



RITA

Alternatiivsed võimalused elukoha andmete fikseerimiseks ja täpsustamiseks.

Mobiilpositsioneerimise ja registriandmete ühendamise seadusandlik kontekst.

TP 2 aruanne

Autorid:
Indrek Keis
Anto Aasa

Tartu 2021

Sisukord

1. Sissejuhatus	4
1.1. Olulisemad mõisted.....	4
1.2. Elukoohaandmete täpsuse olulisus ja tegelik vajadus.....	5
1.3. Elukoohaandmete registrisse kandmise võimalused.....	6
2. Eesti Statistika erinevate registrite ühendamise põhised indeksid	7
2.1. Residentsuse indeks	7
2.2. Partnerluse indeks.....	7
2.3. Paiknemise indeks	8
3. Mobiilsideandmed	8
3.1. ESSnet Big Data II projekt	8
3.1.1. Hollandi katseprojekt	9
3.1.2. Saksamaa kogemused	10
3.1.3. Itaalia kogemused	12
3.1.4. Prantsusmaa katseprojekt.....	12
3.1.5. ESSnet teise etapi kokkuvõte ja osalenud statistikaasutuste ettepanek	13
3.2. Sünteetilised andmed meetoodika arendamiseks.....	13
3.3. Andmetöötamise võimalikud osapooled.....	13
4. Teised alternatiivsed andmeallikad	14
4.1. Elukohtade tuvastamine kolimisfirmade, kinnisvaraportalide ja postiaadresside muutuste ja teiste alternatiivsete andmeallikate näitel	14
4.2. 'Tarkade sõidukite' asukoohaandmed.....	16
4.3. Elektrienergia tarbimise andmed	16
4.3.1. Eriolukorra mõju rahvastiku paiknemisele Elering'i elektritarbimise andmete näitel	18
4.3.2. Elektritarbimise muutuse ruumiline muster.....	20
4.4. Interneti püsiühendused eluruumide asustatuse peeglina	22
4.5. Avaandmed	24
4.6. Perearsti nimistute potentsiaal	25
5. Andmetöötamise õiguslik käsitlus	27
5.1. Statistika- ning teadusuuringutes isikuandmete töötlemise üldised õiguslikud alused.....	27
5.2. Sidevaldkonna eriregulatsioon	28
5.3. Eriregulatsiooni muutumine.....	29
5.4. Siseriiklike seaduste muutmise vajadust tuleb analüüsida	29

6.	Sideandmetega kaasnevate privaatsusküsimuste käsitlemine.....	30
6.1.	Asukohaandmete täpsuse mõju privaatsusele	31
6.2.	Andmete anonüümsus pärast sideoperaatori poolt väljastamist.....	31
6.2.1.	Kas isikud on tuvastatavad kaudsete tunnuste pinnalt?	31
6.3.	Privaatsusriskide käsitlemine krüptograafia arenguga	32
6.4.	Privaatsusriskide käsitlemine valdkondlikul tasemel.....	32
6.4.1.	Andmekaitsealane mõjuhindang õigusloome käigus	32
6.4.2.	Valdkonnapõhine toimimisjuhend.....	33
6.4.3.	Katusorganisatsioonide tegevused	33
7.	Mobiilside andmete kasutamise senine praktika ja vajadused Eestis	34

1. Sissejuhatus

Järgnevalt vaadeldakse rahvastikuregistri ebatäpsuse vähendamise võimalusi erinevate alternatiivsete andmeallikate abil. Eesti Statistika on probleemi lahendamiseks testinud erinevaid viise. Üks suund on olnud erinevate elukoha- ja kontaktaadresse sisaldavate registrite ühendamise põhjal loodud elukohaga seonduvate indeksite arendamine. Testitud on ka elektritarbimise andmeid ja analüüsitud on elektritarbimise andmeid kui inimeste mingis kohas elamise märke.

Lisaks on käesolevasse ülevaatesse kaasatud ESS BigData II projekti tulemused, mille üks töörühm tegeles mobiilpositsioneerimise andmete rakendatavuse uurimisega. Töörühma tegevus keskendus eelkõige riiklike statistikaametite tegevuse mobiiliandmete kasutamiskogemuse uurimisele. Seetõttu on Eesti puhul näiteks välja jäänud Eesti Panga statistika, kus välisreisi statistikat arvutatakse alates 2008. aastast. Seega võivad projekti tulemused olla ebatäpsed ka muude riikide puhul ning positiivseid juhtumeid on veelgi.

Eesti kohta on läbi viidud ka eraldi analüüsid, milles uuritakse, kas inimeste elukohtade täpsustamiseks on võimalik kasutada elektritarbimise ja interneti püsühenduste andmeid. Etteruttavalt võib öelda, et mõlemad andmestikud iseseisvalt elukohtade täpsustamiseks ei sobi, pigem on tegemist abiandmetega näiteks mitme registri ühendamise täiendusega ja/või modelleerimisesisendiga mobiilpositsioneerimise andmete ruumiliseks interpoleerimiseks.

Eesti mobiilpositsioneerimise andmete analüüsi kui alternatiivset võimalust elukohaandmete täpsustamiseks käesolevas dokumendis ei käsitleta, vaid sellele pööratakse ülevaadets tegevuste 2.2 ja 2.3 kohta.

1.1. Olulisemad mõisted

Asjassepuutuvate mõistete puhul on oluline arvestada mõiste kasutamise konteksti, kuna põimuvad nii õiguslikud kui ka statistikatöodes määratletud tehnilised terminid. Seaduses defineeritud mõisted ei pruugi olla samatähenduslikud statistika- ning teadustöodes kasutatud mõistetega (näiteks elukoht) ning mõnel juhul võib mõiste rakendamine olla suhteline, sõltudes huvipoolse hetkehinnangust (näiteks andmete anonüümsus).

Elukoht – rahvastikuregistri seaduse (RSS) kohaselt on elukoht aadress, kus isik alaliselt või peamiselt elab.¹ Tsiviilseaduse üldosa seaduse kohaselt võib elukoht olla üheaegselt mitmes kohas. Elukoht loetakse muutunuks, kui isik asub mujale elama viisil, millest võib järeldada isiku tahet oma elukohta muuta. Kui isiku elukohta ei saa kindlaks määrata, loetakse tema elukohaks tema igakordne viibimiskoht.² Euroopa Liidu õiguses defineeritakse elukohana kohta, kus inimene tavaliselt veedab igapäevase puhkeaja, olenemata ajutisest eemalolekust meelelahutuse, puhkuse, sõprade ja sugulaste külastamise, töö, ravi või palverännaku eesmärgil.³

Sideettevõtja - ettevõtja, kes osutab elektroonilise side teenust (Eestis mobiilsideteenust osutavad suurimad sideettevõtjad on Telia, Elisa, Tele2).

¹ Rahvastikuregistri seadus § 65

² Tsiviilseadusiku üldosa seadus § 14

³ [EL määrus nr 763/2008](#)

Isikuandmed - igasugune teave tuvastatud või tuvastatava füüsilise isiku kohta. Tuvastatav füüsiline isik on isik, keda saab otseselt või kaudselt tuvastada, eelkõige sellise identifitseerimistunnuse põhjal nagu nimi, isikukood, asukohateave, võrguidentifikaator või selle füüsilise isiku ühe või mitme füüsilise, füsioloogilise, geneetilise, vaimse, majandusliku, kultuurilise või sotsiaalse tunnuse põhjal.⁴ Isikuandmed võivad olla nii otseselt tuvastataval (näiteks nimi, isikukood ja sellega seotud tunnused) kui ka pseudonüümsel kujul (näiteks isikuga seostatav koodnimetus ja sellega seotud tunnused).

Anonüümsed andmed – andmed, mida ei ole võimalik siduda ühegi konkreetse üksikisikuga.⁵ Anonüümsuse tase võib olla suhteline ja ajas muutuv, arvestades, et konkreetsetel ajahetkedel võivad andmed olla anonüümsed ühe andmetöötleva jaoks, kuid hiljem täiendavate andmete kättesaadavuse ning nende analüüsi korral võib mõni teine andmetöötleva evida võimekust algselt anonüümseid andmeid siduda mõne konkreetse indiviidiga.

1.2. Elukohaandmete täpsuse olulisus ja tegelik vajadus

Sõltuvalt kohaliku omavalitsuse ja riiklikest avalikest teenustest, võib elukoha määratlus tulla kasuks teenuste osutamise ja planeerimise ning poliitikate kujundamise seisukohast. Korrektne registreeritud elukoht mõjutab rahvaloenduse andmeid ning avab võimalusi teadusuuringute kvaliteedi tõstmiseks ja erinevate meetodikate arendamiseks. Registripõhise elukoha täpsus mängib olulist rolli ka eriolukorra ajal, kuna see võib otseselt mõjutada isiku võimalusi liikuda, kui on kehtestatud liikumispõrangud. Andmete täpsus võimaldab hinnata eriolukorrast tulenevaid asjaolusid või anda asjakohast infot nii avalikule sektorile kui ka avalikkusele tervikuna (näiteks Terviseameti tegevus kinnitatud COVID-19 juhtumite avalikustamine nakatunu elukohajärgse kohaliku omavalitsuse üksuse tasemel).

Andmekaitseõuete seisukohast tuleb arvestada isikuandmete töötlemise põhimõtted, eeskätt eesmärgipärasuse ning andmete võimalikult vähest kogumise printsiipi. Minimaalsuse printsiibi rakendamine eeldab, et andmeid kogutakse ainult eesmärgi täitmiseks vajalikus ulatuses. Seega peab andmetöötleva eelnevalt määratlema, kas eesmärgi saavutamiseks on vaja teada inimese elukohta riigi, kohaliku omavalitsuse, linna, asumi, tänava ning maja ja eluruumi täpsusega.

Peamine ebatäpsete elukohaandmete põhjus on selles, et inimesed ei pea registripõhise elukoha ajakohasena hoidmist vajalikuks.⁶ Seetõttu võib järeldada, et mitte igasugune ebatäpsus ei ole inimese enda vajaduste või avalike teenuste kasutamise ning statistika tegemise seisukohast oluline, kui see ei mõjuta ei inimest ennast ega elukohaga seotud avaliku teenuse pakkujat. Samas viitab see ka sellele, et ka eluruumi täpsusega fikseeritud elukoht ei mõjuta alati isiku elukohapõhiseid avalikke teenuseid.

Elukohaandmete koguse ja kasutamise eesmärkide kirjeldamisel ning andmevajaduste hindamisel tuleb eristada, kas ning miks on vajalik teada inimese elukohta juriidilises tähenduses või on selle asemel võimalik kasutada statistiliste meetodite abil tuletatud elukohta. Seejuures ei saa tähelepanuta jätta ka asjaolu, et juriidiliselt võib inimesel olla mitu erinevat elukohta ning inimese valida jääb, millist neist ta soovib kanda elukohana rahvastikuregistrisse.

⁴ [Isikuandmete kaitse üldmäärus](#), artikkel 4 punkt 1

⁵ AKIT sõnastiku definitsioon

⁶ Vt ka TP1 käsitlus

1.3. Elukohaandmete registrisse kandmise võimalused

Elukohaandmete registreerimist reguleerib rahvastikuregistri seadus ning selle alusel kehtestatud siseministri määrus “Elukohaandmete rahvastikuregistrisse kandmise kord, vormid ja nende täitmise juhend”.

Elukohaandmeid on võimalik registrisse kanda kahel viisil. Üldreeglinas kantakse elukoht rahvastikuregistrisse isiku enda poolt esitatud elukohateate alusel. Lisaks saab elukohateate alusel kanda registrisse ka isiku alaealiste laste ning eestkostetavate elukoha aadressi, mille õigsuse tagamise kohustus isikul endal.⁷ Mõnel juhul nõuab on elukohateate alusel kande tegemise eelduseks ka mõne teise isiku nõusolekut. Nõusolekut küsitakse nii ruumi omanikult ning alaealise lapse elukoha registreerimisel teise hooldusõigusliku vanema käest.⁸ Püsivalt välisriigis elav Eesti kodanik ja konsulaarseaduses sätestatud välismaalane võivad välisriigi elukoha aadressi kandmiseks rahvastikuregistrisse esitada taotluse Eesti välisesindusele konsulaarseaduses sätestatud korras.⁹

Teiseks võimalusena võib elukohaandmete kanne toimuda kohaliku omavalitsuse üksuse algatusel. RRS § 85 sätestab, et KOV algatuse aluseks võib olla sotsiaalhoolekande-, meditsiini-, pääste-, politsei- või muu asutuse teade kohaliku omavalitsuse üksusele oma ülesannete täitmise käigus kohaliku omavalitsuse üksuse territooriumil avastatud püsivalt viibivast isikust, kelle elukoha aadress ei ole rahvastikuregistrisse kantud. RRS § 86 kohaselt võib KOV algatuse aluseks olla ka olukord, kus kohaliku omavalitsuse üksuse territooriumil püsivalt viibiva rahvastikuregistri subjektiks oleva isiku elukoha aadress ei ole kantud rahvastikuregistrisse või isiku kinnitusel ei kasuta ta enam rahvastikuregistrisse kantud aadressijärgset elukohta.

Lisaks eelkäsitletud võimalustele võib RSS kohaselt juba registrisse kantud elukoha aadressi muuta ruumi omaniku nõudmisel, vastavasisulise kohtulahendi või üürikomisjoni otsuse alusel, haldusterritoriaalse korralduse või aadressikohaks olevate objektide nimede ja numbrite muutumisel või välisriigi pädevalt asutuselt saadud andmete alusel, kui inimese elukoht on välisriigis.

Kohalikule omavalitsusele on elukohateadet on võimalik esitada kohapeal, posti teel, digitaalselt e-kirjaga kui ka vastava e-teenuse (e-rahvastikuregister) kaudu. Elukohateate või KOV algatatud menetluse tulemusel tehtud elukohaandmete rahvastikuregistrisse kandmise otsus tehakse elukohakandena rahvastikuregistrisse.

Elukohaandmete registreerimine võib teatud juhul mõjutada ka elukohariigi määramist. Eesti ja Soome valitsuste ning rahvastikuregistrupidajate plaanitava andmevahetuslepingu kohaselt viimase alalise elukoha muutmise fakt olla aluseks elukohariigi määramisel, kui eelnevalt ei saa kindlaks teha isiku enda arvamust elukohariigi valikul.¹⁰

⁷ RSS § 68

⁸ RSS § 75

⁹ RSS § 95

¹⁰ Andmevahetusleping ja menetluse staatus kättesaadav

<https://eelnoud.valitsus.ee/main/mount/docList/195b5880-005b-4e7c-8668-6c7440a4b644#CiLvVkid>

2. Eesti Statistika erinevate registrite ühendamise põhised indeksid

Registripõhiselt määratletud elukoha tuvastamise alternatiividena on Statistikaameti poolt testitud erinevaid meetodeid, mis võtavad sisendiks erinevate registrite andmekoosseise.

2.1. Residentsuse indeks

Alates 2016. aastast kasutab Statistikaamet rahvaarvu analüüsimisel ja avaldamisel residentsuse indeksil tuginevat meetodikat. Indeksi aluseks on tees, et Eestis elades kasutab isik riigi pakutavaid teenuseid, mistõttu on tekivad isiku kohta andmed peale Rahvastikuregistri ka teistesse registritesse. Indeks väärtus on 0 kuni 1. '0' väärtuse korral võib kindlalt sedastada, et isik on mitteresident. Mida suurem isikuga seotud indeksi väärtus, seda tõenäolisemalt võib järeldada, et isik on Eesti resident.

Residentsuse indeksi alusel on teatud tõenäosusega võimalik fikseerida asjaolu, kas isik omab Eestis püsivat elukohta. Statistikaamet tõi esile, et indeksi arvutamisel kasutati 2015. aastal 14 Eesti administratiivset registrit ja alamregistrit, sh Eesti Hariduse Infosüsteemi, riiklikku pensionikindlustuse registrit, ravikindlustuse andmekogu jne. Aktiivsust registrites mõõdetakse nn elumärkide abil. Iga register või alamregister annab inimesele ühe elumärgi, kui inimene sooritab aasta jooksul mingi aktiivse sammu, mis kantakse sellesse registrisse. Keskmiselt koguneb Eesti elanikele aasta jooksul kuni 5 elumärki ja elumärkide summa alusel õnnestub enamik Eesti elanikest jaotada kahte ossa. Piisava arvu elumärkide alusel võib oletada, et tegemist on Eesti elanike ehk residentidega. Teise osasse liigituvate isikute puhul elumärke pole ning need isikud tõenäoliselt ei ela Eestis ja nad on mitteresidendid. Statistikaamet toob esile, et on selline jaotus praktiliseks kasutamiseks liiga ebatäpne, sest nende vahepeale jäävad ühe elumärgiga isikud, kelle kohta ei saa usaldusväärset otsust langetada. Samuti, mõnel Eesti elanikul ei teki mõnel aastal ühtegi elumärki ning neid ei ole õige kohe residentide seast välja arvata.¹¹

Indeksi kujunemisel võetakse arvesse inimese kohta teadaolevaid elumärke möödunud aastal. Indeks võimaldab eristada nii siserännet ja välisrände korral sisse- ning väljarännet.¹²

2.2. Partnerluse indeks

Sarnaselt residentsuse indeksile, kasutatakse partnerluse indeksi leidmiseks täiendavaid registreid. Indeksi eesmärk on vähendada rahvastikuregistri ebaõigetel elukohaandmetest tekkivaid vigu. Indeksi eesmärk on seostada isikuid nende tegelike täisealiste vastassoost partneritega.¹³

Partnerluse indeksi meetodi aluseks on kahte liiki andmed - esiteks tuleb leida andmed, mis seovad isikud üheks leibkonnaks ja/või perekonnaks, ning teiseks ka andmed, mis seovad isikuid ja eluruume. Meetodit võib kasutada kahel erineval viisil - kitsama ja laiema ülesandepüstitusena. Statistikaamet eristuse järgi määratakse kitsama ülesandepüstituse korral määratakse partnerid vaid üksikvanematele (kui see osutub võimalikuks), jälgides sealjuures, et üksikvanematele määratavad partnerid ei tohi kuuluda partnerina olemasolevatesse registripõhisesse leibkondadesse (see tähendab, et olemasolevaid registripõhiseid perekondi ei lõhuta). Laiema ülesandepüstituse korral

¹¹ <https://www.stat.ee/dokumendid/399211>, lk 2

¹² Samas, lk 5

¹³ Constructing families using administrative registers, lk 3

partnerite valimisel ei piirdata üksnes eespool nimetatud „vabade“ isikutega, vaid vaadeldakse kõiki olemasolevatesse eluruumipõhistesse leibkondadesse kuuluvaid isikuid.¹⁴

Statistikaamet toob esile, et partnerluse indeksil baseeruva meetodi järjepidev kasutamine tõstab indeksi täpsust, kui otsustusprotsessis on võimalik kasutada eelmiste aastate andmeid.¹⁵ Meetodi täpsust parandatakse täiendavate andmeallikate kasutamisega (nt 2019 aastal lisati andmed eeskoste kohta).¹⁶ Partnerlusindeksi metoodikat on põhjalikult tutvustatud samas kvartalikirjas 2018 esimeses kvartalis.

2.3. Paiknemise indeks

Leibkondade elukohta määramiseks arendab Statistikaamet paiknemise indeksi. Selleks kasutatakse rahvastikuregistri andmeid kõigi pereliikmete registreeritud elukohta kohta ning kinnistusraamatust nähtuvaid kinnisomandeid. Leibkonna määramisel arvestatakse partnerlusindeksi väärtust. Otsitakse paiknemismärke (registreeritud pereliikmete arv, omanikus pereliikmete arv, seotud eluruumis elavate lähisugulaste ja mittelähisuguluaste arv, info mugavuse kohta ja tubade arvu piisavus) ning tuvastatakse parim võimalik elukoht.¹⁷ Leibkondade kontrolluuringu käigus märgiti, et puuduvad kõik vajalikud partnerluse märgid ning indeks ei ole veel lõplikult välja töötatud. Statistikaameti hinnangul tuleks indeksi koostamisel kasuks ka mobiiliandmete ja elektriandmete kasutamise võimalus.¹⁸ Statistikaamet jätkab eelnimetatud meetodite arendamist ning testimist vähemalt kuni 2021 lõpuni.¹⁹

3. Mobiilsideandmed

Lisaks eelnimetatud Statistikaameti alternatiivsetele meetoditele, on isiku elukoht võimalik määratleda ka mobiilsideandmete pinnalt. Sideandmete kasutamine nõuab mitmete juriidiliste ning metoodikapõhiste sõlmküsimate lahendamist. Seetõttu pole see täielikult väljakujunenud valdkond, kuid esmased sammud on selleks siiski tehtud. Alljärgnevalt kirjeldame mobiilsideandmete potentsiaalseid kasutusjuhte, kaasnevaid takistusi ning ka ohte.

3.1. ESSnet Big Data II projekt

Euroopa Statistikasüsteemis (ESS) osalevad riigid alustasid 2016 aastal ühise suurandmete projektiga. Projekti üldine eesmärk oli käsitleda suurandmete integreerimise võimalusi riiklikustatistika koostamises, uurides erinevaid potentsiaalseid andmeallikaid. Projektis on 12-st kitsama suunitlusega töopaketti. Peamised uurimisteemad hõlmasid turismistatistikat, finantstehingutega kaasnevaid

¹⁴ Statistikaameti kvartali ajakiri, 2018 Q1 lk 32, 33, kättesaadav https://www.stat.ee/publication-download-pdf?publication_id=44733&publication_lang=et

¹⁵ Samas, lk 39

¹⁶ [Constructing families using administrative registers](#), Helle Visk, Kristi Lehto, lk 14

¹⁷

<https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwixu5qR54HuAhUwxosKHfncDUMQFjABegQIARAC&url=https%3A%2F%2Fwww.stat.ee%2Fdokumentid%2F1413081&usq=AOvVaw0ugheR2Nfp6VbDwzQMaeQd>

¹⁸

<https://www.stat.ee/sites/default/files/202011/Leibkondade%20kontrolluuringu%20tulemused%202018.pdf>, lk 6

¹⁹ Statistikaameti 27.01 e-kiri

andmeid, ruumiandmed (nt satelliidi- ja aerofotode teel saadavad andmed) ning mobiilsideandmeid. Projekti [esimene etapp](#) kestis 2016 kuni 2018 maikuuni, [teine etapp](#) algas 2018 novembris ning kestab kuni 2021 juunini.

Esimese etapi raames uuriti [tööpaketi](#) nr 5 käigus mobiilsideandmete kasutamist andmeallikana ametliku statistika tarbeks. Teostati mõned pilootuuringud, mille tulemusel tunnustati sideandmete potentsiaali ning [leiti](#), et sideandmete kasutamiseks on vajalik statistikaasutuste ning sideoperaatorite koostöö. Vastastikusest koostööst võivad kõik osapooled, sest mobiilside andmed on statistika seisukohast relevantsete ning andmeanalüüsist võib kasu lõigata ka sideoperaator.²⁰

Teise etapi raames käsitleb mobiilsideandmetele keskenduva [tööpaketti](#) juba detailsemalt andmete ligipääsetavust, statistika asutuste ja mobiilsideoperaatorite koostöövorme, andmete kvaliteedi ning andmete kasutamist toetavate ühtsete standardite küsimusi. Teise etapi üldine eesmärk oli luua standardne raamistik, mis käsitleks sideandmete kasutamist ametliku statistika tarbeks ning peaks vastavalt esimese etapi suunistele olema modulaarse ülesehitusega. Selleks teostasid tööpaketi osalenud riigid uusi katseprojekte või vaadati vastavas kontekstis üle varasemalt tehtud uuringuid.

3.1.1. Hollandi katseprojekt

Hollandi katseprojekti käigus kasutati meetodikat, mis võimaldas tunnipõhiselt hinnata rahvastiku liikumist Hollandi omavalitsuste vahel tuginedes agregeeritud anonüümsetele mobiilsideandmetele. Meetodika tuvastas töönädala ja nädalavahetuse liikumised ning samuti suuremate üritustega kaasneva liikumise. COVID-19 aspektis järeldati, et meetodika võimaldab jälgida riiklikult kehtestatud piirangute mõju ning teatud määral tuvastada võimalikku nakatumiste ahelat. Tõdeti ka mitmeid puudusi ja statistiliste järelduste ebatäpsusi (nt mitme seadme kasutamine, ebaõiged registripõhised elukohaandmed), mis mõjutavad meetodika täpsust.

Katseprojektis kasutati CDR koosseisust mitmekesisemaid sideandmeid, mille hulk ja kogumise intensiivsus sõltub mh tarbija kasutatavast seadmest, selle operatsioonisüsteemist ning ka levisagedusest (2G kuni 5G). Seejuures tõdeti, et sama meetodika on kohaldatav ka CDR põhine andmestiku kasutamisel.

Sideandmete kogumist kasutati eelnevalt räsitud *imsi* koodi (International Mobile Subscriber Identity), mis tagas, et projektis osalevad statistikaasutuse ega sideoperaatori analüütikud ei saaks seostada koodi kindla seadmega. Lisaks koguti ajatemplina aeg, millal toimus kommunikatsioon seadme ning sidemasti vahel ning masti unikaalne identifikaator. Elukohana määratleti sidemasti piirkond, millega mobiilseade oli 30 päevase jälgimisperioodi jooksul ühendatud kõige pikemalt.

Projektis kasutati kuut erinevat andmeallikat ja -sisendit. Esmaseks andmeallikaks oli ühe sideoperaatori suurandmete keskkonnas olevad 4G võrgu sideandmed. Teiseks koguti sideoperaatoritelt ajakohastatud sidemastide asukohad ning seadistused. Kolmanda allika näol oli tegemist statistikaasutuse varasemalt koostatud statistikaga, mis kajastas omavalitsuse elanike arvu. Neljandaks sisendiks võeti kaardivõrgustik 100 x 100 meetrit ruudustikuga. Viidandaks oli tarvis arvestada omavalitsuste geograafilisi piire. Samuti kasutati Hollandi kõrguskaarti. Töövoos vaates toimus enamus andmeallikate kombineerimine sideoperaatori valduses.²¹

²⁰ Belgia NSI ja Proximus (MNO), Eurostat, lk 43

²¹ Samas, lk 7

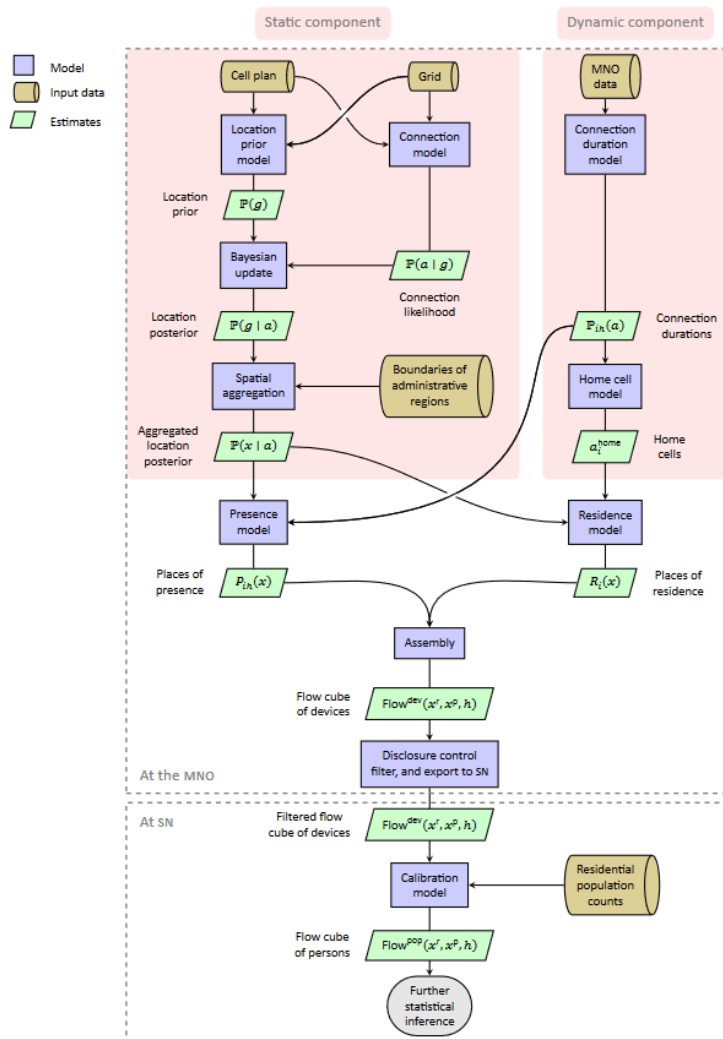


Figure 2.1: The processing pipeline

Joonis 1. Hollandi katseprojekti töövoog²²

3.1.2. Saksamaa kogemused

Saksamaa statistikaasutusel (Destatis) on koostöö kokkulepe kahe sideoperaatoriga kolmest ligikaudu võrdse turuosakaaluga sideoperaatorist. Koostöölepete alusel ostetakse sideoperaatoritelt agregeeritud anonüümseid andmeid ning teostatakse järjepidevalt erinevaid statistikatöid. Erinevate turuosaliste andmeid ühildades saab meetodikaid katsetada tuginedes hinnanguliselt kahe kolmandiku lõppkasutajate agregeeritud sideandmetele. Sideoperaator seob sideandmetega ka lõppkliendi sotsiaal-demograafilised tunnused nagu sugu, vanusegrupp ja rahvus.²³

Agregeerimise tase võib sõltuvalt sideoperaatorist erineda ning sõltub muuhulgas sideoperaatori enda vastavast protsessist ning operaatori ja Saksamaa andmekaitseasutuse (BfDI) vahelisest kokkulepest.²⁴

²² Samas, lk 10

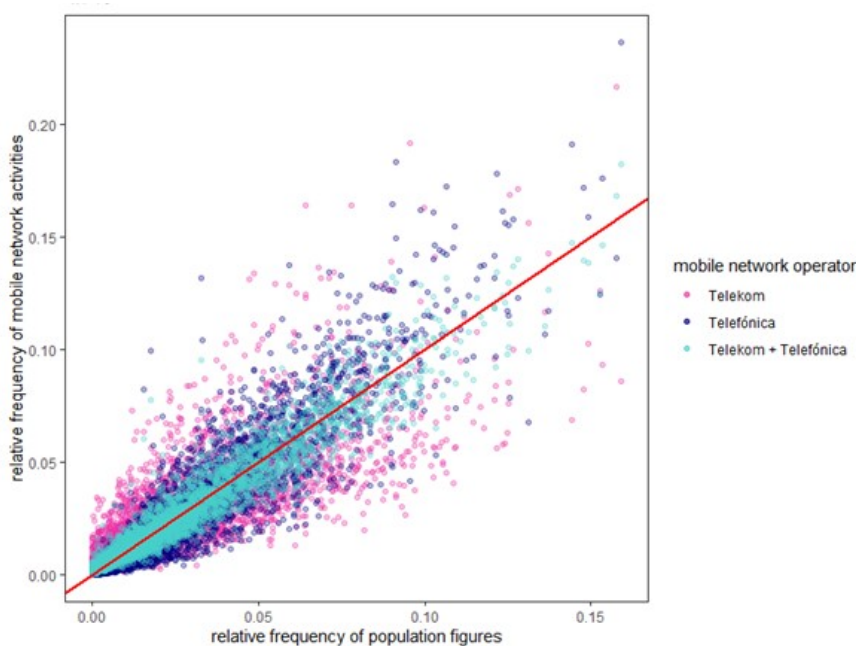
²³ Samas, lk 21

²⁴ Samas, lk 22

Minimaalselt võetakse agregeerimise aluseks vähemalt viie mobiilseadme kasutamisest tekkivad andmed. Sideoperaator kasutab sideandmete ning kliendi kohta teadaolevate andmete ühendamiseks ajutisi pseudo-identifikaatoreid. Ühendatud andmed seejärel agregeeritakse ning edastatakse statistikaasutusele.

Varasemate ja käimasolevate projektide käigus uurib Destatis sideandmete võimalikke kasutusvõimalusi peamiselt rahvastiku-, pendelliiklus- ja turismistatistika koostamiseks.²⁵

Rahvaloenduse kontekstis kontrolliti sideoperaatorite andmestiku kasutamise täpsust, võrreldes 2011.a rahvaloenduse tulemusi ning kahe sideoperaatori käest 2018/2019 kogutud sideandmeid. Eesmärk oli kontrollida, kas ühe või kahe sideoperaatorite andmestikust nähtuva mobiilside kasutamise andmed (mis viitavat populatsiooni tihedusele) kattuvad rahvastikuloenduse alusel teadaoleva populatsiooni tihedusega. Kuigi sideoperaatorite turuosad on enam-vähem võrdsed, tõstab kahe erineva sideoperaatori andmestiku kombineerimine oluliselt statistikatöö täpsust.



Joonis 2. Saksamaa kahe sideoperaatorite andmestikest nähtuva populatsioonitiheduse kattuvus rahvaloendusega (joonisel punane joon)

Tuginedes eelnevalt kontrollitud sideoperaatorite andmestike kasutatavusele, uurib Destatis edaspidiselt, kas sideandmeid saaks kasutada rahvaloenduste vahelisel ajal populatsiooni kohta järelduste tegemiseks. Et seostada vaid residentide andmeid 1 x 1 km ruudustikuga, võetakse esmalt arvesse iga seadme päeva esimese ja viimase signaali asukohta, mille kattuvuse korral eeldatakse, et tegemist on seadme kasutaja elukohaga. Samuti arvestatakse pühapäeva öhtuseid asukohaandmed, mis on praktikas näidanud silmatorkavat kattuvust varasema rahvaloenduse andmetega, mistõttu võib agregeeritud andmete puhul eeldada sama. Andmete üldistamiseks laiemale piirkonnale määratakse andmetele erineva kaaluga väärtused. Lisaks on üks eesmärke uurida, mis ulatuses on erinevate väärtuste kaalumise korral võtta sisendiks täiendavaid andmeid nime ning sugu.²⁶

²⁵ Samas, lk 23

²⁶ Samas, lk 24

Kokkuvõtvalt saab Saksamaa kogemustele viidates järeldada, et sideandmeid kasutades on võimalik saada hea ülevaade rahvastikust ning populatsiooni liikumisest ning sideandmed on statistika koostamisel väärtuslikuks andmeallikaks.²⁷

3.1.3. Itaalia kogemused

Itaalia Riiklik Statistikaamet ([Istat](#)) algatas juba 2015. aastal koostööprojekti sideoperaatoriga, eesmärgiga uurida CDR andmete pinnalt koostatud liikumismudelit võrdluses ametliku statistika pinnalt koostatud mudelitega. Projekti käigus väljastati andmed sideoperaatori poolt tasuta ning 5 järjestikuse nädala kohta. Lisaks lisati projekt riikliku statistika plaani ning andmevahetusele eelnevalt teostati andmekaitsealane mõjuhindang privaatsusriskide tuvastamiseks. Kasutati kolme osapoolega andmevahetust, kus esialgu edastas sideoperaator andmed anonüümsel kujul projekti kaasatud Pisa Ülikoolile, kes analüüsis andmeid statistikaasutuse antud meetodika järgi. Seejärel kaasati statistikaasutus tulemuste hindamiseks ning verifitseerimiseks.²⁸

Sideandmetest tuletava elukoha ja teadaoleva ametliku statistika korrelatsiooni hindamisel võeti aluseks öötundidel (öhtul kella 8st hommikul kella 7ni) toimuva mobiilside kasutajate arv. Tulemused on positiivsed, näidates sideandmetest tuletatavate elukohtade tugevat seost ametliku statistikaga. Kuna Itaalia statistikaasutus on samuti loobumas traditsioonilisest rahvaloendusest, peetakse sideandmete kasutamist ka heaks võimaluseks tuvastada ning hinnata probleeme rahvaloenduse läbiviimisel. Näiteks võib sideandmete pinnalt tekkida võimalus hinnata piirkonnapõhiselt ohtu loenduse raames andmete üle- või alaesindatuse osas, võrreldes selleks sideandmetest tuletatavat populatsiooni piirkonnas ametlikult teadaolevate elanike arvuga.²⁹

Itaalia kogemustest nähtub, et sideandmed on kasutatavad rahvaloenduse kombineeritud meetodil läbiviimisel ning neid võib kasutada nii täiendava kui ka peamise andmeallikana.³⁰

3.1.4. Prantsusmaa katseprojekt

Prantsusmaa statistikaasutus ([INSEE](#)) uuris koostöös sideoperaatoriga sideandmete kasutamist tuginedes algselt agregeerimata andmetele, kuid uurides ka agregeeritud andmeid kasutava meetodika kohaldamist. Erinevalt teiste riikide kogemustest, oli INSEE-l võimalik tugineda nõrke sidevõrgu toorandmetele, mis võrrelduna CDR andmetega sisaldavad oluliselt rohkem sündmusi, fikseerides seadme liikumise oluliselt täpsemalt. Kasutatud andmestikus sisaldas üks kuupäev ligikaudu 20 miljardit sidevõrgu sündmust (INSEE tõi võrdluseks sama sideoperaatori 2007 aastast pärit CDR andmestiku, mis sisaldas ühe kuupäeva kohta 100 miljonit sündmust.)

Kõnealused andmed olid eelnevalt kogutud varasema uurimistöo raames ning olid enne nende kustutamist agregeerimata kujul kasutatavad vaid paari kuu jooksul ning ainult sideoperaatori valduses. INSEE proovis andmete pinnalt koostada esialgset statistikat, et võrrelda selle koosõla ametliku rahvastikustatistikaga.³¹

Arvestama peab seadme väljalülitamisest kasutaja poolt kuni lõppkasutaja sideoperaatori vahetamiseni, mis on lõppkasutaja poolt hõlpsasti teostatav. 2018 esimese poolaasta jooksul vahetas

²⁷ Samas, lk 32

²⁸ Samas, lk 35

²⁹ Samas, lk 38

³⁰ Samas, lk 36

³¹ Samas, lk 45

Prantsusmaal 4 miljonit numbriomanikku oma sideoperaatorit, võttes teise sideoperaatori juurde kaasa oma algse telefoninumbri. Lisaks võib esineda tehnilisi tõrkeid ning kasutaja lahkumist sideoperaatori teeninduspiirkonnast. Need asjaolud tingivad ³² INSEE kogemustest nähtub, et täiemahulise metoodika arendamiseks on primaarse tähtsusega ligipääs andmetele, õiguslik alus nende töötlemiseks ning sideoperaatori koostöö. Populatsiooniga seotud usaldusväärsete hinnangute andmiseks on vajalik teatud ligipääs longituudsetele andmetel.³³ Statistika kvaliteeti ja lõpptulemust mõjutavad ka sideoperaatori tegevused, võimekus ning kasutusel olevad lahendused. Siiski näib INSEE hinnangul sideandmete kasutamine ja vastavate metoodikate arendamine perspektiivikas, kuid see nõuab tihedat koostööd sideoperaatoriga ning andmevahetusel tuleb arvestada sideoperaatori klientide privaatsuse kaitsega.

3.1.5. ESSnet teise etapi kokkuvõtte ja osalenud statistikaasutuste ettepanek

Erinevate riikide katseprojektid ja kogemused näitavad, et sideandmeid on võimalik kasutada peamise või täiendava andmeallikana kombineeritud rahvaloenduse läbiviimiseks, kuid kvaliteetse statistika loomine eeldab asjakohaste ning sideandmete puudustega arvestavate metoodikate arendamist ning tihedat koostööd sideoperaatoritega.

ESSnet teise etapi ühe väljundina koostati ettepanek, mis annab edasi vastavate statistikaasutuste nägemuse raamistikust ja protsessidest, millega arvestada sideandmete kasutamisel statistika koostamisel.³⁴

3.2. Sünteetilised andmed metoodika arendamiseks

Ligipääs mobiilandmetele on maailmas olnud väga lünklik ja takistatud. Samas on teada nende andmete väärtus ning loodetakse, et ligipääs lähemal ajal paraneb. Kuna väärtusahel kõnetoimingute ja/või mobiilse Interneti kasutamise andmete konverteerimisel elu- ja muudeks tähenduslikeks kohtadeks on väga pikk ning metoodika keeruline, siis on hakatud arendama sünteetilisi mobiilpositsioneerimise andmeid. Ühelt poolt loodetakse sellega kiirendada metoodika arendamist, teisalt aga luua ka platvormi, mis aitaks vähendada privaatsusriske, mida päris andmed alati endas kätkevad (Salgado *et al.*, 2021).

3.3. Andmetöötlemise võimalikud osapooled

Sõltuvalt andmetöötlemise eesmärkidest ning osapoolte vajadustest ja võimalustest, saab esile tuua kolm erinevat lähenemisviisi sideandmete statistilisel eesmärgil töötlemisele.

Andmetöötlus sideoperaatori valduses

Statistika aluseks olevad andmed jäävad sideoperaatori valdusesse. Enne andmete edastamist rakendatakse sideoperaatori klientide privaatsust tagavaid abinõusid nagu näiteks anonümiseerimine³⁵ ja agregeerimine³⁶.

Seejärel statistikaamet kombineerib erinevate operaatorite tulemused. Antud variant on statistikaameti jaoks ressursisäästlikum ning sideoperaator saaks agregeeritud andmeid kasutada ka

³² Samas, lk 47

³³ Samas.

³⁴

https://ec.europa.eu/eurostat/cros/system/files/wpi_deliverable_i6_a_proposal_for_a_statistical_production_process_with_mobile_network_data_18_03_2021_final.pdf

³⁵ Anonümiseerimise mõiste -

³⁶ Agregeerimise mõiste -

oma huvides. Õiguslikult ja privaatsuse kaitse seisukohast kõige lihtsam teostada. Miinuspoolelt saab esile tuua, et agregeeritud andmetele tuginedes on keerulisem rakendada ühtset metoodikat. Sideoperaatorite arusaam ning hinnang privaatsust tagavatest abinõudest võib olla erinev, mistõttu erinevate sideoperaatorite andmed ei pruugi olla üheselt võrreldavad ega sama kvaliteediga.

Andmetöötlus statistikaameti valduses

Statistikaamet vastutab ning korraldab kogu andmetöötluse ahela toimimise, sideoperaatori tegevus piirdub peamiselt andmete edastamisega. Antud variant evib õiguslikke kitsaskohti andmete edastamise osas ning ühtlasi nõuab statistikaametilt enim ressursse. Variandi eeliseks on võimalus rakendada ühtset metoodikat.

Kolmanda osapoole kaasamine

Andmetöötluse korraldab sideoperaator ja riikliku statistika asutuse poolt usaldatav kolmas osapool. Ressursikulu poolest on see mõistlik nii sideoperaatori kui statistika asutuse seisukohast, kuid ohuks on õiguslikud takistused ja kolmanda osapoole kasutatava metoodika usaldusväärsus. Samas võimaldab kolmanda osapoole kaasamine suurendada läbipaistvust, olles oluline osa teaduspõhise metoodika arendamise seisukohast.

Kolmandaid osapoolte kaasamise teel on võimalik suurendada ka sideoperaatorite lõpp-klientide privaatsuse kaitset. Nii saab teostada vajaliku andmeanalüüsi, kui samal ajal tagades, et ükski osapool ei saa andmeid omavahel täiendavalt ühendada.³⁷

Kes otsustab andmete väljastamise teadusuuringute tarbeks?

Mobiiltelefoniteenust osutavad sideoperaator on sideandmete eest vastutav nii isikuandmete kaitse üldmääruse kui ka elektroonilise side seaduse (ESS) §§ 101 lg 1 ja 102 lg 1 tähenduses.³⁸ Muuhulgas hõlmab vastutus andmete korrektset kogumist, kasutamist, säilitamist ning edastamist.

4. Teised alternatiivsed andmeallikad

4.1. Elukohtade tuvastamine kolimisfirmade, kinnisvaraportaalide ja postiaadresside muutuste ja teiste alternatiivsete andmeallikate näitel

COVID-19 pandeemiast lähtuvalt on hakatud rääkima erinevate riikide näitel, et inimesed lahkuvad massiliselt linnadest ning kolivad elama maapiirkondadesse. Enamasti on selliseid kiireid muutusi rahvastiku paiknemise väga raske nii kiiresti tuvastada: traditsioonilistesse andmebaasidesse nagu näiteks rahvastikuregister jõuavad elukohavahetused alati hilinemisega. Lisaks „segab“ kiirete muutuste analüüsi ka asjaolu, et püsielukohta defineeritakse 1 aasta põhiselt. Vajadust kiireks rände tuvastamiseks nähti tingituna COVID-19 epideemiast: kui enne epideemiat oli füüsiline lähedus keskustele (headele töökohtadele ja teenustele) normiks, siis sotsiaalse distantseerumise tingimustes, kui liikumist piirati, kesklinnade tühjenesid, kodukontorist tehti kaugtööd ning füüsiline suhtlus kolis Internetiplatvormidele, tekkis hüpotees, et inimesed lahkuvad tihedalt asustatud linnadest. Hüpotees väitis, et linnaline eluviis on väga kallis ning sotsiaalse distantseerumise tingimustes ka väga ebamugav. Seega tekib ilmselt ka soodne pinnas inimeste liikumiseks soodsamasse ja mugavamasse äärelinna keskkonda.

Ameerika Ühendriikides viidi läbi mitmeid uurimusi, mis otsisid samuti alternatiivseid võimalusi migratsiooni (st ka elukohtade) kiiremaks kirjeldamiseks. Selleks testiti erinevaid andmeallikaid.

³⁷ Positium/Cybernetica - 2018 Indonesia projekt

³⁸ Isikukaitse üldmääruse artikkel 4 punkt 7, [Elektroonilise side seadus](#), RT I, 08.01.2020, 4

Analüüsiti näiteks kolimisfirmade numbreid. Uuriti kinnisvaraotsingute andmebaaside päringute ajalugu ([link](#)), tehti küsitlusi. Ükski neist andmeallikatest ei osutunud piisavaks, kuid andsid hea esmase ülevaate toimuvast ([link](#)). Näiteks lükati kolimisfirmade andmetel kiiresti ümber arvamus, et inimesed lahkuvad kesklinnadest massiliselt. Küll aga selgus hoopiski, et inimeste kolimine langes pandeemia tingimustes, mitte ei kasvanud. Lisaks selgus ka, et ei kolitud mitte niivõrd äärelinnalisemasse keskkonda, nagu arvati, vaid samasugusesse suurlinna, kust lahkuti. Seejuures oli muutuste number väga varieeruv ulatudes 1,3% langusest Arizonas 66,1% New-Hampshire'is.

Kinnisvaraportalide otsinguandmete põhjal väideti, et kodude müük langes juunis 2020 üle 10%, mis peaks olema tõendiks, et inimeste liikuvus vähenes. Samal ajal kui juulis 2020 täheldati rekordilist kasvu. Samas tõdetakse ka, et kinnisvara müügitehingud ei pruugi näidata mitte niivõrd inimeste elukohavahetusi vaid peegeldavad pigem majanduslikku olukorda. Nii näiteks oli migratsioon enne COVID-19 lainet langenud ajaloolisse madalseisu ([link](#)). Lisaks kimbutab seniseid tulemusi asjaolu, et suurandmete analüüs ei anna enamasti tulemuseks tegelikke numbreid vaid muutuste trende. Otsustajatel ning statistika jaoks on aga vaja just viimaseid.

Eestis on kolimisfirmade andmetel inimeste liikuvust ilmselt väga keeruline tuvastada: teenus pole levinud ja turul pole suuri tegijaid, kelle kaudu oleks võimalik ligi pääseda esinduslikule andmehulgale. Nii pole ka käesoleva tööpaketi läbiviijatele teadaolevalt Eestis vastavat andmestikku mida uurida.

Eeltoodud alternatiivsete andmeallikate otsingute kõrval viidi läbi ka postiaadresside muutmistaotluste analüüse (nt [allconnect](#)). Selleks päriti Ameerika Ühendriikide postiteenistusest (U.S. Postal Service) postiaadresside muutmistaotluste andmed. Uuringu aluseks olnud hüpotees, et inimese postiaadress näitab tema tegelikku elukohta seisab väga tugeval alusel. Tegemist on unikaalse ja detailse andmestikuga, mitte ainult oma esinduslikkuse pärast vaid ka seetõttu, et eristatud on ajutised (< 6 kuud) ja püsivad postiaadressi muutused. Samas ei saa siiski kindlalt väita, et kõnealune andmestik sobib püsiva statistika tegemiseks elukohtade täpsustamiseks, kuna vaatlusalused uurimused tegelesid eelkõige COVID-19 tõttu linnadest lahkumise hüpoteesi kinnitamise või ümberlükkamisega.

Eesti riiklik postiteenuse osutaja Omniva eraldi andmestikku aadresside ja isikute seostega ei oma. Postisaadetiste kätetoimetamisel lähtutakse ainult aadressist.³⁹ Sarnaselt USA postiaadresside muutuste analüüsile võib ka Eesti puhul olla kasulik, kui e-riigiga seotud portaalid süsteemselt ja tihedamalt inimeste kontaktaadresse päriks. See eeldaks mõistagi vastavate andmestike linkimist ja seostamist Rahvastikuregistri elukohtade andmebaasiga.

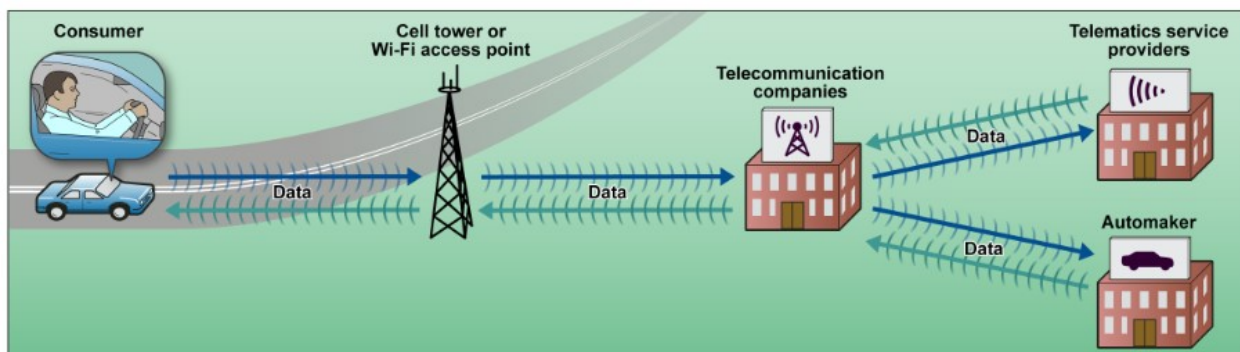
Lisaks eeltoodule on kaalutud, kas inimese elukohta on võimalik määratleda retseptiravimite väljastamise alusel, hüpoteesiga, et retsepte lunastatakse elukoha lähedases apteegis. Esialgsel hinnangul ei ole see perspektiivne alternatiiv, kuna andmete tekkimise tihedus ajas ning asukoht ei paku andmestiku esindatuse ning kvaliteedi osas konkurentsi muudele andmestikele. Lisaks on see seotud inimeste terviseandmetega, mis on seadusega määratletud kui eriliigilised isikuandmed ning mis kätkevad eelduslikult alati suuremat riski inimeste privaatsusele. Lävend eriliigiliste andmete kasutamiseks teadus- ja statistikauuringute kontekstis on suurem, kui mitte eriliigiliste isikuandmete puhul.

³⁹ Omniva esindaja 09.09.21 vastuskiri

4.2. 'Tarkade sõidukite' asukohaandmed

2020 seisuga oli müüdi maailmas hinnanguliselt 30 miljonit sõidukit, mis omasid andmevahetuse võimekust. Nõ 'tarkade' sõiduautode osakaal uute autode müügis kasvab iga-aastaselt.⁴⁰ Lisaks turvalisuse ja juhiabi funktsioonidele ning juhi tervise, väsimuse jmt näitajate kohta koguvad uued sõidukid erinevatel viisidel andmeid ka sõiduvahendi asukoha kohta. Seejuures on andmed sõiduki asukoha kohta üks põhilisi andmestikke, mida autotootjate näitel tarkade sõidukite puhul alati kogutakse.⁴¹ Asukohaandmed võivad talletuda sõiduki tarkvaras, olles ligipääsetavad vaid kaabli vahendusel, aga samuti võib täpseid GPS põhiseid asukohaandmeid koguda sõidukisse sisseehitatud navigatsiooni funktsionaalsuse tarkvara. Kuna andmevahetus toimub üldjuhul üle mobiilsidevõrgu, on sõidukites ka vastav SIM-kaart, mille vahendusel sõiduk ise internetti ühendub või ka kasutajatele internetiühendust pakub. Just SIM-kaardi olemasolu muudab targad sõidukid huvipakkuvaks andmeallikaks ka rahvastiku mobiilsuse uurimise seisukohast.⁴²

Targa sõiduki asukohaandmetele võib erineval kujul olla ligipääs paljudele osapooltele (alates autotootjast kuni kasutaja valitud kindlustus- või muu teenusepakkujani). Andmestiku kvaliteeti ja ulatust silmas pidades oleks statistika- ning teadusuuringute seisukohast mõistlik uurida selliste andmestike eristamist, kasutatavust ning kättesaadavust sideoperaatorite tasemel. Sideoperaatoritel tasemel tekib andmestik erinevate sidemastide piirkonnas sideühendust kasutanud tarkade sõidukite osas. Samas ei ole hetkel teada, kas ja kuidas ning millise kvaliteediga tarkade sõidukite sideühenduse pinnalt sõidukite asukohaga seotud andmestik tegelikkuses on. Lisaks on sõidukeid ning nende mobiilsust on raskem seostada ühe isikuga, kuna erinevalt telefonist sõiduki kasutajad varieeruvad alates ühest püsivast omanikust kuni peresiseste kasutajate ning ühekordsete kasutajateni. Kuna sisseehitatud SIM-kaardi puhul ei pruugi sõiduki müügi korral alati toimuda ka SIM-kaardi vahetust, on keeruline eristada ka andmestikuga seotud isiku vahetumist, mida ekslikult võib pidada elukoha muutuseks.



Source: GAO. | GAO-17-656

Joonis 3. Mobiilivõrku ühendatud auto andmeside põhimõte ([link](#))

4.3. Elektrienergia tarbimise andmed

Elektri tarbimise andmeid elukoha peegliks on analüüsinud ka Eesti Statistika. Täpsemalt uuriti, ka elektritarbimise abil on võimalik tuvastada asustamata ja hooajaliselt kasutatavaid eluruumi. Viiteid projekti kohta võis ja võib leida mitmelt poolt (nt [link](#)). Paraku pole käesoleva projekti ajal selle kohaseid viiteid Eesti Statistika kodulehelt enam näha (ei eksperimentaalstatistika käimas olevate projektide ega ka valminud tööde sektsioonist).

⁴⁰ <https://smartcar.com/blog/connected-cars-worldwide/>

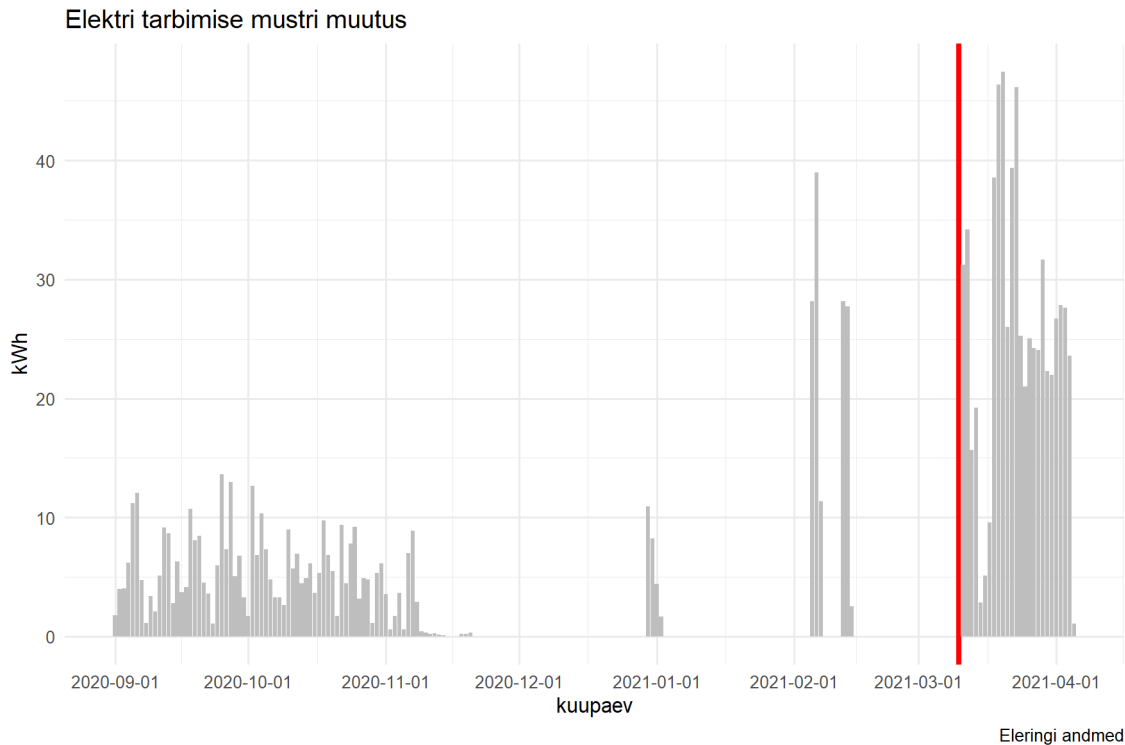
⁴¹ <https://www.gao.gov/assets/gao-17-656.pdf>, lk 17

⁴² https://fpf.org/wp-content/uploads/2017/06/2017_0627-FPF-Connected-Car-Infographic-Version-1.0.pdf

Käesoleva RITA projekti raames päriti vastavaid andmeid Eleringilt uuesti. Paraku ei õnnestunud ligi pääseda täpsematele andmetele, kui Omniva sihtnumbri täpsusele agregeeritud elektri tarbimise tunniandmed perioodist 1. jaanuar – 31. märts 2020. Postiindeksite taset kui suurimat võimalikku täpsust andmete väljastamisel põhjendas Elering andmekaitsete piirangutega. See tähendab, et andmed saadi ca 4000 ruumiüksuse kohta. Elering eemaldas andmetest postiindeksid, kus oli vähem kui 10 tarbijat. Lisaks eemaldati suurtarbijate (tarbimisvõimsus > 8 MW ja aastane tarbimine > 40 GWh) elektritarbimise info, mille põhjuseks toodi börsisaladuse kaitsmise vajadus. Peale selle ilmnes, et väljastatud andmete sihtnumbritest oluline hulk ei kattunud selle perioodi ametlike sihtnumbritest andmebaasiga ([Omniva link](#); [Maa-ameti link](#)). Andmete täpsustamise võimaluste uurimisel, sai teatavaks nõ „võta-või-jäta“-olukord ja täpsematele andmetele ligipääsu ei võimaldatud. Samas on käesoleva projekti raames läbi viidud vestlustest selgunud, et Eesti Statistikal väljastati elektri tarbimise andmed hoone täpsusega. Seega võib väita, et taas on tegemist hea näitega valdkonnast, kus andmete väljastamise reeglid on määratlemata ning standardid puuduvad. Sellistel juhtudel on tulemuseks üksiku inimese kõhutunde pealt tehtud otsused, mille tulemusena väljastatakse erineva sisuga andmeid. Ruumianalüüsi seisukohast võetuna on väljastatavate andmete ruumiüksuse kirjeldamiseks kasutatav sihtnumber leebelt öeldes ebatraditsiooniline ning suuresti ka sobimatu.

Eeltoodust lähtuvalt on selge, et sellise ruumilise ebatäpsusega andmeid pole võimalik kasutada asustamata ja sesoonselt kasutatavate eluruumide tuvastamiseks. Ka sattus andmete väljavõtte periood COVID-19 esimese laine aega, mis tähendab, et inimeste paiknemine erines tavapärasest oluliselt. Samas püstitati hüpotees, et elektri tarbimise andmed (ka sihtnumbri tasemele agregeeritud) võivad olla väga heaks indikaatoriks rahvastiku kiire ümberpaiknemise tuvastamiseks ja mahtude hindamiseks. See info on väärtuslik mitte niivõrd püsiva elukoha selgitamiseks kuivõrd *de facto* rahvastiku paiknemise kohta.

Samas võib mingi maja/korteri elektritarbimine kiiresti muutuda ka näiteks uute elektriseadmete paigaldamise tulemusena. Nii näiteks on õhksoojuspumpade paigaldajad väitnud, et COVID-19 tulenenult kasvas märgatavalt maakodudesse paigaldatavate õhksoojuspumpade hulk (Parimkliima müügiesindaja, isiklik kontakt, märts 2021). See võib aga oluliselt muuta ka elektritarbimise mustrit (Joonis 4).



Joonis 4. Elektrienergia tarbimise muutus õhksoojuspumba paigaldamise (punane joon) tulemusena.

Nagu eelnevalt jooniselt näha võib elektrienergia tarbimise muster pärast uue seadme paigaldamist oluliselt muutuda. Soov maja nõ igaks juhuks soojana hoida võib jätta mulje, et tegemist on püsielukohaga. Lisaks on elektrienergiat kütmiseks kasutamise puhul selge sesoonne muster, mis võib omakorda moonutada tegeliku pildi, kas eluruumis elatakse või mitte.

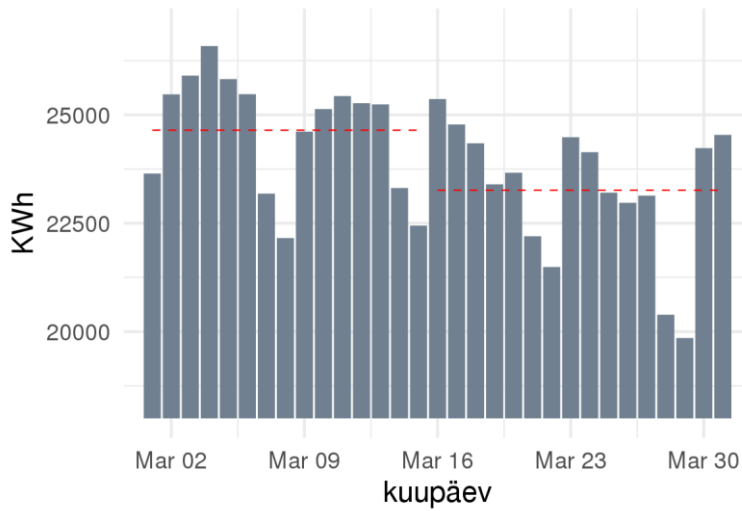
Samas võib elektritarbimise ööpäevase mustri muutus olla väga heaks indikaatoriks inimeste kiire ümberpaiknemise tingimustes. Seda võimalikku näidet käsitletakse järgnevas alapeatükis 2020. aasta märtsi näitel, kui vabariigi valitsus kehtestas seoses uue koroonaviiruse epideemiaga välja eriolukorra.

4.3.1. Eriolukorra mõju rahvastiku paiknemisele Elering'i elektritarbimise andmete näitel

Märtsi teises pooles vähenes kogu Eesti elektritarbimine võrreldes esimese poolega 6%. Seda ei põhjendanud kevade soojem ilm, sest märtsi esimene pool oli isegi veidi soojem kui teine pool (+2,3°C vs +1,8°C). Tarbimise vähenemist põhjustas asutuste (teenindus, haridus jmt), aga ka tootmisettevõtete sulgemine.

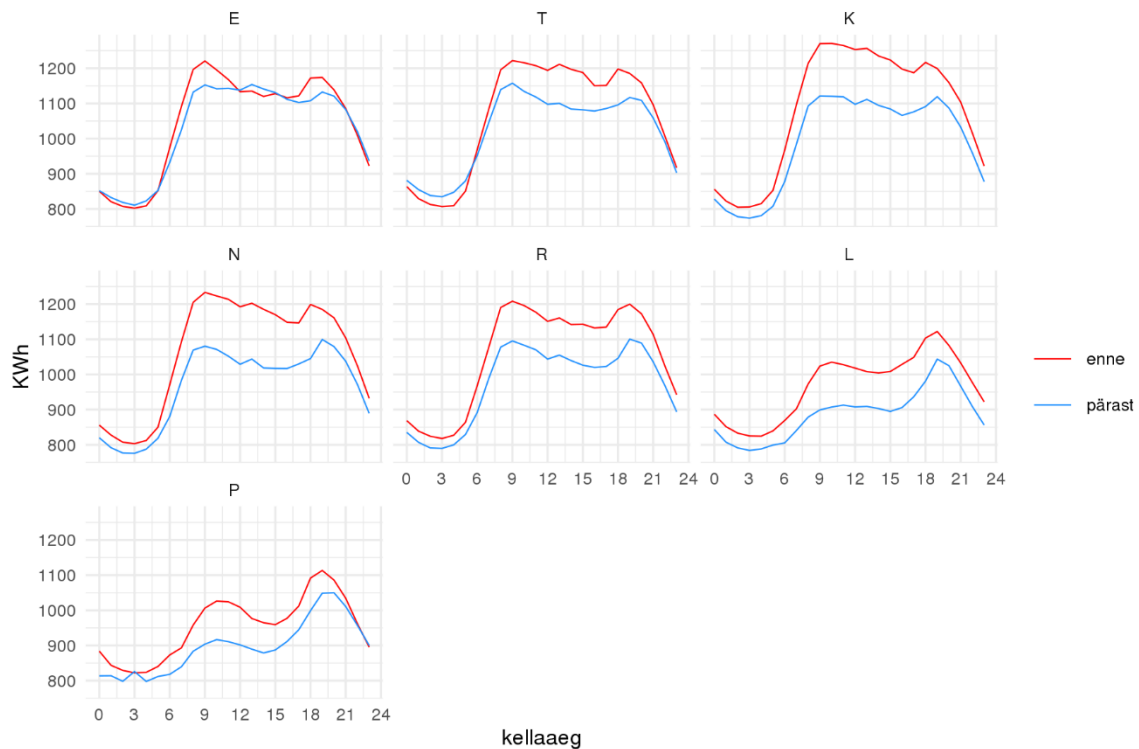
Üldine ööpäevane elektri tarbimine langes Eestis märtsi esimese (1.-15. III) ja teise poole (16.-31.III) võrdluses 6%. Vastavalt 24 647 KWh pealt 23262 KWh peale (Joonis 5). Elektri tarbimises ilmneb selge nädalane rütm: nädalavahetusel on tarbimine selgelt madalam kui tööpäevadel. Eriolukorra kehtestamise järel langes tarbimine kõigile nädalapäevadel suhteliselt ühtlaselt.

Elektri tarbimine Eestis märts 2020

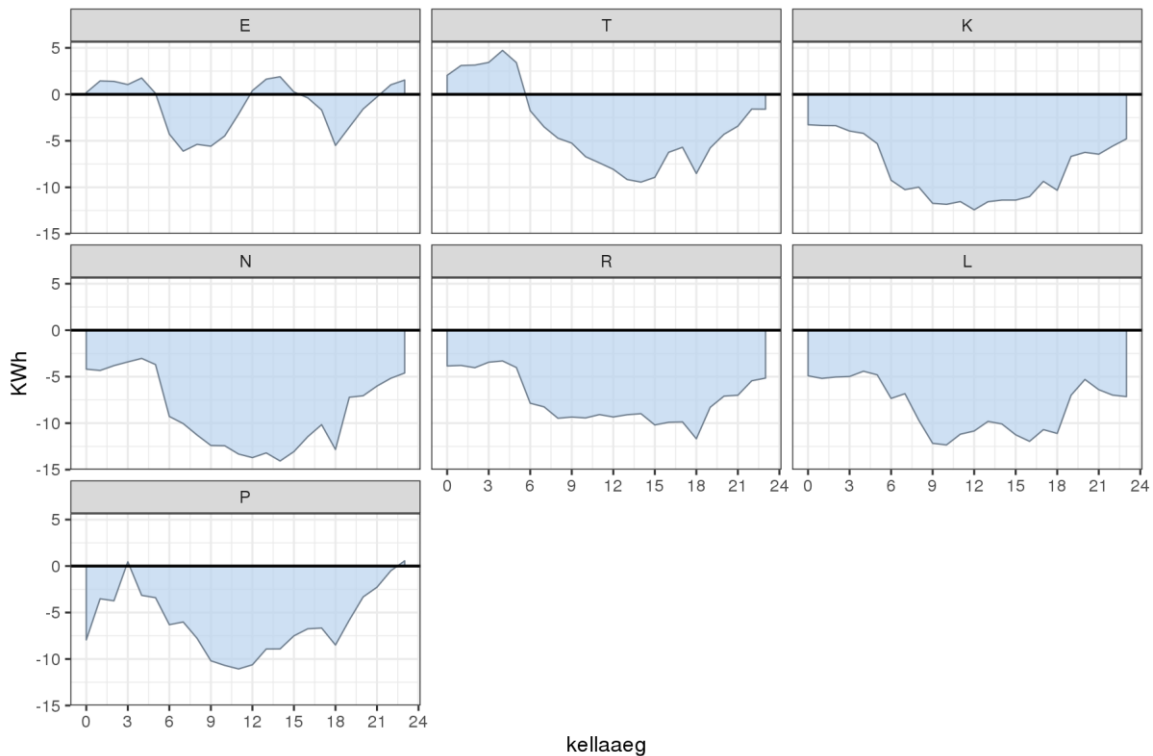


Joonis 5. Elektri ööpäevane tarbimine Eestis märtsis 2020 Eleringi andmetel.

Vaadates tarbimise muutust ööpäeva lõikes nädalapäevade keskmisena (Joonis 6), näeme, et kõige väiksem on olnud tarbimise langus esmaspäevade võrdluses. Langus on selgem hommikuti ja õhtuti, 9:00 kuni 18:00 erinevused peaaegu puuduvad. Teistel nädalapäevadel on näha tarbimise langus eriolukorra järel peaaegu kogu ööpäeva lõikes. Erinevused on väiksemad öötundidel (Joonis 7).



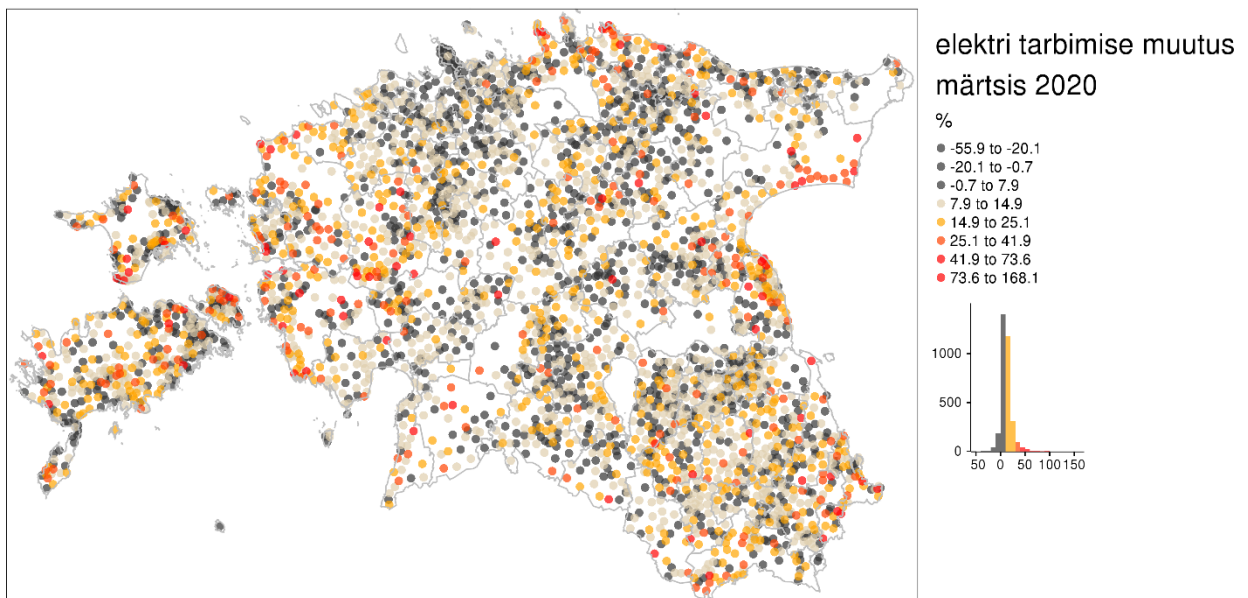
Joonis 6. Elektrienergia ööpäevase tarbimine rütm enne ja pärast eriolukorra kehtima hakkamist.



Joonis 7. Elektri tarbimise muutus Eestis märtsis 2020 perioodide 1.-15 ja 16.-31 märts võrdlusena.

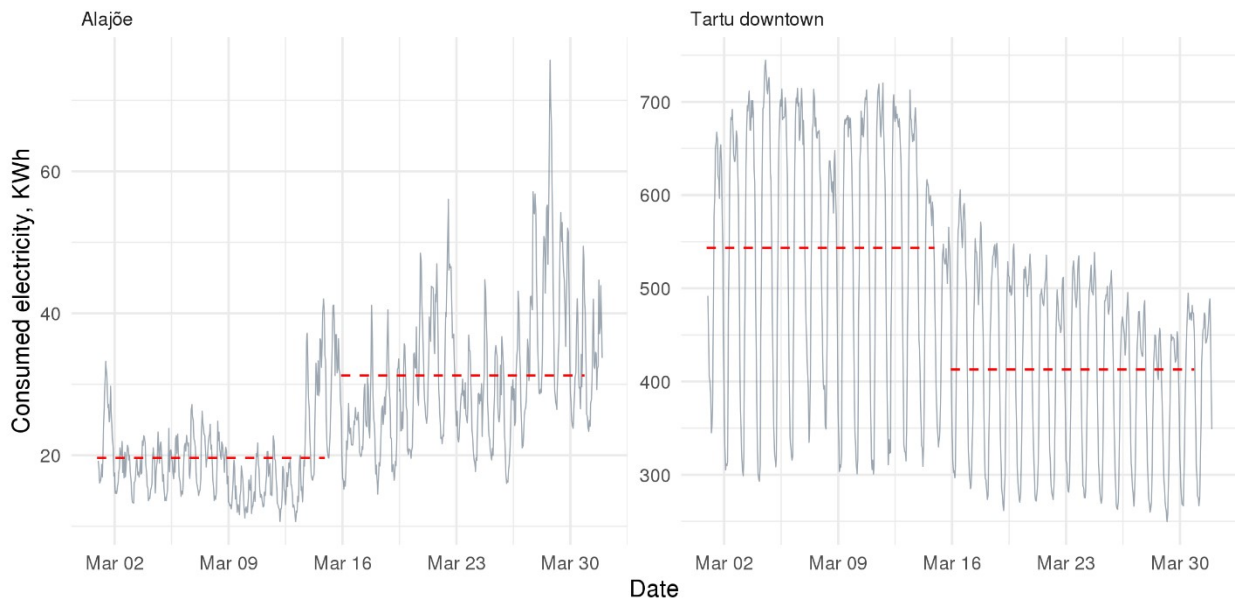
4.3.2. Elektritarbimise muutuse ruumiline muster

Muutuste paremaks esile toomiseks on kasutatud protsentuaalset muutust: kui palju muutus elektri tarbimine pärast eriolukorra kehtestamist (Joonis 8)?



Joonis 8. Elektri tarbimise protsentuaalne muutus asustusüksuse keskmisena

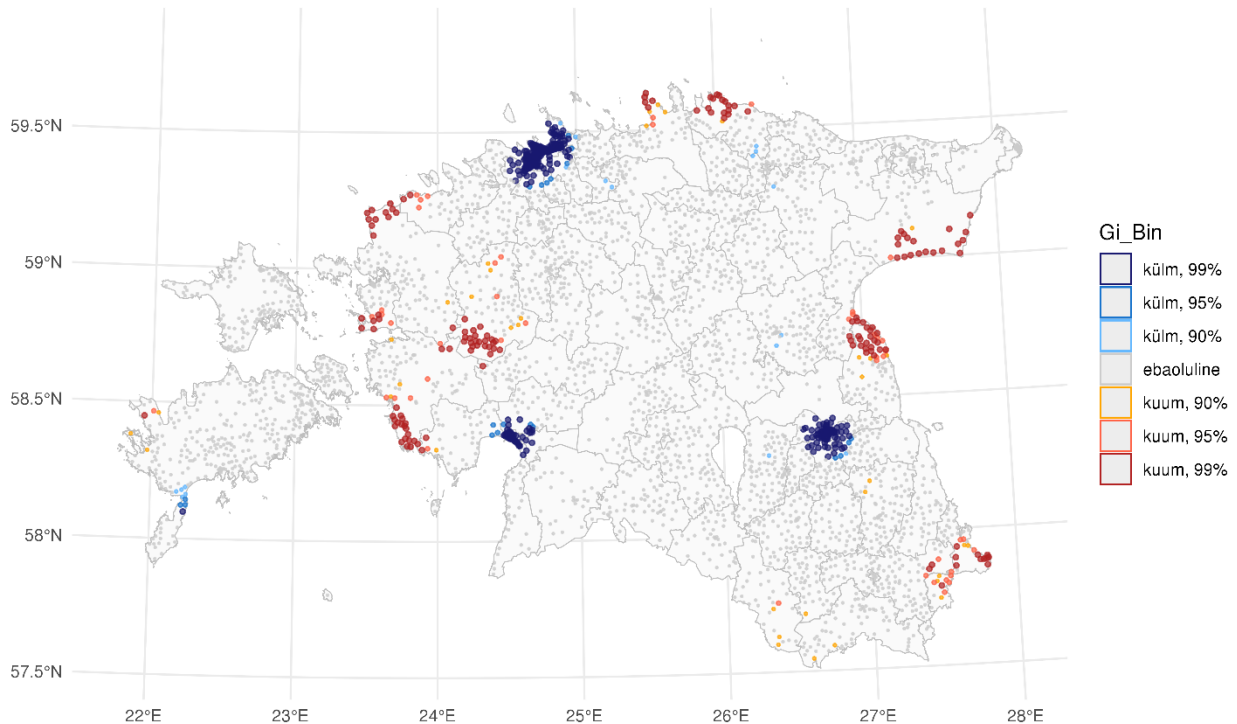
Ruumiline muster on üsna kirju. Siiski on näha, et teatud piirkondades on võimalik täheldada elektri tarbimise tõusu (Peipsi põhja- ja läänerannik, Lahemaa jm) samal ajal kui teistes piirkondades on elektri tarbimine langenud (linnalsed alad). Elektritarbimise järsu muutuse illustreerimiseks on valitud Tartu kesklinn ja Peipsi põhjaranniku suvitusaalal asuv Alajõe (Joonis 9). Et muutuse põhjuseks oli eriolukorra kehtestamine, kinnitab ka konkreetsete postindeksite võrdlus.



Joonis 9. Elektri tarbimise dünaamika ja trend märtsis 2020 Alajõe ja Tartu kesklinna võrdlusena.

Muutuste selgemaks esitamiseks rakendati *hot spot* (kuumkohtade) analüüsi (täpsemalt: *Optimized Hot Spot Analysis*). Analüüsi tulemusena leitakse Getis-Ord G_i^* statistiku abil statistiliselt olulised ruumilised klastrid: kuumkohtades on valdavaks positiivsed väärtused (antud juhul elektri tarbimise tõus) ja külmpohtades valdavad negatiivsed väärtused (tarbimise langus). Klastritesse kuuluvuse tugevus arvutatakse kolmel erineval olulisuse nivool (99%, 95% ja 90%). Analüüsi rakendati nii postiindeksite kaupa (Joonis 10). Postiindeksite kaupa läbi viidud analüüsi tulemusena võib väita, et langus oli tugevam Tallinna, Tartu ja Pärnu kandis. Tarbimise suurem tõus oli iseloomulik Lahemaa, Peipsi põhjarannikule, Mustvee - Kallaste vahel, Eesti kagunurgas, Lääne-Eestis Varbla ümbrusele, Matsalu lahe põhjakaldale ning Vihterpalu ümbrusele ja Vigala vallale. Sõrve poolsaare tarbimise langust näitav klaster pole tarbimismahude pooldest ilmselt märkimisväärne.

Elektri tarbimise muutuse Hot spot (kuumkohtade) analüüs andmed: Elering



Joonis 10. Sarnase tarbimise muutusega klastrid Hot spot (kuumkoht) analüüsi tulemusena postindeksite kaupa.

Kokkuvõtvalt võib tõdeda, et märtsis 2020 muutusid oluliselt nii inimeste paiknemise kui ka elektri tarbimise muustrid. Energia tarbimine langes teenuseid pakkuvates keskustes (eelkõige suuremate linnade meelelahutuse ja kaubanduse piirkonnad) ja tootmisaladel. Samas kui näiteks magalates (nt Tallinna Lasnamäe) energia tarbimine pisut kasvas. Reljeefselt tuleb aga välja traditsiooniliste suvituspriikondade elektrienergia tarbimise kasv.

Kahjuks ei võimaldanud andmete puudumine testida neid muutuseid üksikute tarbimiskohtade täpsusega. Siiski võib väita, et elektri tarbimise muster võib olla üheks abiandmete kihiks mobiilpositsioneerimise andmete põhjal rahvastiku paiknemise ruumiliseks interpoleerimiseks. Seda käsitletakse ja testitakse täpsemalt tegevuste 2.2 ja 2.3 raames.

4.4. Interneti püsiühendused eluruumide asustatuse peeglina

Esmane hüpotees interneti püsiühenduste andmestiku kasutamisel lähtus faktist, et Eestis on interneti kasutajate hulk väga kõrge (2020 aastal 98%), lastega leibkondadest on kodus internet ([link](#)). Seega võiks justkui eeldada, et ka enamuse eluhoonete puhul on olemas internetiühendus. Arvestades, et püsiühendus on parema kvaliteedi, kiiruse ja hinnaga, võiks omakorda eeldada, et inimesed eelistavad püsiühendust ja mitte mobiilset interneti. Seega püstitati hüpotees, et eluruumide asustatust on võimalik hinnata Interneti püsiühenduste andmestiku abil.

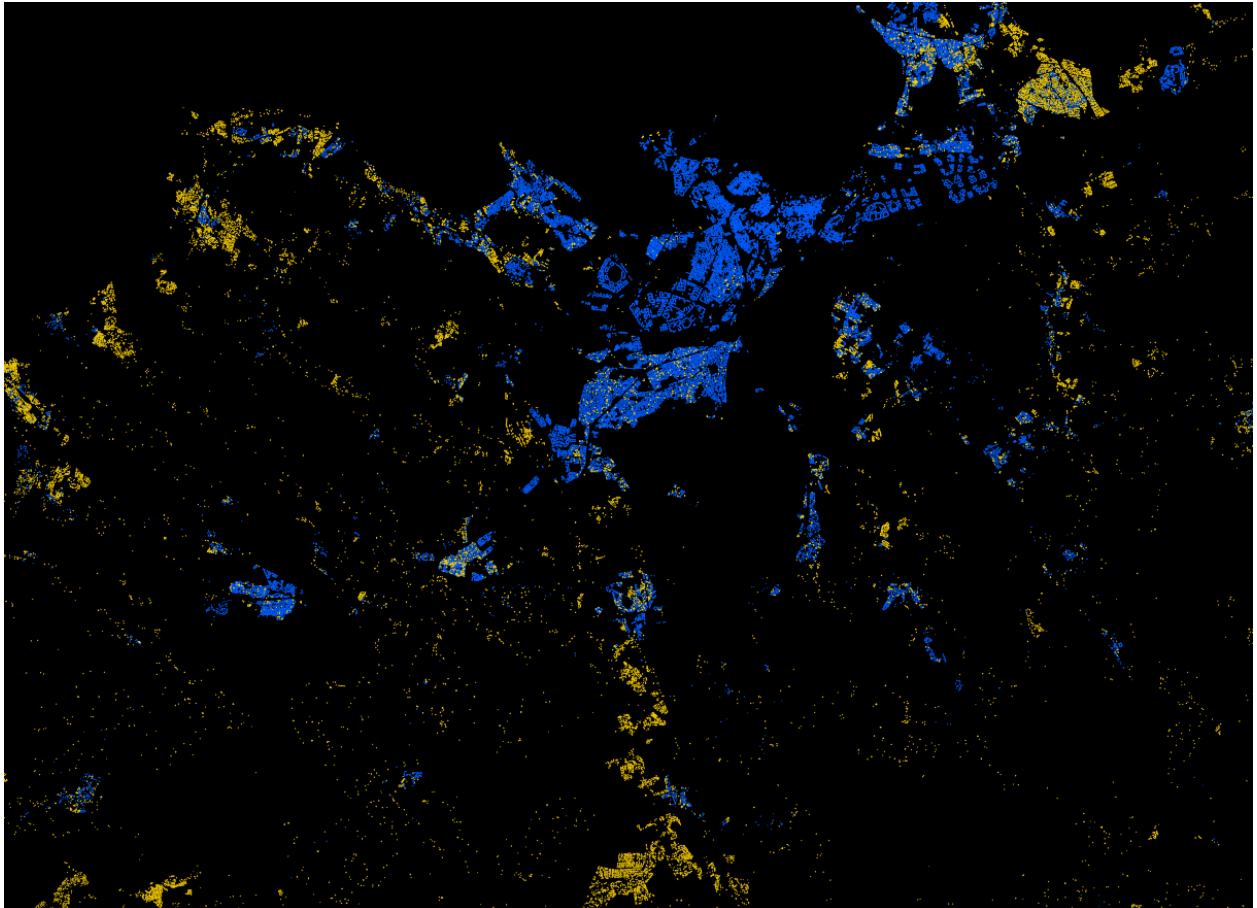
Interneti püsiühenduste andmed päriti Tarbija ja Tehnilise Järelevalve Ametist (tavakasutaja jaoks kättesaadav veebilehena: <https://www.netikaart.ee/tsaApp/>). Andmestiku fragment on esitatud (Tabel 1).

Tabel 1. Fragment interneti püsiühenduste tabelist.

ETAK_ID	ADS_OID	TÄISAADDRESS	VIITEPUNKT	VIITEPUN_1	Max_fix_up	Max_fix_do
607753	EE00972920	Harju maakond, Keila linn, Pargi tn 2b	523306	6574787	50	500
7802452	EE03657924	Harju maakond, Keila linn, Jõe tn 30	524066	6574227	0	0
613306	EE01159044	Harju maakond, Keila linn, Jõe tn 7a	524529	6574175	1	16
5342393	EE01659507	Harju maakond, Harku vald, Tabasalu alevik, Sarapuu tn 70	531998	6587188	0	0
613516	EE01148209	Harju maakond, Keila linn, Metsa tn 12	523428	6574705	5	20
614893	EE01152958	Harju maakond, Keila linn, Vahtra tn 12	523862	6573776	10	40
738154	EE02711419	Saare maakond, Muhu vald, Liiva küla, Lepiku	455513	6496811	0	0
173096	EE02717732	Saare maakond, Muhu vald, Piiri küla, Metsa	452803	6496009	0	0
175022	EE03808805	Saare maakond, Muhu vald, Kallaste küla, Uuesaadu	456271	6503082	0	0
173780	EE00805519	Saare maakond, Muhu vald, Hellamaa küla, Antsu	460583	6495980	10	40
174354	EE00805880	Saare maakond, Muhu vald, Külasema küla, Kadaga	453081	6498639	0	0
173693	EE00805536	Saare maakond, Muhu vald, Ridasi küla, Vesiaia	454314	6494057	0	0
168788	EE00818174	Saare maakond, Muhu vald, Võlla küla, Jüri	462535	6495391	0	0
5698011	EE01875895	Saare maakond, Muhu vald, Võlla küla, Saadu	461159	6494542	1	12
170026	EE00806141	Saare maakond, Muhu vald, Võlla küla, Mardi	461643	6495247	1	12

Andmed on seisuga 21. detsember 2019. Baasis on üle 326 tuhande hoone. Andmed on esitatud hoone, mitte aga eluruumi (nt korter) täpsusega. Tabelis on talletatud Eesti Topograafilise Andmekogu (ETAK) hoone kood, Eesti Aadress-süsteemi kood (ADS-OID), tekstiline täisaadress, koordinaadid LEST-97 süsteemis, maksimaalne üles- ja allalaadimise kiirus.

Interneti ühenduste andmestikust säilitati kõik need hooned, milles vastavalt Ehitisregistri andmete oli eluruumide pindala suurem kui 0. Selle lõike tulemusena jäi alles 239 921 hoonet. Kui vaatame, millisel osal nendest hoonetest on olemas interneti püsiühendus, siis on pilt nukker nii e-riigi seisukohast vaadates kui ka käesoleva projekti raames püstitatud lootusest, et vastav andmestik on hea allikas püsieluruumide määramiseks: 57% eluhoonetest puudub Interneti püsiühendus. Ilmselt on tegelikkuses olukord pisut parem, kuna internetiandmed on pärit alles 2019. aasta lõpust ning vahepeal on Internetti ühendatud hulk hooneid. Lootust andmestiku paranemiseks annab ka EstWin kiire interneti projekt, mille raames on alustatud ka ühenduste lõpptarbijani viimisega (nt [Eesti Energia](#)). Seega on tegemist loodetavasti juba lähitulevikus väga hea andmestikuga eluruumide asustatuse määramisel. See eelduse tugineb siiski ka sellele, et ehitisregistri andmed eluruumide kohta on väga kõrge kvaliteediga. Fragment eluruumidest, mis on interneti püsiühendusega ja mis mitte on esitatud Joonis 11.



Joonis 11. Fragment internet püsiühendusegs eluruumide (sinine) paiknemisest Tallinnas ja selle ümbruses.

Kuigi Interneti püsiühenduste andmestik pole vähemalt täna veel kasutatav eluruumi asustatuse määramiseks, siis lähiaastatel on see võrk kiiresti laienemas ning võivad avaneda uued perspektiivid eluruumide asustatuse määramisel.

Vaatamata Internetiandmete ebatäiusele testitakse nii selle kui ka analüüsi kaasproduktina tekkinud ehitisregistri eluhoonete andmestiku kasutatavust mobiilpositsioneerimise andmete ruumilise interpoleerimise abiandmete kihina tegevuste 2.2 ja 2.3 raames.

4.5. Avaandmed

Avaandmed on tasuta ligipääsetavad andmed, mis on kõigile antud vabalt kasutamiseks, taaskasutamiseks ja jagamiseks, mida saab kasutada äriliste ja mittetulunduslike ettevõtmiste käivitamiseks, uuringute läbiviimiseks ning andmepõhiste otsuste tegemiseks. Avaliku sektori andmetöötajal on kohustus avaldada enda poolt töödeldavad avaandmed. Avaandmete kontseptsiooni üldisem eesmärk on olemasolevate andmete kasutamisega luua uusi väärtusi, suurendada innovatsiooni ning 'saada kasu digitaalsetest dividendidest'. Avaandmetel puuduvad kasutamist takistavad piirangud, kuid teave, millele kohalduvad juurdepääsupiirangud või mille kasutamine on muul viisil reguleeritud, ei kvalifitseeru avaandmeteks. Näiteks isikuandmed ei ole avaandmed, kuivõrd nende kasutamisele laienevad isikuandmete kaitse regulatsioonist tulenevad piirangud.

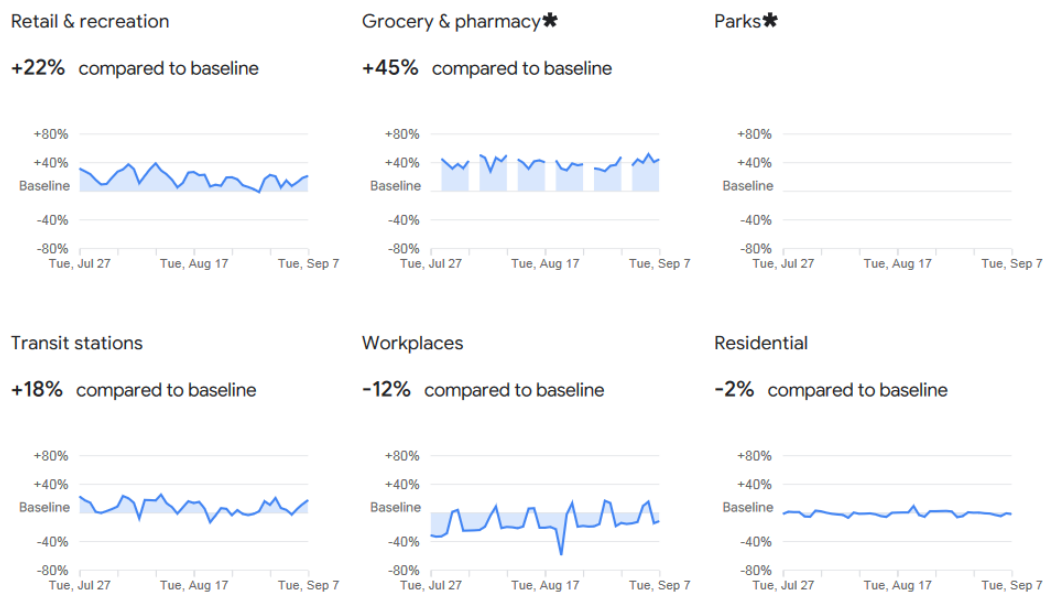
Eesti avaandmete portaalis on 2021 septembri seisuga viidatud 717 andmestikule ja 2206 teabevaldajale.⁴³ Näiteks on avaandmete näol igal huvilisel ligipääs Tallinna ühistranspordivahendite

⁴³ <https://avaandmed.eesti.ee/>

asukohtadele reaajas (<https://avaandmed.tallinn.ee/nimistu?id=35>) kui ka Terviseameti registrile, mis annab ülevaate perearstide nimistutest ning nendes täituvusest ja vabade kohtade olemasolust (<https://mveeb.sm.ee/Nimistud/?>).

Eesti sideoperaatorid on sisuliselt avaandmetena avaldanud agregeeritud asukohtaandmeid, kus veebipõhise rakenduse kaudu on võimalik seadistada ka lihtsamaid parameetreid. Agregeerimise aluseks olevaid andmeid ei saa sideoperaatorid avaldada. Ka Google poolt avaldatav statistika inimeste mobiilsuse kohta toimub piiratud tingimustes. Google poliitika kohaselt kasutatakse andmete anonümiseerimiseks nii agregeerimist kui ka 'müra' lisamist. Statistika aluseks on seadmed, millega ühendatud Google kontod lubavad asukohtaandmete kasutamist. Valimi piirangute tõttu ei kuvataks andmeid valdavalt suuremate linnade osas. Väiksemate asulate või maakondade lõikes märgitakse, et puuduvad andmed või neid ei saa avalda privaatsuskriteeriumite tõttu (Joonis 12). Seetõttu on meetodikate arendamine, katsetamine või täiendava statistika tegemine erasektori poolt avaldatud avaandmete pinnalt piiratud.

Ida-Viru County



* The data doesn't meet quality and privacy thresholds for every day in the chart.

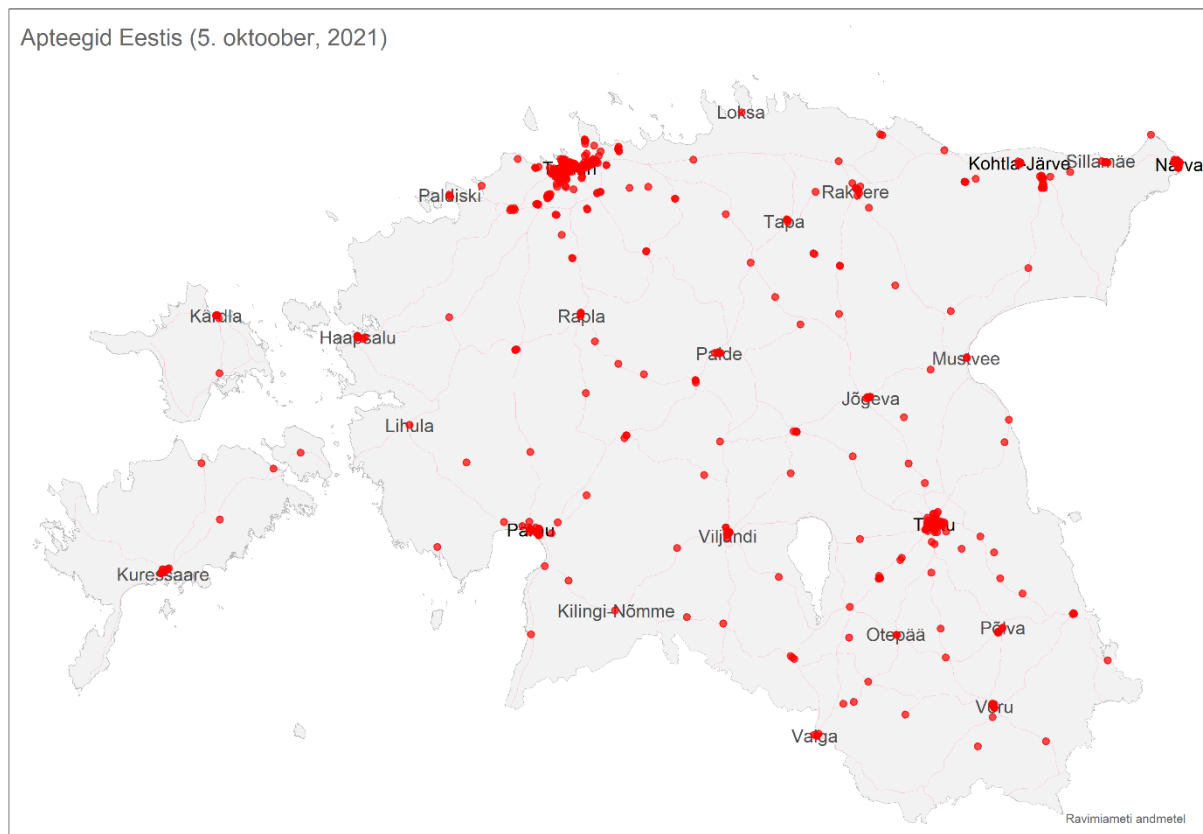
Joonis 12. Google liikuvuse avaandmete näidis.

Euroopa andmestrateegia toob ühtse andmeruumi ühe olulise aspektina esile andmed liikuvuse kohta. Tänapäeval tekib juurde ühe enam sõidukeid, millel on andmesideühendus, samuti tekib andmeid teiste liikumisvahendite tekkedest ja kasutamisest (nt liinibusside liikumine, täituvus, renditavad elektrirattad, takso- või sõidujagamisteenuste kasutamise andmed). Kui kogutavad andmed liigituvad avaandmeteks ning ka algandmetele tekib ligipääs, võib neid andmeid kasutada erinevate statistiliste meetodikate täiendamiseks, testimiseks või täiesti uute väljatöötamiseks.

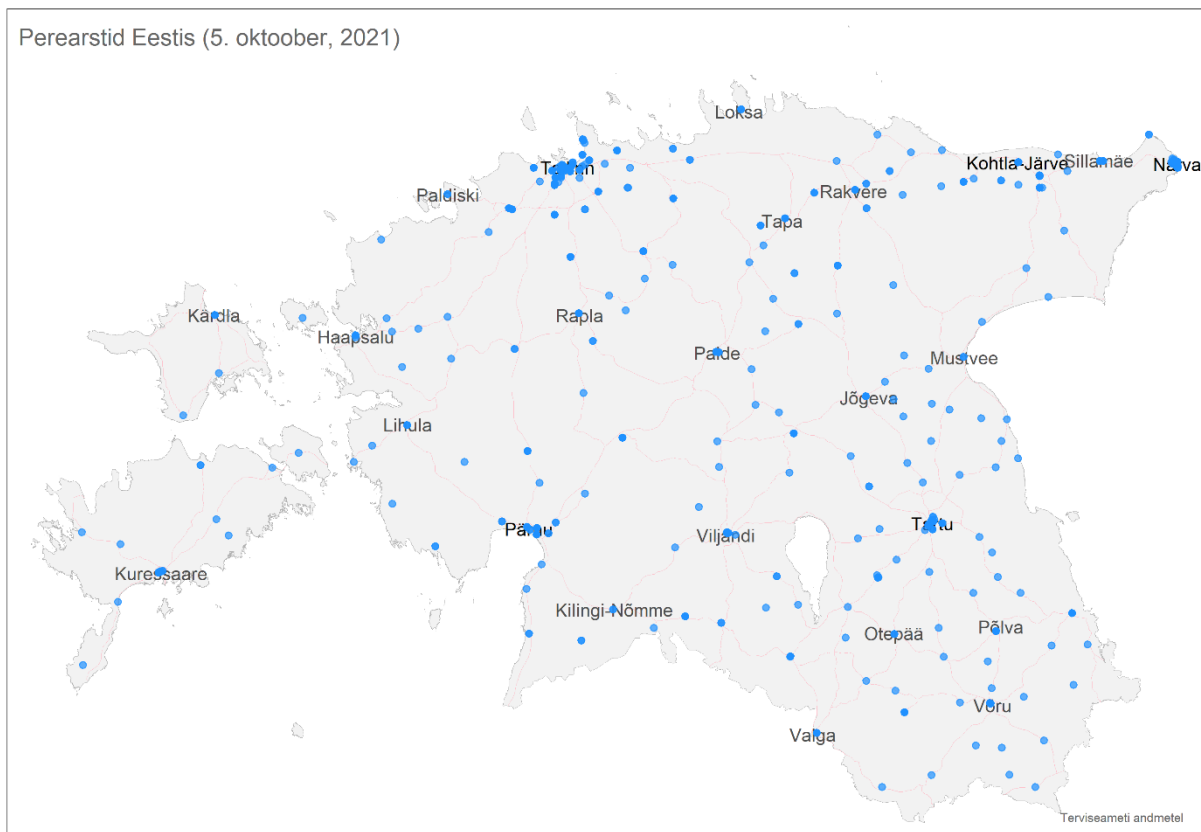
4.6. Perearsti nimistute potentsiaal

Avaandmete näol kättesaadavad perearsti nimistute suurused ning asukohad võivad teatud juhtudel viidata populatsiooni paiknemisele, eeldusel, et inimene võib olla reeglina perearsti nimistus, mille piirkond hõlmab inimese elukohta. Esmaste perearsti registreeritud asukohtade pinnalt võib teha vaid üldistatud järeldusi piirkonna asustustihedusega seoses (suurema rahvastikuga piirkondades on

perearste rohkem). Täpsemate järeltuste tegemine avaandmete pinnalt ilma täiendavaid andmestikke kasutamata on keeruline. Võrreldes mobiilpositsioneerimise andmetega puudub perearsti ja apteekide andmestike sidumisel teiste andmestikega ka niivõrd selge potentsiaal inimeste mobiilsuse ning elukoha määratlemise seisukohast (Joonis 13 ja Joonis 14). Ühest küljest võib püsiva elukoha täpsustamise puhul vaadelda, millise perearsti nimistusse kuulatakse või millisest apteegist ostetakse retseptiravimeid. Teisalt koguvad perearstid ka inimeste kontaktaadresse. Mõlema tervisteenuse ruumilise paiknemise muster on liiga hõre, et nende kohtade külastamine viitaks inimese püsielukoha lähedusele.



Joonis 13. Apteekide paiknemine Eestis.



Joonis 14. Perearstide vastuvõtukohtade paiknemine Eestis.

Kuna terviseandmeid sisaldavad registrid on teistest käesolevas analüüsis käsitletud andmetest veelgi tundlikuma iseloomuga, siis on nende kaasamine inimeste püsiva elukoha selgitamiseks veel ebatõenäolisem, kui teiste käesolevas uuringus analüüsitud andmestike puhul. Samuti puudub põhjenduslik alus, et just perearsti juures registreeritud aadress viitab inimese püsielukohale.

5. Andmetötluse õiguslik käsitlus

5.1. Statistika- ning teadusuuringutes isikuandmete töötlemise üldised õiguslikud alused

Statistika- ja teadusuuringute käigus toimuv isikuandmete töötlemine on reguleeritud isikuandmete kaitse üldmäärusega (IKÜM) kui ka siseriiklike õigusaktidega. Üldiselt saab statistika- ning teadusuuringute raames isikuandmeid töödelda mitmel IKÜM-s sätestatud õiguslikul alusel, sõltuvalt uuringust ning selle läbiviijast.

Esiteks võib andmete töötlemine toimuda isiku nõusoleku alusel isikuandmete kaitse üldmääruse artikkel 6 lg 1 punkt a tähenduses. Sellisel juhul annab isik eelnevalt selge nõusoleku oma andmete kasutamiseks. Kuid statistika- või teadustöö teostamine isikutelt kogutavate nõusolekute alusel andmete kogumine on mõeldav vaid väiksema mahu projekti korral. Nõusolekute kogumine ja haldamine olla ressursikulukas ning seda on keeruline teha suurte mahtude korral. Kui seda peab tegema põhitegevuse kõrvalt, võib kannatada statistika- või teadustöö enda läbiviimine. Samas ei pruugi väike andmeandjate arv tagada piisavat andmehulka ega võrdlusbaasi, et teha kvaliteetseid järeldusi. Kui statistika- või teadustöö eesmärgi saavutamiseks või uurimisküsimuse lahendamiseks on

vajalik koguda andmeid suures mahus ja/või teatud täpsusastmega, siis ei ole nõusolek selleks sobiv õiguslik alus.

Alternatiivselt nõusolekule, võimaldab seadus teatud juhtudel töödelda statistika- või teadustöö tarbeks isikuandmeid ka inimese enda nõusolekuta. Avalik-õiguslik andmetöötaja võib vastavasisulise siseriikliku õigusakti alusel isikuandmeid töödelda avalikes huvides oleva ülesande täitmiseks või oma avaliku võimu teostamiseks. Erasektori andmetöötaja saab tugineda enda õigustatud huvile. Isikuandmete kaitse seadus sätestab täiendavalt, et statistika- või teadusuuringu käigus võib inimese nõusolekuta tema andmeid töödelda vaid juhul, kui samaaegselt on täidetud kolm tingimust:

- 1) pärast tuvastamist võimaldavate andmete eemaldamist ei ole andmetöötluse eesmärgid enam saavutatavad või neid oleks ebamõistlikult raske saavutada;
- 2) teadus- või ajaloouringu või riikliku statistika tegija hinnangul on selleks ülekaalukas avalik huvi;
- 3) töödeldavate isikuandmete põhjal ei muudeta andmesubjekti kohustuste mahtu ega kahjustata muul viisil ülemäära andmesubjekti õigusi.

Eeltoodud õigusnormid kohalduvad juhul, kui tegemist on isikuandmetega. Kui sideoperaator väljastab andmed anonüümsel kujul, ei pea uuringu teostaja tuginema mõnele IKÜM-is õiguslikule alusele ega lähtuma isikuandmete kaitse seaduse täiendavatest kriteeriumidest. Statistikaametil peab lisaks eeltoodule lähtuma riikliku statistika seadusest (RStS), mille vastavad sätted võivad olla ka isikuandmete töötlemise õiguslikuks aluseks IKÜM artikkel 6 lg 1 punkt (e) tähenduses.

5.2. Sidevaldkonna eriregulatsioon

Sideandmete statistika- ja teadustöös kasutamise korral võivad andmetöötluse osapooled olla tihti erinevatel seisukohtadel. Kuigi sideoperaator ei vastuta andmete töötlemise eest uuringu käigus, peab sideoperaator oma tegevuses lähtuma ka valdkonnapõhisest eriregulatsioonist. Siseriiklikult reguleerib sideandmeid elektroonilise side seadus, millega tagatakse Euroopa Liidu direktiivi 2002/58/EÜ rakendamine.

Täna kehtiv elektroonilise side seadus õiguslikult piirab isikuandmete edastamise statistika ja teadusuuringute tarbeks ilma sideoperaatori lõpp-kliendi nõusolekuta. ESS § 102 näeb sideoperaatorile ette kliendiandmete saladuses hoidmise kohustuse, võimaldades neid avaldada üksnes seaduses loetletud juhtudel ning saajatele või kliendile endale ja kliendi nõusolekul ka kolmandale isikule. Teisisõnu saab klient anda volituse alusel andmete edastamiseks kolmandale isikule. Seadus ei konkretiseeri ning puudub ühtne praktika, milliseid andmeid ning millises vormingus saab klient ise nõuda ning milline peab olema sideoperaatori jaoks kehtiv kliendi nõusolek andmete avaldamiseks kolmandatele isikutele. Seetõttu nõuab ka kliendi volituse alusel toimuv andmeedastus eelnevat koostööd sideoperaatori ning uuringuteostaja vahel, et tagada andmete kogumine soovitud ajaperioodi vahemikus ning vältimaks arusaamatusi andmete edastamise lubatavuse ja vorminõuete vastavuse osas.

ESS § 105 sätestab põhimõtte, et kliendi asukohaandmeid võib sideoperaator üksikutest seaduses ettenähtud olukordadest muul eesmärgil töödelda ainult siis, kui need on eelnevalt muudetud anonüümseks. ESS ei kätke asukohaandmete töötlemist sideoperaatorite poolt seoses andmete väljastamisega statistika- või teadustöö eesmärgil. Eelnevast tuleneb, et juhul, kui statistika- või teadustöö teostajal on vajalik töödelda isikuandmeid eesmärgil, mille saavutamiseks on andmete anonümiseerimine välistatud, siis praktikas ei ole võimalik ega mõistlik kasutada muid isikuandmete töötlemise õiguslikke aluseid peale isiku enda nõusoleku.

Viidatud ESS sättes märgitud kliendi nõusolek ei ole samatähenduslik IKÜM nõusoleku kui õigusliku alusena. Samas kätkeb ESS kohase nõusoleku (volituse) küsimine sarnaseid kitsaskohti nagu seda on IKÜM tähenduses nõusoleku küsimine statistika- või teadustöö alusena.

5.3. Eriregulatsiooni muutumine

Euroopa Liidu direktiivi 2002/58/EÜ asemel on väljatöötamisel uus valdkonnaülene määrus. Kuigi määrus on otsekohaldatav, jääb liikmesriikidele võimalus teatud osas määrust täpsustada ning selle kohaldamist mõjutada. Määruse viimane kavand käsitleb ka andmete töötlemist statistika- ja teadustöö vajadusteks. Üldpõhimõtte jääb samaks – kui võimalik, tuleb andmeid töödelda anonüümselt, kuid muudatusena tekkida võimalus töödelda asukoha ja muid andmeid ka anonümiseerimata kujul, kui andmete anonüümseks muutmisega ei saa täita andmetöötamise eesmärki. Asukohaandmete puhul nähakse täiendavate kaitsemeetmena ette pseudonümiseerimine ning lisaks nõue andmed kustutada või muuta anonüümseks, kui vastupidine vajadus ära langeb. Määruse kavand näeb ette ka piirangu, et isikustataval kujul asukohaandmeid töödeldes ei tohi neid kasutada kliendi isiksuse hindamiseks või profileerimiseks.

Eraldi tuuakse esile võimalus kasutada asukoha ning muid andmeid riikliku ja Euroopa Liidu statistika koostamiseks kooskõlas siseriikliku ja liiduõigusega. Kokkuvõtlikult võimaldaks määruse tänasel kujul kehtestamine lahendada õigusliku kitsaskoha andmete saamisel sideoperaatori poolt. Seega võib tulevikus ka regulatsioon toetada sideoperaatoritel näiteks asukohaandmete saamist viisil, et neid on võimalik siduda registriandmetega. Kavandis sätestatud piirangutest tulenevalt vajaks aga eraldi analüüsimist, kas seeläbi oleks asukohaandmeid võimalik siduda ka residentsuse indeksiga, mis võib tähendada määrusega keelatud kliendi profileerimist.

5.4. Siseriiklike seaduste muutmise vajadust tuleb analüüsida

Määruse kehtima hakkamisel tuleks tänane ESS vähemalt määrusega kattavas osas tunnistada kehtetuks. Kuid ESS reguleerib vaid sideoperaatori tegevust ning statistika- ja teadustöö teostaja peab lähtuma ka isikuandmete kaitse seadusest. Õigusselguse poolest võib endiselt probleeme valmistada isikuandmete kaitse seaduse nõue, mille kohaselt peab isikustavate andmete kasutamiseks statistika- või teadusuuringu tegija hinnangul olema selleks ülekaalukas avalik huvi. 'Avalik huvi' ja selle 'ülekaalukus' on määratlemata õiguskäsitlused, mistõttu eeldab nende sisustamine alati teatud annuses subjektiivsust. Mõnel juhul on avalikku huvi ja selle ülekaalukust järeldada lihtsam (nt pandeemia kontekstis), kuid eriala spetsiifilise uuringute korral ei pruugi ülekaalukat avalikku huvi esineda. Seeläbi võib siseriiklik seadusenõue piirata määruse rakendamist. Kuna isikuandmete kaitse seaduse vastavad piirangud on kehtestatud mitte ainult sideandmete töötlemise kontekstis, tuleks hoolikalt analüüsida, kas antud ülekaaluka avaliku huvi nõue jääb asjakohaseks ka edaspidi või võib see põhjendamatult pärssida sideandmete põhiseid statistika- ja teadusuuringuid.

Määrus toob põhimõttena esile, et teatud andmetöötlus võib olla lubatav, kui siseriiklikus õiguses nähakse ette asjakohased kaitsemeetmed. Selles valguses tuleb hinnata, kas isikuandmete kaitse seaduses sätestatud meetmed on piisavad või tuleks määruses sätestatud võimaluste teostamiseks siseriiklikke õigusakte täiendada (nt viidatakse määruses krüpteerimisele, mida siseriiklikult täna eraldi kaitsemeetmena esile ei tooda).

6. Sideandmetega kaasnevate privaatsusküsimuste käsitlemine

Täna kehtivate õiguslike piirangute peamine eesmärk on maandada ohte inimeste privaatsusele, eelkõige andmete konfidentsiaalsuse ning eesmärgipärase kasutamise tagamise teel. Asukohaandmete töötlemisel riikliku statistika ja teadusuuringu eesmärgil kerkivad mitmed privaatsusküsimused.

Sideoperaatori lõppkliendi seisukohast on oluline tagada inimese põhiseaduslikku õigust eraelu puutumatusse. Ühelt poolt tuleb arvestada inimese õigust otsustada, kes ja milliseid tema andmeid töötleb. See õigus ei ole absoluutne, kuid selle mistahes piiramine peab toimuma põhjendatult. Teisalt tuleb arvestada, et inimesega seotud sideandmete väljastamine võib kaasa tuua põhjendamatuid negatiivseid tagajärgi, kui kliendi ja sideoperaatori vahelises suhtes kogutavaid andmeid kasutatakse muude isikute poolt.

Privaatsuse tagamisel on kesksel kohal inimese teadlikkus tema andmete kasutamisest. Vältida tuleb olukorda, kus sideoperaatori klient ei ole tema telefoninumbritega seotud andmete edastamisest statistika- ja teadustööde tarbeks teadlik ning selline andmetöötlus pole talle ootuspärane ega eeldatav. Kitsaskohta saab ületada, kui andmetöötlusest või selle võimalikkusest klienti teavitatakse sideoperaatorit ning statistika koostaja poolt.

Privaatsuse tagamisel on oluline võimalusel anda isikule kontroll temaga seotud andmete kasutamise üle. Kontrolli säilitamine oma andmete töötlemise üle võib toimuda eelneva nõustumise alusel (opt-in) või vastuväite esitamise (opt-out) teel. Mõlemal juhul on eelduseks, et inimene on andmetöötlusest teadlik, sest vastasel juhul ei saa sellega eelnevalt nõustuda ega esitada vastuväidet. Opt-out lahendust on näiteks kasutatud Ühendkuningriikide statistikaasutuse poolt, kui sideoperaatorilt koguti agregeeritud andmeid.

Kui andmeid kogub riik seadusest tuleneva ülesande täitmiseks, ei ole eelneva nõusoleku küsimine asjakohane, kuid sõltuvalt andmetöötluse eesmärgist ning asjaoludest tuleb kaaluda vastuväite esitamise võimaldamist. Autorid peavad siinkohal silmas analoogset lahendust rahvastikuregistrile, kus isik saab esitada vastuväite ning keelata tema andmete kasutamist statistika- ja teadustööde tarbeks. Vastuväite esitamise võimaluse tekitamist tuleb kaaluda, kui kõikide inimeste andmete kogumine pole eesmärgi arvestades vajalik ning eesmärk on võimalik saavutada ka juhul, kui osa inimestest kasutavad seda võimalust.

Sideoperaatori seisukohast jäävad sideoperaatori kanda isikuandmete töötlemisega seotud riskid, sh riskide realiseerumisel märkimisväärsed rahalised- ning mainekahjud. Kui sideoperaatori poolt enne andmete väljastamist andmed anonümiseerida, saab riske vältida või maandada. Ilmekaim näide anonüümsete andmete kasutamisest pärineb 2020 alguses, kui Covid-19 tõi teravalt esile vajaduse saada infot isikute liikumise osas. Statistikaamet koostöös sideoperaatoritega asus analüüsima inimeste liikumist, et mõista 2020.a kevadel väljakuulutatud eriolukorra mõju ning kehtestatud piirangute järgimist. Seejuures ei edastanud sideoperaatorid analüüsi aluseks olevaid algandmeid, vaid teostasid analüüsi igaüks eraldiseisvalt ning edastasid saadud tulemused agregeeritud kujul. Kuid ka anonümiseerimise puhul jääb teatud risk alles, sest anonümiseerimine teostamine võib ebaõnnestuda ning aja jooksul täiendavate andmete tekkimise korral võib anonüümsus kaduda.

Andmete anonümiseerimise praktikat kasutatakse ka globaalsete teenuseosutajate poolt, kellel on ligipääs kasutajate asukohaandmetele. Näiteks teostab inimeste liikuvusanalüüsi ning avalikustab Google, kellel on ligipääs inimeste asukohaandmetele läbi kaasaskantavate seadmete ja tarkvara.

Statistika koostaja ning teadusuuringute läbiviija seisukohast ei pruugi anonüümsete andmete kasutamine täita andmetöötluse eesmärgi. Näiteks, kui statistika koostaja soovib sideandmed

seostada konkreetse indiviidiga, siis anonüümsete andmete korral pole see võimalik (näiteks õiguslikku tähendus omava elukoha määratlemisel). Teadusuuringute puhul pole kindla isiku tuvastamine reeglina vajalik, sest uuringu teostaja eesmärgid ei hõlma isikute suhtes otsuste tegemist. Vaid pikaajalises vahemikus asukohaandmeid vajav teadusuuring eeldab, et andmeid saab siduda sama lõppkliendiga, mistõttu nõuaks see vähemalt sideoperaatori tasemel säilima võimekust sideandmete seostamiseks konkreetse indiviidiga. Kuid ka sellisel juhul ei ole konkreetse indiviidi teadmine teadusuuringu teostajale vajalik, tähtsam on teadmine, et uuringust kasutatavad andmed käivad ühe ja sama inimese kohta.

6.1. Asukohaandmete täpsuse mõju privaatsusele

Mida täpsemat asukohta kasutatakse, seda suurem oht privaatsusele tekib. Passiivse positsioneerimise korral võib asukoha täpsusaste erineda, sõltuvalt piirkonna sidemastide tihedusest. Suurema rahvastikutihedusega piirkondades on rohkem sidemaste, mistõttu masti enda asukohast tulenevalt on võimalik täpsemini määratleda ka lõppseadme asukohta. Hajaasutusega piirkondades, kus sidemaste on vähem ning ühe mastiga kaetakse suuremat ala, on ka asukohaandmete täpsusaste väiksem.

Passiivne positsioneerimine on privaatsussõbralikum võrreldes GPS põhise asukoha fikseerimisega. Sideoperaatorid võimaldavad ka AGPS (Assisted GPS) kasutamist, mis võimaldab lõppkliendil kiiremini fikseerida oma asukoht. AGPS kasutamisel on sidemast eelnevalt tuvastanud satelliidid, mida saab GPS positsioneerimiseks kasutada. Lõppkliendi seade "küsib" need andmed sidemasti vahendusel ega hakka eraldiseisvalt otsima saadaolevaid satelliite. Lisaks võib asukoha täpsuse suurendamiseks kasutada sidemasti ning lõppkliendi seadme signaalitugevust.

6.2. Andmete anonüümsus pärast sideoperaatori poolt väljastamist

Privaatsuse kaitse põhimõtetest lähtuvalt võib andmeid töödelda isikustataval kujul üldreeglina juhul, kui nende töötlemine anonüümsel viisil ei täida töötlemise eesmärke. Kui algselt kogutakse isikustatud andmed ning neid pole mõnel hetkel enam isikustatud kujul tarvis, tuleb andmed anonümiseerida. Andmete anonümiseerimise jaoks tuleb tuvastada, kas ning kui lihtsalt on konkreetset füüsilised isikud andmestiku pinnalt otseselt või kaudselt tuvastatavad. Seeläbi saab hinnata, milliseid anonümiseerimise tehnikaid on vajalik rakendada ning kas ligipääsu andmestikule peaks täiendavalt kaitsma.

Õiguslikult võttes on andmestik anonüümne, kui sellega ei saa seostada ühtegi konkreetset inimest. Praktikas võib tekkida olukord, kus andmed on statistika- ja teadusuuringu teostaja jaoks anonüümsed (st viimane ei saa neid mõistlikke pingutusi tehes siduda ühegi konkreetse isikuga), kuid samaaegselt olla sideoperaatori seisukohast lihtsasti seotavad konkreetsete klientidega. Sellisel juhul on sideoperaator poolt andmete väljastamine siiski käsitatav isikuandmete töötlemisena. Kui ka sideoperaator ei saa andmeid seostada ühegi lõppkliendiga, on tegemist täielikult anonüümse andmestikuga, mille edastamisele ei kohaldu kõik isikuandmete kaitse nõuded.

6.2.1. Kas isikud on tuvastatavad kaudsete tunnuste pinnalt?

Kuna ühe isikuga seotud andmestik võib olla unikaalne, peab arvestama kaudse tunnuste kaudu tuvastamise võimalusega. Mida pikema ajaperioodi kohta andmeid ühe telefoninumbri kohta tekib, seda tõenäolisemalt on nende andmete pinnalt võimalik telefoninumber siduda konkreetse isikuga. Andmete isikustamise tõenäosust saab vähendada anonüümsust suurendavate tehnikate rakendamisega nagu andmete üldistamine ja asendamise teel. Andmete üldistamise korral vähendatakse täpsusastet ning esitatakse andmed agregeeritud kujul. Asendamise eesmärgiks on

tekitada rohkem samade väärtustega andmeid, et vähendada võimalust vähe kasutatud andmete seostamiseks konkreetse üksikisikuga. Praktikas ei pruugi ainult ühe tehnika kasutamine anda soovitud tulemust, kuna tehnikatel on erinevad nõrkused. Seetõttu peaks anonüümsuse tagamiseks hindama erinevaid tehnikate nõrkuseid ning neid vajadusel kombineerima. Andmestiku anonümiseerimine tingib andmestiku kasutusvõimaluste märkimisväärse vähenemise, mistõttu peab eelnevalt hindama anonüümsete andmete kasutamise võimalust teadusuuringu eesmärgi kontekstis.

6.3. Privaatsusriskide käsitlemine krüptograafia arenguga

Täishomomorfse krüptograafia areng võimaldab krüpteeritud andmete pinnalt analüüside teostamist ilma andmeid dekrüpteerimata. Privaatsuse kaitse kontekstis väärib uurimist, kas ning mis määral saab seda koostöös sideoperaatoriga rakendada asukohaandmete osas. Näiteks on huvipakkuv, kas täishomomorfse krüptograafia abil saab analüüsi teostaja jaoks jätta andmed krüpteerituks, kuid samas võimaldades nende analüüsimist. Nii saaks maandada privaatsusriske, kuivõrd analüüsiteostaja seisukohast ei saa krüpteeritud andmeid siduda konkreetsete isikutega. Samuti võib see suurendab andmete kaitstuse taset nende hoidmisel ja liikumisel kolmandate osapoolte volitamata ligipääsu eest. Täishomomorfne krüptograafia võib pakkuda uusi võimalusi longituudsete teadustööde teostamiseks. Samuti võib sellest olla kasu andmestike kvaliteedi ning täpsuse parandamiseks, mida vähendatakse anonüümsuse tagamiseks, kuid tagades seejuures samaväärse kaitse anonümiseeritud andmestikuga.

Teise osapoolte krüpteeritud andmete kasutamine. Täna kättesaadavad tehnoloogilised lahendused võimaldavad vastavate teenusepakujate kasutamise abil ühel osapooltel krüpteeritud viisidel töödelda teiste osapoolte andmeid ilma neid eelnevalt lahti krüpteerimata. Olukorras, kus teise osapoolte andmeid on vaja sisendiks, et nende pinnalt teostada matemaatilisi toiminguid või statistilisi järeldusi, võib selliste tehnoloogiate kasutamine tagada vajaliku privaatsuse isikutele, kelle andmete pinnalt soovitakse statistikat koostada. Erinevad osapooled saavad teostada toimingud ning teha vajalikud järeldused, saamata samas ligipääsu nende aluseks olevatele andmetele, mis võivad olla sideoperaatori klientide sikuandmed. Antud lahendust on kasutatud Indoneesia valitsuse poolt, arvutamaks täpset turismistatistikat tuginedes erinevate mobiilsideoperaatorite andmestikele.

Plokiahela kasutamine. Sõltuvalt andmetöötluse eesmärgist ning seotud osapoolte vajadustest, võib mõnel juhul tulla kõne alla plokiahela loogikale tuginevate krüptograafiliste lahenduste kasutamine. Näiteks kui protsessis on osapoolte jaoks eelkõige oluline andmete õigsus, mitte nende sidumine konkreetse isikuga, st tehniliselt protsessi käigus on osapooltele tarvis teada, et subjektiga seotud andmed on õiged. Sellisel juhul kasutavad andmetöötluse osapooled ainult isikuga seotud ühesuunaliselt krüpteeritud räsikoodi. Kuna seda ei saa tagasi pöörata, ei saa uuesti isikuga seostada, v.a osapool, kelle valduses andmed krüpteeriti. See võib avada uusi võimalusi isiku elukoha kasutamiseks viisil, kus konkreetse isiku elukoha täpne määratlus ei ole oluline andmetöötleja ega isiku õiguste ning kohustuste seisukohast. Näiteks kui andmetöötlejale ei ole oluline täpne aadress, sest sellest ei sõltuks andmetöötleja seisukohast midagi ning andmetöötlus ei mõjuta otseselt kindlat isikut. Seega võib teatud statistika koostamise puhul tõusetuda küsimus, kas statistika saaks koostada ka sisendandmete pinnalt, mis on verifitseeritavalt õiged, kuid mis ei ole seostatud mõne konkreetse aadressi või isikuga.

6.4. Privaatsusriskide käsitlemine valdkondlikul tasemel

Lisaks eelnevate tehniliste meetme rakendamisega andmetöötlejate poolt, saab privaatsusriske käsitleda ka sektori- ja valdkonna põhiselt nii seadusandja kui ka järelevalveasutuste osaluse läbi.

6.4.1. Andmekaitsealane mõjuhinnang õigusloome käigus

Isikuandmete kaitse üldmäärus artikkel 35 kohase mõjuhinna alusel saab andmetöötluse eest vastutav osapool hinnata töötlemise mõju ning kaasnevaid ohte privaatsuse seisukohast. Kuigi üldmäärus sätestab kohustuslikud sisuelemendid, on hinna andmine ise pigem subjektiivne ning võib erinevate vastutavate töötajate korral erineda nii kaalutluste poolest kui ka järelduste osas. Samas näeb üldmäärus ette, et kui andmetöötlus toimub seadusest tuleneva kohustuse või avalikes huvides oleva ülesande täitmiseks ning seadus käsitleb konkreetselt isikuandmete töötlemist nt teadus- ning riikliku statistika tarbeks, võib vastutav töötaja tugineda seaduse koostamise käigus tehtud mõjuhinna.

Panustades seaduse väljatöötamisel ka andmekaitsealase mõjuhinna tegemisele, on seadusandjal võimalik soodustada seaduse alusel planeeritud andmetöötlust. Seaduse loomise käigus teostatud mõjuhinna tagab andmetöötluse osapoolte jaoks parema selguse andmetöötlusega kaasnevate riskide osas, kui mõjuhinna käsitleda andmete väljastamisega kaasnevaid privaatsusriske. Seeläbi on vastutaval töötajal lihtsam maandada oma riske seoses andmete jagamisega, mis võimaldab efektiivsemat koostööd sideoperaatori ning statistika koostaja vahel.

6.4.2. Valdkonnapõhine toimimisjuhend

Üldmääruse kohaselt julgustatakse nii riiklikult kui ka EL asjakohaste instantside tasemel toimimisjuhendite koostamist, mis arvestatakse sektoripõhiseid eripäraseid ning ettevõtjate konkreetseid vajadusi.⁴⁴ Seonduva andmetöötluse toimimisjuhend hõlmab kõiki olulisi andmekaitsealaseid aspekte, alates andmete kogumisest ja isikute teavitamisest kuni pseudonümiseerimise viisi ja taseme ning standardsete turvameetmete detailse kirjelduseni. Toimimisjuhendiga võib käsitleda kõiki asjassepuutuvaid andmekaitsealaseid küsimusi, vajadusel ka detailsemalt, eesmärgiga tagada sektoripõhine areng ning innovatsioon samaaegselt privaatsuse kaitsega.

Toimimisjuhendite vastuvõtmine on aeganõudev ning eeldab järelevalveasutuse heaks kiitu. Kui toimimisjuhend hõlmab andmetöötlust vaid ühes riigis, sõltub selle heaks kiitmine vaid kohalikest järelevalve asutusest. Eesti näitel võiks siseriiklikult piirduva andmetöötluse kohase toimimisjuhendi heaks kiita Andmekaitse Inspeksioon. Kui andmetöötlus on seotud mitme riigiga, esitab kohalik järelevalveasutus selle enne heakskiitmist Euroopa Andmekaitse nõukogule. Euroopa Andmekaitse nõukogu on seni avaldanud heakskiidu kahele toimimisjuhendile, mis hõlmavad pilveteenuste teenusepakkujate andmetöötlust.⁴⁵

Valdkonnapõhise toimimisjuhendi olemasolu annab osapooltele praktilised juhised ja suunised ning vähendab andmetöötlusega kaasnevate vastavusnõuete ala- või ülehindamist. Toimimisjuhendi olemasolu ning selle järgimine soodustaks valdkonnas ühetaolist privaatsusriskide teadvustamist, mõistmist ning käsitlemist, mis omakorda võiks tagada juurdepääsu andmetele statistika- ja teadusuuringute hüvanguks. Siinkirjutajatele teadaolevalt ei ole spetsiifiliselt sideoperaatorite ning sideandmetega seoses ega statistika valdkonnas toimimisjuhendeid avaldatud.

6.4.3. Katusorganisatsioonide tegevused

Lisaks seadusandja ja järelevalveasutuste pingutustele võib valdkonnapõhiselt parimaid praktikaid kujundada mittetulundusliku eesmärgiga algatused ja organisatsioonid. Heaks näiteks on FPF (Future

⁴⁴ IKÜM art 40 lg 1

⁴⁵ https://edpb.europa.eu/news/news/2021/edpb-adopts-opinions-first-transnational-codes-conduct-statement-data-governance-act_en

of Privacy Forum) avaldatud juhised privaatsusriiskide käsitlemiseks seoses erinevate organisatsioonide poolt liikuvusandmete jagamisega.⁴⁶ Juhised suunavad organisatsiooni võtma arvesse andmete jagamise konteksti, seaduslikkust, privaatsuskäsitlust ning kontrollmeetmeid. Antud juhised ei ole üks-ühele kohaldatav sideandmete jagamisega, kuid ilmestab privaatsusküsimuste kompleksust ning võimalust läbi ühtse valdkonnapõhise käsitluse tekitada ühetaolisi praktikaid, mida erinevad huvipooled saaksid arvesse võtta.

7. Mobiilside andmete kasutamise senine praktika ja vajadused Eestis

Eestis on mobiilpositsioneerimise massandmeid kasutatud teaduslikus uurimistöös inimeste paiknemise ja mobiilsuse uurimiseks juba üle 15 aasta (Ahas *et al.*, 2010; Novak *et al.*, 2013). Eesti Panga, OÜ Positiumi ja Tartu Ülikooli mobiilsusuuringute labori koostöö tulemusena on välja arendatud meetodika, mille alusel teeb Eesti Pank välisreisi statistikat (Saluveer *et al.*, 2020). Nii on Eesti üks kahest riigist maailmas, kus mobiilpositsioneerimise andmeid ka tegelikult riikliku statistika tegemiseks kasutatakse. Kaardistamiseks seniseid mobiiliandmete kasutamise praktikaid, võimalusi, probleeme ja tulevikuvaateid viidi käesoleva projekti raames läbi vestlusi mobiiliandmeid kasutavate asutuste ja ettevõtetele. Vabas vormis vestlused viidi läbi Eesti Statistika, Majandus ja Kommunikatsiooniministeeriumi, Eesti Panga, OÜ Positium ja Telia Eesti AS inimestega.

Positium. Mobiilside andmetega omab laialdast praktilist kogemust Positium OÜ, kes on mobiilpositsioneerimise andmeid teadustöödeks ja statistika tootmiseks kasutanud alates aastast 2005. Intervjuust Positiumi juhatuse liikmest saab esile tuua alljärgnevat.

Sarnaselt *ESSnet Big Data* projektis osalejatega on ka Positiumi hinnangul sideandmete kasutamisel suur potentsiaal. Passiivse mobiilpositsioneerimise andmetele tuginevad otsused on andmepõhised ning kuluefektiivsed. Näitena saab esile tuua ettevõtte kogemuse Omaanist, kus mobiilpositsioneerimisele tuginevate andmete abil viidud uuring kujunes hinnanguliselt 150-200 korda odavamaks võrreldes traditsioonilise küsitlusega läbiviidava projektiga. Inimeste paiknemise ja mobiilsuse statistika tootmisel ning seonduvate valdkondade uuringutes pole mobiilpositsioneerimise andmete kasutamisele võrreldavaid alternatiive.

Erinevate koostöömudelite osas võib Positiumi kogemustest järeldada, et pikas plaanis on nii statistika tootmise kui ka teadustööde teostamise huvides kahepoolne koostöö sideoperaatoriga, kus mõlemad osapooled saavad koostööst tuntavat kasu. Vältida tuleks andmete jagamise kohustuslikkust, sest kohustusliku 'koostöö' korral võib sideoperaatori huvi andmete jagamisel olla piiratud ning ressursi säästmise eesmärgil võib operaator piirduda vähima võimaliku tegevusega. See võib omakorda pärssida meetodikate arendamist ning teadustööde läbiviimist, sest kohustusliku andme jagamise korral soovib sideoperaator maandada ärisriske ning kaitsta andmetega seotud väärtusahelat. Positiumi kogemuse näitab, et sideoperaator võib olla huvitatud andmete jagamisest, kui koostöö tulemusel saab sideoperaator ka ise otsest kasu, näiteks valideerides seeläbi ka oma mudelite ja andmetoodete täpsust.

Laias plaanis on oluline standardiseerimine, kuna see tagab statistika ning teadustööde meetodilise järjepidevuse ja läbipaistvuse. Standardiseeritud peab olema nii ligipääs andmetele, andmete sisu kui ka edastamine ja töötlemine.

⁴⁶ <https://fpf.org/wp-content/uploads/2021/08/2-MDSA-Operators-Manual.pdf>

Eesti Pank. Lisaks muudele seadusest tulenevatele ülesannetele on Eesti Pank maksebilansistatistika koostamise osas riikliku statistika tegija.⁴⁷ Passiivse mobiilpositsioneerimise andmeid on Eesti Pank kasutanud alates 2012. aastast, mille alusel arvutatud statistika algab tagasiulatuvalt 2008. aastast.⁴⁸ Eesti Pank toob esile, et varasemate ning suhteliselt väikese valimiga küsitlusuuringutega võrreldes võimaldab mobiilpositsioneerimise andmete kasutamine märkimisväärselt kiiremini, kvaliteetsemalt ning odavamalt saada andmeid välisreiside kohta. Lisaks kasutatakse mobiilpositsioneerimise andmeid ühe sisendina mudelis tööjõutulu arvutamisel, maksebilansi kiirhinnangu koostamisel ning turistide liikumise kohta hinnangute tegemisel. Põhjaliku ülevaate passiivse mobiilpositsioneerimise andmete kasutamise metoodikast välisreiside statistika arvutamiseks leiab Eesti Panga kodulehelt.⁴⁹

Eesti Pank kasutab mobiilpositsioneerimise andmeid üksnes statistilistel eesmärkidel ning need avaldatakse ainult agregeeritud kujul, identifitseerimist võimaldavate tunnusteta. Tulevikuperspektiivis võib Eesti Pangale huvi pakkuda võimalus passiivse mobiilpositsioneerimise andmetele tuginedes arvestada residentsust, mis ühtlasi võimaldaks pankadel täpsemini ja kiiremini otsustada, millise residentsusega nende erakliendid on. Inimeste residentsus on oluliseks sisendiks ka Eesti Panga koostatavale riigi maksebilansi arvutusele.

Kokkuvõtvalt on Eesti Panga hinnangul mobiilpositsioneerimise andmetes sisalduva rikkaliku informatsiooni kättesaadavus Eestis halb. Õiguslikult on tarvis selgust, kuid seaduste muutmiseks oleks esmalt vaja laiapõhjalist ühiskondlikku arutelu, mis käsitleks kuidas mobiilpositsioneerimise andmetega edasi minna. Ühelt poolt on need andmed ülimalt informatiivsed, aga samas on ka tegemist riivega inimeste privaatsusele. Seda, kas andmetest tõusev ühiskondlik kasu kaalub üles riive ja sellest lähtuvad ohud inimeste privaatsusele, tuleb ühiskondlikul kokku leppida.

Majandus- ja Kommunikatsiooniministeerium. Majandus- ja Kommunikatsiooniministeerium (MKM) tegeleb muu hulgas transpordipoliitikaga, mille eesmärk on tagada elanikele ja ettevõtetele mugavad, ohutud, kiired ja kestlikud liikumisvõimalused ([link](#)). Tegevusvaldkondade hulgas ka turism. Vastavalt MKM kodulehele on Eesti turismipoliitika põhieesmärk suurendada Eesti turismi rahvusvahelist konkurentsivõimet ja tagada turismiteenuste ekspordi ning siseturismi arendamisega jätkusuutlik majanduskasv. ([link](#))

MKM-i üks eesmärke transpordipoliitika teostamisel on riikliku transpordimudeli koostamine. Selle üks osa on tervet Eestit kattev kvaliteetne liikuvus- ja liiklusandmete andmebaas. Kõnealune andmebaas peaks sisaldama kõrgekvaliteedilist lähtekoht – sihtpunkt maatriksit (*Origin – Destination matrix*), kus lisaks liikumismahtudele ruumis on teada ka liikujate sotsiaal-majanduslik taust ning liikumisviis. Andmebaasi ruumiline täpsus võib tiheasustusaladel ulatuda transporditsoonideni, hõredamalt asustatud aladel piisab asustusüksuse või kandi täpsusest. Kui turismiuuringutes on tihti piisav üldine OD-maatriks, kus on teada vaid ruumiüksuste vahel liikuvate inimeste hulk, siis transpordimudeli jaoks on vaja märksa suuremat detailsust.

Riigimaanteede võrk on kaetud piisavalt tiheda liiklusloenduse sensorite võrguga, mis võimaldab täpselt hinnata liikumiskoormusi, kuid sellest andmestikust ei piisa OD-maatriksi loomiseks. Lisaks toimuvad regulaarselt nn voolikuloendused, kuid need on tihti korraldatud kohalike omavalitsuste poolt (nt piirkondliku liikumiskava koostamiseks). Usaldusväärseid loendusandmeid kogutakse tänasel päeval peamiselt induktsioon-, radar-/lidarsensoritega ning voolikuloendust abil. On testitud ka

⁴⁷ Riikliku statistika seadus §8 lõige 1

⁴⁸ <https://www.eestipank.ee/press/eesti-pank-hakkab-avaldama-valisreiside-statistikat-06032012>

⁴⁹ <https://statistika.eestipank.ee/failid/mbo/valisreisid.html>

uuemaid tehnoloogilisi saavutusi, nagu näiteks kaamerapõhine numbrimärgiloendus jt, kuid need pole soovitud kvaliteeti ning täpsust andnud.

MKM on teadvustanud, et uuematest andmekogumise meetoditest on kõige suurema potentsiaaliga mobiilpositsioneerimine. Mobiilpositsioneerimise andmete kasutatavust transpordiuuringutes on testitud erinevate projekti raames (nt ESPON projekt „[Potentials of big data for integrated territorial policy development in the European growth corridors](#)“, mille ühe tulemusena valmis [MP-andmete põhine regulaarliikumiste andmestik](#) Eesti kohta). Paraku ei sisalda selline anonüümsete andmete põhine OD-maatriks respondentide sotsiaal-majanduslikku taustainfot ning sellest ei piisa detailse transpordimudeli sisendiks. Seejuures on oluline toonitada, et eesmärk pole koguda isikuandmeid vaid nende sotsiaalmajanduslikke tunnuseid.

Vestluse toimumise päeval on käimas MKMi korraldatud riigihange detailse OD-maatriksi loomiseks mobiilpositsioneerimise andmete abil. Loodav OD-maatriks oleks andmesisendiks transpordimudelisse. Enne mobiilpositsioneerimise andmete kasutamist on OD-maatriksite loomiseks kasutatud sotsiaal-majanduslike andmete kogumiseks transpordiameti korraldatud liikuvusuuringuid. Tegemist on olnud traditsioonilise küsitlusuuringuga, mille käigus värvatakse vähemalt 5000 osalejat, kes vastavad ankeedile ning täidavad 2 päeva (1 töö- ja 1 puhkepäev) jooksul liikumispäevikut. Ka see uuring on vestluse toimumise päeval riigihanges.

Liikuvusuuring on oma olemuselt väga sarnane Eesti Statistika korraldatava ajakasutuse uuringuga ([link](#)). MKM on kaalunud kahe vähe sarnase uuringu ühendamise võimalusi, kuid selle peamiseks takistuseks on mõlema uuringu suur mahukus, mis ühenduuringuna ületaks vastajate/osalejate taluvust. Andmekogumise meetodite arenedes, kui uuring muutuks IKT-põhiseks ning automatiseerituks, mille tulemusena langeks ka vastamiskoormus, siis võiks tulevikus tegemist olla ühenduuringuga. Sel juhul oleks tegemist paneeluuringuga, milles ankeetküsitlusega kogutaks osalejate sotsiaal-majanduslikke andmeid, kuid liikumiste kohta käiv info salvestuks automaatselt vastava (nutitelefoni) rakenduse abil, millest algoritmide abil arvutatakse liikuvuse, paiknemise ja ajakasutuse statistikat.

Mobiilpositsioneerimise põhise liikuvusuuringu osas nendin MKM, et kuigi loodetavasti on lõpptulemuse näol tegemist täiesti uue kvaliteediga, siis kaasneb sellise lähenemisega ka erinevaid probleeme ja riske.

Üheks suuremaks ohuks on nõ monopoolse musta kasti teke. Täna seisus on Eesti suurim mobiilioperaator (Telia) välja arendanud oma liikuvus- ja paiknemisandmete toote ning turustab seda ise. Huvi ja soov algandmeid konkurentidega jagada puudub. See tähendab, et mobiilpositsioneerimise põhiseid liikuvusuuringute hankeid saab võita (kui hind jääb alla kehtestatud limiidi) ainult see sama operaator. Teistel soovijatel oma toote ja meetodikate arendamise võimalus puudub. On küll võimalus, et huvilised saavad kokkuleppele andmete ligipääsuks, kuid jääb risk, et ka teised operaatorid tahavad mobiilpositsioneerimise põhiseid andmetooteid ise turustama hakata. See omakorda tähendab, et ka teine operaator arendab välja oma meetodika, mille abil paiknemise ja liikuvuse andmeid mobiilpositsioneerimisest arvutatakse. Juhul kui uus tulija peaks võitma sama uuringu järgmise hanke, siis on tellija silmitsi probleemiga, et uue hankega saadavad andmed ei lähe kokku eelmise hanke võitjaga. Kaob võimalus adekvaatselt hinnata muutuseid ajas.

Probleem on hangete olemus ka selles aspektis, et kui täna osaleks hangetel mobiilioperaatorid ise või nende lepingupartnerid, siis on uuringutes ja statistika tootmiseks võimalik kasutada vaid ühe, hanke võitnud, operaatori andmeid. Samas tõstaks mobiilpositsioneerimise põhise liikuvusandmete kvaliteeti, kui ühte uuringusse oleks kaasatud mitme operaatori andmed. Mobiilioperaatori

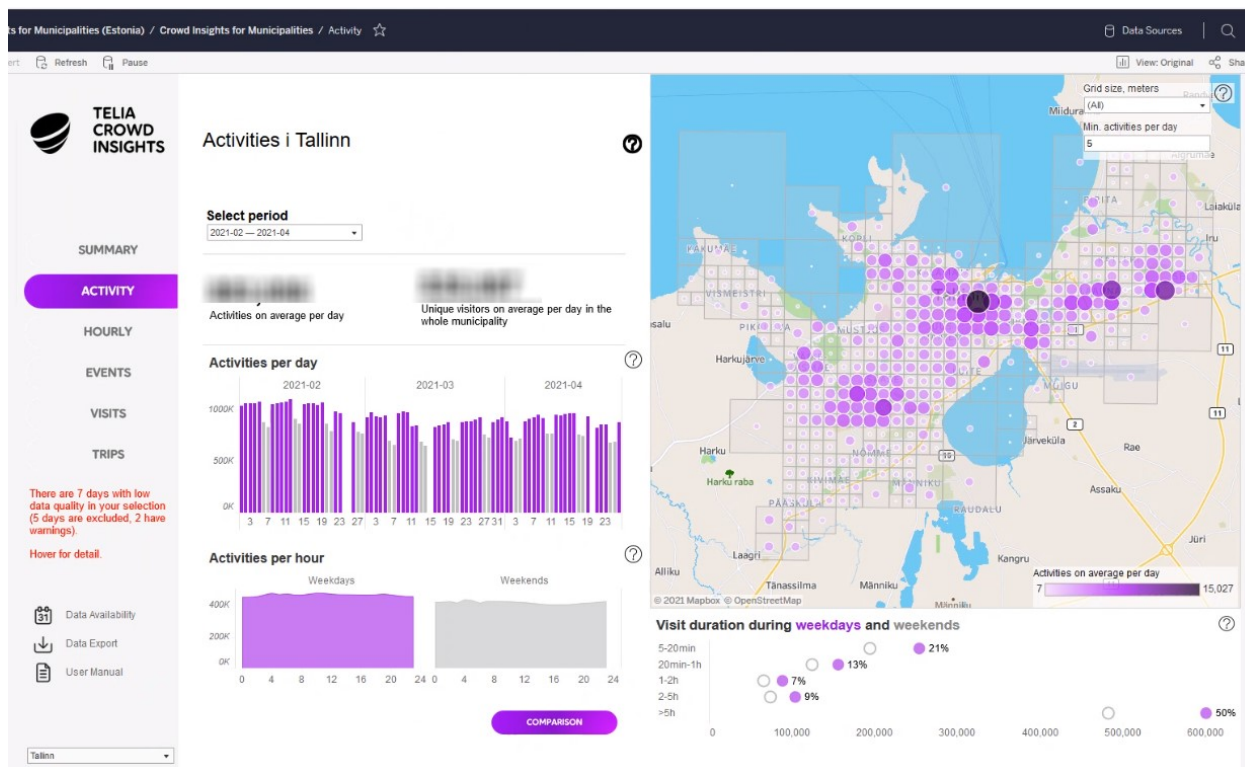
osalemisega sellistes uuringutes ja statistika tootmises kaasnevad ka ärihüvid, mis realiseeruvad autoriõiguste ja ärisaladustena. Sellega minnakse aga vastuollu statistika tootmise põhiprintsiibiga, milleks on läbipaistvus igas etapis. Kui aga näiteks OD-matriksi arvutamise meetodika kuulutatakse ärisaladuseks, siis ei saa ka valideerida selle korrektsust ja objektiivsust. Samuti aeglustub või peatub konkurentsi puudumise tingimustes meetodiline areng: milleks kulutada kui raha tuleb. Uute andmete kasutusele võtmine uuringute ja statistika tootmisel aga eeldab tihedat tellija ja tööde teostaja vahelist koostööd, suureks kasuks tuleks koostöö vastava eriala teadlastega. Uute andmete kasutuselevõtu algstaadiumis ei oska ka tellija sõnastada täpselt, mida täpselt tellida soovitakse. Või siis sõnastatakse traditsioonilisi statistikuid ja parameetreid, mille arvutamine on ebatäpsem, samas kui uus detailirikas andmestik võimaldaks arvutada mitmeid teisi liikuvust ja paiknemist kirjeldavaid parameetreid.

Tihti soovib mobiiloperaator müüa nõ karbitoodet koos graafilise kasutajaliidesega. Sellise liidese arendamine on kallis ja ei vasta tellija tegelikele andmeanalüüsi nõuetele ning ei ühildu teiste tellija poolt kogutavate ning loodavate andmestikega.

Et tulevikus mobiilpositsioneerimise andmete rikkalik sisu nii statistikasse kui ka uuringutesse jõuaks oleks vaja likvideerida seadusandluse vastuolud ning lõpetada temaatiline vaakum, mis takistab andmetele ligipääsu ning ei lase kujuneda andmevahetuse, andmekaitse ning arvutusmeetodikate standarditel. Kõige paremini võimaldaks seda MKM hinnangul saavutada olukord, kus mobiilpositsioneerimise andmeid haldaks neutraalne organisatsioon, mis suhtleks kõigi operaatoritega samadel standardsetel alusetel. See kindlustaks protsessi läbipaistvuse, statsionaarsuse ja meetodilise arengu.

Telia. Telia on välja arendanud ja turule toonud oma mobiilandmete põhise inimeste paiknemise ja liikuvuse andmete visualiseerimise platvormi „Crowd Insights“, mida pakutakse liikuvusstatistikast huvitatud klientidele. Koostööd on teiste hulgas alustatud näiteks Transpordiametiga liikuvusstatistika andmebaasi loomiseks.

Telia sõnul on tegemist mobiilivõrgu anonüümsete koondandmetega, mille abil saab asukohapõhise ülevaate inimeste liikumismustritest. Lisaks mobiilivõrgule kasutatakse andmesisendina ka wifi-võrgu andmeid. Seejuures kasutab Telia ise 24-tunnist kasutaja identifikaatorit. See tähendab, et iga isik on sama identifikaatoriga vaid 24 tunni jooksul, mille järel seda uuendatakse. Ajas nii lühike ID teeb paljud uuringud võimatuks ja kaotab suure hulga andmestiku algsest potentsiaalist. Lisaks andmetele endale on välja arendatud ka visualiseerimiskeskond (Joonis 15).



Joonis 15. Telia Crowd Insights kasutajaliidese ekraanitõmmis.

Kokkuvõttes tuleb tõdeda, et tegemist on nn karbitootega, mille metodika on suuresti nn läbipaistmatu must kast, mis on kallid ja jätab enamuse andmetes peituvast potentsiaalst rakendamata. Kaheldav on ka, et konkurentsi suudetakse pakkuda teaduspõhimõttel arendatavale metodikale. Vestluse ajal oli Telia poolne arvamus, et lähitulevikus nad mobiilpositsioneerimise andmeid teistsugusel kujul jagada ei plaani ning keskendutakse Crowd Insights toote arendamisele.