

KUIDAS MÕISTA ANDMESTUNUD MAAILMA?

GIGANTUM HUMERIS

SARJA KOLLEGIUM

Airi-Alina Allaste (Tallinna Ülikool)

Karsten Brüggemann (Tallinna Ülikool)

Tiina Elvisto (Tallinna Ülikool)

Indrek Ibrus (Tallinna Ülikool)

Marju Kõivupuu (Tallinna Ülikool)

Mihhail Lotman (Tallinna Ülikool, Tartu Ülikool)

Rain Mikser (Tallinna Ülikool)

Katrin Niglas (Tallinna Ülikool)

Hannes Palang (Tallinna Ülikool)

Ülar Ploom (Tallinna Ülikool)

Kristjan Port (Tallinna Ülikool)

Jaan Puhvel (California Los Angelese Ülikool)

Tõnis Pöder (Tallinna Ülikool)

Rein Raud (Tallinna Ülikool)

Raivo Stern (Keemilise ja Bioloogilise Füüsika Instituut)

Marek Tamm (Tallinna Ülikool)

Peeter Torop (Tartu Ülikool)

Jaan Valsiner (Aalborgi Ülikool)

Anna Verschik (Tallinna Ülikool)

Airi Värnik (Tallinna Ülikool)

Tallinna Ülikool

KUIDAS MÕISTA ANDMESTUNUD MAAILMA?

METODOLOOGILINE TEEJUHT

Koostanud ja toimetanud
Anu Masso, Katrin Tiidenberg
ja Andra Siibak

TLÜ Kirjastus
Tallinn 2020



GIGANTUM HUMERIS

Gigantum Humeris
Kuidas mõista andmestunud maailma?
Metodoloogiline teejuht

Keeletoimetaja Sirje Ratso
Korrektor Kai Nurmik
Küljendaja Sirje Ratso
Sarja makett: Rakett

Autoriõigus: peatükkide autorid, 2020
Autoriõigus (koostamine): Anu Masso, Katrin Tiidenberg ja Andra Siibak, 2020
Autoriõigus: Tallinna Ülikooli Kirjastus, 2020

ISSN 2228-1029
ISBN 978-9985-58-889-5

TLÜ Kirjastus
Narva mnt 25
10120 Tallinn
www.tlupress.com

Trükk: Folger Art

SISUKORD

Eessõna	7
---------------	---

Sissejuhatus (<i>Anu Masso, Katrin Tiidenberg, Andra Siibak</i>)	11
--	----

1. Maaailma andmestumine

1.1. Andmepõhine muutuste juhtimine (<i>Anu Masso, Triin Vihalemm, Leno Saarniit</i>)	41
1.2. Andmete õiguslik kaitse ja kasutamine teadustöös (<i>Aleksei Kelli, Irene Kull, Age Värvi</i>)	70
1.3. Avaandmete kasutusvõimalused ja piirangud (<i>Mai Beilmann, Ave Roots</i>)	97
1.4. Eetika ja privaatsus (<i>Katrin Tiidenberg, Andra Siibak</i>)	119

2. (Suur)andmete mõtestamine

2.1. (Suur)andmete visuaalne esitamine (<i>Anto Aasa</i>)	149
2.2. Sotsiaälvõrgustike analüüs (<i>Innar Liiv</i>)	178
2.3. Agendipõhine modelleerimine (<i>Kuldar Taveter</i>)	210
2.4. Masinõppe meetodid ja rakendused suurandmete töötlemisel (<i>Toomas Kirt</i>)	241

3. Teksti- ja tajuandmete analüüs

3.1. Tekstikaeve (<i>Kristel Uihoaed</i>)	283
3.2. Digitaalsed tekstiandmed ja korpuslingvistika (<i>Kadri Muischnek, Liina Lindström</i>)	306
3.3. Automatiseeritult hoiakute mõõtmine ja meelestatuse analüüs (<i>Mare Koit, Haldur Õim</i>)	340
3.4. Pilgujälgimine (<i>Kristian Pentus, Andres Kuusik</i>)	368

4. Lugude jutustamine andmetest ja andmetega

4.1. Digiajaloo ja -arheoloogia uurimismeetodid (<i>Marek Tamm, Hembo Pagi</i>)	409
--	-----

4.2. Ajakirjandus (suur)andmete ajastul (<i>Ragne Kõuts-Klemm, Marju Himma-Kadakas</i>)	434
4.3. Diskursusanalüüs andmestunud ühiskonnas (<i>Katrin Tiidenberg, Anu Masso, Maili Pilt, Liisi Laineste</i>)	462
4.4. Narratiivid sotsiaalmeedias (<i>Maili Pilt, Liisi Laineste</i>)	494
5. Platvormid ja andmetaristud uurimisvahendi ja -objektina	
5.1. Audiovisuaalne kultuur, metaandmed ja lingianalüüs (<i>Indrek Ibrus, Maarja Ojamaa</i>).	531
5.2. Internetiandmete kasutamine tööturu-uuringuteks (<i>Anna Veremchuk, Jaan Masso</i>)	569
5.3. Sotsiaalmeedia andmete sotsiaal-ruumiline analüüs (<i>Olle Järv, Kerli Müürisepp</i>)	597
5.4. Mobiiltelefonid ühiskonna ajalis-ruumilises analüüsis (<i>Siiri Silm, Olle Järv</i>)	622
6. Iseteadlik ja teovõimeline inimene andmeühiskonnas	
6.1. Küsitlusuuringud internetis (<i>Mare Ainsaar, Indrek Soidla, Ave Roots</i>).	653
6.2. Etnograafia andmestunud ühiskonnas: osalusvaatlus ja välitöömärkmep (Katrin Tiidenberg).	680
6.3. Veebiintervjuud, projektiivtehnika ja loovuuringis- meetodid (<i>Maria Murumaa-Mengel</i>).	707
6.4. Autoetnograafia andmestunud ühiskonnas (<i>Pille Pruulmann-Vengerfeldt</i>)	739
Terminisõnastik	763
Aineloend	779
Autorid.	787

EESSÕNA

Mõte koostada raamat andmestunud maailma uurimisest ja mõistmisest sai alguse sihtasutuse Archimedes kõrgkooliõpikute kirjutamise kutsungist. Selle raamatu kolmel koostajal tekkis teineteisest sõltumatult mõte panna kokku raamat, mis tutvustaks uusimaid suundi ühiskonna- ja humanitaarteaduslikes, kuid ka teiste seotud distsipliinide uuringutes, ja seda ajal, mil digitehnoloogiad ja internet aina enam uurimishuvi suunavad ja uurimisobjekte vormivad. Jõudude ühendamisel sündis unikaalne meeskond, kelle koostöö tulemusena on nüüd võimalik tutvustada laiemale lugejaskonnale Eestis kasutatavaid originaalseid andmeid, meetodeid ja vaatenurki.

Raamatu kirjutamisel on olnud kolm eesmärki: esiteks, avada andmestunud maailma analüüsimise mitmekesiseid võimalusi ja põhimõtteid; teiseks, tutvustada uudseid Eesti uurimusi ja siin arendatud analüüsimetodeid; kolmandaks, tekitada huvi uute meetodite kasutamise, katsetamise ja edasiarendamise vastu. Kõik need on eelduseks, et ühiskondlikke muutusi tõendus- ja teadmuspõhiselt juhitaks. Raamatu ambitsioon on käsitleda andmestumisega seotud metodoloogilisi muutusi ning alustada peamiste andmestunud maailma uurimise ja mõistmise kaasnevate nihete ja suundade kaardistamisega Eestis.

Raamatu kirjutamise käigus selgitasime välja Eestis praegu käibivad andmestunud maailma uurimisega seotud arusaamad ning pakume välja asjakohase eestikeelse terminoloogia. Tutvustame rohket rahvusvahelist tähelepanu pälvinud käsitlusi, ent toome lugejani ka innovaatilisi vaatenurki – mitmed kogumikus kajastamist leidvad meetodid on uuenduslikud ning oma valdkonnas teedrajavad nii Eestis kui ka maailmas. Raamat kajastab Eesti unikaalseid algatusi andmevaldkonnas ning arvestab Eesti kohalike õiguslike ja sotsiaalsete traditsioonide ja normidega. Raamatust leiab nii sammsammulisi juhendeid eri meetodite kasutamiseks kui ka diskussioone andmete õiglase ja eetilise loomise ja kasutamise kohta.

Sellisenä on raamat mitte klassikaline õpik, vaid tekstikogumik, mis avab andmeühiskonnas toimuvat ja selle uurimise telgitaguseid oma mitmekesisuses. Kogumikus on teadlikult välditud varasemaid

metodoloogilisi jaotusi ja klassifikatsioone (nt kvalitatiivne vs. kvantitatiivne, arvutuslik vs. interpretatiivne, deduktiivne vs. induktiivne). Rahvusvahelised metodoloogilistele uuendustele suunatud diskussioonid on valdavalt päri, et konventsionaalsed klassifikatsioonid piiravad ja risustavad andmestuva maailma mõistmist, millele on omane pigem olemasolevate vahendite segunemine ja brikolaaž. Ehkki käesolev kogumik toetub osaliselt rahvusvahelises kirjanduses pakutud diskussioonidele, oli meie eesmärk pakkuda just Eesti konteksti sobituvat vaadet. Kogumikus pakutud jaotustega – ühiskonna andmestumine, suurandmete mõtestamine, teksti- ja tajuandmete analüüs, andmetega ja andmetest lugude jutustamine, platvormid ja andmetaristud ning andmeühiskonnas tegutsev inimene – soovime panna lugejaid mõtlema valikute ja võimaluste üle, mis seoses andmemaaailmas toimuvate muutustega avanevad.

Andmestunud maailma üheks eripäraks on, et meie kõigi igapäevane elu on andmetest läbi põimunud. Sageli oleme enesele teadmata osa ühiskonna andmestumisprotsessist, seetõttu on raamat mõeldud laiale lugejaskonnale. Mahukast kogumikust leiavad kasulikku lugemist nii andmetega igapäevaselt töötavad või seda teha soovivad analüütikud, eksperdid ja otsustajad, andmetega tööd alustavad bakalaureuse- või magistritaseme üliõpilased kui ka juba kogenud eksperdid, kes soovivad oma analüütilist repertuaari laiendada või end andmestunud maailmas toimuvate muutustega paremini kurssi viia. Loodame, et kogumik pakub mõtlemisainet ka kõigile neile, kes ise otseselt andmete analüüsimisega kokku ei puutu, kuid kes loovad oma igapäevases elus ja rutiinsete tegevuste käigus analüüsiks kasutatavaid andmeid ja tahaksid andmestumisega kaasnevaid protsesse paremini mõista, samuti neile, kes sooviksid kodanikuaktiivsuse vormis ja avatud andmete abil panustada avalike väärtuste loomisesse. Kogumiku lai sihtrühm ja temaatiline jaotus peegeldab jällegi andmestunud maailma mitmekesisust, ilmestades koostöövajadust eri valdkondades tegutsevate ekspertide ja aktiivsete kodanike vahel.

Andmestuvat maailma uurivad raamatus eri distsipliinide esindajad: peatükkide autoriteks on 38 meedia- ja kommunikatsiooni, sotsioloogia, avaliku halduse, geoinformaatika, inimgeograafia, keeleteaduse, keeletehnoloogia, digilingvistika, folklooriuuringute, arvutiteaduste, mate-

maatika, kultuuriajaloo, majandusteaduste, psühholoogia ja õigusteaduste eksperti, õppejõudu ja praktikut kolmest Eesti ülikoolist – Tartu Ülikoolist, Tallinna Ülikoolist ja Tallinna Tehnikaülikoolist – ning mitmest teadus- ja arendusasutusest (nt Eesti Kirjandusmuuseum), eraettevõtetest ja välisülikoolidest (Malmö Ülikool, Helsingi Ülikool).

Selline multidistsiplinaarsus väljendub muu hulgas paralleelselt kasutusel olevates, aga oluliste tähenduslike nüansierinevustega terminite kasutamises. Ehkki terminoloogia ühtlustamine oli oluline osa kogumiku toimetajate ja autorite tööst, ei surunud me peale täielikku ühtlustamist, vaid proovisime erinevad käibel olevad mõtteviisid üksteisele lähemale tuua ning lugejale arusaadavamaks muuta. Teisiõnu – eelistasime lausühtlusele täpsust. Nii on näiteks ingliskeelse termini *dashboard* jaoks eesti keeles kasutusel kaks täpsemat terminit: andmetöötajatele ja analüütikutele mõeldud n-ö *backend*-lahenduse puhul „kiiranalüütika paneel“ (ptk 1.1) ja sotsiaalmeediaplatvormide kasutajatele mõeldud n-ö *frontend*-lahendus, milleks on „kasutaja koondpaneel“ (ptk 4.3). Seesuguseid keelelisi ja distsiplinaarseid tõlkimise ja tõlgendamise valikuid on kogumikus veelgi.

Kuidas seda raamatut lugeda? Raamat on loetav kaanest kaaneni, kuid ka üksikute peatükkide kaupa. Lisaks võib raamatu lõpus olevat märksõnastikku kasutada suunaandjana, et lugeda raamatut lõikude või alateemade kaupa. Lugemisel tekkivatele terminoloogilistele küsimustele leiab vastused raamatu lõpus olevast leksikonist. Iga peatüki alguses on esitatud kolm kõige olulisemat kirjandusviidet käsitletava teema või meetodi kohta. Peatükkides kasutatud tekstikastid annavad praktikulist infot ühe või teise meetodi kasutamisest või tarkvaralahendustest, samuti meetodi sammhaaval sooritamise kohta. Raamatu osade vahele lisatud „eksperdisäutsud“ (Twitteri säutsu mõõtmes kommentaarid) loovad konteksti ning ärgitavad mõtlema eri vaatenurkade ja põhimõtete, meetodite ja tehnikate päritolu, ajaloo ja rolli üle laiemas ühiskonnauurimuslikus ökosüsteemis.

Kuna andmed, ühiskond ja andmestunud ühiskond on pidevas muutumises, siis soovime raamatu koostajatena teatepulga lugejatele edasi anda – julgustame lugejaid uute meetoditega katsetama, neid testima ja edasi arendama ning pakkuma aina uusi metoodilisi viise andmeühiskonna mõistmiseks ja mõtestamiseks.

Tähusõnad

Raamatu koostajad tänavad kõiki raamatu kirjutamises pühendumisega osalenud autoreid (tähestikulises järjekorras): Anto Aasa, Mare Ainsaar, Mai Beilmann, Marju Himma, Indrek Ibrus, Olle Järv, Aleksei Kelli, Toomas Kirt, Mare Koit, Ragne Kõuts, Irene Kull, Andres Kuusik, Liisi Laineste, Liina Lindström, Innar Liiv, Anu Masso, Jaan Masso, Kadri Muischnek, Maria Murumaa-Mengel, Kerli Müürisepp, Haldur Õim, Maarja Ojamaa, Hembo Pagi, Kristian Pentus, Maili Pilt, Pille Pruulmann-Vengerfeldt, Ave Roots, Leno Saarniit, Andra Siibak, Siiri Silm, Indrek Soidla, Marek Tamm, Kuldar Taveter, Katrin Tiidenberg, Kristel Uiboaed, Age Värv, Anna Veremchuk ja Triin Vihalemm.

Samuti täname raamatu osade vahele siduvad kommentaarid kirjutanud oma ala eksperte Airi-Alina Allastet, Veronika Kalmust, Dagmar Kutsarit, Ene-Margit Tiitu, Liina-Mai Toodingut, Leo Vöhandut ja Mikko Lagerspetzi. Raamatu illustreerimise eest oleme tänulikud osade vahetiitlitele karikatuurid joonistanud Ave Taavetile ja kaanepildi teinud Janne Liasele.

Raamat on saanud võimalikuks tänu sihtasutuse Archimedes, Eesti Keele Instituudi, Tartu Ülikooli digihumanitaaria ja infoühiskonna keskuse, Tallinna Ülikooli kirjastuse, Tallinna Tehnikaülikooli arenguprogrammi Astra ning teiste grantide toetusele. Suur aitäh selle eest!

Konstruktiivsete tähelepanekute ja soovitude eest täname raamatuprojekti anonüümseid retsensente. Meie eriline tänu kuulub lõpliku käsi kirja retsensentidele Mikko Lagerspetzile ja Kadri Rootalule, kelle ülimalt detailsed ja sisukad retsensioonid on olnud suureks abiks raamatu viimistlemisel. Täname keeleteoimetajat ja kujundajat Sirje Ratsot ning Tallinna Ülikooli kirjastust eesotsas peatoimetaja Rebekka Lotmaniga suurepärase koostöö eest raamatu väljaandmisel.

Meie eriline tänu kuulub Liina Deutschlerile, kes lisaks suuremahulisele tööle terminoloogia ühtlustamisel aitas kaasa raamatu leksikoni ja indeksi koostamisele, peatükkide tehnilisele ühtlustamisele ja seeläbi raamatu selgusele ja loetavusele.

Anu Masso, Katrin Tiidenberg, Andra Siibak

SISSEJUHATUS¹

Anu Masso, Katrin Tiidenberg,
Andra Siibak

LUGEMISSOOVITUSED

- Rob Kitchin, *The Data Revolution: Big data, Open Data, Data Infrastructures and Their Consequences*. Thousand Oaks CA, Sage 2014.
- Richard Rogers, *Doing Digital Methods*. Los Angeles, London: Sage 2019.
- Mirko Tobias Schäfer, Karin van Es (eds.), *The Datafied Society: Studying Culture through Data*. Amsterdam University Press 2017.

Käesolev raamat keskendub andmestumise (*datafication*)² kui ühe viimaste aastate olulisima, samas vastuolulise nähtuse uurimisele. Kui täpsem olla, on raamat pühendatud andmestumise uurimise ja mõistmise viisidele – tegemist on kogumikuga, mille keskmes on meetodid ja metodoloogilised võtted.

Andmestumine tähendab lihtsustatult seda, et igapäevased, tava-päraselt valdavalt kaduvad, haihtuvad ja „nähtamatud“ tegevused ning ühiskondlikud protsessid teisendatakse enamasti arvulisteks, vähemalt osaliselt süstematiseeritud (nt andmebaasi koondatud), kvantitatiivseteks ja analüüsivalmis formaadis olevateks andmeteks, mille põhjal ühiskonnas olulisi otsuseid tehakse. Bussisõit, kinopiletite ost ja uudiste lugemine olid mõnikümmend aastat tagasi praktilised, sotsiaalsed ja kommunikatiivsed tegevused, kuid praegusel ajal on need suuresti tänu infotehnoloogia levikule, internetiühenduse valdavusele ja arvutusliku võimekuse kasvule lisaks ka andmeid loovad tegevused.

Andmestumine on hoomatav pea kõigis elu- ja tegevusvaldkondades. Helistamine lähedastele, sammuloendamise rakenduse kasutamine,

¹ Peatüki kirjutamist on toetanud Tallinna Tehnikaülikooli arenguprogramm Astra aastateks 2016–2022 (2014–2020.4.01.16-0032).

² Kuna enamik teaduskirjandust ilmub inglise keeles, siis on üldjuhul siin ja edaspidi eesti-keelsete terminite juures esitatud ka ingliskeelne vaste.

sotsiaalmeedias suhtlemine, ostutehingud pangakaardiga ja internetikauplustest – need on vaid mõned näited tegevustest, millest jäävad maha andmed. Sellised andmed võimaldavad mõista ühiskonnas toimuvaid protsesse, kuid andmetega läbi põimunud igapäevaelu muutub ka uueks mõistmist ja selgitamist vajavaks nähtuseks.

Järgnevalt anname ülevaate peamistest nihetest, meetoditest, mõtteviisidest ja muutustest seoses maailma andmestumisega. Tuginedes varasematele diskussioonidele ning siinses kogumikus käsitletud meetoditele pakume välja omapoolse lähenemise andmestunud maailma uurimisele.

Kuidas mõista andmestunud maailma?

Esmased sotsiaalteaduslikud diskussioonid andmepöörde teemal (Kitchin 2014a) olid üsna skeptilised selles osas, kuivõrd olemasolevad traditsioonilised sotsiaalteaduslikud põhimõtted uute, suurandmeteks nimetatud andmete konteksti sobituvad. Ühelt poolt ennustati radikaalseid epistemoloogilisi ja paradigmaatilisi nihkeid senises sotsiaalteaduslikus mõtlemises, teisalt otsiti n-ö kolmandat teed traditsiooniliste statistiliste ja uuemate arvutuslike meetodite vahel. Samal ajal kerkisid esile andmeuuringud, mis kutsusid üles kriitiliselt mõtestama andmete

TEKSTIKAST I. ANDMETE JA MEETODITE TEEMAL ILMUNUD EESTIKEELNE KIRJANDUS

Mai Beilmann, Ave Roots, Kadri Rootalu (toim.), Sotsiaalse analüüsi meetodite ja metodoloogia õpibaas. 2020. <http://samm.ut.ee>.

Meri-Liis Laherand, Kvalitatiivne uurimisviis. Tartu: Sulesepp 2010.

Mikko Lagerspetz, Ühiskonna uurimise meetodid: Sissejuhatus ja väljajuhatus. Tallinna Ülikooli Kirjastus 2017. <https://www.tlu.ee/pood/home/239-uhiskonna-uurimise-meetodid-sissejuhatus-ja-valjajuhatus.html>.

Katrin Niglas, Statistiline andmeanalüüs paketi SPSS for Windows 14.0. Tallinna Ülikooli Kirjastus 2008. http://www.cs.tlu.ee/~katrin/wp/wp-content/uploads/2013/11/SPSSI4_pealkirjaga.pdf.

Kadri Rootalu, Veronika Kalmus, Anu Masso, Triin Vihalemm (toim.), Sotsiaalse analüüsi meetodite ja metodoloogia õpibaas. 2014. <http://samm.ut.ee>.

Ene-Margit Tiit, Liina-Mai Tooding, Statistikalleksikon. Tartu Ülikooli Kirjastus 2019.

Liina-Mai Tooding, Andmete analüüs ja tõlgendamine sotsiaalteadustes.

2., täiend. vlj. Tartu Ülikooli Kirjastus 2015.

ühiskondlikke ja kultuurilisi tagajärgi ning andmete loomise, kogumise ja kasutamise loodud reaalsust (Dalton *et al.* 2016).

Siinse sissejuhatus eesmärk pole korrata varasemaid sotsiaalteadusliku metodoloogia diskussioone: tekstikast 1 pakub põgusa ülevaate suurepärastest eestikeelsetest teostest, mis keskenduvad just sellele. Käsitleme varasemaid seisukohti ja pikaajalisi dispuute peamiselt muutuste võtmes (vt alajaotus „Metodoloogilised nihked“), mis on vorminud andmestunud maailma uurimist puudutavate diskussioonide pinnase.

Omalt poolt pakume välja uudse lähenemisviisi andmestunud maailma uurimiseks, esitades selleks kümme postulaati (vt tekstikast 2), millest soovitame andmestunud maailma tõlgendamisel ja mõistmisel lähtuda. Postulaatide loomisel tuginesime nii rahvusvahelistele teadusdiskussioonidele, raamatu peatükkides esitletule kui ka meie enda uurimistööst tõusetunud teadmistele.

TEKSTIKAST 2. ANDMESTUNUD MAAILMA UURIMISE POSTULAADID

1. Andmestumine on keeruline protsess, andmesuhted on hierarhilised.

Andmete loomine, kogumine ja kasutamine on vastastikku sõltuv, keeruline ja läbipõimunud sotsiaalne protsess. Andmetöös on oluline mõista andmete sageli vastuolulistest ja hierarhilistest suhetest läbipõimitud eluteed.

2. Andmestunud maailmas valitsevad võrgustunud seosed.

Andmeühiskonna mõistmiseks ei piisa tavapärastest lineaarsetest põhjuslikest tõlgendustest. Andmeühiskonna mõistmiseks on vaja keskenduda võrgustike seostele.

3. Andmetaristu on avalik hüve. Andmete mitmekesisus ja võrdne juurdepääs andmetele on ülioluline. Selle eest peaks võitlema nii andmesubjektid, andmekodanikud, andmeaktivistid, andmetöötajad kui ka andmeühiskonna uurijad.

4. Eetika ja andmeõiglus peaks olema andmestunud ühiskonna alustalad.

Andmete kogumine, analüüsimine ja tõlgendamine, aga ka selle tegemata jätmine peab olema andmesubjekti ning ühiskonna seisukohalt eetiline, õiglane ja õigustatud.

5. Kontekst on kuningas. Andmed vajavad alati tõlgendamist. Tõlgendused peavad lähtuma sotsiaal-kultuurilisest kontekstist. Analüüsitulemused on enamasti üldistatavad vaid konkreetses ajas ja ruumis toimuvatele nähtustele.

6. Andmed ei teki iseenesest. Andmetöötajal ja uurijal on aktiivne roll andmekogumise planeerimisel ja andmete kogumisega seotud valikute tegemisel. Kõiki andmeid, mida on võimalik luua või koguda, ei ole tingimata vaja luua või koguda. Andmeid tuleks luua ja tõlgendada lähtuvalt praktilisest vajadusest ning andmeõigluse ja -eetika põhimõtetest.

7. **Teooria pole surnud.** Andmestunud maailma uurimine algab asjakohase ja haritud küsimuse püstitamisest. Andmete tõlgendamine peab olema dialoogis huvitava nähtuse kohta juba teadaoleva infoga.
8. **Käes on metodoloogilise pluralismi aeg.** Mõistlik on meetodeid paindlikult ja uurimisküsimusele vastavalt kombineerida.
9. **Paradigmasõjad jäägu minevikku.** Andmeühiskonna mõistmisel tuleks lähtuda kasulikkuse põhimõttest, ühelgi distsipliinil ei ole andmeühiskonna mõtestamise monopoli. Traditsioonidest ja distsiplinaarsetest erisustest olulisem on uurimisprobleemi pragmaatiline käsitus.
10. **Andmetöötaja või uurija on kõige olulisem epistemoloogiline tööriist.** Andmetöötaja või uurija peaks olema iseteadlik ja teovõimeline ning hindama alati kriitiliselt oma tegevuse tagajärgi.

Kui mõni sõnastatud soovitudest on sotsiaal- ja humanitaarteaduslikus uurimistöös ja ka andmestunud ühiskonna uurimisel juba juurdumas (nt andme-eetika), siis mõne põhimõtte konkreetsed rakendusviisid ja lahendused on alles välja kujunemas (nt juurdepääs andmetaristule, andmeõiglus). Nii pakume neile kümnele postulaadile toetudes välja kuus andmestunud maailma mõistmise viisi, mille järgi oleme peatükid koondanud raamatu osadeks: 1) maailma andmestumine, 2) suurandmete mõtestamine, 3) teksti- ja tajuandmete analüüs, 4) andmetega ja andmetest lugude jutustamine, 5) platvormid ja andmetaristud ning 6) andmeühiskonnas tegutsev inimene (vt alajaotus „Raamatu ülesehitus: meetodid, mõtteviisid ja muutused“). Selline jaotus on kahtlemata üks paljudest võimalikest. Seetõttu loodame raamatu koostajate ja kaasautoritena, et diskussioonid andmeühiskonna uurimise parimate lahenduste leidmise nimel jätkuvad.

Lisaks on pakutud postulaadid Eesti kontekstis esimene samm andmestumisega seotud muutuste mõtestamise suunas. Loodame, et see suunab ja toetab andmestumist mõtestavaid tegevusi ja toob kaasa järgnevad sammud nende edasiarendamiseks. Meie hinnangul loob nende kümne põhimõtte ning kuue käsitusviisi läbikaalumine vundamendi heade andmepõhiste lahenduste loomiseks ning aitab loodetavasti ära hoida soovimatuid tagajärgi.

Järgnevalt tutvustame andmestunud ühiskonna uurimise postulaate ja nendega kaasnevat pisut lähemalt.

Andmestumine ja andmepööre

Käesolevas kogumikus mõistame andmeid ja andmestumist kui ontoloogilist (andmed kui loodud reaalsus või selle konstrueerimise vahend), epistemoloogilist (andmed kui teadmise viis) ja sageli ka ideoloogilist protsessi. Rahvusvahelistes uuringutes on eksperdid defineerinud ühiskonna andmestumist eelkõige kui kõigi ja kõige viimist struktureeritud andmeformaati, mis omakorda võimaldab ja soosib kvantifitseerimist (vt nt Dijck *et al.* 2018; Schäfer 2016; Lupton 2020). Sageli väidetakse, et andmestumine kaasneb paratamatult teiste maailmas toimuvate muutustega, nagu digitaliseerimine ja võrgustamine.

Lihtsustatult võib väita, et infotehnoloogia vahendusel sooritatud tegevusest jäävad maha „andmepurust rajad“. Ehkki nende teket on paljudes avalikes ja ka professionaalsetes keskusteludes kombeks paratamatuks pidada, soovitavad kriitilised käsitlused (D’Ignazio, Klein 2020) alati küsida, kelle kätte koonduvad andmed, kas sellega kaasneb võimu koondumine, ning kas seesinane koondumine ja seda soosiv süsteem peaks ikka eksisteerima. Eriti aktuaalseks on sellised küsimused muutunud näiteks USA ja Hiina näotuvastustehnoloogiate ja ennustava politseitöö teemalistes vestlustes. Lisaks võiks iga andmeagent (andmete looja, kasutaja, nendega töötav või neid uuriv isik) endalt küsida, kas andmepuru kogumine, arhiveerimine, agregeerimine ja teatud eesmärkidel kasutamine on paratamatu ja vajalik. Kui midagi on võimalik andmeteks muuta, ei tähenda see tingimata, et see on hea idee. Andmestumise protsessilisusest tulenevalt olemegi sõnastanud andmestunud maailma mõistmise esimese postulaadi: **andmete loomine, kogumine ja kasutamine on keeruline sotsiaalne protsess, kus andmeanalüütik (andmete põhjal otsustaja) peab arvestama andmete loodud sotsiaalse reaalsusega, sh sageli vastuoluliste ja hierarhiliste andmesuhetega.**

Andmete ja ühiskonna vahelised seosed on keerulised. Ühest küljest luuakse andmeid inimeste elus ja ühiskonnas toimuvate protsesside põhjal ning neid kasutatakse samade protsesside analüüsimiseks ja mõtestamiseks; nii vaadates on andmed ühiskonna osa. Teisalt tekivad andmestumise baasilt täiesti uued ühiskondlikud nähtused. Seega on oluline, millisest perspektiivist lähtuvalt tõlgendatakse muutuste selgitamisel andmete ja ühiskonna omavahelist suhet (vt nt Beraldo, Milan

2019) – kas ühiskondlikud protsessid on andmete ja andmetehnoloogiate poolt juhitud (*data-driven*), andmetel põhinevad (*data-based*) või andmetest informeeritud (*data-informed*) (vt ka ptk 1.1). Esimesel juhul eeldatakse, et andmed ja andmetehnoloogiad tingivad teatud nähtusi (näiteks kui algoritmid filtreerivad ja vormivad inimeste juurdepääsu informatsioonile, millel võivad olla laiaulatuslikud sotsiaalsed, poliitilised ja võrdset kohtlemist puudutavad tagajärjed). Teisel juhul on andmed vaid formaalne infrastruktuur, mis loob aluse otsuste tegemiseks, näiteks teadlike valikute tegemisel algoritmide poolt filtreeritud informatsiooni kontekstis. Nii oleme sõnastanud andmestunud maailma uurimise teise postulaadi: **võrgustunud ja omavahel tihedas seoses olevad protsessid ei pruugi olla selgitatavad lineaarsete tõlgenduste ja meetodite kaudu, vaid üha enam tuleb tähelepanu pöörata tegutsejate võrgustikule, mis võimaldab selgitada andmestumisega seotud keerukaid nähtusi ja võimalusi mitmetasandiliste ja üksteisega põimunud seoste kaudu.**

Ühiskondlikult oluliste tegevuste kvantifitseerimises pole iseenesest midagi uut: valitsejad on maksukogumiseks andmeid kogunud tuhandeid aastaid, kirikud kogusid ja avaldasid aastasadu surnute andmebaase. Ka nende näidete puhul võib väita, et andmesubjektide kohta teadmiste koondamisega kaasnes võimu koondumine nende kätte, kelle kätte koondusid andmed (D'Ignazio, Kelin 2020). Siiski võib tänase andmestumise üheks alguspunktiks märkida arvutustehnoloogiate kasutuselevõtu rahvaloendusandmete töötlemisel. Herman Hollerithi 1890. aastal USA-s välja töötatud arvutusmasin (vt nt Whitby 2020) võimaldas rahvaloenduse tulemustest kokkuvõtte teha mõne kuuga varasema kaheksa aasta asemel. Seega võib andmestumist selgitada lainetena toimuva tehnosotsiaalse muutusena (Schot, Kanger 2018), mis digitaliseerumise (Archer 2015), meediastumise³ (Hepp 2020) ja ühenduvusega (*connectivity*) (Couldry, Meijas 2019) seoses on järjekordse transformatsiooni läbi teinud. Käimasolevat andmestumiselainet peetakse eripäraseks (Just, Latzer 2017) johtuvalt selle skaalast ja ulatusest, mistõttu on praeguse andmestumisega kaasnenud suurem killustumine, kaubastumine, ebavõrdsus ja riigipiiride hägustumine ning vähenenud andmetega seotud

³ Meediastumine on sotsiaalse maailma ja selle eri protsesside aina tihenev seotus meediaga.

protsesside läbipaistvus, kontrollitavus ja ennustatavus. Andmestumine ei ole passiivne (*datafied*), vaid aktiivne protsess (*datafying*), milles on oma võitjad ja kaotajad.

Eestis on andmestumine nähtavamaks muutunud seoses Euroopa andmekaitseseadusega (EP 2016), mille jõustumise ning kaasnevate piirangute tõttu on era- ja avaliku sfääri institutsioonid, aga ka n-ö andmesubjektideks peetavad indiviidid, kelle isikuandmeid kogutakse, hoitakse või töödeldakse (sageli mh nende enese teadmata), saanud teadlikumaks andmekasutamisega seotud võimalustest, ohtudest ja piirangutest. Selline areng on Eestis tõstatanud järgmised küsimused: Kas ja mis tingimustel on lubatud erafirmadel andmeid kasutada ja analüüsida? Kuidas teha andmed kättesaadavaks teadusuuringute jaoks? Kuidas tagada avatud andmete liikumine avaliku sektori institutsioonidele ja kodanikuaktivistidele? Kas ärilistel ja teaduslikel eesmärkidel andmete kogumisele peaks rakenduma erinevad piirangud? Ka siinses raamatus käsitletakse andmete kasutamise printsiipe, näiteks andmete õiguslikku kaitset ja kasutamist (ptk 1.3), andmete kasutamisega seotud eetilisi põhimõtteid (ptk 1.4) ning andmetega seotud sotsiaalse õigluse küsimusi (ptk 1.1). Sellest lähtuvalt oleme sõnastanud ka andmestunud maailma uurimise kolmanda postulaadi: **andmestunud maailmas pole andmetaristu üksnes institutsionaalne omand, vaid võimalus avaliku hüve loomiseks, andmetes leiduva mitmekesisuse kaudu sotsiaalse mitmekesisuse tagamiseks ning andmepõhiste muutuste strateegiliseks juhtimiseks.**

Raamatu koostajatena väidame, et andmed ja andmetes esindatud isikud pole teineteise teisikud (*data doubles*; vt Raley 2013; Goriunova 2019). Pigem peegelduvad andmekogudes uurijate väärtused ning (kohati aegunud) arusaamad sotsiaalsete kategooriate (sugu, rass, sisetulek, vanusegrupp, haridustase) tõlgendusjõust. Kuigi andmestumine loob loendamatud andmekogud ja teoreetilise võimaluse haarata ühiskondlikku elu kogu selle mitmekesisuses, jääb andmete mitmekesisuse küsimus sageli tähelepanuta. Andmestumine lähtub suuresti põhimõttest, et sotsiaalne reaalsus jaotatakse piiratud kategooriatesse. Ka klassikaline arvutusteaduslik meetod baseerub dihhotoomsele ehk 0–1 jaotusele ja arusaamale maailmast. Varasemalt kasutatud binaarsed kategooriad (nt 1 – õige, 2 – vale) on andmete puhastamisel,

korrastamisel ja tõlgendamisel endiselt kasutusel, selmet püüelda andmete lõpmatus mitmekesisuses leiduva intersektionaalsuse kajastamise poole (Hopkins 2019). Andmestunud ühiskonna teadmusloome (Shaw 2015; Wagner-Pacifci *et al.* 2015) eeldab andmete käsitlemist nii epistemoloogilise kui ka ontoloogilise nähtusena, ent tekitab lisaks hulga andmeõigluslike (*data justice*) küsimusi, mille eesmärk on kindlustada, et andmed tagaks senisest suurema, mitte väiksema sotsiaalse õigluse. Nende diskussioonide põhjal oleme sõnastanud andmestunud maailma mõistmise neljanda postulaadi: **lisaks tavapärastele teaduseetika nõuetele (uuritavate kahjustamise vältimine) peab andmete kogumine, analüüsimine ja tõlgendamine olema andmesubjekti ja ühiskonna seisukohalt õigustatud ja õiglane.**

Eesti ühiskond pakub suurepärase võimaluse andmestumisega seotud arengusuundade analüüsiks – ühest küljest ollakse Eestis andme- ja tehnoloogiausust kannustatuna alid katsetama uudseid andmetehnoloogiaid (Runnel *et al.* 2009; Shaw 2015), teisalt on Eesti maailma kontekstis olnud suhteliselt tagasihoidlik valitsuse avaandmete arendamisel (McBride *et al.* 2018). Probleemiks on olnud eelkõige era- ja avaliku sektori ebavõrdne juurdepääs andmeressurssidele, aga ka erinevad arusaamad avalikest väärtustest ja sotsiaalsetest normidest andmelahenduste väljatöötamisel (Männiste, Masso 2020). Meil on eeskujulik digitaalne riigihaldus ja e-valitsemise taristu, ent erasektori digiteeritus on ebaühtlane ning kultuuripärandi lausdigiteerimise ja avaandmestamise algatused kipuvad venima (vt ptk 5.1).

Siiski on Eesti olnud teerajaja mitut tüüpi andmete kasutuselevõtmisel ja nende analüüsiks vajalike meetodite arendamisel. Mobiiltelefonide kõnetoiminguandmed ja mobiilpositsioneerimine sotsiaal-ruumilise segregatsiooni analüüsimiseks (Ahas *et al.* 2007) on üks eredaim näide (vt ptk-d 5.4 ja 2.1). Eesti digitehnoloogiatega edusamme ja uuenduslikke andmelahendusi on kasutatud isegi riigi rahvusvaheliseks brändimiseks (Tamppuu, Masso 2018; Männiste, Masso 2020). Selle hiljutiseks näiteks on Eesti tehisintellekti ehk krattide lahendused.

Paljud Eesti teadlased ja selle kogumiku autorid on oma erialal uurimismeetodite edasiarendamises ja nendega eksperimenteerimises esirinnas, luues lahendusi, mida kasutavad kolleegid ja tudengid rahvusvaheliselt. Käesolev raamat esitab ülevaate Eestis välja pakutud

andmestunud maailma uurimise viisidest, mis on kohandatavad rahvusvaheliselt, kuid eelkõige loodud Eesti ühiskonda arvestades. Kuna andmelahendused on loodud konkreetsetes sotsiaalses kontekstis, on ka andmestunud maailma mõistmise viiside korral oluline lähtuda konkreetse riigi kontekstist. Nende eelduste põhjal oleme sõnastanud viienda postulaadi: **andmestunud maailma mõistmiseks – uuringu planeerimisel, andmete kogumisel, analüüsimisel ja tõlgendamisel – on oluline arvestada uuritava nähtuse ja andmete sotsiaal-kultuurilise kontekstiga.**

Andmete mitmetähenduslikkusest

Kui ühiskonna andmestumisest rohkem rääkima hakati, peeti silmas eelkõige n-ö suurandmeid,⁴ mida defineeriti algselt eelkõige kasvanud andmemahu (*volume*) kaudu (ülevaateks vt nt Schäfer 2016). Hilisemad käsitlused on iseloomustanud suurandmeid viie V-ga ehk mahule lisati sellised omadused nagu mitmekesisus (*variety*), kiirus (*velocity*), tõepärasus (*veracity*) ja väärtus (*value*; Gupta *et al.* 2012). Viimastel aastatel on hakatud suurandmeid kirjeldama, mainides suuremat hulka omadusi (Lupton 2015), mida tuntakse ka suurandmete 13 P-na, väites et suurandmed võivad olla kõikehõlmavad (*portentous*), pahelised (*perverse*),⁵ isiklikud (*personal*), loovad (*productive*), osalised (*partial*), praktikaga seotud (*practices*),⁶ ennustavad (*predictive*), poliitilised (*political*), provokatiivsed (*provocative*), privaatsed (*privacy*), mitmetähenduslikud (*polyvalent*), mitmekujulised (*polymorphous*) ja mängulised (*playful*).

⁴ Suurandmeid on eesti keeles nimetatud ka mahtandmeteks; mõlema nimetuse korral on tegemist metafoorse konstruktsiooniga, mis ei anna üheselt edasi nähtuse mitmetähenduslikkust. Sel põhjusel oleme siinses kogumikus eelistanud seotud sotsiaalsetele protsessidele viitavat terminit „andmestumine“.

⁵ Andmete pahelisis tähendab siin andmetega seotud representatsioonide ambivalent-sust – suurandmed ei tekita mitte ainult suurt elevust seoses tekkivate võimalustega, vaid ka hirmu ja ärevust, sest andmete loomine ja kasutamine ei pruugi suure mahu, lakkamatu loomise ja kasutamise tõttu olla kontrollitav.

⁶ Andmete seotus praktikaga tähendab, et suurte andmekogude loomine ja kasutamine hõlmab mitmesugust üksikisikute ja organisatsioonide andmepraktikat, sealhulgas enda kohta teadlikult teabe kogumist (nt nutikellaga), aga sageli ka enesele teadvustamata andmete loomist näiteks sotsiaalmeedia lehele sisu lisades.

Seetõttu käsitleme kogumiku teises osas lähemalt peamisi suurandmete mõtestamise viise – nii masinõppe ja visuaalsete meetodite kasutamise kasvu suurtes andmemassiivides tähenduslike mustrite leidmiseks kui ka tunnustepõhiselt analüüsilt individipõhistele meetoditele liikumist (sotsiaalsete võrgustike analüüs ja agendipõhine modelleerimine).

Nagu paljusid uusi nähtusi, selgitati ka suurandmeid algselt arvukate metafooride kaudu (Puschmann, Burgess 2014) – suurandmeid kirjeldati allutamist vajava loodusjõu ja tarbitava ressursina. Eriti levinud on (suur)andmete nimetamine naftaks, mis loob Luke Starki ja Anna Lauren Hoffmani (2019) sõnul spetsiifilise ettekujutuse sellest, mida andmetega peaks tegema, ja tekitab mitmeid andme-eetika ja andmeõigluse probleeme, kuna on suunitletud konkurentsile, olelusvõitlusele ja võimuahnusele. Maavarametafooride asemel soovivad kriitikud lähtuda pigem „katkiste andmete“ metafoorist, mis rõhutab, et andmed pole kunagi täiuslikud (Pink *et al.* 2018). Kui siiski loodusvara metafooride juurde jääda, peaks õppima ehk looduskaitse või metsamajanduse valdkonnast ja andmekaevandamise ning väärindamise asemel mõtlema näiteks andmehooldusest (*data stewardship*) või pakkuma välja sootuks alternatiivseid metafoore (Stark, Hoffman 2019).

Teine grupp olulisi turumajanduslikke metafoore, mis andmesuhteid vormivad, on andmeõigluse kontekstis kasutatavad ekspluateerimise (Mühlhoff 2019), isegi koloniseerimise (Chun 2018) metafoorid. Sotsiaalmeediaplatformide kasutajaliidesed näiteks innustavad inimesi sisu jagama, laikima ja postitama, kuna see võimaldab maksimeerida turundus- ja reklaamiklientidele kasulike andmete teket. Kui tavakasutaja jaoks on andmed kõrvalprodukt, siis platvormiomanike ärimudeli perspektiivist on suhtlus väärtuslik seetõttu, et selle tagajärjel tekivad müüdavad andmed. Samas ei ole tavakasutajatel võimalik nn andmedoonorlusest loobuda või sellest mingit kasu saada. Käesoleva kogumiku kolmandasse ossa oleme koondanud individide kognitiivse ekspluateerimise tagajärjel loodud tekstiandmete analüüsimise meetodite ja ka nende tajuprotsesside süstemaatilisele hindamisele keskenduvate meetodite peatükid, kuna neis toetutakse peamiselt nn andmedoonorluse käigus tekkinud andmete analüüsile.

Andmetega seotud ühiskondlikke protsesse on tabavalt kirjeldatud kui sotsiaal-kultuurilist koodi (Chun 2018) või andmeteekonda (Bates

et al. 2016), mille mõtestamiseks peame aduma andmete loomise, kasutamise ja mõtestamise terviklikku protsessi. Selliselt pole andmed mitte üksnes nähtuste konstrueerimise ja mõistmise vahend, vaid ka oluline sisend lugude jutustamisel maailmas toimuvast. Andmete kui koodi mõistmine ja selle põhjal lugude jutustamine on muutunud oluliseks uurimisvahendiks, aga ka -objektiks. Sel põhjusel oleme kogumiku neljanda osa pühendanud lugude jutustamisele andmetest ja andmetega. Andmed ei ole kunagi neutraalsed ega teki „iseenesest“, neid loovad konkreetsetes ajalistes ja ruumilises kontekstis inimesed. Tunnustatud informaatikaproffessor Geoffry Bowker (2005: 183–184) on nentunud, et „toorandmed on oksüümoron“, väites, et andmed on alati mingil moel „küpsetatud“. Kohe, kui määratleme mingite fenomenide vaatlemise, mõõtmise ja märkamise tulemused andmetena, tegeleme nende andmeteks „küpsetamisega“, see tähendab, et andmetesse on alati küpsetatud teatud tähendused, eeldused, küsimused ja muidugi võimusuhted.

Iga andmetöötaja ja uurija peaks seega olema teadlik oma rollist andmete „küpsetamisel“ – nii siis, kui ta andmeid „kogub“, kui ka siis, kui ta andmetest sööda ja kasuliku produkti loob (vt ka ptk 1.1). Bowkeri väidet edasi arendades kirjutavad Lisa Gitelman ja Virginia Jackson (2013), et toorandmete mõiste mõjub analüütikutele peibutavalt, sest sellesse on peidetud pika ajalooga ja ideoloogiliselt võimas positivistlik eeldus (vt ka Markham 2016), et arvulised andmed on kuskil meist sõltumata olemas, et andmed eelnevad faktidele, et need on meie teadmiste alus, midagi objektiivset, ilmselget ja läbipaistvat, mis on vaja üksnes kokku korjata ja hoolikalt ära mõõta. Paljudes era- ja ka uurimissituatsioonides räägitakse lisaks „isetekkelistest“ (mõnikord ka „loomulikult esinevatest“) andmetest, mida kriitikute sõnul samuti tegelikult olemas ei ole. Nendele diskussioonidele tuginedes oleme sõnastanud andmestunud maailma mõistmise kuuenda postulaadi: **andmed ei teki iseenesest, vaid luuakse uurija valikute tulemusel, mis tähendab, et uuringu ülesehitusest sõltub, milliseid järeldusi on uuringu põhjal võimalik teha.**

Andmestunud ühiskonna kontekstis pole muutunud niisiis mitte ainult andmed, nende loomine ja kasutamine, vaid andmetöö üldiselt (Fuchs 2018; Rossi 2019), hõlmates lisaks traditsioonilisele andmeanalüütiku tööle aina enam ka internetikasutajate valdavalt teadvustamata tööd ja veebipõhiste andmelahenduste testijate (n-ö klikitööliste)

sageli halvasti tasustatud panust. Andmeid loovad platvormid on seega muutunud nii oluliseks andmeallikaks kui ka sotsiaalseks reaalsuseks. Selle kogumiku viiendas osas käsitlemegi lähemalt nihet platvormipõhisele uurimusele ning näitlikustame seda, millist rolli võivad platvormid jt andmetaristud mängida uurimisyvahendi ja -objektina.

Metodoloogilised nihted

Andmestumise kontekstis on teadmusloome aruteludesse põimitud kaks kesksel teemat (Veltri 2017; Thylstrup *et al.* 2019): 1) vaidlused mõõtmis- tehnikate üle, st loodetakse, et uued andmed võimaldavad objektiivse- malt mõõta inimeste loodud reaalsust, ning 2) vaidlused traditsiooniliste (nt statistiliste) ja arvutuslike (nt masinõppe) meetodite üle. Tuntu- mad näited on siin käsitlused „teooria lõpust“ (Anderson 2008) ning „kirjeldavast empirismist“ (Kitchin 2014b), kus väidetakse, et hüpotee- side ja mudelite testimise ning teoreetiliste mudelite kinnitamise meetod on aegunud ja selle asemel tuginevad andmetest juhitud (*data-driven*) analüüsid korrelatiivsetele seostele, selgitamata nende seoste aluseks ole- vaid sotsiaalseid mehhanisme (Anderson 2008). Selle arusaama kohaselt väheneb teooria roll uuringutes märkimisväärselt. Hiljutises empiirilises uuringus, kus analüüsiti teooria lõpu hüpoteesist inspireeritult tead- miste loomise praktikaid, need arengusuunad siiski kinnitust ei leidnud (Masso *et al.* 2020).

Vastusena teooria lõpu hüpoteesile ja arvutuslike meetodite kaitseks on rõhutatud arvutuslike meetodite suurt varieeruvust (Hindman 2015), mis peaks võimaldama igale uurimisprobleemile vastamiseks sobivaima lahenduse leidmise. Arvutuslikud meetodid võivad olla nii deduktiivsed kui ka induktiivsed. Mõni autor väidab koguni, et teatud nähtuste esinemise põhjusi selgitada võimaldavate arvutuslike meetodite populaarsusega kaasneb induktiivne hüpe sotsiaalteadustes (Bengio *et al.* 2019). Üks induktiivsel loogikal põhinevaid arvutuslikke meetodeid on masinõppe kasutamine analüüsis (vt ptk 2.4; aga ka mujal, nt ptk-d 2.1, 2.2, 2.3, 3.3). Masinõppet peetakse ideaalseks lahenduseks komplekssete nähtuste selgitamisel, sest see ei testi hüpoteesi, vaid genereerib küsimused varasemate kogemuste süstemaatilise hindamise tulemusena (Breiman 2001; Bengio *et al.* 2019). Hiljutised edusammud masinõppe meetodites (Bengio *et al.*

2019) püüavad lisaks mustrite tuvastamisele andmetes selgitada ka kauasaalseid seoseid. Teisisõnu, selle asemel et vastata küsimusele *mis*, püütakse leida vastuseid küsimusele *miks* ehk siis selgitada sisust lähtuvaid, seni vastamata küsimusi. Nende diskussioonide ning varasema empiirilise uurimistöö põhjal oleme sõnastanud andmestunud maailma uurimise seitsmenda postulaadi: **teooria pole surnud, ehk vaatamata uutele andmetele, analüüsitehnikatele, tarkvarale ja meetoditele algab andmestunud maailma uurimine endiselt küsimuse püstitusest.**

Pluralism meetodites

Esimeseks vastuseks uute andmete tekkele ja andmemahu suurenemisele on olnud arvutuslik sotsiaalteadus (*computational social science*; vt nt Cioffi-Revilla 2014), kus (sageli suure võimsusega) arvutustehnoloogiad kasutatakse sotsiaalsete nähtuste analüüsimiseks, modelleerimiseks ja simuleerimiseks. Arvutusliku sotsiaalteaduse rakendamisest on arvukalt näiteid (Cioffi-Revilla 2014; Park *et al.* 2015). Näiteks töötasid Emmanuel Lazega ja Tom Snijders (2016) välja võrgustikuanalüüsi meetodi dünaamiliste ja suuremahuliste andmete analüüsimiseks; Daniel Dellaposta koos kolleegidega (2015) meetodid veebiandmete abil poliitilise orientatsiooni ruumilise ja ajalise dünaamika analüüsiks; Dirk Helbing (2013) simulatsioonimeetodid võrgustunud riskidega⁷ toimetuleku analüüsimiseks.

Neid algselt absoluutse tõena esitatud arvutuslikke meetodeid on sageli kritiseeritud, sest toimunud nihked konkreetsetes analüüsitehnikates või -meetodites pole toonud kaasa loodetud metodoloogilisi uuendusi. Nn kolmanda tee otsingud on lisaks traditsioonilistele statistilistele ja uuematele arvutuslikele meetoditele pakkunud mitmeid alternatiive. Lisaks andmeteaduse universaalsete põhimõtete ühtlustamisvajadusele (Slota *et al.* 2020) või arvutuslike põhimõtete kohandamisele sotsiaal- (Cioffi-Revilla 2014) ja humanitaarteadustele (Schäfer, Es 2017) pakkus Lev Manovich (2017) alternatiivina välja kultuurianalüütika, mis rakkud suuremahuliste kultuuriandmestike analüüsimisel arvutuslikke

⁷ Võrgustunud risk on olukord, kus omavahel ühendatud globaalsed võrgustikud loovad üksteisest sõltumatud süsteimid, mida on keeruline mõista ja kontrollida.

analüüsitehnikaid. Samas on ka Manovichi üldistusi ja järeldusi korraldvalt kahtluse alla seatud, nentides, et Instagrami pildid võimaldavad siiski järelduste tegemist Instagrami-põhise eneseesitluse sotsiaalsete normide ja sellega seotud tõlgenduste, mitte aga kultuurigeograafiliste (nt millises linnas elavad kõige mornimad inimesed), arvuliselt eristatavate fenomenide kohta.

Suhteliselt pika traditsiooniga on Richard Rogersi meeskonna (Amsterdami Ülikoolis) digimeetodid (Rogers 2013, 2019), mis kasutavad andmete loomiseks ja analüüsiks veebikeskkondade rakendusliideseid jm tehnilisi vahendeid ning kategoriseerivad andmed ja meetodid n-ö digitaalsena sündinuks ja digiteerituteks.⁸ Digimeetodid püüavad kasutada internetti ja sotsiaalmeediat uurimismeetodi ja uurimisvahendina, vastandudes selgelt n-ö virtuaalsetele meetoditele, mis kohandavad varasemalt kasutatud sotsiaalteaduslikud meetodid veebikeskkonnas kasutatavaks. Teemast lähemalt huvitatud lugejatel tasuks tutvuda Richard Rogersi monograafiatega (2013, 2019).

Niisiis valitseb uurimustes endiselt meetodite ja käsitlusviiside pluralism – tekkinud arvutuslike meetodite kõrval kasutatakse paralleelselt ja kombineerituna klassikalisi kvalitatiivseid ja kvantitatiivseid meetodeid, mis võimaldavad lisaks põhjuslike seoste kirjeldamisele ja ennustamisele ka komplekssete võrgustunud nähtuste põhjuste mõistmist. Vahepealsete vaidluste käigus pakuti näiteks „suurte“ ja „väikeste“ andmete kõrvale ka rikaste ja tihedate andmete kategooriat. Kvalitatiivse ja eriti etnograafilise uurimisega tegelevad ühiskonnateadlased leidsid, et nende tõlgendatavad materjalid ei kvalifitseeru mitte mingil juhul väikesteks andmeteks, ent on suured teisel moel kui suurandmed, keskendudes uuritava nähtuse kontekstile ja tihedale tõlgendusele⁹ (vt ka ptk 6.2). Sellest lähtuvalt oleme sõnastanud andmestunud maailma uurimiseks olulise kaheksanda postulaadi: **segameetodite paindlik kasutamine**

⁸ Digitaalsena sündinud (*natively digital*) meetodid on näiteks soovitusalgoritmide süsteemid ja folksonoomiad (*folksonomy*) ehk meetodid, mis põhinevad interneti toimimisel. Digiteeritud (*digitalized*) on n-ö traditsioonilised sotsiaalteaduslikud meetodid, mis on veebi viidud (ehk virtuaalsed meetodid, nt veebipõhine küsimustik).

⁹ Tihedus viitab siin etnograafia terminile „tihe kirjeldus“ (*thick description*), mis tähendab uuritava fenomeni detailirohkust ja kontekstitundlikku kirjeldust.

võimaldab tagada tervikliku pildi andmestunud maailmas toimuvatest komplekssetest nähtustest.

Andmestunud maailma mõistmise hõlbustamiseks ning paradigmaatiliste metodoloogiliste diskussioonide ja meetodite paljususe kontekstis on pakutud ühe lahendusena kriitilist realismi (Bhaskar 2008; vt ka Kennedy, Moss 2015) ja pragmatismi (vt nt Eklund *et al.* 2019). Kui kriitiline realism lähtub metodoloogilise pluralismi ideest ja toob uurimisprotsessis kesksele kohale inimese kui aktiivse toimija (*agency*), siis pragmatism soovitab varasemate kvalitatiivsete vs. kvantitatiivsete või andmetest tuletatud (*data-driven*) vs. interpretatiivsete käsitluste ning „lõpliku tõe“ otsimise asemel lähtuda „lõpliku kasulikkuse“ põhimõttest. Mõlemal juhul soovitatakse (uurimis)probleemide lahendamiseks kasutada segameetodeid, kus konkreetsete analüütiliste sammude asemel lähtutakse algsest uurimisküsimusest, probleemiseadest ja uurimuse eesmärkidest. Nende diskussioonide ja käesolevas raamatus esitatud praktiliste Eesti näidete varal oleme pakkunud andmestunud maailma uurimise üheksanda postulaadi: **uurimistööks sobivate meetodite valikul on paradigmasõdade asemel mõistlik lähtuda kasulikkuse põhimõttest.**

Andmestumise mõistmine eeldab nii andmesubjektiga arvestamist kui ka andmetöötaja või uurija refleksiivse võimekuse arendamist, hindamaks andmete kogumise ja analüüsiprotsessi asjakohasust. Ka siinses kogumikus oleme pidanud oluliseks rõhutada uurija refleksiivsust, mistõttu on sellele pühendatud kogumiku kuues osa, mis keskendub meetoditele, mille keskmeks on iseteadlik ja teovõimeline inimene. Rakendades kriitilise realismi põhimõtteid andmestumisele, muutuvad andmeprotsessis oluliseks ka aktiivsed andmekodanikud – inividid, kellel on võimalik avatud andmete abil andmesubjektide huve kaitsta ja ka ise analüütilisse tegevusse panustada. Selline tõusev andmeaktivism (Milan, Velden 2016) või uued andmekodakondsuse vormid (Hintz *et al.* 2019) annavad muu hulgas võimaluse kodanikuosaluse täiustamiseks. Eeltoodust ajendatult oleme sõnastanud andmestunud maailma uurimise kümnenda postulaadi: **uurija refleksiivsus ehk kriitiline võime hinnata oma tegevuse tagajärgi on oluline valikute tegemisel andmete, meetodite ja tõlgendusviiside mitmekesisuse kontekstis, uuringu ühiskondliku kasu tagamisel ning võimalike kahjude vältimisel.**

Soovime rõhutada, et andmestunud ühiskonna uurimiseks kogu tema mitmekesisuses peab olema avatud kombineerimisele ning sõltuvalt uuringu eesmärgist võivad olla kasulikud nii digitaalsed, virtuaalsed, arvutuslikud kui ka interpretatiivsed meetodid. Vaatamata nihetele andmetes, meetodites ja ühiskonnas on andmestunud ühiskond endiselt ühiskond, inimesed endiselt inimesed ja sedalaadi komplekssete süsteemide mõtestamiseks on mõistlik kombineerida asjakohaseid meetodeid ning kasutada mitmekesiseid kättesaadavaid tööriistu ja luua uusi. Teooriat on andmeühiskonna mõistmiseks samuti vaja enam kui iial varem.

Raamatu ülesehitus

Raamat koosneb 24 sisulisest üksteisega tihedalt seotud peatükist, mis on jaotatud kuude temaatilisse ossa:

1. Maailma andmestumine
2. (Suur)andmete mõtestamine
3. Teksti- ja tajuandmete analüüs
4. Lugude jutustamine andmetest ja andmetega
5. Platvormid ja andmetaristud uurimisvahendi ja -objektina
6. Iseteadlik ja teovõimeline inimene andmeühiskonnas

Sellise temaatilise jaotusega sõnastame kuus peamist muudatust, mis meie hinnangul andmestumist ja selle uurimist iseloomustavad, ning soovime rõhutada, et ranged meetodite klassifikatsioonid, näiteks kvalitatiivne vs. kvantitatiivne, positivistlik vs. interpretatiivne, ei pruugi andmeühiskonna uurimisel kehtida ning andmestumisega seotud nähtuste mõistmiseks ja selgitamiseks võib olla vajalik rakendada kõiki neid meetodeid. Jaotus koorus välja raamatu koostamise, toimetamise ning kaasautoritega toimunud arutelude ja ühiste põhimõtete otsingute käigus.¹⁰

Raamatu esimene osa „**Maailma andmestumine**“ keskendub andmeühiskonna uurimise eetilistele, sotsiaalsetele ja õiguslikele nüanssidele. Kas see, et mingi informatsioon on andmeteks muudetav või

¹⁰ Nt Eesti sotsiaalteaduste aastakonverentsil 2019. aastal aprillis toimunud kahepäevasel ettekande- ja arutelude sessioonil „Andmestunud ühiskonna uurimise meetodid“ ning 2020. aasta jaanuaris toimunud andmestunud ühiskonna uurimise meetodite terminoloogia arendamise seminaril.

tehniliselt andmetena kättesaadav, tähendab tingimata, et seda peaks sellisena kasutama? Millised on andmestumise ja andmeanalüütika võimalikud kasud ja kahjud ning kuidas neid kaaluda? Kas kõikide andmetoimingute taga peaks seisma õigluse ja kahju vältimise ideaal? Esimese osa avavad Anu Masso, Triin Vihalemm ja Leno Saarniit peatükiga „Andmepõhine muutuste juhtimine“. Andmepõhine muutuste juhtimine tagab teadlikud otsused, vähendab määramatust ja kiirendab otsuste langetamist. Ohuks on väärtuspõhised eriarvamused, mille vältimiseks soovivad peatüki autorid lähtuda andmeõigluse – rakendada sotsiaalse õigluse põhimõtteid andmetöös. Üksikisikute huvide kaitsele keskendub ka järgmine peatükk „Andmete õiguslik kaitse ja kasutamine teadustöös“ (autorid Aleksei Kelli, Irene Kull, Age Värvi), mis soovib teadusuuringu korraldajal arvestada andmetöös peamiselt nelja kaasneva õigusliku aspektiga: 1) vajadus arvestada andmebaasi looja õigustega, 2) vajadus kaitsta isikuandmeid, 3) erandlik õigus kasutada andmeid teadustöö eesmärgil ja 4) nõue andmete kasutamisel kaitsta andmesubjekti huve. Peatükk „Avaandmete kasutusvõimalused ja piirangud“ (autorid Mai Beilmann, Ave Roots) rõhutab, et tänu avaandmetele on kvaliteetsetele andmetele juurdepääs varasemast märksa enamatel inimestel. Teisalt ei sobi andmete avamine kõigile uuringu raames toodetud andmetele ja sellega kaasnevad ka ohud – avaandmete kasutamine eeldab lisaoskusi andmete kvaliteedi hindamiseks ja andmete eetiliseks kasutamiseks. Raamatu esimese osa võtab kokku peatükk „Eetika ja privaatsus“ (autorid Katrin Tiidenberg, Andra Siibak), mille peamine eesmärk on innustada uurijaid ja andmetöötajaid küsima endalt teatud põhimõttelisi küsimusi ning sellest johtuvalt otsustama võimalikult eetilise andmete kogumise, puhastamise, talletamise, analüüsamise ja hävitamise viisi kasuks. Eetika ja privaatsus on ka raamatu kõiki peatükke läbiv ühine teema.

Raamatu teine osa „**(Suur)andmete mõtestamine**“ keskendub arvutusliku sotsiaalteaduse nurgakiviks kujunenud andmetele ja meetoditele. Juttu tuleb visualiseerimise abil andmetes leiduvate mustrite hindamisest, tunnusepõhisel analüüsilt indiviidipõhisele analüüsile (võrgustiku- ja agendipõhise modelleerimise meetodid) liikumisest ning viimastel aastatel sotsiaal- ja humanitaarteaduslikes uuringutes valdavaks muutunud masinõppe meetodite rakendustest. Anto Aasa peatükk

„(Suur)andmete visuaalne esitamine“ annab ülevaate visualiseerimistehnikatest, mis aitavad andmeid mõtestada ja neid väärtuslikuks ressursiks muuta. Analüüsi kiiruse ja keerukuse kasvuga, ent ka interaktiivsete visualiseerimisvahendite rohkusega kaasneb üks suurandmete visualiseerimise peamisi kitsaskohti – pakutavate vahendite ülekülluse kontekstis peab uurija suutma teha mõistlikke valikuid värvi- ja sümbolilahenduste osas. Innar Liivi peatükk „Sotsiaälvõrgustike analüüs“ annab detailse ülevaate sotsiaalsete struktuuride tehnikatest, vahenditest ja meetoditest mustrite avastamiseks. Sotsiaalsete võrgustike meetodid aitavad mõista ja leida uudseid lahendusi näiteks sellele, kuidas institutsioonid ja riik inimestega suhtlevad. Sarnaselt sotsiaälvõrgustike meetodiga keskendub ka Kuldar Taveteri peatükk „Agendipõhine modelleerimine“ keerukatele sotsiaalsetele suhtevõrgustikele, milles osalejad ei pruugi käituda ratsionaalselt, mistõttu nende käitumine ei pruugi olla ennustatav. Agendipõhine modelleerimine selgitab keeruliste nähtuste (nt Brexiti mõju majandusele, ingliskeelsete elanike arvu kasvu mõju eesti keele elujõulisusele või koroonaviiruse leviku kiirus) aktiivseid olemeid ehk agente ja nende vastastikust mõju. Raamatu teise osa lõpetab Toomas Kirdi peatükk „Masinõppe meetodid ja rakendused suurandmete töötlemisel“. Peatükis tutvustatud masinõppe meetodeid kasutatakse näiteks klientide ostukorvi analüüsimiseks äriettevõtte müügi suurendamise eesmärgil või inimeste linna kolimise põhjuste selgitamiseks individuaalsete profiilide analüüsi kaudu. Masinõppe leviku ja meetodite paljususe kontekstis on käsitluse peamisteks ülesanneteks kvaliteetsete andmete tagamine, tööprotsessi kiirendavate töövahendite paralleelne kasutamine ja treeningandmete kättesaadavus.

Kolmas osa „**Teksti- ja tajuandmete analüüs**“ pakub sissejuhatuse meetoditesse, mis tegelevad struktureerimata (teksti)andmete viimisega struktureeritud kujule. See osa näitab ilmekalt, et andmed võivad tekkida inimeste kognitiivsete ressursside n-ö ekspluateerimise tagajärjel sotsiaalmeedia keskkondades (inimestele meeldivaid emotsioone loovate nn klikitegevuste tulemusel) (Mühlhoff 2019). Teisalt on meil üha enam võimalusi kasutada erinevaid tööriistu samade tajuprotsesside uurimiseks (nt kuidas inimesed tajuvad erinevaid tekste, andmetehnoloogiaid ja neis sisalduvat meelestatust). Selliselt kajastab see osa sotsiaal- ja humanitaarteaduste n-ö kognitiivset pööret – tekkinud on võimalused

uurida tekstides leiduvaid hoiakuid ja meelestatust, aga ka meelestatuse kujunemise kognitiivseid mehhanisme. Raamatu kolmanda osa juhatab sisse Kristel Uihoaia peatükk „Tekstikaave“, mis näitlikustab Eesti erakondade valimisprogrammide põhjal, kuidas tekstikaave meetodid võimaldavad otsida suurtest tekstikogudest uut informatsiooni, mustreid ja korrapära. Peatükk annab tänuväärse ülevaate eripäradest, mis on seotud eestikeelsete andmete töötlemisega, ning tutvustab tekstikaave projekti eri etappe ja enim kasutatavaid meetodeid. Kadri Muischneki ja Liina Lindströmi peatükk „Digitaalsed tekstiandmed ja korpuslingvistika“ annab ülevaate sellest, kuidas rakendada suuri tekstikollektsioone ehk -korpusi, saamaks süstematiseeritud infot nii keele arengusuundade kui ka veebikorpustes kujutatud ühiskonna kohta. Tekstiandmete üheks levinumaks analüüsiviisiks on hoiakute automatiseeritult mõõtmine ja meelestatuse analüüs, mida käsitlevad järgmises peatükis põhjalikumalt Mare Koit ja Haldur Õim. Autorid näitavad, et vaatamata meelestatuse hindamise automatiseerimisele, näiteks masinõppe meetodeid rakendades, tuleb eri tüüpi tekstides positiivse või negatiivse meelestatuse väljendamiseks kasutatavaid keelelisi erinevusi enamasti hinnata ka lähilugemise abil. Kolmanda osa lõpetab Kristian Pentuse ja Andres Kuusiku peatükk „Pilgujälgimine“, mis kirjeldab, kuidas pilgujälgimise abil on võimalik selgitada seda, kuidas inimene maailma näeb ja kogeb, kuidas on fikseeritud inimese tähelepanu ning kuidas optimeerida andmekuvamise lahendusi.

Neljas osa „**Lugude jutustamine andmetest ja andmetega**“ näitab, kuidas kogutud ja töödeldud andmed võimaldavad jutustada veenvaid, haaravaid, mobiliseerivaid, reaalsust, ajalugu ja tulevikku ümberkujundavaid lugusid nii tänapäeva- kui ka minevikuühiskondadest. Oskus andmete põhjal kas visuaalselt (ptk 2.1) või tekstiliselt lugusid jutustada ning andmete (varjatud) diskursiivseid tähendusi mõista on muutunud andmeühiskonna uurijate üheks keskseks kompetentsiks. Marek Tamme ja Hembo Pagi peatükk „Digiajaloo ja -arheoloogia uurimismeetodid“ tutvustab meetodeid, mis aitavad mitmekesistada meie teadmisi minevikuühiskondadest. Näiteks antakse peatükis ülevaade käsikirjaliste tekstide tuvastamise, ajaloolise võrgustikuanalüüsi, aga ka laser-skaneerimise ja konstruktsioonianalüüsi võimalustest. Ragne Kõutsi ja Marju Himma peatükk „Ajakirjandus andmeajastul“ näitab, kuidas

andmeajakirjandus informatsiooni vahendades sotsiaalset tegelikkust konstrueerib. Andmetest küllastunud ühiskonnas lasub andmeajakirjaniku roll kõigil andmete kogumise, töötlemise ja esitamisega tegelevatel analüütikutel, kes peaks vältima moonutatud pildi loomist tegelikkusest. Nii andmeajakirjanikud kui ka teised andmetega tegelevad inividid ja institutsioonid vastutavad valeinfo leviku ja andmetega manipuleerimise piiramise eest. Järgmises peatükis „Diskursusanalüüs“ (autorid Katrin Tiidenberg, Anu Masso, Maili Pilt, Liisi Laineste) on lähemalt käsitletud kolme meetodit: multimodaalne kvalitatiivne, kriitiline tehnokultuurline diskursusanalüüs ja Q-metodoloogia. Peatükk näitab, et kriitilised tähendusi, tegevusi-praktikaid, identiteete, suhteid, põhimõtteid, seoseid, märgisüsteeme ja teadmisi analüüsivad võtted on andmestunud ühiskonna mõistmisel ja selgitamisel hädavajalikud. Samas on uued analüüsi tööriistad, andmekogumise ja -analüüsi keskkonnad (nt Ken-Q Analysis), aga ka kvalitatiivse analüüsitarkvara kasutamine loonud võimaluse analüüsiprotsessi kiirendamiseks ja läbipaistvamaks muutmiseks. Sarnaselt käsitleb järgmine peatükk „Narratiivid sotsiaalmeedias“ (autorid Liisi Laineste, Maili Pilt) sotsiaalmeedia levikuga kaasnevate uute multimodaalsete narratiivsete nähtuste analüüsimise võimalusi, loomaks inimeste argisuhtlusest ja kogemustest mitmekihilisi tõlgendusi.

Viies osa võtab lähema vaatluse alla „**Platvormid ja andmetaristud uurimisvahendi ja -objektina**“. Osa keskendub nii teiste andmebaaside (nt filmiandmebaas) rollile ühiskonna ja kultuuri mõtestamisel kui ka kommertsteenuste tulemusena loodud andmete (nt sotsiaalmeedia-, mobiili- ja internetiandmed) kasutamisele ajaliste ja ruumiliste ning sotsiaalsete protsesside mõistmisel. Indrek Ibrus ja Maarja Ojamaa näitavad peatükis „Audiovisuaalne kultuur, metaandmed ja lingüüsilisus“ unikaalse kultuurianalüüsi võimalusi Eestis ja rahvusvaheliselt, tuginedes filmiandmebaasi metaandmetele. Filmide autorite seotuse ja filmide märksõnavõrgustike ajaperioodide kaupa võrdlemine loob uued võimalused varem märkamata jäänud tendentside, aga ka ootamatute anomaaliade avastamiseks. Olle Järve ja Kerli Müüriseppa peatükk „Sotsiaalmeedia andmete sotsiaal-ruumiline analüüs“ avab lisaks eelnevalt käsitlemist leidnud tekstiliste sotsiaalmeediaandmete analüüsile (nt ptk-des 3.1 ja 3.2) lugejale ka platvormiandmete potentsiaali ruumilisi ja sotsiaalseid struktuure käsitlevate uuringute jaoks. Platvormiandmete

analüüsile keskenduvad ka Anna Veremchuk ja Jaan Masso peatükis „Internetiandmete kasutamine tööturu-uuringuteks“. Mõlema peatüki autorid näitavad sotsiaalmeedia ja CV Keskuse veebikeskkonna näitel platvormiandmete unikaalset potentsiaali näiteks tööturu-uuringuteks või inimeste ruumilise käitumise selgitamiseks. Platvormide eelis uuringutandrina peitub selles, et uurijal on suhteliselt odavalt võimalik koguda uurimistööks vajalikke andmeid, muidugi juhul, kui andmete omanik seda lubab. Samuti annavad internetiandmed teiste andmestikega kombineerituna võimaluse saada täpsemaid ja usaldusväärsemaid tulemusi. Siiri Silma ja Olle Järve peatükk „Mobiiltelefonid ühiskonna ajalis-ruumilises analüüsis“ rõhutab, et mobiilpositsioneerimise meetodite paljud tugevused (reaalajas suhteliselt täpne liikuvuse jälgimine) kaaluvad selgelt üle nõrkused (nt valimi kallutatus), mistõttu meetod leiab üha enam kasutust nii passiivse (anonüümsed suuremahulised kõnetoimingute asukohapõhised andmed) kui ka aktiivse positsioneerimise (väikeses mahus nutitelefoni positsioneerimine) kaudu loodud andmete kujul. Meetod annab võimaluse reaalses ühiskonnas dünaamiliselt jälgida ning targalt juhtida, näiteks liikluse reguleerimise, kriiside juhtimise ja asukohapõhiste teenuste pakkumise kaudu.

Raamatu viimane osa „**Iseteadlik ja teovõimeline inimene andmeühiskonnas**“ rõhutab seda, et ka andmestunud ühiskonnas, näiliselt isetek kivite andmete ja iseõppivate masinate kontekstis, on sotsiaalse elu keskseks figuuriks iseotsustav, sageli ebaratsionaalne inimene, kes võib, aga ei pruugi andmeloome tehnoloogiaid eesmärgipäraselt kasutada. Sellises kontekstis on uurija jaoks jätkuvalt oluline mõista enda ja uuritava vahelisi suhteid ning olla oma uurimisküsimuste, uurimistegevuste ja andmeprotsesside suhtes süstemaatiliselt refleksiivne. Mare Ainsaar, Indrek Soidla ja Ave Roots näitavad peatükis „Küsitlusuuringud internetis“, et üha kasvav veebipõhiste uuringute osakaal võimaldab küll kvaliteetseid küsitlustulemusi, ent veebiuuringuil on ka omad puudused ning nendega kaasneb terve hulk müüte. Veebipõhiste küsimustike juures tuleb mees pidada, et kõrget vastamismäära tasub oodata vaid kõrge internetikasutusega rühmades, et mobiiltelefonis on pikale küsimustikule vastamine ebamugav ning et veebiküsimustik ei ole sageli odavam kui mõni traditsiooniline küsitlusmeetod. Uurija aktiivset rolli ning meetodi potentsiaali tundlike teemade ja rühmade uurimisel

rõhutab Maria Murumaa-Mengel peatükis „Veebiintervjuud, projektiivtehnikad ja loovuurimismeetodid“. Peatükk tutvustab veebipõhiste, voog- ja viivisintervjuudega seonduvaid võimalusi, riske ja eetilisi mõttekohti. Lisaks pakub peatükk täiendavaid mõtteid veebipõhise kvalitatiivse andmekogumise rikastamiseks Eestis endiselt vähe kasutatud projektiiv- ja loovuurimismeetoditega. Katrin Tiidenbergi peatükk (6.2) ja Pille Pruulmann-Vengerfeldti peatükk (6.4) käsitlevad mõlemad andmestunud ühiskonna etnograafilist uurimist. Andmeühiskonna etnograaf tegeleb tehnoloogia kasutamise, veebipõhiste gruppide teguviiside, platvormimajanduse või andmeühiskonda korraldavate ja taastootvate automatiseeritud protsesside uurimisega. Teemast sõltumata ühendab etnograafilisi käsitusi eeldus, et indiviidide, gruppide ja nende loodud tehnoloogiate mõistmiseks on oluline teada, kuidas nad tähendusi loovad ja tähendusi omistavad. Katrin Tiidenberg keskendub andmeühiskonna etnograafilise uurimise praktilistele detailidele, kirjeldades välitöid, uurimisvälja piiritlemist, vaatlust ja väljamärkmete tegemist. Pille Pruulmann-Vengerfeldt kirjeldab autoetnograafilisi ehk uurimistöö teostajast lähtuvaid ja uurija kogemuste nüansirikkal, ent süstemaatilisel analüüsil põhinevaid arusaamu.

VIIDATUD KIRJANDUS

- Ahas, R.; Aasa, A.; Silm, S.; Aunap, R.; Kalle, H.; Mark, Ü. 2007. Mobile Positioning in Space–Time Behaviour Studies: Social Positioning Method Experiments in Estonia. – *Cartography and Geographic Information Science* 34, 4, 259–273. <https://doi.org/10.1559/152304007782382918>.
- Anderson, C. 2008. The End of Theory. – *Wired* 16, 7, 108.
- Archer, M. S. 2015. *Generative Mechanisms Transforming the Social Order*. Springer International Publishing. http://sfx.ethz.ch/sfx_locator?sid=ALEPH:EBI01&genre=book&isbn=9783319137728.
- Bates, J.; Lin, Y.-W.; Goodale, P. 2016. Data journeys: Capturing the socio-material constitution of data objects and flows. – *Big Data & Society* 3, 2. <https://doi.org/10.1177/2053951716654502>.
- Bengio, Y.; Deleu, T.; Rahaman, N.; Ke, R.; Lachapelle, S.; Bilaniuk, O.; Goyal, A.; Pal, C. 2019. A Meta-Transfer Objective for Learning to Disentangle Causal Mechanisms. – *ArXiv.Org*. <http://search.proquest.com/docview/2174081487/?pq-origsite=primo>.

- Beraldo, D.; Milan, S. 2019. From data politics to the contentious politics of data. – *Big Data & Society* 6, 2, 2053951719885967. <https://doi.org/10.1177/2053951719885967>.
- Bhaskar, R. 2008. *A Realist Theory of Science*. Verso.
- Bowker, G. C. 2005. *Memory Practices in the Sciences*. MIT Press.
- Breiman, L. 2001. Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). – *Statistical Science* 16, 3, 199–231. <https://doi.org/10.1214/ss/1009213726>.
- Chun, W. H. K. 2018. Queerying Homophily Muster der Netzwerkanalyse. – *Zeitschrift Für Medienwissenschaften* 10, 1, 131–148. <https://doi.org/10.14361/zfmw-2018-0112>.
- Cioffi-Revilla, C. 2014. *Introduction to Computational Social Science: Principles and Applications*. Springer.
- Couldry, N.; Mejias, U. 2018. Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. – *Television and New Media*, 1–14.
- Couldry, N.; Mejias, U. 2019. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford University Press.
- Dalton, C. M.; Taylor, L.; Thatcher, J. 2016. Critical Data Studies: A dialog on data and space. – *Big Data & Society* 3, 1. <https://doi.org/10.1177/2053951716648346>.
- Dellaposta, D.; Shi, Y.; Macy, M. 2015. Why do liberals drink lattes? – *American Journal of Sociology* 120, 5, 1473.
- Dijck, J. van; Poell, T.; Waal, M. de 2018. *The Platform Society: Public Values in a Connective World*. Oxford University Press.
- D'Ignazio, C.; Klein, L. F. 2020. *Data Feminism*. Cambridge, MA: MIT Press.
- Eklund, L.; Stamm, I.; Liebermann, W. K. 2019. The crowd in crowdsourcing: Crowdsourcing as a pragmatic research method. – *First Monday* 24, 10. <https://doi.org/10.5210/fm.v24i10.9206>.
- EP 2016 = European Parliament and Council of the European Union. Regulation on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (Data Protection Directive), L119, 4 May 2016, implementation date 25 May 2018.
- Fuchs, C. 2018. Capitalism, Patriarchy, Slavery, and Racism in the Age of Digital Capitalism and Digital Labour. – *Critical Sociology* 44, 4/5, 677–702. <https://doi.org/10.1177/0896920517691108>.
- Gitelman, L. ja Jackson, V. 2013. „Raw Data“ is an Oxymoron. Cambridge MA: MIT Press.
- Goriunova, O. 2019. The Digital Subject: People as Data as Persons. – *Theory, Culture and Society* 36, 6, 125–145. <https://doi.org/10.1177/0263276419840409>.

- Gupta, R.; Gupta, H.; Mohania, M. 2012. Cloud computing and big data analytics: What is new from database perspective? – *Big Data Analytics: Proceedings of First International Conference, BDA 2012, New Delhi, India, December*, Springer, 42–61.
- Helbing, D. 2013. Globally networked risks and how to respond. – *Nature* 497 (7447), 51–59. <https://doi.org/10.1038/nature12047>.
- Hepp, A. 2020. *Deep Mediatization*. Routledge.
- Hindman, M. 2015. Building Better Models: Prediction, Replication, and Machine Learning in the Social Sciences. – *The Annals of the American Academy of Political and Social Science* 659, 1, 48–62. <https://doi.org/10.1177/0002716215570279>.
- Hintz, A.; Dencik, L.; Wahl-Jorgensen, K. 2019. *Digital citizenship in a datafied society*. Polity.
- Hopkins, P. 2019. Social geography I: Intersectionality. – *Progress in Human Geography* 43, 5, 937–947. <https://doi.org/10.1177/0309132517743677>.
- Just, N.; Latzer, M. 2017. Governance by algorithms: Reality construction by algorithmic selection on the Internet. – *Media, Culture and Society* 39, 2, 238–258. <https://doi.org/10.1177/0163443716643157>.
- Kennedy, H.; Moss, G. 2015. Known or knowing publics? Social media data mining and the question of public agency. <http://eprints.whiterose.ac.uk/91180/1/2053951715611145.full.pdf>.
- Kitchin, R. 2014a. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Sage.
- Kitchin, R. 2014b. Big Data, new epistemologies and paradigm shifts. – *Big Data & Society* 1, 1. <https://doi.org/10.1177/2053951714528481>.
- Lazega, E.; Snijders, T. A. B. 2016. *Multilevel Network Analysis for the Social Sciences: Theory, Methods and Applications*. Methodos Series Book 12. Springer. http://sfx.ethz.ch/sfx_locator?sid=ALEPH:EBI01&genre=book&isbn=9783319245201.
- Lupton, D. 2015. The Thirteen Ps of Big Data. <https://simplysociology.wordpress.com/2015/05/11/the-thirteen-ps-of-big-data/>.
- Lupton, D. 2020. *Data Selves: More-than-human Perspectives*. Polity.
- Manovich, L. 2017. *Cultural Analytics, Social Computing and Digital Humanities*. – *The datafied society: Studying culture through data*. Eds. Mirko Tobias Schäfer, Karin van Es. Amsterdam University Press, 55–68.
- Markham, A. N. 2016. Troubling the Concept of Data in Qualitative Digital Research. – U. Flick (ed.), *The Sage Handbook of Qualitative Data Collection*. Sage, 511–524.

- Masso, A.; Männiste, M.; Siibak, A. 2020. 'End of Theory' in the Area of Big Data: Methodological Practices and Challenges in the Social Media Studies. – *Acta Baltica Historiae et Philosophiae Scientiarum* 8, 1, 33–61.
- McBride, K.; Toots, M.; Kalvet, T.; Krimmer, R. 2018. Leader in e-Government, Laggard in Open Data: Exploring the Case of Estonia. – *Revue Française d'administration Publique* 167, 3, 613–625. <https://doi.org/10.3917/rfap.167.0613>.
- Milan, S.; Velden, L. van der 2016. The Alternative Epistemologies of Data Activism. – *Digital Culture and Society* 2, 2, 2364–2114. <https://doi.org/10.14361/dcs-2016-0205>.
- Männiste, M.; Masso, A. 2020. 'Three Drops of Blood for the Devil': Data Pioneers as Intermediaries of Algorithmic Governance Ideals. – *Mediální Studia | Media Studies* 14, 1, 55–74.
- Mühlhoff, R. 2019. Human-aided artificial intelligence: Or, how to run large computations in human brains? Toward a media sociology of machine learning. – *New Media and Society*, 1461444819885334. <https://doi.org/10.1177/1461444819885334>.
- Park, G.; Schwartz, H. A.; Eichstaedt, J. C.; Kern, M. L.; Kosinski, M.; Stillwell, D. J.; Ungar, L. H.; Seligman, M. E. P. 2015. Automatic Personality Assessment Through Social Media Language. – *Journal of Personality and Social Psychology* 108, 6, 934–952. <https://doi.org/10.1037/pspp0000020>.
- Pink, S.; Ruckenstein, M.; Willim, R.; Duque, M. 2018. Broken data: Conceptualising data in an emerging world. – *Big Data and Society* 5, 1. <https://doi.org/10.1177/2053951717753228>.
- Puschmann, C.; Burgess, J. 2014. Metaphors of big data. – *International Journal of Communication* 8, 1690–1709.
- Raley, R. 2013. Dataveillance and countervailance. – Lisa Gitelman (ed.), 'Raw Data' Is an Oxymoron. MIT Press.
- Rogers, R. 2013. *Digital Methods*. MIT.
- Rogers, R. 2019. *Doing Digital Methods*. Sage.
- Rossi, U. 2019. The common-seekers: Capturing and reclaiming value in the platform metropolis. – *Environment and Planning C: Politics and Space* 37, 8, 1418–1433. <https://doi.org/10.1177/2399654419830975>.
- Runnel, P.; Pruulmann-Vengerfeldt, P.; Reinsalu, K. 2009. The Estonian Tiger Leap from Post-Communism to the Information Society: From Policy to Practice. – *Journal of Baltic Studies* 40, 1, 29–51. <https://doi.org/10.1080/01629770902722245>.

- Schot, J.; Kanger, L. 2018. Deep transitions: Emergence, acceleration, stabilization and directionality. – *Research Policy* 47, 6, 1045–1059. <https://doi.org/10.1016/j.respol.2018.03.009>.
- Schäfer, M. 2016. Challenging Citizenship: Social Media and Big Data. – *Computer Supported Cooperative Work* 25, 2, 111–113. <https://doi.org/10.1007/s10606-016-9255-8>.
- Schäfer, M. T.; Es, K. van 2017. *The Datafied Society: Studying Culture through Data*. Amsterdam University Press.
- Shaw, R. 2015. Big Data and reality. – *Big Data and Society* 2, 2, 1–4. <https://doi.org/10.1177/2053951715608877>.
- Slota, S. C.; Hoffman, A. S.; Ribes, D.; Bowker, G. C. 2020. Prospecting (in) the data sciences. 0 *Big Data and Society* 7, 1, 1–12. <https://doi.org/10.1177/2053951720906849>.
- Stark, L.; Hoffmann, A. L. 2019. Data Is the New What? Popular Metaphors & Professional Ethics in Emerging Data Culture. – *Journal of Cultural Analytics*, May 1, 1–22. <https://doi.org/10.22148/16.036>.
- Tammpuu, P.; Masso, A. 2018. 'Welcome to the virtual state': Estonian e-residency and the digitalised state as a commodity. – *European Journal of Cultural Studies* 21, 5, 543–560. <https://doi.org/10.1177/1367549417751148>.
- Thylstrup, N. B.; Flyverbom, M.; Helles, R. 2019. Datafied knowledge production: Introduction to the special theme. – *Big Data and Society* 6, 2, 1–5. <https://doi.org/10.1177/2053951719875985>.
- Wagner-Pacifi, R.; Mohr, J. W.; Breiger, R. L. 2015. Ontologies, methodologies, and new uses of Big Data in the social and cultural sciences. – *Big Data and Society* 2, 2, 1–11. <https://doi.org/10.1177/2053951715613810>.
- Veltri, G. A. 2017. Big Data is not only about data: The two cultures of modelling. – *Big Data and Society* 4, 1, 1–6. <https://doi.org/10.1177/2053951717703997>.
- Viik, K. 2015. Sõnastik: interseksionaalsus. Feministeerium. <https://feministeerium.ee/nadala-sona-interseksionaalsus/>.
- Whitby, A. 2020. *The Sum of the People: How the Census Has Shaped Nations, from the Ancient World to the Modern Age*. Hachette UK.