet kui […] mahukate toimingute kiiret läbiviimist läbi innovaatilise tarkvara arendamise ja digitõendite eelistamise […]. Õiguskaitses tuleb lisaks traditsioonilistele uurimis- ja õigusteadmistele eeldada ka tehnoloogiateadmisi, väga tihedat rahvusvahelist koostööd, tänasest automatiseeritumat analüüsi, mis eeldab suuri arvutusvõimsusi suurte andmemahtude analüüsimiseks […] (Justiitsministeerium, 2019b, lk 11).

lähenemiste väljatöötamine võtmetähtsusega, kuid praegu ei olda uute tehnoloogiliste lahenduste kasutuselevõtuks valmis. See tähendab, et olemasolevad lahendused on puudulikud ning suuremat rõhku tuleb pöörata andmete analüüsi võimaluste rakendamisele, et võimalikke probleeme ennetada ja lahendada (Siseministeerium, 2020b, lk 35; Siseministeerium, 2019, lk 11). Arengukava koostamise määruses seisab, et „Arengukava elluviimisel on vaja […] tähelepanu pöörata siseturvalisuse teenuste nutikale ja mõjusale pakkumisele, sh tehnoloogia ja innovaatiliste lahenduste arendamisele." (Siseministeerium, 2019, lk 9).

Lisaks on tehnoloogia kasutuselevõtu olulisus ja ajakohasus välja toodud „Siseturvalisuse arengukava 2020–2030 koostamise määruses" kui ka arengukava kavandis kui meede tööjõu puuduse leevendamiseks. Määruses tõdetakse, et „rahvastiku vananemine, sh tööjõuturule sisenevate noorte arvu vähenemine ja eriteenistujate pensionile siirdumine, toovad kaasa raskusi siseturvalisuse teenuste pakkumiseks vajalike teenistujate leidmisel" (Siseministeerium, 2019, lk 11). Nii märgitakse, et üks oluline personalipoliitika osa peaks olema uute optimaalsete tehnoloogiate kasutuselevõtt sisejulgeoleku teenuste tagamiseks; tööde automatiseerimine (Siseministeerium, 2019, lk 11; Siseministeerium, 2020b, lk 9). Personali nappus on võimalik ohukoht kriisiolukordades, mil töökätest võib situatsiooni lahendamisel puudus tulla (Siseministeerium, 2020b, lk 9).

Lühidalt, korrakaitse on just selline avaliku sektori osa, kus võiks olla suur abi suurandmete analüüsist (eriti isikustatud andmehulka silmas pidades) eeldatud optimeerimisel ja assisteerimisel. Mõne tehisintellekti süsteemi kasutuselevõtt, nt digitaalsed menetlustoimingud, on juba plaanis. Samas ei ole kitsad tehisintellekti süsteemid aga mitte alati autonoomsed, vaid kõigest toetavad inimeste tööprotsessi. Seetõttu ei tohiks neid vaadelda kui lõplikku lahendust demograafilistele aga ka hariduses toimuvatele muutustele, vaid pigem näha neis ajakohast täiustust toimivale praktikale.

4.3. SUURANDMETE KASUTUSELEVÕTU KITSASKOHAD

Suurandmete analüüsi potentsiaal ning võimalike krattide kasutuselevõtt on seotud ootusega optimeerida nii tööjõu kui ka aja kulu ja rahalisi vahendeid ning planeerida teenuseid paindlikumalt. Kuigi kratitehnoloogia integreerimine tööprotsessidesse võib luua lisaväärtust näiteks uute teadmiste või hoogsama töövoo näol, ei tohi seda pidada tööjõu- ja ressursiprobleemide üheseks lahenduseks, mis avaliku sektori teenuste toimimist määrab. Võimalik, et see automatiseerib teatud teenused ja tegevused ning võimaldab töötajad suuremat lisaväärtust tootvate tegevuste juurde suunata. Küll aga ei võida kõik teenused automatiseerimisest, vaid püsima peab jääma paindlikkus: kasutajasõbralikkus ja inimesekesksus peavad määrama teenuse olemuse, mitte ei tohi kasutuselevõttu planeerides keskenduda kõigest tehnoloogia rakendamise võimalikkusele ning kasu(mi) prognoosile.

Pealegi võimaldavad suurandmed leida küll suhteseoseid, mille tuvastamine digitaalsete andmete ning andmebaaside koostoimeta on olnud võimatu, kuid suurandmete analüüsi tulemus võib olla ka eksitav. Kui tavaliselt käsitleb andmeanalüüs lähteülesandega määratud ning metoodiliselt sobival viisil kogutud andmeid, kus valim on esinduslik, siis suurandmete puhul esitatakse uurimisküsimus kogutud ja kättesaadavate andmete pinnalt. Analüüsitulemus lähtub andmetest, mis on, mitte andmetest, mida vaja. Nii ei pruugi tulemus aga lähteülesandega kokku minna.

Eestis on senisele digitaliseerimisele tuginedes ootus, et avalik sektor on krattide kasutuselevõtul teerajaja, kuid teenuste väljatöötamine jääks erasektori kanda. Era- ja avaliku

sektori tehnoloogia arendamise ning rakendamise eeldustes ja ootustes on aga ebakõla, mis on tingitud lisaks arendushuvide erinevusele ka riigi- ja erasektori erinevatest võimalustest andmeid kasutada. Riigi kasutuses olevate suurandmete olemust vaadates ilmneb, et kuigi koostoimivaid teenuseid ja nende vahendusel ning nende jaoks tekkivaid andmeid, ka isikustatud andmeid, on palju ehk inimeste käitumisest saaks kokku panna keerukaid ja terviklikke profiile, on õigus ja ligipääs andmeid (nii) kasutada piiratud. Samuti nõuab erinevatest allikatest, k.a interneti rakenduste kasutamisest ja nutistust tekkivate andmete analüüs lisatööd, sest mitte kõik andmed ei ole üheti kodeeritavad või liigitatavad. Kõikidele tekkivatele digitaalsetele andmetele pole hetkel kasutustki. Lisaks ilmneb juba toimivate ja uute rakenduste plaanimise valguses küsitavusi nii selle kohta, kas teenused on omavahel koostoimivad, kui ka selle kohta, kas nende teenuste ülalpidamine ja andmete jätkusuutlik haldamine ning talletamine on võimalik.

Suurandmete rakendamise entusiasmi valguses ei võimalda ei olemasolev riiklik rahaline ressurss, andmete kasutusvõimalus ega ka olemasolevate teenuste olemus seda, et kõik teenused kratilahendustega kaetakse või lausa asendatakse. Olgugi et kõik digitaalsed teenused toodavad andmeid, ei tähenda see üheselt, et teenus kratilahendust rakendades kohe optimaalsem või sobilikum oleks. Nii tuleb riigil välja töötada teenused, mille puhul selle kaasajastamine ja tööprotsessi täiustamine kratilahenduste abil kõige tulemuslikum on, toimivaid protsesse parandab ning sellega ühiskonda lisaväärtust toodab. Nagu (rahaline) optimeerimine ei tohi olla lisaväärtus iseenesest, ei tohi seda olla ka soov omada suuremat kontrolli ühiskonnaliikmete käitumise üle lisanormide seadmise ning nende täitmise hindamise näol. Nt poe lojaalsusprogrammi andmetest tulenevalt saaks ostlemisharjumusi ja meditsiiniandmeid analüüsides tuvastada võimalikke probleemseid käitumismustreid ja vajadusel teha inimestele ettekirjutusi.

Selliste arengute ennetamiseks on juba praegu tarvis mõtestatud protsessi kratitehnoloogia kasutuselevõtuks.

Alljärgnevalt on toodud välja mõned kitsaskohad, mis on Eestis suurandmete analüüsi potentsiaali rakendamiseks olulised ning vajavad uue tehnoloogia tulemuslikuks, kasutajasõbralikuks ja jätkusuutlikuks rakendamiseks tähelepanu.

Õigusliku raami puudused

Virtuaalne ja füüsiline ruum on küll juba paarkümmend aastat paralleelselt eksisteerinud, kuid nende kahe ruumi omavaheline suhe on õiguslikult siiani mõneti määramata – aegajalt on need kaks ruumi samaväärsed, kuid on ka olukordi, kus virtuaalne on ainult füüsilise pikendus ning seal toimuv ei oma samaväärset kaalu.[51]

Kratistrateegia näeb ette avaliku sektori teenuste suuremat suurandmete kasutamist, sh korrakaitse töös. Suurandmete puhul, millest kratistrateegia tõukub, on tegemist digitaalsete andmetega, mis eksisteerivad virtuaalses ruumis. Teoreetiliselt on ligipääs ja andmete kasutus tänu andmete virtuaalsele kujule väga lihtne, kuid teenusepakkujatel on nii kohustus andmeid heaperemehelikult hoida ja kasutada kui ka võimalus neid endale sobivalt rakendada.

Nii on riigile teenuseid arendavatel erafirmadel väga keeruline ülesanne. Näiteks praegu on korrakaitseorganite andmekasutus fikseeritud ning andmetele ligipääs limiteeritud. Et korrakaitses saaks rakendada suurandmete analüüsi, oleks vaja üle vaadata ning fikssee-

[51] Näiteks virtuaalses ruumis toimuvate väärtegude puhul lähtutakse tõsiduse hindamisel tihti konkreetsest situatsioonist, mitte väärteo olemusest.

rida, missugustele andmetele, millisel alusel ja millises mahus on ligipääs võimalik. Lisaks on andmehulga suurenemise ning pideva täienemise valguses oluline mõtestada, kas ja kuidas on võimalus virtuaalseid andmeid kasutada ning menetlustoimingutes analüüsi rakendada. Pelgalt teoreetilisest võimalusest andmete analüüsi tulemusel uute teadmiste ja praktikateni jõudmiseks ei piisa. Et see ka praktiliselt teostatav oleks, peab andmekasutuse jätkusuutliku õiguspärasuse ja tulemuste õigusjärgsuse tagama õiguslik raam.

Lisaks võib digitaalsete andmete kasutusvajadus muutuda, nt kriisiolukorras on vaja ligipääsu ja analüüsivõimalust andmetele, mida tavaolukorras kõnealune ametkond kasutada ei või. Nii ei saa suurandmete kasutamise õigusliku raami väljatöötamisel piirduda kõigest tavaolukorras andmete kasutamise piiritlemisega. St õiguslikult ei piisa, kui suurandmete kasutamiseks antakse volitused ametkondadele, kelle töö on digitaliseerimise tõttu muutunud suurandmete haldamiseks ja kasutamiseks (nt maksuamet) või institutsioonidele, kus proportsionaalselt (lähi)tulevikus on suurandmete kasutamine tänu uutele kratilahendustele võimalik või paratamatu. Täpsete volituste andmine on küll andmete ja nii ka inimeste kaitsmise seisukohalt kahtlemata ainuõige lahendus, kuid ilma mehhanismita antud volitusi vajadusel laiendada või muuta võib olemasolevate andmete potentsiaal jääda rakendamata. Digitaalsed andmed pakuvad määramatu hulga võimalusi, kuidas neid uutes olukordades rakendada. See tähendab, ka kehtiv seadusandlus ja andmepoliitika peab olema valmis uuteks suurandmete kasutuseks, et vajalikke rakendusi õigel ajal arendada ja töösse võtta.

Oskusliku tööjõu vähesus

Tehisintellekti süsteemi käsitlemine tööjõupuuduse leevendajana võib olla lühinägelik ning kui süsteemide arendamisega paralleelselt ei koolitata tööjõudu, võib see viia pikaajaliste organisatsioonisiseste ja operatiivsust pärssivate arenguteni. Tehisintellekti süsteemi kasutuselevõtt eeldab nii juba töötavate inimeste täiend- ja ümberõpet kui ka uute oskustega töötajaskonda. Pelgalt tehnoloogiliste lahenduste arendamisest ei piisa, et tagada rakenduse sihtotstarbeline ja kasumlik kasutus – vaja on väljaõppele lisada nii rakenduse kasutusoskus kui ka mõistmine, kuidas andmed kogunevad ning erinevad tehnoloogilised lahendused neid töötlevad; kuidas andmete ristkasutus toimib.

Korrakaitses näiteks on kõrgest riigi digitaliseeritusest hoolimata erialases väljaõppes suured puudujäägid tehnoloogiliste lahenduste mõistmisel ning andmekogumisest ja koostoimest arusaamisel. Piret Pernik (2019, lk 97) toob oma uurimuses välja, et Sisekaitseakadeemia kadettidel ei ole head arusaama, kuidas andmed tekivad, kuidas need võiksid või peaksid omavahel ristkasutatavad olema ning missugused on võimalused ja piirangud nende andmete kasutamisel. Samuti arvab Pernik, et esmatasandi väljakutsetele vastajad (nt politsei patrulli töötajad) peavad oskama ka vähem tehnoloogiateadlikele selgitada, missugune vastutus e-lahendustega kaasneb; millised andmed neid lahendusi kasutades tekivad, kuidas neid kasutatakse jne (Pernik, 2019, lk 99). Ka sellisest väljaõppest on praegu vajaka.

Seega tähendaks uute tehnoloogiliste lahenduste kasutuselevõtt nt andmeanalüüsis vajadust koolitada mitte ainult tehnoloogiaalaseid spetsialiste-arendajaid, vaid muuta kõikide riigisektori teenistujate väljaõpet. Inimesel on nii andmebaaside toimimisel kui ka kitsaste tehisintellekti süsteemide rakendamisel võtmeroll. Kui kasutajad ei ole kompetentsed, muutub ka arendatud süsteem kasutuks või algoritmi vastus vigaseks.

Erafirmade ja riigi partnerlus

Tehnoloogilised lahendused, mille riigid võtavad kasutusele selleks, et suurandmete potentsiaali realiseerida, ei ole (suurandmete kogumisega sarnaselt) alati avaliku sektori välja töötatud. Pigem on tegemist tehnoloogiliste lahenduste ja rakendustega, mille on teinud-koostanud erafirmad. Sellist erafirmade ja riigi kootööd propageerib ka krati projekt – avalik sektor on tellija, kes oma vajadustest ja võimalustest erafirmat teavitab ning firmad arendavad vastava lahenduse.

Erafirmadel on tehnoloogia arendamisel selge eelis, sest kasumlikkusest lähtudes müüginumbrite või kasutajate arvu silmas pidades, saavad nad selle investeeringute puhul eesmärgiks seada: nii kaua kuni arendus on kasumlik, on ka selle arendamiseks vajalikud kulutused õigustatud. Riigil on tihti lihtsam tarbida erafirmade teenust kui ise vajalikku lahendust välja töötada. Teenust sisse ostes pole riigil vajadust investeerida tehnoloogia arendamisse: riiklikult arendatud teenusel peab olema märkimisväärne kasutegur; kulutused peavad olema proportsionaalsed. Ostes saab riik vastavalt vajadusele ja võimalustele valida pakutavate lahenduste vahel. Nii on lihtsam vajadusel suunda muuta; uut tehnoloogiat või metodoloogiat rakendada; andmestiku muutumisega kohaneda. Sõltudes riikliku tähtsusega teenuse puhul erafirma teenusest – eriti veel mõne domineeriva riigi erafirma teenusest –, seab riik end aga olukorda, millega kaasnevad riskid.

Esmalt on erafirmadel võimalus otsustada, kellele ja mis alustel nad oma teenust osutavad. Erafirmade rakenduste kasutamine on ülemaailmselt laialdast kasutamist leidnud praktika. Ameerika Ühendriikide korrakaitses näiteks on kasutusel rakendused, mis (I) hindavad võimalikke seoseid toimepandud kuritegude vahel (rakendus Patternizr) (Griffard, 2019, p. 76); (II) hindavad andmestiku alusel kuritegevuse võimalust erinevates piirkondades (rakendused CrimeCentral või PredPol) (Kirkpartick, 2017); (III) aitavad välja arvutada võimalikku korduvrikkumist (COMPAS robot; Turk ja Pind, 2019, lk 52). Veelgi enam, mõne rakenduse puhul on võimalus, et rakenduse tarbijad on ühteaegu nii riigi kui ka erasektori ettevõtted. Näiteks näotuvastusprogrammi Clearview AI[52] kasutajad on nii riikide julgeolekuinstitutsioonid kui ka suured eraettevõtted (Lyons, 2020).

Riigile, kes teenust sisse ostab, tähendab see, et on olemas ka riigist väljaspool olev või eraldiseisev firma, mis on teadlik ja omab ligipääsu andmetele ning tehnoloogiatele, millega neid analüüsitakse. Ehk kellelgi teisel on võim riigis kasutatava tehnoloogia üle ning võimalus otsustada, kes teab protsessidest ja andmetest ning kes mitte. Pealegi, ka teenuse pakkumise lõppedes ei tähenda see alati automaatselt, et andmed kustutatakse või edastatakse, võimalused teenus lahti kodeerida hävitatakse jne. Seega, teenust sisse ostes võtab riik riski, sest ligipääs toimivatele protsessidele ei ole kõigest nende enda hallata. Näiteks ID-kaardi tootjafirma Gemalto muutus turvariskiks Eesti elanikele ja ka Eesti riigile. Kuigi teenuse tarbimine või sisseostmine lõpetati, toimus selle tulemusel aga mitu kohtuvaidlust. Lahendist hoolimata on ja jääb riik sõltuvaks firmast, kes on kasutanud ja hallanud suure hulga eestlaste andmeid. Kuigi Eestil on Gemalto juhtumi varal olemas kogemus, millega riik elutähtsaid teenuseid tagavaid lahendusi sisse ostes riskib, pole see oluliselt muutnud kasutusel olevat praktikat. Ka andmesaatkonna puhul toetavad riiki mitu firmat: Cybernetica, Dell EMC, Ericsson, Open Node ja Telia (e-eesti, 2020). Siingi sõltub teenuse ladus toimimine kõigi osapoolte koostöövalmidusest ning jagatud vastutusest ja arusaamadest. Ennetamaks juba olemasolevate ja ka tulevikus kasutusele võetavate tehnoloogiate puhul sarnaste probleemide kordumist ning täpsustades teenuse tellija nõudmisi ning teenuse pakkuja kohustusi, oleks Gemalto juhtumi valguses vaja

[52] Clearview fuktsioon on avalikest allikatest, nt Facebookist, pilte koguda, et siis tehnoloogia abil vajadusel inimesi tuvastada.

olemasolevad lepingud üle hinnata. Nii saab riik end ja enda kodanikke kaitsta sisseostetud teenusega kaasnevate turvariskide eest, eriti andmete käitlemisel.

Lisaks pakuvad (globaalsed) erafirmad rakendusi, mis on riigile hädavajalike teenuste osutamiseks ning julgeoleku tagamiseks vajalikud. Rakendusi pakutakse tihti alguses tasuta, et hiljem teenus tasuliseks muuta. Ajal, mil teenus tasuliseks muutub, on see riigis aga juba kasutusele võetud ja vajalik; süsteemi toimimiseks rakendunud. Selline taktika tagab teenust pakkuvatele erafirmadele kliendibaasi (Morozov, 2018), kuid muudab riigid erafirmade ees jõuetuks. Kui sama teenust pakutakse aga lausa mitmes riigiasutuses või ka sarnaselt eraettevõtluses, muudab see teenuse tarbijad, nii otsesed kui ka kaudsed, (veelgi) haavatavamaks – ühest rakendusest ning selle toimimisest, turvalisusest ning selle arendanud firma diskreetsusest sõltub väga mitme erineva valdkonna asutuste töö.

Pealegi on suur võim koondunud üksikute riikide, eriti USA ja Hiina, erafirmade kätte – Euroopa Liidu Tehisintellekti Süsteemide strateegia tunnistab, et Euroopa Liit peab suunama rohkem ressursse konkurentsivõime parandamiseks, ajude äravoolu tõkestamiseks ning spetsialistide Euroopasse meelitamiseks. See on hädavajalik, et püsida tehnoloogia arenguga kaasas. Et nii üksikud riigid kui ka EL tervikuna saaks kõik vajaliku toodetud, tuleb teenuseid sisse osta; mingit osa teenusest või kogu teenust võib pakkuda väljaspool Euroopat olev riik või erafirma.

Selliste ühisest majandus- ja koostööruumist väljastpoolt sisse ostetud teenuste puhul võib ülalmainitule lisaks hakata mängima aga rolli ka riik, kust see pärit on. Näiteks 2019. aastal tekkinud USA ja Hiina RV kaubandussõja tulemusena kehtestasid mõlemad riigid vastastikku sanktsioone kaubavahetusele ning ka teatud telekommunikatsioonifirmadele ning nende teenustele.[53] Riikidevaheliste suhete halvenedes võib riik end taas leida situatsioonist, kus erafirma, kes on neile seni teenust pakkunud, pole enam sobilik või võimalik partner. Jälle on riigil suur risk, sest andmed ja toimingud on käinud läbi erafirma teenuse ning missugused ligipääsud teenusesse sisse ehitatud on, on kõigest teenusepakkuja teada. Pealegi võib ebaõnnestunud riiklike suhete puhul olla erafirmal motivatsioon oma positsiooni kasutada ja andmeid kuritarvitada isegi suurem. Teenuste sisseostu planeerides tuleb selliste võimalustega arvestada.

Lõpetuseks, kui riik jääb sõltuma erafirmade pakutavatest teenustest, võib väikese ja üsna omanäoliste tehnoloogiliste lahendustega silmapaistev Eesti jääda vaeslapse rolli. Nimelt ei hõlma suurfirmade arendatavad lahendused alati riiklike andmebaaside või andmete kogumise eripärasid. Nii võib Eesti leida end situatsioonist, kus tehisintellekti süsteemid ei ühti juba toimiva isikustatud andmete süsteemiga ning hulk arendatud rakendusi on kas kasutud või nõuavad suuri muutusi ja lisarahastust.

Jätkusuutlikkus andmepoliitikas

Kratiprojekti aruanne tuvastas, et tehnoloogia arendajatel puudub ülevaade võimalikest kasutatavatest andmetest. Samas pole ka üheselt kindel, kas praegu toimivad andmed ja andmevahetuse lahendused uutele rakendustele sobivad. Et tehisintellekti süsteemid toimivad andmetel ning jätkusuutlik toimimine nõuab süsteemset andmete kogumist ja talletamist, siis on juba kratiprojekte arendades oluline välja töötada, milliseid andmeid

[53] USA-Hiina suhete jahenemise üks põhjus oligi Hiina firma Huawei. Firmal on globaalne haare: teenuseid pakutakse nii eratarbijatele, nt nutistu seadmed, firmadele, nt riistvara, kui ka riikidele, k.a terviklikud telekommunikatsiooni lahendused, nt 5G-sidevõrk. Tehnoloogialahendustesse on ehitatud tagauksed, mis jätavad kolmandatele osapooltele, k.a Hiina RV luureteenistusele, ligipääsu kasutaja andmetele. Huawei läbipaistmatuse tõttu ei suuda teised firmad Huawei huvisid tuvastada (Kaska, Bechvard & Minarik, 2019).

on tarvis koguda, et plaanitava teenuse pikaajaline ja tõrgeteta toimimine oleks võimalik. Samuti on oluline, et plaanitav ei läheks vastuollu juba hetkel toimiva praktikaga, vastasel juhul võivad teatud andmed kasutuks osutuda. Liigne innovatsiooni ambitsioon võib saada Eesti digijätkusuutlikkusele saatuslikuks.

Hetkel räägitakse kratiprojektide raames andmetest – ka nendest, mida tulevikus vaja on – pelgalt abstraktselt. Suurandmete eriilmelisusest hoolimata viidatakse neile kui suure potentsiaaliga homogeensetele andmetele. Teenused on ju alles mõtte või varases arendamise järgus.

Samas on riik juba andmesaatkonda luues teinud valiku, milliseid andmeid eelistada. Samuti on riigil ainult limiteeritud võimekus tagada andmete talletamine. Seetõttu võivad jätkusuutlikkus ja innovatsioon uute tehnoloogiliste lahenduste väljatöötamisel vastanduda: andmed, mida on tarvis koguda ja talletada täna või tulevikus, et elutähtsad teenused toimiksid, võivad olla omavahel vastuolus või pole koostoimivad. Sellistest võimalikest vastuoludest teadlik olles tuleb riigil langetada kaalutletud ning tulevikku vaatavad otsused andmete talletamiseks aga juba lähiajal.

Arusaadavalt on riigi ees suur küsimus andmepoliitika kujundamisest – kas koguda ja hoida tallel kõik võimalikud andmed või piiritleda juba praegu, mida kogutakse, ning riskida võimalusega, et tulevikus vajaminevaid või kasulikke andmeid ei ole. Esimene võimalus on mahukas ja kulukas, teine aga loob võimaluse, et teenuseid ja isegi teenuse pakkujaid on vaja vahetada ning arendatud ja toimiv süsteem ümber muuta. Ühtne andmepoliitika ning läbimõeldud tehnoloogia kasutuselevõtt on ainsad võimalused, kuidas juba täna toimiv ka tulevikuks kindlustada. Vastasel juhul võib aastaid kestnud arendus kasutuks osutuda.

Lisaks tuleb riikliku andmepoliitika puhul silmas pidada andmebaaside jagamise kohustust ning võimalikke edasisi koostalituse projekte, mis Eestil Euroopa Liidu liikmena on. Euroopa Liit töötab mitmes valdkonnas, nt sise- ja välisjulgeoleku eesmärgil, andmebaaside jagamist võimaldavate lahenduste väljatöötamise kallal.[54] Iga liikmesriigi panus on süsteemi toimimiseks ja jätkusuutlikkuseks vajalik ning riigis kehtestatud andmepoliitika ei tohi plaanitud koostööd ja rakendust kehtetuks või sisutuks muuta. Ka siin on võtmeküsimusteks riigis kasutatava uue tehnoloogia läbimõtlemine ja andmete edasine kogumine, sest kasutatavad lahendused peavad võimaldama koostoimet EL-i lahendustega.

Innovatsiooniretoorika

Teadus- ja majandusarengust räägitakse uue tehnoloogia arendamise valguses kui üksteist toetavatest ja teineteisest tõukuvatest protsessidest. Nii on see ka erinevates tehnoloogia arengu tulevikuvisioonides nii Euroopa Liidu kui ka riigi tasandil kirja pandud – ühiskonna jätkusuutlik majandusareng on võimalik vaid juhul, kui investeeritakse tehnoloogiaalasesse teadustegevusse (Majandus- ja Kommunikatsiooniministeerium ja Riigikantselei, 2019). Eesti kontekstis tähendab see, et innovatiivsete tehnoloogiliste riist- või tarkvara lahedusteni, mida saaks kas Eestis (või veel parem, maailmas) turustada, on suurem tõenäosus jõuda juhul, kui teadusesse ja tehnikaarendusse piisavalt investeeritakse ning tagatakse arenemiseks sobilikud tingimused. Erilise tähelepanu all on infotehnoloogilised lahendused, nt IKT-idufirmade arendatavad teenused, kus kasu- või kasumi-

[54] Näiteks EUCRIS-TNC alustab üle Euroopa 2022 või 2023 (Särekanno, 2020).

tegur võivad olla proportsionaalselt suuremad ja kiiremini saavutatavad.[55] Konkreetsete teenuste arendamise sobilikkusest ja vajalikkusest on täna olulisem mõista üldisemalt seda kausaalset seost, mida sellise retoorikaga majanduselt-teaduselt eeldatakse ning mis eesmärgil toimuvat ka toetatakse.

Loomulikult on globaliseerunud teadus- ja majandusruumi valguses jätkusuutlikkuse ning jätkusuutliku arengu tagamiseks olulised nii uudsus kui ka kasumlikkus. Selline lähenemine (või prioriteetide seadmine) aga pärsib laiemas ühiskondlikus ja riiklikus kontekstis mõistmist, kui oluline on teadus- ja arendustegevus ning vajalik rakendada innovaatilisi lahendusi. See tähendab, et vaid kasumlike tehnoloogiliste lahenduste arendamist soosides võivad jääda märkamata, rahastamata ja tegemata uuendused ja teadustööd teemadel, mis ei vasta juba eos ootusele panustada majandusarengusse. Riiklik IT-teenuste arendamisvajadus ning globaalsetel trendidel tuginevad ootused ei ühti alati.

Esmalt vajab Eesti kõrge digitaliseeritusega teaduslikku lähenemist ning jätkusuutlikku tehnoloogilist arendustegevust, mis jätkab töös olevate lahenduste mõtestamist, toetamist ja edasiarendamist. Tegemist ei ole sel juhul alati turustatavate ja kasumlike põnevate lahenduste, vaid hädavajaliku – süsteemi tööd tagava – uurimise ja arendusega. Digitaalsete lahenduste jätkusuutlikuks ning seaduspäraseks kasutamiseks on see oluline, kuid pelgalt konkreetsete lahenduste ja riigi kontekstis.

Teisalt on Eesti koostoimivad teenused ning kogutavad ja tekkivad suurandmed omanäolised, nagu juba III alapeatükis sai kirjeldatud. See tähendab, et ka neid andmeid kasutavate uute tehnoloogiliste lahenduste väljatöötamisel ei saa eeldada, et valminud tehnoloogiline lahendus oleks laiemalt turustatav või kasutatav. Samas on selline arendustegevus oluline, sest nõnda rakendatakse Eesti suurandmete potentsiaali.

Kolmandaks, nagu ka eelnevast era- ja avaliku sektori suhte kohta selgus, on ootus, et avalik sektor oleks tellija ja tarbija rollis, samas kui erasektori ettevõtted peaksid teenuseid arendama ja pakkuma. Samas on selline lähenemine aga teadus- ja majandusarengu koostoimimise valguses poolik. Kui eeldatakse, et tehnoloogia areng tagab majanduskasvu, siis peab ka tehnoloogiast saadav tulu (nii otsene kui ka kaudne), selleks võimaluse andma. Riigisektori teenuste sisseostmine käib aga parima, st majanduslikult optimaalsema pakkumise põhimõttel. Kas saab siis eeldada, et erasektori teenusepakkujad oleksid majandust silmas pidades riigile vajalike teenuste arendamisest huvitatud. See tähendab, riigil endal võib olla vajadus tehnoloogilistesse lahendustesse ja uutesse analüüsimeetoditesse palju rohkem panustada.

Lühidalt, Eesti digitaalsete lahenduste püsimajäämiseks ning uute lahenduste väljatöötamiseks peaks riik arendama maailmatrendide järgimisele alternatiivse lähenemise ja seda toetava rahastusmudeli. See on teadustegevus ja tehnoloogiaarendus, mis tegeleb riigi ja kodanike jaoks hädavajaliku ning jätkusuutliku arendamisega. Siin ei ole eesmärk mitte digitaalse eduloo edasiarendamine, vaid riigi (digi)turvalisus ja püsimajäämine. Digitaalsete teenuste kasutuskindlus ning riskide hindamine ja elanikkonna teavitamine nendest on usaldusväärse e-riigi alus (Vaks, 2019). Olgugi digitaalne, on e-teenuste puudumise korral häiritud ju kogu füüsilise ühiskonna töö, k.a elutähtsad teenused (Koort, 2019).

Lisaks ei saa uut tehnoloogiat pidada pelgalt uueks innovatsioonivõimaluseks ja edulooks, vaid suuremat tähelepanu tuleb pöörata ka tehnoloogiaga kaasnevatele riskidele. Innovatsiooni jätkusuutlikkus sõltub ka valmisolekust oma teenuseid kaitsta ning või-

[55] Näiteks Covid-19 pandeemia tagajärjel tekkinud majanduslanguse/majanduskasvu languse valguses pakkus Eesti president Kersti Kaljulaid, et IT-firmadele (kui Eesti turundusartiklile ja visiitkaardile) võiks rakendada maksusoodustusi, mis antud sektorit elavdada võiks.

malikke ohte ennetada. Seetõttu on oluline, et teaduse ja innovatsiooni diskursuse kõrval oleks nähtav ka turvalisuse diskursus. Digitaalne ei ole iseenesest turvaline, vaid vastupidi, haavatav ka distantsilt. Nii on uue tehnoloogia küsimus ka küsimus ühiskonna muutuvatest turvavajadustest ning kuritegevuse muutuvast näost. Eesti on suurt rõhku pannud riiklikule küberjulgeolekule ja -turvalisusele ning üha kasvava küberkuritegevuse ennetamisele ja tuvastamisele[56] ning on riigina maailmapoliitikas küberteemade suunanäitaja (nt küberdiplomaatia teema on Eesti välispoliitika visiitkaart (Välisministeerium, 2019)). Oluliselt vähem on tähelepanu pööratud aga kõrge digitaliseerituse ja uute tehnoloogiliste lahendustega kaasnevatele ülesannetele ja arengutele, mis Eesti ühiskonda ja selle liikmeid ees ootavad, nt muutuvad ootused tulevastele töötajatele või mõjud elanikkonnale uue tehnoloogia kasutuselevõtul. Ka seda tuleb arengukavades meeles pidada ning tagada vastavad meetmed vajalikuks teadus- ja arendustegevuseks.

Pealegi on digitaliseerimisega ilmnenud suurandmete kasutamiseks ja andmeanalüüsist lähtuvate muutuste tegemiseks teoreetiliselt mõõtmatu potentsiaal, sest seni pole suhteseoseid sellises mahus ja põhjalikkuses hinnata saanud. Andmete positsioon on oma tähtsuses muutumas. Samas ei tohi aga pelgalt suurandmete analüüsi väljatöötamisele panustada ning vajalike kratilahenduste arendamist toetada. Tehnoloogia areng on kiire ning uusi väljakutseid ja võimalusi on ning tekib pidevalt juurde. Nii on jätkuvalt oluline mitmekesise tehnoloogia rakendamine ning vastavasisuliste uurimuste toetamine.

[56] Vaata näiteks Cybersecurity in Estonia 2020 (Riigi Infosüsteemi Amet, 2020a); Riigi Infosüsteemi Ameti Aastaraamat, 2020 (Riigi Infosüsteemi Amet, 2020b).

KOKKUVÕTE

Suurandmete analüüsil põhinevate tehisintellekti süsteemide ehk kratilahenduste arendamise ja rakendamise potentsiaalile olemasolevate teenuste täiustamiseks ning uute lahenduste väljatöötamiseks pööratakse nii riiklikult kui ka rahvusvahelises kontekstis üha enam tähelepanu.[57] Riikide ja erafirmade kasutuses olevad ning pidevalt kasvavad digitaalsete andmete hulgad võimaldavad täna analüüsides leida suhteseoseid, mis kuni IKT arengu ning interneti laialdase leviku ja kasutuseni polnud võimalikud. Nii eeldatakse suurandmete analüüsilt ja kasutuselt andmekogude uudsusest ja hetkel veel väga limiteeritud rakendamisvõimalustest hoolimata suurt majanduslikku ja ühiskondlikku kasu. Levinud on arusaam, et andmeanalüüsil põhineva kratitehnoloogia laialdasem kasutuselevõtt lubaks pakkuda nii vähem kulukaid kui ka asjakohasemaid või täpsemaid lahendusi ja teenuseid. Suurandmetele, mis antud võimalused loovad, viidatakse sellises innovatsiooniretoorikas samas aga kui homogeensele massile andmetele, mille kasutamist erinevate valdkondade töös juba praegu takistab ennekõike vaid sobiliku metodoloogia, tehnoloogia või rakenduse puudumine.

Selle levinud arusaama valguses keskendus käesolev uurimus suurandmete olemuse avamisele näitamaks, et mõiste „suurandmed“ on vaid ühisnimetaja, mis aga viitab tegelikult väga eriilmelistele andmekogudele. Pelgalt suurandmete kasutamise potentsiaalile keskendumine varjutab andmete eriilmelisusest tingitud analüüsi keerukuse ning haldajate erinevatest globaalsetest positsioonidest tingitud andmete koostoime raskused. Kuigi innovatsiooniretoorikas pööratakse sellele tähelepanu harva, on suurandmed moodustavad andmekogud aga kõik eriilmelised, mis tingitud andmeid tootvate rakenduste, teenuste ja seadmete erinevustest, andmete haldajate erinevatest rollidest ja kohustustest, andmete koostoime võimekusest ning andmefaili enda vormist ja sisust. Suurandmete kasutusvõimalused ning -vajadus on andmete olemusest tulenevalt erinevad. Kõikidele tekkivatele andmetele hetkel asjakohast kasutust ei olegi: andmete kasutamisel võivad saada takistuseks nii andmetöötluse keerukus (nt video) kui ka andmete ebaolulisus või kogutud teabe rakendamise keerukus (nt metaandmed). Lisaks on suurandmete olemuse juures määravaks ka erinevused riikide andmekogumise praktikates, seadusandluses ja kohaliku kogukonna privaatsusnõuetes või -ootustes. Nii leidis uurimus, et vaatamata üha suuremale rahvusvahelisele tähelepanule, mida suurandmed saavad, ning ootustele, mis kratilahendustele pannakse, tuleb riiklikul tasandil rakendusvõimaluste mõtestamisel, andmepoliitika kujundamisel ning teenuste arendamisel lähtuda ennekõike juba kasutusel olevatest tehnoloogilistest- ja digilahendustest. Olemasolevad lahendused määravad nii riiklike suurandmete olemuse kui ka nende edasised kasutus- ja arendusvõimalused. Nii tuleb rahvusvaheliselt tuvastatud tehisintellekti süsteemide potentsiaali kohandada

[57] Ka Eestis tutvustati 2019. aastal „Eesti Tehisintellekti kasutuselevõtu ekspertrühma aruannet“ (Majandus- ja Kommunikatsiooniministeerium ja Riigikantselei, 2019).

kohalike võimaluste ja vajadustega, et tagada riiklikult juba toimivate teenuste jätkusuutlikkus aga ka mõtestatud tehnoloogiline areng ja rakenduste kasutuselevõtt. Andmekasutus, mis on võimalik ja kasumlik rahvusvahelisel suurfirmal, ei pruugi seda alati olla mõnel (väike)riigil.

Uurimus, mis eristas suurandmete komplekssе olemuse selgitamiseks (koostoimivaid) andmebaase, internetiteenuseid ning nutistu seadmete kasutamsel tekkivaid andmeid, leidis, et Eesti suurandmed moodustavad andmekogud on üldise riikidega kaasnevate suurandmete mudeliga suures plaanis sarnased ehk rahvusvaheliselt tuvastatud suurandmete potentsiaalist rääkimine on kohane. Tinglikult moodustavad ka Eesti suurandmed (i) erinevad digitaliseeritud andmebaasid ning digitaalsetest teenustest tekkivad andmebaasid; (ii) riigi residentide ja kodanike internetti kasutavate rakenduste kasutusest tekkivad andmed ning (iii) üha kasvav hulk automaatselt andmeid koguvaid seadmeid, mis on saanud uue võimekuse andmeid salvestada ja jagada ning ka uued nutistu seadmed. Kõigi andmetega käivad lisaks kaasas ka metaandmed, mis edastavad andmefaili tehnilised ja statistilised väärtused. Kõikide suurandmeid moodustavate andmete kogumise ja talletamise määrab teenusepakkuja kasutushuvi ja kogumise kohustus ning vajadus.

Eesti suurandmeid eristab teiste riikide omadest aga suur avaliku sektori ja erafirmade digitaalsete ehk e-lahenduste osakaal, kus isikutuvastamine käib isikukoodiga ja üha enam isikukoodiga seotud tehnoloogiliste rakenduste abil. Nii on paljud andmebaaside andmed isikustatud ning teiste andmebaaside ja teenustega koostoimivad. Ennekõike on isikustatud andmete tekkimine paratamatu ametlikke toiminguid tehes, nt arstivisiidil, pangakontoris või e-maksuametis. Lisaks on võimalus sarnast isikutuvastust kasutada aga ka lihtsamate toimingute puhul, nagu raamatukogukülastus, poe lojaalsusprogramm, kasiinokülastus või netikommentaariumi sõnavõtt. Just selline isikustatud andmete paljusus, eriilmelisus ja koostoimevõime iseloomustavad Eesti suurandmeid. Riigi jätkusuutlikkuse tagamiseks on olulisim andmehulk talletatud andmesaatkonda, kuid see on vaid osa riigi hallatavatest ning väga väike osa üldisest riigi elanike suurandmed moodustavatest andmetest. Suurandmeid, mille analüüsi uute rakenduste arendamisel kasutada võiks, on nii erafirmadel kui ka riigi institutsioonidel oluliselt rohkem, kuid nende olemus on eriilmelisem ning rakendamisvõimalus ja -vajadus erinevad.

Järjest kasvava huvi kratilahenduste vastu ja üha suureneva hulga andmete valguses toob uurimus välja, et just tänu edukale digitaliseerimisele on nii Eesti tehnoloogiaarendajad, poliitikakujundajad kui ka riik kui teenuste tellija ja rakendaja, suurte väljakutsete ees. Seni arendatud koostoimivad digilahendused tagavad e-Eesti eduka toimimise ning teenuste jätkusuutlikkuse jaoks on oluline olemasolevaid süsteeme ja teenuseid arendada ja uuendada. See on ressursimahukas, sest raha on vaja nii arendus- kui ka haldustegevuste jaoks. Samas on kratilaheduste prognoositavad kasutegurid aga tekitanud huvi neid juba olemasolevate teenuste täiustamiseks või asendamiseks arendada. See on samuti ressursimahukas, aga peidab endas ka riski, et juba toimivad lahendused uude süsteemi ei sobi. Veelgi enam, kratilahenduse rakendamine võib ka kasumita jääda või sobimatuks osutuda: andmete töötlemisel tuginev protsess ei ole iseenesest veel kasumlik. Rahvusvaheline teadus-, tehnoloogia- ja majanduskogukond, kes kratilahendusi arendab ja propageerib, aga ka riigid, kes avaliku sektori teenistuses kratilahendusi juba rakendamas, on andmete kasutamise võimaluste ja ka arendatavate tehnoloogiate puhul tihti erineval positsioonil kui Eesti. Tulevikulahendusi planeerides tuleb Eestil silmas pidada, et tänu juba kasutusel olevatele lahendustele võivad mujal toimivad lahendused Eestis sobimatuks või liiga algeliseks osutuda. Nii tuleb rakenduste kohandamisel sobivusele suurt tähelepanu pöörata. Pealegi ei tohi uusi lahendusi planeerides kõrvale jätta ka elanikkonda, kes arendatavaid rakendusi oma töös ja igapäevatoimingutes kasutama peavad hakkama. Ainult nii üldise

ühiskondliku võimekuse tõstmine kui ka vajalike spetsialistide koolitamine võimaldab uute tehnoloogiliste lahenduste eduka kasutuselevõtu; kasutajaskonna koolitamine peab samuti olema osa innovatsiooniliste lahenduste kasutuselevõtu plaanist.

Suurandmete uurimuse tulemusel saab nii järeldada, et Eesti senist digitaalsete lahenduste toimimist silmas pidades on mõtestatud planeerimine ning kogutud andmete pakutavate võimaluste hindamine oluline uudsete, kasulike ning hästitoimivate lahenduste väljatöötamiseks. Seega on tähtis, et arenguvisioonide seadmisel ei keskenduks Eesti riigina kõigest rahvusvaheliselt tunnustatud kratilahenduste arendamisele või arendatu sisseostmisele, mis Eesti digitaalset edulugu kinnistaks, vaid suudaks leida rahastust ja kompetentsi ka Eestile unikaalsete lahenduste väljatöötamiseks, mis olemasolevate ja koostoimivate andmekogudega sobiks ning teenuse pakkumist ja tarbimist parendaks. Et nii digilahenduste kui ka planeeritavate kratilahenduste juures on andmete olemasolu ja kasutusvõimalus olulised, siis tuleb hinnata võimalikke kitsaskohti ja rakendada tulevikuvisioonile vastavat jätkusuutlikku andmepoliitikat. Uurimuse Lisa 1 (samas, lk 35–41) toob välja täpsema riigi kasutatavate lahenduste arendamise, kasutamise ja jätkusuutlikkusega kaasnevate küsimuste ja kitsaskohtade taksonoomia, mis põhineb suurandmete kasutamisel või suurandmete analüüsi rakendamisel riigi toimimisele ning elanikele hädavajalike teenuste osutamisel.

Veel enne kui uute rakendusteni jõutakse, on kratilahenduste edukaks arendamiseks ja kasutuselevõtuks vaja Eestis lahendada mitmed senistes praktikates esinevad vajakajäämised, mille uurimus tuvastas, käsitledes Eestis juba kasutusel olevate tehnoloogiliste lahendustega ilmnenud probleeme, selgitades Eesti suurandmete olemust ja hinnates esitatud tulevikuvisiooni. Ilma tuvastatud vajakajäämisi adresseerimata on kratilahenduste arendamine, rakendamine ning jätkusuutlik toimimine kaheldavad.

1. Andmete kasutamiseks ning kratilahenduste arendamiseks on vaja seadusandlust, mis määraks üheselt virtuaalse ja reaalse suhte ning sellest tulenevalt digitaalsete andmete kasutamise ning tehnoloogiliste uuenduste rakendusvõimalused.
2. Välja on vaja töötada viis, kuidas tõsta üldist digiteadlikkust ja ka tehnoloogiaalast kompetentsi ning uute rakenduste kasutusoskust tööjõu väljaõppe ja/või täiendõppe osana.
3. Riigi kui uute (riigi toimimiseks vajalike) tehnoloogiliste lahenduste (kratilahenduste) tellija ootused ja nõuded ning erafirmade kui teenusepakkujate kohustused on vaja teenuse jätkusuutlikkuse ja andmete turvalise ning konfidentsiaalse kasutamise eesmärgil paremini vormistada, näiteks seadusandluses ära märkida.
4. Riigisektorisse uute tehnoloogiliste rakenduste tellimise vajaduse ja võimalikkuse hindamiseks on vaja välja töötada kord, et oleks tagatud uute tehnoloogiliste lahenduste vajaduspõhine ja jätkusuutlik rakendamine.
5. Kratilahenduste jätkusuutlikkuse tagamiseks on tarvis välja töötada riiklikud andmete kogumise ja talletamise põhimõtted.
6. Poliitikakujundajatel tuleb suuremat tähelepanu pöörata tehnoloogia arenguga kaasnevatele ühiskondlikele muutustele ning võimalikele julgeolekuriskidele, mis ei piirdu vaid küberjulgeoleku ja küberkuritegevuse sektoritega.

Kokkuvõttes juhib uurimus tähelepanu tõsiasjale, et optimeerimise, kasumlikkuse ja tunnustuse lootus, mis suurandmete kasutuselevõtu sooviga (varjatult) kaasnevad, ei tohi varjutada ja varjata tehnoloogiakasutusega kaasnevaid riske: ühiskonna toimimist mõjutavate tehnoloogiate arendamine ning nende kasutuselevõtt ei tohi olla mitte riigi majandus- või tunnustushuvist tiivustatud, vaid peab lähtuma eetilis-moraalsetest kaalutlus-

test. See aga nõuab uute tehnoloogiate võimaluste ja kitsaskohtade analüüsi ning vajalike meetmete rakendamist, nii ühiskondlikus kui ka tehnoloogilises kontekstis, süsteemi jätkusuutlikkuse tagamiseks.

Uurimus toob ka välja, et hetkel käsitletakse tehisintellekti süsteeme ja nende kasutuselevõttu ennekõike haridus-, teadus- ja majandusinnovatsioonina, mitte kui protsessi, mis mõjutab kõikide ühiskonna liikmete elu ja sektorite tööd ning riigi turvalisust tervikuna. Tehisintellekti süsteem muudab ju nii organisatsiooni või teenust, kus see kasutusele võetakse, kui ka seda, kuidas ühiskond sellele tehnoloogiale vastama peab; sellega koos toimima hakkab. Käsitledes uut tehnoloogiat kõigest optimeeriva innovatsioonina, võib juhtuda, et selle vajalikkus ja sobitumine teiste teenustega, juba toimivate teenuste jätkusuutlik arendamine ning haldamine ja ka tekkivad uued probleemid ja turvariskid jäävad õigel ajal märkamata. St ennetamaks võimalikke uute tehnoloogiliste lahendustega kaasnevaid probleeme, tuleb Eestil edaspidi arendada tehnoloogiaalast kompetentsi lisaks küberturvalisusele ka teistes seotud valdkondades. Just ühiskonna teavitamisel ja harimisel on suur tähtsus, et uusi rakendusi kohusetundlikult kasutataks ning uute lahendustega kaasnevaid ohtusid mõistetaks; tehnoloogia toimimise õpetamine võimaldab kasvada põlvkonnal, kes on uute lahenduste arendamisel leidlikud ning kasutamisel enesekindlad. Uute tehnoloogiliste lahenduste kasutuselevõtu puhul nagu tehisintellekti süsteemid seda on, on tarvis kujundada kogu ühiskonna suhtumine tehnoloogiasse, et selle roll ühiskonnaprotsesside osana ja kasutusviisid oleksid üheti mõistetavad. Mõtestatud ja sujuv üleminek on siinkohal riigi võimuses.

Nii vajavad uurimistulemustest lähtuvalt ning ennetamaks võimalikke uusi väljakutseid, mis suurandmete analüüsil põhinevate rakendustega kaasneda võivad, edasist uurimist kaks omavahel seotud valdkonda: (i) kratilahenduste rakendamise vajalikkus ja võimalikkus riigi teenustes, kus kratilahenduste kasutuselevõtt peaks olema teerajaja või eeskuju. (Et uue tehnoloogia kasutuselevõtt on siseturvalisuse arengukavas märgitud, siis on üks potentsiaalne uurimisvaldkond just siseturvalisuse valdkonnas uute tehnoloogiliste lahenduste, k.a kratilahenduste, kasutamise võimalikkus ja vajalikkus.); (ii) võimalikest kratilahendustest tingitud muutused ja arengud ühiskonnas ning tekkivad uued ülesanded, k.a siseturvalisuse valdkonnas.

KASUTATUD KIRJANDUS

Blank, G. & Dutton, W. H., 2014. Next Generation Users: A New Digital Divide. Rmt: M. Graham & W. H. Dutton, toim-d. *Society & the Internet. How Networks of Information and Communication are Changing Our Lives.* Oxford: Oxford University Press, p. 36–52.

Bundeskriminalamt, 2020. Organised Crime – National Situation Report 2018. [võrguteavik] Leitav: https://www.bka.de/SharedDocs/Downloads/EN/Publications/AnnualReportsAndSituationAssessments/OrganisedCrime/organisedCrimeSituationReport2018.html;jsessionid=A4BB945212C9EF8A053856A3723FA529.live2301 [kasutatud: 6.04.2020].

Chapple, M., 2020. *What is metadata? Lifewire* [võrguteavik] Leitav: https://www.lifewire.com/metadata-definition-and-examples-1019177 [Kasutatud: 5.02.2020].

De Saulles, M., 2019. Internet of Things Statistics. Informationmatters: data-driven innovation news & analysis. [võrguteavik] Leitav: https://informationmatters.net/internet-of-things-statistics/ [Kasutatud: 5.02.2020].

Dewer, D. & Miladinova, V., 2017. The Big Data Challenge: Ompact and opportunity of large quantities of information Under the Europol Regulation. *Computer Law & Security Review.* 33, pp. 298–308.

Drechsler, W., 2018. Pathfinder: e-Estoania as the β-version. *JeDEM,* 10(2), pp. 1–22.

e-Estonia, 2020. Data embassy. [võrgumaterjal] Leitav: https://e-estonia.com/solutions/e-governance/data-embassy/ [Kasutatud: 6.04.2020].

e-Estonia, 2018. What we learned from the eID card security risk? [võrgumaterjal] Leitav: https://e-estonia.com/card-security-risk/ [Kasutatud: 3.06.2020].

Eliko, 2020. *Kalaranna. SmartStreet* [võrguteavik] Leitav: http://www.eliko.ee/smartstreet/ [Kasutatud: 6.04.2020].

Euroopa Komisjon, 2018. *Lisa järgmise dokumendi juurde: Komisjoni teatis Euroopa parlamendile, Euroopa Ülemkogule, Nõukogule, Euroopa majandus- ja sotsiaalkomiteele ning regioonide komiteele. Tehisintellekti käsitlev kooskõlastatud kava*. [võrguteavik] Leitav: https://ec.europa.eu/transparency/regdoc/rep/1/2018/ET/COM-2018-795-F1-ET-MAIN-PART-1.PDF [Kasutatud: 5.02.2020].

Euroopa Parlamendi ja Nõukogu määrus (2016) (EL) 2016/679.

ERR uudised, 2020. Estonian coronavirus hackathon transforms into global movement [võrguteavik] Leitav: https://news.err.ee/1072355/estonian-coronavirus-hackathon-transforms-into-global-movement [kasutatud: 16.05.2020].

Freedom House, 2020. Freedom on the Net 2019 — Estonia Country Report. [võrgumaterjal] Leitav: https://freedomhouse.org/country/estonia/freedom-net/2019 [Kasutatud: 6.04.2020].

Griffard, M., 2019. A bias-free predictive policing tool?: An evaluation of the NYPD's Patternizr. Fordham Urban Law Journal, 47(1), pp. 43–83.

Hanson Robotics Ltd., 2020. *Sophia.* [võrguteavik] Leitav: https://www.hansonrobotics.com/sophia/ [Kasutatud: 6.04.2020].

Heldna, E., 2016. Julgeolekuasutuste kogutud informatsiooni kasutamine kriminaalmenetlustes ja jagamine uurimisasutustes. *Juridica.* XXIV (10), lk 718–726.

Hädaolukorra seadus (2017) RT I, 03.03.2017, 1.

Internet World Statistics, 2020. [Võrgumaterjal] Leitav: https://internetworldstats.com/stats.htm [Kasutatud: 6.04.2020].

Isikukoodide moodustamise ja andmise kord (2017) RT I, 22.12.2017, 17.

Isikut tõendavate dokumentide seadus (2000) RT I, 31.01.2020, 14.

Justiitsministeerium, 2019a. Eelnõu: Kriminaalpoliitika põhialused aastani 2030. [võrguteavik] Leitav: https://www.kriminaalpoliitika.ee/et/kriminaalpoliitika-arengusuunad/kriminaalpoliitika-pohialused-aastani-2030-eelnou [Kasutatud: 6.04.2020].

Justiitsministeerium, 2019b. Seletuskiri Riigikogu otsuse „Kriminaalpoliitika põhialused aastani 2030“ eelnõu juurde. [võrguteavik] Leitav: https://www.kriminaalpoliitika.ee/et/kriminaalpoliitika-arengusuunad/kriminaalpoliitika-pohialused-aastani-2030-eelnou [Kasutatud: 6.04.2020].

Kaska. K., Bechvard, H. & Minarik,T. 2019. Huawei, 5G, and China as a Security Threat. [võrguteavik] Leitav: https://ccdcoe.org/uploads/2019/03/CCDCOE-Huawei-2019-03-28-FINAL.pdf [Kasutatud: 27.05.2020].

Khandelwal, S. 2017. Serious Crypto-Flaw Lets Hackers Recover Private RSA Keys Used in Billions of Devices. [võrguteavik] Leitav: https://thehackernews.com/2017/10/rsa-encryption-keys.html [Kasutatud: 3.06.2020].

Kirkpartick, K., 2017. It's Not the Algorithm, It's the Data. *Communications of the ACM.* 60 (2). p. 21–23.

Kitchin, R., 2014. Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, p. 1–12.

Kitchin, R. & Dodge, M., 2019. The (in)security of smart cities: vulnerabilities, risks, mitigation and prevention. *Journal of Urban Technologies.* 26, pp. 47–65.

Kivistik, O., 2018. Digitaalse menetluse tulevik. Prokuratuuri aastaraamat 2018. [võrgumaterjal] Leitav: https://www.prokuratuur.ee/et/digitaalse-menetluse-tulevik [Kasutatud: 6.04.2020].

Koort, E. 2017. E-resistentsus annab tunda ehk kui palju kulub e-riigi kaitsele. *Postimees.* [Võrgumaterjal *Postimees*: 3.12.2019] Leitav: https://leht.postimees.ee/6840537/erkki-koort-e-resistentsus-annab-tunda-ehk-kui-palju-kulub-e-riigi-kaitsele [Kasutatud: 3.06.2020].

Kotka, T., 2014. Ehitame riigi, mis mahub igaühele taskusse. Äripäev. [võrguteavik] Leitav: https://www.aripaev.ee/uudised/2014/04/10/kotka-ehitame-riigi-mis-mahub-igauhele-taskusse [Kasutatud: 5.02.2020].

Kurbalija, J. 2017. *The impact of (big) data on geopolitics, negotiations, and the diplomatic modus operandi*. [võrgumaterjal] Leitav: https://www.diplomacy.edu/blog/impact-big-data-geopolitics-negotiations-and-diplomatic-modus-operandi [Kasutatud: 6.04.2020].

Laaring, M., 2015. *Eesti korrakaitseõigus ohuennetusõigusena.* Tartu: Tartu Ülikooli Kirjastus.

Lee, I. & Estivill-Castro, V., 2011. Exploration of massive crime data sets through data mining techniques. *Applied Artificial Intelligence.* 25, pp. 362–379.

Levira, 2020. Asjade internet. [võrgumaterjal] Leitav: https://levira.com/teenused/asjade-internet/#iot-tehnoloogiad [Kasutatud: 5.02.2020].

Lyons, K., 2020. Clearview AI's client list includes 2,200 organizations spanning law enforcement to universities. The Verge. [võrgumaterjal] Leitav: https://www.theverge.com/2020/2/27/21156678/clearview-ai-client-macy-fbi-doj-twitter-facebook-youtube [Kasutatud: 6.04.2020].

Lõhmus, U., 2016. Elektroonilise side andmete säilimise saaga sai lahenduse, Eestis siiski veel mitte. *Juridica.* XXIV (10), lk 698–708.

Majandus- ja Kommunikatsiooniministeerium ja Riigikantselei, 2019. *Eesti Tehisintellekti kasutuselevõtu ekspertrühma aruanne.* [võrgumaterjal] Leitav: https://www.riigikantselei.ee/sites/default/files/riigikantselei/strateegiaburoo/eesti_tehisintellekti_kasutuselevotu_eksperdiruhma_aruanne.pdf [Kasutatud: 6.04.2020].

Majandus- ja Kommunikatsiooniministeerium, 2019a. Küberturvalisuse strateegia. [võrguteavik] Leitav: https://www.mkm.ee/sites/default/files/kuberturvalisuse_strateegia_2019-2022.pdf.

Majandus- ja kommunikatsiooniministeerium, 2019b. *kratid.ee* [võrguteavik] Leitav: https://www.kratid.ee/mis-on-kratt [Kasutatud. 6.04.2020].

Majandus- ja Kommunikatsiooniministeerium, 2018. *Eesti infoühiskonna arengukava 2020. Uuendatud 2018.* [võrguteavik] Leitav: https://www.mkm.ee/sites/default/files/eesti_infouhiskonna_arengukava_2020.pdf [Kasutatud. 6.04.2020].

Majandus- ja Kommunikatsiooniministeerium, 2013. *Eesti infoühiskonna arengukava 2020. VK nr 509.* [võrguteavik] Leitav: https://www.valitsus.ee/sites/default/files/content-editors/arengukavad/eesti_infouhiskonna_arengukava_2020_0.pdf [Kasutatud. 6.04.2020].

Maran, K. ja Seppel, I. 2020. [võrguteavik] Eesti e-riik valmistub WHO kaudu maailma vallutama. Leitav: https://leht.postimees.ee/6973751/eesti-e-riik-valmistub-who-kaudu-maailma-vallutama [Kasutatud 16.05.2020].

McCullagh, D., 2004. Database Nation. *Reason.* 36 (2), pp. 26–35.

Mehozay, Y. & Fisher, E., 2019. The epistemology of algoritmic risk assesmentand the pathway towards a *non-penology penology. Punishemtn & Society, 21 (5),* p. 523–541.

Morozov, E., 2018. Digikapitalismi tulevik. *Vikerkaar*, 5, lk 50–59.

National Crime Agency, 2019. *National Strategic Assessment of Serious and Organised Crime.* [võrguteavik] Leitav: https://nationalcrimeagency.gov.uk/who-we-are/publications/296-national-strategic-assessment-of-serious-organised-crime-2019/file [Kasutatud: 6.04.2020].

NIS COOPERATION GROUP. 2019. EU Coordinated risk assessment of the cybersecurity of the 5G networks. [võrguteavik] Leitav: https://mkm.ee/sites/default/files/content-editors/failid/E_riik/eu_coordinated_risk_assessment.pdf [Kasutatud: 27.05.2020].

O'Neill, P. H., 2020, Apple and Google are building coronavirus tracking into iOS and Android [võrguteavik] Leitav: https://www.technologyreview.com/2020/04/10/999213/apple-and-google-are-building-coronavirus-tracking-into-ios-and-android/ [Kasutatud: 17.05.2020].

O'Neill, P.H. Ryan,-Mosely, T. & Johnson, B., 2020. A flood of coronavirus apps are tracking us. Now it's time to keep track of them. [võrguteavik] Leitav: https://www.technologyreview.com/2020/05/07/1000961/launching-mittr-covid-tracing-tracker/ [kasutatud: 17.05.2020].

Pernik, P., 2019. Cybersecirty Education in Estonia: Building Competences for Internal Secity Personnel. Estoanin Academy of Security Sciences, Proceedings. 18, lk 71–108.

Riigi Infosüsteemi Amet, 2019a. Küberturvalisus 2019. [võrguteavik] Leitav: https://www.ria.ee/sites/default/files/content-editors/kuberturve/kuberturvalisus-2019.pdf [Kasutatud: 6.04.2020].

Riigi Infosüsteemi Amet, 2019b. Küpsised 2019. [võrguteavik] Leitav: https://www.ria.ee/et/ametist/kupsised.html [Kasutatud: 6.04.2020].

Riigi Infosüsteemi Amet, 2020a. Cybersecurity in Estonia 2020. [võrguteavik] Leitav: https://www.ria.ee/sites/default/files/cyber_aastaraamat_eng_web_2020.pdf [Kasutatud: 3.06.2020].

Riigi Infosüsteemi Amet, 2020b. Riigi Infosüsteemi Ameti Aastaraamat. [võrguteavik] Leitav: https://www.ria.ee/sites/default/files/ria_aastaraamat_2020_48lk_est_veeb.pdf [Kasutatud: 3.06.2020].

Sikkut, S., Velsberg, O. ja Vaher, K., 2020. *#Bürokratt: digiriigi järgmine arenguaste e-Eestis. Visioon ja kontseptsioon.* [võrguteavik] Leitav: https://f98cc689-5814-47ec-86b3-db505a7c3978.filesusr.com/ugd/7df26f_02237229bbea453d97161d109f1b31d2.pdf [Kasutatud: 6.04.2020).

Siseministeerium, 2020a. *RAND Europe küberkutitegevuse raporti valguses: võimekust tuleb*

tõsta. [võrguteavik] Leitav: https://www.siseministeerium.ee/et/uudised/rand-europe-kuberkuritegevuse-raporti-valguses-voimekust-vaja-tosta [kasutatud: 6.04.2020].

Siseministeerium, 2020b. Siseturvalisuse arengukava 2020–2030 kavand. [võrguteavik] Leitav: https://www.siseministeerium.ee/et/STAK2030 [Kasutatud: 6.04.2020].

Siseministeerium, 2019. Siseturvalisuse arengukava 2020–2030 koostamise ettepanek. VK nr 217. [võrguteavik] Leitav: https://www.valitsus.ee/sites/default/files/content-editors/arengukavad/stak_koostamise_ettepanek_09.2019.pdf [Kasutatud: 6.04.2020].

Siseministeerium, 2016. Täiendatud „Siseturvalisuse arengukava 2015–2020". VK nr 388. [võrguteavik] Leitav: https://www.valitsus.ee/sites/default/files/content-editors/arengukavad/taiendatud_siseturvalisuse_arengukava_2015-2020.pdf [Kasutatud: 6.04.2020].

Smith, K., 2019. *53 Incredible Facebook Statistics and Facts. Brandwatch.* [võrguteavik] Leitav: https://www.brandwatch.com/blog/facebook-statistics/ [Kasutatud: 6.04.2020].

Statistikaamet, 2020. Internetiühendus leibkondades. [võrgumaterjal] Leitav: https://www.stat.ee/29991 [Kasutatud: 5.02.2020].

Statistikaamet, 2019. Internetti kasutatakse üha enam ostlemiseks. [võrgumaterjal] Leitav: https://www.stat.ee/internetti-kasutatakse-uha-enam-ostlemiseks [Kasutatud: 5.02.2020].

Särekanno, U. 2020. *EU-Lisa. Ettekanne.* Tallinn, EU-Lisa 7.02.2020. kohtumine: EU-LISA meeting with the Heads of Diplomatic Missions and Senior Estonian Officials.

Sõltumatu kõrgetasemeline tehisintellekti eksperdirühm; Sidevõrkude, sisu ja tehnoloogia peadirektoraat (Euroopa Komisjon), 2019. *Eetikasuunised usaldusväärse Tehisintellekti arendamiseks.* [võrguteavik] Leitav: https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2019/11-06/Ethics-guidelines-AI_ET.pdf [kasutatud: 6.04.2020].

Tam, S.-M. & Kim,J-K., 2018. Big Data ethics and selection-bias: An official statistician's perspective. *Statistical Journal of the IAOS 34.* pp. 577–588.

Turk, K. ja Pild, M., 2019. Kratiga või kratita – see on küsimus. Robotitest ja tehisintellektist tsiviilõiguslikult. *Juridica,* I, lk 43–55.

Vaarik, D., 2015. Where Stuff Happens First. White Paper on Estonia's Digital Ideology. [*Presentatsioon. Majandus-ja Kommunikatsiooniministeeriumi veebileht.*] Leitav: https://www.mkm.ee/sites/default/files/digitalideology_final.pdf. [Kasutatud. 6.04.2020].

Vaks, T. 2019. Toomas Vaks: kas ja kes peaks kartma küberohtu? *Postimees.*[võrguteavik. Leitav: https://leht.postimees.ee/6840530/toomas-vaks-kas-ja-kes-peaks-kartma-kuberohtu?_ga=2.61368044.1947091724.1590355982-884318551.1563271269 [Kasutatud: 3.06.2020].

Veistennson, K., 2020. US Government Challenges Apple on Encryption (Again).[võrguteavik] Leitav: https://www.hrw.org/news/2020/01/16/us-government-challenges-apple-encryption-again [Kasutatud: 3.06.2020].

Välisministeerium, 2019. Välisministeeriumisse luuakse uus küberdiplomaatia osakond. [võrgumaterjal] Leitav: https://vm.ee/et/uudised/valisministeeriumisse-luuakse-uus-kuberdiplomaatia-osakond [Kasutatud: 3.06.2020].

WhatsApp, 2020. [võrgumaterjal] Leitav: https://faq.whatsapp.com/general/security-and-privacy/information-for-law-enforcement-authorities [Kasutatud: 5.06.2020].

Zwitter, A., 2014. Big Data ethics. *Big Data & Society.* pp. 1–6.

1.1. Riigile kättesaadavad suurandmed ja riigi suurandmed

KESKNE TEEMA	KAASNEVAD KÜSIMUSED	KÜSIMUSE OLEMUS	ENNETAMISEKS VAJALIK	RISKID
ANDMETE KOGUMINE	Kogumise õiguspärasus – kuidas tohib andmeid koguda?	• Missuguste andmete kogumine on lubatud? • Millisel viisil on andmete kogumine lubatud? • Kes tagab kogumise õiguspärasuse?	• Seadusandlus, mis määrab kogumise viisid. • Mehhanism, mis kontrollib kogumise seaduspärasust.	• Võimalik risk eksida andmete kogumisel kehtestatud seadusandluse vastu.
	Kogumise meetod – kuidas käituda andmete paljususega?	• Kas koguma peab ainult vajalikke andmeid? • Kas koguma peab kõikvõimalikke andmeid, k?a metaandmeid; nutistu toodetavaid andmeid; interneti teenuste vahendusel tekkivaid andmeid? • Kes otsustab andmete kogumise meetodi?	• Ühtlustatud andmepoliitika, k.a » visioon vajaminevatest andmetest, » visioon tulevikuteenustest. • Mehhanism, mis tagab andmete kogumise.	• Võimalik risk ebavajalike andmetega koormata süsteemi ja kasutada riigi ressurssi ebaotstarbekalt.
	Kogumise vajadus – mida on vaja koguda?	• Kas koguma peab andmeid, mis on vaja vaid hetkel kasutusel olevate teenuste toimimiseks? • Kas koguma peab andmeid, mis on vajalikud ka võimalike tulevikuteenuste toimimiseks? • Kes vastutab andmete kogumise eest?	• Ühtlustatud andmepoliitika, k.a » Visioon vajaminevatest andmetest. » Visioon tulevikuteenustest. • Mehhanism, mis tagab andmete kogumise.	• Võimalik risk tulevikus mitte omada vajalikke andmeid, et teenuseid pakkuda ja arendada.
ANDMETE TALLETAMINE	Talletamise õiguspärasus – kuidas tohib andmeid talletada?	• Missuguste andmete talletamine on lubatud? • Millisel viisil on andmete talletamine lubatud? • Kes tagab talletamise õiguspärasuse?	• Seadusandlus, mis määrab andmete talletamise aja ja meetodi. • Mehhanism, mis kontrollib andmete talletamise seaduspärasust.	• Võimalik risk eksida andmete talletamisele kehtestatud seadusandluse vastu.

KESKNE TEEMA	KAASNEVAD KÜSIMUSED	KÜSIMUSE OLEMUS	ENNETAMISEKS VAJALIK	RISKID
ANDMETE TALLETAMINE	Talletamise võimalikkus – kuhu on võimalik andmeid talletada?	• Missugused võimalused andmete talletamiseks on? • Kes hindab andmete talletamiseks sobilikku kohta?	• Ühtlustatud andmepoliitika. • Mehhanism, mis andmete talletamise viise ja võimalusi riigis ja välisriigis hindab ja kasutamise otsustab.	• Võimalikud füüsilised ja virtuaalsed riskid andmete säilimisele; andmete ligipääsetavuse säilimisele.
	Talletamise jätkusuutlikkus – kuidas tagada andmete jätkusuutlik talletamine?	• Kui kaua on vaja andmeid talletada? • Missuguses mahus on andmeid vaja talletada? • Kuidas ebavajalikke andmeid kustutada/arhiveerida? • Kes kontrollib jätkusuutliku talletamise protsessi?	• Ühtlustatud andmepoliitika, k.a » visioon vajaminevatest andmetest, » visioon tulevikuteenustest. • Talletamiseks vajaliku ressursi olemasolu. • Mehhanism, mis jätkusuutlikku talletamist haldab.	• Võimalik risk tulevikus mitte omada vajalikke andmeid, et teenuseid pakkuda ja arendada. • Võimalik risk ebavajalike andmetega koormata süsteemi ja kasutada riigi ressurssi ebaotstarbekalt.
	Talletamise turvalisus – kuidas tagada andmete turvaline talletamine?	• Missugused riskid kaasnevad andmete talletamisega? • Kes vastutab riskide hindamise ja hajutamise eest ning tegeleb probleemide tekkimisel nende lahendamisega?	• Ühtlustatud andmepoliitika. • Teadlikkus turvariskidest. • Teadlikkus riskide arengust/muutumisest. • Mehhanism, mis turvalisuse tagab.	• Võimalikud virtuaalsed ja füüsilised riskid andmete säilimisele; andmete ligipääsetavuse säilimisele.

KESKNE TEEMA	KAASNEVAD KÜSIMUSED	KÜSIMUSE OLEMUS	ENNETAMISEKS VAJALIK	RISKID
ANDMETE KASUTAMINE	Kasutamise õiguspärasus – kuidas tohib andmeid kasutada?	• Missuguste andmete kasutamine on lubatud? • Millistel tingimustel on andmete kasutamine lubatud? • Kellele on andmete kasutamine lubatud? • Kes kontrollib andmete seaduspärast kasutamist?	• Ühtlustatud andmepoliitika, k.a » visioon vajaminevatest andmetest, » visioon tulevikuteenustest, » visioon riikidevahelisest koostööst ja vajalikest andmetest. • Andmete kasutamist lubava mehhanismi olemasolu. • Andmete kasutamist lubava mehhanismi paindlikkus.	• Võimalik risk eksida andmete kasutamiseks kehtestatud seadusandluse vastu. • Vajalike volituste ületamine või puudumine.
	Kasutamises erandite tegemine – kellel on õigus andmete kasutamise viisi muuta?	• Missuguses olukorras võib andmete kasutamise viisi muutmine olla (häda)vajalik? • Kellel on õigus ja vastutus andmekasutuse üle otsustada?	• Ühtlustatud andmepoliitika, k.a » visioon andmete potentsiaalist, » visioon võimalikest olukordadest, kus ligipääsu peab tagama. • Mehhanism, mis kontrollib andmekasutamise seaduspärasust k.a volitatud isikute käitumise seaduspärasust.	• Puudub võimlaus (häda)vajalikus olukorras muutusi teha. • Vajalike volituste ületamine või puudumine.
	Kasutamise vajalikkus – missuguseid andmeid tuleks kasutada?	• Missuguste andmete kasutamine on vajalik? • Kes otsustab andmete kasutamise vajalikkuse üle?	• Ühtlustatud andmepoliitika, k.a » teadlikkus tehnoloogia võimekusest, » visioon vajaminevatest andmetest, » visioon tulevikuteenustest. • Mehhanism, mis kasutamise vajalikkust hindab.	• Võimalik risk koormata süsteemi ebavajalike teenustega ja kasutada riigi ressurssi ebaotstarbekalt.

KESKNE TEEMA	KAASNEVAD KÜSIMUSED	KÜSIMUSE OLEMUS	ENNETAMISEKS VAJALIK	RISKID
ANDMETE KASUTAMINE	Kasutamise viis – kuidas tuleks andmeid kasutada?	• Millistel juhtudel ja kuidas tuleks andmeid kasutada? • Kes otsustab andmete kasutamise viisi üle?	• Ühtlustatud andmepoliitika, k.a » teadlikkus tehnoloogia võimekusest, » visioon tulevikuteenustest. • Teadlikkus analüüsi võimalustest ja piiratusest. • Mehhanism, mis kasutamise viisi hindab.	• Võimalik risk ebavajalike teenustega koormata süsteemi ja kasutada ebaotstarbekalt riigi ressurssi, ka inimressurssi.
	Andmete kasutamine uutes rakendustes – kuidas riigi arendatavad/hallatavad teenused andmeid kasutavad/kasutada saaksid?	• Millistel juhtudel tuleks andmeid kasutades uus tehnoloogiline lahendus arendada? • Millistel juhtudel tuleks andmeid kasutades juba toimivat tehnoloogilist lahendust uuendada/arendada? • Kes otsustab uue ja/või toimiva arendamise üle?	• Ühtlustatud andmepoliitika, k.a » teadlikkus tehnoloogia võimekusest, » visioon vajaminevatest andmetest, » visioon tulevikuteenustest. • Mehhanism, mis jätkusuutlikku kasutamist, toimimist ja arendamist haldab.	• Võimalik risk ebavajalike teenustega koormata süsteemi ja kasutada riigi ressurssi, ka inimressurssi, ebaotstarbekalt.
	Andmete kasutamine uutes rakendustes – kuidas erasektori arendatavad teenused riigi hallatavaid andmeid kasutada saavad/võivad? (Ehk riik kui erasektori teenuse sisse ostja.)	• Millistel juhtudel tuleks rakendus tellida/sisse osta erasektori ettevõttelt? • Kes otsustab erasektori teenuste puhul riigi hallatavate andmete kasutamise vajalikkuse üle?	• Ühtlustatud andmepoliitika, k .a » teadlikkus tehnoloogia võimekusest, » teadlikkus erasektori võimekusest. • Mehhanism, mis otsustab riigi sisse ostetavate erateenuste vajalikkuse/andmete kasutamise vajalikkuse; kontrollib teenuste toimimist.	• Võimalik risk kasutada riigi ressurssi ebaotstarbekalt. • Võimalik risk sõltuda riigile vajaliku teenuse pakkumisel erasektorist.

2.1. Suurandmed, millele ligipääsu võib riigil vaja minna k.a interneti teenuste ja nutistu andmed

KESKNE TEEMA	KAASNEVAD KÜSIMUSED	KÜSIMUSE OLEMUS	ENNETAMISEKS VAJALIK	RISKID
ANDMETE KOGUMINE	Kogumise õiguspärasus – kuidas tohib andmeid koguda?	• Missuguste andmete kogumine on lubatud ja võimalik? • Millisel viisil on andmete kogumine lubatud? • Kes tagab kogumise õiguspärasuse?	• Seadusandlus, mis määrab kogumise viisid. • Riigi väljatöötatud metodoloogia vajalike andmete seaduspäraseks kogumiseks. • Visioon riikidevahelisest koostööst ja vajalikest andmetest. • Mehhanism, mis seaduspärasust kontrollib.	• Võimalik risk eksida andmete kogumisel kehtestatud seadusandluse vastu. • Vajaliku rahvusvahelise koostöö puudumine.
ANDMETE KASUTAMINE	Kasutamise õiguspärasus – kuidas tohib andmeid kasutada?	• Missuguste andmete kasutamine on lubatud? • Millistel tingimustel on andmete kasutamine lubatud? • Kellele on andmete kasutamine lubatud? • Kes tagab kontrolli õiguspärase kasutamise üle?	• Andmete kasutamist lubava mehhanismi olemasolu. • Andmete kasutamist lubava mehhanismi paindlikkus. • Riigi väljatöötatud metodoloogia vajalike andmete seaduspäraseks kasutamiseks. • Visioon riikidevahelisest koostööst ja vajalikest andmetest. • Mehhanism, mis kasutust kontrollib.	• Võimalik risk eksida andmete kasutamiseks kehtestatud seadusandluse vastu. • Vajalike volituste ületamine või puudumine. • Vajaliku rahvusvahelise koostöö puudumine.
	Kasutamises erandite tegemine – kellel on õigus andmete kasutamise viisi muuta?	• Missuguses olukorras võib andmete kasutamise viisi muutmine olla (häda)vajalik? • Kellel on õigus ja vastutus andmekasutuse üle otsustada?	• Visioon andmete potentsiaalist. • Visioon võimalikest olukordadest, kus ligipääsu peab tagama. • Visioon riikidevahelisest koostööst ja vajalikest andmetest. • Mehhanism, mis kontrollib andmekasutamise seaduspärasust k.a volitatud isikute käitumise seaduspärasust.	• Puudub võimlaus (häda)vajalikus olukorras muutusi teha. • Vajalike volituste ületamine või puudumine. • Vajaliku rahvusvahelise koostöö puudumine.

KÄESOLEVA UURIMUSE EESMÄRK ON SELGITADA LÄHEMALT SUURANDMETE OLEMUST.

Andmete eriilmelisusest lähtuvalt kaardistab uurimus samuti võimalikud kitsaskohad andmeanalüüsil tuginevate rakenduste arendamiseks ja kasutuselevõtuks juba digitaliseeritud ning koostoimivaid teenuseid omavas Eestis.

Esmalt selgitab uurimus suurandmete potentsiaali ilmnemist maailma ja Eesti tehnoloogia arengu valguses. Teiseks selgitab uurimus suurandmete mitmekülgset olemust ning sellest tulenevat kasutamise ja koostoime keerukust. Kolmandaks avatakse raportis, milliseid andmeid saab Eesti kontekstis lugeda suurandmeteks ning miks on need teiste riikide suurandmetega võrreldes mõneti erinevad. Viimases alapeatükis avatakse kitsaskohti, mis võivad suurandmete rakendamisel ja teenuste arendamisel Eesti jaoks murekohaks saada.

sisekaitse.ee