

Kas andmetest kujuneb uus varaklass?

Internetti ühendatud seadmete arvu hüppeline kasv ning üha suurenenud digitaalsete lahenduste kasutamine majanduses on tekitanud olukorra, kus igapäevaselt luuakse aina enam virtuaalseid andmeid. Hinnanguliselt kahekordistub andmete hulk kogu maailmas iga kahe aastaga. Andmesalvestuse odavnemine, internetivõimsuse kasv, analüütiliste vahendite areng ning pilvetehnoloogia teke on tinginud selle, et tänapäeval on võimalik oluliselt rohkem andmeid talletada, edastada ja töödelda ning teha seda odavamalt, kiiremini ja efektiivsemalt kui kunagi varem. Siiski ei leia valdav osa andmetest rakendust, mistõttu jääb suurandmetest saadav potentsiaalne kasu praegu veel marginaalseks.

Mis on suurandmed?

Suurandmed ehk *Big data* on oma definitsiooni kohaselt suured ja keerulised andmekogumid, mida on olemasolevate infotehnoloogiliste rakendustega keeruline analüüsida. Suurandmete puhul on tegu suurte andme- mahtudega, mis on heterogeensed ehk mitmekesised (andmed kogutakse

mitmest allikast ja nad esinevad erineval kujul, sh videosalvestused, e-kirjad, tekst, numbrid jne) ning mida kogutakse ja analüüsitakse reaalsajas (või võimalikult selle lähedal). Mõnede allikate kohaselt on suurandmete üks iseloomulik omadus ka tõenduspõhisus (4).

Suurandmete analüüsimeetodid (*data analytics*) võimaldavad analüüsida kasvavat hulka struktureerimata andmeid, mis on erinevates keeltes ja formaatides ning kogutud erinevatest allikatest. Suurandmete analüüsi keskne idee on leida algandmete seast tähendust omavaid mustreid ja korrelatsioone, kasutamaks saadud tulemusi paremate otsuste langetamiseks, produktiivsuse tõstmiseks ja innovatsiooni loomiseks. Suurandmete analüüs hõlmab endas enamjaolt kogutud andmetele suure hulga algoritmide rakendamist tuvastamiseks neis andmetes seaduspärasusi. See erineb oluliselt traditsioonilisest andmeanalüüsist, kus fookus on peamiselt hüpoteesi testimisel. Suurandmete analüüsi eeliseks on eelkõige võimalus teostada analüüsi reaalsajas ehk jooksvalt, samal ajal kui andmeid luuakse kogu aeg juurde (5).

Ligipääs üha suurenevatele andmekogudele on saanud võimalikuks eeskätt tänu digitaliseerimisele (võimekus koguda ja analüüsida üha rohkem andmeid digitaalselt, nutikad tarkvaralised lahendused, mobiilirakendused jms), andmekeskuste (kiire andmemahtude hoiustamine, pilveteenused) ja interneti levikule (ligipääs andmetele sõltumata geograafilisest asukohast, andmete ülekanded reaalsajas jms).

Millised on suurandmete kasutusvõimalused?

(18). Suurandmed annavad meile senisest detailsemat informatsiooni näiteks turunduse ja finantside valdkonnast, olulist teavet strateegiate paikapanemiseks ja otsuste langetamiseks ning tervikuna paremaid andmeid majanduse ja ühiskonna toimimise kohta. 2012. aastal läbi viidud rahvusvaheliste ettevõtete juhtide küsitlusest selgus, et üheksa juhti kümnest peab andmeid vähemalt sama oluliseks ressursiks, kui seda on füüsiline vara, tööjõud ja kapital (4).

OECD uuringu tulemuste kohaselt saavutavad ettevõtted, kes oma ärioluliste otsuste tegemisel kasutavad suurandmete analüüsi, 5–10% suurema produktiivsuse tõusu võrrelduna nende ettevõtetega, kes seda ei tee. McKinsey Global Institute'i hinnangul võimaldaks täiemahuline andmeanalüüsi rakendamine vähendada Euroopa 23 suuremas riigis halduskulusid 15–20% võrra (5).

Tänapäeval kasutatakse suurandmetel põhinevat andmeanalüüsi peamiselt järgmistes valdkondades:

Valdkond	Kasutusvõimalus
Kindlustus	Riskianalüüs ja kindlustuspettuste ennetamine (9).
Meditiin	Suuremate valimite analüüs, mis võimaldab edukamalt sümptomitele vastavat ravi leida ja haigusi ennetada (10).
Turism	Turismimahtude ennustamine ja personaliseeritud turismitoodete turustamine (11). Mobiiliandmetepõhiste statistikate ja analüüside koostamine sise- ja väliturismi uurimise tarbeks (7).
Linnaplaneerimine	Transpordivõrkude tõhustamine ja energiatarbimise vähendamine (12).
Tööstus	Tootmisahelate optimeerimine (13).
Pangandus ja finantssektor	Uued võimalused krediitdireitingute loomiseks, teiste seas ka sotsiaalmeedia andmete toel (6). Uued meetodid finantsanalüüsis (14).
Turundus ja kaubandus	Klientide profileerimine – personaliseeritud pakkumiste koostamine vastavalt klientide kohta kogutud andmetele (15).
Avalik sektor	Avalike teenuste pakkumise tõhustamine analüüsi toel, maksupettuste tuvastamine, kriminaalstatistikal põhineva politseitöö tugevdamine (16).

Suurandmete analüüs on leidnud tänaseks rakendust väga paljudes erinevates valdkondades. Nii näiteks on rakendatud suurandmete analüüsi USA tervisekindlustusametites võimalike **kindlustuspettuste** kontrollimise eesmärgil. See kindlustuspettuste ennetuse süsteem teenis juba esimese aasta jooksul kolmekordselt tasa oma maksumuse (4). **Jaekaubanduses** kasutatakse suurandmeid eelkõige kliendibaasi efektiivsemal profileerimisel ning seeläbi suunatud pakkumiste ja efektiivsema turunduse korraldamisel. Jaekaubanduses kasutab kliendikaartide andmete põhjal teostatud suurandmete analüüsi näiteks Suurbritannia supermarketite kett Tesco (5). Lisaks ettevõttesisesele kasutusele soosib suurandmete analüüsi ärilist kasutamist ka asjaolu, et ühe ettevõtte poolt kogutud andmed võivad osutada kasulikuks ka teise valdkonna ettevõttele. Nii näiteks kogutakse andmeid sotsiaalmeediaplatformidelt ning müüakse neid edasi **туру-uuringute** läbiviimiseks. Samuti võidakse sotsiaalmeediast saadud andmeid, juhul kui isik on sellest teadlik ja sellega nõus, kasutada krediitdireitingute hindamiseks (6).

Suurandmete analüüs on lootustandev meetod ka [meditsiini valdkonnas](#), võimaldades varasemast täpsemat haiguste uurimist, nakkushaiguste ennetamist ja patsiendi vajadustele vastavat parimat ravi. 2014. aastal avalikustatud Massachusetts Institute of Technology (MIT) ja Harvardi Ülikooli koostööprojekti kasutati skisofreenia uurimiseks suurandmete analüüsi, mis võimaldas leida skisofreenia tekke eelsoodumuseks olevaid geneetilisi mutatsioone. Uuringust selgus, et ka pärilike haiguste analüüs eeldab senisest oluliselt suuremaid valimeid ning tulemusteni jõudmine polnuks võimalik suurandmete analüüsi metoodikata ning võimekuseta töödelda niivõrd suurt andmekogu (4).

[Töötleva tööstuse](#) jaoks loovad suurandmed võimaluse optimeerida ettevõtte tootmisahelaid ja parandada sisekommunikatsiooni. Ühelt poolt lubavad suurandmed integreerida teavet teadus- ja arendusosakonna ning inseneride ning tootmisüksuse vahel, lühendades nii uute toodete väljaarendamise aega. Teisalt võimaldab erinevate sensorite kasutamine tootmisliinil koos suurandmete kogumise ja jooksva analüüsiga leevendada ebakindlust tootmisprotsessis, mis võib tuleneda komponentide kättesaadavuse probleemidest ning nende konstantse kvaliteedi puudumisest.

Muudes valdkondades leiab suurandmete analüüs aktiivset rakendust [liikluse](#) ning [nutikate transpordivõrkude](#) planeerimise valdkonnas, mille abil on võimalik vähendada ummikuid ning seeläbi ka õhusaastet ja kütusekulu. Lisaks nähakse suurandmete potentsiaali ka linnade energiatõhusamaks muutmisel tänu põhjalikumale tarbimismustrite analüüsile (3). Samuti võimaldab andmete kogumine ja andmeanalüüs tõsta ühiskonna turvalisust, näiteks hoiatavad sensorid saabuva tormi või veetaseme tõusu eest.

Kindlasti ei saa mööda vaadata ka [suurandmete julgeolekust](#) – tänapäeva kuritegevus on samuti kolinud küberruumi ning andme- ja identiteedivargused annavad selgelt mõista, et andmetel on väärtus. Suurimateks väljakutseteks on suurandmete logide kaitsmine häkkerite eest, reaaliajase kogutavate ja esitatavate andmete kaitse, aga ka andmetele ligipääsu ning andmeallikate kaitsmine. Iga andmebaasi omanik on kindlasti teadlik andmete väärtusest, kuid neid mitte kaitstes võib see väärtus kiiresti kahaneda.

Eesti on Euroopas esimene riik, kes on koostanud riiklikke statistilisi uuringuid, mis põhinevad suurandmetel. Nendeks on Eesti Panga ja Positium LBS koostöös teostatud üleriigilised turismistatistikad, mis on loodud mobiilpositsioneerimise andmeid rakendades. Suurandmete kasutamine turismistatistika loomisel võimaldab koguda andmeid odavamalt ja kiiremini ning saada terviklikuma pildi turistide liikumismustritest (7). Eestis tegutsevad suurandmetve analüüsi ning nende ärilisse konteksti viimisega teadaolevalt veel Nortal (8), Swedbank ja Telia.

Andmemajanduse väljakujunemine, andmed kui varaklass

Andmemajanduse all mõeldakse süsteemi erinevate turuosaliste, tootjate, teadlaste ning taristu haldajate vahel, kelle koostöö tagab andmete kättesaadavuse ning kasutatavuse. Turuosalised saavad andmetest kasu lõigata, kasutades neid erisuguste rakenduste tarbeks, mis aitavad erinevaid protsesse optimeerida (17). **Andmed** (juriidilises mõttes saab küll ainult andmebaase osta ja müüa) on kasvava väärtusega varaklass, mida on võimalik ka osta ja müüa, ehkki valdav osa väärtuse loomisest toimub tänapäeval läbi andmevaldaja enda, kes kujundab teenuseid toorandmete põhjal. Selle põhjuseks võivad olla isikuandmekaitsega seotud riskid ning ärisaladuse hoidmine, mistõttu eelistatakse hoida andmeid koduseinte vahel (5).

Tavaliselt on andmemajanduses kolm osapoolt: eraisik (või juriidiline isik või n-ö intelligentne masin), andmekoguja (sotsiaalmeediaplatform, mobiiloperaator, mobiilirakendus (GPS, telekom), riik jms) ja andmete kasutaja (pangad, ettevõtted, riik). Riiki kui seadusandjat võib käsitleda ka andmemajanduse neljanda osapoolena. Samas on tema roll piiratud seadusloomega, kui riik ise ei ole just andmete koguja rollis. Igal juhul on igal osapoolel oma roll (kuidas andmeid esitatakse, kogutakse, ligipääsu võimaldatakse, kaitsakse jms). Majanduslik kasutegur tekibki tavaliselt andmete kogumisel/omamisel (andmebaas kui müüdiv vara) ja andmete kasutamisel (turunduses, müügistrateegiate koostamisel, tarbijakäitumise analüüsimisel jms). Andmete salvestamist (*data storage*) ja andmeanalüüsi võib vaadelda kui andmemajanduse n-ö lisategevusi, mis võimaldavad andmeid efektiivsemalt analüüsida ja kasutada. Enamikul juhtudel on äriidee üles ehitatud nii, et see võimaldab ligipääsu andmebaasidele, mis teebki suurandmetest väärtusliku kauba – olles kord andmebaasi omanik, saab andmeid suhteliselt madalate kuludega korduvalt müüa. Ainuke tingimus on see, et andmed oleks uuendatud ning asjakohased (näiteks kui tarbijad ei kasuta enam Google'i otsimootorit või Facebooki, muutuvad ka viimaste poolt kogutud andmed väärtusetuks).

Lisaks nende põhjal pakutavatele teenustele võib suurandmetel ettevõttesiseselt olla väärtus lahenduste pakkujana nii kulude vähendamiseks kui ka tulude suurendamiseks. Näiteks võib tuua tootmisliinide optimeerimise, turvastandardite tõstmise, tootedefektide tuvastamise, uute toodete ja teenuste väljaarendamise ning ka vanade edasiarendamise. Siiski ennustatakse, et andmeturg avaneb senisest rohkem ning kõiki funktsioone ei täida enam üks ettevõtte iseseisvalt, vaid kujuneb välja andmemajanduse struktuur, kus erinevaid rolle täidavad andmete esitlejad, kes pakuvad tavaklientidele teenuseid, analüütikud ja statistikud, andmeplatvormide omanikud ning andmete haldajad, loojad ja omanikud (18). Praeguseks

eksisteerib üksikuid ettevõtteid, kelle põhitegevus on olla andmeturuks ehk osta ja müüa andmeid. Tuntuimaks näiteks on Qlik DataMarket.¹

Suurandmete käsitlemisel varaklassina on primaarne küsimus see, mis on andmete õigused, ese ja kohustused. Tänapäevane seadusandlus ei anna andmetele ei õiguste ega esemete omadusi ega luba neid seega varade hulka liigitada. Samas on diskussioon andmete liigitamise kohta varade hulka esile kerkimas (24). Suurandmete analüüs annab neid andmeid töötlevatele ettevõtetele võimaluse luua suuremat lisandväärtust, mis omakorda loob tugevama konkurentsieelise.⁽¹⁾ Ka andmeanalüüsi tulemuste osas käib vaidlus, kellele kuulub andmete töötlemisest saadav tulemus. Näiteks leiab Georgetowni ülikooli professor Jeffrey Ritter, et kuigi isikuandmeid saab igaüks olulisel määral kontrollida ning oma andmete kasutust piirata, siis nende andmete töötlemise tulemus – nt röntgenipilt või mõni muu mingi protsessi tulemusel tehtud salvestus võiks kuuluda sellele isikule/ühingule/organisatsioonile, kes töötlemist sooritas, nt haigla, uurimisasutus jne. Siin sisaldub omakorda risk andmemajanduse oligopolide tekkeks, mis tuleneb tänapäevastest populaarseimast internetiplatvormidest. Omades suurt osalust turul, on neil ettevõtetel või organisatsioonidel suur mõjujõud väiksemate konkurentide üle.

Tulenevalt asjaolust, et paljude andmete tootmise protsessis on oma digitaalse tegevusega kohal **indiviid** kui andmete looja, tekib küsimus, kuidas peaks antud inimene saama õiglase kompensatsiooni või tasu oma tegevuse eest. Seni on kehtivaks reeglits olnud see, et toimub vahetus, kus inimene kasutab tasuta tema kohta kogutavate andmetega platvormi (19). Andmed, mis on üksikisiku internetikasutuse põhjal kogutud, omavad praegu eraldi-seisvatena vaid marginaalset väärtust, ulatudes mitte üle 50 eurosendi ühe andmeühiku kohta. Inimene saab ise oma andmeid uuringute tarbeks müüa. See võimaldab andmekogujal hankida paremat infot ja inimesel saada rahalist kompensatsiooni, ehkki selle tulemusel peab ta oma privaatsusest veelgi enam loobuma (20).

1 Samas võib eeldada, et, näiteks nii Facebooki, Google'i, Waze'i, Amazoni kui ka teiste platvormimajanduse suurte tegijate üks põhiäri on kas otsene või kaudne andmete müük.

Millised on suurandmetega seonduvad väljakutsed?

Ehkki suur osa mõttekodasid ja eksperte on suurandmete analüüsi tuleviku osas optimistlikud, on valdkonnas ka ohtralt kitsaskohti ja probleeme, mida ei saa tähelepanuta jätta. Esmatähtsaks ja [suurimaks ohuks suurandmete kasutamisel peetakse privaatsuse ja isikuandmete kaitset](#). Kasvav andmete hulk tähendab ka suurenevat isikuandmete hulka ning nende aktiivsem kasutamine loob uusi riske. Inimeste kohta võidakse tihti koguda andmeid ka olukorras, kus nad ise sellest otseselt teadlikud ei ole. Isikuandmete aktiivsem kasutus analüüsi tarbeks eeldab ettevõtetelt teadlikumat ja ettevaatlikumat ümberkäimist delikaatsete isikuandmetega (1). Andmete kommertsialiseerumine tähendab kasvavat vajadust kaitsta isikuandmete anonüümsust, mis tuleneb andmete aktiivsemast ringlusest organisatsioonide vahel (3). Debateerimise teemaks on see, kas andmemajandust aitab edendada regulatsiooni tugevdamine või lõdvendamine. Euroopas on arutelu keskpunktiks 2018. aasta mais jõustuv ELi andmekaitseriform, mis esindab seisukohta, et isikuandmete ja privaatsuse kaitse eeldab eraettevõtluselt suuremat ettevaatust ja rangemaid eeskirju. Riskiks peetakse sedagi, et tarbijate kasvav mure oma isikuandmete kaitse pärast toob kaasa usalduse kaotuse ning seeläbi ärivõimaluste ja tulude vähenemise eraettevõtja jaoks (21). Regulatsiooni vähendamise pooldajad väidavad, et andmekaitseenormide rangemaks muutmine pärsib andmete vaba liikumist ning innovatsiooni.

Andmete vaba liikumise ja isikuandmete kaitse kõrval on teiseks suuremaks kitsaskohaks [valdkonda tundvate ekspertide ja analüütikute vähesus](#), kes omavad vastavaid oskusi ja kogemusi suurandmete analüüsimisel. Euroopas on tänapäeval hinnanguliselt kuus miljonit inimest, kelle töö seonduv andmete haldamise, töötlemise või analüüsiga. Aastaks 2020 ennustatakse nende inimeste arvu kasvu 7,1 miljonini. Siiski vaevab sektorit suur kvalifitseeritud tööjõu puudus ning pea nelisada tuhat töökohta on jäänud täitmata. Tulenevalt kasvavast tööjõuvajadusest andmeanalüüsi valdkonnas nõuab konkurentsivõime säilitamine suuremaid investeeringuid vastavasse haridusse, sealhulgas ümber- ja täiendõppesse. Lisainvesteeringuid vajaksid niisugused erialad nagu matemaatika, statistika ja infotehnoloogia. Suurandmete analüüs ei nõua ainult IT-töötajaid ja statistikke, vaid vajab ka tööjõudu, kes oskaks tulemusi tõlgendada ehk saadud mustrite põhjal seoseid leida, ning neid inimesi, kes mõistaks, kuidas tulemusi ärilisse konteksti panna (4). Tehniliseks ja juriidiliseks kitsaskohaks on [suurandmete kuuluvuse küsimus](#). Tihti ei kuulu suurandmed vaid ühele kindlale inimesele või organisatsioonile ning mõningatel juhtudel on andmete kuuluvust üldse keeruline välja selgitada. Juriidilise puuduse lahendamiseks on välja pakutud idee käsitleda andmeid kui kaasomandit, mille puhul oma

digitaalse tegevusega andmeid loovat inimest vaadeldakse kui andmete kaasomanikku (22). Selle idee eesmärk oleks tagada olukord, kus andmete haldamise ja töötlemisega tegeval organisatsioonil säiliks võimalused andmeid aktiivselt kasutada ja taaskasutada ning indiviidi jaoks valitseks senisest suurem läbipaistvus, mis võimaldaks näha, milliseid andmeid kogutakse ja milleks neid kasutatakse.

Veel üks juriidiline kitsaskoht on [andmete lokaliseerimine](#), mis seab geograafilised piirangud, mis määravad, millises riigis tohib konkreetseid andmeid töödelda ning säilitada. Andmete lokaliseerimine võib olla tingitud soovist kaitsta nii isikuandmeid kui ka ärisaladust, kuid see võib takistada näiteks isikustamata andmete, näiteks kliimaandmete levikut. Andmete lokaliseerimine võib seega pärssida suurandmete vaba liikumist ja jagamist. Teema kerkis üles eelkõige 2013. aastal ühenduses NSA pealtkuulamis-skandaaliga, millega seoses ilmnes, et USA välisluureteenistus kasutas andmete vaba liikumist Euroopa Liidu riikide kodanike jälgimiseks ning kuulus pealt poliitiliste liidrite telefonikõnesid, teiste seas Saksamaa liidukantsleri ning Prantsusmaa presidendi omi. Sellest tingituna on Saksamaa valitsus alustanud rangemate turvanõuete sisseviimist ning oma andmete lokaliseerimist, millega seoses liidetakse tihedalt kokku riigi IT-taristud ning soovitakse vähendada Saksamaa sõltuvust teiste riikide serveritest, kus andmed asuvad hoiustamise ja töötlemise eesmärgil. (23). Seda loetelu sobib jätkama ka Eesti e-residentsuse ja andmesaatkondade algatus. Viimasega tahetakse eeskätt kaitsta riigi julgeolekut – Eesti riigile olulised andmed dubleeritakse ja salvestatakse teises riigis (Luksemburgis) asuvas serveris, mis ohu korral tagab Eesti riigi jätkumise *de jure*. Eesti e-residentsuse idee ei ole midagi muud kui soov suurendada Eestit virtuaalselt; suurendades virtuaalselt residentide arvu, saab Eesti endale juurde rohkem e-teenuste kasutajaid, mis võimaldab ka teenuseid efektiivsemaks muuta. Toodud näited on digimaailmas pigem innovaatilised, nihutades andmete vaba liikumist kasutades ka rahvusvahelise õiguse piire.

Digitaalsete andmete tähtsuse kasv tingib ohu, et tekib niinimetatud “[andmete lõhe](#)” nende inimeste vahel, kes kasutavad, ja nende vahel, kes ei kasuta infotehnoloogiat. Mittekasutajate arvamus, soovid ja väärtused võivad jääda arvesse võtmata, kui disainitakse uusi tooteid või koostatakse uusi seadusi (4). Inimesed, kes ei osale digitaalmeedias, võivad oma puuduva “digitaalse jalajälje” tõttu osutada ühiskondlikes arengutes kõrvalejäetuks.

Mõjud Eestile

Eesti oma digisõbraliku ühiskonnaga on kindlasti **hea platvorm andme-majanduse arendamiseks**. Digitaalrajanduse ja ühiskonna indeks (*Digital Economy and Society Index, DESI*)² asetab Eesti ELi 28 liikmesriigi seas kõrgele seitsmendale kohale. Üldjoontes kuulub Eesti suure edumaaga riikide hulka. Seejuures on Eesti tulemused üle ELi keskmise ja need tulemused on võrreldes 2016. aastaga kiiresti paranenud. Eesti on avalike digiteenuste pakkumise ja kasutamise poolest esirinnas ning eestlased on oskuslikud digitehnoloogia kasutajad, samuti kasutatakse agaralt erinevaid interneti poolt pakutavaid võimalusi. Edusammudest hoolimata seisab ettevõtjate ees väljakutse digitehnoloogiat veelgi enam rakendada (s.h. tööstuse automatiseerimine) (25).

Eesti ettevõtetest on suurandmeid analüüsinud 13%, enamikus ettevõtetes tegid seda oma töötajad. Enim on suurandmete analüüsiga tegelenud info- ja sidevaldkonna ettevõtted (29%), järgnevad veevarustuse, kanalisatsiooni, jäätm- ja saastekäitlusega tegelevad ettevõtted (28%) ning finants- ja kindlustusettevõtted (21%)(27).

McKinsey toob välja **viis valdkonda, kus lähitulevikus on näha suurandmete plahvatuslikku kasutust**: tervishoid, avalik haldus, jaekaubandus, tootmine ja isiku asukohaandmed (nagu näiteks GPSi abil tuvastatav asukoht ning personaliseeritud pakkumised mobiilile) (3). Ka Eestil on nendes valdkondades võimalik tulevikus kasu lõigata.

Kõik täna **riigi poolt pakutavad e-teenused** koguvad andmeid ning moodustavad suurandmete kogumi – nii e-tervis, e-maksuamet, e-häälletamine kui ka näiteks e-kataster on tegelikult suurandmed. Riik küll kogub andmeid, kuid ei analüüsi neid veel sellel tasemel, mida suurandmete analüüs võimaldaks. Näiteks e-tervise raames kogutakse suurel hulgal isikuandmeid, mille analüüs võimaldaks paremini planeerida Haigekassa eelarvet. Samuti on inimesed ise väga aktiivsed andmekogujad – näiteks võiks tulevikus üksikisiku nutikella abil kogutud terviseandmed (vererõhk, pulss, taastumine füüsilisest tegevusest jms) olla automaatselt edastatud e-tervise andmekogusse. See eeldab n-ö hajaandmete kogumise võimalust ning annab hea võimaluse andmeanalüüsiks ja personaalmeditsiini arenguks. Ka geenivaramu on suurandmete kogum, mida analüüsid saab tuvastada

2 Digitaalrajanduse ja -ühiskonna indeks (*Digital Economy and Society Index, DESI*) on Euroopa Komisjoni sidevõrkude, sisu ja tehnoloogia peadirektoraadi poolt välja töötatud liitindeks, millega hinnatakse ELi liikmesriikide arengut digitaalrajanduse ja -ühiskonna suunas. See koondab endas teatavad asjakohased näitajad viiest valdkonnast: ühenduvus, inimkapital, internetikasutus, digitehnoloogia integreerimine ja avalikud digiteenused. Indeks liigitab riigid nelja rühma: suure edumaaga, väheneva edumaaga, järelejäädvad ja mahajäävad riigid. Lisateabe saamiseks DESI kohta vt <https://ec.europa.eu/digital-single-market/en/desi>

mustreid, kuidas inimeste haigused on seotud nende geenidega, mis omakorda võimaldab välja töötada vastavaid ravimeid või anda inimestele teada nende terviseriskidest võimaldamaks läbi teadlikkuse ennetada haigusi. Kuna terviseandmete puhul on tegu tundlike isikuandmetega, siis, kombineerides suurandmeid *blockchain*-tehnoloogiaga, on võimalik andmebaase veelgi tõhustada ning muuta andmeid ka tõenduspõhisemaks ja turvalisemaks.

Eesti X-tee platvorm kogub täna samuti suurel hulgal andmeid, mida analüüsid saaks nii tõhustada [avaliku sektori poolt pakutavaid e-teenuseid](#) kui ka teha riigi edasise digitaliseerumise alaseid strateegilisi otsuseid. Näiteks võimaldab suurandmete analüüs leida mustreid maksude kõrvalehoidmises või prognoosida ettevõtete maksekäitumist. Ennetav statistiline analüüs (*predictive analytics*³) võimaldab näiteks prognoosida, millised on tulevikus potentsiaalsed muutused e-teenuste tarbimismustris või milliseid e-teenuseid mis ajahetkel kõige tõenäolisemalt kasutatakse. Põhimõtteliselt on võimalik kõikide X-tee platvormil pakutavate avalike teenuste andmeid analüüsid muuta need kasutajasõbralikumaks ja efektiivsemaks, kuid ka ressursitõhusamaks. Elektrooniliste pangamaksete analüüs annab näiteks infot nii elanike geograafilisest asukohast, tarbimisharjumustest, pangaeelistustest või kaupluste valikust ja üldisest maksekäitumisest kui ka makse sooritamise vahenditest (mobiil või arvuti), mis omakorda võimaldab teenusepakujal prognoosida, kas tulevikus võiks näiteks enamik makseid toimuda mobiilirakenduse kaudu, ning seda juba varakult tarbija jaoks mugavamaks muuta.

Üheks suureks tarbijakäitumist analüüsida võimaldavate andmete kogujaks on [teleteenuste](#) pakujad, kes on turul muutunud omamoodi oligopolideks. Telekom-ettevõtted koguvad igapäevaselt miljoneid ühikuid andmeid, mida analüüsid saab suurepärase pildi tarbijakäitumise mustritest. See omakorda võimaldab neil oma tooteid ja teenuseid paremini sihtgruppide vajadustega kohandada, suunates nii ka teiste valdkondade tarbimiskäitumist. Sarnaselt lõikavad suurandmete omamisest kasu ka [jaekaubanduse](#) ettevõtted (sh ka kommertspangad), kes üha rohkem kliendikaarte (või tarbimis-/eluasemelaene) välja andes kinnistavad tarbijaid oma kaubamärgi külge, kontrollides nii nende tarbijakäitumist. Jaekaupluste põhiline kliendi kinnistamise võte on allahindlused, mis peaksid tarbijale looma tunde, et kauplus neist hoolib. Samas annab iga kliendikaardi registreerimine kauplusele teavet tarbija ostetavate kaupade, kaupluste asukoha, ostusummade jms kohta, mis omakorda võimaldab kauplustel oma pakkumisi personaliseerida. Samamoodi võimaldab [isiku kohaandmete](#)

3 Ennetav statistiline analüüs on sisuliselt regressioonanalüüs, kus kogutud andmeid mingi kordajaga läbi korrutades saadakse võimalik tulevane käitumismuster. Mida pikem on aegrida ning mida rohkem on andmeid, seda tõepärasem on prognoos.

analüüs tuvastada inimeste liikumistrajektoore ning muuta avalikud teened mugavamaks ja ökonoomsemaks nii turismi- kui ka transpordisektoris ning arendada regionaalpoliitikat.

Üheks suure kasvupotentsiaaliga valdkonnaks, kus suurandmete analüüs võiks anda konkurentsieelise, on [targa linna](#) kontseptsioon. Tallinn kogub tänapäeval suurel hulgal andmeid (hariduses, transpordis, ettevõtluses, elamuehituses jms), mida analüüsid oleks võimalik luua näiteks tarku tänavavalgustusvõrkusid, optimeerida transpordisüsteemi, suunata paremini sotsiaaltoetusi erinevatele sihtgruppidele, kasutada efektiivsemalt energiat või korrastada haridusvõrku.

Autorid



Katre Eljas-Taal

Technopolis Group
Eesti juhataja



Anne Veerpalu

NJORD Advokaadibüroo
partner, Tartu Ülikooli info-
tehnoloogia õiguse doktorant
ja IT-õiguse magistriprogrammi
õppejõud



Jari Romanainen

Technopolis Group poliitika-
nõunik, Soome innovatsiooni-
agentuuri Tekes endine nõunik

Kasutatud allikad

1. OECD (2017): [Bridging the Gap: Inclusive Growth 2017 Update Report](#)
2. Ernst & Young (2014): [Big data: changing the way businesses compete and operate](#)
3. McKinsey & Company (2011): [Big data: The next frontier for innovation, competition, and productivity](#)
4. European Parliamentary Research Service (2016): [Big data and data analytics. The potential for innovation and growth](#)
5. OECD (2014): [Data-driven Innovation for Growth and Well-being](#)
6. Andmekaitse Inspektsioon (2017): [Suurandmed ja privaatsus](#)
7. Statistikaamet (2014): [Eesti statistika kasutamise võimalused](#)
8. Finantsuudised (2017): [Lauri Ilison: väärtus tekib tehnoloogia ühendamisel äriaga](#)
9. IBM (2012): [IBM Smarter Analytics: Big Data and Insurance](#)
10. Big Data Value Association (2016): [Big Data Technologies in Healthcare. Needs, opportunities and challenges](#)
11. Song & Liu (2017): [Predicting Tourist Demand Using Big Data](#)
12. Michael Batty (2013): [Big data, smart cities and city planning](#)
13. Accenture (2014): [Big Data Analytics in Supply Chain: Hype or Here to Stay?](#)
14. Accenture (2016): [Exploring Next Generation Financial Services: The Big Data Revolution](#)
15. McKinsey & Company (2012) : [Creating competitive advantage from big data in retail](#)
16. IDB 2016 (2016): [Big Data in the Public Sector](#)
17. Euroopa Komisjon (2017): [Building a European Data Economy](#)
18. IBM (2016): [The Rise of the Data Economy: Driving Value through Internet of Things Data Monetization](#)
19. World Economic Forum (2011): [Personal Data: The Emergence of a New Asset Class](#)
20. Billy Ehrenberg (2014): [How much is your personal data worth?](#)
21. Euroopa Komisjon. Õigus- ja tarbijaküsimuste peadirektoraat (2016): [Eli andmekaitse reform ja suurandmed](#)
22. Lisbon Council (2017): [Making Europe a Data Economy: A New Framework for Free Movement of Data in the Digital Age](#)
23. Albright Stonebridge Group (2015): [Data Localization: A Challenge to Global Commerce](#)
24. Anna Mayer & Jeffrey Ritter: *Regulating Data as Property: A New Construct for Moving Forward.*
25. Digitaalmajanduse ja ühiskonna indeks (2016): <https://ec.europa.eu/digital-single-market/en/desi>
26. Digiarengu indeks (2017): <https://hbr.org/2015/02/where-the-digital-economy-is-moving-the-fastest>
27. Eesti statistika kvartalikiri 2/2017: https://www.stat.ee/valjaanne-2017_eesti-statistika-kvartalikiri-2-17