

SOSIAALISEN MEDIAN MONITOROINTI POLICY-ANALYYSISSA

Kuinka digitaalista dataa voidaan käyttää policy-prosessien ymmärtämiseen?

**Ville Hakala
Maisterintutkielma
Yhteiskuntapolitiikka
Yhteiskuntatieteiden
ja filosofian laitos
Humanistis-
yhteiskuntatieteellinen
tiedekunta
Jyväskylän yliopisto
kevät 2021**

TIIVISTELMÄ

SOSIAALISEN MEDIAN MONITOROINTI POLICY-ANALYYSISSA

Kuinka digitaalista dataa voidaan käyttää policy-prosessien ymmärtämiseen?

Ville Hakala

Yhteiskuntapolitiikka

Pro gradu -tutkielma

Yhteiskuntatieteiden ja filosofian laitos

Humanistis-yhteiskuntatieteellinen tiedekunta

Jyväskylän yliopisto

Ohjaaja: Mikko Jakonen

Kevät 2021

Sivumäärä: 92

Uusien digitaalisten datalähteiden sekä analyysitekniikoiden on väitetty muuttavan yhteiskuntatieteitä radikaalisti. Tässä tutkielmassa tätä muutosta lähestytään tarkastelemalla siirtymää policy-analyysistä policy-analytiikkaan. Tarkemmin ottaen tutkitaan, kuinka kansalaisten mielipiteitä voidaan monitoroida sosiaalisesta mediasta, ja kuinka sosiaalisen median data voi auttaa ymmärtämään policy-prosesseja. Datafikaatiosta käyty keskustelu on usein konseptivetoista ja kaukana konkretiasta, eikä sosiaalisen median monitoroinnillekaan ole kehitetty käytännöllistä toteuttamistapaa.

Tutkimustehtävänäni on selvittää, kuinka sosiaalisen median monitorointia voidaan käytännössä toteuttaa. Tätä varten luon tutkielmassa datankäsittelyn eri vaiheille analyttisen typologian, joka mahdollistaa monitoroinnin prosessin analyttisen tarkastelun. Sosiaalisen median monitoroinnin käytännön toteutus vaatii työkaluja useilta erilaisilta policy-analyysin suuntauksilta ja perinteiset jakolinjat, esimerkiksi positivismin sekä tulkinnallisuuden suhteen, eivät välttämättä ole kovinkaan toimivia.

Tämän ”käytännöllisen sosiaalisen median monitoroinnin syklin” toimivuutta testaan noin 12 miljoonan twiitin muodostamalle aineistolle, joka on kerätty Christchurchin terrori-iskujen tutkimista varten keväällä 2019. Heti terrori-iskujen jälkeen Uudessa-Seelannissa uudistettiin aselakeja ja myös tämä policy-prosessi sisältyi aineistoon. Tarkastelen aselakeja koskevan keskustelun rajaamiseen sekä kansalaisten mielipiteiden tunnistamiseen liittyviä sosioteknisiä valintatilanteita edellä kehitellyn analyttisen typologian pohjalta mahdollisimman seikkaperäisesti.

Tutkielmassa havaittiin esimerkiksi pienien erojen aineistonrajauksen avainsanojen valinnassa vaikuttavan huomattavasti lopullisen aineiston koostumukseen. Case-tyyppisestä tutkimusasetelmasta johtuen tulosten yleistettävyyttä sinällään on hankala tulkita, mutta tuloksista voidaan päätellä, että datankäsittelyn vaiheisiin perustuva typologia on hedelmällinen lähestymistapa sosiaalisen median monitoroinnin käytännön toteuttamiseen.

Vastaavaa datalähtöistä lähestymistapaa voidaan suositella myös muissakin tutkimuksissa. Digitaalinen tutkimus mahdollistaa monien sellaisten tutkimusvaiheiden empiirisen tarkastelun, jotka aiemmin ovat tapahtuneet vain ”tutkijan pään sisässä”. Näiden vaiheiden dokumentointi on ensimmäinen askel vertailevaan empiiriseen tutkimukseen.

Avainsanat: policy-analyysi, laskennallinen yhteiskuntatiede, policy-analytiikka, Twitter, Christchurch

SISÄLTÖ

1. JOHDANTO.....	2
1.1 Data tulee.....	2
1.2 Aiempi tutkimus	4
1.3 Tutkielman rakenne sekä tutkimusasetelmat ja -kysymykset.....	7
2. DIGITAALINEN DATA YHTEISKUNTATIETEISSÄ.....	10
2.1 Digitaalinen on aina laskettavissa.....	10
2.2 Digitaalisen datan erityispiirteet yhteiskuntatieteiden kannalta.....	13
2.3 Aineiston ja analyysin välissä.....	16
2.4 Datan käsittelyn analyttinen viitekehys.....	19
3. POLICY-ANALYTIikka JA SOSIAALISEN MEDIAN MONITOROINTI.....	23
3.1 Analyysistä analytiikkaan.....	23
3.2 Poliittikaprosessien tunnistaminen.....	26
3.3 Kansalaisten demokraattinen toimijuus.....	29
3.4 Sosiaalisen median monitorointi käytännössä.....	33
4. TWITTER SOSIAALISEN MEDIAN MONITOROINNIN DATALÄHTEENÄ.....	36
4.1 Twitter datalähteenä	36
4.2 Twiitti digitaalisena havaintoyksikkönä.....	37
4.3. Twitter proxyna policy-prosessiin.....	41
4.4 Muistilista Twitterin monitoroinnille.....	43
4. CASE CHRISTCHURCH.....	46
4.1 Kuinka suurelta osin ilmiö tapahtuu verkossa ja mikä on digitaalisten menetelmien tuoma lisäarvo?.....	47
4.2 Kuinka Twitterin tavat antaa data vaikuttavat analyysiin?.....	52
4.3 Oatko huomioon käyttäjien tavat käyttää mediumia?.....	58
4.4 Löytääkö analyysimenetelmä ”oikean maailman” tapahtumat? Entä kansalaisten mielipiteet ja preferenssit?.....	66
6. JOHTOPÄÄTÖKSET JA POHDINTA.....	77
6.1 Johtopäätökset.....	77
6.2 Pohdinta ja päätäntö	80
7. LÄHTEET	83

1. Johdanto

1.1 Data tulee

Vuonna 2020 maailmassa arvioitiin olevan noin neljäkymmentäneljä zettatavua digitaalista dataa. Lukua on suhteellisen hankala laittaa perspektiiviin, mutta ”zetta” tarkoittaa miljardia miljardia ja näin ollen digitaalisia tavuja on noin neljäkymmentä kertaa enemmän kuin havaittavia tähtiä maailmankaikkeudessa. Datan määrän kasvu on eksponentiaalista ja vuonna 2025 ihmiskunnan on arvioitu tuottavan digitaalista dataa noin sata kertaa enemmän kuin puhuttua kieltä on aikojen saatossa tuotettu yhteensä. (Vuleta 2021.)

Osittain datan määrän kasvu johtuu yksinkertaisesti internetin käyttäjien määrän kasvusta. Nykyäänkin vasta kuusikymmentä prosenttia ihmiskunnasta käyttää internetiä ja määrän on arvioitu kasvavan yhdeksäänkymmeneen prosenttiin vuoteen 2030 mennessä. Esimerkiksi Afrikassa internet-käyttäjien määrä yli satakertaistui edellisen kahdenkymmenen vuoden aikana. Osittain datan määrän kasvu on myös laadullista, ja johtuu muuttuneista viestintäkäytännöistä: tätä nykyä päivittäin katsotaan lähes viisi miljardia YouTube-videota ja esimerkiksi Facebookilla on jo melkein kaksi miljardia käyttäjää, jotka tuottavat neljä petatavua dataa päivittäin. (Vuleta 2021.)

Yhteiskuntatieteilijöistä pisimpään digitaalisen datan parissa ovat työskennelleet internet-tutkijat, jotka ovat tutkineet digitaalista vuorovaikutusta yli kolmen vuosikymmenen verran ja käynnistivät keskustelun erityisesti internetin tutkimista varten tarkoitetuista menetelmistä vuosituhannen lopulla. Tämän ajan tutkimus lähestyi kuitenkin internetiä palveluineen erillisenä tilana ja paikkana, jonka vasta ’oikea maailma’ teki todelliseksi. (Rogers 2009, 22–23.) Digitaalista maailmaa kuvattiin virtuaalisena todellisuutena tai kyberavaruutena ja myöhemmin tätä erontekoa on kutsuttu *online/offline*-jakolinjaksi (Rogers 2019, 261).

Vuosituhannen vaihteessa internetin käyttö yleistyi ja samalla huomattiin, etteivät vanhat tutkimusasetelmat enää riittäneet. Tämä käynnisti muutoksen, jota on retrospektiivisesti kutsuttu laskennalliseksi käännteeksi. Laskennallisen käänteen aikana havaittiin, ettei

internetin tutkiminen omana todellisuutenaan ollut enää mielekästä, vaan sitä oli mahdollista lähestyä datanlähteenä, jonka avulla voidaan tutkia yleisiä sosiaalisia ja kulttuurillisia muutoksia. (Rogers 2019, 261.)

Laskennallisen käänteiden aikana monet tutkijat vertasivat käynnissä olevaa muutosta luonnontieteisiin. Esimerkiksi Venturinin ja Latourin (2010, 2) mukaan digitaalisten teknologioiden tuoma muutos yhteiskuntatieteille on verrannollinen kirjapainotaidon keksimiselle 1400-luvulla, joka osaltaan mahdollisti tieteellisen vallankumouksen. Watts (2007) taas argumentoi digitaalisen datan mahdollistavan yhteiskuntatieteilijöille ensimmäistä kertaa miljoonien ihmisten vuorovaikutuksen havainnoinnin tavalla, joka on herkkä myös yksilötason muutoksille. Näin yhteiskuntatieteissäkin voitaisiin luoda luonnontieteiden tapaan yksi koherentti viitekehys, joka on aiemmin ollut mahdotonta yksinkertaisesti siitä syystä, että sosiaaliset ongelmat ovat hankalimpia tieteellisiä ongelmia ratkaista.

Ottamatta kantaa siihen, ovatko digitaaliset teknologiat kirjapainotaitoon verrattavissa oleva mullistus, tulee niiden vaikutus varmasti olemaan suurempi yhteiskunta- kuin luonnontieteissä. Yhteiskuntatieteille kyse ei ole ainoastaan tutkimusmenetelmien muuttumisesta tai aineiston digitoinnista, vaan myös tutkimuksen *kohde* eli yhteiskunta digitalisoituu. Sosiaalisen median palveluista esimerkiksi Twitterin on sanottu muuttuneen olemassaolonsa aikana banaalista ”mitä söin tänään lounaaksi”-tyyppisestä sosiaalisesta mediasta alustaksi, jossa käsitellään maailman suuria uutistapahtumia (Rogers 2019, 170). Viimeistään *Cambridge Analytica* -skandaali osoitti, miten näennäisten arkipäiväistenkin asioiden jakaminen sosiaaliseen mediaan, voi jopa muokata vaalien lopputuloksia (Hern 2018).

Digitaalisen datan poliittinen merkitys onkin saanut sitä hallinnoivat yritykset yhä enenevässä määrin kiristämään ulkopuolisten pääsyä tietokantoihinsa. Facebookin sivuilta ei pysty enää lataamaan koneluettavassa muodossa viestejä kirjottaneiden käyttäjien nimiä ja ryhmistä saa ladattua tietoa vain näiden ylläpitäjän luvalla. Jokaisen Twitterin rajapintaa käyttävän sovelluksen on saatava palvelulta erikseen hyväksyntä. YouTuben käyttöehtojen on jopa tulkittu kieltävän palvelun käyttämisen tutkimustarkoituksiin. (Laaksonen ja Salonen 2018)

Samaan aikaan, kun digitaaliset aineistot sulkeutuvat tutkijoilta, julkinen sektori tukeutuu enenevässä määrin uusiin digitaalisiin analyysitapoihin ja rakentaa analyysinsä yksityisen sektorin alustojen päälle. Viime vuonna Suomen hallitus alkoi koronakriisistä johtuen seuraamaan kansalaisten liikkeitä yhteistyössä teleoperaattoreiden kanssa (Linnake 2020), Suomen Pankki käyttää työttömyyskorvausten googlaamista talousmallinuksissaan (Parviala 2020) ja ulkoministeriö monitoroi Twitter-keskustelua yksityisen 'tekoälyn' avustuksella – ja sama teknologia aiotaan ottaa käyttöön muissakin ministeriöissä. (Strömberg 2020).

Yhtä aikaa on siis käynnissä kolme kehityskulkua, joiden yhteisvaikutusta voi pitää suhteellisen huolestuttavana. Yhä suurempi osa yhteiskunnan informaatiovirroista on digitaalisessa muodossa, tutkijoilla on tähän informaatioon kaiken aikaa rajatumpi pääsy, ja kuitenkin tämä informaatiota vaikuttaa yhä enenevässä määrin demokraattisiin prosesseihin.

Ongelmallisimpia ovat ”tekoälyn” tapaiset konseptit, jotka sinällään kertovat hyvin vähän siitä käytännöllisestä työstä, jota tällaiset analyysit pitävät sisällään. Kehittyneidenkin koneoppimismenetelmien taustalla on paljon näkymätöntä työtä, kuten datan keräämistä, puhdistamista, kuratointia, trenisettien rakentamista sekä algoritmien valintoja, ja kehittämistä. Tämä työ voi vaikuttaa puhtaasti tekniseltä, mutta se sisältää lukemattomia valintatilanteita ja vaihtokauppoja, ja juuri näissä valinnoissa kulttuurilliset arvot liitetään systeemeihin. (Elish and boyd 2018, 70.) Tämän tutkielman tavoitteena on osaltaan valottaa sitä, miten tällainen näkymätön työ vaikuttaa analyysien tuloksiin.

1.2 Aiempi tutkimus

Elish ja Boyd ovat huomauttaneet, että aiemmin mainittuun *Cambridge Analytica* -skandaaliin liittyi myös paljon *hypeä* ja pelkoa. Keskustelua uusien teknologioiden mahdollisuuksista käytiin ja käydään idealistisella tasolla, josta puuttuvat niin teknologiset kuin analyttiset realiteetit. Tekniset asiantuntijat kyllä tietävät, että teknologinen todellisuus on hyvin kaukana idealisoiduista narratiiveista, mutta usein teknologioita implementoivat ei-tekniset asiantuntijat esimerkiksi koulutuksen, lääketieteen ja oikeustieteen aloille. (2018, 69-70.)

Eniten kritiikkiä Elishiltä ja Boydilta (2018) kirvoittavat ”hype-termit” *big data* ja *AI*, joihin heidän mukaansa usein suhtaudutaan kuin taikuuteen konsanaan. Myös Flyverbom ja Madsen ovat todenneet, että keskustelua emergoituvasta data-yhteiskunnasta dominoivat konseptit ja teoriat, joiden tavoitteena on identifioida datafikaation yleisiä piirteitä. Konsepteihin ja teorioihin nojaava lähestymistapa kuitenkin pitää abstraktiotason liian kaukana datankäsittelyn konkreettisesta työstä. (2015, 144.)

Digitaalisen datan tuomaa muutosta onkin lähestytty useilla eri ihmis- ja yhteiskuntatieteiden aloilla ja yritetty konseptualisoida useamman erilaisen termin alle. Tässä tutkielmassa pääroolissa on käsite *policy-analytiikka*, mutta politiikan tutkimuksen piirissä on käytetty rinnakkaisia termejä *'policy analytics'*, *'big data in public affairs'*, *'policy informatics'* ja laskennallinen yhteiskuntatiede (Longo ja McNutt 2018, 368). Laskennallinen yhteiskuntatiede on luultavasti käytetyin käsite yhteiskuntatieteissä yleensä, mutta sekin jakautuu useampiin erilaisiin lähestymistapoihin (Nelimarkka, 2021). Lisäksi humanistisissa tieteissä puhutaan yleensä *digitaalisista ihmistieteistä*, mutta käsitettä kuitenkin sovelletaan myös yhteiskuntatieteisiin (Helsinki Centre for Digital Humanities 2017).

Policy-analytiikan ideana on, että uudet digitaalisen datan lähteet sekä datan analysointi tekniikat mahdollistavat policy-analyysin perinteisen vaihemallin yhdistämisen yhdeksi jatkuvaksi syklikseksi prosessiksi. Prosessin tärkeimpinä piirteinä ovat reaaliaikaisuus, skaalautuvuus ja kokeellisuus, jotka myös mahdollistaisivat monen positivistinen policy-analyysin perinteisen kipukohdan parantamisen. (Longo ja McNutt 2018, 374.) Policy-analytiikkaan on useita erilaisia käytännön lähestymistapoja, joista tässä tutkimuksessa käsitellään kansalaisten mielipiteiden monitorointia sosiaalisesta mediasta (Grubmüller, Götsch, ja Krieger 2013). Kuten edellisessä alaluvussa todettiin, itseasiassa esimerkiksi ulkoministeriö tekee tällaista analytiikkaa jo, mutta sen käytännön toteutukseen liittyviä lukuisia valintoja ei ole dokumentoitu.

Yhdynkin tässä Flyverbomin ja Madsenin ajatuksiin, että konsepteihin ja teorioihin nojaavaa lähestymistapa uusiin datan lähteisiin pitää abstraktiotason liian kaukana konkreettisesta datan tuottamisesta, organisoimisesta, jakamisesta sekä visualisoinnista ja näin aiheuttaa riskin, että oleellisesti erilaiset datankäsittelyn käytännöt sekoittuvat samaksi paradigmaksi

(2015, 144). Tästä syystä tässä tutkielmassa painotus onkin paradigmojen sijaan datankäsittelyyn liittyvissä konkreettisissa vaiheissa, joita myös sosiaalisen median monitorointiin liittyy.

Sosiaalisen median monitorointiaakaan ei voi käsitellä yhdessä maisterintutkielmassa tarpeeksi yksityiskohtaisesti, joten tässä fokusoidaan datalähteenä toimivan Twitterin teknisiin ominaisuuksiin suhteellisen spesifisti. Käytännössä Twitter on myös ainoa sosiaalisen median palvelu, jonka aineistot ovat tällä hetkellä tutkijoiden saatavilla siinä laajuudessa, että sosiaalisen median monitorointi olisi mahdollista.

Koska tutkimuksessa käytetään sosiaalisen median dataa, käytännön lähestymistapa mukailee hyvin pitkälti *digitaalisia metodeja*, joka on Rogersin (2009, 2019) muotoilema metodologinen lähestymistapa internetin *kanssa* toteutettavaan tutkimustyöhön. Digitaalisten metodien kantava idea on *online groundedness*. Tällä hankalasti suomentuvalla käsitteelle tarkoitetaan, että digitaalisella datalla ja menetelmillä voidaan perustella yleisiä väitteitä kulttuurillisesta tai sosiaalisesta muutoksesta, joita ei tarvitse todentaa perinteisin menetelmin. Venturini ym. (2018, 4198) kuitenkin huomauttavat, että digitaalisten metodien soveltuminen tutkimukseen riippuu hyvin pitkälti siitä, missä määrin tutkittava ilmiö tapahtuu mediumilla, jota tutkitaan.

Tässä mielessä nimenomaan policy-analytiikka tarjoaa sosiaalisen median datalle mielenkiintoisen tulkinnallisen viitekehyksen. Policy-analyysin klassisen määritelmän mukaan analyysin tavoitteena on edesauttaa parempien päätösten syntymistä kuin ilman sitä syntyisi (Quade 1975). Tästä syystä policy-analytiikassa ontologiset kysymykset digitaalisen datan luonteesta eivät saa samanlaista painoarvoa kuin monessa muussa tutkimusasetelmassa. Näin ollen tämänkin tutkielman näkökulma on yhteiskuntapoliittisen käytännöllinen: policy-analytiikka tehdään jo, joten miten sitä tulisi parantaa?

Toisaalta sosiaalisen median monitoroinnin ideana on päästä käsiksi kansalaisten mielipiteisiin (Longo ja McNutt 2018, 373) ja tämä luo äärimmäisen tarkat raamit Twitterille *proxyna* kuten Venturini ym (2018, 4201) kutsuvat mediumia, jota käytetään välineenä jossain muualla tapahtuvan ilmiön tutkimiseen. On aivan eri asia tutkia Twitter-käyttäjien Twitter-keskustelua kuin käyttää Twitteriä *proxyna* kansalaisten mielipiteiden

monitoroimiseen, ja tämä myös herättää aivan uudenlaisia kysymyksiä. Keitä ovat kansalaiset Twitterissä? Onko esimerkiksi tykkäys mielipide?

1.3 Tutkielman rakenne sekä tutkimusasetelmat ja -kysymykset

Twitterin soveltuminen proxyksi policy-analytiikkaan ei ole itsestään selvyyttä. Vaikka tässä tutkimuksessa *online-groundedness* näytteleekin tärkeää osaa, ei sitä digitaalisten metodien tapaan käytetä epistemologisen kannanottona, vaan sen testaaminen on itseasiassa koko tutkimuksen aihe. Tästä johdetaan tutkimusongelma:

Kuinka sosiaalisen median monitorointia toteutetaan käytännössä, ja kuinka se voi auttaa ymmärtämään policy-prosesseja?

Policy-analytiikka ja sosiaalisen median monitorointi sen osana ovat määritelmällisesti syklisiä prosesseja, ilman alkua tai loppua. Tästä syystä luvussa kaksi luodaan Flyverbomia ja Madsenia (2015) mukaillen datankäsittelylle typologia, joka mahdollistaa sosiaalisen median monitoroinnin analyttisen tarkastelun. Luvun toissijaisena tavoitteena on löytää erilaisista lähestymistavoista datafikaatioon sellaisia piirteitä, jotka ovat yhteisiä kaikille digitaalisen datan kanssa työskenteleville yhteiskuntatieteilijöille, eri tieteenalojen teoreettisista oletuksista riippumatta. Tämän tyyppisille ”kauppa-alueille” on ilmaistu olevan tarvetta laskennallisen yhteiskuntatieteiden harjoittajien parissa (Nelimarkka 2021) ja datankäsittelyn analyttinen typologia voidaan nimenomaan mieltää tällaiseksi kauppa-alueeksi.

Luvussa kolme tarkastellaan, kuinka aikaisempi tutkimus on nähnyt uusien datalähteiden muuttavan policy-analyysin tekemistä sekä käsitellään siirtymää policy-analyysistä kohti policy-analytiikkaa. Edellisessä luvussa kehitettyä analyttistä typologiaa sovelletaan sosiaalisen median monitorointiin, joka toisaalta luo typologialle tulkinnallisen viitekehyksen, ja toisaalta typologia mahdollistaa sen tarkastelun, kuinka sosiaalisen median monitorointia voidaan käytännössä toteuttaa. Luvussa tuodaan vahvasti esille sitä, että digitaaliseen dataan perustuva tutkimus ei noudata perinteisiä jakolinjoja esimerkiksi positivismin sekä tulkinnallisuuden suhteen, ja sosiaalisen median monitorointi tarvitsee toteutuakseen työkaluja useilta erilaisilta policy-analyysin suuntauksilta.

Neljännessä luvussa tarkastellaan spesifisti Twitteriä ja sen roolia niin sosiaalisena mediana, datalähteenä kuin poliittisen keskustelun areenana. Twitterillä on ehdottomasti ylikorostunut rooli niin laskennallisen tutkimuksen kuin politiikan tutkimuksenkin osalta, mutta tälle on myös päteviä perusteluja. Twitterin suhteellisen avoin datapolitiikka mahdollistaa ylipäättään tällaisen tutkimuksen tekemisen, mutta myös datankeruun puutteiden huomioimisen - ainakin jossain määrin. Siksi tämänkin tutkimuksen aineisto on sieltä kerätty. Twitter-tutkimuksen on myös argumentoitu mahdollistavan tutkimusmenetelmien kehittämisen sosiaaliselle medialle yleisesti (Jungherr 2016, 81) ja tähän ajatukseen pohjaten luku neljä päättyy ”muistilistaan” sosiaalisen median monitoroinnille, joka samalla on ikään kuin tämän tutkielman tutkimusongelmaa tarkentavat tutkimuskysymykset:

- Missä määrin tutkittava ilmiö tapahtuu verkossa ja mikä on digitaalisen datan tuoma lisäarvo?
- Kuinka operationalisaatiossa huomioon mediumin tavat antaa data?
- Kuinka operationalisaatiossa huomioon käyttäjien tavat käyttää mediumia?
- Löytääkö analyysimenetelmä tunnetut tapahtumat? Entä kansalaisten mielipiteet?

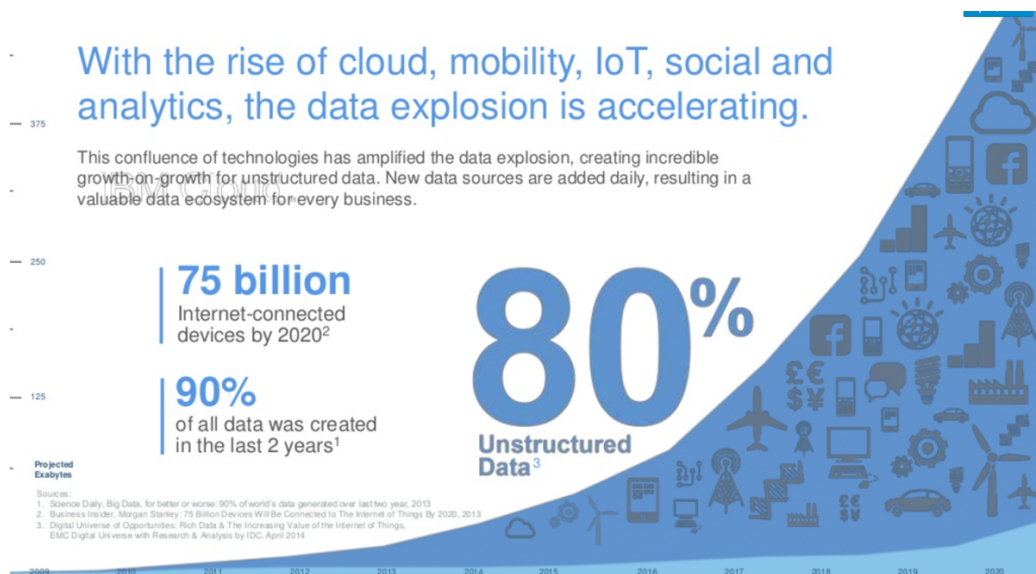
Näihin kysymyksiin vastaaminen muodostaa pohjan luvulle viisi, jossa *case*-esimerkkinä toimii Uuden-Seelannin aselakiuudistus, joka toteutettiin reaktiona Christchurchin terrori-iskuihin. Terrori-iskuihin liittyen HYTE-tutkimuskonsortiossa kerättiin vajaan kahdentoista miljoonan twiitin aineisto keväällä 2019, josta tätä tutkimusta varten täytyy ensin filttäroidä aselakeja koskeva keskustelu erikseen ja sitten tunnistaa tästä keskustelusta kansalaisten mielipiteet. Case-esimerkin tavoitteena on yhtä aikaa testata luvuissa kaksi ja kolme luotua ”sosiaalisen median monitoroinnin käytännöllistä viitekehystä” ja toisaalta tuoda esiin ne lukuisat näennäisen merkityksettömät valintatilanteet, jotka usein vastaavassa tutkimuksessa sivuutetaan, mutta joilla on mahdollisesti merkittävä vaikutus analyysin lopputulokseen. Tämän tutkimuksen tapauksessa aineiston suuri määrä mahdollistaa poikkeuksellisen hyvin erilaiset vertailevat asetelmat ja näin ollen edellä esitellyn muistilistan kysymyksiin vastaamisen.

Pohdinta ja päätäntö osiossa käydään läpi tutkimuksen kulkua onnistumisineen ja puutteineen. Lisäksi osiossa pohditaan yleisesti sekä digitaaliseen dataan että policy-analytiikkaan liittyviä avoimia kysymyksiä sekä mahdollisia jatkotutkimuksen aiheita. Pohdinnan lopuksi vielä mietitään, voisivatko myös tieteen tekemisen käytännöt olla diginatiiveja.

2. DIGITAALINEN DATA YHTEISKUNTATIETEISSÄ

2.1 Digitaalinen on aina laskettavissa

Cd-levyn keksiminen 1980-luvulla tarjosi ensimmäistä kertaa halvan tavan tallentaa digitaalista tietoa ja samalla teki sen jakamisen helpoksi. Tästä käynnistyi kehitys, jossa digitaalinen tallennustila alkoi hiljalleen syrjäyttää analogisia vastineitaan kuten vinyylejä, kasetteja ja filminauhoja. Seuraavana vuosikymmenenä korkearesoluutioiset skannerit tulivat laajasti saataville, minkä voidaan yleisesti ottaen katsoa aloittaneen laajamittaisen analogisten aineistojen digitoinnin. (Savić 2019, 38.) Kuten johdannossa todettiin, nykyisellään digitaalisesti tallennettuja tavuja on jo enemmän kuin havaittavia tähtiä universumissa, ja määrä kasvaa eksponentiaalisesti.






KUVIO 1. Digitaalisen datan määrän kasvu (Michael Beatty 2017)

Kasvaneesta digitaalisen datan määrästä huolimatta sama prosessi, jossa ihmisten ja objektien analogista toimintaa muunnetaan kvantitatiiviseksi ja binääriseksi datavirraksi on edelleenkin ensimmäinen vaihe dataa käsitellessä (Flyverbom ja Madsen 2015, 145–46). Yhteiskuntatieteilijälle tämä saattaa jäädä usein näkymättömiin. Siinä missä esimerkiksi monen luonnontieteilijän pitää edelleen nähdä aineiston digitoinnin vaiva, ovat yhteiskuntatieteelliset aineistot kuten puolueohjelmat tai sanomalehtiartikkelit usein

valmiiksi digitaalisessa muodossa (Nelimarkka 2021). Näin ollen yhteiskuntatieteellisessä tutkimuksessa ei välttämättä ole oleellista edes ottaa kantaa aineiston digitoinnin prosessiin

Samanlaiset digitaaliset objektit voivat kuitenkin kätkeä taakseen hyvin erilaisia analogisia prosesseja. Sata vuotta vanhan valokuvan skannaamisessa ja älypuhelimien näytön painamisessa kuvan ottamisen hetkellä on molemmissa kyse analogisen informaation muuttamisesta digitaaliseen muotoon ja lopputuloskin voi olla bitilleen sama. Toisaalta myös digitaalisten informaatiovirtojen merkityksen kasvu voi laskea analogisen informaation relevanssia (Loukasmäki ja Makkonen 2019, 133)

Ehkä edellisen kaltaisissa tapauksissa digitoinnista puhumisen sijasta pitäisikin puhua kokonaisten prosessien digitalisoinnista. Digitointia ja digitalisointia käytetään usein synonyymeinä, vaikka digitalisoinnilla tarkoitetaan spesifisti kokonaisten (liiketoiminta)prosessien muuttamista digitaaliseen muotoon (Savić 2019, 38). Digi-kieli ja -käsitteet elävät kuitenkin jatkuvasti ja käsitteitä määritetään usein pitkälti liiketoiminnan perspektiivistä. Tässä tutkielmassa termin digitointi ymmärretään tarkoittavan analogisen muuttamista digitaalseksi, oli kyseessä sitten ilmiökokonaisuus tai yksittäinen objekti. Digitalisaatiolla taas viitataan prosessiin, jossa jatkuvasti suurempi osa yhteiskunnan, kulttuurin ja talouden tuottamasta informaatiovirrasta on digitaalisessa muodossa.

	DIGITIZATION	DIGITILIZATION	DIGITAL TRANSFORMATION
Focus	Data conversion	Information processing	Knowledge leveraging
Goal	Change analog to digital format	Automate existing business operations and processes	Change company's culture, the way it works and thinks
Activity	Convert paper documents, photos, microfilms, LPs, films, and VHS tapes to digital format	Creation of completely digital work processes	Creation of a new digital company or transformation to a digital one
Tools	Computers and conversion/encoding equipment	IT systems and computer applications	Matrix of new (currently disruptive) digital technologies
Challenge	Volume <i>Material</i>	Price <i>Financial</i>	Resistance to change <i>Human resource</i>
Example	Scanning paper-based registration forms	Completely electronic registration process	Everything electronic, from registration to content delivery
			

KUVIO 2. Digitoimisen, digitalisoimisen ja digitaalisen muutoksen väliset eroavaisuudet Savićin mukaan (2019, 37)

Rogers on lähestynyt digitoimisen ja digitalisoimisen välistä rajan vetoa hieman toisesta kulmasta. Hänen mukaansa voidaan nähdä ontologinen ero niiden objektien, sisältöjen, laitteiden ja ympäristöjen välillä, jotka ovat syntyneet digitaalisesti, ja jotka ovat digitoituja. (2009, 21–22). Tämän tutkimuksen käytännöllisestä lähestymistavasta johtuen ontologisten kysymysten pohdintaan digitaalisen datan luonteesta ei voida käyttää sitä tilaa, mitä se ansaitsisi. Kysymys siitä, onko digitaalisessa datassa jotain oleellisesti uutta, säteilee silti läpi tämänkin tutkimuksen.

Alun perin digitoimisen hyödyn ajateltiin olevan digitaalisten objektien pitkäkestoinen säilyvyys verrattuna analogisiin vastineisiin. Digitointi ei ehkä täyttänyt tätä lupautta, mutta se toi muunlaisia hyötyjä liittyen käytettävyyteen, saatavuuteen, siirrettävyyteen sekä prosessointimahdollisuuksiin. (Savić 2019, 38) Prosessointimahdollisuuksilla tarkoitetaan sitä, että kaikki digitaalisessa muodossa oleva tieto on pohjimmiltaan binääristä eli ykkösiä ja nollija, joita tietokoneet pystyvät laskemaan. Tietokonepohjainen ”laskeminen” on kuitenkin huomattavasti laajempi ilmiö kuin laskeminen siinä mielessä kuin kvantitatiivinen yhteiskuntatiede sen käsittää ja mielestäni suomen kielen sana ”tietojenkäsittely” kuvaisikin tietokoneiden roolia yhteiskuntatieteissä paremmin. Kuitenkin vuonna 2009 Lazer ym. esittelivät termin *computational social science*, joka yleisesti suomennetaan laskennalliseksi yhteiskuntatieteeksi.

Laskennallisen yhteiskuntatieteen esiinmarssiin on vaikuttanut datan määrän kasvun lisäksi nimenomaan tietokoneiden kasvanut laskentakapasiteetti. Nykyään älypuhelimissakin on laskentakapasiteettia 1990-luvun supertietokoneen verran (Bookman 2017). Näin tutkijat voivat analysoida dataa suuremmissa mittakaavassa ja nopeammin kuin koskaan aiemmin ja tehdä sen omilla laitteillaan, sen sijaan, että resursseja pitäisi varata suuresta yhteisestä supertietokoneesta (Nelimarkka 2021). Esimerkiksi eräässä tutkimuksessa pyydettiin tutkittavia vastaamaan älypuhelimilla otetuilla kuvilla, jotka tutkijat sitten koodasivat Googlen Vision -rajapinnalla, joka osaa kategorisoida kuvia niiden sisällön perusteella. Perinteiseen kvalitatiiviseen koodaukseen verrattuna tämä oli 520 kertaa nopeampaa ja 150 kertaa halvempaa (Bosch, Revilla, ja Paura 2018).

Keskustelua siitä, tekevätkö suuremmat määrät dataa ja nopeampi laskenta menetelmistä sinällään uusia, on käyty luultavasti yhtä pitkään kuin näitä menetelmiä on käytetty.

Esimerkiksi Watts (2007, 489) argumentoi, että monet menetelmät, joita mainostetaan uusina, eivät sitä ole, ja tietokoneavusteisia tutkimusmenetelmiä onkin käytetty ainakin 1980-luvulta alkaen (Saari 2011b, 76). Myös Nelimarkan mukaan laskennallisen yhteiskuntatieteen uusia menetelmiä painottavalla lähestymistavalla löytyy historiallista yhteyttä tietokoneavusteisiin kvantitatiivisiin menetelmiin ja usein sen edustajilta löytyy taustaa tilastotieteestä (2021). Tässä tutkimuksessa painotus on nimenomaan datassa, joten menetelmien uutuusarvosta käytävä keskustelu ei ole niin oleellinen.

Menetelmien ja aineiston suhteen kyse on kuitenkin pitkälti painotuksista, eikä näitä voida täysin erottaa toisistaan. Kun digitalisaatio etenee ja digitaalisen datan määrä kasvaa, tarvitaan väistämättä uusia laskennallisia menetelmiä sen käsittelyyn. *Big data* määritelläänkin tässä tutkimuksessa, niin suureksi määräksi dataa, että tarvitaan tietokoneen (*computer*) apua sen käsittelyssä (*compute*) ja yhteiskuntatieteellisten kysymysten kohdalla nimenomaan laskennallista yhteiskuntatiedettä (*computational social science*). *Big datan* sijasta tässä tutkielmassa puhun usein digitaalisesta datasta, mutta sillä viitataan nimenomaan niin suureen määrään dataa, ettei sitä ole mielekäästä käsitellä kuin laskennallisesti.

2.2 Digitaalisen datan erityispiirteet yhteiskuntatieteiden kannalta

Edellä puhuttiin datasta yhtenä kokonaisuutena ja tämä on myös yleisempi tapa lähestyä ilmiötä, kun digitalisaatiosta tai yhteiskunnan datafikaatiosta on kyse. On kuitenkin argumentoitu, että *big datasta* puhuminen yhtenä ilmiönä häivyttää nyansseja sekä eroavaisuuksia, joita erilaiset datalähteet ja analyysitavat pitävät sisällään. (Flyverbom and Madsen 2015, 4)

Esimerkiksi erilaisten sensorien tuottama data on erittäin merkittävää myös laskennallisen yhteiskuntatieteen kannalta (Nelimarkka 2021) ja sensoreiden dataan perustuvia tutkimusasetelmia on käytetty jo viidentoista vuoden ajan (Eagle ja Pentland 2006). Sensoreista esimerkkinä voisi mainita lukuisat älylaitteet, jotka nykyään mittaavat ihmisten unen määrää ja laatua. Digitaalisten sensoreiden määrä kasvaa myös jatkuvasti ja vuonna 2025 on arvioitu 26 miljardin yksittäisen esineen olevan yhteydessä internetiin. Niin kutsuttu

esineiden internet mahdollistaa myös ihmisten kommunikoinnin esineiden kanssa ja näin esineilläkin voidaan argumentoida olevan jonkinlaista toimijuutta. (Hoffman ja Novak 2017.) Tällaiseen tutkimukseen liittyy kuitenkin aivan omanlaisiaan kysymyksenasetteluja ja tästä syystä joudun rajaamaan ne tämän tutkimuksen ulkopuolelle.

Laskennallisen yhteiskuntatieteen dataa painottava lähestymistapa onkin alusta asti hyvin pitkälti painottanut nimenomaan ihmisten välisestä vuorovaikutuksesta syntyvää dataa. Digitaalisen datan on yleensä nähty mahdollistavan vuorovaikutuksen tarkastelun tavalla, joka pakenee perinteisiä jaotteluja kvalitatiiviseen ja kvantitatiiviseen tai mikeroon ja makroon. Lisäksi digitaaliset tutkimusympäristöt mahdollistavat tällaisen tarkastelun reaaliajassa. (mm. Watts 2007, Lazer ym. 2009, Venturini ja Latour 2010.) Tätä tarkastelutapaa on kutsuttu useammalla erilaisella käsitteellä, mutta kutsun sitä tässä yksinkertaisesti *zoomaukseksi*. Teknisesti *zoomauksessa* on hyvin pitkälti kyse tarkasteltavan datan muodosta.

Kuviossa 1 kuvattiin datan määrän kasvua, mutta jaoteltiin data myös strukturoituun ja ei-strukturoituun. Strukturoiduksi kutsutaan dataa, joka on valmiiksi muodossa, jossa sitä voidaan käsitellä relaatiotietokannoissa, kuten kauppojen myyntejä. Suurin osa nykyisin tuotetusta datasta on kuitenkin ei-strukturoitua kuten esimerkiksi kirjoitettu teksti. Näiden välissä on kuitenkin datatyyppi, joka pakenee perinteisiä jaotteluja eli semistrukturoitu data, josta tyyppiesimerkkejä ovat älypuhelimilla otetut kuvat. Kuvat ovat itsessään strukturoimatonta tietoa, mutta jokainen älypuhelimella otettu kuva sisältää lisäksi esimerkiksi aikaleiman ja paikkatietoja strukturoidussa muodossa. Tällaisia semistrukturoituja tiedostotyyppejä ovat esimerkiksi JSON, CSV sekä XML. (Smallcombe 2020.)

Yhteiskuntatieteiden perspektiivistä mielenkiintoista on, että sosiaalinen vuorovaikutus digitaalisissa ympäristöissä tuottaa lähes aina tällaista strukturoitua metadataa. Jos esimerkiksi verrataan sähköpostia perinteiseen kirjeeseen, voidaan niiden viestinnällisen funktion argumentoida olevan likipitään sama. Kuitenkin, kun sähköposti lähetetään, liitetään siihen automaattisesti aikaleima, joka myös välittyy vastaanottajalle. Tämä on nykyään jo niin itsestään selvää, ettei sitä usein tule ajatelleeksi, mutta perinteisessä kirjeessä ajan lisääminen on vapaaehtoista, aikaleima ei ole niin tarkka kuin digitaalinen vastineensa ja kirjeen lähetyksen sekä vastaanottamisen ajat eroavat toisistaan huomattavasti.

Useimmat sähköpostiohjelmat myös organisoivat viestejä näiden aikaleimojen perusteella ja näyttävät uusimmat viestit ensimmäisenä. Metadata yhdessä digitaalisten teknologioiden kanssa mahdollistaakin myös uudenlaisia tapoja tiedon tarkasteluun. Sähköpostiohjelmaa edustavampana esimerkkinä voitaisiin käyttää jotain lukuisista digitaalisista karttapalveluista. Voidaan argumentoida, etteivät nämä digitaaliset kartat ole yhtään analogisia vastineitaan tarkempia ja niissä esiintyy aivan samanlaisia tulkinnallisuuksia, esimerkiksi sen suhteen, mihin valtioon joku konfliktialue katsotaan kuuluvaksi. Toisaalta voitaneen pitää kiistattomana, että digitaalisten karttapalvelujen mahdollisuudet katsella maailmaa kokonaisuudessaan ja tarpeen tullen zoomata yksityiskohtaisesti tiettyyn pisteeseen, auttavat saamaan paremman kokonaiskuvan tarkastelun kohteena olevasta ilmiöstä kuin analogiset vastineensa. Ja koska älypuhelimella otettujen kuvien metadata sisältää sijaintitiedot, voidaan esimerkiksi nämä kuvatkin sijoittaa kartalle vastaavalla tavalla. Digitaalista tutkimusta onkin usein verrattu kartoittamiseen (esim. Marres 2015, Marres ja Moats 2015, Burgess ja Matamoros-Fernández 2016, Alinejad ym. 2019).

Zoomaamisen ohella jälki (*trace*) on hyvin käytetty konsepti digitaalista dataa käsittelevässä yhteiskuntatieteellisessä tutkimuksessa (esim. Venturini 2010, Alinejad ym. 2019, Flyverbom ja Madsen 2015, Venturini ja Latour 2010). Digitaalisen jäljen on määritely olevan mikä tahansa digitaalisen mediumin kollektiivista toimintaa välittävä viesti, kuten blogissa julkaistu postaus, verkkokaupan loki kaupankäynnistä tai hyperlinkki, joka yhdistää kaksi nettisivua toisiinsa. Käsitteellä ”jälki” viitataan nimenomaan siihen, että yleensä näitä jälkiä ei ole erityisesti luotu akateemista tutkimusta varten ja ne ovat monelta osin kolmansien osapuolien hallinnassa. (Venturini ym. 2018, 4200.) Jäljessä on siis kyse hyvin pitkälti myöskin datan strukturoinnista, mutta sen määritelmä painottaa myös datan tuottamista sekä kolmansien osapuolten vaikutusta sen käytössä.

Jäljistä on myös todettu, etteivät ne ole perinteisessä yhteiskuntatieteellisessä mielessä tilastollisesti edustavia otoksia yhteiskunnasta. Tämä on kuitenkin ongelma vain siinä tapauksessa, jossa digitaalista dataa halutaan kohdella kuin kyselytutkimuksia ja käyttää tilastollisessa aggregaatiossa. Digitaalisten menetelmien uutuusarvoa painottavat tutkijat korostavat, etteivät nämä menetelmät noudata perinteistä jakoa kvantitatiivisiin ja kvalitatiivisiin, vaan ovat ’kvali-kvantitatiivisia’. Heille yhteiskuntatieteellisten menetelmien suurimpana puutteena näyttäytyy, että makro-tason ilmiöiden tarkastelu

tehdään tilastollisin menetelmin, joilla kyetään selittämään kattavasti jo olemassa olevia ja normalisoituneita ilmiöitä, mutta jotka soveltuvat heikosti vasta muodostumassa olevien ilmiöiden tarkasteluun. (Venturini ja Latour 2010).

Tässä mielessä suurien digitaalisten aineistojen kanssa työskentelevät tutkijat joutuvat aivan uudentyypisten kysymysten eteen. Aineistot eivät ole tilastollisesti edustavia perinteisessä kvantitatiivisessa mielessä, mutta toisaalta datan määrä ja strukturointi erottavat sen radikaalisti laadullisista aineistoista. Tästä johtuen digitaalista dataa käyttävän yhteiskuntatieteellisen tutkimuksen parissa tunnutaan käyvän jatkuvaa keskustelua kvalitatiivisesta ja kvantitatiivisesta, objektiivisuudesta ja subjektiivisuudesta, positivismista ja relativismista sekä erilaisista konsepteista, joilla edellä mainitut vastinparit koitetaan kiertää. Kuten digitaalisen datan ontologian kohdalla, myös tämän keskustelun kaikuja kuuluu kaikkialla tässäkin tutkimuksessa.

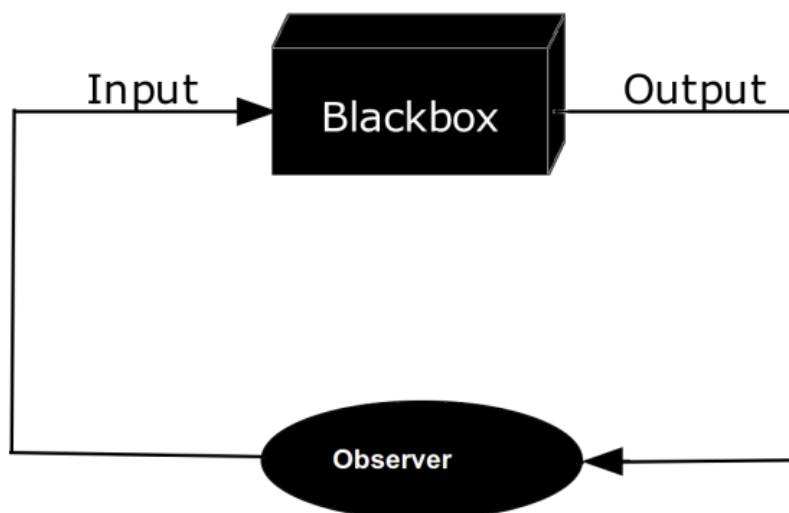
2.3 Aineiston ja analyysin välissä

Tähän mennessä tässä tutkielmassa on pitäyditty eron tekemisessä digitaalisen datan ja uusien menetelmien välille, mutta kuten edellä huomattiin, ei esimerkiksi karttapalvelujen tapa järjestää ja visualisoida dataa oikeastaan oikein sovi kumpaankaan lokeroon. Tämä kummallinen välimaasto datan ja menetelmien välissä myös osaltaan aiheuttaa ongelmia sen suhteen, kuinka totuudellisina digitaalisesta datasta tuotettuihin tuloksiin voi ylipäätään suhtautua

Esimerkiksi Elish ja Boyd ovat argumentoineet, että kaksi datafikaation *hype*-käsitettä ”big data” ja ”AI” ovat oikeastaan osa samaa ilmiötä (2018, 59). Ajatuksessa ei ole sinällään mitään mullistavaa, ja myös datan kanssa työskentelevien yritysten on tapana kuvata datan käsittelyään samaan tapaan (Sense Corp 2018). Toisaalta on pidettävä mielessä, että digitaalisen datan avulla voidaan tehdä erilaista analytiikkaa ja visualisaatioita ilman, että tarvitaan mitään tekoälyyn viittaavaa tai edes varsinaista datatiedettä. Tästä syystä tässäkin tutkimuksessa data on kaiken lähtökohta.

Elishin ja Boydin mukaan uusiin datalähteisiin ja niiden käsittelyyn käytettyihin menetelmiin suhtaudutaan kuin taikuuteen ja tästä syystä datan käsittelyyn liittyvät arkiset käytännöt jäävät vaille huomiota. Kehittyneidenkin koneoppimismenetelmien taustalla on paljon näkymätöntä työtä, kuten datan keräämistä, puhdistamista, kuratointia, treenisettien rakentamista sekä algoritmien valintoja ja kehittämistä. Tämä työ voi vaikuttaa puhtaasti tekniseltä, mutta se sisältää lukemattomia valintatilanteita, ja juuri näissä valinnoissa kulttuurilliset arvot liitetään digitaalisiin systeemeihin. (2018, 70.)

Elishin ja Boydin (2018) mukaan käytännön työssä digitaalisen datan parissa on paljon samaa kuin etnografiassa ja tästä syystä heidän mielestään pitäisikin puhua nimenomaan laskennallisesta etnografiasta. Myös etnografit ympäröivät itsensä datalla, valitsevat mitä nähdä ja olla näkemättä sekä kehittävät koherentin mentaalimallin kiteyttämään tuloksia. Tästä näkökulmasta datankäsittely siis näyttäytyy hieman kuin mustana laatikkona, jossa “*big dataa*” syötetään tutkijan mentaalimallille, joka sitten tuottaa tulosteita. Ajatukselle on sikäli perusteita, että digitaalisen datan kanssa työskentelevien yhteiskuntatieteilijöidenkin vaikuttaa yleisesti olevan tapana jättää raportoimatta datan keräämiseen, käsittelyyn ja siivoamiseen liittyviä valintoja (Jungherr 2016), asioita tehdään “näppituntuman varassa” (Makkonen ja Loukasmäki 2019, 141) tai aineistoja vain “tuotetaan koneellisesti” (Vainikka ym. 2020, 252)



KUVIO 3. Mustan laatikon toiminta (Krauss 2014)

Tämä ei kuitenkaan tarkoita, että näin olisi oltava, etenkin tieteen tekijöiden parissa. Sama jäljitettävyyden, mikä aiemmin todettiin pätevän digitaalisiin interaktioihin, pätee myös tutkijoihin. Datatiedettä harjoittava tutkija joutuu väistämättä koodaamaan oman mentaalimallinsa tietokoneelle, eikä voi jättää sitä päänsä sisään. Tämä ei tietenkään tarkoita, etteikö malli tällöin sisältäisi tulkinnallisia elementtejä, mutta reflektio näiden tulkinnallisten elementtien osalta voidaan hoitaa yhteisöllisesti, koska malli ei ole pään sisällä piilossa.

Täsmälleen samaa keskustelua käytiin 50 vuotta sitten *Limits to Growth* -raportin suhteen, jota kirjoittajat perustelivat tavalla, joka on oleellinen vielä tänäkin päivänä. Mallilla tarkoitetaan ainoastaan järjestettyä joukkoa oletuksia kompleksisesta systeemistä ja ihmiset yleisesti lähestyvät ongelmia mallintamalla. Päätöksentekijät kaikilla tasoilla käyttävät tiedostamattomia mentaalimalleja tehdessään poliittisia päätöksiä ja ihmisaivojen rajallisesta kapasiteetista johtuen mentaalimallit ovat aina yksinkertaistuksia oikeasta maailmasta. Nimenomaan tästä syystä mallintamiseen kannattaakin käyttää tieteellistä metodologiaa ja nykyaikaisia tietokoneita. (Meadows ja Club of Rome 1972, 20-21.) Myös *Limits to Growth* -raportissa esitellyn mallin kritisoitiin sisältävän satunnaisia oletuksia, joita se sitten käsittelee ja tuottaa satunnaisia tuloksia. Mallin siis sanottiin toteuttavan tietojenkäsittelytieteen vanhinta maksimia: roskaa sisään, roskaa ulos. (Passell ja Ross 1972). Koska raportti ennusti tapahtumia pitkälle tulevaisuuteen, käydään siitä debattia edelleen. Viime vuosina sävy on ollut positiivisempi ja mallin ennusteiden on todettu olevan monella tapaa oikeassa (Turner ja Alexander 2014).

Toisaalta ”laskennallinen etnografia” on vaan konseptin korvaamista toisella konseptillä. Flyverbom ja Madsen ovat argumentoineet, että nimenomaan konsepteihin ja teorioihin nojaavaa lähestymistapa uusiin datan lähteisiin pitää abstraktiotason liian kaukana konkreettisesta datan tuottamisesta, organisoimisesta, jakamisesta sekä visualisoinnista ja näin aiheuttaa riskin, että oleellisesti erilaiset datankäsittelyn käytännöt sekoittuvat samaksi paradigmaksi. Yleisesti keskustelua emergoivasta datayhteiskunnasta dominoivat konseptit ja teoriat, joiden tavoitteena on identifioida datafikaation yleisiä piirteitä. Erilaiset dataprojektit eroavat toisistaan kuitenkin niin paljon, ettei niitä ole mielekäästä käsitellä osana yhtä paradigmaa. (Flyverbom ja Madsen 2015, 144–146)

Uuden konseptin sijasta Flyverbom ja Madsen ehdottavat analyyttistä viitekehystä, jossa datankäsittelyyn liittyviä projekteja tarkastellaan erilaisten teknisten vaiheiden kautta. He hahmottavat tätä datan arvoketjuna, jossa data, kuten muutkin resurssit, muutetaan ja järjestellään muotoon, josta siitä saa mahdollisimman paljon arvoa. Tässä tapauksessa arvolla tarkoitetaan toiminnallista tietoa (Flyverbom and Madsen 2015, 145)

Nimenomaan teknisiin valintoihin keskittyminen myös voisi mahdollistaa eri tieteenalojen keskustelemisen keskenään. Laskennallisen yhteiskuntatieteen kehityksen kannalta on ollut ongelmallista, että eri tieteenaloilta tulevat tutkijat puhuvat erilaisilla konsepteilla ja Nelimarkka (2021) onkin ehdottanut tähän ratkaisuksi *kauppa-alueita*, joissa tutkijat vaihtavat keskenään pienempiä informaation osia laajojen ja monimutkaisten teoreettisten pakettien sijasta. Tämä mahdollistaisi informaation jakamisen myös tilanteessa, jossa tutkijat ovat eri mieltä tutkimuksen konsepteista tai taustaoletuksista.

Vaikka tässä tutkimuksessa painotetaan nimenomaan teknologista reflektiota, on kuitenkin todettava, ettei tähän aina ole mahdollisuuksia ja tätä ei missään nimessä pitäisi käyttää selityksenä sille, ettei reflektiota ole laisinkaan. Jos teknologiseen reflektioon ei ole mahdollisuutta, on metodologisen reflektion mukailtava hyvin pitkälle Ellishin ja Boydin ajatuksia, toki teknistä dokumentaatiota unohtamatta. Tätä problematiikka käsitellään tarkemmin Twitterin kohdalla luvussa neljä.

2.4 Datan käsittelyn analyyttinen viitekehys

Flyverbomin ja Madsenin analyyttinen viitekehys pohjaa ajatukseen järjestelystä (*sorting*) sateenvarjokonseptina erilaisilla analyyttisille prosesseille, kuten klassifiointille, kategorisoinnille, kvantifiointille, laskemiselle, arvottamiselle ja suhteellistamiselle. Käsitteellä yritetään tavoittaa sitä, kuinka pinnallisesti tekniset ja arkipäiväiset datanmurskausprosessit tarjoavat uudenlaisia oivalluksia datafikaation organisatorisiin ja sosiaalisiin implikaatioihin. Lähestymistapa tarjoaa täsmällisemmän ymmärryksen järjestelyn roolista datan käsittelyssä, problematisoi laajalle levinneitä oletuksia datasta suoran tarkastelun kohteena ja tietona, joka näennäisesti ”puhuu itse”, sekä voi lopulta tehdä tilaa entistä vivahteikkaammille konsepteille datafikaation ja yhteiskunnallisten muutosten välisestä suhteesta. (Flyverbom ja Madsen 2015, 144–45)

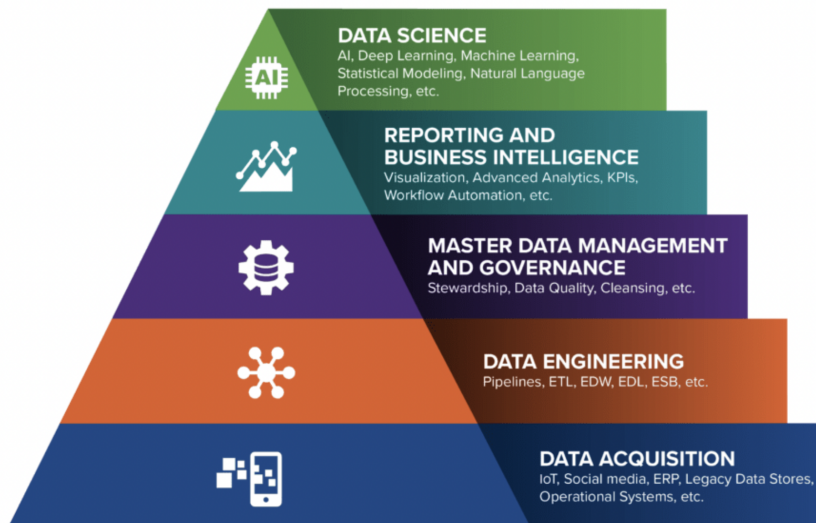
Näistä järjestelyprosesseista ja -vaiheista muodostuu konseptualisaatio *big datan* analyysista, jota voidaan ajatella myös ”datan arvoketjuna”, jossa data, kuten muutkin materiaalit, hankitaan, jalostetaan ja muunnetaan arvokkaiksi tuotteiksi. Nimenomaan fokusointi tiedon tuottamiseen datasta ja ymmärrys tiedosta tuotettavana, on olennaista tälle *big datan* konseptualisaatiolle. (Flyverbom ja Madsen 2015, 144–45.) On huomattava, että ensimmäisessä alaluvussa mainittu sana ”tietojenkäsittely” tietyllä tapaa sisältää nämä prosessit ja voisi olla jälleen suomeksi hyvä termi.

Viitekehyksessä reaaliaikaisen datan muuttaminen toiminnalliseksi tiedoksi jaetaan neljään erilaiseen analyttiseen vaiheeseen. Ensimmäinen vaihe koskee datan tuottamista. Tässä vaiheessa ihmisten ja objektien toiminta muunnetaan kvantitatiiviseksi sekä binääriseksi datavirraksi, jota voidaan tallentaa ja prosessoida tietokoneella. Toinen vaihe koskee datan järjestämistä ja tämä vaihe sisältää valintoja tietokannoista, luokittelusysteemeistä ja metadatasta, joiden avulla dataa voidaan järjestää ja lukea systemaattisesti. Kolmas liittyy datan jakamiseen ja siinä keskitytään erityisesti tapaan, jolla tietokantoihin päästään käsiksi ja digitaalisten jälkien distribuutiosta neuvotellaan datan omistajien ja loppukäyttäjien kesken. Neljännessä vaiheessa pohditaan datan visualisaatiota. Se sisältää valintoja siitä, kuinka saatavilla oleva data visualisoidaan kiteyttämään se aspekti maailmasta, josta ollaan kiinnostuneita, sekä harkintaa algoritmien käytöstä projektissa. (Flyverbom ja Madsen 2015, 145–146.)

Neljään vaiheeseen perustuva typologia integroi konseptuaalisia ja empiriisiä löydöksiä tavalla, jossa myös datafikaatioprosessien nyanssit tulevat esille. Flyverbom and Madsen kuitenkin huomauttavat, ettei malli ole universaali, eikä sovi kaikille asetelmille tai projekteille. Typologia on ennemminkin arvokas lähtökohta tutkimuksille, jotka keskittyvät *big data* -projekteissa tapahtuvaan järjestelyyn ja he toivovat, että sitä laajennetaan muiden toimesta. (Flyverbom and Madsen 2015, 146) Tässä tutkielmassa viitekehystä muokataan nimenomaan digitaalisia jälkiä koskevan tutkimuksen tarpeisiin.

Typologiasta voidaan kuitenkin tehdä universaaleja huomioita. Nähdäkseni vaiheisiin perustuva lähestymistapa jättää huomioimatta sen, että datasta tietoa tuottaessa aiemmat vaiheet rakentuvat edellisten päälle. Aiempiin vaiheisiin liittyy myös enemmän työtä ja siinä missä visualisointia on helppo vaihtaa lennosta, ei sama päde datan keräykseen. Asiaan

erikoistuneilla yrityksillä onkin tapana kuvata tätä pyramidilla, josta eräs esimerkki kuviossa neljä.



KUVIO 4. Eräs muotoilu ”datatieteen pyramidista” (Sense Corp 2018)

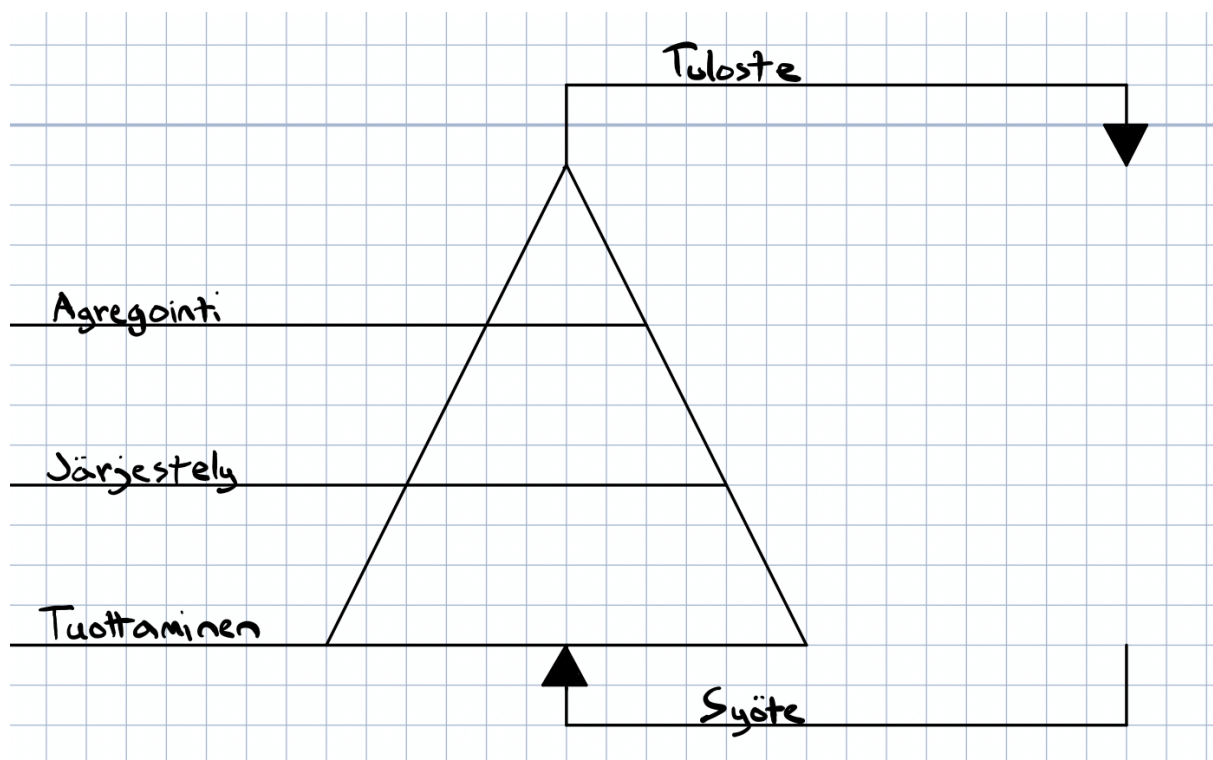
Ei pidä myöskään paikkaansa, että visualisointi olisi sinällään välttämätöntä datankäsittelyssä. Datasta voidaan myös aggregoida erilaisia numeroita, yksinkertaisimmillaan vaikka laskea kyseessä olevien dokumenttien määrä. Aggregointi toki pitää sisällään visualisoinnin ja onkin aika pitkälti semanttista, missä näiden kahden välillä on ero. Joka tapauksessa aggregointi on mielestäni visualisointia parempi termi kuvaamaan tätä vaihetta datan kanssa työskentelyssä.

Toisaalta sekä visualisaatioihin että aggregointeihin liittyy oleellisesti se, kenelle tietoa tuotetaan ja näin ollen en näe tarkoituksen mukaisena kuvata datan jakamista omana vaiheenaan. Jos datasta tuotetun tiedon loppukäyttäjät ovat osaamiseltaan vastaavalla tasolla kuin dataa analysoivat, voi jakamiseen riittää kommentoidut otteet koodista. Toisaalta Powerpoint-esitystä varten voi olla tarpeen tuottaa jotain visuaalisesti erityisen näyttävää. Näin ollen tätä osaa datankäsittelystä määrittävät vahvasti ne tulosteet, joita datasta halutaan.

Kuten aiemmin todettua digitaaliset jäljet ovat määritelmällisesti kolmansien osapuolien välittämiä ja muissakin tapauksissa digitaalisia yhteiskuntatieteitä harjoitetaan usein tilanteessa, jossa aineistoin digitoimisen prosessiin ei ole mahdollista päästä käsiksi. Dokumentoinnin puutteessa johtuen tähän osaan datan tuottamisesta ei voida soveltaa saman

tasoista teknistä reflektiota kuin esimerkiksi datan järjestämiseen. Datan tuottamista määrittävät ennen kaikkea syötteet eli se, minkälaista dataa halutaan ja tässä tullaan hyvin lähelle aiemmin esitettyjä Elishin ja Boydin ajatuksia.

Tässä tutkielmassa käytetty data-analyysin typologia on kuvattu kuviossa 5. Tässä muodossaan viitekehys on nähdäkseni universaali kaikessa *big dataan* liittyvässä yhteiskuntatieteellisessä tutkimuksessa teoreettisista oletuksista tai tulkinnallisista viitekehyksistä riippumatta. Tämä ei tarkoita, että dataa voisi koskaan tarkastella ilman tulkintaa ja tästä erilaiset teoreettiset käsitykset toimivatkin syötteinä ja tulosteina, jotka sitten vaikuttavat koko datankäsittely prosessiin. On myös huomattava, että esimerkkinä aiemmin mainitut digitaaliset karttapalvelut toteuttavat tätä typologiaa.



KUVIO 5. Tässä tutkimuksessa käytettävä datankäsittelyn typologia

3. POLICY-ANALYTIikka JA SOSIAALISEN MEDIAN MONITOROINTI

3.1 Analyysistä analytiikkaan

Policy-analyysia on kutsuttu demokratian politiikkatieteeksi, jolla viitataan yhtäältä sen tieteelliseen lähestymistapaan ja toisaalta sitoutumiseen edustuksellisen demokratian instituutioita kohtaan (Haikio ja Heino 2014, 18–19). Perinteisesti analyysin tehtävänä on ollut tukea demokraattista päätöksentekoa tarjoamalla mittasuhteita poliittisten ongelmien määrittelyyn, kerätä ja analysoida todisteita ohjeeksi päätöksenteolle, uskollisesti implementoida päätöksiä ja objektiivisesti arvioida poliittisten interventioiden tehokkuutta. Policy-analytikoille on annettu tehtäväksi tarjota skaalaa ja tarkkuutta poliittisten ongelmien määrittelyyn, kerätä ja analysoida todisteita, tukea päätöksentekoa pelottomilla ohjeilla, uskollisesti implementoida päätöksiä ja objektiivisesti arvioida poliittisten interventioiden tehokkuutta. (Longo ja McNutt 2018, 367.)

Policy-analyysi muotoutuu policy-syklin pohjalta, joka koostuu erilaisista vaiheista: ongelman määrittelystä, ratkaisuvaihtoehtojen pohdinnasta ja analyysistä, ratkaisun tunnistamisesta ja lähestymistavan valinnasta, toimeenpanosta sekä policyn arvioinnista (Longo ja McNutt 2018, 374; Sihto 2015, 336). Ongelman määrittelyn vaiheessa pohditaan ongelman kehystämistä, eri väestöryhmille koituvia hyötyjä ja haittoja sekä vaikutusten tunnettavuutta. Ratkaisuvaihtoehtoja analysoidaan sen suhteen, minkälaisia ratkaisuja on tarjolla ja toisaalta ratkaisun vaikutuksia ja hyväksyttävyyttä eri ryhmien suhteen. Ratkaisun lähestymistavan valinnassa joudutaan pohtimaan, mitkä ovat poliittisia, tieteellisiä, hallinnollisia, taloudellisia ja muita tekijöitä, jotka vaikuttavat valintaan, ja millä tavalla ne vaikuttavat. Implementaatiossa punnitaan sitä, onko valittu politiikka mahdollista ja voidaanko sitä toimeenpanna käytännössä. Evaluaatiossa analysoidaan, onko määriteltyyn poliittiseen ongelmaan pystytty puuttumaan ja kuinka politiikka voidaan parantaa, muuntaa tai lopettaa. (Sihto 2015, 336.)

Vaiheita ei voi tarkasti erottaa toisistaan, vaan ne ovat osin päällekkäisiä. Poliitiikan menestyksellisyyden kannalta tärkeää on erityisesti ongelman määrittelyn ja toimeenpanon

välinen suhde, jonka tulkitaan usein perustuvan päätöksentekijän ylhäältä-alas -malliin ja samalla rationaalis-lineaariseen päätöksentekomalliin (Sihto 2015, 336-337). Policy-analyysiä onkin vaikeaa luonnehtia yksiselitteisenä menetelmänä (Sihto 2015, 335) ja tässä alaluvussa käsitellään lähinnä policy-analyysin positivistista traditiota.

Policy-analyysin positivistinen traditio on kehittynyt yhdistelemällä yhteiskuntatieteellistä teoriaa operaatiotutkimuksesta ja taloustieteistä lainattuihin kvantitatiivisiin menetelmiin (Longo ja McNutt 2018, 369) ja on perinteisin policy-analyysin muoto. Policy-analyysia on historiallisesti kritisoitu liian hitaaksi ollakseen relevanttia päätöksenteolle (Longo ja McNutt 2018, 375) Toinen kritiikin kohde on ollut tarpeeksi laajan datan puute. (Longo ja McNutt 2018, 370). Lisäksi positivistisen politiikka-analyysin kykyä operoida demokratian kontekstissa on kyseenalaistettu kohta puolen vuosisadan verran (Longo ja McNutt 2018, 369)

Longon ja McNuttin mukaan uudet digitaalisen datan lähteet sekä datan analysointitekniikat mahdollistaisivat politiikka-analyysiin vaiheiden yhdistämisen yhdeksi jatkuvaksi prosessiksi, jota he kutsuvat reaaliaikaiseksi ja kokeelliseksi politiikka-analytiikan sykliksi. Tämän prosessin tärkeimpinä piirteinä ovat reaaliaikaisuus, skaalautuvuus ja kokeellisuus ja näin se vastaisi osaltaan myös perinteiseen positivistisen policy-analyysin kritiikkiin (2018, 374) He mainitsevat policy-analytiikan olevan tietyllä tapaa rinnakkaiskäsite laskennalliselle yhteiskuntatieteelle ja luultavasti tästä syystä näiden kahden käsitteen ympärillä käydyissä keskusteluissa onkin paljon samaa.

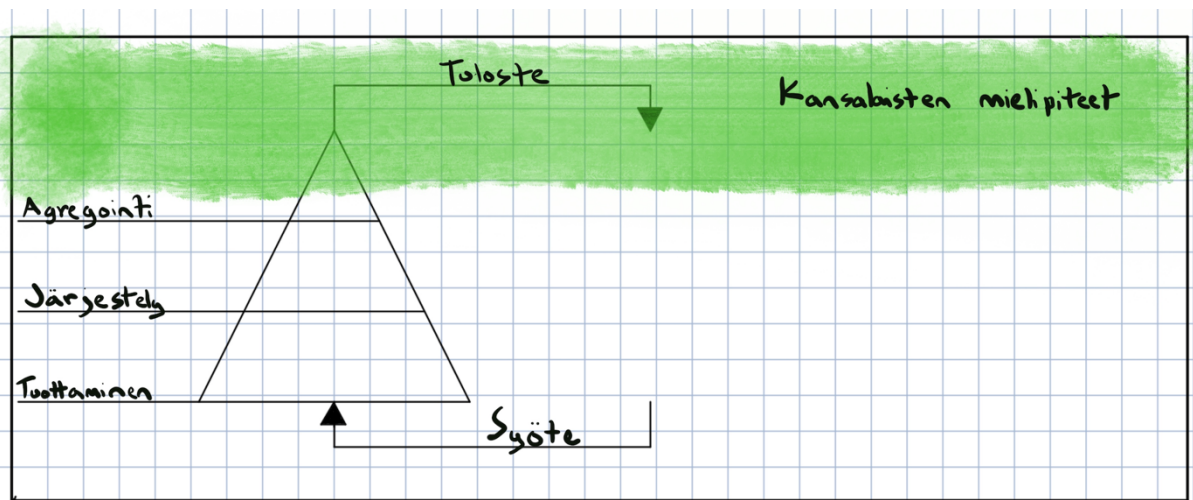
Policy-analytiikka on jälleen laaja konsepti, joka voi kätkeä sisälleen hyvin paljon erilaisia käytäntöjä. Näistä käytännöistä ja lähestymistavoista Longon ja McNuttin (2018, 373) mukaan helpoin tapa policy-analytiikan käytännön toteuttamiseen on kansalaisten preferenssien, kokemusten, arvojen ja käyttäytymisen monitorointi sosiaalisen median avulla, kuten Grubmüller, Götsch, ja Krieger asian esittävät (2013). Tässä tutkimuksessa huomio kohdistuu spesifisti kansalaisten mielipiteisiin.

Näin ollen policy-analytiikka sekä sosiaalisen median monitorointi luovat edellisessä luvussa luodulle datankäsittelyn typologialle yhden mahdollisen tulkinnallisen viitekehyksen, jossa datasta halutaan tulosteena kansalaisten mielipiteitä. Syötteenä

viitekehys taas saa sosiaalisen median dataa. Nämä syötet ja tulosteet taas sitten määrittävät osaltaan sekä datan tuottamista, järjestämistä että siitä luotavia aggregaatioita.

Longo ja McNutt myös huomauttavat, ettei kattavakaan datan määrä poista sellaisia ontologisia kysymyksiä tiedon luonteesta, jotka ovat vaivanneet politiikka-analyysin positivistista traditiota jo vuosikymmeniä. Ensinnäkään suurikaan määrä empiiristä dataa ei koskaan tarjoa täydellistä tai kattavaa kuvaa maailmasta. Toisekseen faktojen ja arvojen (tosiasioiden ja tulkintojen) välinen rajanveto ei ole kovinkaan yksinkertaista ja arvot vaikuttavat myös datankeruu- ja analyysiprosesseihin. (2018, 379) Tässäkin mukailaan hyvin pitkälti edellisen luvun keskustelua sekä datan ontologiasta että reflektiosta. Kuten todettua, näihin kysymyksiin ei tarvitse suhtautua pelkästään ontologisina, vaan niitä voidaan lähestyä myös empiirisesti teknisestä näkökulmasta.

Tämän tutkielman käytännöllisestä lähestymistavasta johtuen huomio kiinnittyy ennemminkin siihen, että sosiaalista mediaa ei ole mahdollista mitenkään yksiselitteisesti vain monitoroida. Tämä johtuu ennen kaikkea siitä, että dataa on yksinkertaisesti liikaa, eikä mikään laskentakapasiteetti mahdollista sen käsittelyä kokonaisuudessaan. Lisäksi suurin osa sosiaalisen median alustoista on tavalla tai toisella suljettuja, eikä niiden dataan ole kolmansilla osapuolilla vapaata pääsyä. Kolmanneksi on huomattava, ettei tietävästi mikään sosiaalisen median palvelu tue kansalaisuuden tunnistamista, joten datavirtaa ei pysty myöskään rajaamaan kansalaisuuden perusteella.



KUVIO 6. Kansalaisten mielipiteet datankäsittelyn typologiassa

Näin ollen, vaikka fokusointi sosiaalisen median monitorointiin antaakin edellisessä luvussa luodulle typologialle osaltaan tulkinnallisen viitekehyksen ja määrittää sen tulosteet ei se oikeastaan tarjoa tällaisenaan käytännön apua syötteiden määrittämisen suhteen. Koska sosiaalisen median palvelut eivät sinällään tue kansalaisuuden tunnistamista, on kansalaisuuden tunnistamisen itseasiassa oltava osa datan järjestämistä. Mutta millä perustein sosiaalisen median datavirtaa sitten rajataan?

3.2 Poliitiikkaprosessien tunnistaminen

Policy-analyysia on kutsuttu myös politiikka prosessien tutkimukseksi (Sabatier 2011, Sihto 2015 335 mukaan). Poliitiikkaprosessien tutkimuksessa keskeisessä asemassa on tapahtumakeskeinen tutkimus, jossa mielenkiinto kohdistetaan tapahtumiin, sosiaalisten prosessien rakentumiseen, toimijoihin sekä ajalliseen toimintaympäristöön (Tirronen 2011, 26). Tällöin policy-sykliä lähestytään vaiheiden sijaan tapahtumina ja poliitiikkaprosessien nähdään muodostuvan toisiaan seuraavista ja toisiinsa kytkeytyneistä tapahtumista. Siinä missä vaiheet havainnollistavat poliitiikkaprosessia jonkinlaisena normatiivisena ideaalityyppinä, tapahtumakeskeinen lähestymistapa painottaa enemmän poliitiikkaprosessien tosiasiallista ”juoksutusta”, jossa tapahtumaketjun muodostamat tapahtumat ehdollistavat toisiaan vaikeasti ennakoitavissa olevalla tavalla.

Yhteiskuntatieteellisen tutkimuksen tavoite on yhteiskunnan muutoksen kuvaaminen ja analyysi (Tirronen 2011, 26) ja jokainen yhteiskunnallisen uudistus muodostuu käytännössä monivaiheisesta ja -polvisesta tapahtumaketjusta. Yksittäisten uudistusten kohdalla tapahtumaketjut voivat kuitenkin poiketa toisistaan merkittävästi sekä rakenteeltaan että kestoltaan. Toiset perustuvat enemmän poliittiseen harkintaan kuin toiset ja vastaavasti tapahtumaketjujen pituudet voivat vaihdella kuukausista vuosikymmeniin. Tapahtumaketjuista voidaan kuitenkin löytää myös toistuvia säännönmukaisuuksia. (Saari 2011a, 14.)

Tapahtumaketjuanalyysilla voidaan sekä selittää muutosprosessin vähittäistä rakentumista että kuvata sen kehitysvaiheita. Tavoitteena on selvittää muutosprosessi, tapahtumaketjut sekä keskeiset käännekohtat ja selittää näiden avulla, miten muutos prosessoitui ja mitkä tekijät vaikuttivat muutosprosessin etenemiseen. Muutoksen selittämiseen on olemassa

vaihtoehtoisia tapoja, joista yhdessä muutosprosessit tapahtuvat tornadomaisesti lyhyiden muutoshetkien ja tasaantuvien kehitysjaksojen vuorotteluna. Tämän prosessin dynamiikan selittäminen edellyttää muutosprosessin intensiteetin havaitsemista eli millä aikavälillä ja millä voimakkuudella muutosprosessi toteutuu. (Tirronen 2011, 39)

Politiikkaprosessien tutkimuksessa keskeistä on huomion kohdentaminen muutosprosessin rakentumiseen ja tapahtumien historialliseen rakenteeseen. Tapahtumaketjuanalyysissä muutosta selitetään tapahtumien ajoituksella, tapahtumien välisillä yhteyksillä ja järjestyksellä. Prosessien arvioinnissa on tärkeää tunnistaa prosessin rakenne ja hahmottaa, mitä muutosprosessissa tapahtuu. Tapahtumilla sekä niiden välisillä sidoksilla ja järjestyksellä on keskeinen merkitys tapahtumaperusteisessa tutkimusotteessa. (Tirronen 2011, 26–27)

Tämän tutkimuksen kontekstissa tapahtumakeskeisistä lähestymistavoista erityisen mielenkiintoinen on analyttinen narratiivi. Analyttisessä narratiivissa tapahtumille luodaan formaali rakenne tekemällä tarkastelun kohteena olevien ilmiön kannalta tärkeimpien syiden ja seurausten väliset mekanismit mahdollisimman näkyviksi ja vähemmän tärkeät rajataan tarkastelun ulkopuolelle. (Saari 2011b, 72.) Siinä missä sosiaalisen median datavirran rajaaminen ei ole kansalaisuuden perusteella mahdollista, tarjoaa analyttinen narratiivi siihen kirjaimellisesti analyttisiä välineitä. Metodilla on myös ilmeisiä yhteyksiä digitaalisiin menetelmiin, ja tämä yhteys muistuttaa taas digitaalisten menetelmien uutuusarvosta käydystä keskustelusta.

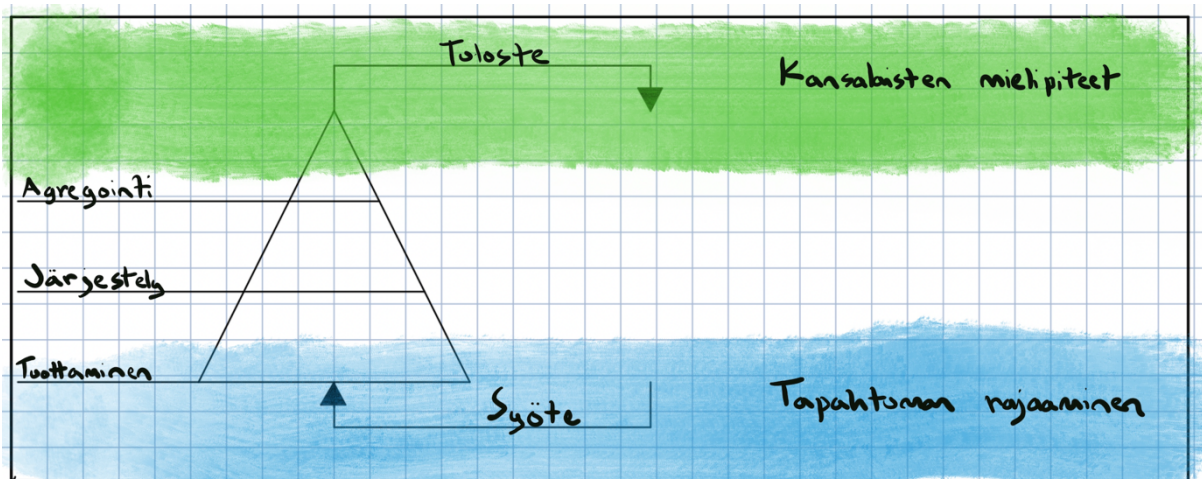
Analyttisten narratiivien luomisen apuna on jo vuosikymmenien ajan käytetty tietokoneohjelmia ja tapahtumaketjujen syy-seuraussuhteiden koodaamisen onkin sanottu vaativan sekä tietokoneohjelmien hallintaa että laajaa tietovarantoa (Saari 2011b, 92). Yleisesti politiikkaprosessien tutkimisen on sanottu olevan työlästä muuttujakeskeiseen tutkimusotteeseen verrattuna, koska tutkija tarvitsee tutkimustaan varten runsaasti tapahtumien keskinäisiä suhteita valottavaa informaatiota, jonka kerääminen on varsin aikaa vievää (Saari 2011a, 19). Digitaalinen data ja sosiaalisen median monitorointi voisivat helpottaa tätä informaation keräämistä huomattavasti.

Myös sosiaalisen median datalla on itseasiassa tehty vastaavaa tutkimusta, jossa Twitteriä on käytetty vaalien etäanalyysiin (*remote event analysis*) ja valikoiduista viesteistä on luotu

visualisaatioita. Tällöin Twitteristä on sanottu luotavan nimenomaan ”analyttinen tarinankerronta kone”. (Rogers 2019, 168.) Tällainen tapahtumiin perustuva tutkimusote vaikuttaakin ainoalta mielekkäältä tavalta toteuttaa sosiaalisen median monitorointia. Kuten ensimmäisen luvun esimerkeistä kävi ilmi, aikaleimat ovat hyvin yleinen metadatan muoto digitaalisissa objekteissa. Viestien sisällön analyysin ja aikasarja-analyysin on myös arveltu olevan robustimpia erilaisille datan puutteille, moniin muihin menetelmiin verrattuna (Jungherr 2016, 84).

Tällaiset menetelmät, kuten suurin osa tapahtumaketjujen analysointiin käytetyistä menetelmistä keskittyvät eri tavalla kielen rooliin tapahtumien kantajana. Lähestymistapaa kutsutaan diskursiiviseksi institutionalismiksi ja se perustuu ihmisten ja organisaatioiden rajalliseen kykyyn käsitellä informaatiota monimutkaisissa päätöksentekotilanteissa. (Saari 2011b, 72–73.) Tämä ei kuitenkaan täysin vastaa sosiaalisen median monitoroinnin ajatusta, jossa kansalaisten mielipiteet ovat kaiken lähtökohta. Tapahtumakeskeisyys on lähinnä vallitsevan teknologisen todellisuuden sanelema käytännöllinen ratkaisu.

Kuten kuviosta 7 voidaan huomata, ei tapahtumakeskeinen lähestymistapa myöskään tarjoa työkaluja kuin datan rajaamiseen tapahtumakohtaisesti. Se, miten dataa järjestellään tavalla, jossa kansalaisten mielipiteet tulevat esille, on paljon hankalampi kysymys.



KUVIO 7. Tapahtuman rajaaminen yhdistettynä datankäsittelyn typologiaan

3.3 Kansalaisten demokraattinen toimijuus

Kontrastina positivistiselle politiikka-analyysille on tulkitsevan politiikka-analyysin sanottu luonteeltaan olevan normatiivista demokratian politiikkatiedettä. Normatiivisuudella tarkoitetaan yhteiskunnan jäsenten toimijuuden merkityksen korostamista ja kansalaisten toimijuuden sekä vaikutusmahdollisuuksien vahvistamista hallinnon asettamiin politiikkatavoitteisiin sitoutumisen sijasta. (Häikiö ja Heino 2014b, 18–19.) Tutkimustapa on kehittynyt reaktiona yhteiskunnan hallintakäytäntöjen ja politiikkaprosessien normittavien sääntöjen muuttumiseen, joka on haastanut tutkijat tarkastelemaan hallintoinstituutioita uudella tavalla ja laajentamaan politiikan ja hallinnon tarkastelua vakiintuneiden instituutioiden ulkopuolelle. Kulttuurisia, diskursiivisia ja materiaalisia käytäntöjä tarkasteleva sekä todellisuuden tulkinnallista ja hybridistä luonnetta korostava yhteiskuntatutkimus on myös muokannut tulkitsevan politiikka-analyysin kehittymistä (Häikiö ja Heino 2014b, 14-15).

Traditional policy analysis versus deliberative policy analysis		
	Traditional policy analysis	Deliberative policy analysis
Philosophy	Positivism	Post-positivism
Rationality	Instrumental or technical rationality	Communicative or cooperative rationality
Social background	A modern society where professionalism prevails	An era of public participation and collaborative governance
Role of policy analysts	Advisors to policy makers	Supports both policy makers and citizens, assists to create forums and conditions for dialogue and negotiation
The way to handle value/interest	Separation of fact/value, focuses on facts and technical analysis, leaves value judgment to decision makers or develops inquiries on the analysts' preference	Acknowledges and respects different values and interpretations, takes values/interest as central part of analysis
Relations with public participation	Separation of analysis from public participation	Policy analysis serves and gets inputs from public participation
Mode of problem solving	Analysts as technical problem solvers invent options and suggest favorite solution for policy makers	Analysts assist the communication and negotiation among citizens, stakeholders, and decision makers to help them find their own solutions
Basic units of analysis	Data, models	Dialogue, argumentation, and deliberation
Knowledge and capacity used	Expertise and professional abilities of analysts and experts; information is used for strengthening professional analysis	Both expertise and local knowledge; and capacity of citizens and stakeholders is employed

KUVIO 8 Perinteisen policy-analyysin erot tulkitsevaan (deliberatiiviseen) policy-analyysiin (Li 2015)

Myös tulkitsevan politiikka-analyysin piirissä on havahduttu uusien teknologioiden tuomaan murrokseen. Demokraattisen toimijuuden ymmärtämiseen on nähty edellyttävän uudenlaisia tutkimustapoja, koska yhteiskunnallisia asioita ei enää käsitellä fyysisesti näkyvissä julkisissa tiloissa, vaan näkemysten vaihtaminen tapahtuu virtuaalisesti. Kun vaikuttaminen on siirtynyt verkkoon, on yhteiskunnallisuudesta samalla siirrytty enenevässä määrin henkilökohtaisuuteen sekä suosimaan projektimaista tarkkaan rajattua toimintaa pitkäjänteisen aatteelliselle arvopohjalle rakentamisen sijaan. (Häikiö ja Heino 2014a, 249–251)

Digitaalisten teknologioiden on sanottu mahdollistavan aivan uudenlaista demokraattista osallistumista, jossa edustuksellisen demokratian ja julkisvallan rooli vaihtelee luvan myöntäjästä mahdolliseen tukijaan ja merkityksettömästä toimijasta vastustajaan. Esimerkiksi Internetin ja sosiaalisen median hyödyntämistä pidetään tärkeimpänä yhteiskunnallisena muutoksena uuden itseorganisoituvan ja verkostomaisen kansalaistoiminnan taustalla ja määritelmällisenä piirteenä esimerkiksi kaupunkiaktivismille. (Mäenpää, Faehnle, ja Schulman 2017, 240.)

Yhteiskunnan institutionaaliset rakenteet eivät kuitenkaan ole kadonneet ja poliittinen valta jaetaan edelleen edustuksellisen demokratian periaattein järjestetyissä vaaleissa (Häikiö ja Heino 2014a, 247). Digitaalisten teknologioiden demokratisoiviin vaikutuksiin on liitetty paljon odotuksia ja esimerkiksi Rasmus Nielsen on argumentoinut, että monet näistä odotuksista ovat jääneet toteutumatta, koska visiot eivät ota huomioon olemassa olevia edustuksellisen demokratian realiteetteja. Toisaalta hänkin toteaa, hieman samaan tapaan kuin Flebom ja Madsen ensimmäisessä luvussa, että kaikilla on nykyään jonkinlainen näkemys digitaalisten teknologioiden roolista demokratiassa, mutta harvoille on empiirisesti selvää, kuinka näiden teknologioiden ja demokratian vuorovaikutus oikeasti toimii. (Nielsen 2014.) Myös Häikiö ja Heino ovat todenneet demokraattisen toimijuuden ymmärtämisen on kuitenkin nähty edellyttävän uudenlaisia tutkimustapoja, koska yhteiskunnallisia asioita ei enää käsitellä fyysisesti näkyvissä julkisissa tiloissa, vaan näkemysten vaihtaminen tapahtuu virtuaalisesti (2014a, 249–250).

Kuten jo aiemmin totesin pakenevat uudet digitaaliset menetelmät vanhoja jaotteluja esimerkiksi suhteessa positivismiin. Tulkitsevan policy-analyysin piirissä käydyistä keskustelua voisi tulkita niin, että nimenomaan sosiaalisen median monitoroinnin kaltaiselle

empiiriseen menetelmälle olisi tilausta. Tulkitseva policy-analyysi lähestyy monilta osin myös tieteen ja teknologian tutkimuksen kenttää (Saikkonen 2014, 73). On huomattavaa, että monilla digitaalisia metodeja kehitelleillä tutkijoilla on itseasiassa taustaa nimenomaan tieteen- ja teknologian tutkimuksen alueilla ja alalla kehitetty myös itsenäisesti tutkimusmetodeja, joissa on paljon samoja piirteitä kuin digitaalisessa ympäristössä tapahtuvassa tulkitsevassa politiikka-analyysissä voisi kuvitella olevan. (Venturini 2010b; 2010a.) Näiden menetelmien käyttämistä 'demokratian edistämiseen' on myös kokeiltu. Demokratian käsite jää tässä kuitenkin hieman ohueksi politiikan tutkijan näkökulmasta ja toisaalta kokeilussa huomattiin, että tutkimustulosten poliittisen relevanssin viestintä oli ongelmallista (Venturini ym. 2015, 74).

Tieteen ja teknologian tutkimuksen käsitteistöllä tulkitsevaa politiikka-analyysia on kehitetty muun muassa sosio-materiaalisten käytäntöjen näkökulmasta (Bamberg 2014). Sosio-materiaaliset käytännöt pohjautuvat materiaaliseen semiotiikkaan, jonka mukaan myös teknologisilla artefakteilla voidaan nähdä olevan toimijuutta ja näin ne voivat sekä jäsentää että muuntaa inhimillisten toimijoiden suhteita (Bamberg 2014, 205). Materiaaliseen semiotiikkaan pohjaavaa käsitystä artefaktien toimijuudesta on myös kritisoitu esimerkiksi ympäristökysymysten kohdalla siitä, ettei materiaalista toimijuutta voi käsitellä symmetrisesti inhimillisen toimijuuden kanssa, koska materiaalista toimijuutta joudutaan joka tapauksessa tarkastelemaan inhimillisten toimijoiden kautta (Åkerman, Maria 2009, 253–54). Tämä kritiikki ei kuitenkaan nähdäkseni koske teknologisia sovelluksia, joille jo lähtökohtaisesti suunnitellaan erilaisia toiminnallisuuksia.

Esimerkiksi Bambergin tutkimuksessa teknologisen artefaktina on verkkosovellus, joka kehitettiin osallistamaan pienjulkisoa asuinalueen suunnitteluprosessiin (2014). Tämän teknologisen artefaktin toimijuus sekä sen suunnittelu näkyi esimerkiksi siinä, että sovelluksesta päätettiin tehdä kaikille avoin. Lopulta sovelluksen käyttäjistä suurin osa asui suunnitteluprosessin kohteena olevan asuinalueen ulkopuolella ja tämä muokkasi keskustelua odottamattomaan suuntaan. (Bamberg 2014, 212). Jos osallistaminen olisi toteutettu kyseisellä asuinalueella fyysisessä tilassa ja siitä olisi ilmoitettu esimerkiksi asuinalueen lähikaupan seinällä, olisi suunnittelusta käytävä keskustelu ollut luultavasti hyvin erilaista.

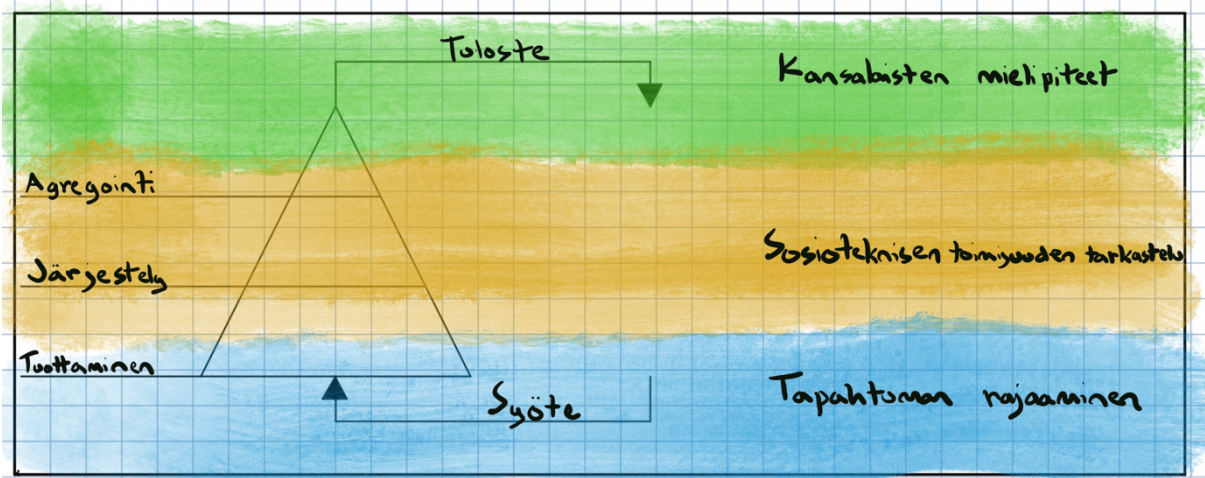
Saman tyyppisiä kysymyksiä tulee vastaan myös sosiaalisen median monitoroinnissa. Onko somepalvelun algoritmit viritetty niin, että vain samaa mieltä olevat keskustelijat löytävät toisensa? Puhuvatko käyttäjät samoista asioista samoilla termeillä? Edistääkö alusta vain vahvimpien toimijoiden näkyvyyttä? Missä määrin sen yhteisöt ovat suljettuja? Näiden ja muiden vastaavien kysymysten tarkempi teoreettinen tarkastelu joudutaan rajaamaan tämän tutkimuksen ulkopuolelle, mutta siltikin jokainen sosiaalisen median monitorointia tekevä joutuu jollain tavalla ottamaan niihin kantaa.

Teknologian ja sen käyttäjien vuorovaikutusta voi toki tarkastella monesta muustakin teoreettisesta näkökulmasta, eikä tieteen ja teknologian tutkimuksen käsitteistö ole välttämätön. Uusia digitaalisia tutkimusmenetelmiä käyttävillä tutkijoilla on paikoittain tapana nähdä teknologialla jonkinlaista toimijuutta (Nelimarkka 2020, 7) ja esimerkiksi kriittinen algoritmitutkimus on kaiken aikaa kasvava tutkimusalue (Algoritmien valta 2017). Human-Computer Interaction on monitieteinen tutkimusalue, jossa tutkitaan ihmisen ja koneen välistä vuorovaikutusta ja joka on jo vakiintunut akateemiseen maailmaan siinä määrin, että alalla on Suomessa oma maisteriohjelmakin (Aalto University 2020). Luultavasti näillä aloilla on paljon sellaista tietoa, josta olisi hyötyä myös sosiaalisen median monitoroinnissa.

Todettakoon myös, että sosiaalisen median monitorointia tekevän tutkijaposition on huomattavasti lähempänä esimerkiksi edellä mainittua Bambergin tutkimusta kuin ”ylhäältä-alas”-tyyppistä tutkimusasetelmaa. Käytännössä jokainen sosiaalisen median käyttäjä monitoroi jollain tasolla sosiaalista mediaa ja joutuu seulomaan sieltä relevantiksi kokemaansa sisältöä ylitsepursuavasta informaation määrästä. Sama tilanne on tutkijoillakin.

Idealistisimmillaan sosiaalisen median monitorointi voitaisiinkin nähdä kansalaisten osallistamismenetelmänä demokraattiseen päätöksentekoon. Joka tapauksessa, huolimatta normatiivisesta suhtautumisesta kansalaisten toimijuuden merkityksen korostamiseen, ei positivistisen policy-analyysin traditio tarjoa keinoja kansalaisten mielipiteiden tunnistamiseen sosiaalisen median datavirrasta. Tästä syystä tulkitsevan policy-analyysin näkökulma on ensisijaisen tärkeä. Myös tieteen ja teknologian käsitteistö mahdollistaa monien sellaisten sosioteknisten vaikutusten kuvaamisen ja kriittisen tarkastelun, joita

sosiaalisen median monitoroinnissa tulee vastaan, ja joiden käsittelyyn positivistisen policy-analyysin traditiosta ei löydy työkaluja.



KUVIO 9. Sosioteknisen toimijuuden tarkastelu yhdistettynä datankäsittelyn typologiaan

3.4 Sosiaalisen median monitorointi käytännössä

Kuvioon 9 on koostettu, mitä sosiaalisen median monitoroinnin ajatellaan tässä tutkimuksessa olevan. Tätä voisi kutsua vaikka ”käytännölliseksi sosiaalisen median monitoroinnin sykliseksi”. Tämä ei ehkä ole teoreettisesti kaikista kaunein rakennelma, mutta nähdäkseni se kuvaa hyvin pitkälti sitä, miten sosiaalisen median monitorointia on käytännössä toteutettava. Tässä syklissä perustavanlaatuisena ajatuksena ovat edellisessä luvussa esitellyt digitaaliset jäljet ja toisaalta digitaalisten metodien tutkimusparadigma, jossa nämä jäljet ja ”internetin elävyys” tietyllä tapaa hyväksytään osaksi tutkimusstrategiaa ja tutkimusta tehdään ”internetin kanssa” (Rogers 2009).

Policy-analytiikan näkökulmasta tämä on kaksipiippuinen juttu. Jos kansalaiset haluavat käydä poliittista keskustelua yritysten hallinnoimilla alustoilla, on keskustelun demokraattinen regulaatio haastavaa. Toisaalta esimerkiksi Facebook on pyytänyt, että julkinen sektori nimenomaan reguloisi sosiaalisen median yrityksiä enemmän, koska niillä itsellään ei ole tähän osaamista (Macias 2020). Joka tapauksessa julkinen sektori on niin vahva toimija, ettei policy-analytiikassa tarvitse vain hyväksyä olemassa olevaa sosioteknistä todellisuutta, vaan sitä voidaan myös muuttaa. Esimerkiksi koronakriisin

aikana Suomen valtio monitoroi Twitter-keskustelua, mutta myös kehitti monitorointiin Koronavilkku sovelluksen, jonka otti käyttöön lähes puolet Suomen aikuisväestöstä (Knapp 2021).

Onkin aina syytä pitää mielessä, ettei internetiä tai digitaalisia teknologioita ole lähtökohtaisesti luotu tutkimustarkoitukseen, saatikka edistämään demokratian toteutumista. Internet on alun perin armeijan kehittämä järjestelmä ja tätä nykyä sitä kehitetään lähinnä kaupallisten toimijoiden ehdoilla. Nämä toimijat ovat maksaneet huikeita summia internetin infrastruktuurin pystyttämisestä, lähtien kaapeleista merenpohjalla, jotka tällä hetkellä yhdistävät jo suurimman osan ihmiskunnasta. Samalla nämä samat toimijat myös toimivat portinvartijoina digitaalisille jäljille. (Venturini ym. 2018, 4197) 'Big data aikakauden' paradoksin on sanottu olevan, että huolet yksityisyydestä rajoittavat julkisen sektorin mahdollisuuksia kerätä tietoa kansalaisista, kun samalla yritysten datapankit kasvavat kasvamistaan (Longo ja Mcnutt 2018, 379). Näistä syistä myös digitaalisella datalla operoivat tutkijat joutuvat datankerauksessa sopeutumaan yritysten määrittelemiin sääntöihin ja rajapintoihin (Laaksonen ja Salonen 2018).

Digitaaliset menetelmät eivät myöskään sovellu kaikkeen tutkimukseen ja toisaalta saman datan perusteella voidaan esittää hyvin erilaisia tutkimuskysymyksiä, josta Venturini ym. käyttävät esimerkkinä Steam-pelipalvelun dataa. Millaisia ovat Steam-käyttäjien pelaamistavat? Kuinka verkkopelaamisen trendit käyttäytyvät? Minkälaisia ovat tietokonepelaamisen kognitiiviset vaikutukset? Mikä on pelaamisen sosiaalinen rooli yleisesti? Kaikkiin näihin kysymyksiin voidaan tietysti vastata Steam-pelipalvelun datan perusteella, mutta datan selitysarvo ei missään nimessä ole yhtä suuri Steam-käyttäjien pelaamistapoja tai tietokonepelaamisen kognitiivisia vaikutuksia tutkiessa. (2018, 4200–4201.)

Kansalaisten mielipiteiden monitorointi luokin sosiaaliselle medialle äärimmäisen tiukat rajoitteet proxyna. Yleisesti voisi ajatella, että sosiaalisen median globaalista luonteesta johtuen, se soveltuisi myös datalähteenä parhaiten mahdollisimman globaalien yhteiskunnallisten kysymysten käsittelyyn. Toisaalta tätä nykyä monet talonyhtiötkin puivat sisäistä politiikkaansa Facebook-ryhmissä. Diskursiivisen institutionalismin näkökulmasta voitaisiin myös ajatella, että lyhyellä aikajänteellä kansalaisten asiantuntijuus korostuu ja näin ollen myös sosiaalisen median monitoroinnin painoarvo korostuu. Ts. mitä lyhyempi

politiikkaprosessi, sitä varmemmin oleellinen tieto siitä löytyy nimenomaan sosiaalisen median kansalaiskeskusteluista, eikä esimerkiksi työryhmien muistioista.

4. TWITTER SOSIAALISEN MEDIAN MONITOROINNIN DATALÄHTEENÄ

4.1 Twitter datalähteenä

Kuten toisessa luvussa todettiin, erilaisten dataprojektien käsittely yhden konseptin alla, piilottaa helposti tärkeitä datankäsittelyyn liittyviä nyansseja. Vaikka policy-analytiikka ja sosiaalisen median monitorointi antoivatkin datankäsittelylle tulkinnallisen viitekehyksen, eroavat sosiaalisen median palvelut niin paljon toisistaan, ettei niitä ole mielekästä käsitellä yhtenä kimpuna. Tässä tutkimuksessa fokus on Twitterissä, joka Flyverbomia ja Madsenia mukaillen ei ole ainoastaan teknologisesti deterministinen valinta, vaan käsillä olevien teknologioiden ja muiden vaikutusten summa. (2015, 148)

Vaikka Twitter on muiden sosiaalisen median alustojen tapaan kiristänyt datapolitiikkaansa viimeisen vuosikymmenen aikana, on sitä kutsuttu 'viimeiseksi Web 2.0 palveluksi' (Rogers 2019, 157). Periaatteessa jokainen Twitter-käyttäjien interaktio on dokumentoitu ja kolmansien osapuolien saatavilla Twitterin rajapintojen sallimissa rajoissa. Tämä avaa mielenkiintoisia mahdollisuuksia tutkijoille, jotka ovat kiinnostuneita digitaalisten välineiden käytöstä, viestien julkisuuden vaikutuksista tai Twitter-käyttäjien vuorovaikutuksesta syntyvästä informaatiovirtojen verkostosta. (Jungherr 2016, 80). Koska Twitteristä on saatavilla paljon dataa poliittisen kommunikaation analyysiin, on palvelusta tullut todella suosittu tieteenekijöiden parissa, jotka työskentelevät yhteiskunta- ja tietojenkäsittelytieteiden rajapinnassa. (Jungherr 2016, 81)

Tämä ei kuitenkaan tarkoita, että Twitteristäkään olisi kaikkea dataa saatavilla. Vielä vuosikymmen sitten Yhdysvaltalainen Library of Congress Twitterin kanssa yhteistyössä keräsi kokoelmaansa lähes kaikki koskaan lähetetyt twiitit ja tarjosi niitä myös tutkijoiden käyttöön. Ajan kuluessa aineisto kuitenkin kasvoi niin suureksi, ettei sitä enää pystytty järjestämään tutkimuskäyttöön soveltuvaan muotoon. Lopulta vuonna 2018 kirjasto ilmoitti, ettei sen ole mahdollista enää edes säilyttää kaikki twiittejä ja tätä nykyä kirjasto kuratoi kokoelmaansa, arkistoiden vain tiettyjä teemoja ja tapahtumia. (Rogers 2019, 158.) Näin

ollen myöskään Twitterin monitorointi, siinä mielessä, että kaikki siellä tapahtuva vuorovaikutus dokumentoitaisiin, ei ole mahdollista.

Twitteristä voidaan kuitenkin kerätä dataa laskennallisesti kahdella erilaisella tavalla. Ensimmäisessä tähän tarkoitukseen erikseen koodatuilla skripteillä tehdään kyselyjä Twitterin rajapinnoista tai ”skreipataan” eli raavitaan Twitterin sivustoa, ja toisessa käytetään kolmannen osapuolen ohjelmistojä datan keräämiseen (Jungherr 2016, 81.) Tämän tutkimuksen kannalta ei ole oleellista käsitellä skreippausta sen tarkemmin. Oleellista on ymmärtää, että skreippaus tuottaa HTML-tiedostoja, jotka ovat tapoja esittää tietoa, joka on tietokannassa JSON-muodossa. Näihin JSON-tiedostoihin taas pääsee käsiksi vain Twitterin omien rajapintojen kautta, joihin myös kolmannen osapuolen ohjelmistot tavalla tai toisella pohjaavat.

Twitter tarjoaa kehittäjille rajoitetun pääsyn tietokantoihinsa kahden erilaisen rajapinnan avulla. REST-rajapinnan kautta tehdään kyselyjä Twitterin tietokantoihin tiettyjen hakuparametrien perusteella, kuten etsitään twiittejä, jotka ovat tietyn käyttäjän lähettämiä tai sisältävät tiettyjä hakusanoja tai aihetunnisteita. Streaming-rajapinta taas tuottaa reaaliaikaista datavirtaa Twitterissä lähetetyistä viesteistä, joita voidaan myös rajata hakuparametreilla. (Jungherr 2016, 82.) REST-rajapinnan kautta on mahdollista hakea vain rajattu määrä dataa (Toivanen ym. 2020, 5).

Koska REST-rajapinnan toiminallisuus on rajattua ja toisaalta policy-analytiikassa tavoitellaan reaaliaikaisuutta, on Streaming-rajapinta luonteva vaihtoehto Twitter-datan monitorointiin. Mutta minkälaista dataa rajapinta tuottaa?

4.2 Twiitti digitaalisena havaintoyksikkönä

Twiitin määrittelyn voi olettaa olevan näennäisesti helppoa, mutta itseasiassa hyvin harvassa tutkimuksessa on tätä tehty. Twiitit ovat selvästikin ensimmäisessä luvussa määriteltyjä digitaalisia jälkiä, joita Twitterin rajapinnat välittävät. Tämä ei kuitenkaan ole tämän tutkimuksen kannalta tarpeeksi kattava määritelmä.

Tässä tutkimuksessa Twitteriä lähestytään nimenomaan datalähteenä tai proxyna kansalaisten mielipiteisiin, mutta esimerkiksi viestinnän tutkimuksessa tutkittava ilmiö tapahtuu usein alustalla itsessään. Vuorovaikutuksesta Twitterissä sekä sosiaalisessa mediassa yleensä onkin julkaistu paljon tutkimusta. Helander (2017) on maisterintutkielmassaan eritellyt erilaisia vuorovaikutuksen muotoja, joita Twitter-tutkimuksessa on nostettu esille ja päätynyt seuraavan listaukseen:

1. twiitit
2. suorat vastaukset (@reply)
3. käyttäjän mainitseminen (@mention)
4. yksityisviestit (Direct message)
5. tykkäykset (like)
6. uudelleentwiittaukset ja uudelleentwiittaukset maininnalla (retweet ja quote retweet)

Lista vuorovaikutuksen muodoista on hyvin pitkälti samanlainen kuin edellisessä luvussa esitellyn tulkinnallisen policy-analyysin näkökulma Twitterin teknologisista toiminnallisuuksista voisi olla. Kuitenkin, kun Twitteriä tarkastellaan nimenomaan rajapintojen kautta, eivät esimerkiksi twiitti ja uudelleentwiittaus eroa toisistaan, vaan molemmat ovat JSON-objekteja, joita rajapinnasta saadaan. Onkin jälleen huomattava, että erilaiset tutkimussuuntaukset voivat käsitellä teknologisesti samanlaista ilmiötä hyvinkin erilaisilla tavoilla, ja tämä tekee tutkimuksien keskenään keskustelemisen haastavaksi.

Myös esimerkiksi historian tutkimuksen piirissä on myös käyty keskustelua Twitterin säilyttämisestä kulttuurillisena artefaktina tulevaisuuden tutkijoiden käyttöön ja todettu, etteivät JSON-tiedostotkaan riitä tähän. Twitterissä oleellisia ovat myös erilaiset audiovisuaaliset sisällöt sekä linkit toisille sivustoille. Vaikka esimerkiksi linkki itsessään tallentuisi JSON-tiedostoon, ei se välttämättä myöhemmin avattaessa sisällä sitä sisältöä, mitä viestin lähettäessä. Toisaalta Twitter itsessään on myös vuorovaikutteinen ympäristö, jossa käyttötapoja kehittyy ajan kuluessa. Onkin likipitään mahdotonta säilyttää tätä käyttökokemusta kokonaisuutena, mutta siihen voidaan pyrkiä esimerkiksi tutkimalla Twitterin käyttöohjeita laskennallisen tutkimuksen rinnalla. (Bruns ja Weller 2016, 4.)

Tässäkin tutkimuksessa tarkastellaan linkkejä, jota twiitteihin sisältyy, mutta koska fokus on reaaliaikaisessa datankäsittelyssä, ei näiden linkkien säilyvyys ole vastaava ongelma. Linkit, kuten twiitin tekstisisältökin, ovat kuitenkin vain pieniä osia twiitistä, joka tässä tutkimuksessa määritellään pelkäksi JSON-tiedostoksi. Twitterin rajapinnan tuottamat JSON-tiedostot sisältävät noin sata riviä tekstiä, joissa on runsaasti hyvin yksityiskohtaista tietoa sekä digitaalisesta interaktiosta että sen lähettäjistä, kuten lähettäjän profiilin taustaväri.

Siinä missä esimerkiksi HTML-tiedostomuoto tai Twitterin käyttöliittymä ovat vain tapoja kuvata twiittiä, JSON-tiedosto on tietyllä tapaa twiitin perimmäinen olemus. Jos ajatellaan Twitteriä tutkittavan laadullisesti, voisi tutkija esimerkiksi tarkastella edellä mainittuja profiilien taustavärejä. Tässä tapauksessa värin ja tutkijan välissä olisi kuitenkin useita tulkinnallisia kerroksia: erilaiset näytöt toistavat värit hyvin erilaisilla tavoilla ja toisaalta myös ihmiset näkevät värit aidosti erilaisina. Mutta se väri, jonka toinen tutkija tulkitsisi punaiseksi, toinen siniseksi ja kolmas violetiksi on JSONn tiedostossa yksiselitteinen värikoodi.

Tämä ei tietenkään tarkoita, etteikö JSON-tiedoston käsittelyyn itsessään sisältyisi tulkintaa. Samaten lähestymistapa jättää huomiotta sen, että Twitter-käyttäjät toki itse valitsevat itselleen taustavärin, eivät värikoodia. Tähän intentioon ei kuitenkaan ole mahdollista päästä käsiksi siinäkin tapauksessa, että tutkittaisiinkin värejä värikoodien sijasta.

```

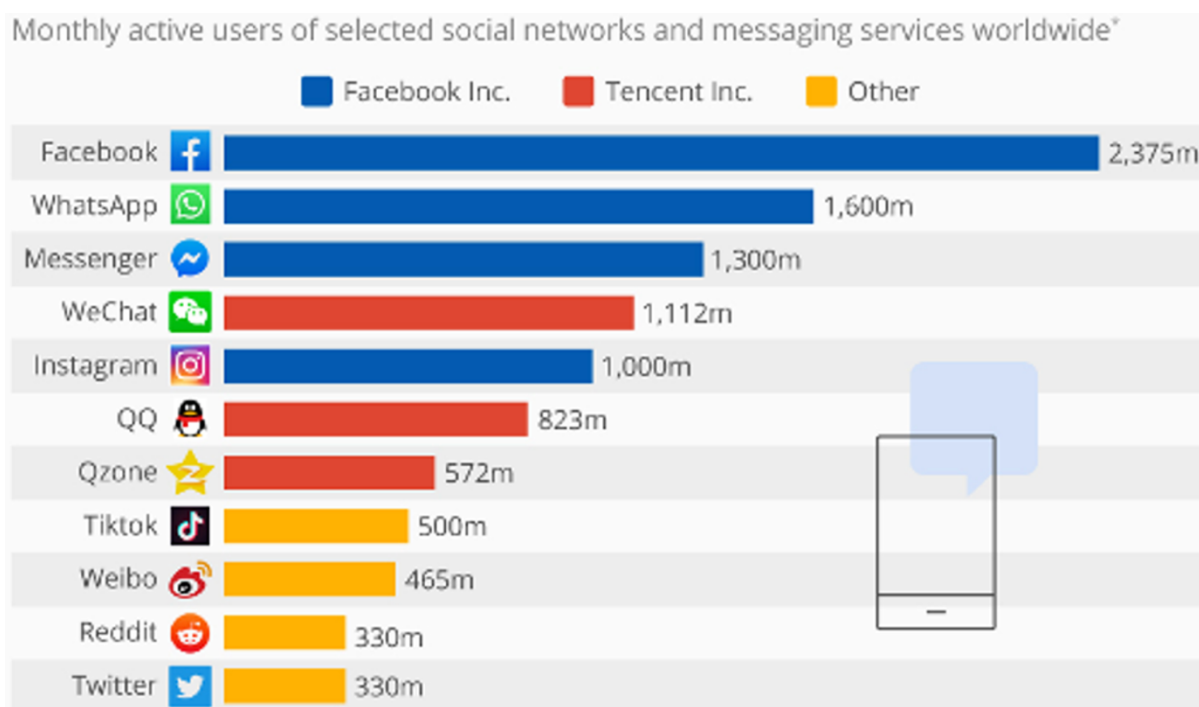
1  {}
2  "extended_tweet": {}
3  "display_text_range": [
4  0,
5  225
6  ],
7  "entities": {}
8  "urls": [],
9  "user_mentions": [],
10 "hashtags": [],
11 "symbols": []
12 },
13 "full_text": "What has happened in Christchurch is an extraordinary act of unprecedented violence. It has no place in New Zealand.
14 Many of those affected will be members of our migrant communities – New Zealand is their home – they are us."
15 },
16 "created_at": "Fri Mar 15 03:33:02 +0000 2019",
17 "source": "<a href='\"http://twitter.com\"' rel='\"nofollow\"'>Twitter Web Client</a>",
18 "in_reply_to_screen_name": null,
19 "in_reply_to_user_id_str": null,
20 "id": "1106397870628847600",
21 "contributors": null,
22 "in_reply_to_status_id_str": null,
23 "retweet_count": 16453,
24 "in_reply_to_status_id": null,
25 "filter_level": "low",
26 "place": null,
27 "favorited": false,
28 "quote_count": 1664,
29 "geo": null,
30 "text": "What has happened in Christchurch is an extraordinary act of unprecedented violence.
31 It has no place in New Zealand... https://t.co/X0HF9hHe8H",
32 "is_quote_status": false,
33 "in_reply_to_user_id": null,
34 "reply_count": 2034,
35 "coordinates": null,
36 "truncated": true,
37 "retweeted": false,
38 "favorite_count": 44761,
39 "user": {}
40 "notifications": null,
41 "created_at": "Thu Mar 05 18:57:11 +0000 2009",
42 "protected": false,
43 "follow_request_sent": null,
44 "statuses_count": 6930,
45 "name": "Jacinda Ardern",
46 "profile_background_tile": false,
47 "utc_offset": null,
48 "id_str": "22959763",
49 "profile_sidebar_fill_color": "F0F0F0",
50 "profile_sidebar_border_color": "FFFFFF",
51 "favourites_count": 707,
52 "profile_image_url_https": "https://pbs.twimg.com/profile_images/820351342464016384/_otHuDCr_normal.jpg",
53 "time_zone": null,
54 "listed_count": 878,
55 "following": null,
56 "default_profile_image": false,
57 "friends_count": 4408,
58 "profile_link_color": "FF0000",
59 "profile_text_color": "333333",
60 "followers_count": 216900,
61 "default_profile": false,
62 "profile_image_url": "http://pbs.twimg.com/profile_images/820351342464016384/_otHuDCr_normal.jpg",
63 "is_translator": false,
64 "description": "Prime Minister of NZ. Leader @nzlabour. Won't tweet what I ate for breakfast–make no promises beyond that.
65 Auth by Rt Hon Jacinda Ardern, Parlt Buildings, WLG.",
66 "profile_use_background_image": true,
67 "screen_name": "jacindaardern",
68 "id": "22959763",
69 "geo_enabled": true,
70 "profile_banner_url": "https://pbs.twimg.com/profile_banners/22959763/1501620205",
71 "profile_background_color": "131516",
72 "url": "http://jacinda.co.nz",
73 "contributors_enabled": false,
74 "profile_background_image_url_https": "https://abs.twimg.com/images/themes/theme14/bg.gif",
75 "lang": "en",
76 "translator_type": "none",
77 "verified": true,
78 "location": "Auckland, New Zealand",
79 "profile_background_image_url": "http://abs.twimg.com/images/themes/theme14/bg.gif"
80 },
81 "lang": "en",
82 "id_str": "1106397870628847617",
83 "entities": {}
84 "urls": [
85  {}(4 items)
86 ],
87 "user_mentions": [],
88 "hashtags": [],
89 "symbols": []
90 }
91 }

```

KUVIO 10. Esimerkki JSON-tiedostosta

4.3. Twitter proxyna policy-prosessiin

Twitter kuuluu ryhmään sosiaalisen median alustoja, jotka nousivat suosioon 2000-luvun ensimmäisen vuosikymmenen aikana. Näiden alustojen ominaispiirteenä on suuri määrä rekisteröityjä käyttäjätilejä sekä aktiivisia käyttäjiä. (Bruns ja Weller 2016.) Aiemmin sosiaalisen median alustojen käyttäjämääriä on vertailtu kuukausittaisten aktiivisten käyttäjien suhteen, mutta vuodesta 2019 eteenpäin Twitter on ilmoittanut päivittäisten aktiivisten käyttäjien määrän, joita oli tuolloin 145 miljoonaa ja määrä kasvanut jatkuvasti (Twitter 2019). Erilaiset tavat ilmoittaa käyttäjien aktiivisuutta tekevät eri alustojen vertailusta hankalaa, mutta joka tapauksessa Twitterin käyttäjämäärät jäävät selvästi suurimmista sosiaalisen median alustoista, kuten kuviosta 11 voi nähdä.



KUVIO 11. Suurimpien sosiaalisen median palveluiden käyttäjämäärät (Zia 2019)

Onkin todettu, että Twitter-dataan perustuvaa laskennallista tutkimusta tehdään huomattavasti enemmän kuin sen käyttäjämäärät antaisivat aiheita. (Toivanen ym. 2020, 5) Twitterin käyttöä policy-analytiikkaan kuitenkin puoltaa useampi tekijä. Siinä missä monet sosiaalisen median palvelut ohjaavat käyttäjiään ainakin jossain määrin suljettuihin yhteisöihin, on Twitterin lähestymistapa päinvastainen.

Twitterin kommunikaatiostrukturi on ainutlaatuinen suurten sosiaalisen median palveluiden joukossa. Toisen käyttäjän toiminnan seuraaminen ei lähtökohtaisesti vaadi tältä hyväksyntää, vaan yhdeksänkymmentäviisi prosenttia Twitterin käyttäjätileistä on julkisia ja kenen tahansa seurattavissa. Tämä luo verkostorakenteen, jossa merkittävät määrät käyttäjiä seuraavat globaalisti tunnettujen yksilöiden ja organisaatioiden käyttäjätilejä, kuten poliitikkojen tai uutismedioiden. Tämä kommunikaatiostruktuurin erityisyys korostuu erityisesti ajankohtaisten ja merkittävien uutistapahtumien sekä muiden vastaavien ”informaatiokaskaadien” kohdalla, jolloin informaatio leviää Twitterissä poikkeuksellisen nopeasti poikkeuksellisen laajalle yleisölle. (Bruns ja Weller 2016)

Twitterille onkin muodostunut erityinen asema sosiaalisen median alustojen joukossa, mitä tulee uutistapahtumien käsittelyyn (Toivanen ym. 2020, 5). Samalla se tarjoaa monenlaisia mahdollisuuksia poliittisen viestinnän suhteen, kun poliittiset vaikuttajat voivat viestiä suoraan kohderyhmälleen ilman perinteisen median välittävää roolia. Tästä syystä palvelusta on myös tullut merkittävä tietolähde toimittajille (Strippel 2014, Helanderin 2017, 9 mukaan). Twitter-käyttäjien on myös sanottu osaltaan osallistuvan kollektiiviseen merkitysneuvotteluun poliittisen eliitin, toimittajien sekä toisten käyttäjien kanssa (Jungherr 2016, 83)

Vaikka Twitter-tutkimusta tehdäänkin enemmän kuin sen käyttäjämäärät antavat aiheita, voidaan asetelma nähdä toisinkin päin. Twitterin on myös argumentoitu tarjoavan pääsyn datatyyppeihin, joka edustaa samalla muiden digitaalisten palvelujen, kuten Facebookin tai Googlen, keräämää dataa. Tästä syystä esimerkiksi poliittisen käytöksen tutkiminen Twitter-aineiston pohjalta ei välttämättä kerro vain Twitteristä, vaan mahdollistaa tutkimusmenetelmien kehittämisen vastaaville digitaalisille aineistoille yleisesti. (Jungherr 2016, 81.)

Joka tapauksessa policy-analytiikan ja sosiaalisen median monitoroinnin kannalta vaikuttaa selvältä, että *pelkästään* Twitter-dataan tukeutuminen ei anna kattavaa kuvaa kansalaisten mielipiteistä. Kuten monessa muussakin tutkimuksessa, joudutaan tässäkin vain hyväksymään se, että Twitteriä tutkitaan, koska sieltä on dataa tarjolla.

4.4 Muistilista Twitterin monitoroinnille

Huomattavaa on, että Twitter-tutkimukselle ei ole kehittynyt yhteistä standardia datankeruun raportointiin ja isossa osassa tutkimuksia jätetään kokonaan ilmaisematta, minkälaista datankeruun metodia on käytetty. Tästä syystä vertailevaa systemaattista tutkimusta erilaisten datankeruumetodien, rajapintojen, ohjelmistojen tai datan rajaamisen suhteen ei ole myöskään tehty. Koska datankeruuta ei raportoida, ei vinoutuneen otoksen vaikutusta tutkimukseen ei ole vielä arvioitu. On kuitenkin saatu alustavia tuloksia siitä, että tapa, jolla datankeruu suoritetaan, voi vaikuttaa datasettien koostumukseen ja näin ollen myös tutkimuksen tuloksiin. (Jungherr 2016, 81–83)

Esimerkiksi relevantin datan tunnistamisessa aihetunnisteita tai hakusanoja käytetään yleensä indikaattorina siitä, että käyttäjä on kontribuoinut viestillään kyseessä olevaan keskustelunaiheeseen. Silti näissä kahdessa lähestymistavassa on eroa. Voidaan olettaa, että aihetunnisteiden käyttö painottaa sellaisia Twitter-käyttäjiä, jotka ovat tietoisempia palvelun viestinnällisistä konventioista ja näiden perusteella luotu aineisto voi olla vinoutunut. Toisaalta hakusanojen käyttö voi johtaa aineiston ”laimentumiseen”, koska niiden kautta kerätty aineisto sisältää mahdollisesti paljon viestejä, jotka eivät liity kyseessä olevaan aiheeseen. Molemmissa lähestymistavoissa tutkijoiden täytyy myös identifioida relevantit aihetunnisteet tai hakusanat ennen kuin he aloittavat datan keräämisen. Tämä ei välttämättä ole ongelma esimerkiksi etukäteen suunniteltujen kampanjatapahtumien, kuten tv-debattien suhteen, mutta kompleksisten ja moniulotteisten sosiaalisten tapahtumien, kuten vaalien suhteen, tilanne on paljon hankalampi. Myöskään aihetunnisteiden ja hakusanojen käytön eroista datasetin muodostumisessa ei ole systemaattista vertailua ja tämä tekee hankalaksi arvioida, johtavatko nämä kaksi erilaista lähestymistapaa erilaisiin datasetteihin. (Jungherr 2016, 82–83)

Toinen ilmeinen tapaus tutkimuksien konsolidaatiolle olisi keskittyä lähestymistapoihin datan keräämisessä sekä valinnassa. Kuten yllä todettiin, tutkijat keräävät dataa Twitteristä hyvin erilaisilla lähestymistavoilla. Rajapintakyselyjen lisäksi käytetään erilaisia kolmannen osapuolen sovelluksia sekä manuaalisesti kopioidaan tietoa twiiteistä ja käyttäjien fiideistä Twitterin nettisivulta. Vaikka nämä lähestymistavat voivat johtaa vertailtaviin ja tasapainoisiin tuloksiin, tätä harvoin systemaattisesti testataan. Vaikkakin tässä tutkimuksessa käytetty lähestymistapa on teoreettisesti perusteltu, on tärkeää huomioida, ettei erilaisten tutkimusasetelmien vaikutusta datasettien muodostumiseen yksinkertaisesti tunneta (Jungherr 2016, 84)

Ensimmäiset vuodet Twitterin poliittisessa tutkimuksessa ovat luoneet rikkaan pohjan empiiristä evidenssiä, vaikkakin vain heikosti keskenään kytköksissä olevia tutkimuksia. Jotta ala voisi kehittyä pidemmälle, pitää tulevana vuosina konsolidoida evidenssiä, datankeruun sekä -valinnan metodeja ja yhdistää perinteisiä metodeja uusiin. Tämän konsolidaation kautta on mahdollista vertailla uusia tapaustutkimuksia sen suhteen, miten ne eroavat vakiintuneista käytännöistä ja kuinka paljon nämä käytännöt ovat sidonnaisia tiettyihin kontekstuaalisiin faktoreihin. Lisäksi traditionaalisten ja uusien tutkimusmenetelmien käyttö rinnakkain mahdollistaa twitter-tutkimuksen yhdistämisen vakiintuneeseen sosiaalitieteiden tutkimukseen. Tämä integraatio on välttämätöntä tarkastellessa Twitterin roolia ei erillisenä digitaalisenä välineenä, vaan yhtenä monista välineistä, jotka muuttavat politiikkaa. (Jungherr 2016, 85.)

Kuten ensimmäisessä luvussa todettiin, pitäisi digitaalisen datan suhteen pyrkiä ennen kaikkea teknologiseen reflektioon, mutta aina se ei ole mahdollista, kuten Twitter-tutkimuksen tapauksessa valitettavasti on. Tällöin digitaalisetkin tutkimusasetelmat alkavat hyvin pitkälti muistuttaa etnografista tutkimusta. Tästä syystä tämän tutkimuksen empiirisessä osassa painotetaan teknisten detaljien mahdollisimman tarkkaa dokumentointia, jotta vertailevalla laskennalliselle tutkimukselle olisi tulevaisuudessa mahdollisuuksia.

Samasta syystä tämä luku päättyy, mukaillen *reality checklist for digital methods* -artikkelia (Venturini ym. 2018), sellaisiin kysymyksiin, joita Twitterin monitoroinnissa on otettava huomioon. Kysymykset on kuitenkin muotoiltu tavalla, joka on yleistettävissä myös muille sosiaalisen median alustoille. Hahmottelen myös näille kysymyksille ideaaliset

laskennalliset testausasetelmat, jotka voisivat olla mahdollisia, jos Twitter-tutkimukselle olisi kehittynyt enemmän empiiristä todistusaineistoa. Kuten todettua, tällä hetkellä näin ei kuitenkaan ole.

TAULUKKO 1. Muistilista Twitterin monitoroinnille ja mahdolliset tutkimusasetelmat

Muistilista Twitterin monitoroinnille	Mahdollinen empiirinen tutkimusasetelma
Missä määrin tutkittava ilmiö tapahtuu verkossa ja mikä on digitaalisen datan tuoma lisäarvo?	Kerättäisiin jatkuvasti dataa poliittisesta keskustelusta Twitterissä, jolloin tutkimuksen kohteena olevan politiikkaprosessin keskustelun määrää voitaisiin verrata normaaliin poliittiseen keskustelun määrään.
Otatko operationalisaatiossasi huomioon mediumin tavat antaa data?	Vertailtaisiin esimerkiksi Twitterin rajapintojen vaikutuksia erilaisten tutkimusasetelmien datankeruussa.
Otatko operationalisaatiossasi huomioon käyttäjien tavat käyttää mediumia?	Vertailtaisiin esimerkiksi aihetunnisteiden käyttöä keskustelussa suhteessa erilaisiin tapahtumiin.
Löytääkö analyysimenetelmä tunnetut tapahtumat? Entä kansalaisten mielipiteet?	Semanttinen algoritmi osaisi itsenäisesti tunnistaa mielipiteet ja tutkijan tehtävänä olisi algoritmin toiminnan parantaminen

4. CASE CHRISTCHURCH

Tässä luvussa testataan luvuissa kaksi ja kolme luotua sosiaalisen median monitoroinnin viitekehystä keväällä 2019 tapahtuneeseen Uuden-Seelannin aselakiuudistukseen. Myös aselakien uudistamiseen oleellisesti liittyviä Christchurchin terrori-iskuja käsitellään siinä määrin kuin tutkimuksen kannalta on tarpeellista. Luvun rungon muodostaa edellä esitelty muistilista kysymyksistä, joihin Twitteriä monitoroidessa olisi kyettävä vastaamaan. Tässä joudutaan kuitenkin jatkuvasti sen ongelman äärelle, että policy-analytiikka on määritelmällisesti jatkuva ja syklinen prosessi. Kysymyksiin vastaaminen lineaarisesti etenevän tekstin muodossa tarjoaa aina vain hetkittäisen tilannekatsauksen tähän prosessiin.

Kuten edellisessä luvussa todettiin, muistilistan kysymyksiin pitäisi ideaalitulanteessa löytää vastaukset empiirisen vertailevan tutkimuksen keinoin. Vertailukohteiden puutteen vuoksi, tähän ei kuitenkaan kyetä myöskään tämän tutkimuksen osalta. Pyrkimyksenäni kuitenkin on avata tutkimusprosessissa tehtyjä valintoja niin paljon kuin mahdollista, ja näin mahdollistaa vertailevaa tutkimusta tulevaisuudessa.

Ensimmäisessä alaluvussa käsitellään Christchurchin terrori-iskuja sekä Uuden-Seelannin aselakien muutosta suhteessa Digitaalisten metodien tutkimusperinteessä kehitettyyn ”*online groundedness*” käsitteeseen (Rogers 2009, 2019). Hankalasti suomentuvalla käsitteelle tarkoitetaan, että digitaalisella datalla ja menetelmillä perustellaan yleisiä väitteitä kulttuurillisesta tai sosiaalisesta muutoksesta, joita ei tarvitse todentaa perinteisin menetelmin. Venturini ym. (2018, 4198) kuitenkin huomauttavat, digitaaliset menetelmät sopivat tutkimukseen vain silloin kuin tutkittava ilmiö tapahtuu myös digitaalisesti, ainakin osittain. Toisessa alaluvussa sama tarkastelu ulotetaan spesifisti Twitterin tapoihin ”antaa” dataa ja sen mahdollisiin vaikutuksiin ilmiön tutkimisessa.

Kolmannessa tarkastellaan käyttäjien tapoja eli aselakikeskustelun rakennetta muun muassa aihehallinnuksen (Toivanen ym. 2020; Nelimarkka 2019; Makkonen ja Loukasmäki 2019) ja visuaalisen analytiikan (visual-analytics.eu 2017) avulla. Näillä keinoin aineistosta, joka on kerätty terrori-iskuja silmällä pitäen, saadaan erotettua nimenomaan aselakeja koskeva keskustelu.

Neljännessä alaluvussa aselakeja koskevista twiiteistä valikoidaan osa lähiluettavaksi, jonka perusteella muodostetaan analyttinen narratiivi (Saari 2011b; Rogers 2019) ja käytetään määrällistä sisällönanalyysia kansalaisten mielipiteiden kartoittamiseen (Seitamaa-Hakkarainen 2014).

TAULUKKO 2. Muistilista Twitterin monitoroinnille ja miten tässä tutkimuksessa vastataan

Muistilista Twitterin monitoroinnille	Tässä tutkimuksessa vastatataan
Missä määrin tutkittava ilmiö tapahtuu verkossa ja mikä on digitaalisen datan tuoma lisäarvo?	Käydään läpi Christchurchin iskuja ja aselakiprosessia oleellisin osin ja tarkastellaan tapahtumien digitaalisia erityispiirteitä
Otako operationalisaatiossasi huomioon mediumin tavat antaa data?	Käsitellään Twitterin rajapintojen, aikeleimojen ja uudelleentwiittausten merkitystä aineiston muodostumisessa
Otako operationalisaatiossasi huomioon käyttäjien tavat käyttää mediumia?	Käytetään aihemallinnusta aineiston kartoittamiseen ja tarkastellaan käyttäjien keskustelua visuaalisen analytiikan keinoin
Löytääkö analyysimenetelmä tunnetut tapahtumat? Entä kansalaisten mielipiteet?	Luodaan aineistosta analyttinen narratiivi, jonka tuloksia verrataan tunnettuihin tapahtumiin. Tarkastellaan määrällisen sisällön analyysin keinoin analyttista narratiivia sekä kansalaisten mielipiteiden osuutta sen muodostumisessa.

4.1 Kuinka suurelta osin ilmiö tapahtuu verkossa ja mikä on digitaalisten menetelmien tuoma lisäarvo?

4.1.1 Christchurchin terrori-iskujen online groundedness

Maaliskuun viidentenätoista päivänä 2019 aseistettu henkilö hyökkäsi kahteen moskeijaan perjantaisen rukoushetken aikaan Uuden-Seelannin Christchurchissa. Hyökkääjä tappoi 49 ihmistä ja myöhemmin vielä kaksi haavoittunutta kuoli saamiinsa vammoihin. Kyseessä oli kiistatta vakavin terrori-isku Uuden-Seelannin historiassa. Aiemmin terrorismia oli pidetty maassa lähinnä latenttina uhkana, mutta iskun myötä siitä tuli elettyä todellisuutta. (Macklin 2019, 18.)

Ampuja kirjautui hieman ennen ensimmäistä iskuä tunnetulle kuvalaudalle ja kirjoitti viestin, jossa kertoi aikomuksistaan hyökkäyksen toteuttamiseen. Hän kertoi viestissä myös Facebook-käyttäjätilinsä ja sanoi lähettävänsä sitä kautta terrori-iskuista live-videota. Lisäksi hän linkitti useita aiempia kuvalaudalle kirjoittamiaan tekstejä sekä 74-sivuisen manifestin, jossa selitettiin iskujen taustalla olevia ajatuksia ja ideologiaa. Kuvalaudalle lähettämässään viestissä hyökkääjä myös pyysi muita käyttäjiä levittämään hänen ajatuksiaan nimenomaan internetille ominaisilla tavoilla: ”—*please do your part spreading my message, making memes and shitposting as you usually do--*”. (Macklin 2019, 18.)

Hyökkääjä oli hyvin tietoinen meemien visuaalisesta vallasta ja niiden roolista poliittisten ajatusten kulttuurillisessa sopeutumisessa. Joissakin online-yhteisöissä hyökkääjää juhlittiinkin sankarina ja hänen toiveitaan viestin levittämisestä toteutettiin kuuliaisesti. Makaabereimmillaan tämä näkyi, kun meemillä, jossa hyökkääjän ja viiden muun äärioikeistolaisen terroristin kuvat oli muokattu keskiaikaisiksi pyhimyksiksi, myytiin t-paitoja ja mukeja. (Macklin 2019, 25.)

Hyökkääjä myös toteutti kuvalaudalla antamansa lupauksen ja kuvasi terrori-iskun päähän kiinnitettävällä kameralla ja iskujen aikana hän ikään kuin selosti tapahtumia lähetyksensä yleisölle. Noin 200 ihmistä katsoi videota reaaliajassa, eikä yksikään heistä raportoinut videosta Facebookille. Yhteensä videota oli keretty katsomaan jo 4000 kertaa siinä vaiheessa, kun Facebook poisti sen sivustoltaan ja seuraavan vuorokauden aikana yhtiö joutui tekemään vastaavia poistoja noin 1,5 miljoonaa kertaa. Huolimatta teknologiayhtiöiden yrityksistä poistaa video, tuli siitä nopeasti viraali ilmiö, joka levisi ympäri internetiä. (Macklin 2019, 19-20.)

Hyökkääjä suunnitteli videonsa visuaalisesti näyttämään ensimmäisen persoonan ampumispeliltä ja manifestissaan hän viittasi vitsinomaisesti useisiin eri videopeleihin. *Spyro: Year of the Dragon* ”opetti hänet etnonationalistiksi” ja *Fortnite* ”koulutti hänestä tappajan”. Iskuja videon välityksellä seuraavia hyökkääjä kehotti seuraamaan erästä maailman tunnetuimmista Youtube-käyttäjistä, joka tekee pelivideoita (Macklin 2019, 19.) Myöhemmin pelien ja iskujen vuorovaikutus näkyi myös niin päin, että iskujen tapahtumia on mallinnettu uudelleen tunnettujen pelien käyttäjien itse modaamiin lisäosiin ja iskujen

perusteella on jopa tehty omia pelejä. Tämän voidaan argumentoida hämärtävän simuloitun ja todellisen väkivallan välistä eroa. (Hrehorová 2019.)

Aiemmin mainittuun digitaalisten menetelmien *online groundedness* -keskusteluun liittyy olennaisesti kysymys siitä, missä määrin medium ja viesti edes voidaan erottaa toisistaan, ja tätä samaa keskustelua on käyty analogisen median osalta huomattavasti pidempään (McLuhan 1964, Venturini ym. 2018 mukaan). Macklinin (2019, 19) mukaan Christchurchin terrori-iskujen tapauksessa live-video ei ollut medium viestille, vaan nimenomaan viesti itsessään, jopa enemmän kuin hyökkääjän manifesti. Niin olennainen ja erottamaton osa iskuja digitaalinen teknologia oli.

Iskujen jälkeen Uuden-Seelannin pääministeri yritti rajata tekijän samaa huomiota ja ilmoitti, ettei aio lausua tai toistaa tekijän nimeä missään yhteydessä. Hän kehotti muitakin tekemään samoin sekä pitämään keskustelun painopisteen uhreissa ja suremisessa. (Wahlquist 2019.) Maan media noudatti pääministerin toivomaa linjaa ja jo ennen tätä ilmoitusta Christchurchin kaupunki oli kollektiivisesti päättänyt unohtaa tekijän ja jättää tämän vaille huomiota, jota hän halusi (Weaver 2019). Voidaan kuitenkin perustellusti kysyä, missä määrin hyökkääjä edes halusi huomiota perinteiseltä medialta taikka Christchurchin kaupungilta, kun hänen viestinsä oli niin selkeästi kohdennettu internetin yleisölle.

Vaikka iskujen *online-groundedness* vaikuttaa harvinaisen selvältä, on Twitterin rooli ongelmallisempi. Twitter ei lähtökohtaisesti ole videopalvelu ja voidaan hyvin syin olettaa, että digitaalinen vuorovaikutus iskujen ympärillä on keskittynyt nimenomaan videopalveluiden ympärille tai sitten kuvalaudoille, jonne tekijä jakoi viestiään. Kuitenkin kuten aiemmin todettua, on Twitter nimenomaan se Some-palvelu, jossa poliittista keskustelua käydään, eikä tämän tutkimuksen fokus ole iskuissa sinällään, vaan niistä seuranneessa aselakien uudistuksessa.

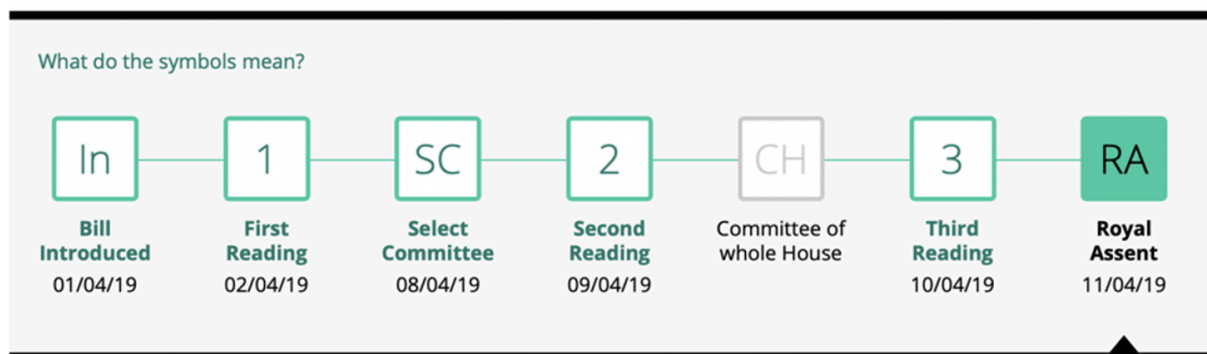
4.1.2 Digitaalisten menetelmien tuoma lisäarvo Uuden-Seelannin aselakiuudistuksen ymmärtämisessä

Christchurchin terrori-iskujen tekijä harrasti ampumista paikallisessa asekerhossa ja pidätyksen hetkellä hän omisti viisi erilaista asetta. Hänellä oli voimassa oleva aselupa ja

iskuissa käyttämänsä aseet hän oli hankkinut laillisia reittejä pitkin, vaikkakin yhtä aseista hän oli modifioinut laittomasti. (Macklin 2019, 19.) Heti iskuja seuranneena aamuna Uuden-Seelannin pääministeri ilmoittikin, että maan aselakeja tullaan muuttamaan.

Uuden-Seelannin lainsäädäntöprosessissa on normaalisti seitsemän vaihetta, joista ensimmäinen on lain esittely (Introduction), jonka jälkeen on oltava vähintään kolme päivää taukoa ennen prosessin seuraavaa vaihetta, jotta parlamentin jäsenet ehtivät tutustua aloitteeseen. Toisessa vaiheessa (First Reading) jäsenet debatoivat ja äänestävät siitä, tuleeko aloitteen käsittelyä jatkaa. Kolmannessa vaiheessa (Select Committee) aloite annetaan siihen soveltuvalla komitealla käsiteltäväksi, jolla on kuusi kuukautta aikaa kuulla yleisön ehdotuksia ja esityksiä, joiden pohjalta komitea tekee lakialoitteeseen muutosesityksiä. Muutosesitysten esittelyn jälkeen on jälleen pidettävä kolmen päivän tauko, jonka jälkeen edustajat jälleen debatoivat ja äänestävät, jatkuuko aloitteen käsittely (Second Reading). Tämän jälkeen aloite etenee komiteaan, joka muodostuu kaikista parlamentin jäsenistä, jotka haluavat osallistua komiteatyöskentelyyn (Committee of the Whole House) ja tämän komitean käsittelyssä laki saa yleensä lopullisen muotonsa. Kahden viimeisen vaiheen aikana (Third Reading ja Royal Assent) lakia ei enää muuteta ja hyvin harvoin hylätään. (New Zealand Parliament, 2020.)

Progress of the bill



KUVIO 12. Uuden-Seelannin aselakiuudistuksen tapahtumarakenne (New Zealand Parliament 2019b)

Vain viisitoista päivää Christchurchin iskujen jälkeen, parlamentille esiteltiin lakialoite *Arms (Prohibited Firearms, Magazines, and Parts) Amendment Bill*. Lakialoitteen esittelystä kului vain kymmenen päivää, kun aloite oli jo edennyt lainsäädäntöprosessin läpi

ja hyväksytty äänin 119-1. (New Zealand Parliament 2019.) Nopea aikataulu oli mahdollista, koska lakialoite julistettiin kiireelliseksi (*urgency*) ja tämä mahdollistaa poikkeamisen normaalista lainsäädäntöjärjestyksestä (Buchanan 2019). Kiireellisten lakialoitteiden kohdalla tiettyjä lainsäädäntövaiheita voidaan jättää välistä, eikä aiemmin esiteltyjä odotusaikoja, jolloin edustajat voivat tutustua lakiin, tarvitse noudattaa (New Zealand Parliament 2019a).

Onnistuneita institutionaalisia uudistuksia mahdollistavan politiikan ytimessä ovat politiikkaprosessit, jotka muodostuvat toisiaan seuraavista ja toisiinsa kytkeytyneistä tapahtumista, ja normaalisti politiikkaprosessit ovat kestoaltaan muutamasta kuukaudesta vuosikymmeneen (Saari 2011, 9-10). Vastaavia aselakien uudistamisprosesseja kuin Uudessa-Seelannissa on ollut esimerkiksi Suomessa, kun aselakeja lähdettiin uudistamaan nopeasti Jokelan ja Kauhajoen koulusurmien jälkeen, mutta uudistukset veivät kuitenkin useamman vuoden (Lindström ym. 2011, 260.) Tähän verrattuna sekä Uuden-Seelannin hallituksen reagointia terrori-iskuihin että lainsäädäntöprosessin vauhtia itsessään voidaan pitää poikkeuksellisen nopeana. Nimenomaan politiikkaprosessin nopeus nostaa digitaalisten menetelmien painoarvoa aselakiuudistuksen ymmärtämisessä usealla eri tavalla.

Ensinnäkin politiikkaprosessien tutkimisessa on keskeistä, että huolimatta tapausten aikajänteestä, tapahtumasarja voidaan tutkijoiden toimesta jakaa yksittäisiin tapahtumiin, josta aikaisemmat lisäävät myöhempien toteutumisen todennäköisyyttä (Saari 2011, 10). Voidaan kuitenkin olettaa tapahtumien erottamisen tapahtumasarjasta käyvän sitä haastavammaksi, mitä enemmän tapahtumia tapahtuu lyhyellä aikajänteellä. Aikaan liittyvää problematiikka käsitellään myöhemmin lisää ja tässä vaiheessa on riittävää todeta Twitterin olevan ajan suhteen tarkempi kuin monet analogiset tai digitoidut tietolähteet. Esimerkiksi lehti jutusta selviää ainoastaan, minä päivänä se on julkaistu, mutta twiitin julkaisuhetki on dokumentoitu sekunnilleen. Tämä mahdollistaa hienovaraisemman tapahtumien erittelyn kuin perinteiset lähteet.

Tapahtumien nopea aikajänne myös mahdollistaa tapahtumaketjun rajaamisen aineistosta tarkemmin kuin pidemmän tapahtumaketjun tapauksessa. Jos aselakien uudistamisesta keskusteltaisiin esimerkiksi vuosien ajan, on todennäköisempää, että keskustelussa käytetyt käsitteet muuttuisivat ja tapahtumien rajaaminen samoilla parametreilla ei välttämättä olisi

mielekästä. Myös median rooli itsessään voi muuttua, jos lähdeaineiston aikajänne on pitkä (Makkonen ja Loukasmäki 2019, 133). Lyhyt aikajänne mahdollistaa sen, että sekä Twitterin roolin että poliittisessa keskustelussa käytettyjen termien voidaan hyvin olettaa pysyneen samoina. Kuten edellisessä luvussa todettiin, voidaan politiikkaprosessin nopeuden nähdä nostavan sosiaalisen median monitoroinnin painoarvoa myös diskursiivisen institutionalismin näkökulmasta.

4.2 Kuinka Twitterin tavat antaa data vaikuttavat analyysiin?

4.2.1 Rajapintojen vaikutus

Tässä tutkimuksessa aineistona käytetyt twiitit kerättiin osana HYTE-tutkimuskonsortiota, jossa kehitetään uudenlaisia monialaisia tutkimusmenetelmiä terroristista väkivaltaa koskevien globaalien mediatapahtumien tutkimiseen. Projektissa oli kehitetty jo aiemmin valmis käytäntö datankeruuseen sitä silmällä pitäen, että jotain sopivaa tapahtuisi ja tätä mallia sovellettiin Christchurchin iskuihin. Tarkoituksena oli kerätä reaaliajassa suuria määriä dataa Twitteristä, niin että digitaaliset mediaetnografit seuraavat tapahtumien kulkua ja hakuehtoja päivitetään heidän ehdotustensa mukaisesti.

Reaaliaikainen datankeräys, eli aiemmin mainitun streaming-rajapinnan käyttö, nähtiin ainoana tarkoituksenmukaisena keinona kerätä suuria määriä dataa. Tutkimusta varten kehitetty keräin otti parametriksi joukon hakuehtoja, joita datan keräämisestä vastaava tutkija päivitti digitaalista mediaetnografiaa tekevien tutkijoiden ehdotusten mukaisesti. Tapahtuman akuutin vaiheen ajan mediaetnografit seurasivat tapahtumaa erilaisten medioiden kautta, datan keräämisestä vastaava tutkija seurasi keräimen toimintaa ja hakuehtoja kehiteltiin osana Facebookin pikaviestimessä käytyä keskustelua. Lopulliseksi hakuehtolistaksi muodostui: *'christchurch'*, *'christchurchmosqueattack'*, *'christchurchshooting'*, *'christchurchattack'*, *'JeSuisChristchurch'*, *'PrayForChristchurch'*, *'Al Noor'*, *'mosque shooting'*, *'mosque massacre'*, *'BrentonTarrant'*, *'NewZealandShooting'*, *'JeSuisHuman'*, *'JeSuisMuslin'*, *'HelloBrother'*, *'NewZealandTerroristAttack'*, *'NewZealandStrong'*. (Toivanen ym. 2020, 5.)

Reaaliaikasta datankeruuta on perusteltu sillä, että pyyntöihin perustuvalla keräyksellä eli aiemmin mainitulla REST-rajapinnalla voidaan kerätä vain rajallinen määrä dataa (Toivanen ym. 2020, 5). Kuitenkin myös reaaliaikainen datankeräys on jossain määrin rajoitettua ja streaming-rajapinnan tekninen maksimi on 3000 twiittiä minuutissa (Green 2020). Keräimen toiminnasta on saatavissa tietoa vain tuntitasolla, joten ei voida olla varmoja, onko tähän rajaan törmätty. Iskun akuuteimman vaiheen aikana twiittejä kuitenkin kerättiin lähes 180 000 tunnissa, joten vaikuttaa selvältä, että maksimiraja on tullut ainakin välillä täyteen ja näin ollen tuntematon määrä hakuehtoja vastaavia twiittejä on jäänyt keräämättä. Tämän tuntemattoman kokoluokkaa ei kuitenkaan ole mahdollista selvittää, mutta kyse on joka tapauksessa vain muutamasta tunnista.

Christchurch: Captured tweets per hour



KUVIO 13. Keräimen keräämät twiitit aikavälillä 16.3.2019-9.4.2019

On myöskin huomionarvoista, että keräin käy lähellä 180 000:n twiitin maksimirajaa, mutta ei kuitenkaan koskaan saavuta sitä. On esitetty epäilyjä, että Twitter jollain tasolla moderoi myös streaming-rajapinnan toimintaa ja tämä erotus voisi mahdollisesti olla tuo sensuurin määrä.

Keräin myös käynnistettiin vasta useita tunteja iskujen tapahtumisen jälkeen, koska iskujen tapahtuessa Uudessa-Seelannissa oli Suomessa yö ja keräämisestä vastaava tutkija sai tiedon iskuista vasta seuraavana aamuna (Toivanen ym. 2020, 5). Lisäksi merkittävä osa aineistosta kerättiin sinä aikana, kun hakusanalista vielä päivitettiin. Kun esimerkiksi ”HelloBrother”

lisättiin hakusanaksi vasta kymmenen aikaan illalla, on se oletettavasti aliedustettuna aineistosta, josta viidesosa oli tähän mennessä jo kerätty. On vielä erikseen painotettava, että tämä aineistonkeräyksen ”elävyys” johtuu nimenomaan Twitterin teknisistä rajoitteista, jotka eivät käytännössä mahdollista aineistonkeruuta muilla tavoin.

4.2.2 Aikaleimat ja -vyöhykkeet

Keräin oli lopulta päällä syyskuuhun 2019 asti ja keräsi yhteensä 11 884 723 twiittiä, joskin twiittien määrä painottuu hyvin pitkälti iskujen akuuttiin vaiheeseen. Koska keräin tallentaa myös uudelleentwiitatut twiitit, sisältää aineisto myös twiittejä ajalta ennen keräimen käynnistymistä. Vanhimmat uudelleentwiittaukset ovat niinkin kaukaa kuin vuodelta 2018.

Aikaleimat ovat eräs yleisimpiä ja samalla tärkeimpiä metadatan muotoja, ja ovat osa lähes jokaista digitaalista interaktiota. Tässä tutkielmassa niiden merkitystä korostaa aineiston pitkä aikajänne, joka pakottaa tekemään jonkinasteisia aikarajauksia. Esimerkiksi yksinkertaisistakin aikasarja-visualisaatioista tulee tarpeettoman monimutkaisia, jos niissä pitää ottaa huomioon myös edellisen vuoden twiitit, jotka eivät kuitenkaan tarjoa tapahtumarakenteeseen itsessään mitään lisäarvoa.

Kuten todettua keräin käynnistyi myöhässä johtuen Uuden-Seelannin ja Suomen välisestä aikaerosta. Samaten reaaliaikaisesta datankeräyksen realiteeteista johtuen hakusanalista muuttui paljon ensimmäisen päivän aikana ja näyttää selvältä, ettei aineisto ole tällä ajanjaksolla samalla tavalla edustava kuin myöhemmin. Lisäksi samalla ajanjaksolla keräimen kapasiteetti ylittyi oletettavasti useamman ja tämän takia tuntematon määrä twiittejä jäi keräämättä. Tämän tutkielman kannalta oleellinen kysymys on, että onko tällä merkitystä aselakeja koskevan keskustelun suhteen.

Nämä kysymykset myös liittyvät oleellisesti Twitterin luonteeseen globaalina somepalveluna, joka toimii ”internet-ajassa”. Twitterissä ei sinällään toimi missään yhteisesti sovitussa aikavyöhykkeessä, vaan se lokalisoi jokaisella käyttäjälle ajan tämän sijainnin mukaan. Twitterin rajapinnasta haettujen twiittien aikaleimat ovat kuitenkin Greenwichin ajassa (GMT+0), kun taas datankeräys on dokumentoitu Suomen ajassa (GMT+2) ja Uuden-Seelannin tapahtumia on tarpeen tarkastella paikallisessa ajassa

(GMT+11). Lisäksi sekä Uudessa-Seelannissa että Suomessa on käytössä kesäaika, johon siirrytään eri vaiheessa kevättä.

Kesäajan tapaiset tunnin epätarkkuudet aikaleimoissa eivät välttämättä ole tutkimuksen kannalta kriittisiä, mutta suuremmat erot voivat olla, koska tällöin aineiston tapahtumat voivat sijoittua jo täysin eri vuorokauden puolelle. Toisaalta digitaaliset aikaleimat voivat mahdollistaa täysin uudenlaista tarkkuutta tutkimusmenetelmiin, joten pieniäkään eroja ei ole syytä ohittaa olankohautuksella.

Tämän tutkielman kannalta oleellista oli tarkastella, vaikuttivatko aineistonkeräyksen epäselvyydet ensimmäiseltä päivältä aselaeista käytyyn keskusteluun. Kun hakusanalista oli muodostunut kymmenen jälkeen illalla, elettiin Uudessa-Seelannissa jo seuraavan päivän aamua. Samana aamuna maan pääministeri Jacinda Ardern myös ilmoitti aselakien uudistamisesta (CNN 2019). Tietoa siitä, *mihin aikaan* Ardern ilmoitti aselakien uudistamisesta, on kuitenkin hyvin hankala löytää. Keräimen aikaleimasta voidaan taas sanoa sekunnin murto-osan tarkkuudella, milloin hakusanalista oli lopullisessa muodossaan, kuten kuvioista 14 näkyy.

2019-03-15T22:21:41.818438+02:00 Streaming data

... search terms: christchurch, christchurchmosqueattack, christchurchshooting, christchurchattack, JeSuisChristchurch, PrayForChristchurch, Al Noor, mosque shooting, mosque massacre, BrentonTarrant, NewZealandShooting, JeSuisHuman, JeSuisMuslin, HelloBrother, NewZealandTerroristAttack

KUVIO 14. Aikaleima keräimestä

Tämä on jälleen esimerkki siitä, miten analogiset tai digitoidut lähteet eivät yksinkertaisesti tarjoa samanlaisia ominaisuuksia kuin natiivisti digitaaliset ja tästä syystä tutkimuksen juurruttaminen niihin on ylipäätään kyseenalaista. Paras keino saada selville Ardernin ilmoituksen tarkka aika, on itseasiassa tarkistaa se samaisesta aineistosta, jonka reliabiliteettia tässä on tarkoitus todentaa. Toisaalta tämä myös on osoitus digitaalisen tutkimusprosessin syklisestä luonteesta: jotta aineistosta pystyttiin etsimään kyseinen ajankohta, oli aineiston käsittelyä varten luotava jo analyysityökalut, joita käsitellään myöhemmin tässä tutkimuksessa.

Aineistosta kävi ilmi, että ensimmäinen twiitti Ardernin ilmoituksesta aselakien muuttamiseksi, oli lähetetty vain 20 minuuttia sen jälkeen kuin lopullinen hakusanalista oli muotoiltu. Näin ollen tässä tapauksessa kesäaikaan siirtymisen kaltaisilla näennäisen pienillä epätarkkuuksilla olisi ollut väliä. Tämä huomio oli myös oleellinen myöhemmin tutkimuksessa, kun analyyttistä narratiivia muotoiltiin.

4.2.3 Uudelleentwiittaus

Uudelleentwiittaus on erityisen tärkeä ominaisuus aineiston muodostumisen kannalta, koska se mahdollistaa aineistojen keräämisen myös ajalta ennen keräimen käynnistymistä. Kuten todettua aineistoa kerättiin yhteensä lähes 12 miljoonaa twiittiä, mutta tietokannassa nämä jakautuvat uudelleentwiittauksien perusteella kolmeen erilaiseen twiittikokoelmaan: live-twiitteihin, live-uudelleentwiitteihin sekä alkuperäisiin twiitteihin. Osittain näissä kokoelmissa voi myös esiintyä päällekkäistä sisältöä.

Live-twiitit ovat keräimeen tarttuneita twiittejä, joissa ei uudelleentwiitata toista twiittiä. Jos live-twiittiä on kuitenkin myöhemmin uudelleentwiitattu, esiintyy sama twiitti myös alkuperäistwiittien kokoelmassa. Koska näissä twiiteissä ei uudelleentwiitata sisältöä, on niiden tekstisisältö lähtökohtaisesti uniikkia. Live-twiittejä on tietokannassa noin miljoona.

Live-uudelleentwiitit ovat keräimeen tallentuneita twiittejä, joissa uudelleentwiitataan toista twiittiä. Samoin kuin edellä myös live-uudelleentwiittiä on voitu itsessään uudelleentwiitata, jolloin se esiintyy lisäksi alkuperäistwiittien kokoelmassa. Näissä twiiteissä voidaan joko kommentoida alkuperäistä sisältöä tai sitten vain twiitata se sellaisenaan. Ei ole selvää, kuinka suuri osa twiiteistä sisältää uniikkia tekstisisältöä ja kuinka suuri osa kierrättää vain alkuperäisen twiitin sisältöä. Suurin osa aineistosta on live-uudelleentwiittejä eli noin kymmenen miljoonaa.

Alkuperäistwiitit ovat twiittejä, joita keräimen tallentamissa twiiteissä on uudelleentwiitattu ja nämä poikkeavat oleellisesti kahdesta muusta aineiston osasta. Ensinnäkään alkuperäistwiittien ei tarvitse sisältää keräämisessä käytettyjä hakusanoja, vaan riittää, että hakusanoja käytetään siinä twiitissä, minkä osana alkuperäistwiittiä on uudelleentwiitattu. Toisekseen näitä twiittejä on myös ajalta ennen keräimen käynnistämistä.

Alkuperäistwiittien tekstisisältö on myöskin lähtökohtaisesti uniikkia. Alkuperäistwiitit myös sisältävät tiedon siitä, kuinka monta kertaa kyseistä twiittiä on uudelleentwiitattu. Näitä twiittejä on aineistossa kaikista pienin määrä eli 461 421.

”Lähtökohtaisesti uniikilla tekstisisällöllä” tarkoitetaan tässä, että Twitterin omat toiminnallisuudet eivät tässä kohtaa tue tekstisisällön kierrätystä. Mikään ei tietystikään estä käyttäjää kopioimasta tekstiään toiselta käyttäjältä ja kierrättämään näin samaa sisältöä. Myöhemmin analyysissa myös huomataan, että ätä myös tapahtuu ainakin suurten mediayhtiöiden kohdalla: ne saattavat jakaa samaa sisältöä erikseen esimerkiksi globaalilla Twitter-tilillään ja erillisellä Uuden-Seelannin tilillään. Koska suurten mediayhtiöiden twiittejä myös uudelleentwiitataan paljon, voi tällä mahdollisesti olla merkittäväkin vaikutus, riippuen käytetystä analyysimuodosta.

TAULUKKO 3. Twiittien jakautuminen erilaisiin osajoukkoihin

	Live-twiitit	Live-uudelleentwiitit	Alkuperäistwiitit
Määrä	Noin miljoona	Noin 10 miljoonaa	Noin puoli miljoonaa
Sisältää hakusanoja	Kyllä	Kyllä	Ei välttämättä
Aikaväli	Keräimen <u>alku-</u> <u>loppu</u>	Keräimen <u>alku-</u> <u>loppu</u>	2018-keräimen loppu
Uudelleentwiitaukset	Ei	Ei	Kyllä
Lähtökohtaisesti uniikkia tekstisisältöä	Kyllä	Ei välttämättä	Kyllä

Twiitit ovat tietokannassa JSON-objekteina ja on painotettava, että monet myöhemmin esitellyistä analyyseista ovat tutkimusekonomisesti hyvin hankalia tehdä koko aineistolla. Kun jokaisessa JSON-tiedostossa on noin sata riviä tietoa ja twiittejä on lähes kaksitoista miljoonaa, pitää yksinkertaista sanahakuakin varten siis käsitellä miljardin rivin verran informaatiota. Tästä syystä tässä tutkimuksessa päädyttiinkin rajaamaan tarkastelu ainoastaan alkuperäistwiitteihin.

Alkuperäistwiittien käyttämistä aineiston aggregaattina puoltaa se, että jokainen uudelleentwiittaus myös sisältää alkuperäistwiitin ja näin ollen alkuperäistwiitit myös kertovat aineiston sisällöstä laajemmin. Lisäksi niiden käyttämistä puoltaa aiemmin mainitut

epäselvyydet aineiston keräyksessä. Vaikka ei voida tietää, kuinka suuri osa hakuehdot täyttävistä twiiteistä jäi keräyksen ulkopuolelle, mitä enemmän twiittiä on uudelleentwiitattu, sitä todennäköisemmin se myös jotain kautta on aineistoon päätenyt. Näin ollen aineistossa olevien alkuperäistwiittien osuus kaikista mahdollisista alkuperäistwiiteistä on oletettavasti suurempi kuin esimerkiksi uudelleentwiittausten osuus kaikista mahdollisista uudelleentwiittauksista.

Pelkästään alkuperäistwiittien käyttö analyysissa ei kuitenkaan missään nimessä ole ongelmatonta, sikäli kun tavoitteena on kansalaisten mielipiteiden tavoittaminen. On täysin mahdollista, että fokuointi alkuperäistwiitteihin nimenomaan jättää huomioimatta niiden ”tavallisten kansalaisten” äänen, joiden twiittejä ei uudelleentwiitata. Toisaalta alkuperäistwiitit ovat nimenomaan twiittejä, jotka ovat herättäneet keskustelua, kannatusta tai miten ikinä uudelleentwiittamisen haluaakaan tulkita. Tämän tutkielman kontekstissa ei ole mahdollista tarkastella tätä kysymystä, mutta uudelleentwiittauksen funktion tarkempi määrittely olisi ehdottoman tärkeää.

4.3 Otatko huomioon käyttäjien tavat käyttää mediumia?

4.3.1 Aihemallinnus

Twitterin suhteen on veteen piirretty viiva, missä teknologiset käytännöt loppuvat ja käyttäjien tavat alkavat. Esimerkiksi aihetunnisteet ovat olivat alun perin nimenomaan käyttäjien tapa organisoida keskustelu ja myöhemmin Twitter lisäsi toiminnallisuuden virallisesti osaksi palveluaan (Eriksson Krutrök ja Lindgren 2018). Aineiston rajaaminen uudelleentwiittauksien perusteella voitaisiin nähdä myös käyttäjien tapojen mukailemisena, vaikkakin tässä tutkielmassa se tehtiin nimenomaan teknisistä syistä. Tässä tutkielmassa Twitterin käyttäjien sosioteknisillä konventioilla kuten toisten käyttäjien mainitsemisella tai aihetunnisteilla ei kuitenkaan ole erityistä roolia, vaan niiden sijasta käsitellään hyvin pitkälti sanoja.

Aineiston keräämisessä käytettiin sanoja aihetunnisteiden sijasta, koska terrori-iskujen kaltaisissa mediatapahtumissa käyttäjät eivät käytä aihetunnisteita normaaliin tapaan

(Toivanen ym. 2020, 5). Sanoihin perustuvaa lähestymistapaa edustaa myös aineistolle suoritettu aihehallinnus, josta Toivanen ym. ovat kirjoittaneet tarkemmin toisaalla (2020). Lisäksi esimerkiksi Nelimarkka (2019) sekä Makkonen ja Loukasmäki (2019) ovat käsitelleet tarkemmin aihehallinnuksen sekä ohjaamattomien koneoppimismenetelmien käyttöä yhteiskuntatieteissä ja politiikan tutkimuksessa.

12	Uuden-Seelannin pikainen puoliautomaattisten aseiden kielto	gun, ban, law, weapon, day, change, week, use, rifle, assault
----	---	---

KUVIO 14. Aihemallinnuksen tuloksena syntynyt aselaki-aihe (Toivanen ym. 2020, 13).

Tämän tutkielman kannalta on oleellista, että aselait olivat eräs aihehallinnuksen tuottamista aiheista ja aihehallinnuksen tuloksena saadaan myös aiheen merkitsevimmät sanat. (Toivanen ym. 2020, 13) Näin tämän tutkielman tekoa aloittaessa oli olemassa jo jonkinlainen kuva siitä, mitä tutkielman kannalta oleellisia piirteitä aineisto sisältää, ja kuinka sitä lähteä rajaamaan. On huomattava, että periaatteessa aineistoa voitaisiin rajata myös aihehallinnuksen itsensä perusteella ja hakea aineistosta aselaki-aihetta eniten kuvaavat twiitit. Koska aihehallinnuksen syötteenä käytetty aineisto oli kuitenkin rajattu hieman eri tavoin kuin tässä tutkielmassa, hylättiin tämä lähestymistapa ja tästä syystä aihehallinnusta ei ole myöskään tarpeen tässä tarkemmin käsitellä.

Oletuksena tässä tutkielmassa kuitenkin oli, että nimenomaan aihehallinnuksen tuottamat neljä merkitsevintä sanaa eli ”gun”, ”law”, ”ban” ja ”weapon” kuvaavat, millä termein Twitterin käyttäjät käyvät keskustelua. Luin myös aiheeseen liittyviä lehtijuttuja, joissa toistuivat fraasit ”change gun laws” sekä ”ban semi-automatic weapons”. Tämän pohjalta aineistoa oli tarpeen rajata niin, että aselakeihin liittymätön keskustelu olisi mahdollisimman tehokkaasti karsittu pois, mutta aineiston laajuus olisi silti edelleen tuhansia twiittejä, koska tämän kokoisella aineistolla Twitter-tutkimusta on aiemminkin toteutettu (Helander 2017, Hellsten ja Leydesdorff 2019).

4.3.2 Työkalut käyttäjien tapojen seuraamiseen

Tässä vaiheessa aineisto sisälsi edelleen lähes puoli miljoonaa JSON-objektia, joissa oli myös paljon tämän tutkielman kannalta epärelevanttia informaatiota. Aineiston muodosta ja koosta johtuen se oli liian raskas esimerkiksi Excelin tai muiden vastaavien taulukkolaskentasovellusten käsiteltäväksi. Toisaalta kaupalliset data-analyysiin keskittyvät ohjelmistot ovat usein kalliita, rakenteeltaan hyvin suljettuja ja omien tietokantojen saattaminen niiden kanssa yhteensopiviksi on työlästä. Näin ollen oli tarkoituksenmukaisinta tehdä itse datankäsittelyyn vaadittavat ohjelmat. Tässä käytettiin apuna pääosin seuraavia teknologioita:

- *Python* on monipuolinen, mutta yksinkertaisen syntaksin omaava, tulkettava ohjelmointikieli ("General Python FAQ — Python 3.9.0 documentation", 2021)
- *Funktio* on teknisesti ottaen Python-koodin pätkä, jota voidaan kutsua muualta ohjelmointikoodista ("Ohjelmoinnin perusteet Python-kielillä", 2021)
- *Pandas* on Pythonille kehitetty kirjasto datan analyysiin ja muokkaamiseen ("pandas - Python Data Analysis Library" 2021)
- *NLTK* on Pythonille kehitetty kirjasto luonnollisen kielen käsittelyyn ("Natural Language Toolkit — NLTK 3.6.2 documentation" 2021)
- *Plotly* on Pythonille kehitetty kirjasto interaktiivisten visualisaatioiden tekemiseen (Plotly 2021)
- *Jupyter-työkirjat* ovat selaimessa toimivia interaktiivisia kehitysympäristöjä, joita käytetään ohjelmakoodin kirjoittamiseen, ajamiseen ja jakamiseen ("Project Jupyter" 2021)

Kaikkia aineiston siivoamisen vaiheita ei ole tarpeen käsitellä tässä yksityiskohtaisesti, mutta esimerkin omaisesti kerron, mitä esimerkiksi twiitin tekstin sisältävän sarakkeen aikaansaamiseksi piti tehdä. Twitter nosti muutama vuosi sitten viestien maksimimerkkimäärän 140:stä merkistä 280:een ja tästä historiallisesta detaljista johtuen Twitterin rajapinta antaa twiiteille kaksi eri tekstikenttää "text" ja "full text". Jokaisen twiitin

”text”-kentällä on arvo, mutta siinä tapauksessa, että twiitti on yli 140 merkkiä pitkä, ei tämä kenttä sisällä koko tekstiä, vaan kokonainen teksti on ”full text”-kentässä. Tätä varten Pythonilla täytyi kirjoittaa funktio, joka tarkistaa, onko ”full_text”-kentässä sisältöä ja siinä tapauksessa, että kenttä on tyhjä, kopioi se perusmuotoisen tekstikentän sisällön ”full_text”-kenttään ja lopuksi poistaa kaikki normaalit tekstikentät.

user_screen_name	text	full_text
NafEnesate	Evidence #3 that they use creating fake SSID with a NodeMCU ESP8266 Of target's location so it can be mobbed, sensi... https://t.co/RiqsWrv11o	Evidence #3 that they use creating fake SSID with a NodeMCU ESP8266 Of target's location so it can be mobbed, sensitization, noise campaign from first responders, directed conversations, anchoring etc.\n\n#gangstalking #gangstalkers https://t.co/bdl210YK8e

KUVIO 15. Esimerkki päällekkäisistä kentistä

Lisäksi erilliset funktiot piti kirjoittaa muun muassa poistamaan sellaisia kenttiä, jotka eivät ole tämän tutkielman kannalta oleellisia; järjestämään aineisto tiettyjen sarakkeiden mukaan, jotta voidaan tarkastella esimerkiksi uudelleentwiitatuimpia twiittejä tai aikajärjestystä; muuttamaan aineiston aikaleimoja eri aikavyöhykkeiden mukaan; etsimään käytetyimpiä sanoja ja sanapareja sekä rajaamaan aineistoa erilaisilla hakusanoilla.

On kuitenkin tarpeen korostaa, ettei aineistosta varsinaisesti haeta sanoja, vaan *merkkijonoja*. Haku voidaan toki kohdentaa niin, että esimerkiksi sanaa ”law” haetaan eksaktisti tuossa muodossa. Tämä kuitenkin jättää huomiotta viestit, joissa sana esiintyy monikossa ja aselakien uudistamisesta puhuttiin usein monikkomuodossa ”gun laws”. Toisaalta ”ban” merkkijono esiintyy niin monessa englannin kielen sanassa (esim. **Brisbane**), että sen suhteen eksaktin haun käyttöä voitaisiin pitää perusteltuna.

```
def fill_the_full(df):
    df['full_text'] = np.where(df['full_text'].isnull(), df['text'], df['full_text'])
    df['user_mentions_full'] = np.where(df['user_mentions_full'].isnull(), df['user_mentions'], df['user_mentions_full'])
    df['hashtags_full'] = np.where(df['hashtags_full'].isnull(), df['hashtags'], df['hashtags_full'])
    return df
```

KUVIO 16. Esimerkki Pythonilla kirjoitetusta funktiosta

Haut voivat myös tarvittaessa huomioida isojen ja pienten kirjainten väliset erot. Käytännössä tämä ei ole kovinkaan olennaista rajattaessa aineistoa avainsanoilla, mutta esimerkiksi käyttäjänimien kohdalla tällä on väliä. Twitter ei tee käyttäjänimissä eroa isojen

ja pienten kirjainten välillä, mutta käyttäjät voivat viesteissään kirjoittaa nimiä monella eri tavalla. Tekstissä voidaan siis mainita esimerkiksi ”@JacindaArdern” tai ”@jacindaardern”, mutta molemmat näistä maininnoista viittaavat samaan käyttäjätiliin.

Lisäksi erilaisia hakusanoja voi yhdistellä hyödyntäen loogisia operaattoreita (”Propositiologiikka” 2020). Aineistossa esimerkiksi viitataan ampujaan usein termillä ”gunman” ja tästä johtuen sanalla ”gun” tehdyt haut tuottavat paljon aselakeihin liittyttämiä viestejä, vaikka sana olikin aihehallinnuksen edustavin sana. Hakua voidaan tarkentaa loogisilla operaattoreilla esimerkiksi hakemalla viestejä, joissa esiintyy merkkijono ”gun”, mutta ei merkkijonoa ”man”. Haun tarkkuutta voidaan myös parantaa yhdistämällä hakusanoja, niin että haetaan viestejä, joissa esiintyvät merkkijonot ”gun” ja ”law”.



KUVIO 17. Esimerkki Jupyter-työkirjan toiminnasta

Näin Jupyter-työkirjat siis mahdollistivat ympäristön, jossa noin puolesta miljoonasta twiitistä voidaan ajallisesti sekä merkkijonojen perusteella raja reaaliaikaisesti pienempiä osia aineistosta tarkasteluun. Lisäksi työkirjoihin luotiin funktioita, jotka mahdollistivat interaktiivisten visualisaatioiden tekemisen esimerkiksi viestien jakautumisesta aikajanalla. Tämä ympäristö vastaa hyvin pitkälti sitä, mistä tutkielman toisessa luvussa käytettiin termiä kartoittaminen: käyttäjien keskustelu voidaan tarkastella reaaliajassa, vaihdellen mikro- ja makrotasojen välillä.

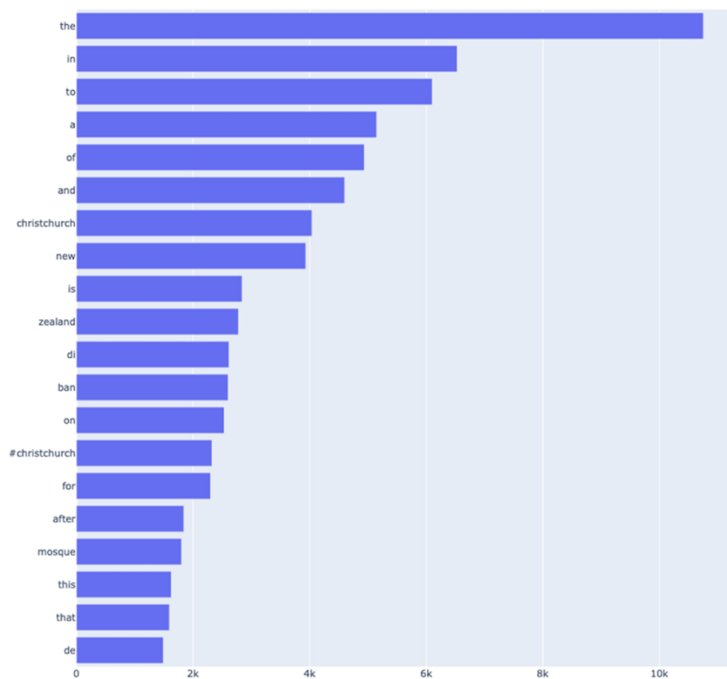
4.3.3 Aselaista käytävän keskustelun etsiminen

Aluksi hakuehdot oli syytä pitää mahdollisimman laajana ja hain aineistosta kaikki viestit, joissa avainsanat esiintyivät. Sanoilla tarkoitetaan tässä merkkijonoja ja esimerkiksi ”ban”-haku sisälsi edellä mainitun BrisBANen sekä sanan pienellä ja isolla kirjoitetut eri muodot. Lisäksi sanojen esiintyvyydet toistensa kanssa ristiintaulukoitiin (AND-operaattori). Yhteensä näistä viesteistä muodostui hieman vajaan kahdenkymmenen neljän tuhat tviitin aineisto, jota pystyi käsittelemään analyysityökaluilla reaaliaikaisesti.

TAULUKKO 4. Twiittien määrä erilaisilla avainsanoilla ja niiden yhdistelmillä

s	Gun	Law	Weapon	Ban
Gun	18035	2886	1128	1635
Law		6265	373	572
Weapon			3514	1554
Ban				13789

Kuten todettua, eivät aihehallinnuksen tulokset täysin vastanneet tämän tutkielman tarpeita. Tästä syystä oli myös tarpeen tarkastella, voisiko aineistossa olla joitain muita mahdollisia avainsanoja, jotka ovat syystä tai toisesta jääneet aihehallinnuksessa huomaamatta. Tätä varten käytin funktiota, joka näyttää aineiston käytetyimmät sanat. Funktio ajettiin jokaisen neljän avainsanan rajaamalla aineistolla erikseen, josta esimerkki kuviossa 18.



KUVIO 18. Käytetyimmät sanat rajauksella ”ban”

Pelkkä käytetyimpien sanojen listaaminen ei kuitenkaan osoittautunut kovinkaan hyödylliseksi, koska listat sisältävät hyvin pitkälti ”täytesanoja” (*stopword*) sekä aineiston keräyksessä ja rajaamisessa käytettyjä sanoja (*collection words*). Englannin kielelle on kuitenkin olemassa *Natural Language Toolkit*-kirjasto, joka sisältää muun muassa listan yleisemmistä täytesanoista ja näin helpottaa niiden poistamista. Funktioon lisättiin toiminnallisuus, joka poistaa sekä täytesanat että keräyksessä käytetyt sanat, jotka täytyi kuitenkin lisätä käsin.

Funktiolla tarkasteltiin myös sanapareja eli sanoja, jotka esiintyvät toistensa kanssa samassa viestissä. Myös tässä tapauksessa poistettiin turhat täytesanat, mutta keräyssanojen poisto on sanaparien kohdalla hieman monimutkaisempaa. Tästä syystä sanalistoille jäi paljon turhia sanapareja, joista esimerkkejä kuviossa 21. Näitä kaikkia ei ollut mielekäästä lähteä käsin poistamaan, vaan jätin ne vain yksinkertaisesti huomiotta listoja läpikäydessä.

Periaatteessa näitä sanahakuja olisi ollut hyödyllistä tehdä myös koko aineistolle, mutta koska poistettavia sanoja on niin paljon, ei tämä ollut tutkimusekonomisesti järkevää. Sanahakujen perusteella ei löytynyt yhtään uutta avainsanaa, joten aineistorajauksia jatkettiin neljällä jo olemassa olevalla avainsanalla.

	bigram	count
0	(new, zealand)	1429
1	(prime, minister)	439
2	(jacinda, ardern)	325
3	(new, zealand's)	272
4	(zealand, prime)	218
5	(mosque, shooting)	192
6	(minister, jacinda)	168
7	(two, mosques)	166
8	(mass, shooting)	142
9	(zealand, pm)	130
10	(new, zealand's)	114
11	(social, media)	108
12	(50, people)	106
13	(terror, attack)	98
14	(terrorist, attack)	95
15	(new, zealand.)	94
16	(10, days)	93
17	(pm, jacinda)	90
18	(new, zealand.)	87
19	(mosque, shootings)	86

KUVIO 21. Useimmiten toistensa kanssa esiintyvät sanat rajauksella ”law”

Hakuja tehtiin kaikilla mahdollisilla neljän avainsanan yhdistelmillä sekä erilaisilla rajauksilla ja sekä aineistoon että sieltä poimituihin linkkeihin tutustuttiin lähiluvun keinoin. Lopulta aineistoa päädyttiin rajaamaan niin, että siihen sisältyvät enää viestit, joissa esiintyy merkkijono ”gun” ja ”law” tai ”ban” ja ”weapon”. Tätä tuki se, että useissa viesteissä ja lehtijutuissa esiintyivät fraasit ”change gun laws” sekä ”ban semiautomatic weapons” eri muodoissa. Toisaalta näin hakuheitojen määrittely ei myöskään ole tarpeettoman monimutkaista ja aineiston määrä on kattava.

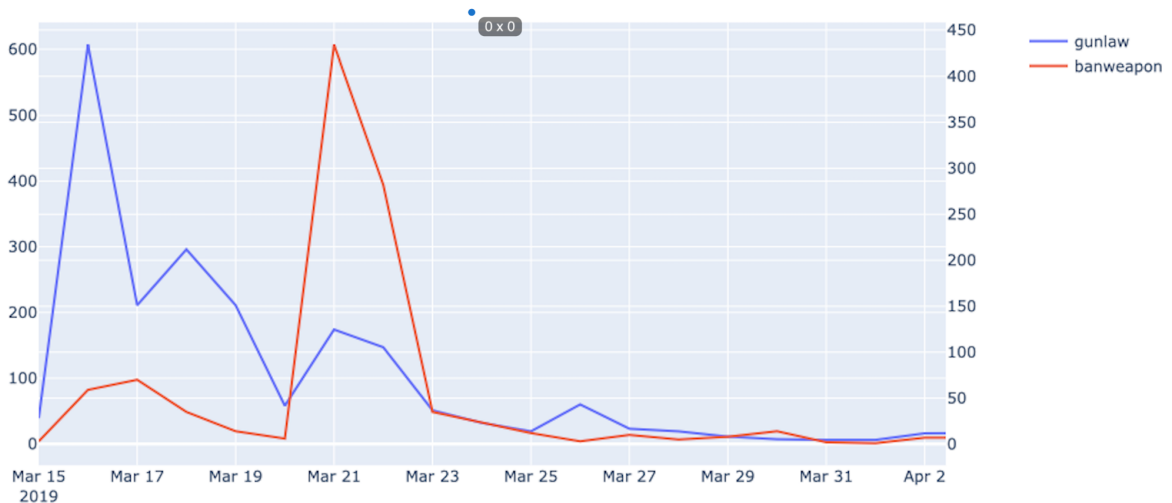
TAULUKKO 5. Lopullinen aineiston rajaus

gun	AND	law	2886
ban	AND	weapon	1554
	OR	Yhteensä	4440

Aineistosta piti vielä poistaa monikerrat eli viestit, joissa esiintyvät kaikki avainsanat. Tämä tehtiin *Pandasin* aggregointifunktiolla käyttäen twiittien yksilöivää ID-arvoa uniikkiuden

tunnuksena. Monikertojen poiston jälkeen aineiston kooksi jäi 4321 twiittiä eli twiittejä, joissa kaikki avainsanat esiintyivät, oli alle kolme prosenttia aineistosta. Tämän voi osaltaan olettaa indikoivan, että rajaus oli onnistunut.

Kuten kuviosta 22 näkee, noudattavat ”ban + weapon” sekä ”gun + law” rajauksin tehdyt aineistot myös täysin erilaista logiikkaa keskustelun volyymin suhteessa aikaan. Tätä voitaneen pitää jonkinasteisena evidenssinä siitä, että aihemallinnuksesta on hyötyä aineiston kartoittamisessa myös siinä tapauksessa, että tutkittava aihe on tiedossa. Aihemallinnuksesta puuttuessa olisi aineistoa mahdollisesti lähdettyä alusta asti rajaamaan vain aselait eli ”gun + law” tyyppisesti (tällöinkin ”ban+weapon” olisi tietty voitu löytää visuaalisen analytiikan/lähiluvun keinoin).



KUVIO 22. Kahden eri rajauksen erot aikasarjalla

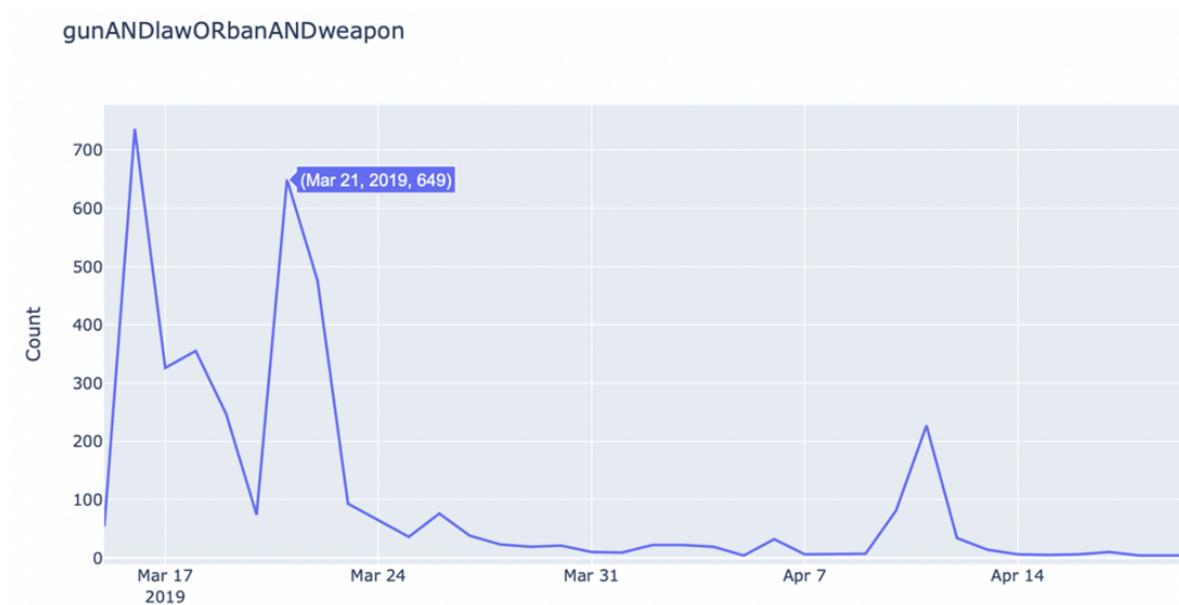
4.4 Löytääkö analyysimenetelmä ”oikean maailman” tapahtumat? Entä kansalaisten mielipiteet ja preferenssit?

4.4.1 Analyttisen narratiivin muodostaminen

Analyttisesti suuntautuneessa narratiivisessa tutkimuksissa tapahtumille luodaan formaali rakenne, jossa tarkastelun kohteena olevan ilmiön kannalta tärkeimmät syiden ja seurausten väliset mekanismit tehdään mahdollisimman näkyväksi ja vähemmät tärkeät rajataan

tarkastelun ulkopuolelle. (Saari 2011b, 72–73) Tässä mielessä tuhansien twiittien aineisto sisälsi vielä paljon toistoa ja vähemmän tärkeitä detaljeja aselaeista, jotka myös hankaloittivat turhaan visuaalisia analyyssejä. Tästä syystä aineisto rajattiin tässä vaiheessa sisältämään vain viestit ensimmäisen kuukauden ajalta, jolloin tärkeimmät aselakeja koskevat tapahtumat tietyvästi olivat.

Plotly-kirjastolla voidaan tehdä interaktiivisia visualisaatioita, joiden etuihin kuuluu, että kuvaajaa pystyy helposti zoomaamaan valittuun ajankohtaan, jolloin pienemmätkin erot kuvaajalla näkyvät tarkemmin. Lisäksi kuvaajan pistettä klikatessa näkee tarkan arvon sekä ajalla että viestien määrälle. Kun tähän toiminnallisuuteen yhdistetään funktio, jolla aineistoa pystytään rajaamaan ajan mukaan, voidaan aineistolle tehdä hyvinkin tarkkaa visuaalista analyysia.



KUVIO 23. Lähetettyjen twiittien määrä suhteessa aikaan ensimmäiseltä kolmeltakymmeneltä päivältä.

Kun kuukautta katsotaan kokonaisuutena, näyttää keskustelun määrä suhteellisen vähäiseltä maaliskuun vaihteesta. Toisaalta tälläkin ajanjaksolla on havaittavissa selkeää päiväkohtaista vaihtelua. Ilmeinen keino rajata aineistoa olisikin keskittyä kuvaajilla näkyviin piikkeihin, eli niihin päiviin, jolloin twiittejä on lähetetty erityisen paljon. Sanomalehtiartikkelien suhteen näin on toimittukin ja näitä piikkejä on tulkittu keskustelun tiheytyksiksi (Vainikka ym. 2020).

Tällainen lähestymistapa olisi kuitenkin kyseenalainen, koska kuten aiemmin todettiin, Twitter toimii ”internet-ajassa”. Aika on tietysti aidosti jatkuva muuttuja, mutta sen pilkkominen päiviksi on osin mielivaltaista. On esimerkiksi aivan mahdollista, että jotain politiikkaprosessin kannalta merkittävää on tapahtunut Uudessa-Seelannissa 20. päivä ja yllä olevan kuvaajan piikki 21. päivä kertoo ainoastaan siitä, että tällöin tapahtumaan on herätty esimerkiksi Yhdysvalloissa, jossa Twitter-käyttäjiä on oletettavasti huomattavasti enemmän.

Näin ollen on nähdäkseni tarkoituksenmukaista ottaa tarkasteluun twiittejä kaikilta päiviltä ja valikoin lopulta jokaiselta päivältä viisi eniten uudelleentwiittauksia kerännyttä twiittiä lähilukuun ja analyttisen narratiivin luomiseen. Tämä on sama rajaus, jota Rogers käyttää oppikirjaesimerkkinä Twitterin käyttämisestä analyttisenä tarinankerrontakoneena (2019, 167). Yhtenä päivänä kolmestakymmenestä viestejä ei ollut lähetetty kuin 4, joten analyttisen narratiivin muodostavien twiittien määräksi tuli lopulta 149.

Edellä mainitussa Iranin vaalien oppikirjaesimerkissä etäanalyysin kohteena oleva tapahtumaketju oli määritetty aiheutunnisteen perusteella ja näin ollen kaikki twiitit olivat määritelmällisesti osa tätä tapahtumaketjua ja myös analyttista narratiivia. Tässä tutkielmassa samoilla perusteilla määritettyä analyttista narratiivia taas käytetään sen todentamiseen, onko käyttämäni aiheanoihin perustuva rajaus toimiva ja löytyvätkö aselakien säätämiseen liittyvät tapahtumat aineistosta.

Näin ollen analyttisen narratiivin luomiseksi täytyi ensin järjestää tarkastelun kohteena olevat 149 twiittiä aikajärjestykseen ja tarkastella, onko joukossa twiittejä, jotka eivät liity tapahtumaketjuun. Näitä twiittejä löytyi ainoastaan kahdeksan, joten näiltä osin rajausta voidaan pitää onnistuneena. Painotin twiittien määrää vielä uudelleentwiittausten määrällä, jotta nähtäisiin, miten rajaus käyttäytyisi laajemmalla aineistolla. Tällöin aselakeihin liittymättömän keskustelun osuus oli vieläkin pienempi.

TAULUKKO 6. Aselakeihin liittyvien twiittien osuus

	Ei liity aselakeihin	Liittyy aselakeihin	Yhteensä
Määrä	8	141	149
Määrä x RT	779	193 586	194 365
Osuus	5,37 %	94,63 %	100%
Osuus x RT	0,49 %	99,60 %	100 %

Aselakeihin liittymättömät viestit koostuivat lähinnä viesteistä, joissa rajaamisessa käytetyt merkkijonot esiintyivät osana jotain toista sanaa. Tältä olisi voitu mahdollisesti välttyä hakemalla sanoja eksaktissa muodossa, mutta tämä olisi toisaalta voinut pudottaa aineistosta jotain oleellista pois. Yksi viesti myös tarttui mukaan ”vahingossa”, kun siinä käytettiin sanaa *ban* eri yhteydessä, mutta silti se sisälsi aselakeihin liittyneen tapahtuman.

TAULUKKO 7. Vahingossa mukaan tullut asiaankuuluva twiitti

Independent 25.3.2019 0:48:43
New Zealand bans ‘manifesto’ of Christchurch mosque gunman as more than 1,000 hand over weapons https://t.co/yUxOcF5Us4

TAULUKKO 8. Asiaan kuulumaton twiitti (*Bannon + weaponry*)

Alexjrowell 15.3.2019 23:06:25
#Christchurch killer referred to 1683 Battle of Vienna and 732 Battle of Tours on his weaponry. In 2014, Steve Bannon cited both as proud moments in "the long history of the Judeo-Christian West struggle against Islam" https://t.co/2tEYZp7aJB https://t.co/X7yO6n4nNp

Aselakeihin liittymättömien viestien poistamisen jälkeen jäljelle jääneet 141 viestiä muodostavat siis analyttisen narratiivin aselakien säätämisen politiikkaprosessista. Tässä ei ole tarpeen käsitellä jokaista näistä viesteistä, vaan lähinnä tarkastella löytyvätkö sieltä ne tapahtumat, jotka lainsäädäntöprosessista esiteltiin aiemmin tiedettävän. Taulukossa 9 näkyvät 10 edustavinta viestiä, jotka joko sisältävät tapahtuman tai oleellista informaatiota aselakiprozessista.

TAULUKKO 9. Analyttinen narratiivi

Lähettäjä	Aika	Viesti	Selitys
spectatorindex	16.3.2019 9:40:22	BREAKING: Prime Minister of New Zealand says gun laws will be reformed following Christchurch attacks	Ilmoitus, että aselakeja muutetaan
shannonrwatts	16.3.2019 17:11:34	New Zealand bans semiautomatic rifles less than 24 hours after the Christchurch mass shooting. Imagine: elected officials putting public safety over gun manufacturers' profits. Americans deserve better than lawmakers who are letting gun lobbyists write our nation's gun laws. https://t.co/TinWP7PSfM	Tieto siitä, että aselakien muuttaminen tarkoittaa ”semiautomatic rifles” kieltämistä
BBCBreaking	18.3.2019 17:23:20	PM Jacinda Ardern says New Zealand's cabinet backs gun law changes 'in principle' after Christchurch shootings https://t.co/I1RH7KU5iC	Hallitus on päässyt sopuun aselakien muuttamisesta

ajplus	19.3.2019 1:44:35	<p>New Zealand PM Jacinda Ardern said she'll announce new gun laws within 10 days after a gunman killed 50 people at two mosques in Christchurch.</p> <p>"To make our community safer the time to act is now."</p> <p>https://t.co/Eb9xBN3DLm</p>	Uudet aselait tullaan esittelemään nopeasti
AJEnglish	22.3.2019 3:28:35	<p>New Zealand's PM Jacinda Ardern bans the sale of assault weapons following the Christchurch mosque attack. The ban includes:</p> <ul style="list-style-type: none"> • all military-style semi-automatics (MSSAs) • related parts used to convert guns into MSSAs • high-capacity magazines <p>https://t.co/GUqEk9gmjt</p>	Aseiden myynti kielletään/lain tarkempi sisältö selviää
NPR	2.4.2019 2:43:00	<p>"This will mean that within four weeks of the Christchurch terrorist attack, New Zealand will have passed legislation banning all military-style semi-automatic weapons and assault rifles," deputy Prime Minister Winston Peters told reporters on Monday.</p> <p>https://t.co/IOUosPJQFc</p>	Ainut asia, jota viesteistä ei suoranaisesti selvinnyt oli "Bill Introduced". Tästä sen voi kuitenkin päätellä.

ChinaDaily	3.4.2019 1:45:09	A new gun law following the #Christchurch shootings on March 15 has passed its first reading in New Zealand's parliament on Tuesday. #GunControl #NewZealand https://t.co/0oGVbux31p https://t.co/HwpReHOEkG	1st reading
racephenom	7.4.2019 23:54:23	New Zealand introduces bill to ban weapons used in Christchurch. New Zealand introduces gun laws that will ban 'military-style' weapons and give owners until October to hand in outlawed firearms https://t.co/PS6J8xfLb6 via @MailOnline	Tieto siitä, että laki tulee myynnin kieltämisen lisäksi kieltämään aseiden hallussapidon.
NZStuffPolitics	9.4.2019 18:55:33	Christchurch mosque attacks: First new gun law passes second reading https://t.co/TamCTAjYPD	2nd reading
abcnews	10.4.2019 22:51:26	New Zealand gun laws pass 119-1 after Christchurch mosque shootings https://t.co/ZYuKKGcB21	3 rd reading
kylegriffin1	13.4.2019 1:00:21	New Zealand's governor general has formally signed into effect sweeping gun laws outlawing military style weapons, less than a month after the Christchurch massacre. https://t.co/xANjUKsVGH	Royal Assent

Analyttinen narratiivi osoittaa, että käyttämäni menetelmä löysi kaikki tunnetut policy-prosessin vaiheet ja lisäksi antoi oleellista lisätietoa esimerkiksi aselakien sisällöstä. Tästä voidaan päätellä, että Twitter toimii suhteellisen luotettavasti proxyna policy-prosesseihin,

ainakin mitä tulee tapahtumien tunnistamiseen. Kansalaisten mielipiteiden tunnistaminen onkin jo huomattavasti hankalampaa.

4.4.3 Kansalaiset ja mielipiteet

Sosiaalisen median monitoroinnin ajatuksena kuunnella kansalaisten mielipiteitä, joten seuraavaksi oli selvitettävä, 1) Kuinka monessa viestissä on mielipiteitä 2) Kuinka suuri osa mielipiteiden lähettäjistä on Uuden-Seelannin kansalaisia. Twiitit luokiteltiin aluksi sen mukaan, onko niissä mielipidettä aselakien muuttamisesta (81 twiittiä) vai ei (60 twiittiä). Jälleen twiittejä painotettiin myös uudelleentwiittausten määrällä, jotta nähdään, missä määrin tulokset pätsivät laajempaan aineistoon.

TAULUKKO 10.

	Ei mielipidettä tai preferenssiä	Mielipide tai preferenssi	Yhteensä
Määrä	60	81	141
Määrä x RT	85 345	108 241	193 586
Osuus	42,55 %	57,45 %	100 %
Osuus x RT	44,09 %	55,91 %	100 %

On huomattava, että luokka ”Ei mielipidettä tai preferenssiä” sisältää mielipiteitä aselaeista, mutta ei välttämättä ainoastaan Uuden-Seelannin aselaeista. Koska viestien kielenä oli lähtökohtaisesti englanti, sisälsivät ne oletetusti myös paljon viittauksia muiden angloamerikkalaisten maiden politiikkaan. Näiden viestien määrä oli kuitenkin yllättävän suuri ja maantieteellistä moninaisuuttakin löytyi: keskustelua käytiin niin Yhdysvaltojen, Iso-Britannian, Kanadan kuin Intiankin aselainsäädännöstä.

TAULUKKO 11. Esimerkki Australian aselakeja käsittelevästä twiitistä

RLBraithwaite 1.4.2019 12:25:37

@LiberalAus @billshortenmp A gun shot. Less than three weeks since an Australian carried out #christchurch massacre. Less than a week since we learnt some LNP members will preference a party willing to trade off our strong gun laws for some filthy lucre. Wtf is wrong with you people? #auspol #AusVotes2019

Viestien sisällön perusteella on kuitenkin useissa tapauksissa haastavaa päätellä, käsittelevätkö ne nimenomaan Uuden-Seelannin politiikkaa. Tämä vaatii samalla sekä ymmärrystä kansainvälisestä politiikasta että Twitterin viestintäkäytännöistä. Esimerkiksi aiheutunnistetta #auspol viitataan yleisesti Australian politiikkaan ja viestit, jotka sisältävät tämän aiheutunnisteen, oli yksinkertaista rajata tarkastelun ulkopuolelle (Bogle 2016). Monissa muissa tapauksissa rajausta oli huomattavasti haastavampaa, ja jos viestien sisällön perusteella ei ollut täysin selvää, että viitataan jonkin muun maan lainsäädäntöön, luokiteltiin se sisältämään mielipide.

TAULUKKO 12. Esimerkki lähes identtisistä twiiteistä, jotka mediayhtiö jakoi kahdella eri tilillään.

cnni	29.3.2019 23:36:36	More than 11,000 people have signed a petition against New Zealand's proposed tougher gun laws in the wake of the Christchurch mosque massacres https://t.co/boDcWACo2U
CNN	30.3.2019 0:50:08	More than 11,000 people have signed a petition against proposed tougher gun laws in the wake of the Christchurch mosque massacres, criticizing the changes as "unjust to law-abiding New Zealand citizens." https://t.co/VnckG2vknN

Huomattavan suuri osa twiittien lähettäjäistä ei myöskään ollut sen kummemmin Uuden-Seelannin kuin minkään muun maan kansalaisia, vaan mediayhtiöiden tilejä. Yhtiöillä saattoi myös olla useita tilejä, joilla ne twiittasivat suunnilleen samaa sisältöä, kuten

taulukosta 12 näkee. Vaikkakin mediayhtiöiden tilit rajautuivat tässä tarkastelusta pois, on huomattavaa, että niilläkin saatettiin kertoa tietoa kansalaisten mielipiteistä, vaikkakaan ei siinä mielessä kuin policy-analytiikassa tarkoitetaan.

Ylipäättään aineistossa on selvästi ylliedustettuna sekä mediayhtiöiden että angloamerikkalaisten julkisuuden henkilöiden tilit. Tähän vaikuttaa osaltaan oletettavasti se, että aineiston rajaamisessa on käytetty uudelleentwiittaamista parametrina, mutta kuten aiemmin todettua, uudelleentwiittaukset ovat myös twiittejä itsessään. Hankalampi kysymys on, missä määrin nämä ovat mielipiteitä. Jos Uusi-Seelantilainen uudelleentwiittaa yhdysvaltalaisen presidenttiehdokkaan twiittiä Yhdysvaltojen aselaista, voidaanko tämä tulkita mielipiteeksi Uuden-Seelannin aselaista?

TAULUKKO 14. Esimerkki twiitistä, josta voisi päätellä lähettäjän olevan Uuden-Seelannin kansalainen

theCarlosMiller 1.4.2019 7:14:44

Gang bosses say weapons won't be handed back after Christchurch mosque shooting.

Maybe Prime Minister @jacindaardern should dress up their gang colors and tattoos and go try to convince them to turn in their guns.

#NewZealand #GunLaws <https://t.co/qe5FfyqzeJ>

Sama problematiikka toistuu myös kansalaisia tunnistaessa ja tekee kansalaisten tunnistamisesta lähtökohtaisesti lähes mahdotonta. Vaikka viestin sisällöstä voisi päätellä lähettäjän olevan Uuden-Seelannin kansalainen, vaatii sen varmistaminen niin verran paljon salapoliisityötä, ettei tätä voida pitää policy-analytiikan automaatiotavoitteiden kannalta tarkoituksenmukaisena. Esimerkiksi taulukossa 14 esitetystä twiitistä voisi päätellä lähettäjän olevan nimenomaan Uuden-Seelannin kansalainen, mutta profiilisivullaan hän itseasiassa kertoo olevansa Yhdysvaltalainen.

Tietysti käyttäjän omakin ilmoitus kansalaisuudestaan ei välttämättä ole totta, vaikka vaikuttaisi epätodennäköiseltä, että käyttäjät järjestelmällisesti ilmoittaisivat väärän kansalaisuuden. On huomionarvoista, että Twitter myös mahdollistaa käyttäjätilin verifioimisen, mutta tämäkään ei koske tavallisia kansalaisia, vaan saadakseen verifioinnin

pitää tilin olla merkittävä (*notable*) (Twitter 2021). Näin ollen, vaikka Twitter toimikin erinomaisesti proxyna policy-prosessin tapahtumiin, on kansalaisten mielipiteiden monitoroiminen sieltä lähestulkoon mahdotonta.

6. JOHTOPÄÄTÖKSET JA POHDINTA

6.1 Johtopäätökset

Tämän pro gradu -tutkielman tavoitteena oli selvittää, kuinka sosiaalisen median monitorointia voidaan käytännössä toteuttaa ja kuinka se voi auttaa policy-prosessien ymmärtämisessä. Tätä varten luvussa kaksi luotiin typologia, joka mahdollisti policy-analytiikan syklisen prosessin analyttisen tarkastelun. Luvussa kolme todettiin, että vaikka policy-analytiikan konsepti kehitettiin vahvasti policy-analyysin positivistisen suuntauksen parissa, ainakin sosiaalisen median monitoroinnin käytännön toteuttamiseen tarvittiin työkaluja myös tulkinnallisemmilta policy-analyysin suuntauksilta. Näin saatiin luotua sosiaalisen median monitoroinnista malli, jonka toimivuutta pystyttiin käytännössä testaamaan. Malli oli kuitenkin edelleen liian laaja, jotta sosiaalisen median teknisiä nyansseja voitaisiin ottaa huomioon, joten se piti fokusoida spesifisti Twitteriin. Twitteriin fokusoidulla taas luotiin muistilista, joka puolestaan muodosti pohjan tutkielman empiiriselle tarkastelulle.

Lähtökohtana tutkielmassa oli, että ideaalitulanteessa kaikkiin kysymyksiin kyettäisiin vastaamaan empiirisesti, mutta tähän pystyttiin vain osittain. Christchurchin terrori-iskuista kerätty lähes 12 miljoonan twiitin aineisto oli kuitenkin poikkeuksellisen laaja ja näin ollen mahdollisti empiirisen tarkastelun esimerkiksi aineiston rajaamisen osalta, joka ei ole usein mahdollista. Usein aineistot kerätään ennalta määrätyillä hakusanoilla, mutta koska tämän tutkielman varsinainen aineisto rajattiin suuremmasta aineistosta, pystyttiin esimerkiksi erilaisten avainsanojen vaikutusta rajaamiseen tarkastelemaan huomattavan yksityiskohtaisesti. Toisaalta aineistolla aiemmin tehty tutkimus, erityisesti aihehallinnus, osoittautui myös erittäin tärkeäksi.

Tutkielmassa saatiinkin useampi empiirinen tulos aikaan, joita voitaneen myös pitää suhteellisen merkittävänä. Ne voidaan tiivistää neljään kohtaan:

- 1) Aihemallinnuksesta on apua suuren aineiston kanssa myös silloin, kun tutkimuksen kohde on etukäteen tiedossa

- 2) Avainsanojen valinta vaikuttaa merkittäväällä tavalla Twitteristä tehtäviin aineistorajauksiin
- 3) Twitter toimii hyvin policy-prosessin tapahtumien tunnistamisessa
- 4) Twitter-keskustelusta on lähes mahdotonta erottaa kansalaisten mielipiteitä

Se, missä määrin tulokset ovat yleistettäviä on kysymys erikseen. Kuten toisessa luvussa todettiin, tämä hankala kysymys tulee eteen yleisesti digitaalisten datalähteiden kanssa. Lisäksi digitaalisten jälkien suhteen joudutaan aina pohtimaan, milloin on kyse mediumin vaikutuksista ja milloin viesti kertoo jotain yleisempää. Tähän hankalaan ontologiseen kysymykseen ei pystytä tässäkään tutkielmassa vastaamaan, mutta joka tapauksessa tutkielman tuloksista on vedettävissä jonkinlaisia suuntaviivoja.

Aihemallinnusta ei käsitelty tässä tutkielmassa kovinkaan tarkasti, mutta ohjaamattomana koneoppimismenetelmänä sitä on luontevaa käyttää silloin, kun ei ole etukäteen tiedossa, mihin aineistossa halutaan huomio kohdentaa (Toivanen ym. 2020). Tässä tutkielmassa aihemallinnuksesta havaittiin olevan hyötyä, vaikka huomion kohdentamisen kohde olikin etukäteen tiedossa. Tarkastelu kuitenkin oli sen verran pintapuolista, että tästä ei voi vetää kovinkaan pitkälle meneviä johtopäätöksiä. Kuitenkin tutkielma antaa osviittaa siihen suuntaan, että aihemallinnusta kannattaa mahdollisuuksien mukaan käyttää myös tunnettujen aineistojen kanssa, ainakin apuvälineenä.

Se, että sopivien avainsanojen valinta vaikuttaa hyvin selvästi aineiston rakentumiseen, on nähdäkseni relevantti tutkimustulos kaiken Twitteriin pohjautuvan tutkimuksen kannalta. Hyvin useissa tutkimuksissa aineistoa vain rajataan joidenkin hakusanojen perusteella, kuten tämän tutkielman varsinaisen aineiston suhteenkin tehtiin. Kuten todettua, tämä on usein teknologisten realiteettisen sanelemaa. Tässä tutkielmassa kahdella näennäisen samanlaisella rajauksella saatiin kuitenkin niin merkittävän erilaisia tuloksia, että vertailevalla tutkimukselle aineiston rajaamisen eri vaiheista, olisi selvästi paljon tarvetta.

Tässä tutkielman ei ollut enää mahdollisuuksia vertailla, minkälaisia tapahtumia tai kansalaisten mielipiteitä muut mahdolliset aineistorajaukset olisivat tuottaneet. Kysymystä ei kuitenkaan voi jättää huomiotta. Olisiko analyttinen narratiivi löytänyt oikeat tapahtumat, jos aineistoa olisi rajattu ainoastaan avainsanoilla ”gun+law”? Vaikka tähän ei pystytä vastaamaan, on huomionarvoista, että tämä kysymys myös alleviivaa ensimmäistä

kertaa luvussa kaksi esitettyä pyramidin mallia. Juuri tästä syystä datankäsittelyn ensimmäisiin vaiheisiin, kuten aineiston rajaukseen, tulisi käyttää erityistä tarkkuutta.

Tuloksien kolme ja neljä yleistettävyyttä ei muutenkaan ole niin vahvaa, mutta sosiaalisen median monitoroinnin, ja oikeastaan kaiken Twitter-dataa käyttävän policy- tai politiikka-analyysin kannalta niitä voidaan pitää keskeisinä. Tulokset kertovat selvästi, että Twitter toimii erinomaisesti policy-prosessin tapahtumarakenteen muodostamisessa, mutta ei niinkään varsinaisen kansalaiskeskustelun seuraamisessa. Kuten todettua, on politiikkaprosessien analyysia perinteisesti pidetty raskaana verrattuna muuttujakeskeisiin tutkimusotteeseen ja digitaalisten aineistojen sekä menetelmien käyttö voisi tehdä tällaisen tutkimuksen tekemisestä huomattavasti helpompaa.

Vaikka tässä tutkielmassa ei saatukaan analyttisesti eroteltua kansalaisten mielipiteitä twiiteistä, ei tämänkään tutkimussuunnan suhteen kannata heittää kirvestä kiveen. On hyvin mahdollista, että tämä johtui uudelleentwiittauksien käytöstä aineistonrajauksessa ja kuten aiemmin totesin, aineistonrajauksen vaikutukset ovat huomattavia ja oletettavasti kertautuvat, kun analyttisten vaiheiden pyramidissa siirrytään ylöspäin. On kuitenkin painotettava, että uudelleentwiittauksien käyttö johtui hyvin pitkälti Twitterin teknisistä rajoitteista ja samat rajoitteet koskevat muutakin tutkimusta.

Longo ja Mcnut määrittivät policy-analytiikan olevan syklinen prosessi, josta tässä pystyttiin käymään läpi ainoastaan tietynlainen tuokiokuva. Vaikka syklisen prosessin luonnetta ei kuvattu tarkemmin, sillä luultavimmin tarkoitetaan, että nimenomaan tällaisista tuloksista opitaan ja prosessia muokataan esimerkiksi niin, että uudelleentwiittauksia ei enää käytetä aineistorajauksiin. Tässä tutkielmassa ei kuitenkaan ollut mahdollisuuksia kuin yhteen iteraatioon.

Toisaalta tutkimustuloksista voidaan joka tapauksessa päätellä, ettei kansalaisten mielipiteiden tunnistaminen ole erityisen helppoa. Tämän takia on syytä suhtautua kriittisesti esimerkiksi johdannossa mainittuihin ministeriöiden projekteihin, joissa sosiaalista mediaa monitoroidaan. Kaiken sen tiedon pohjalta, mitä tämän tutkielman kirjoittajalla erilaisista tekoälyä tavoittelevista algoritmeista on, eivät nekkään olisi pystyneet tunnistamaan kansalaisten mielipiteitä vastaavasta aineistosta. Toisaalta monet tässä tutkimuksessa

kansalaisten mielipiteiden tunnistamiseen liittyvistä ongelmista johtuivat englannin kielestä ja näin ollen on mahdollista, että suomeksi vastaava prosessi olisi huomattavasti helpompi.

Yleisesti tämän tutkimuksen lähestymistapa oli hyvin käytännönläheinen ja painotti datankäsittelyn käytäntöjä silläkin uhalla, että joudutaan keskittymään nyansseihin, joita yhteiskuntatieteellisessä tutkimuksessa ei ehkä perinteisesti olla painotettu. Lähestymistapaa voidaan kuitenkin sinällään pitää onnistuneena, koska nimenomaan näennäisistä nyansseista nousivat edellä esitetyt empiiriset tulokset.

6.2 Pohdinta ja päätäntö

Tämä pro gradu -tutkielma alkoi tarkastelulla digitaalisen datan määrän kasvusta yleisesti. Vaikka tämä tarkastelu jäikin lähinnä anekdoottien tasolle, on se silti tämän tutkielman oleellisimpia osia. Kun yhä suurempi määrä yhteiskunnasta toimii digitaalisten pelisääntöjen mukaan, on myös tutkimus jossain määrin pakotettu seuraamaan tätä logiikkaa. Digitaalista voi tietysti tutkia myös analogisesti, mutta datan suuret määrät lähes väistämättä myös edellyttävät uudenlaisia tutkimusmenetelmiä. Binäärejä voi tietysti tulostaa paperille, mutta tällainen tutkimus olisi yksinkertaisesti niin tehotonta, että sille on hankala nähdä mitään yhteiskunnallista relevanssia.

Kun käsite ”laskennallinen yhteiskuntatiede” lanseerattiin vuonna 2009 (Lazer ym.), esitettiin samalla huomio, että laskennallista yhteiskuntatiedettä harjoitettiin Googlen tapaisissa internet-yrityksissä ja valtiollisten turvallisuuselinten toimesta. Akatemian puolella se oli harvinaisempaa ja laskennallinen yhteiskuntatiede loisti poissaolollaan niin taloustieteen, sosiologian kuin politiikan tutkimuksen julkaisuista. Alaa harjoitetaan edelleen myös yksityisellä puolella ja tätä nykyä myös kuluttajatutkimukseen erikoistuneet yritykset ja teleoperaattorit tekevät yhteiskunnallisia analyyseja hyvin kehittyneillä menetelmillä. Tästä syystä on esitetty suurien datamassojen analysoinnin johtavan siihen, että nykyään kuka vaan voi halutessaan olla yhteiskuntatieteilijä, koulutustaustasta riippumatta. Tämän kehityksen on argumentoitu mahdollisesti johtavan kriisiin akateemisilla yhteiskuntatieteen aloilla ja on herännyt kysymys siitä, tarvitaanko yhteiskuntatieteellistä *teoriaa* ylipäättään yhteiskunnan ymmärtämiseksi. (Savage ja Burrows 2007, 2009, Burrows ja Savage 2014.)

On totta, että laskennallisten menetelmien esiinmarssi on mahdollistanut sosiaalisten ilmiöiden tutkimisen ilman linkkiä yhteiskuntatieteen perinteiseen teoriapohjaan. Laskennallisen yhteiskuntatieteen klassikotutkimuksessa Adamic ja Glance tarkastelivat, kuinka konservatiiviset ja liberaalit blogikirjoittajat linkittävät teksteissään toistensa blogeja. Tutkimuksessa ei kuitenkaan viitata kuin muutamassa kohtaa yhteiskuntatieteelliseen tutkimukseen eikä käytetä esimerkiksi termiä ”polarisaatio” lainkaan. (Adamic ja Glance 2005.) Nelimarkka (2021) on kuitenkin painottanut, että perinteinen yhteiskuntatieteellinen koulutus antaa esimerkiksi kykyjä reflektoida niitä sosiaalisia prosesseja, joissa dataa tuotetaan. Kuten tässäkin tutkielmassa on useampaan kertaan käynyt ilmi, tämä reflektio on erittäin olennaista.

Jo tässä tutkielmassa pystyttiin osoittamaan lukuisia ongelmallisia vaiheita koskien sosiaalisen median monitorointia. Tästä huolimatta näitä menetelmiä käytetään yhä enenevässä määrin sekä yksityisellä että julkisella sektorilla. Etenkin julkisen sektorin kohdalla tulee pitää huolestuttavana myös sitä, että ei olla kerrottu, mihin tätä tietoa itseasiassa käytetään. Koronakriisin aikaan liikkeiden seuranta on perusteltu sillä, että seurataan liikkumisrajoitusten vaikutusta. Sen sijaan ei ole käynyt selväksi, missä määrin koronaan liittyvää sosiaalisen median monitorointia aiotaan käyttää (tai on jo käytetty) poliittisen päätöksenteon pohjana.

Jos yhteiskunnallisista kysymyksistä tehdään päätöksiä ainoastaan datan perusteella, on vaarana, että demokratian ja hyvinvointivaltion tapaiset periaatteet korvautuvat ”Piilaakson logiikalla”. Pahimmassa tapauksessa tämä johtaa siihen, että ”data ilman politiikkaa” ohjaa kaikkia päätöksiä terveydenhuollosta rikosten torjuntaan. (Morozov 2014, Flyverbomin ja Madsenin 2015, 143 mukaan.) Tämä on mielestäni tärkein perustelu sille, miksi yhteiskuntatieteellisen tutkimuksenkin olisi yhä enenevässä määrin fokusoitava huomionsa konkreettisiin datankäsittelyn käytäntöihin.

On myös esitetty, että uudenlaiset datan lähteet sovelluksineen aiheuttavat empirisen sosiologian kriisin, polarisoivat ja jakavat tutkijoita sekä johtavat erilaisiin tutkimussuuntauksiin, jotka eivät enää kykene kommunikoimaan keskenään (Burrows ja Savage 2014). Jos johdannossa esitetystä erilaisen terminologian määrästä voidaan jotain päätellä, uhka tällaisesta kriisistä on todellinen. Kuitenkin terminologian ja konseptien

takana datankäsittelyn käytännöt ovat hyvin pitkälti samat, ja tämä onkin toinen perustelu sille, miksi näihin käytäntöihin olisi keskityttävä.

On myös huomattava, että tässäkin tutkielmassa käsiteltiin hyvin pitkälti vain tekstimuodossa olevaa dataa, mutta yhä suurempi osa myös yhteiskunnallisesti relevanteista informaatiovirroista välitetään kuvien ja videoiden kautta. Siinä missä tekstitiedostot ovat vielä suhteellisen kevyitä analysoida, ovat videot ja kuvat huomattavasti raskaampia. Kuvan on sanottu vastaavan tuhatta sanaa, mutta entä kun videossa on kuusikymmentä kuvaa sekunnissa?

On myös hyvin mahdollista, että perinteiset julkaisukäytännöt tekevät digitaalisen tutkimuksen teosta tarpeettoman hankalaa. Tässäkin tutkielmassa tasapainoiltiin jatkuvasti sen ongelman kanssa, että policy-analytiikka oli määritelmällisesti päättymätön prosessi, josta piti esitellä jonkinlainen tuokiokuva. Ongelma ei kuitenkaan ole luonteeltaan ainoastaan määritelmällinen, vaan hyvin konkreettinen.

Maisterintutkielman ulkoasun ohjeistuksessa, jota tässäkin tutkielmassa noudatettiin, pyydetään erottelemaan kuviot ja taulukot. Tämän tutkielman analyysien suhteen tätä erontekoa voi pitää hieman kummallisena. Tutkielmassa käytetty aineisto oli csv-tiedostomuodossa eli taulukko, mutta sitä analysoitiin hyvin pitkälti visualisoimalla taulukosta erilaisia aggregointeja, jotka määritelmällisesti ovat kuvioita. Luvussa viisi pystyttiin esittelemään tästä analyysista vain murto-osa tai aiempaa ilmaisuani käyttäen, tuokiokuva. Näitä analyysisyklejä tehtiin tutkielman teon aikana lukemattomia määriä, koska käytetyt työkalut mahdollistivat analyysin sekä analyysin parametrien muuttamisen reaaliajassa.

Jos ja kun digitaalisia tutkimusta tulevat jatkossakin määrittämään reaaliaikaisuus ja sykliisyys, on tarpeen myös pohtia, ovatko lineaarisesti etenevät ja pdf-muotoon pakotetut tekstikokonaisuudet se tapa, jolla tällaista tutkimusta tulisi julkaista. Tämän tutkielman teoriaosuuden alussa puhuttiin digitoinnista ja lineaariset tekstikokonaisuudet ovat nimenomaan osoitus tästä: yritys pitää analogisesta toimintalogiikasta kiinni myös silloin, kun tutkimusprosessit ovat alusta loppuun digitaalisia. Onkin mielenkiintoista pohtia, millaista diginatiivi tutkimus voisi olla.

7. LÄHTEET

Aalto University. 2020. Master's Programme in Computer, Communication and Information Sciences – Human-Computer Interaction
<https://www.aalto.fi/en/study-options/masters-programme-in-computer-communication-and-information-sciences-human-computer> Viitattu 12.3.2020

Adamic, Lada A., ja Natalie Glance. 2005. The political blogosphere and the 2004 U.S. election: divided they blog.
https://edisciplinas.usp.br/pluginfile.php/3333574/mod_resource/content/2/p36-adamic.pdf
viitattu 20.1.2020

Algoritmien valta. 2017. Algoritmien valta - Neutraalius ja puolueellisuus koneellisessa päätöksenteossa
<https://algoritmitutkimus.fi/tutkimushanke/>. Viitattu 22.5.2021

Alinejad, Donya, Laura Candidatu, Melis Mevsimler, Claudia Minchilli, Sandra Ponzanesi, ja Fernando N. Van Der Vlist. 2019. Diaspora and Mapping Methodologies: Tracing Transnational Digital Connections with 'Mattering Maps'. *Global Networks* 19 (1): 21–43.

Bamberg, Jaakko. 2014. Kaavoitukseen osallistaminen sosiomateriaalisena käytäntönä. Teoksessa *Tulkinnan mahti: johdatus tulkitsevaan politiikka-analyysiin*. Tampere: Tampere University Press 2014.

Bogle, Ariel. 2016. #auspol: The Twitter Hashtag Australia Can't Live Without. *Mashable*.
<https://mashable.com/2016/03/21/twitter-australia-auspol/>. Viitattu 15.4.2021.

Bookman, Samantha. 2017. 15 Huge Supercomputers That Were Less Powerful Than Your Smartphone.
<https://www.theclever.com/15-huge-supercomputers-that-were-less-powerful-than-your-smartphone/>. Viitattu 20.8.2020.

Bosch, Oriol J., Melanie Revilla, ja Ezequiel Paura. 2018. Answering Mobile Surveys With Images: An Exploration Using a Computer Vision API: Social Science Computer Review 37(5).

Buchanan, Kelly. 2019. New Zealand: Legislation Banning Certain Firearms, Magazines, and Parts Passed.

www.loc.gov/law/foreign-news/article/new-zealand-legislation-banning-certain-firearms-magazines-and-parts-passed/. Viitattu 30.3.2020.

Burrows, Roger, ja Mike Savage. 2014. After the Crisis? Big Data and the Methodological Challenges of Empirical Sociology. Big Data & Society 1 (1).

Lianne Kolirin. 2019. New Zealand Prime Minister says, 'Our gun laws will change'". <https://www.cnn.com/2019/03/15/asia/new-zealand-gun-control-intl/index.html>. Viitattu 20.3.2020

Eagle, Nathan, ja Alex (Sandy) Pentland. 2006. Reality Mining: Sensing Complex Social Systems. Personal and Ubiquitous Computing 10 (4).

Eriksson Krutrök, Moa, ja Simon Lindgren. 2018. Continued Contexts of Terror: Analyzing Temporal Patterns of Hashtag Co-Occurrence as Discursive Articulations. Social Media + Society 4 (4).

Flyverbom, Mikkel, ja Anders Koed Madsen. 2015. Sorting Data out Unpacking Big Data Value Chains and Algorithmic Knowledge Production. Teoksessa Die Gesellschaft Der Daten. Bielefeld: Transcript Verlag.

General Python FAQ — Python 3.9.0 documentation. <https://docs.python.org/3/faq/general.html#what-is-python>. Viitattu 18.11.2020.

Twitter. 2017. Giving you more characters to express yourself. https://blog.twitter.com/en_us/topics/product/2017/Giving-you-more-characters-to-express-yourself.html. Viitattu 18.11.2020.

Green, Adam. Twitter API Tutorial – Aggregating tweets: Search API vs. Streaming API. <http://140dev.com/twitter-api-programming-tutorials/aggregating-tweets-search-api-vs-streaming-api/>. Viitattu 16.9.2020.

Grubmüller, Verena, Katharina Götsch, ja Bernhard Krieger. 2013. Social Media Analytics for Future Oriented Policy Making. *European Journal of Futures Research* 1 (1): 20.

Häikiö, Liisa ja Helena Heino. 2014. Tulkitsevan politiikka-analyysin lähtökohdat. Teoksessa *Tulkinnan mahti: johdatus tulkitsevaan politiikka-analyysiin*. Tampere: Tampere University Press 2014.

Helander, Daniel. 2017. VUOROVAIKUTUS HELSINGIN PORMESTARIEHDOKKAIDEN TWITTER- KESKUSTELUISSA KUNTAVAALEISSA 2017. Tampereen yliopisto. Puheviestinnän pro gradu -tutkielma.

Hellsten, Iina, ja Loet Leydesdorff. 2019. Automated Analysis of Actor–Topic Networks on Twitter: New Approaches to the Analysis of Socio-semantic Networks. *Journal of the Association for Information Science and Technology* 71(3).

Hern, Alex. 2018. Cambridge Analytica: How Did It Turn Clicks into Votes? <http://www.theguardian.com/news/2018/may/06/cambridge-analytica-how-turn-clicks-into-votes-christopher-wylie>. Viitattu 15.5.2021

Hoffman, Donna L, ja Thomas P Novak. 2017. Consumer and Object Experience in the Internet of Things: An Assemblage Theory Approach. *Journal of Consumer Research* 44(6).

Hrehorová, Markéta. 2019. The Ethics of First-Person Shooter Aesthetics. <http://gamephilosophy.org/wpcontent/uploads/confmanuscripts/pcg2019/Mark%C3%A9ta%20Hrehorova%20-%20The%20Ethics%20of%20FPS-Aesthetics.pdf> Viitattu 20.6.2020

Jungherr, Andreas. 2016. Twitter use in election campaigns: A systematic literature review. *Journal of Information Technology & Politics* 13 (1).

Knapp, Noora. 2021. Onko Koronavilkku luotettava ja miksi se vie niin paljon akkua? Selvitimme vastaukset näihin kysymyksiin ja siihen, miksi suomalaiset eivät ole ladanneet sovellusta. <https://www.mtvuutiset.fi/artikkeli/onko-koronavilkku-luotettava-ja-miksi-se-vie-niin-paljon-akkua-selvitimme-vastaukset-naihin-kysymyksiin-ja-siihen-miksi-suomalaiset-eivat-ole-ladanneet-sovellusta/8050644>. Viitattu 15.5.2021

Krauss. 2014. blackbox and observer, the "experiment profile". <https://commons.wikimedia.org/wiki/File:Blackbox3D-obs.png>. Viitattu 22.5.2021

Laaksonen, Salla-Maaria, ja Margareta Salonen. 2018. Kuka Saa Päättää, Mitä Dataa Tutkijalla on Käytössään? Ei Ainakaan Amerikkalainen Suuryritys. <https://rajapinta.co/2018/12/04/kuka-saa-paattaa-mita-dataa-tutkijalla-on-kaytossaan-ei-ainakaan-amerikkalainen-suuryritys/>. 21.1.2020

David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, Albert-Laszlo Barabasi, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, Marshall Van Alstyne. 2009. Computational Social Science. Science 323 (5915).

Li, Ya. 2015. Think Tank 2.0 for Deliberative Policy Analysis. Policy Sciences 48 (3).

Lindström, Kauri, Pekka Räsänen, Atte Oksanen, ja Johanna Nurmi. 2011. Poliittikkaprosessi ja aselainsäädännön uudistaminen Jokelan ja Kauhajoen koulusurmien jälkeen. Teoksessa Poliitiikan polut ja hyvinvointivaltion muutos. Helsinki: Kela.

Linnake, Tuomas. 2020. Hallitus alkaa tutkia suomalaisten liikkeitä Telian mobiilidatalla – seurataan rajoitusten tehoa. <https://www.is.fi/digitoday/mobiili/art-2000006463019.html>. Viitattu 10.7.2020

Longo, Justin, ja Kathleen Mcnutt. 2018. From policy analysis to policy analytics. Teoksessa Policy analysis in Canada.

Macias, Amanda. 2020. Facebook CEO Mark Zuckerberg Calls for More Regulation of Online Content.

<https://www.cnbc.com/2020/02/15/facebook-ceo-zuckerberg-calls-for-more-government-regulation-online-content.html>. Viitattu 20.5.2021

Macklin, Graham. 2019. The Christchurch Attacks: Livestream Terror in the Viral Video Age. *Combating Terrorism Center at West Point* 12 (6).

Makkonen, Kimmo, ja Petri Loukasmäki. 2019. Eduskunnan täysistunnon puheenaiheet 1999–2014: *Politiikka* 61 (2).

Marres, Noortje. 2015. Why Map Issues? On Controversy Analysis as a Digital Method. *Science, Technology, & Human Values* 40 (5).

Meadows, Donella H, ja Club of Rome. 1972. *The Limits to Growth*. New York: New American Library.

Michael Beatty. 2017. ”IBM Cloud Storage - Cleversafe”.

<https://www.slideshare.net/MichaelBeatty/ibm-cloud-storage-cleversafe>. Viitattu 2.5.2021

Mäenpää, Pasi, Maija Faehnle, ja Harry Schulman. 2017. *Kaupunkiaktivismi, Jakamistalous Ja Neljäs Sektori*. Teoksessa *Kansalaiset Kaupunkia Kehittämässä*. Tampere: Tampere Peace Research Institute, University of Tampere.

Natural Language Toolkit — NLTK 3.6.2 documentation.

<https://www.nltk.org/>. Viitattu 13.5.2021.

Nelimarkka, Matti. 2019. Aihemallinnus sekä muut ohjaamattomat koneoppimismenetelmät yhteiskuntatieteellisessä tutkimuksessa: kriittisiä havaintoja. *Politiikka* 61(1).

Nelimarkka, Matti. 2021. *Coding Social Science*.

<http://codingsocialscience.org>. Viitattu 30.6.2020

New Zealand Parliament. 2016. How a Bill Becomes Law.

<https://www.parliament.nz/en/visit-and-learn/how-parliament-works/how-laws-are-made/how-a-bill-becomes-law/>. Viitattu 28. helmikuuta 2020.

New Zealand Parliament. 2019a. What Is 'Urgency'?

<https://www.parliament.nz/en/visit-and-learn/how-parliament-works/fact-sheets/what-is-urgency>. Viitattu 28.2.2020

New Zealand Parliament. 2019b. Arms (Prohibited Firearms, Magazines, and Parts) Amendment Bill

https://www.parliament.nz/en/pb/bills-and-laws/bills-proposed-laws/document/BILL_86423/arms-prohibited-firearms-magazines-and-parts-amendment.

Viitattu 28.2.2020

Nielsen, Rasmus. 2014. Democracy [Draft] [#digitalkeywords] – Culture Digitally.

<http://culturedigitally.org/2014/05/democracy-draft-digitalkeywords/>. Viitattu 10.1.2020

Ohjelmoinnin perusteet Python-kielillä.

<https://www.cs.helsinki.fi/group/linkki/materiaali/python-perusteet/materiaali.html#13>.

Viitattu 18. marraskuuta 2020.

pandas - Python Data Analysis Library.

<https://pandas.pydata.org/>. Viitattu 18. marraskuuta 2020.

Parviala, Antti. 2020. Viisi kuvaa kertoo talouden uudesta käännteestä: Suomen Pankki seuraa nyt tarkasti rekkojen liikkeitä ja näitä Google-hakuja.

<https://yle.fi/uutiset/3-11646678>. Viitattu 15.5.2021

Passell, Roberts ja Ross. 1972. The Limits to Growth.

<https://www.nytimes.com/1972/04/02/archives/the-limits-to-growth-a-report-for-the-club-of-romes-project-on-the.html>. Viitattu 5.1.2021

Plotly. 2021. Chart Studio Docs.

<https://plotly.com/python/chart-studio/>. Viitattu 15.5.2021

Project Jupyter.

<https://www.jupyter.org>. Viitattu 18.11.2020.

Propositiologiikka. 2020.

<https://fi.wikipedia.org/w/index.php?title=Propositiologiikka&oldid=19071517>. Viitattu 16.8.2020.

Quade, Edward S. 1975. Analysis for Public Decisions. New York Elsevier.

Rogers, Richard. 2009. The End of the Virtual: Digital Methods. Amsterdam: Amsterdam University Press.

Rogers, Richard. 2019. Doing Digital Methods. California : SAGE Publications

Saari, Juho. 2011a. Pienten askelten politiikka ja hyvinvointivaltion muutos. Teoksessa Poliitiikan polut ja hyvinvointivaltion muutos. Helsinki: Kela.

Saari, Juho. 2011b. Sosiaaliturvan kokonaisuudistus Tapahtumarakenneanalyysi syksystä 2007 syksyyn 2010. Teoksessa Poliitiikan polut ja hyvinvointivaltion muutos. Helsinki: Kela.

Sabatier, Paul A. 2011. Understanding and influencing the policy process. Policy Sciences 45(1):1-21.

Saikkonen, Paula. 2014. Monimuotoinen kokoomateos tulkitsevasta politiikka- analyysistä. ALUE JA YMPÄRISTÖ 43(1).

Savage, Mike, ja Roger Burrows. 2007. The Coming Crisis of Empirical Sociology. Sociology 41 (5).

Savage, Mike, ja Roger Burrows. 2009. Some Further Reflections on the Coming Crisis of Empirical Sociology. Sociology 43 (4).

Savić, Dobrica. 2019. From Digitization, through Digitalization, to Digital Transformation. Online Searcher 43(1).

Seitamaa-Hakkarainen, Piritta. 2014. Kvalitatiivinen sisällönanalyysi.

<https://metodix.fi/2014/05/19/seitamaa-hakkarainen-kvalitatiivinen-sisallon-analyysi/>

Viitattu 4.7.2020

Sense Corp. 2018. Data Science Pyramid | Improving the Data Process.

<https://sensecorp.com/data-science-pyramid/>. Viitattu 1.2.2021

Smallcombe, Mark. 2020. Structured vs Unstructured Data: 5 Key Differences.

<https://www.xplenty.com/blog/structured-vs-unstructured-data-key-differences/>. Viitattu

15.4.2021

Strippel, Christian. 2014. Dhiraj Murthy (2013): Twitter. Social Communication in the Twitter Age. Medien & Kommunikationswissenschaft 62 (1).

Tirronen, Jarkko. 2011. Poliittikkaprosessin sisällä Polkuriippuvuus, agendan asettaminen ja innovaatioiden leviäminen. Teoksessa Poliittikan polut ja hyvinvointivaltion muutos. Helsinki: Kela.

Toivanen, Pihla, Jukka Huhtamäki, Katja Valaskivi, ja Minttu Tikka. 2020. Aihemallinnus hybridin mediatapahtuman ja merkitysten kierron tutkimuksessa. Media & Viestintä 43(1).

Turner, Graham, ja Cathy Alexander. 2014. Limits to Growth Was Right. New Research Shows We're Nearing Collapse.

<https://www.theguardian.com/commentisfree/2014/sep/02/limits-to-growth-was-right-new-research-shows-were-nearing-collapse>. Viitattu 30.7.2020

Twitter. 2021. Twitter Verification Requirements - How to Get the Blue Check.

<https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts>.

Viitattu 10.4.2021

Vainikka, Eliisa, Antti Kanner, Riikka Era, Auli Harju, Anu Koivunen, Olli Seuri, ja Eetu Mäkelä. 2020. Kilpailukykyloikasta ruutitynnyyriin: Metaforat ja affektiivisuus poliittisen kriisin journalismissa. *Media & viestintä* 43 (4).

Venturini, Tommaso. 2010. Building on Faults: How to Represent Controversies with Digital Methods. *Public Understanding of Science* 21 (7).

Venturini, Tommaso, Liliana Bounegru, Jonathan Gray, ja Richard Rogers. 2018. A Reality Check(List) for Digital Methods. *New media & society* 20 (11).

Venturini, Tommaso, ja Bruno Latour. 2010. The Social Fabric: Digital Traces and Quali-Quantitative Methods. *Proceedings of Future En Seine 2009*. Cap Digital.

visual-analytics.eu. 2017. What Is Visual Analytics?.
www.visual-analytics.eu/faq/. Viitattu 30.1.2021

Vuleta, Branka. 2021. How Much Data Is Created Every Day?.
<https://seedscientific.com/how-much-data-is-created-every-day/>. Viitattu 16.5.2021

Wahlquist, Calla. 2019. Ardern Says She Will Never Speak Name of Christchurch Suspect.
<https://www.theguardian.com/world/2019/mar/19/new-zealand-shooting-ardern-says-she-will-never-speak-suspects-name>. 20.2.2020

Watts, Duncan J. 2007. A Twenty-First Century Science. *Nature* 445(489)

Weaver, Matthew. 2019. Christchurch Mayor on Mass Shooter: 'I've No Idea What His Name Is'.
<https://www.theguardian.com/world/2019/jul/08/christchurch-mayor-shooter-no-idea-name>. Viitattu 20.1.2020

Zia, Muhammad. 2019. Statistics Reveal Facebook's Social Media Monopoly.
<https://www.digitalinformationworld.com/2019/10/monthly-active-users-of-top-social-media-platforms-and-messaging-apps.html>. Viitattu 1.6.2021

Åkerman, Maria. 2009. ”Hybridit ja ympäristön politiikka”. Teoksessa Vihreä teoria: Ympäristö yhteiskuntateorioissa. Helsinki: Gaudeamus