

LIKENNEDATA – HAVAINNOSTA TILANNEKUVAKSI



Ylemmän ammattikorkeakoulututkinnon opinnäytetyö

Tulevaisuuden liikennejärjestelmät, Riihimäki

Syksy 2020

Jonas Kurtto

TIIVISTELMÄ

Tämän opinnäytetyön tavoitteena oli tarkastella Helsingin kaupungin olemassa olevia liikennetiedon lähteitä ja kasvattaa liikennetiedon määrää ja laatua kohdealueella tehtyjen havaintojen pohjalta. Työssä selvitettiin ja kuvattiin liikennetiedon ominaisuuksia, millaisista lähteistä sitä voidaan kerätä, kuinka sitä voidaan käsitellä ja miten sitä voidaan hyödyntää ja jakaa erilaisiin käyttötarkoituksiin.

Helsingin kaupungin tavoitteena on jakaa kerättyä liikennedataa sovelluskehityksen ja uusien palveluiden luomisen tarpeisiin sekä hyödyntää sitä tilannekuvan muodostamiseen.

Liikennedataa hyödynnetään muun muassa liikenteenhallinnassa ja arkipäiväisessä liikkumisessa sekä liikenteen suunnittelun ja operoinnin apuvälineenä. Reaaliaikainen liikennedata mahdollistaa käyttäjälleen senhetkisen tilanteen mukaisen parhaimman reitin löytämisen paikasta toiseen ja se on oleellinen osa myös autonomisten ajoneuvojen toimintaedellytyksiä sekä liikenteen reaaliaikaista tilannekuvaa.

Työssä on esitetty liikennetiedon prosessi havainnosta hyödynnettäväksi informaatioksi sekä selvitetty uusien liikennetiedon lähteiden ja keruumenetelmien potentiaalia asetettujen tavoitteiden saavuttamiseksi. Liikennetiedon määrää ja laatua voidaan kasvattaa perinteisten teknisten ratkaisujen lisäksi myös erilaisilla ihmisten liikkumiseen perustuvilla keinoilla, joihin julkisen sektorin toimijat voivat vaikuttaa.

Avainsanat liikennedata, tilannekuva, datalähde

Sivut 82 sivua

ABSTRACT

The aim of this thesis was to examine the existing sources of traffic information in the City of Helsinki and to increase the quantity and quality of traffic information on the basis of observations made in the target area. In this thesis, the properties of traffic information were investigated and described, from what sources it can be collected, how it can be processed and how it can be utilized and distributed for different purposes.

The objective of the City of Helsinki is to share the collected traffic data for the needs of application development and the creation of new services and to utilize it to form real-time awareness of traffic. Traffic data is utilized in, among other things, traffic management and in everyday mobility as well as a tool for traffic planning and operation. Real-time traffic data allows its user to find the best route from one location to another according to the current traffic situation and it is also an essential part of the operating conditions of autonomous vehicles and real-time awareness of traffic.

In this thesis, the process of traffic information from observation to usable knowledge is presented and the potential of new traffic data sources and collection methods to achieve the objectives of the City of Helsinki are assessed. In addition to traditional technical solutions, the quantity and quality of traffic information can be increased by various means based on the movement of people, which can be influenced by public sector actors.

Keywords traffic data, situational awareness, data source

Pages 82 pages

Sisällys

1	Johdanto	5
2	Työn taustaa, tavoitteet, toteutus ja rajaukset.....	6
3	Liikenteestä kerättävä tieto.....	9
3.1	Datamuotoinen liikennetieto.....	11
3.2	Reaaliaikaisen datan ominaisuudet ja vaatimukset.....	13
3.3	Liikennetiedon kerääminen	16
4	Liikennetiedon lähteet	20
4.1	Tietolähteet tyypeittäin	21
4.2	Perinteiset liikennetiedon lähteet	23
4.2.1	Induktioilmaisimet sekä muut tiehen asennettavat ilmaisimet	24
4.2.2	Tutka- ja infrapunailmaisimet	27
4.2.3	Kameratekniikkaan perustuvat ilmaisimet	28
4.3	Uudemmat ilmaisinteknologiat ja havainnointimenetelmät.....	29
4.3.1	LiDAR	29
4.3.2	Magnetometri-ilmaisimet	30
4.3.3	Koneoppimiseen perustuvat menetelmät	31
4.3.4	Joukkoistettu data.....	32
4.3.5	Joukkoliikennedata.....	33
4.3.6	Langattomat signaalit.....	34
4.3.7	GPS-paikkatieto, GPS-tutkimukset ja matkapuhelindata.....	35
4.3.8	Kelluva ajoneuvodata.....	36
5	Liikennetiedon hyödyntäminen	39
5.1	Liikennetiedon jakaminen.....	41
5.1.1	Yleinen datan jakamisprosessi	44
5.1.2	Avoimen datan konsepti	46
5.2	Jaettu liikennedata.....	48
6	Helsingin kaupungin liikennetieto	51
6.1	Nykytilan kuvaus ja tavoitteet	51
6.2	Olemassa olevat liikennetiedon lähteet	54
6.2.1	Automaattiset laskentalaitteet	54
6.2.2	Manuaalilaskennat	56
6.2.3	Muut liikennetiedon lähteet	57
6.3	Potentiaaliset tietolähteet ja datanmuodostusmenetelmät.....	59

6.3.1 Helsingin kaupungin rooli.....	64
6.4 Liikenteen reaaliaikainen tilannekuva	66
Lähteet.....	71

Kuvat ja taulukot

Kuva 1. Keskeisimmät liikennedataan liittyvät informaation lähteet	11
Kuva 2. Oleellisimmat liikennedatata tuottavat tekijät	12
Kuva 3. Liikennedataan ja tietojärjestelmiin tyypillisesti liittyvät riskit	15
Kuva 4. Liikennedatan hyödyntämisprosessi vaiheineen.....	17
Kuva 5. Kävelyn ja pyöräilyn tietolähteet luokittain	22
Kuva 6. Kiinteiden ilmaisimien hyödyt ja haitat	24
Kuva 7. Havainnollistus kelluvan ajoneuvotarkkailun toiminnasta.....	37
Kuva 8. Päävaiheet datan iteratiivisessa avaamisprosessissa.....	45
Kuva 9. Liikennetiedon elinkaari tavoitetilassa 2030	53
Kuva 10. Esimerkki kulkumuotojen tunnistamisesta koneoppimisen avulla	61
Kuva 11. Esimerkkitoteutus Eco-Counterin kävely- ja pyöräilyilmaisimen käytöstä	63
Kuva 12. Tukholman liikenteen tilannekuvapalvelu.....	68
Taulukko 1. Inventaarion mukaiset automaattiset laskentalaitteet	55
Taulukko 2. Inventaarion mukaiset manuaalilaskennat.....	56

1 Johdanto

Monelle on luultavasti joskus tullut eteen tilanne, jossa on ollut tarve kulkea tiettyyn paikkaan ilman tietoa siitä, kuinka päästä perille kohteeseen. Nykyisin hyvin tavanomainen tapa suunnitella matkaansa varsinkin uusiin paikkoihin on käyttää älylaitteella verkkopohjaista reittiopastuspalvelua tai -sovellusta. Matkan suunnittelussa on hyödyllistä tietää reittiopastuksen lisäksi ainakin tarvittavat kulkuvälineet sekä arvio matkaan tarvittavasta ajasta.

Tällaisen nykyisesti tavallisen palvelun toiminnan taustalla on suuri määrä tietoa, jonka avulla voidaan muodostaa käyttäjälle senhetkisen tilanteen mukaisesti paras reitti paikasta toiseen. Toiminta perustuu aikaisempien ajanhetkien vastaavassa tilanteessa vallinneiden liikennetilanteiden avulla muodostettuun olettamukseen liikenteen nykytilanteesta sekä esimerkiksi julkisen liikenteen aikataulutietoon eli niin sanottuun historiatietoon. Tällöin ei kuitenkaan voida huomioida juuri tällä hetkellä vallitsevia liikenteellisiä poikkeamia, joita on esimerkiksi tämän työn kohdealueella Helsingissä ajoittain runsaasti.

Liikenteen reaaliaikaisella tilannekuvalla tarkoitetaan ajantasaista, juuri tämän hetken liikennetilannetta kuvaavaa otosta, joka ottaa huomioon liikennemäärän ohella liikenteen sujuvuuteen liittyviä erityispiirteitä hyödyntäen erilaisia tietolähteitä. Tällaisista tietolähteistä saatavaa niin sanottua liikennedatata voidaan kerätä moninaisin tavoin aina liikkujan sijainnin paljastavasta GPS-tiedosta koneoppimiseen perustuvaan kuvantunnistukseen.

Työssäni liikenneinsinöörinä Helsingin kaupungilla yhtenä tehtävänäni on osallistua erinäisiin projekteihin, joiden tarkoituksena on liikennedatan määrällinen ja laadullinen kasvattaminen liikenteen reaaliaikaisen tilannekuvan muodostamiseksi. Tämän lisäksi liikennedatata pyritään jakamaan avoimena datana sovelluskehittäjille ja muille siitä kiinnostuneille tahoille.

Helsingin kaupungin roolina on toimia tulevaisuudessakin liikennedatan hyödyntämisen mahdollistajana ja olla liikenteellisesti teknologinen edelläkävijä sekä kaupunkistrategisen visionsa mukaisesti *maailman toimivin ja parhaiten digitalisaatiota hyödyntävä kaupunki*.

2 Työn taustaa, tavoitteet, toteutus ja rajaukset

Helsinki on liikenteellisesti erityinen Suomen mittakaavassa liikenteen määrän ohella siitäkin syystä, että se sijaitsee verrattain hankalasti kapeana alueena veden ympäröimänä. Tämän lisäksi kaupungissa sijaitsee satama, josta aiheutuu tasaisin väliajoin merkittäviä raskaan liikenteen ruuhkauttavia vaikutuksia alueen katuverkolle laivojen saapuessa satamaan. Raitioliikenteen vaikutukset alueen liikenteeseen ovat myös huomattavia ja ainutlaatuisia Suomessa. Nämä alueelliset erityispiirteet yhdistettynä liikenteen suureen määrään aiheuttavat liikenteen sujuvuudelle haasteita, joihin voidaan pyrkiä vastaamaan liikennedatalla hyödyntämällä.

Helsingin kaupungin ja muun pääkaupunkiseudun liikennettä operoidaan Pasilassa sijaitsevasta tieliikennekeskuksesta käsin, johon kuuluu myös pääkaupunkiseudun liikenteenhallintakeskus eli PLH. Sen tehtävänä on muun muassa liikenteen alueellisesta sujuvuudesta huolehtiminen, operatiivinen liikenteen ohjaaminen, liikenteen seuranta sekä häiriötilanteiden hallinta yhteistyössä viranomaisten ja alueen kaupunkien, Helsingin, Espoon ja Vantaan kanssa.

Liikenteen sujuvuuteen vaikuttavia ajantasaista seurantaan vaativia tilanteita voivat olla muun muassa katuverkolla tapahtuvat tietyöt, suuremmat yleisötapaukset, onnettomuudet ja kolarit tai tavanomaisesta poikkeavat muusta liikenteestä johtuvat liikennemäärän muutokset, kuten edellä kuvatut satamaliikenteen purkutilanteet katuverkolle. Suomen vaihtelevat sääolosuhteet aiheuttavat myös vaikeasti ennustettavia muutoksia liikenteeseen.

On myös huomionarvoista, että liikenteen teknologiset innovaatiot ovat hiljalleen saapuneet Suomeen ja Helsingissäkin liikkuu tälläkin hetkellä ilman fyysistä kuljettajaa toimivia automaattisia ajoneuvoja, jotka nekin pystyvät hyödyntämään liikennedatalla monipuolisesti. Tulevaisuudessa täysin autonomisesti toimivien ajoneuvojen toimintaedellytyksinä on riittävän kattava, ajantasainen kuva ympäristöstä ja muista tienkäyttäjistä.

Muun muassa edellä kuvatuista syistä johtuen monipuolisen sekä määrällisesti ja laadullisesti riittävän liikennedatan tarve on Helsingin kaupungille ilmeinen. Tämän työn tarkoituksena on löytää toimivia ratkaisuja tarkastellen Helsingin kaupungin olemassa olevia

liikennedatan lähteitä ja selvittää uusien potentiaalisten datalähteiden ja datankeruumenetelmien tarjontaa liikennedatan määrän ja laadun kasvattamiseksi. Työssä pyritään selvittämään, millaisista lähteistä liikennedatata on mahdollista saada ja kuinka tällaista dataa voidaan hyödyntää muun muassa liikenteen reaaliaikaisen tilannekuvan muodostamisessa. Työssä käsitellään myös liikennedatan elinkaaren lukuisat vaiheet havainnosta lisäarvoa tuovaksi informaatioksi. Työssä pyritään vastaamaan kysymyksiin

- miten liikenteestä voidaan muodostaa hyödyllistä digitaalista informaatiota ja ajantasaista tilannekuvaa?
- millä tavoin liikennedatata voidaan kerätä, käsitellä, hyödyntää ja jakaa?
- kuinka Helsingin kaupungin olemassa olevia liikennedatan lähteitä voidaan hyödyntää paremmin ja mitä uusia tapoja ja menetelmiä voidaan käyttää liikennedatan määrän ja laadun kasvattamiseksi?

Näihin kysymyksiin pyritään vastaamaan perehtymällä asiasta tehtyyn tieteelliseen tutkimukseen, kirjallisuuteen ja ajantasaisiin artikkeleihin sekä kontaktoimalla asiaan perehtyneitä tahoja. Työssä korostuvat ajantasaiset ja eri menetelmien toimintaa kuvailevat lähteet sekä relevanteiksi arvioidut tutkimuksissa ja pilottihankkeissa havaitut kokemukset ja huomiot, joita voidaan hyödyntää Helsingin kaupungin tavoitteiden saavuttamisessa. Käsiteltävien menetelmien tyypillisesti kokeellisesta ja kehitysasteisesta luonteesta johtuen niitä pyritään arvioimaan niiden potentiaalinen näkökulmasta.

Työn teoriaosuudessa kuvataan liikenteestä kerättävän tiedon ominaisuuksia perinteisestä historiatiedosta nykypäivän liikennedataan. Teoriaosuudessa esitetään liikennedatan elinkaari havainnosta käyttökelpoiseksi informaatioksi ja se keskittyy antamaan vastauksia liikennedatan määrän ja laadun kasvattamiseen sekä sen jakamiseen. Työssä tarkastellaan myös reaaliaikaisen liikennedatan erityispiirteitä tieto- ja tietoliikenneteknisestä näkökulmasta, huomioiden muun muassa tiedonsiirtoon liittyviä vaatimuksia.

Työn empiirisessä osuudessa arvioidaan potentiaalisten datalähteiden ja datankeruumenetelmien soveltuvuutta ja ominaispiirteitä peilaten niitä Helsingin kaupungin tarpeisiin ja havaittuihin liikenteen tietolähdepuutteisiin teoriaosuudessa esitettyjen huomioiden pohjalta. Potentiaalisissa liikennedatan muodostustavoissa otetaan myös

kantaa niiden sopivuudesta eri kulkumuotojen liikennedatan keräämiseen huomioiden kohdealueen erityispiirteet. Tämän lisäksi työssä käsitellään esimerkillisiä käyttötarkoituksia liikennedatan hyödyntämiseksi esimerkiksi liikenteen reaaliaikaisen tilannekuvan muodossa.

Työ rajautuu teoriaosuuden, liikennetiedon lähteiden ja niiden hyödyntämisen osalta Helsingin kaupungille työn tavoitteiden kannalta oleellisimmiksi ja potentiaalisimmiksi arvioituihin ratkaisuvaihtoehtoihin. Rajaus perustuu kohdealueen nyky- ja tavoitetilan asettamiin reunaehtoihin, kuten liikennetiedon lähteiden teknisiin ominaisuuksiin ja rajoitteisiin sekä kohdealueen erityispiirteisiin, kuten eri menetelmien arvioituun soveltuvuuteen Helsingin kaltaiselle, tiheästi asutetulle ja runsasliikenteiselle alueelle.

Rajauksessa huomioidaan myös, että erilaisten menetelmien ja teknologioiden käytettävyys ja sopivuus saattavat edellyttää tapauskohtaisesti yhteensopivaa kokonaisinfrastruktuuria muun muassa laite- ja järjestelmätasolla. Täten toisaalla toimiva kokonaisratkaisu ei välttämättä sovellu sellaisenaan tai lainkaan kohdealueen tarpeisiin ilman suuria investointikustannuksia tai kokonaisvaltaista nykytilan muutosta. Oleellisimpia ja potentiaalisimpia ratkaisuvaihtoehtoja pyritään täten havaitsemaan niiden toteuttamiskelpoisuus ja kustannusvaikutukset huomioiden. Ratkaisuvaihtoehtojen soveltuvuutta arvioidaan myös teoriaosuuden havaintojen ja kokemusten perusteella.

Julkisen sektorin toimijat toteuttavat liikennedatan edellyttämiä toimenpiteitä, kuten laite- ja järjestelmätilauksia sekä fyysisiä asennuksia tyypillisesti hankintoina kolmansilta osapuolilta. Työssä huomioidaan myös Helsingin kaupungin nyky- ja tavoitetilan rooli liikennedatan mahdollistajana. Työssä ei täten esitetä esimerkiksi kaikkien datankeruumenetelmien syvällisiä toimintaperiaatteita sähkö- ja tietoteknisine ominaisuuksineen tai avata liikennedatan käsittelyssä ja analysoinnissa käytettävien menetelmien ja algoritmien toimintalogiikkaa tai -periaatteita perusteellisesti.

Käsiteltäviä kokonaisuuksia, kuten liikennedatan hyödyntämistä avataan siinä laajuudessa, kun se on työn tavoitteiden kannalta tarkoituksenmukaista. Kokonaisuutena työssä käsiteltävät aihepiirit on niiden monialaisen, muun muassa tieto-, tietoliikenne-, sähkö- ja liikenneteknisen luonteensa vuoksi rajattu työn tavoitteiden kannalta tarpeelliseen ja soveltamiseen tarvittavan ymmärryksen tuovaan laajuuteen.

3 Liikenteestä kerättävä tieto

Liikenteen ruuhkautuminen on tavallinen ilmiö kaupunkialueilla ja sen vaikutuksia voidaan pyrkiä vähentämään seuraamalla ja ennustamalla liikennevirtaa. Jotta tämä olisi mahdollista, vallitsevasta liikennetilanteesta tarvitaan informaatiota. Tällaista tietoa eli liikennedataa voidaan kerätä monista erilaisista lähteistä. (Wang ym., 2020, s. 1)

Liikennedataa hyödynnetään muun muassa liikenteenhallinnassa ja ihmisten arkipäiväisessä liikkumisessa. Liikenteenhallintaa toteuttavat tahot, kuten julkiset toimijat käyttävät informaatiota liikenteen suunnittelun ja operoinnin apuvälineenä. Ajoneuvojen kuljettajat hyötyvät reaaliaikaisesta informaatiosta, jonka perusteella kuljetun reitin mahdolliset erityistilanteet kiertoteineen ja kestoaikoineen ovat helposti saatavilla matkaa varten. Tämä edellyttää mahdollisimman tarkkaa, luotettavaa, ajantasaista ja monipuolista liikennedataa. (Leduc, 2008, s. 2)

Tie- ja liikennetietoja käytetään toiminnan ja suunnittelun lähtötietoina ja niitä tarvitaan lisäksi liikenteen ohjaustoimintaan, strategiasuunnitteluun sekä ylläpitoon kaikilla toiminnan tasoilla. Liikennetietoja ovat esimerkiksi liikennemäärät, ajonopeudet, matkojen lähtöpisteet ja määränpäätt, liikenteen suoritteet, onnettomuus- ja sujuvuustiedot sekä liikennevirrat. (Niinikoski & Laine, 2005, s. 11) Tämän lisäksi esimerkiksi matka-aika on oleellinen muuttuja liikenteen mallinnuksessa ja liikenneverkon tehokkuuden arvioinnissa (Feng ym., 2014, s. 161).

Niinikosken & Laineen (2005, s. 3) mukaan keskeisimpiä liikennetietoja ovat muun muassa määrä- ja koostumustiedot sekä eri ajoneuvoluokkien KVL-tiedot. Niiden avulla saadaan muodostettua yleinen käsitys tien merkityksellisyydestä ja määritettyä toimenpiteiden laatutasoja. Tämän lisäksi hankesuunnittelun liikenneteknisessä mitoituksessa hyödynnetään liikennemäärätietoja huipputuntien aikana, liikenneturvallisuustyön ja liikennetiedotuksen hyödyntäessä liikenteen nopeus- ja sujuvuustietoa.

Liikennejärjestelmä- ja verkkotason suunnittelussa käytetään usein liikennevirta- ja matkakohtaista tietoa taustatietoineen. Näiden ohella käytetään lisäksi liikennevirta- ja käyttäjätietoja ylläpidon suunnittelussa sekä onnettomuustietoja kaikkien

suunnittelutasojen perustelussa ylläpidon ja hankkeiden osalta. Toimenpiteiden vaikutusten arvioinnissa voidaan hyödyntää liikennevirtatietoa. (Niinikoski & Laine, 2005, s. 3)

Liikennetieto voidaan jakaa kahteen kategoriaan, staattiseen tietoon eli historiatietoon sekä dynaamiseen tietoon eli muuttuvaan tietoon. Liikenteen kontekstissa historiatiedolle on ominaista pysyvyys eli se muuttuu harvoin tai ei ollenkaan. Dynaaminen tieto on tyypillisesti vaihtuvaa. Reaaliaikaisuus kuvaa tiedon välitöntä saatavuutta tai julkaisemista tiedon syntymisen tai keräämisen tapahduttua ja dynaaminen tieto onkin usein reaaliaikaista tai lähes reaaliaikaista. Nämä tietotyypit eivät kuitenkaan ole toistensa synonyymejä. (Pihlajamaa ym., 2020, s. 21)

Liikenteen historiatiedon hyöty perustuu olettamukseen, että liikenne seuraa ajallista trendiä ollen siten tiettyssä paikassa tiettyyn aikaan suhteellisen samankaltaista kuin vastaavana ajanhetkenä menneisydessä. Tarkasteltava ajanjakso on tyypillisesti kalenteriviikko, joka erotellaan viikonpäiviin ja kellonaikoihin. Historiatiedossa tarkastellaan usein tien käyttäjien lukumäärää ja nopeutta keskiarvoisesti. Laadittavan analyysin tarkkuus paranee tiedon luotettavuuden kasvaessa, kun tietoa on kerätty laajemmalla historialliselta ajanjaksolta. (ESRI, 2020)

Liikenteen dynaaminen tieto on usein ja nopeasti muuttuvaa tietoa. Se on tästä syystä johtuen haettava suoraan lähteestään, eikä sitä voida muuttumisnopeutensa takia viedä tietokantaan historiatiedon tapaan. Tällaisia tietoja ovat muun muassa sijainti- ja häiriötiedot sekä muu, ajallisesti nopeasti muuttuva liikenneinformaatio. (Leskinen ym., 2016, s. 11)

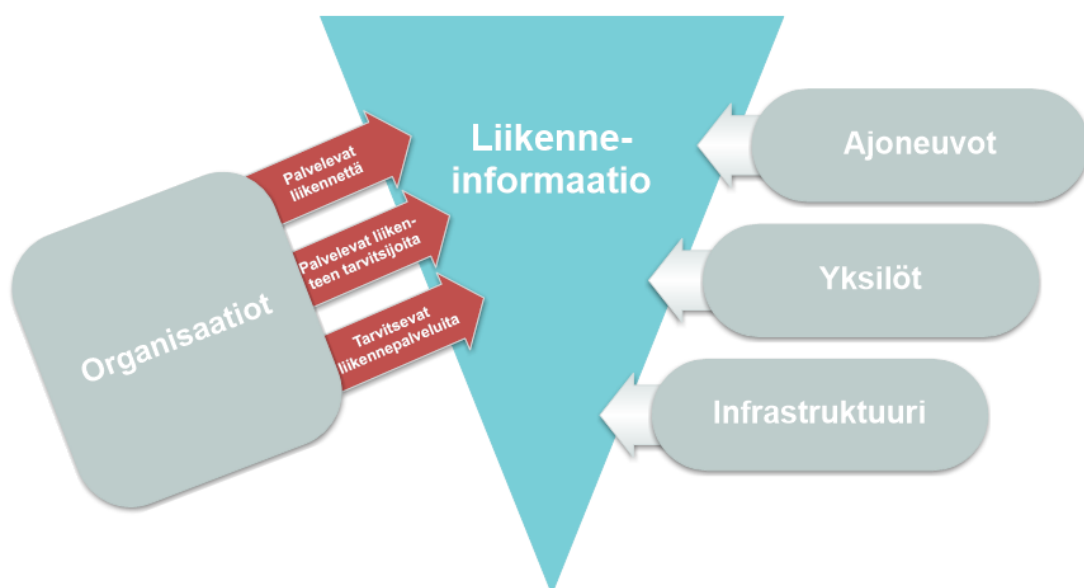
Kerättyä liikennetietoa säilytetään digitaalisessa muodossa erilaisissa valtiollisissa tietojärjestelmissä, kuten liikenteen automaattisen mittauksen LAM-tietokanta (Niinikoski & Laine, 2005, s. 11). Kunnat ja kaupungit keräävät liikennetietoa omia tarpeitaan varten ja tällöin tietoa voidaan säilyttää esimerkiksi kaupungin omilla palvelimilla tai pilvipalvelussa. Eräitä merkittäviä liikennetietoa kerääviä ja varastoivia tahoja ovat erinäiset suuret reittipastuspalveluita tarjoavat yritykset, kuten Google ja HERE (BITRE, 2014, ss. 13-14).

3.1 Datamuotoinen liikennetieto

Data on numeerista tai muuta tietoa, joka koostuu havainnoista. Itsessään ilman kontekstia data ei ole informatiivista. Informaatio on datasta ja sen kontekstista analysoimalla tai tulkitsemalla koostettua informaatiota. Kerätyn informaation avulla voidaan osaamisen ja kokemuksen avustuksella saavuttaa tietoisuus ja ymmärrys asiasta. (Zins, 2007, s. 481) Tämä on sovellettavissa myös vallitsevaan liikennetilanteeseen.

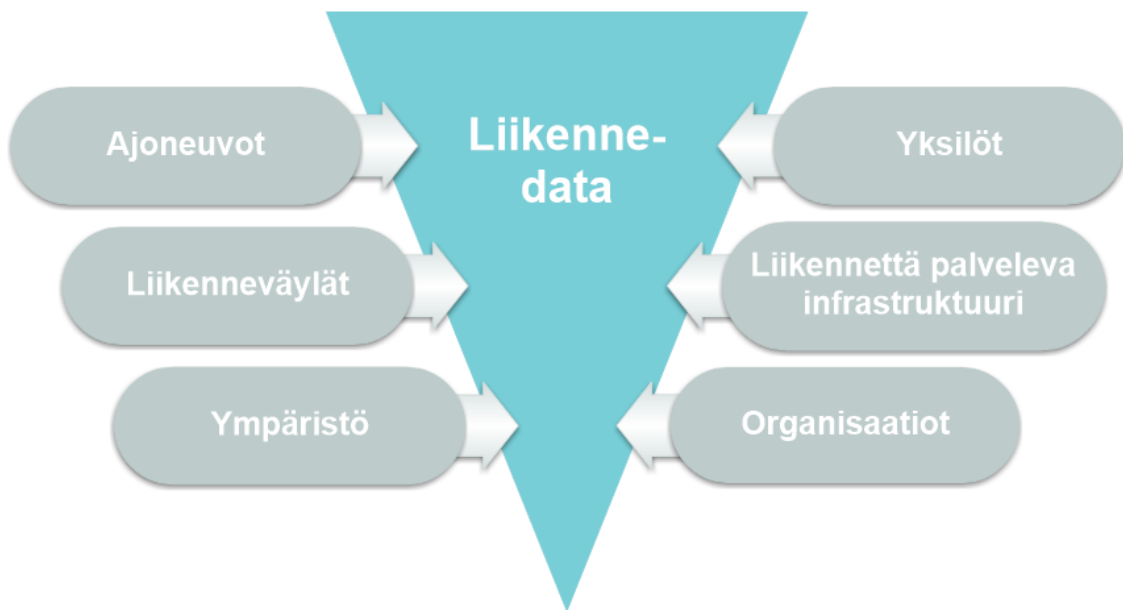
Digitaalinen eli datamuotoinen liikennetieto on olomuodoltaan tiedostomuotoista, kuten mikä tahansa muukin tietotekninen data. Sisällöllisesti liikennedata sisältää havaintoja, jotka ovat peräisin jostakin liikennetiedon lähteestä. Liikennedata on siis toisin sanoen liikenneinformaatiota sisältävää dataa. Linturin & Kuittisen (2014, s. 24) mukaan liikennedataan liittyvää informaatiota syntyy lukuisista erilaisista lähteistä ja jaottelua voidaan tehdä informaation tuottajan mukaan. Keskeisimmät liikennedataan liittyvät informaation lähteet ovat ajoneuvot, yksilöt, infrastruktuuri sekä liikennettä palvelevat, liikenteen tarvitsijoita palvelevat ja liikennepalveluita tarvitsevat organisaatiot. Keskeisimmät liikennedataan liittyvät informaation lähteet on esitetty kuvassa 1.

Kuva 1. Keskeisimmät liikennedataan liittyvät informaation lähteet (mukaellen Linturi & Kuittinen, 2014, s. 24).



Linturi & Kuittinen (2014, s. 24) tarkentavat, että liikennedatata muodostuu erityisesti tienkäyttäjien ja tie-infrastruktuurin kohteista ja niiden välisistä suhteista. Oleellisimpia liikennedatan tuottajia ovat ajoneuvot, yksilöt, liikenneväylät ja muu liikennettä palveleva infrastruktuuri, ympäristö sekä erinäiset organisaatiot. Oleellisimmat liikennedatata tuottavat tekijät on esitetty kuvassa 2.

Kuva 2. Oleellisimmat liikennedatata tuottavat tekijät (mukaellen Linturi & Kuittinen, 2014, s. 24).



Raa'alla datalla tarkoitetaan käsittelemätöntä ja analysoimatonta dataa. Erilaisista tietolähteistä saatava raakadata ei ole itsessään käyttökelpoista esimerkiksi palveluiden kehittämiseen. Sen hyödyntäminen edellyttää jalostamista visualisoimalla tai yhdistelemällä sitä muuhun tietoon. Tämä edellyttää muun muassa yhteensovitettuja protokollia, sovellusten välisiä rajapintoja, kehitettyjä hakualgoritmeja sekä metadataa ja katalogeja. (Sundquist & Oesch, 2014, ss. 16-18) Raakadatan analysointi edellyttää algoritmien käyttöä hyödyllisen informaation, kuten havaintojen muodostamiseksi (NPRA, 2004, s. 16).

Raakadata on usein sisällöltään epätäydellistä ja epäjohdonmukaista ja se saattaa sisältää puuttuvia arvoja, kohinaa tai tarpeetonta informaatiota. Tällainen laadultaan huono data laskee käytettävien hakualgoritmien tehoa. Tähän voidaan vaikuttaa raakadatan

esikäsittelyllä, jolloin raakadatan laatu ja luotettavuus paranee. (Ramírez-Gallego ym., 2017, s. 39)

Tietojoukko (*engl. Data Set*) on termi, joka kattaa kaikki datan olomuodot. Kun raakadataa jalostetaan muuttamalla, editoimalla tai puhdistamalla hyödylliseksi informaatioksi, siitä muodostuu jalostettua dataa. Data voi olla myös muun muassa kuvan, kartan tai taulukon muodossa. Tietojoukko sisältää usein joukon tiedostoja, jotka on kerätty yhdestä lähteestä tai jotka on luotu yhdellä laitteella. (MGDS, 2020)

3.2 Reaaliaikaisen datan ominaisuudet ja vaatimukset

Verkossa tapahtuva tiedonsiirtoliikenne voidaan jakaa elastiseen ja reaaliaikaiseen kategoriaan. Elastinen tietoliikenne kuvaa tiedonsiirtoa, joka ei ole aikasensitiivistä, kuten sähköpostiliikenne tai internetin selaus. Reaaliaikainen tietoliikenne kattaa informaation, joka on käyttökelpoista ainoastaan silloin, kun se vastaanotetaan pienen ajallisen viiveen sisällä. Tällöin esimerkiksi datapakettien uudelleenlähetyks myöhemmin on merkityksetöntä, sillä data saapuisi vastaanottajalle liian myöhään. Esimerkkejä reaaliaikaisesta kategoriasta ovat muun muassa videoneuvottelut ja muut sovellukset, jotka vaativat pientä vasteaikaa. (Tung & Walrand, 2003, s. 1) Liikennedata kuuluu myös tähän kategoriaan, mikäli sitä käytetään reaaliaikaisena, lyhyitä vasteaikoja edellyttävissä tarkoituksissa, kuten esimerkiksi autonomisia ajoneuvoja ohjaavana tietona.

Liikenteen reaaliaikainen tieto on historiatietoa tarkempaa, sillä siinä otetaan huomioon juuri tarkasteluhetkellä vallitseva liikennetilanne. Historiatiedolle tyypilliset keskimääräiset arvot voivat vaihdella huomattavasti esimerkiksi tietyn tapahtuman, kuten satamaan saapuvan laivan, onnettomuuden tai yleisötapahtuman seurauksena. Reaaliaikainen tieto perustuu liikenteen nykytilan havainnointiin erilaisten sensorien avulla. (ESRI, 2020)

Linturin & Kuittisen (2014, s. 34) mukaan nykyinen liikennedata ei ole kuitenkaan suurimmilta osin kriittistä reaaliaikaisuus- tai luotettavuusvaatimuksiltaan, sillä dataan liittyy viiveitä, se ei ole välttämättä erityisen tarkkaa ja sitä käytetään suuntaa antavasti. Autonomisten ajoneuvojen käyttämä data tulee tosin olla käytännössä virheetöntä ja mahdollisimman reaaliaikaista, sillä pienetkin virheet karttatiedoissa tai viiveet

tiedonsiirrossa voivat aiheuttaa vahinkoja. Tämän lisäksi tietoturvallisuuden ja tietosuojan on kiinnitettävä erityistä huomiota.

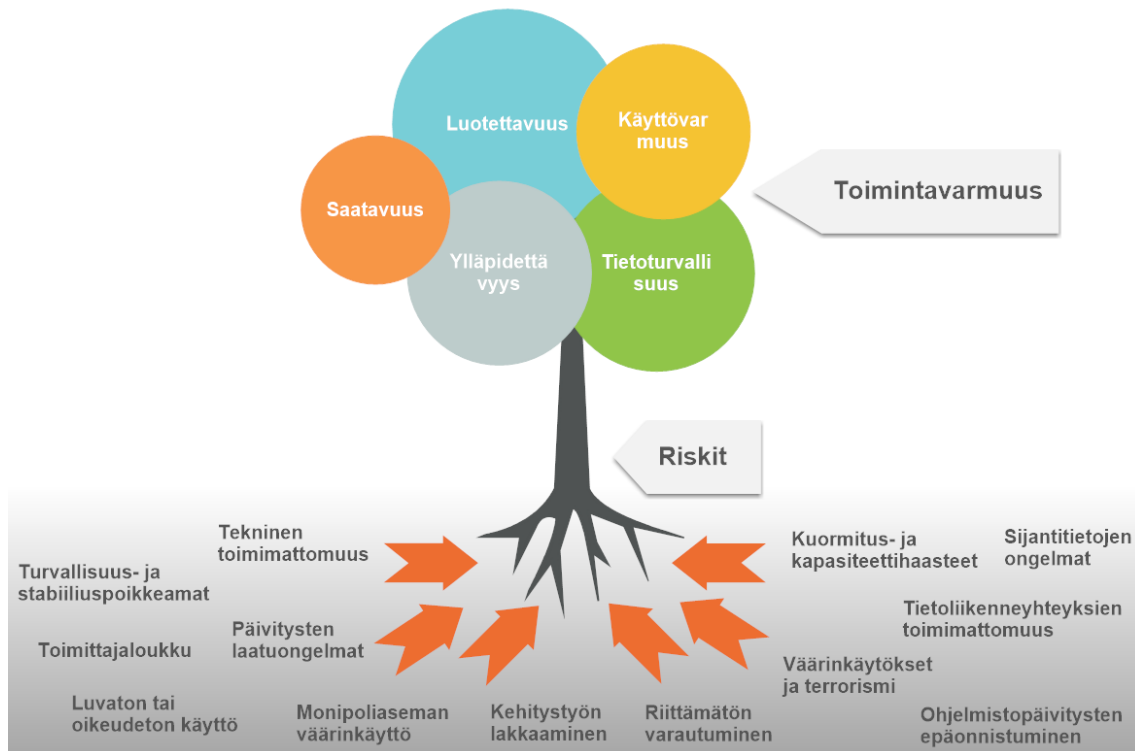
Eryteisesti ajoneuvojen ja matkustajien välittämä tarkka ja visuaalinen verkkoon siirtyvä tieto rasittaa tietoliikenneverkkoa, etenkin ruuhkaisilla alueilla. Tästä syystä raaka liikennedata tulisi käsitellä ajoneuvossa, jotta verkon välityksellä siirrettäisiin mahdollisimman vähän informaatiota normaalitilanteissa. Viiveetöntä ja luotettavuudeltaan korkeaa tietoliikennettä käyttäville toimijoille tulee pyrkiä takaamaan riittävän laadukkaat yhteydet, jotka voidaan kahdentaa tietoturvallisuuden parantamiseksi ja erityisesti ruuhkaisilla alueilla riittävän tiedonsiirtokapasiteetin turvaaminen korostuu. (Linturi & Kuittinen, 2014, s. 34)

Linturi & Kuittinen (2014, ss. 74-76) täydentävät, että reaaliaikaisuus asettaa yleisellä tasolla järjestelmän toimintavarmuudelle erityisiä vaatimuksia ja sen arkkitehtuurin on oltava riittävän vikasietoinen. Toimintavarmuus koostuu luotettavuudesta, käyttövarmuudesta, turvallisuudesta, ylläpidettävyydestä, saatavuudesta ja tietoturvallisuudesta. Liikennedataan ja tietojärjestelmiin liittyy tyypillisiä riskejä, joita ovat muun muassa

- tekninen toimimattomuus
- kehitystyön estyminen tai pysähtyminen teknisistä tai kaupallisista syistä
- alttius väärinkäytöksille ja terrorismille
- yksi toimittaja välttämättömissä toiminnoissa, esimerkiksi tietoliikenneyhteys
- pilvipalveluiden kuormitus- ja kapasiteettiriskit
- sijaintitietojen väärinkäyttötilanteet ja kerääminen yhteen sijaintiin
- ohjelmistopäivitysten virheet ja epäsopivuudet
- dynaamisten päivitysten laadunvarmistus
- käyttäjien autentikointi ja käyttöoikeudet
- tietoliikennekapasiteetti, erityisesti kauko-ohjauksen ja etävalvonnan kannalta
- sovellusten ja palveluiden rajapintastandardien turvallisuus ja stabiilius
- jatkuvuussuunnittelun ja varautumisen riittämätön taso
- dataoperaattorin monopoli

Yleisellä tasolla toimintavarmuuteen vaikuttaa myös muun muassa palvelunlaatu (*engl. Quality of Service, QoS*) (Mohsin ym., 2001, s. 1). Liikennedataan ja tietojärjestelmiin tyypillisesti liittyvät riskit on esitetty kuvassa 3.

Kuva 3. Liikennedataan ja tietojärjestelmiin tyypillisesti liittyvät riskit (mukaellen Linturi & Kuittinen, 2014, ss. 74-76).



Palvelunlaatu on perimmäisessä merkityksessään johdonmukaista ja ennustettavaa tiedonsiirtoa. Reaaliaikaista tiedonsiirtoa edellyttävät järjestelmät ja sovellukset toimivat oikein ainoastaan silloin, kun tiedonsiirto on riittävän ajantasaista. Riittävään ajantasaisuuteen vaikuttavat muun muassa viiveet (*engl. latency*) sekä viiveiden variaatio (*engl. jitter*). (Mohsin ym., 2001, s. 7)

Reaaliaikaisen tiedonsiirron vaatimuksiin voidaan vastata erinäisin teknisin menetelmin, kuten virtuaalisten reititysjonon avulla (Tung & Walrand, 2003, s. 2). Näiden menetelmien tarkempi kuvaus on rajattu tämän työn laajuuden ulkopuolelle.

Suomessa 5G-verkko laajenee koko maan kattavaksi arviolta vuoteen 2025 mennessä (Mikrobitti, 2020). Perinteiset, jo laajasti käytössä olevat langattomat verkot ovat

mahdollistaneet yksittäisten laitteiden ja verkon välisen viestinnän useisiin käyttötarkoituksiin riittävällä tavalla. Yksittäisten laitteiden keskinäinen viestintä on suuressa mittakaavassa mahdollista 5G:n avulla ja tämä mahdollistaa muun muassa autonomisten ajoneuvojen ja IoT-perustaisten (*engl. Internet of Things, esineiden internet, IoT*) sensoripohjaisten sovellusten laajamittaisen ja tehokkaan hyödyntämisen. (Carlson, 2020, s. 725)

Eräs 5G:n eroavaisuus nykyisin laajasti käytössä olevaan 4G-tekniikkaan on sen huomattavasti suurempi tiedonsiirtonopeus, sen ollessa tyypillisesti satoja kertoja edeltäjiänsä nopeampi. Liikennedatankin kannalta toinen hyvin oleellinen eroavaisuus on 5G:n huomattavasti matalampi latenssi sekä korkeampi toimintavarmuus. Näihin eroavaisuuksiin vaikuttavat muun muassa 5G-tekniikan käyttämä edeltäjiään suurempi taajuusalue. (Carlson, 2020, s. 725)

On arvioitu, että 5G kykenee suoriutumaan suurenkin laitetiheyden nopeasta tiedonsiirrosta alueilla, joilla on miljoona laitetta neliökilometrillä. Tämä mahdollistaa ”*älykkään kaupungin*” edellyttämien langatonta tiedonsiirtoa käyttävien anturien laajamittaisen käytön informaation keräämisessä reaaliajassa. 5G:n edistysaskeleet ovat hyödynnettävissä myös liikennetiedon keräämiseen. (Carlson, 2020, ss. 725-726)

Liikenne- ja viestintävirasto Traficom (2020) mukaan liikenteen ja liikkumisen digitalisaatio tapahtuu 5G:n avulla ja se mahdollistaa sovellusten edellyttämän langattoman tiedonsiirron nopeuden, luotettavuuden ja pienen viiveen. Langattomassa 5G-verkossa liikennejärjestelmään tuotetaan informaatiota anturien avulla, joka käsitellään tekoälyyn pohjautuen fyysisen infrastruktuurin, tietoliikenneverkkojen, tiedon ja palveluiden kehittämiseksi ja optimoimiseksi.

3.3 Liikennetiedon kerääminen

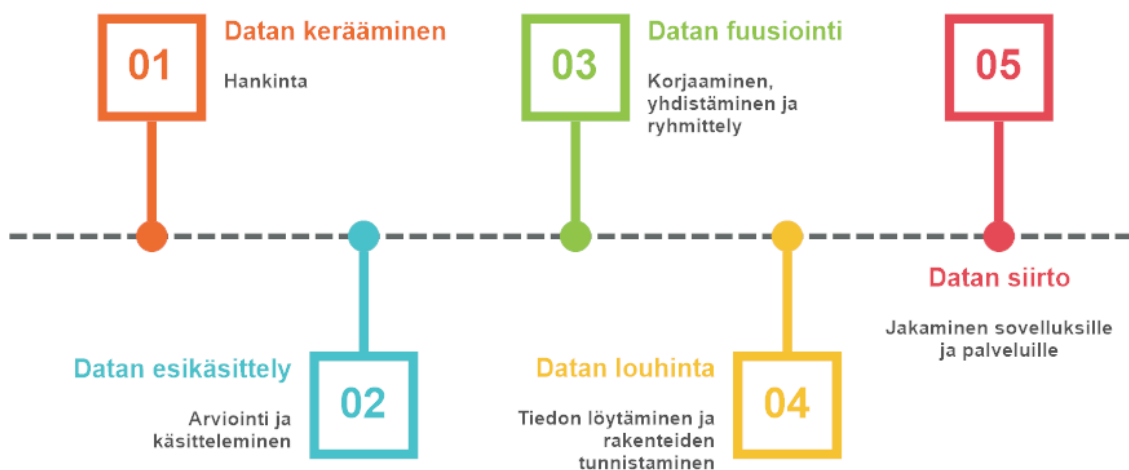
Liikenteestä kerättävä tieto ja sen perusteella laadittavat ennusteet ovat tiestön ja liikenteen suunnittelun ja kehittämisen perusvaatimuksia. Se on myös olennainen osa kuvailevaa kansantaloustiedettä ja sen avulla järkevän liikennepolitiikan toteuttaminen on mahdollista

ihmisten ja tavaroiden liikkumisen sekä julkisen ja yksityisen sektorin kannalta. (NPRA, 2004, s. 12).

Korkealaatuinen ja ajantasainen liikennedata ovat edellytyksiä reaaliaikaisten liikenteenhallinta- ja informaatiojärjestelmien toiminnalle. Aikasarjamuotoista liikennedataa voidaan kerätä runsaasti muun muassa ilmaisimista, liikennevaloista, matkustustiedoista ja liikenteen seurannasta. Tämän tiedon avulla on mahdollista parantaa päätöksentekoa liikennejärjestelmäsuunnittelussa, liikenteenhallinnassa, reittiopastuspalveluissa sekä kaupunkisuunnittelussa. (Lopes ym., 2010, s. 216) Tarkan liikennemäärädatan saaminen on edellytyksenä myös liikenteen operointiin ja suunnitteluun (Suh ym., 2015, s. 2092).

Tällaisen tiedon käyttöön ja hallintaan liittyy kuitenkin haasteita, erityisesti reaaliaikaisissa sovelluksissa ja reaaliaikaisuutta edellyttävissä käyttötarkoituksissa. Liikennedatan kerääminen on oleellinen osa liikennedatan hyödyntämisprosessia, jossa on useita vaiheita. Vakaan, yhtenäisen ja käyttökelpoisen datan muodostus sisältää datan keräämisen, käsittelyn ja yhdistämisen, jolloin data on hyödyntämiskelpoista. (Lopes ym., 2010, s. 216) Liikennedatan hyödyntämisprosessi vaiheineen on esitetty kuvassa 4.

Kuva 4. Liikennedatan hyödyntämisprosessi vaiheineen (mukaellen Lopes ym., 2010, s. 216).



On tärkeää, että mahdollisimman suurta osaa saatavilla olevasta liikennedatasta pystytään hyödyntämään ja tämä on haasteellista niin järjestelmätoimittajien kuin datan käsittelymenetelmienkin kannalta. Liikenteen ja ruuhkautumisen arvioimisessa ja ennustamisessa pyritään käyttämään monipuolisia, toisiinsa näennäisesti liittymättömiä

datalähteitä, joista saadaan korkeatasoista ja käyttökelpoista informaatiota erinäisten kehittyneiden algoritmien ja menetelmien avulla. (Lopes ym., 2010, s. 216) Näiden algoritmien ja menetelmien tarkempi kuvaus on rajattu tämän työn laajuuden ulkopuolelle. Datan jakamisprosessi on kuvattu avoimen liikennedatan osalta tarkemmin alaluvussa 5.1.1.

Liikenne- ja matkustusmäärää sekä jalankulun ja pyöräilyn osuutta kuvaavan datan kerääminen on toimivan kaupunkiliikennejärjestelmän tutkimisen kannalta välttämätöntä. Eri liikkumismuotoja koskeva informaatio määrittelee lähtötiedot liikennettä koskeville hankkeille tieverkoston eri osa-alueilla. Matkustusmäärät toimivat alkutietoina suunniteltaessa toimivaa joukkoliikennejärjestelmää tai analysoitaessa sitä. Kävelyä koskevaa informaatiota voidaan hyödyntää esimerkiksi erilaisissa sosiologisissa tutkimuksissa ja markkinointitutkimuksissa. Liikennejärjestelmää koskevan informaation kerääminen on myös tärkeä osa kaupunkien liikennettä ennustavien mallien rakentamista. (Yakimov, 2018, s. 882)

Älykkäiden liikennejärjestelmien (*engl. Intelligent Transportation System, ITS*) kehittämisen edellytyksenä on reaaliaikainen ja riittävän korkealaatuinen liikennedata. Perinteisiä liikennetiedon lähteitä, kuten induktiosilmukoita tarvitaan osana keräämisen kokonaisuutta, mutta ne eivät kattavuuden puutteen sekä asennus- ja ylläpitokustannusten takia ole itsessään riittäviä. Uudet datalähteet, kuten kelluva ajoneuvodata (*engl. Floating Car Data, FCD*) ovat perinteisiä, kiinteitä lähteitä kustannustehokkaampia ja niillä voidaan välttää joitakin perinteisten menetelmien heikkouksia. (Leduc, 2008, s. 2)

Aikaleimatun sijaintidatan suuren mittakaavan kerääminen edellyttää tiettyjä varastointiin, tutkimiseen ja raportointiin liittyviä metodologioita. Tämän lisäksi haasteita liittyy

- näytteenottoon ja datan tarkkuuteen
- keräämisen jälkeiseen tilastolliseen analysointiin
- tietosuojaan
- immateriaalioikeuksiin
- varastointiin

(BITRE, 2014, ss. 11-12)

Riittävät sijaintidatan tarkkuusvaatimukset riippuvat käyttötarkoituksesta niiden ollessa erityisen suuria esimerkiksi nopeuden määrittämisessä. Tarkasteltaessa esimerkiksi GPS-pohjaista sijaintitietoa sen tarkkuuteen vaikuttavat näytteenottotaajuuden lisäksi muun muassa alueen fyysinen infrastruktuuri, kuten korkeat rakennukset. Sijaintitiedon lähteiden lukumäärä tietyllä alueella voi myös aiheuttaa haasteita, eikä esimerkiksi kahta erillistä signaalia yhdessä ajoneuvossa voida välttämättä erottaa kahdesta signaalista vierekkäisissä ajoneuvoissa. Tähän puolestaan vaikuttavat muun muassa signaalia lähettävien laitteiden käyttämät tiedonsiirto-operaattorit. (BITRE, 2014, s. 11)

Keräämisen jälkeiset analysointihaasteet voivat aiheutua muun muassa sijaintitiedon yhdistämisessä paikkatietoon eli alueen fyysiseen infrastruktuuriin. GPS-pohjaiset laitteet ovat useinmiten myös yksittäisten henkilöiden omistuksessa ja täten niiden datan käyttöön saattaa liittyä tietosuojateknisiä haasteita. (BITRE, 2014, s. 12) Esimerkiksi älylaitteen yksilöivä MAC-osoite on sijaintitietoon yhdistettäessä henkilötietoa (Finlex, 2015).

Varastointiin liittyvät haasteet johtuvat tyypillisesti datan vaatimasta tallennustilan tarpeesta. Datan tallennukseen voidaan käyttää useita menetelmiä, joita ovat muun muassa datan säilyttäminen eri sijainneissa ja yhdistäminen siirtovaiheessa. (BITRE, 2014, s. 12)

Teknologiayritysten uusien datankeruumenetelmien testaaminen on kohdistunut pääasiassa henkilökohtaisiin ajoneuvoihin ja raskaaseen liikenteeseen johtuen niiden suuresta taloudellisesta potentiaalista. Julkisen sektorin laajasti käytetty ja ylläpidetty joukkoliikenne sekä toimintaympäristö ovat kuitenkin myös hyvä testikohde teknologisille kokeiluille. Joukkoliikenteen kustannusten muodostuessa suurelta osin operointikustannuksista se on myös houkutteleva kohde liikenteen automatisaation kannalta. (Polzin, 2016, ss. 11-12)

Liikennetiedon keräämisen eräänä haasteena on sen edellyttämät ilmaisimet, jotka havaitsevat tiellä liikkuvat ajoneuvot. Riittävän kattavia, luotettavia ja tehokkaita ilmaisinjärjestelyjä on tavallisesti vain pienellä osalla teistä (Feng ym., 2014, s. 160). Seuraavassa luvussa esitetään erilaisia ilmaisimia ja menetelmiä, joiden avulla liikennetiedon kerääminen on mahdollista.

4 Liikennetiedon lähteet

Liikenteen hallintaan on käytetty ajan kuluessa erilaisia ilmaisimia, joiden avulla liikennettä on voitu ohjata halutulla tavalla. Perinteisten tiellä sijaitsevien ilmaisimien, kuten induktiosilmukoiden käytön lisäksi liikenteen havainnointitavat ovat kehittyneet liikennedatan keräämiseen, laskentaan ja tiedonsiirtoon soveltuviksi. (Lopes ym., 2010, ss. 216-217)

Liikennedatan keräämiseen on olemassa lukuisia teknisesti erilaisia menetelmiä. Käytettävän menetelmän valintaan vaikuttavat datan tarve ja käyttötarkoitus sekä käytettävissä olevat resurssit. Laajan alueen liikennevirtatieto ei välttämättä edellytä reaaliaikaista informaatiota samaan tapaan kuin esimerkiksi häiriötiedot ja ruuhkautumisilmoitukset. On myös huomioitava, että erilaisia tietolähteitä voidaan yhdistää tiedon kattavuuden parantamiseksi ja tiedon keräämisen kustannusten pienentämiseksi. (BITRE, 2014, s. 3)

Liikennelaskennan laatu on kriittinen tekijä liikenteen seurannassa ja suunnittelussa. Tarkan liikennelaskentatiedon kerääminen on kuitenkin usein haastavaa suuriliikenteisillä kaupunkialueilla ilmaisimien asentamisen ja ylläpidon kannalta. (Suh ym., 2015, s. 2093)

Liikennettä on perinteisesti laskettu käsinlaskentana ja automaattisin laskimin. Käsinlaskennassa nimensä mukaisesti suoritetaan fyysisesti laskentapaikalla kirjaten ajoneuvojen lukumäärän sekä mahdollisesti ajoneuvotyyppin laskentaa. Risteysalueella ajoneuvojen kulkusuuntaa voidaan myös laskea. Laskennan jälkeen kerätty tieto muutetaan manuaalisesti haluttuun muotoon. Automaattiseen laskentaan ja liikenteen mittaamiseen soveltuvia teknologioita ja menetelmiä on lukuisia ja ne soveltuvat eri käyttötarkoituksiin vaihtelevasti. (NPRA, 2004, ss. 13-14).

Kävelyn ja pyöräilyn merkitys on lisääntynyt viime vuosina runsaasti tämän työn kohdealueella (Helsingin kaupunki, 2019c, s. 7). Tästä syystä myös tarve tarkalle ja ajantasaiselle kävely- ja pyöräilydatalle on kasvanut. Näille kulkumuodoille on ominaista maastoriippuus, sääolojen vaikutus kulkureittiin sekä keskimääräisesti lyhyemmät matkat verrattuna moottoroituun liikenteeseen. (Lee & Sener, 2019, s. 2)

Näiden kulkumuotojen datan kerääminen on perinteisesti rajoittunut paikkalaskentoihin ja liikkumiskyselyihin. Koska tällaiset keräysmenetelmät vaativat verrattain paljon ihmistyötä ja aikaa, on data yleisesti ollut pieniotoksista ja harvoin päivitettyä. Paikkalaskennoissa voidaan käyttää fyysistä läsnäolevaa henkilöä suorittamassa laskentaa, videokuvaan perustuvaa jälkikäteistä tai automaattista laskentaa sekä erilaisia ilmaisimiin perustuvia laskentatapoja, kuten infrapuna- tai induktioilmaisimia. (Lee & Sener, 2019, s. 2)

Kävelyn ja pyöräilyn määrää voidaan myös karkeasti arvioida suojateiden painonappien käytön perusteella. Henkilön suorittaman paikkalaskennan vahvuus on lähes sadan prosentin kattavuus ja liikkumiskyselyillä saadaan puolestaan usein kattavaa informaatiota, joskin työmäärän ja kustannusten hinnalla. (Lee & Sener, 2019, s. 2)

4.1 Tietolähteet tyypeittäin

Leduc (2008, s. 3) jaottelee liikennetietoa mittaavat teknologiat kahteen kategoriaan: fyysisesti tiealueelle asennettaviin ja etäältä hallittaviin lähteisiin. Tiealueelle asennetaan tyypillisesti ilmaisimien ja dataa tallentava laite, etähallittavat teknologiat perustuvat etähavaintoihin.

Jaottelu voidaan tehdä myös jakamalla ilmaisimien- ja havainnointityypit perinteisiin, jo pitkään käytössä olleisiin menetelmiin ja uusiin teknologioihin. Perinteiset menetelmät kattavat erilaiset tiealueelle asennettavat fyysiseen havaintoon perustuvat ilmaisimet, liikennekamerat rekisterinumeron tunnistuksineen sekä liikkumista koskevat kyselytutkimukset. Uudet teknologiat ovat langattomia ilmaisinteknologioita, kuten esimerkiksi tietoliikenneteknisiä signaaleja hyödyntävät havainnointimenetelmät sekä erilaiset tutkat. (BITRE, 2014, s. 6)

Liikennedatan kerääminen tietolähteistä voidaan lisäksi kategorisoida kolmeen pääasialliseen menetelmään, jotka ovat paikkadata, kelluva ajoneuvodata ja laaja-alainen data. Jokaisella näistä menetelmistä on erilaiset tekniset ominaisuudet ja toimintaperiaatteet, joihin sisältyvät muun muassa kerätyn liikennedatan rakenne, havaintojen tarkkuus sekä peittoalue. (Lopes ym., 2010, s. 217)

Paikkadata kuvaa tiellä sijaitsevien, eri teknologioin toimivien ilmaisimien keräämää dataa. Tällaisia ilmaisimia ovat muun muassa induktioilmaisimet, paineentunnistavat pneumaattiset putket, pietsosähköiset kaapeli-ilmaisimet sekä tutkat. Näihin menetelmiin kuuluvat myös ultraääneen ja akustiseen havainnointiin perustuvat ilmaisimet, kuten magnetometri-ilmaisimet ja LiDAR. (Lopes ym., 2010, s. 217)

Kelluvaa ajoneuvodataa voidaan ajatella ajoneuvoperustaisena havainnointina. Sillä viitataan sijaintitiedon keräämiseen paikantamalla ja tunnistamalla ajoneuvoja liikenneverkon eri alueilla joko fyysisillä tai virtuaalisilla ilmaisimilla. (Lopes ym., 2010, s. 217)

Erityisesti kävelyyn ja pyöräilyyn soveltuvat liikennetiedon lähteet voidaan myös luokitella kategorioihin. Lee & Sener (2019, s. 3) jakavat tällaiset ilmaisim- ja havainnointityypit perinteisiin tietolähteisiin ja uudempiin menetelmiin. Perinteiset tietolähteet jaotellaan laskentamenetelmiin ja liikkumiskyselyihin. Uudemmat menetelmät koostuvat vain kävelyyn ja pyöräilyyn tarkoitetuista menetelmistä sekä kaikille liikkumismuodoille soveltuvista menetelmistä. Luokittelu kävelyn ja pyöräilyn tietolähteistä on esitetty kuvassa 5.

Kuva 5. Kävelyn ja pyöräilyn tietolähteet luokittain (mukaellen Lee & Sener, 2019, s. 3).



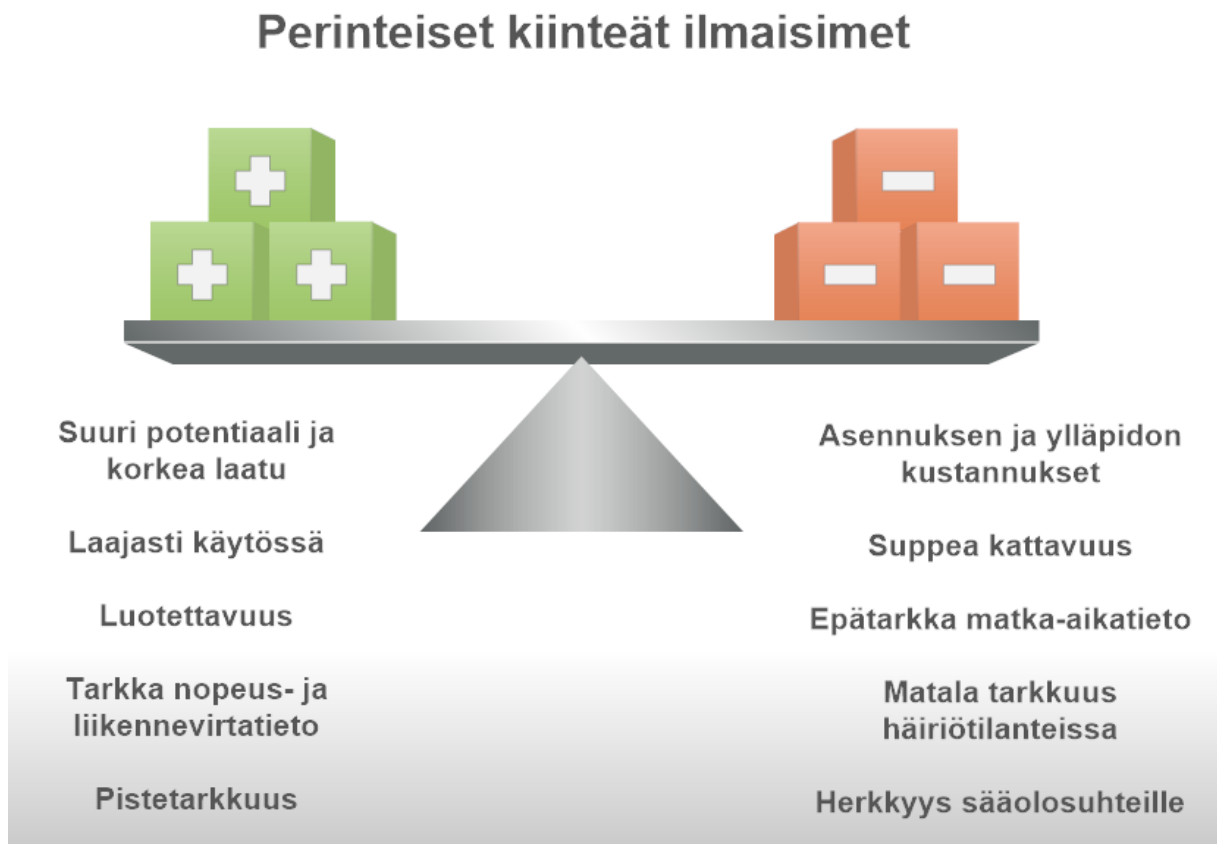
Kokonaisuutena liikennetietoa voidaan saada lukuisista erilaisista lähteistä lukemattomin erilaisin menetelmin. Seuraavissa alaluvuissa käsitellään tämän työn tavoitteiden kannalta oleellisimmiksi arvioituja tapoja. Erilaisia liikennetiedon lähteitä ominaisuuksineen on esitelty tyypeittäin seuraavissa alakappaleissa jaoteltuna perinteisiin liikennetiedon lähteisiin sekä uudempiin ilmaisinteknologioihin ja havainnointimenetelmiin. Jokaisen tietolähteen osalta esitetään sen toimintaperiaate lyhyesti sekä huomioita muun muassa ilmaisintyyppin soveltuvuudesta liikennedatan keräämiseen. Erityistä huomiota on kiinnitetty työn tavoitteiden kannalta potentiaalisimmiksi arvioituihin, uudempiin ilmaisimiin ja teknologioihin.

4.2 Perinteiset liikennetiedon lähteet

Perinteisten ilmaisinteknologioiden tehokkuutta rajoittavat tavallisesti ilmaisimien vähäinen peittoalue ja sijoittelu. Havaintojen siirtäminen ilmaisimilta erilaisiin järjestelmiin saattaa edellyttää myös työläitä toimenpiteitä. (Gu ym., 2016, s. 321) Fyysisesti tiealueelle asennettavat ilmaisimet ovat olleet teknologiana käytössä jo pitkään ja ne tuottavat todistetusti tarkkaa ja relevanttia tietoa vallitsevasta liikennetilanteesta esimerkiksi nopeus- ja liikennevirtatiedon muodossa. Julkinen sektori hyödyntää tällaisia ilmaisimia laajasti ja niiden tuottama data tulee olemaan avainasemassa myös tulevaisuudessa. (Leduc, 2008, s. 8)

Fyysisesti tiealueelle asennettavat ilmaisimet muodostavat kuitenkin kustannuksia johtuen infrastruktuuriin vaadittavista asennuksista ja ylläpidosta. Kiinteän luonteensa vuoksi ne eivät myöskään pysty kattamaan suuria alueita ilman massiivisia asennuksia. (Leduc, 2008, s. 8) Kokonaisuutena perinteisiin, kiinteisiin ilmaisimiin liittyy hyötyjä ja haittoja, jotka on esitetty kuvassa 6.

Kuva 6. Kiinteiden ilmaisimien hyödyt ja haitat (mukaellen Leduc, 2008, s. 8).



4.2.1 Induktioilmaisimet sekä muut tiehen asennettavat ilmaisimet

Induktioilmaisimet eli *silmukat* ovat perinteisin tapa kerätä liikennedatata. Menetelmää käytetään laajasti Euroopassa ja se soveltuu säänkestonsa ansiosta hyvin myös Suomen olosuhteisiin. Kustannukset muodostuvat suurilta osin asennuksen edellyttämistä töistä ja ylläpidosta vikatilanteissa. (Leduc, 2008, s. 3) Induktioilmaisimien vikatilanteet ovat verrattain yleisiä ja niiden kiinteä, maahan kaivettava asennustapa lisää ylläpidon kustannuksia (Houbraken ym., 2018, s. 2).

Induktioilmaisimien koostuu maahan, esimerkiksi ajorataan upotetusta metallisilmukasta, johon johdetaan oskilloivaa jännitettä magneettikentän muodostamiseksi. Pääosin metallista koostuvan ajoneuvon ylittäessä ilmaisimen magneettikenttä muuttuu ja ajoneuvo voidaan tällöin havaita. Menetelmä on herkkä ulkoisille vaurioille, kuten kaapelin tai liittimien vioittumiselle ja niiden asentamisessa täytyy ottaa huomioon lähettyvillä olevat suuret metallimassat, kuten siltarakenteet tai kaivonkannet. (Dalglish & Hoose, 2008, ss. 123-124)

Jotta ilmaisin tunnistaa tiellä liikkuvan ajoneuvon, täytyy ajoneuvon koostumuksen olla riittävän induktiivinen eli metallia sisältävä. Käytännössä tunnistustarkkuus riittää moottoripyörien havaitsemiseen, joskin pyöräilijöiden havaitseminen ei ole aina mahdollista. Ilmaisua ohjaavien laitteiden säätömahdollisuudet mahdollistavat ilmaisuherkkyden säätämisen ja ilmaisimien sijoittelulla ja asennettavan ilmaisinsilmukan muodolla voidaan myös vaikuttaa ilmaisuherkkyteen ja havaitsemistehokkuuteen. (Dagleish & Hoose, 2008, ss. 123-124)

Automaattisten liikennelaskimien toiminta perustuu yleensä induktioilmaisimiin, muihin tiehen asennettaviin ilmaisimiin tai tutka- ja infrapunailmaisimiin. Erilaisia langattomia laskentamenetelmiä on myös lukuisia. (NPRA, 2004, s. 14).

Dagleish ja Hoose (2008, ss. 121-122) määrittelevät liikennelaskimilla tehtävät mittaukset kahteen luokkaan: määrällisiin ja kategorisiin mittauksiin. Määrällinen luokka sisältää numeerisia arvoja, kuten mittauspisteen ohittavien ajoneuvojen nopeuden. Kategorinen luokka kattaa luokkakohtaisia mittaustuloksia, kuten linja-autojen osuutta kaikista ajoneuvoista. Saman määritelmän mukaan liikenteestä kerättävä tieto jaotellaan kolmeen luokkaan: liikennevirtatietoon, yksittäisen ajoneuvon tietoon ja linkkitietoon.

Liikennevirtatieto on jatkuvan liikennevirran tietoa mittauspisteestä ja se esitetään usein aikasarjana. Esimerkkejä liikennevirtatiedosta ovat muun muassa silmukan varausaste, ajoneuvojen välinen etäisyys, keskinopeus ja tietyn persentiilin nopeus. Yksittäisen ajoneuvon tieto on tietoa tietystä ajoneuvosta, kuten esimerkiksi ajoneuvon luokka ja nopeus. Linkkitieto on tietyn liikenneverkon osan tietoa, jota on esimerkiksi keskimääräinen matka-aika. Linkkitieto esitetään myös usein aikasarjana ja sitä hyödynnetään muun muassa ruuhkautumisen mittaamisessa. (Dagleish & Hoose, 2008, ss. 121-122)

Suomessa käytetään laajasti induktioilmaisimiin perustuvia automaattisia liikennelaskimia, jotka tunnetaan myös lyhenteellä LAM. Mittauspisteitä on Suomessa noin 500 ja niistä saadaan avointa liikennedatata raaka- ja raporttimuodossa, jonka jakelusta vastaa Traffic Management Finland. LAM-mittauspiste sisältää tiedonkeruuyksikön lisäksi kaistakohtaiset induktioilmaisimet, jotka havaitsevat ajoneuvojen ajosuunnan, ajokaistan, nopeuden, pituuden, ajoneuvoluokan sekä peräkkäisten ajoneuvojen aikaeron ja havaintoajan. Havaitut

ajoneuvot voidaan jaotella seitsemään luokkaan. Menetelmä soveltuu tällä hetkellä nelipyöräisten ajoneuvojen havaitsemiseen ja kokeiluasteella on myös moottoripyörät havaitseva laitetyyppi. (Väylävirasto, 2020)

Ajoneuvoja voidaan havaita myös paineentunnistukseen perustuvilla pneumaattisilla putkilla. Niiden toiminta perustuu ajoneuvon painon aiheuttamaan ilman liikkumiseen putkessa, joka voidaan muuntaa havainnoksi. Menetelmä soveltuu liikenteen määrälaskentaan ja sen avulla voidaan myös tunnistaa ajoneuvoluokka. Toistuvan fyysisen rasituksen myötä ilmaisimet saattavat kuitenkin vikaantua ja huomiota on kiinnitettävä ilmaisinsijoitteluun sekä ylläpitoon. (NPRA, 2004, s. 14). Leduc (2008, s. 3) huomauttaa myös, että pneumaattiset putket ovat herkkiä lämpötilan muutoksille ja niiden tarkkuus ei välttämättä ole riittävä alhaisilla nopeuksilla.

Tiealueelle asennettavilla liikkeessä painon tunnistamiseen perustuvilla ilmaisimilla voidaan havaita ilmaisinalueen ylittävien ajoneuvojen määrä, paino ja ajoneuvoluokka. Tällaiset toimintaperiaatteiltaan erilaiset ilmaisimet saavat havaintoja perustuen ajoneuvojen painoon. (NPRA, 2004, s. 15).

Mitattavalle alueelle voidaan asentaa ajoneuvojen akselien painosta venyviä liuskoja. Liuskoihin lähetettävät sähköiset signaalit muuttuvat ajoneuvojen painon vaikutuksesta, jolloin ne voidaan havaita. Havaintoja voidaan saada myös kapasitiivisen anturin avulla, jonka toiminta perustuu ajoneuvon painon aiheuttamaan kapasitanssin muutokseen ilmaisimessa. Tämä menetelmä soveltuu myös väliaikaiseen mittaamiseen kapasitiivisen anturimaton muodossa, joka on helposti siirrettävissä lyhytkestoisia mittauksia varten. Myös pietsosähköinen kaapeli-ilmaisim perustuu ajoneuvon painon aiheuttamaan jännitemuutokseen ilmaisimessa. Se on ominaisuuksiltaan hyvin samankaltainen kiinteän induktioilmaisimen kanssa ollen verrattain taloudellinen ja huoltovapaa ilmaisim. (NPRA, 2004, s. 15).

Pietsosähköisen kaapeli-ilmaisimen toiminta perustuu mekaanisen energian muuntamiseen sähköiseksi havainnoksi. Pietsosähköisen materiaalin muovautuminen mekaanisesta rasituksesta, kuten ajoneuvon painosta aiheuttaa ilmaisimen pinnan jännitetiheyden muutoksen, jolloin elektrodien välille muodostuu potentiaaliero. Havainnon voimakkuus

riippuu muutoksen suuruudesta ja taajuudesta, joten ilmaisimella voidaan havaita ajoneuvon nopeus ja paino eli ajoneuvoluokka. Ilmaisimien asennus toteutetaan tiealueelle tehtäviin uriin mitattaville kaistoille. (Leduc, 2008, s. 3)

Tällaisia kiinteästi tiehen asennettuja ilmaisimia käytetään liikenteen määrälaskentaan, jonka avulla voidaan laskea esimerkiksi erinäisiä päivä-, viikko- tai kuukausikohtaisia liikenteen keskiarvoja sekä havaita muutoksia ja trendejä liikenteen määrässä eri ajanjaksoina. Jatkuvat määrälaskennat ovat oleellisia liikenteen suunnittelutarkoituksiin, sillä liikennemäärään perustuvat huipputuntilaskelmat määrittellään usein tällaisten kiinteiden ilmaisimien havaintojen perusteella. (Suh ym., 2015, s. 2092)

On myös huomioitava, että ilmaisinjärjestelyt vaikuttavat saatavan liikennetiedon määrään. Yksittäisten ilmaisimien avulla voidaan havaita ajoneuvojen lukumäärä, mutta esimerkiksi ajoneuvon nopeustieto tai ajoneuvon pituuden määrittäminen edellyttää vähintään kahta ilmaisinta riittävän etäällä toisistaan. (NPR, 2004, s. 14).

4.2.2 Tutka- ja infrapunailmaisimet

Tutkailmaisimien toiminta perustuu laitteen lähettämään säteeseen, joka kimpoaa mittauspisteen ohi ajavasta ajoneuvosta takaisin laitteeseen. Tällä menetelmällä pystytään laitteesta riippuen havaitsemaan ajoneuvojen lukumäärä, nopeus, kulkusuunta sekä ajoneuvojen pituus eli ajoneuvoluokka. Infrapunailmaisimet havaitsevat mittausalueen lämpötilaeroja, jolloin lämpimämmät kohteet, kuten ajoneuvot voidaan tunnistaa. Ilmaisimet soveltuvat myös jalankulkijoiden liikkeen tunnistamiseen. (Dalglish & Hoose, 2008, ss. 124-125)

Tämän tyyppisten ilmaisimien haasteena on ilmaisualueen katveet eli ilmaisimen havaintoalueen estyminen esimerkiksi suuren linja-auton toimesta. Kaistakohtaisissa ilmaisuissa viereisten kaistojen liikenne saattaa myös aiheuttaa virheellisiä heijasteilmaisuja mitattavalle alueelle. Näihin haasteisiin voidaan vaikuttaa ilmaisimien sijoittelulla, korkeuden ja asennuskulman ollessa merkitseviä tekijöitä. (Dalglish & Hoose, 2008, ss. 124-125)

Infrapunailmaisimet voidaan jakaa kahteen kategoriaan, aktiivisiin ja passiivisiin ilmaisimiin. Niiden haasteina ovat matala havainnointikyky haastavissa sääolosuhteissa sekä kapea

mittausalue. (Leduc, 2008, s. 3) Passiiviset infrapunailmaisimet havaitsevat lämpötilan muutoksia ja ne tunnetaan paremmin liiketunnistimen nimellä. Aktiiviset infrapunailmaisimet lähettävät silmälle näkymättömän säteen, jonka katketessa muodostetaan havainto. Menetelmällä voidaan havaita kaikki lämpöä säteilevät kulkumuodot, mikäli lämpötilaero ympäristöön on riittävän suuri. (Diamond Traffic Products, 2020)

4.2.3 Kameratekniikkaan perustuvat ilmaisimet

Kameraperustaisten ilmaisinteknologioiden avulla ajoneuvoista voidaan havaita lukumäärä, ajoneuvoluokka ja nopeus sekä käytettävästä teknologiasta riippuen myös esimerkiksi pysähtymisviivan ylitys. Menetelmä on altis säteen ja valon vaikutuksille, niiden vaikuttaessa negatiivisesti havaintoihin. (Leduc, 2008, s. 4)

Kamerateknologiaan perustuvien ilmaisimien sijainti on merkitsevä havaitsemistarkkuuden kannalta, sillä auringonvalo, varjot ja muut visuaaliset häiriötekijät vaikuttavat havaintoihin negatiivisesti. (Dalgleish & Hoose, 2008, s. 125) Kvantunnistukseen perustuvien ilmaisimien haasteena ovat myös tiheästi liikennöidyt alueet ja sääolosuhteiden aiheuttamat ongelmat. Tämän lisäksi katvealueet ja valaistusolosuhteet vähentävät havaintojen tarkkuutta. Ilmaisimien tarkkuutta voidaan pitää hyväksyttävänä, kun niiden tarkkuus on 10 % etäisyydellä käsinlaskentana suoritetusta kontrollilaskennasta. Menetelmän luotettavuus on hyvin riippuvainen käyttöolosuhteista ja asennustavasta, johon voidaan vaikuttaa asennusteknisin ja ylläpidollisin keinoin. (Suh ym., 2015, s. 2093)

Menetelmään käytettävät kameralaittejärjestelmät ovat monesti hyvin heterogeenisiä ja niiden ominaisuudet, kuten resoluutiot, ruudunpäivitysnopeudet sekä asennuspaikat ja -tavat vaihtelevat suuresti. Kameravalvontaa suoritetaan myös tyypillisesti jatkuvasti, joka tuottaa hyvin suuria datamääriä. Heterogeeninen laitekanta tuottaa tarkkuudeltaan ja sisällöltään hyvin vaihtelevaa videomuodossa olevaa dataa, jonka ohjelmallinen tulkinta on täten haastavaa. Suurten datamäärien tiedonsiirto sekä tallennuksen vaatima tallennustila aiheuttavat myös haasteita. (Fedorov ym., 2019, ss. 1-2)

4.3 Uudemmat ilmaisinteknologiat ja havainnointimenetelmät

Markkinoille on saapunut ajan kuluessa perinteisiä ilmaisinteknologioita haastavia liikennetiedon lähteitä. Niiden avulla pystytään välttämään kiinteille ilmaisimille tyypillisiä rajoitteita ja haasteita. (Leduc, 2008, s. 8)

4.3.1 LiDAR

LiDAR eli valotutka (*engl. Light Detection and Ranging*) on aktiivinen ilmaisinteknologia. Aktiivisuudella tarkoitetaan laitteen itse luomaa energiaa, valoa, joka havaitsee etäisyyden päässä olevia kohteita. Toimintaperiaate perustuu nopeisiin laserimpulsseihin, jotka kimpoavat takaisin laitteeseen. Impulssien matkaan kuluvan ajan perusteella voidaan selvittää etäisyys kohteeseen, joka voidaan muuttaa korkeusarvoksi. Laittekokonaisuudessa voidaan hyödyntää GPS-paikannusta sekä erillistä mittayksikköä, jolloin kohteesta saadaan X-, Y- ja Z-koordinaatit, jotka muodostavat kolmiulotteisen kuvan. Valoimpulssin aallonpituutta muuttamalla voidaan mitata erilaisia kohteita. (Wasser, 2020)

Wu ym. (2020, s. 1) mukaan LiDAR:n havaitsemiskykyä huonoissa sääolosuhteissa, kuten lumisateella voidaan parantaa hyödyntämällä kehittyneitä datankäsittelyalgoritmeja sekä datan suodatus- ja ryhmittämismenetelmiä. Tällöin ajoneuvot voidaan havaita suhteellisen tarkasti myös Suomen kaltaisissa haastavissa sääoloissa.

Myös tuuliset ja sumuiset olosuhteet heikentävät havaintojen tarkkuutta, mutta häiriöitä voidaan suodattaa taustafiltoinnilla. Tällöin häiriötekijöiden, kuten lumen aiheuttama kohina poistetaan ohjelmallisesti ja siten voidaan saavuttaa jopa 90 % tarkkuus. Menetelmä mahdollistaa lisäksi menneiden havaintojen hyödyntämisen nykytilanteessa ja tämä parantaa datan tarkkuutta merkittävästi. (Wu ym., 2020, s. 3)

Menetelmän etuna on laitteen kyky muodostaa kolmiulotteinen malli ympäristöstä, jonka avulla on mahdollista suodattaa häiriötekijöistä, kuten lumesta aiheutuvia virheellisiä ilmaisuja. Kolmiulotteinen malli mahdollistaa myös etäisyyden ja nopeuden tarkemman mittaamisen. Ulkoisen valaistuksen muutokset eivät myöskään vaikuta laitteen toimintaan lainkaan. Menetelmään liittyviä haasteita ovat korkea hinta ja matala tarkkuus verrattuna

kamerateknologisiin menetelmiin. Monien muiden ilmaisimien tapaan asennustavalla, kuten asennuskulmalla ja -korkeudella on suuri merkitys laitteen toimintakykyyn. (Hokajärvi & Nieminen, 2018, s. 13)

4.3.2 Magnetometri-ilmaisimet

Wang ym. (2020, ss. 1-3) mukaan magnetometri-ilmaisimiin perustuvan ajoneuvojen havainnoinnin toimintaperiaatteessa pääosin eri metalleista koostuvat ajoneuvot aiheuttavat muutoksia maan magneettikentässä liikkeessään mittausalueella. Näiden muutosten perusteella ajoneuvot voidaan havaita 10 metrin säteellä toisistaan. Menetelmän etuina pidetään skaalautuvuutta, matalia rakennus-, ylläpito- ja käyttökustannuksia sekä sietokykyä häiriöille. (Wang ym., 2020, s. 8).

Menetelmä mahdollistaa luotettavan, mukautettavan ja kustannustehokkaan tavan kerätä liikennedatata langattomasti. Ilmaisimien havaitsee ajoneuvot noin 99 % tarkkuudella reaaliajassa ja sen sähköistykseen voidaan käyttää esimerkiksi julkisen valaistuksen sähkönsyöttöä tai aurinkopaneeleita. Induktioilmaisimiin verrattuna magnetometri-ilmaisimet ovat yksinkertaisempia asentaa ja niiden huolto on monia perinteisiä ilmaisintyyppisiä helpompaa johtuen asennustavasta. (Urbiotica, 2020)

Menetelmän avulla liikenteestä saadaan selville ajoneuvojen lukumäärä ajanjaksolla, tien käyttöaste, keskimääräinen nopeus, ajoneuvoluokka, ajoneuvojen välinen aikaero sekä ruuhkautuminen liikenteen tiheyden muodossa. Ilmaisimien havainnot voidaan lähettää esimerkiksi liikennevalojen ohjauskojeisiin tai pilvipalveluun datana ja se soveltuu käytettäväksi myös liikennedatan lähteenä rajapinnan kautta. (Urbiotica, 2020)

Jokainen mittalaittekokonaisuus koostuu magnetometri-ilmaisimesta ja langattoman tiedonsiirron mahdollistavasta moduulista. Ilmaisimia ja moduuleja sijoitetaan lähelle ajorataa tasaisin välimatkoin. (Wang ym., 2020, ss. 2-3)

4.3.3 Koneoppimiseen perustuvat menetelmät

Videokuvasta voidaan havaita ohjelmallisesti paljon liikenteellisesti mielenkiintoista informaatiota, kuten ajoneuvoja. Erilaisia tunnistusmenetelmiä on runsaasti, sillä tunnistuksen kohteesta riippuen on käytettävä juuri siihen tarkoitukseen tarkoitettua menetelmää. (Turán ym., 2020, s. 127) Liikenteen seurantaan tarkoitettujen kamerajärjestelmät ovat tärkeä osa älykästä liikennejärjestelmää. Ne sisältävät tyypillisesti automaattisia liikenteenseurantakameroita, jotka kuvaavat ajoneuvoja ja muita liikkuvia kohteita. Perinteisesti tallennettua materiaalia on käyty lävitse manuaalisesti ja tästä johtuen menetelmä on ollut työläs ja aikaa vaativa. (Tang ym., 2015, s. 5817)

Ajoneuvojen tunnistaminen videokuvasta on haastavaa, sillä kuviin liittyy vääristäviä tekijöitä ja erilaisia ja -muotoisia ajoneuvoja on liikenteessä runsaasti. Käytettyjen kameralaitteistojen ominaisuudet, valaistus ja säätila vaikuttavat myös tunnistustarkkuuteen. (Tang ym., 2015, s. 5819)

Menetelmän toiminta perustuu kahteen vaiheeseen, ajoneuvon havaitsemiseen ja sen tunnistamiseen. Havaitsemisvaiheessa käytetään muun muassa koneoppimisalgoritmeja, joiden perusteella kuvasta voidaan havaita kiinnostava kohde (*engl. region-of-interest, ROI*), kuten ajoneuvo. Tunnistamisvaiheessa hyödynnetään muun muassa esimerkillisiä kuvia tunnistamisen kohteesta ja vääristä vaihtoehtoja, joiden avulla järjestelmä voidaan *opettaa* päättämään havaitun kohteen olevan halutun kaltainen, kuten ajoneuvo. (Tang ym., 2015, s. 5820)

Käsitteenä koneoppiminen on yksi tekoälyn keskeisiä osa-alueita. Sillä on lähes rajattomasti käyttökohteita erilaisista suunnittelutehtävistä ja ongelmanratkaisusta autonomisiin ajoneuvoihin. Ajoneuvojen havaitseminen videokuvasta on niin sanottua heikkoa tekoälyä, jolla tarkoitetaan laitteiden tai ohjelmien kykyä analysoida niille annettu tehtävä annetun logiikan mukaan. Valvoton koneoppiminen voi olla liikenteen kontekstissa esimerkiksi järjestelmän kyky tunnistaa ajoneuvoja itsenäisesti ilman erillistä valvontaa. (tekoäly.info, 2020)

Tekoälyn avulla voidaan käsitellä suuria, erimuotoisia ja epäyhtenäisiä heterogeenisiä datamääriä. Se soveltuu täten esimerkiksi erilaisiin datalähteiden yhdistämistehtäviin, kuten

liikennetiedon ja paikkatiedon avulla muodostettavan liikenteen reaaliaikaisen tilannekuvan luomiseen. Sitä voidaan käyttää myös muun muassa suurten datamäärien syy-seuraussuhteiden ja toistuvien trendien havaitsemiseen ennusteiden luomiseksi esimerkiksi liikenneväylien ylläpidon tarpeisiin. On oleellista, että tekoälyä hyödyntäviin tarkoituksiin saadaan runsaasti laadukasta dataa, jota voidaan saada esimerkiksi erilaisista liikenteen datalähteistä. (Korpela, 2019, s. 12)

Koneoppimiseen liittyvät teknologiat tehostavat useita älykkään liikennejärjestelmän edellyttämiä toimenpiteitä. Tällaisia ovat muun muassa

- liikenneinformaation, kuten liikennevirta-, määrä-, nopeus- ja häiriötiedon kerääminen, jota tarvitaan kattavaan, reaaliaikaiseen ja luotettavaan tilannekuvaan vallitsevasta liikennetilanteesta
- liikenteenhallinnan edellyttämän päätöksenteon parantaminen kerätyn liikennedatan sekä historiatiedon perusteella
- liikenneinformaation levittäminen, jonka avulla tienkäyttäjät saavat reaaliaikaista tilannetietoa liikennetilanteesta ja jota voidaan hyödyntää esimerkiksi liikennesuunnittelussa ja reittiopastuksessa

(Amin-Naseri, 2018, ss. 2-4)

4.3.4 Joukkoistettu data

Joukkoistettu data koostuu liikennedatan tapauksessa tienkäyttäjien vapaaehtoisesti ja tarkoituksella keräämästä datasta, joka perustuu ihmisten havaintoihin vallitsevasta liikennetilanteesta. Tienkäyttäjien havaitsemat poikkeustilanteet, kuten onnettomuudet ja ruuhkautuminen ovat esimerkkejä joukkoistetusta liikennedatasta. (U.S. Department of Transportation, 2019, s. 12)

Tällaista dataa voidaan hyödyntää liikenneinformaation tietolähteenä. Sitä voidaan kerätä lukuisista lähteistä, kuten esimerkiksi sosiaalisesta mediasta ja älylaitteiden reittisovelluksista. Joukkoistettu data on luonteeltaan raakadataa, joten se edellyttää esikäsitteilyä ollakseen hyödynnettävissä. On myös havaittu, että joukkoistettu data sisältää tyypillisesti runsaasti tarpeettomia ilmoituksia samoista häiriöistä. Tällöin voidaan toisaalta

mahdollisesti päätellä häiriön vaikutuksen laajuutta ja vakavuutta. Tällaisten tarpeettomien ilmoitusten reaaliaikainen suodatus on eräs joukkoistetun datan käytön suurimmista haasteista datan luotettavuuden selvittämisen ohella. (Amin-Naseri, 2018, ss. 5-7)

Tällä tavalla kerätty data on myös luonteeltaan hyvin heterogeenistä, sillä dataa kerätään lukemattomista eri lähteistä, joilla kaikilla on oma muotonsa, tarkkuutensa ja rajoitteensa. Moninaisen joukkoistetun datan yhdistäminen merkitykselliseksi, luotettavaksi informaatioksi on hankalaa ja tämä on aktiivisen tutkimuksen kohde tällä hetkellä. (Amin-Naseri, 2018, s. 7)

4.3.5 Joukkoliikennedata

Joukkoliikenteen reaaliaikaista liikennetilannetta koskevia tietoja ovat muun muassa ajoneuvojen sijainti reaaliajassa sekä saapumis- ja lähtötiedot pysäkeiltä (Pihlajamaa ym., 2020, s. 23). Muita joukkoliikenteen reaaliaikaisia tietoja ovat muun muassa liikenneverkon häiriö- ja muutostiedot, joiden merkitys palvelutason ylläpidossa ja joukkoliikenteen käyttöasteen kasvattamisessa on oleellista (Pihlajamaa ym., 2020, s. 52).

Joukkoliikenteen reaaliaikaista sijainti- ja häiriötietoa voidaan hyödyntää liikenteen hallinnassa, matka- ja reitti-informaatiopalveluissa sekä liikennöitsijöiden toimesta häiriö- ja toimivuusongelmien havaitsemisessa. Nämä tiedot ovat oleellisia joukkoliikenteen sujuvuudelle ja seurannalle sekä matkustajille varmistaen reitin suunnittelun kulkuvälineen vaihtoineen ja aikatauluineen. (Pihlajamaa ym., 2020, s. 48)

Laadukkaaseen matkatietoon, kuten esimerkiksi reitti-, aikataulu- ja pysäkkietoon liittyy useita laatutekijöitä. Tällainen tieto on luonteeltaan

- oikeamuotoista ja standardinmukaista
- oikeellista ja täsmällistä
- johdonmukaista
- täydellistä
- yhteentoimivaa

(Pihlajamaa ym., 2020, s. 72)

Pääkaupunkiseudulla Helsingin seudun liikenne (HSL) toimii yhtenä avoimen joukkoliikennedatan tarjoajana ja dataa on saatavilla sen kautta rajapintojen avulla ja datapaketteina. Avoimesti jaettua dataa ovat esimerkiksi reittitiedot, aikataulut, häiriö- ja poikkeustilannetiedot sekä paikkatiedot muun muassa joukkoliikenteen pysäkeistä. (HSL, 2020)

4.3.6 Langattomat signaalit

Langattomiin signaaleihin, kuten Bluetooth:iin perustuvat ilmaisimet ovat verrattain edullisia ja niitä voidaan asentaa joustavasti ja huomaamattomasti ilman kaupunkikuvallisia haittoja. Tästä syystä niitä voidaan käyttää suurissa määrin kaupunkialueilla pienin asennus- ja käyttökustannuksin. Ilmaisiverkko koostuu tyypillisesti eri kohteisiin, kuten liikenne- tai katuvalopylväisiin asennetuista antureista. Niiden avulla voidaan kerätä tietoa liikenteestä ja ympäristöstä jopa reaaliaikaisesti. (Narh ym., 2016, ss. 648-649)

Tällaiset ilmaisimet toimivat verkkovirralla tai akuilla ja niiden avulla saadaan tietoa esimerkiksi lämpötilasta, ilmankosteudesta, ilman hiilidioksidi-, typpi- tai typpioksiditasosta tai melusta. Ilmaisintyyppistä riippuen niiden avulla voidaan mitata myös liikennemäärää, kaistakohtaista tilaa ja ajoneuvon nopeutta. (Narh ym., 2016, ss. 648-649)

Bluetooth-perustainen liikennedatan perustuu MAC-osoitteiden (*engl. Media Access Control*), ainutlaatuisuuteen. Bluetooth-laitteet käyttävät kommunikointiin ja datan siirtoon tiettyä taajuusalueita, 2,45 gigahertsiä. Jokaisella verkkoa käyttävällä laitteella on uniikki, laitekohtainen MAC-osoite, joka liikennedatan kontekstissa yksilöi jokaisen ajoneuvon mitta-alueella. Ajoneuvoissa sijaitsevat Bluetooth-signaalia lähettävät laitteet, kuten älypuhelimet, kannettavat tietokoneet ja ajoneuvo-kohtaiset laitteet havaitaan mitta-alueella ja niistä muodostuu liikennettä kuvaavaa informaatiota. (Narh ym., 2016, ss. 649-650)

Eri mittauspisteiden välillä havaittu tietty, identtinen MAC-osoite mahdollistaa tietyn ajoneuvon seurannan ja täten liikennevirran sujuvuuden, kaistan varausasteen, nopeuden ja matka-ajan mittaamisen. Vastaavaa menetelmää käytetään muun muassa Australiassa, Yhdysvalloissa ja Englannissa. (Narh. ym., 2016, ss. 649-650)

4.3.7 GPS-paikkatieto, GPS-tutkimukset ja matkapuhelindata

GPS-tekniikan yleistyminen ja langattomien tiedonsiirtoteknologioiden kehittyminen mahdollistaa ajoneuvojen toimimisen eräänlaisina luotaimina. Riittävä määrä GPS-seurannassa olevia ajoneuvoja voi toimia luotettavana ja tarkkana matka-ajan tietolähteenä. Tämän lisäksi GPS-perustaisen tiedon avulla voidaan selvittää vallitsevia liikennetilanteita, kuten esimerkiksi ruuhkia. Tällaista tietoa voidaan yhdistää myös liikennevalojen ajoitustietoon ja tiegeometriaan matka-aikojen tarkentamiseksi. (Feng ym., 2014, s. 161)

GPS-tietoa voidaan kerätä esimerkiksi ajoneuvojen navigaattorilaitteista ja kojelautajärjestelmistä, matkapuhelimista ja älylaitteista sekä näitä yhdistämällä (Majchrakova, 2016).

Ihmisten liikkumissyklejä ja yleistä liikkumisaktiivisuutta voidaan seurata usean päivän kestoilla GPS-tutkimuksilla. Tutkimusajan on oltava kestoiltaan riittävän pitkä, sillä suuri osa matkoista ei toistu päivittäin eikä liikkumisen syklejä voida muutoin havaita. Dataa kerätään sekunnin intervalleilla huomioiden kiihtyvyyssanturin kyky havaita käytetty kulkumuoto. (Pastinen, 2012, s. 4)

Matkapuhelinten tukiasemat keräävät paikka- ja aikasidonnaista tietoa laitteiden sijainnista. Laitteen siirtyessä toisen tukiaseman alueelle voidaan päätellä myös henkilön liikkuneen sinne ja täten menetelmä soveltuu yksittäisten ja suurempien henkilömäärien liikkeiden seurantaan. Tukiasemia on lukumäärällisesti enemmän alueilla, joilla asutusta on tiheästi ja joilla siten myös liikutaan useammin. GPS-tutkimuksilla saadaan suurempi tarkkuus henkilön liikkeistä, mutta matkapuhelindatan käyttöön liittyy useita etuja. (Pastinen, 2012, ss. 4-5)

Matkapuhelinten viestintään perustuva seurantatieto on tarkkuudeltaan GPS-sijaintitietoa heikompaa. Matkapuhelinpaikannus perustuu laitteen sijainnin kolmiomittaukseen tukiasemien välillä ja menetelmän monimutkaisuus hankaloittaa sen potentiaalia sijaintitiedon lähteenä. (Houbraken ym., 2018, s. 2)

4.3.8 Kelluva ajoneuvodata

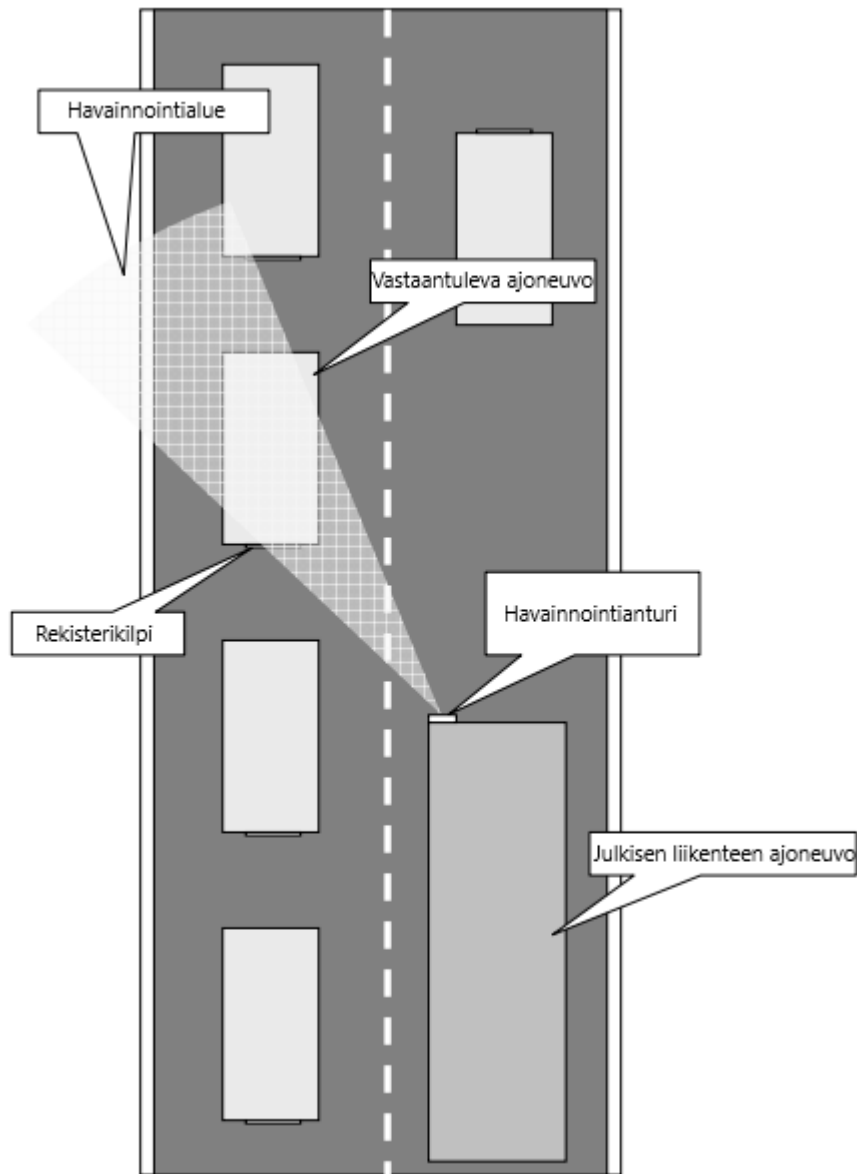
Kelluva ajoneuvodata (*engl. Floating Car Data, FCD*) on sijaintitietoon perustuva liikenteen havaitsemismenetelmä. Kelluvan ajoneuvodatan käyttö liikennetiedon lähteenä edellyttää huomattavasti vähemmän tienvarsilaitteiden kustannuksia kuin perinteiset ilmaisimet. Menetelmää käytetään tavallisesti matka-ajan historiatiedon saamiseen sekä muun liikennedatan ja algoritmien arvioimiseen, mutta se soveltuu käytettäväksi myös liikenteenhallintajärjestelmissä. (Houbraken ym., 2018, s. 1)

Menetelmän toiminta perustuu ihmisten mukana oleviin älylaitteisiin, jotka voivat lähettää ja vastaanottaa liikennetietoa ihmisten liikkeessä. Laitteiden käyttäjät hyötyvät henkilökohtaisesti optimaalisista kulkureiteistä lisäten samalla liikennetiedon laatua lähettämällä omaa dataansa liikennedatan tarkentamiseksi. (Houbraken ym., 2018, s. 1)

On huomionarvoista, että menetelmän avulla havaitaan tavallisesti ainoastaan yksittäiset ajoneuvot, jotka keräävät ja lähettävät kelluvaa ajoneuvodataa. Täten esimerkiksi ruuhkaisilla alueilla hitaasti liikkuvat yksittäiset ajoneuvot eivät anna laajaa kuvaa vallitsevista liikennetilanteista. Menetelmään voidaan kuitenkin yhdistää muita, muuta liikennettä havainnoivia antureita ja teknologioita. (Naumann ym., 2010, ss. 19-20)

Menetelmän toimintaa voidaan laajentaa kelluvan ajoneuvotarkkailun (*engl. Floating Car Observer, FCO*) periaatteella. Tällöin kelluvaa ajoneuvodataa keräävät ja lähettävät ajoneuvot havaitsevat lähiympäristönsä langattomia signaaleja, kuten esimerkiksi Bluetooth- tai Wi-Fi -yhteyksiä ja toimivat eräänlaisina luotaimina. Tämä mahdollistaa myös muiden kulkumuotojen, kuten kävelijöiden ja pyöräilijöiden havaitsemisen niiden toimesta. (Ruppe ym., 2012, ss. 1525-1527) Luotaimina voivat toimia esimerkiksi julkisen liikenteen ajoneuvot, joihin voidaan asentaa esimerkiksi kuvantunnistukseen tai langattomien signaalien havaitsemiseen perustuvia antureita (Naumann ym., 2010, s. 20). Havainnollistava piirros julkisen liikenteen ajoneuvon toimimisesta luotaimena on esitetty kuvassa 7.

Kuva 7. Havainnollistus kelluvan ajoneuvotarkkailun toiminnasta (mukaellen Naumann ym., 2010, s. 20).



Liikkuvia kohteita seurataan tyypillisesti GPS- tai älylaitteiden sijaintitiedon avulla. GPS:n ominaisuuksia ovat suuri tarkkuus ja älylaitteiden sijaintitieto on lukumäärällisesti suurta johtuen laitteiden yleisyydestä. Menetelmä sopii erityisen hyvin reaaliaikaisen ajoneuvojen sijaintitiedon keräämiseen ja se on lisäksi verrattain taloudellinen tapa tuottaa liikennedatata. (Leduc, 2008, s. 8)

Laitteiden sijainti voidaan havaita erilaisilla teknisillä toteutuksilla, joiden tärkeimpiä ominaisuuksia on näytetaajuus. Tieverkolla liikkuvat laitteet tallentavat ja lähettävät

sijaintitietonsa tasaisin väliajoin prosessoitavaksi ja yhdistettäväksi. Tämä näytetaajuus vaihtelee riippuen käytettävissä olevan tietoliikenneyhteyden kaistanleveyden ja laitteen energiankulutuksen mukaan, joka vaihtelee esimerkiksi kiinteissä ajoneuvolaitteissa ja kannettavissa älylaitteissa. (Houbraken ym., 2018, s. 2)

Korkean näytetaajuuden omaavissa laitteissa sijaintitieto päivitetään 1-10 sekunnin intervalleilla, joka mahdollistaa hyvin tarkan seurantatiedon saamisen sijainnin vaihdellessa näytteiden välillä vain vähän. Matalataajuuksisissa laitteissa sijainti päivitetään 30-60 sekunnin välein tai jopa harvemmin ja tästä syystä tarkan seurantatiedon saaminen on hankalaa. Tällöin tarkan kulkureitin arviointi edellyttää arvioiden ja reitin rekonstruktioonin tekemistä, sillä laite saattaa liikkua jopa useita kilometrejä sijaintipäivitysten välillä. Arviointiin tarvitaan tällaisissa tapauksissa monimutkaisempia algoritmeja. (Houbraken ym., 2018, s. 2)

Menetelmän käyttö liikennetiedon lähteenä edellyttää riittävää määrää kelluvaa ajoneuvodataa käyttäviä laitteita. Määrä vaihtelee riippuen käyttötarkoituksesta ja laitteiden näytetaajuudesta. Arviot riittävästä määrästä vaihtelevat 1 % - 10 % välillä, pienemmän osuuden ollessa mahdollisesti riittävä korkean näytetaajuuden laitteiden tapauksessa. (Houbraken ym., 2018, s. 2)

Kelluvaan ajoneuvodataan liittyy haasteista huolimatta suurta potentiaalia, sillä arviolta 1 % osuus tienkäyttäjistä vastaa toiminnalliselta tarkkuudeltaan induktiosilmukoihin perustuvaa ilmaisinjärjestelmää esimerkiksi jonoutumisen havaitsemisessa. Suurin hyöty kelluvasta liikennedatasta voidaan saavuttaa käyttämällä sitä yhtenä liikennedatan lähteenä muiden menetelmien ohella yhdistäen sitä eri tietolähteiden dataan, jolloin myös pienempi osuus on riittävä. On myös huomionarvioista, että menetelmään vaikuttavat tietoliikennetekniset laatutekijät, kuten latenssi. (Houbraken ym., 2018, s. 3)

5 Liikennetiedon hyödyntäminen

Tieverkon ruuhkautuminen voi aiheutua toistuvasta liiallisesta kysynnästä tai kertaluonteisten häiriöiden, onnettomuuksien tai muiden sattumanvaraisten tilanteiden, kuten haastavan säätilan aiheuttamista viiveistä. Ruuhkautumista voidaan pyrkiä ehkäisemään muun muassa pitkän aikajänteen suunnittelulla, liikenteenhallinnallisilla muutoksilla tai reaaliaikaisilla adaptiivisilla toimenpiteillä, kuten esimerkiksi liikennevalojen tai nopeusrajoitusten muutoksilla. (Leurent & Nguyen, 2010, s. 2)

Sattumanvaraisia tilanteita ei voida täysin ennustaa, mutta tällaisista syistä aiheutuvaan ruuhkautumiseen voidaan varautua havaitsemalla poikkeavat tilanteet varhaisessa vaiheessa liikennedatan avulla. Ajantasaisen tiedon avulla liikenteessä vallitseviin olosuhteisiin ja muutoksiin reagoiminen on mahdollista ja tällöin voidaan parantaa koko tieverkon toimintaa häiriötilanteiden aikana. (Leurent & Nguyen, 2010, s. 2)

Liikennevirta jakautuu vaihtelevasti tieverkolle ja siihen vaikuttavat muun muassa erilaisten ajoneuvojen suhteellinen osuus, liikenne-infrastruktuurin vaihtelevuus sekä muuttuvat sääolosuhteet. Liikenteestä saatavaa tietoa hyödynnetään muun muassa

- liikenteenhallintaan
- liikennesuunnittelun ja -projektien priorisointiin ja aikataulutukseen
- projektisuunnitteluun
- ylläpidon suunnitteluun
- liikenteen laskentaan
- liikenneturvallisuuden parantamiseen

(NPRA, 2004, s. 12)

Korkeatasoinen liikenteen analysointi, kuten liikennevirtojen ennakointi on oleellinen osa älykästä liikennejärjestelmää. Liikenneanalyysityökalujen kehittämiseksi on myös usein hyödyllistä käyttää runsaasti analysoimalla saatua liikennetietoa. (Choi ym., 2018, ss. 1-2)

Jotta liikennetietoa voidaan hyödyntää tehokkaasti, tulee sen olla digitaalisessa muodossa. Tietoa hyödyntävien palveluiden suurin tarve koskee laadukasta ja hyvin saatavilla olevaa

digitaalista tietoa. Suuristakaan tietovarannoista ei ole myöskään mahdollista muodostaa kattavaa ja digitaalista kokonaisuutta, mikäli niitä tuottavilla tahoilla ei ole siihen työkaluja, vaatimusta tai motivaatiota. (Pihlajamaa ym., 2020, s. 57).

Lukuisista erilaisista lähteistä kerätty liikennedata edellyttää usein jonkin asteista datan yhdistämistä, keskiarvoistamista tai skaalausta sekä uudelleentulkintaa ollakseen kelvollista ja hyödynnettävissä. Datojen yhdistämisessä käytettävät menetelmät perustuvat tyypillisesti erinäisiin tilastollisiin keskiarvoistamismenetelmiin, suodattimiin ja algoritmeihin. Datan reaaliaikaisuusvaatimukset riippuvat datan käyttäjästä, sillä esimerkiksi liikennesuunnittelussa tarvittava tieto on useinmiten vähemmän aikakriittistä kuin esimerkiksi häiriötilanteita valvovien tahojen. (BITRE, 2014, ss. 10-11)

Ennen kuin data on hyödynnettävissä, sitä voidaan joutua valmistelemaan. Datan valmistelu (*engl. Data Preparation*) tapahtuu tyypillisesti vaiheittain. Aluksi datasta tai tietokannasta kerätään ja otetaan näytteitä, datan ollessa mahdollisesti useissa sijainneissa. Tämän kerätyn datan ja näytteiden eheys tarkistetaan, jonka jälkeen ne siirretään yhteen sijaintiin, kuten tietokantaan. Lopuksi käsitelty data muunnetaan ja siitä luodaan uusia analyttisiä muuttujia ja poistetaan sellaiset muuttujat, jotka eivät enää tuo lisäarvoa. (Mamdouh, 2007, s. 13)

Tiedonlouhinta (*engl. Data Mining*) on menetelmäjoukko, joiden tarkoituksena on suorittaa matemaattisiin malleihin ja datan manipulointiin liittyviä toimenpiteitä uuden informaation muodostamiseksi datasta. Nämä toimenpiteet voidaan jakaa niiden tavoitteiden perusteella ryhmiin, jotka ovat luokittelu, arviointi, ennustus, ryhmittely ja affiniteettianalyysi. (Mamdouh, 2007, s. 7)

Luokittelussa pyritään valitsemaan tiedolle luokka ennalta määritellyistä arvoista, kuten esimerkiksi tietty riskitaso akselilta matala, keskitasoinen ja korkea. Arviointimenetelmän avulla voidaan pyrkiä selvittämään vaihtuvan muuttujan tietty arvo, kuten kotitalouden vuosittain muuttuva nettotulo. Ennustusta käytetään arvioimaan tietyn arvon tulevaa arvoa, kuten ilmanlaatua. Ryhmittely sopii informaation löytämiseksi suuresta datamäärästä jakamalla se pienempiin osiin. Affiniteettianalyysillä puolestaan pyritään löytämään toisiinsa liittyviä arvoja eli arvojen assosiaatioita, esimerkiksi joukkoliikenteen matkustajamäärien

vaihdellessa julkisen liikenteen ajoneuvojen määrä saattaa myös muuttua tähän liittyen. (Mamdouh, 2007, ss. 7-8)

5.1 Liikennetiedon jakaminen

Linturi & Kuittinen (2014, ss. 73-74) ovat havainneet, että liikennetiedon alusta on kriittinen resurssi ja siitä voi tulla myös strateginen tekijä, sillä liikennettä voi hallita liikennedatan avulla. Liikennedatan alustan tulisi olla esimerkiksi MyData -tyyppinen, jotta se olisi avoimesti kehittyvä sekä toimijoille avoin. Tällöin ihmiset voivat hyödyntää, hallita ja luvittaa synnyttämäänsä dataa esimerkiksi liikkumisestaan ja tällainen datan hallinta mahdollistaa sen siirrettävyyden ja uudelleenkäytön sekä vähentää toimittajariippuvuuksien syntymistä (Poikola ym., 2018, s. 4).

On tärkeää hallinnoida tietojen omistussuhteet, vastuut ja rajapinnat sekä huomioida standardien käytön mahdollisesti aiheuttamat riskit kehityksen pysähtymiselle. Tämän ohella myös teknologiriippuvuutta on vältettävä. Eri lähteistä saatavan datan yhdistämisellä saavutetaan eniten lisäarvoa myös liikennedatan tapauksessa. (Linturi & Kuittinen, 2014, s. 77)

Jotta jaettua liikennedatata voidaan hyödyntää tehokkaasti, sen on oltava koneluettavassa muodossa. Koneluettavuus merkitsee liikennedatan tapauksessa tallennusmuotoa, josta on luotavissa palveluita tiedon perusteella, esimerkiksi reittiopastuspalvelu perustuen joukkoliikenteen reitteihin ja aikatauluihin. Tällaisessakin tapauksessa tiedon ollessa hajaantuneena eri tietovarantoihin on olennaista, kuinka hajanaiset eri tietolähteiden tiedot ovat yhdistettävissä. (Pihlajamaa ym., 2020, s. 47)

Jaettavan liikennedatan laadulliset tekijät voivat aiheuttaa haasteita muun muassa tiedon heterogeenisen luonteen vuoksi. Eri lähteistä peräisin oleva tieto voi olla sisällöltään samankaltaista, mutta muodoltaan yhteensopimatonta esimerkiksi fyysisten sijaintikoordinaatistojen tai nimeämiskäytäntöjen osalta. Oleellimmat tekniset haasteet liittyvät heterogeeniseen ja osittain puutteita sisältävään tekniseen ympäristöön, järjestelmien yhteensopivuusongelmiin ja prosesseihin. Myös vanhan staattisen tiedon digitoimisessa voi olla puutteita. (Pihlajamaa ym., 2020, ss. 60-62)

Haasteita voidaan ehkäistä luomalla tietoa kokonaisuutena hallitseva kokonaisarkkitehtuuri, jotta erilliset järjestelmät erillisine tietovarantoineen ovat keskenään yhteensopivia (Pihlajamaa ym., 2020, ss. 60-62). Tiedon muodolliset haasteet voidaan pyrkiä ratkaisemaan standardien käytön avulla, joskin standardien mahdollinen joustavuus ja tulkinta voivat tuottaa myös yhteensopimatonta standardoitua tietoa (Pihlajamaa ym., 2020, s. 57).

Laadukkaaseen digitaaliseen tietoon liittyy useita ominaisuuksia. Näitä ominaisuuksia ovat muun muassa

- muodollinen ja merkityksellinen oikeamuotoisuus ja standardinmukaisuus, joka mahdollistaa tiedon automaattisen käytön ja tulkinnan
- sisällöllisesti yhdenmukainen tieto kaikista tietolähteistä
- oikeellinen ja täsmällinen tieto, jolloin tiedon suunniteltuun tai reaaliseen toimintoon tai ominaisuuteen voidaan luottaa
- ajantasainen, paras nykytilanteen tieto kuvatuista asioista
- tiedon täydellisyys, jolloin kaiken tarvittavan tiedon voidaan luottaa olevan käytössä

(Pihlajamaa ym., 2020, s. 57)

Tiedon sisällöllisten kriteerien ohella palveluiden rakentaminen hyötyy keskeytymättömästi ja viiveettömästi saatavilla olevasta tiedosta. Tähän tarkoitukseen voidaan käyttää tiedon haun kyselyrajapintoja, joista tietoa voidaan hakea tarpeen mukaan. Kokonaisuutena on tarpeellista, että tietoa on saatavilla staattisena ja dynaamisena lisäarvon luomiseksi. Palvelut hyötyvät myös liikkujien niin sanotuin MyData -periaattein itse tuottamasta tiedosta. Näiden tietojen käyttö mahdollistaa älykkäiden ja personoitujen palvelujen luomisen. Tällöin on myös huomioitava yksityisyydensuoja ja EU:n tietosuoja-asetus GDPR. (Pihlajamaa ym., 2020, s. 58)

MyData -periaatteet ovat yksityisyydensuojan ja tiedon pirstaleisuuden haasteita ehkäisemään pyrkiviä henkilötietojen organisointitapoja, joissa henkilö itse hallitsee häntä koskevia tietoja. Henkilöstä eri tahojen toimesta keräämä data on tällöin henkilön itsensä käytettävissä ja hän pystyy itse päättämään häntä koskevan tiedon jakamisesta, vaihtamisesta tai myymisestä. Ihmisten itse muodostama tieto, kuten paikka- ja reittitieto

ovat erityisen hyödyllisiä monissa käyttötarkoituksissa, kuten liikenteen tietolähteenä.
(Poikola ym., 2018, s. 5)

Linturi & Kuittinen (2014, ss. 77-78) painottavat MyData -periaatteiden mahdollistavan henkilötietoa hyödyntävien sovellusten, palveluiden ja liiketoimintarakenteiden syntymisen ja kehittymisen. Selkeät, koneluettavaa dataa sisältävät rajapinnat parantavat palveluiden luomisen helppoutta. On oleellista, että henkilö voi helposti saada haltuunsa itseään koskevan datan sekä käyttää ja siirtää sitä niin sanotun data portability -käsitteen mukaisesti. Tämän lisäksi henkilön on voitava hallita datansa käyttöä ja jakelua läpinäkyvästi esimerkiksi hallintapalvelun kautta.

Yksityisyyden suojaan sisältyy tietosuoja-asetus, joka suojelee yksilökohtaisia henkilötietoja. Tämän ohella henkilökohtaisen tiedon käytössä on huomioitava Euroopan unionin asetukset. (Ylitepsa, 2018, ss. 4-5) GDPR eli Euroopan unionin yleinen tietosuoja-asetus (*engl. General Data Protection Regulation*) säätelee muun muassa henkilökohtaisen datan eli yksilöön liittyvän henkilötiedon käyttöä esimerkiksi sijaintitiedon osalta ja sen tarkoituksena on suojata yksittäisen henkilön identiteettiä tunnistamiselta. GDPR kattaa myös datan käsittelyn sekä sen kohteen, hallinnoijan ja prosessoijan oikeudet ja vastuut. (Wolford, 2020)

Yksityisyyden suojaaminen voi edellyttää jaetun datan anonymisointia henkilötietojen poistamiseksi. Anonymisoinnissa henkilön tunnistaminen tehdään mahdottomaksi peruuttamattomasti dataa käsittelemällä. (Tietosuojavaltuutetun toimisto, 2020)

Eyupoglu ym. (2018, s. 373) jaottelevat anonymisointimenetelmät kategorioihin. Niitä ovat yleistys-, poistamis-, anatomisaatio-, uudelleenjärjestely- ja häiriömenetelmät, joissa

- yleistäminen korvaa datan arvoja geneerisillä arvoilla
- poistaminen hävittää tiettyjä arvoja datasta
- anatomisaatio erottaa tunnisteen ja määritteiden suhteet
- uudelleenjärjestely toteuttaa anatomisaation luomalla dataryhmiä ja sekoittamalla niiden sisältämää dataa toisiinsa
- häiriömenetelmä vaihtaa datan sisällön paikkoja tai lisää sattumanvaraista dataa tai kohinaa sisältöön

Anonymisointimenetelmät voidaan luokitella myös muun muassa satunnaistamiseen, kohinan lisäämiseen, permutaatioon ja luokituksen karkeistamiseen. Satunnaistamisessa muutetaan datan sisällön totuudenmukaisuutta henkilön ja tiedon yhteyden heikentämiseksi, jolloin niitä ei voida enää yhdistää. Kohinan lisääminen toteutetaan muuttamalla datan tarkkuutta säilyttäen kuitenkin kokonaisjakauma ennallaan. Permutaatioissa datan arvoja siirretään toisiin yksilöihin ja tällöin arvojen ja yksilöiden korrelaatiota ei voida havaita. Luokituksen karkeistaminen tapahtuu yleistämällä esimerkiksi arvojen suuruusluokkaa, jolloin yksilöä ei voida enää erottaa joukosta. (EU:n tietosuojaryhmä, 2014, ss. 12-16)

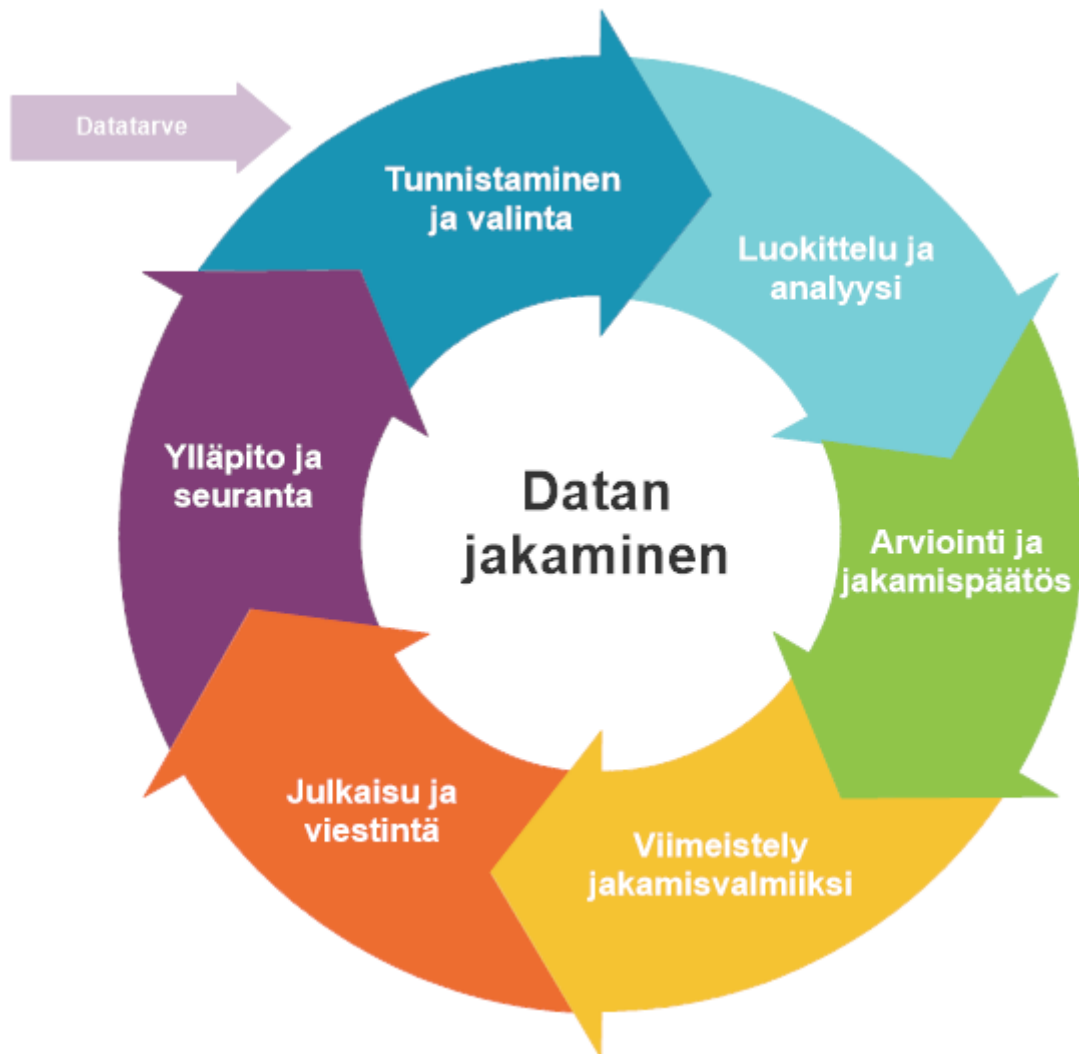
5.1.1 Yleinen datan jakamisprosessi

Julkinen hallinto muodostaa suuria määriä dataa, joka on usein suljetuissa tietojärjestelmissä. Tällainen data voi kuitenkin olla hyödyllistä lukuisille tahoille ja sen avulla voi muodostua innovaatioita käyttötarkoituksiin, joihin dataa ei alun perin ole suunniteltukaan käytettävän. Datan sisältämää informaatiota voidaan hyödyntää laajasti jakamalla data avoimesti. (Helsinki Region Infoshare, 2017)

Jaettu julkisen hallinnon data voi hyödyttää koko yhteiskuntaa parantamalla läpinäkyvyyttä ja demokratiaa, mahdollistamalla uusia markkinoita ja innovaatioita sekä tehostamalla sisäistä hallintoa. Olemassa olevissa toimintatavoissa ei välttämättä osata myöskään hyödyntää digitaalisuuden tuomia etuja, jotka voidaan havaita paremmin jaetun datan avulla. (Helsinki Region Infoshare, 2017)

Datan avaaminen eli julkaiseminen tai jakaminen voidaan nähdä prosessina. Jakamisprosessi on iteratiivinen ja se soveltuu myös liikennedatan jakamiseen. Datan jakamisprosessin päävaiheet on esitetty kuvassa 8.

Kuva 8. Päävaiheet datan iteratiivisessa avaamisprosessissa (mukaellen Espoon kaupunki, 2019, s. 4).



Datan jakamisprosessi alkaa datatarpeesta, joka voi saada alkunsa esimerkiksi yksittäiseltä ihmiseltä tai yhteiskunnalta. Kiinnostus ja tarve tietynlaista dataa kohtaan voi muodostua monille tahoille. Datatarpeen perusteella on aluksi tunnistettava ja rajattava haluttu data ja selvitetävä mahdolliset lupatekniset asiat muun muassa datan omistajuuteen ja jakamisen rajoituksiin liittyen. Tämän jälkeen data luokitellaan ja analysoidaan jakamisen mahdollisten hyötyjen, haittojen ja kustannusten selvittämiseksi. (Espoon kaupunki, 2019, s. 6)

Luokittelu- ja analyysivaiheessa tarkastellaan datan jakamisen mahdollisia hyötyjä ja riskitekijöitä sekä laaditaan kustannuksia alustavasti arvioiva analyysi. Kustannuksia saattaa syntyä esimerkiksi datan laadun varmistamisesta ja anonymisoinnista sekä ylläpidollisista

toimista, kuten datan päivittämisestä tai jakamisprosessin aiheuttamista henkilöstökuluista. Analyysin perusteella jakamisprosessissa jatketaan seuraavaan vaiheeseen tai se keskeytetään. (Espoon kaupunki, 2019, s. 7)

Laadittua analyysia arvioidaan suhteutettuna datan jakamisen aiheuttamaan riskitasoon. Riskitaso määritellään alhaiseen, keskitasoon tai korkeaan riippuen datan jakamisen riskeistä esimerkiksi tietosuojateknisistä syistä. Riskitason kasvaessa voidaan hyödyntää organisaation eri alojen erikoisosaajia jakamispäätöstä tehtäessä, esimerkiksi riskienhallinnan, tietoturvallisuuden tai tietosuojan osalta. (Espoon kaupunki, 2019, s. 8)

Arvioidun riskitason perusteella dataa tarkastellaan ja käsitellään tarvittaessa esimerkiksi anonymisaation avulla. Tässä vaiheessa huomioidaan myös EU:n tietosuoja-asetus GDPR. Anonymisaatiossa datasta poistetaan kaikki henkilötiedot ennen sen jakamista ja tunnistaminen on estettävä kaikilta osapuolilta. (Espoon kaupunki, 2019, s. 9)

Data viimeistellään jakamiskelpoiseksi valitsemalla datan lisenssi ja muoto siten, että koneluettavuus, avoimuus ja vapaa käytettävyys huomioidaan. Tämän lisäksi datan metatiedot sekä linkitys muuhun dataan dokumentoidaan. Datana jakamisesta viestitään datan julkaisualustoilla ja on suositeltavaa, että datan käyttäjät voivat myös antaa palautetta. (Espoon kaupunki, 2019, s. 10)

Jaetulle datalle määritetään sitä ylläpitävä ja seuraava taho sekä nimetään vastuutaho, joka voi olla yksittäinen henkilö tai esimerkiksi organisaation tietty yksikkö. Vastuutaho huolehtii datan ylläpidollisista ja seurannallisista toimista riskiarvioinnin perusteella, jota tulee päivittää jatkuvasti. Datana laatua ja hyödyntämistä seurataan esimerkiksi palautteiden perusteella. Havaitun hyödyllisyyden tai riskien perusteella datana laatua voidaan kasvattaa tai sen jakaminen voidaan päättää poistamalla data. (Espoon kaupunki, 2019, s. 11)

5.1.2 Avoimen datan konsepti

Avoimella datalla tarkoitetaan dataa, jota voidaan hyödyntää koneellisesti ja joka on avoimella lisenssillä lisensoitua. Avointa dataa voidaan hyödyntää esimerkiksi erilaisissa sovelluksissa ja tutkimuksessa sekä julkisella, että yksityisellä sektorilla. (Saarinen & Partanen, 2020)

Avoimeen dataan liittyy kolme ominaispiirrettä, jotka ovat saatavuus ja pääsy, uudelleenkäyttö ja -jako sekä yleinen osallistuminen. Datan on oltava saatavilla kokonaisuudessaan, tarkoituksenmukaisessa ja muokattavassa muodossa ilmaiseksi tai kohtuullisin kustannuksin. Dataa on voitava käyttää ja jakaa uudelleen ja sen on oltava koneluettavassa muodossa. Näiden asioiden on myös toteuduttava käyttäjistä riippumatta, mihin käyttötarkoitukseen tahansa. (Open Knowledge Foundation, 2020b)

Data voidaan jaotella omistajuutensa ja julkisuutensa perusteella kolmeen kategoriaan. Suljettu data on vain sen tahon saatavilla ja käytettävissä, joka sen omistaa. Jaettu data on tietyille ryhmille tiettyihin tarkoituksiin jaettua suljettua dataa, joka ei kuitenkaan ole kaikille avointa. Avoin data on nimensä mukaisesti kaikkien saatavissa, käytettävissä ja jaettavissa. (de Beer, 2016, s. 5)

Avoin data on dataa, joka on kaikkien käytettävissä ja julkaistavissa vapaasti ilman rajoituksia. Sen käyttö mahdollistaa datan laajan hyödyntämisen, yhteisöllisen datan luomisen ja sovellusten kehityksen suuren käyttäjäkunnan kautta. Avoimen datan konseptiin liittyy myös haasteita. (U.S. Department of Transportation, 2019, s. 116)

Yksityisyyden määrittäminen on haasteellista avoimen datan osalta, sillä jonkun on tehtävä päätöksiä siitä, mitä dataa on saatavilla. Tällöin on tärkeää huomioida yksityisyyden suoja sekä datan laadulliset tekijät, jotka voivat kärsiä dataa karsiessa. Avoimen datan tulkintaan liittyy myös haasteita, jotka voivat johtua datan huonosta dokumentoinnista, monimutkaisuudesta tai sisällöltään huonolaatuisesta datasta. Tämä voi aiheuttaa sekaannusta ja jopa laillisia seuraamuksia datan ollessa väärinkäytösten kohteena tai sen aiheuttaessa esimerkiksi virheiden vuoksi onnettomuuksia. (U.S. Department of Transportation, 2019, s. 116)

Tämän lisäksi avoimen datan kerääminen, prosessointi, säilytys, hallinta ja levittäminen aiheuttavat kustannuksia. Yksityisen sektorin toimijat eivät tavallisesti ole valmiita tekemään taloudellisia tappioita ja julkisen sektorin toimiessa avoimen datan lähteenä voi olla kyseenalaista, voidaanko tällaisella vastikkeettomalla tiedolla tehdä taloudellista voittoa käyttämällä sitä kaupallisissa sovelluksissa. (U.S. Department of Transportation, 2019, s. 116)

Avointa liikennedataa on Suomessa saatavilla useilta tahoilta. Tällaisia ovat muun muassa Traffic Management Finland, Väylä sekä useat muut julkiset organisaatiot sekä kaupungit ja kunnat. Liikennedataa jaetaan tiedostomuotoisena ja koneellisesti luettavassa muodossa rajapintojen kautta. Saatavilla oleva tieto vaihtelee tarjoajasta riippuen ajantasaisesta liikennetiedosta tilastolliseen historiatietoon. (Saarinen & Partanen, 2020)

5.2 Jaettu liikennedata

Laajamittainen reaaliaikaisen liikennedatan jakaminen on oleellista älykkäiden liikenteenhallintaratkaisuiden kehittämiseksi. Reaaliaikaisen liikennetiedon hyödyntämisen edellytyksenä on jaettu ja saataville koottu liikennedata, joka on peräisin erilaisista liikennetiedon lähteistä. Avoin data houkuttelee uusien palveluiden muodostumista ja palveluista itsestään voidaan saada edelleen uutta dataa palveluiden käyttäjiltä. (Strom, 2019)

Dataa voidaan jakaa useita kanavia käyttäen, joita ovat muun muassa verkkosivustot, FTP-palvelimet ja rajapinnat. Näistä kanavista erityisesti rajapintojen käyttö soveltuu liikennetiedon tyyppisen datan jakamiseen, sillä niiden avulla on mahdollista ladata tietoa osittain kokonaisten, suurten tietomäärien sijaan. Rajapinta on tyypillisesti yhdistetty tietokantaan, jota päivitetään reaaliaikaisesti. Tällöin saatavilla oleva tieto on myös ajantasaista. (Open Knowledge Foundation, 2020a)

Dataa tulisi jakaa raakamuotoisena ja kokonaisuudessaan. Tällaisen datan avulla

- alkuperäiseen datan tarjoajaan ei ole riippuvuutta, eikä datan saatavuus muutu esimerkiksi tarjoavan tahon rakennemuutosten johdosta
- kuka tahansa voi käsitellä ja levittää dataa uudelleen, jolloin datan jakelusta aiheutuvat kustannukset hajautuvat eikä yksittäisen tahon ongelmat estä jakelua
- kuka tahansa voi kehittää palveluita datan avulla, sillä datan jatkuva saatavilla olo voidaan taata

(Open Knowledge Foundation, 2020a)

Raakamuotoinen ja kokonaisuudessaan jaettu data mahdollistaa sen käyttämisen myös tarkoituksiin, joihin sitä ei ole alun perin tarkoitettu käytettävän. Tällöin datasta voi muodostua uudenlaista hyötyä yhdistettäessä muuhun dataan. Suurissa tietomäärissä rajapinnan tulisi tarjota datan viimeisin versio, mutta myös kokonainen raakadata tulisi päivittää säännöllisin väliajoin. (Open Knowledge Foundation, 2020a)

Rajapinta eli API (*engl. Application Programming Interface*) on määritelmä tietojen tai palveluiden tarjontatavasta muille tietojärjestelmille. Se sisältää määritellyt säännöt, joiden mukaan tietokoneet tai sovellukset keskustelevat keskenään. Rajapinta sijoittuu sovelluksen ja internetpalvelimen väliin, toimien välityskerroksena kahden järjestelmän välisessä tiedonsiirrossa. (IBM, 2020)

Rajapinta toimii käytännössä siten, että sovellus lähettää rajapintaan pyynnön saadakseen tietoa. Rajapinta välittää tiedon pyynnöstä toiseen sovellukseen tai palvelimeen, joka lähettää vastauksen ja pyydetyn tiedon takaisin rajapintaan. Lopuksi rajapinta lähettää tiedon alkuperäiselle sovellukselle. (IBM, 2020)

Rajapinta on avoin, mikäli sen kaikki ominaisuudet ovat julkisia ja sitä voidaan käyttää rajoituksetta. Tällöin myös rajapintakuvauksen ja sen dokumentaation on oltava avoimesti saatavilla ja rajapintaa on voitava käyttää vapaasti, maksuttomasti ja lupaa kysymättä mihin tahansa tarkoitukseen. (avoinrajapinta.fi, 2014)

Datarajapinta mahdollistaa palvelun sisältämän datan lukemisen ja käytön toisissa järjestelmissä. Rajapinta voi olla myös toiminnallinen, jolloin se sisältää myös järjestelmän tietojen muuttamisen ja laskenta-algoritmien käytön. (avoinrajapinta.fi, 2014) Rajapinnat käyttävät erilaisia protokollia, jotka määrittelevät käytettävät datatyypit ja komennot. Käytössä olevia protokollia ovat muun muassa SOAP, XML-RPC, JSON-RPC ja REST. (IBM, 2020)

SOAP (*engl. Simple Object Access Protocol*) on XML-pohjainen rajapinta, joka mahdollistaa datan lähettämisen ja vastaanottamisen SMTP- ja HTTP-protokollilla. SOAP-rajapintojen avulla tietoa voidaan vaihtaa erilaisilla kielillä kirjoitetuissa tai erilaisissa ympäristöissä toimivien sovellusten välillä. (IBM, 2020)

XML-RPC käyttää tiettyä XML-tietoformaattia. Se on yksinkertaisempi, kevyempi ja vanhempi versio SOAP-protokollasta. JSON-RPC on XML-RPC:n kaltainen yksinkertainen protokolla, mutta se käyttää JSON-tietoformaattia tiedonsiirrossa. REST (engl. *Representational State Transfer*) on kokoelma verkkorajapintojen arkkitehtuuriperiaatteita, eikä sillä siten ole standardoitua muotoa kuten protokollilla. (IBM, 2020)

Liikennedatan jakamisessa on huomioitava yhteistyön merkitys yhteisten toimintatapojen sopimiseksi, joka koskee myös rajapintojen määrittelyä. Useiden toimijoiden keskenään toteuttama datan avaaminen edellyttää muun muassa tiedostomuotojen ja rajapintojen sopimista. Optimaalitulanteessa liikennedatalla on saatavilla yhdessä sovituin tavoin ja muodoin, jolloin se on myös varmemmin käyttökelpoista ja helposti yhdistettävissä esimerkiksi eri alueiden liikennetiedon lähteistä. (Isotalo, 2017, ss. 14-15)

Julkisen sektorin toimijoiden, kaupunkien ja kuntien roolina on luoda edellytyksiä ja mahdollisuuksia innovaatioille ja palveluille kehityksen ja liiketoimintamahdollisuuksien luomiseksi sekä toimia esimerkin näyttäjänä muun muassa avoimuudessa ja tietosuojasioissa. Julkisen sektorin tulisi lähtökohtaisesti avata kaikki sellainen data, joka on avattavissa. On kuitenkin huomioitava, että tämä edellyttää resursseja avaamisen lisäksi myös ylläpidon osalta. Toiminnan kannattavuus riippuu datan käytöstä, sen jatkojalostusmahdollisuuksista sekä sen hyödyntämisestä. (Kulmala, 2019)

6 Helsingin kaupungin liikennetieto

Tämän työn kohdealueena on Helsingin kaupunki. Liikennedatata kerätään Helsingin kaupungin toimesta tällä hetkellä muun muassa kävelyn, pyöräilyn ja ajoneuvoliikenteen osalta.

Kappaleessa 6.1 kuvataan tiivistetysti Helsingin kaupungin nykytila ja tavoitteet liikennedatan osalta. Nykyiset, olemassa olevat tietolähteet ja käytettävät menetelmät esitellään kappaleessa 6.2 jaoteltuna liikennetiedon lähdeinventaarion mukaisiin kategorioihin. Kappaleessa 6.3 esitetään uusia potentiaalisia datalähteitä ja datan muodostamiseen käytettäviä menetelmiä, joilla kappaleessa 6.1 kuvatut tavoitteet voidaan saavuttaa. Kappale 6.4 käsittelee liikenteen reaaliaikaista tilannekuvaa, joka on yksi Helsingin kaupungin tavoitteista.

6.1 Nykytilan kuvaus ja tavoitteet

Helsingin kaupungilla liikennedatan kerääminen sijoittuu tämän työn kontekstissa organisatorisesti erityisesti kaupunkiympäristön toimialan, maankäytön ja kaupunkirakenteen palvelukokonaisuuden ja liikenne- ja katusuunnittelupalvelun alle. Sitä koskevat erityisesti Helsingin älyliikenteen kehittämissuunnitelma 2030 sekä Helsingin datastrategia.

Helsingin älyliikenteen kehittämissuunnitelma 2030 määrittelee kehittämistoimenpiteet liikennedatata koskien. Näitä ovat liikennetietojen kerääminen ja hyödyntäminen, liikenteen hallinta digitaalisin keinoin, kaupungin rooli uusissa palveluissa sekä liikenteen automatisoituminen. (Helsingin kaupunki, 2019b, s. 3)

Helsingin datastrategia on kaupungin johtoryhmässä 2019 hyväksytty kaupungin digitalisaatio-ohjelman tavoitteet ja linjaukset kuvaava dokumentti. Kaupungin tavoitteena on, että kerättyä dataa jaetaan ulkopuolisille toimijoille sovelluskehityksen ja uusien palveluiden luomisen tarpeisiin kaupungin toimiessa mahdollistajan roolissa. (Helsingin kaupunki, 2019a, s. 10)

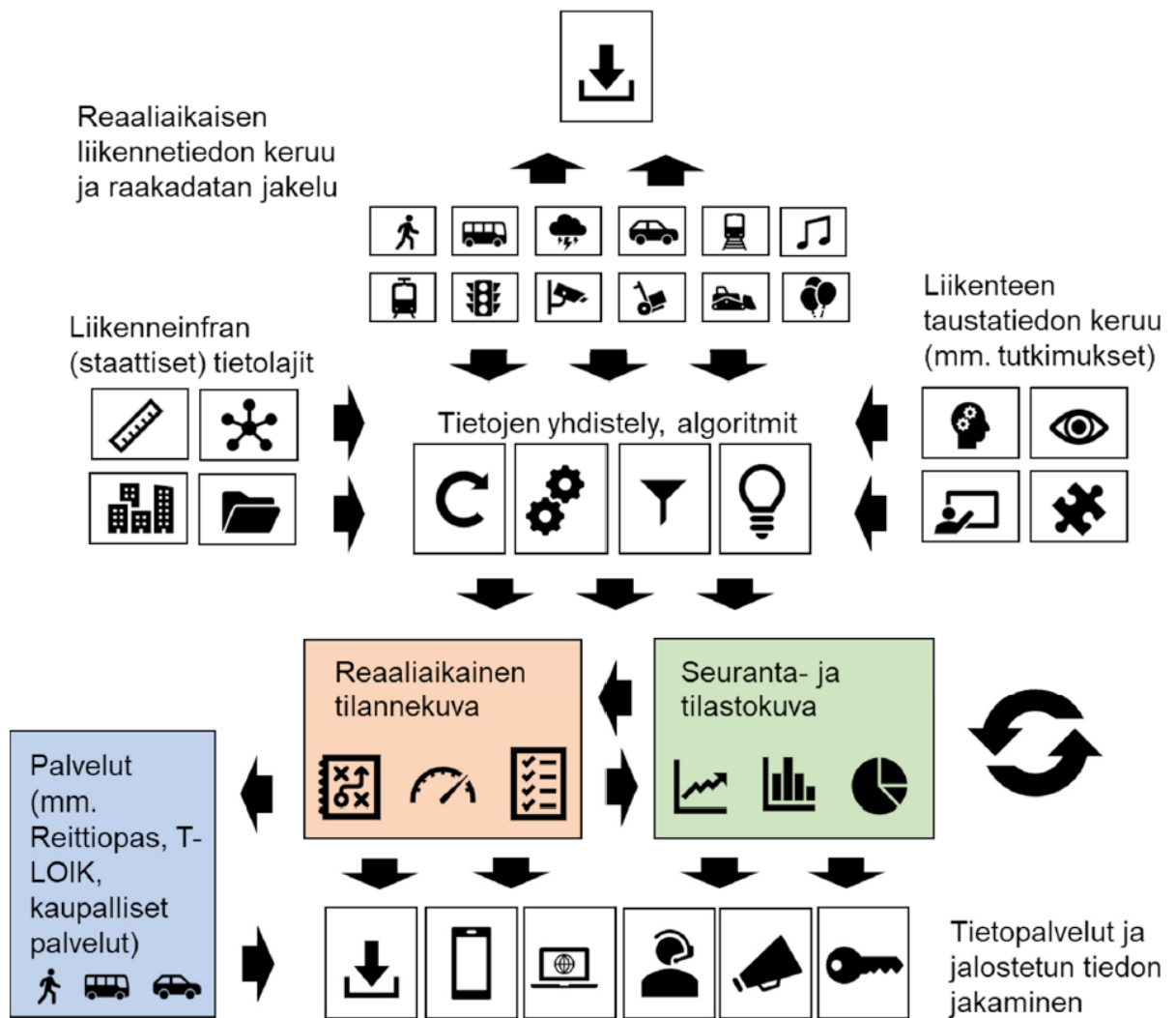
Marraskuussa 2019 pidetyssä ”*Helsinki Loves Developers*” -tilaisuudessa esiteltiin kaupungin avoimen liikennedatan ja sen hyödyntämismahdollisuuksien kirjoa sekä suunnitelmia tulevaisuuden älyliikenteestä Helsingissä. Liikennedatan lisääntymisellä pyritään luomaan lukuisia mahdollisuuksia ihmisten ja yritysten toimille parantaen samalla liikenteen sujuvuutta ja liikenneturvallisuutta. Liikennedatalla pyritään saamaan reaaliaikaisena risteyksistä, pysäköintijärjestelmistä ja joukkoliikenteestä sekä staattisena muun muassa liikenne-, sää- ja tapahtumatietona. Useita datalähteitä yhdistämällä luodaan liikenteen reaaliaikainen tilannekuva. (Partanen, 2019)

Helsingin kaupungin tavoitteena on rakentaa avoin arkkitehtuuri ja kaupungin roolina on luoda puitteet palvelujen rakentamiseen sekä varmistaa datan saatavuus ja avoimuus. Tällöin yrityksillä on hyvät edellytykset rakentaa avointa liikennedatalla hyödyntäviä palveluita. (Partanen, 2019)

Tavoitteena on, että vuonna 2030 liikennetiedon perusteella voidaan seurata liikennejärjestelmän kehitystä ja suunnitella toimintaa sekä mahdollistaa laadukkaan tilannetiedon saaminen liikkujille sekä kuljetus- ja liikkumispalveluyrityksille toiminnan suunnitteluun ja tehokkaampaan operointiin. Eri tietolähdetyypeistä saatava informaatio muodostaa ajantasaisen tilannekuvan informaatiopalveluiden ja operatiivisen toiminnan palvelukseen sekä seuranta- ja tilastokuvan liikennesuunnittelun ja sen toimenpiteiden avuksi sekä liikennejärjestelmän kehittymisen seurantaan. (Helsingin kaupunki, 2019b, s. 15)

Kehittämishojelman mukaisessa vuoden 2030 tavoitetilassa kerätään kattavaa ja ajantasaista informaatiota suunnittelun ja seurannan tarpeisiin kaikkien kulkumuotojen osalta. Oleellisimpia tietoja ovat määrätiedot kulkumuotokohtaisesti, nopeus- ja palvelujen varaustiedot tarvittavilta osin sekä liikenneinfrastruktuurin olosuhdetiedot, kuten säätilaa, ylläpitoa ja poikkeustilanteita kuvaavat tiedot. Kerätty informaatio kootaan ja jalostetaan nykyaikaiselle data-alustalle, jossa muodostuu tilannekuva sovelluskehittäjien, palveluntarjoajien ja yhteistyökumppanien käyttöön sekä seuranta- ja tilastokuva etenkin kaupungin omiin tarpeisiin. (Helsingin kaupunki, 2019b, s. 16) Havainnollistava kuva liikennetiedon elinkaaresta tavoitetilassa 2030 on esitetty kuvassa 9.

Kuva 9. Liikennetiedon elinkaari tavoitetilassa 2030 (Helsingin kaupunki, 2019b, s. 17).



Liikennedatan määrän ja laadun merkitys on Helsingin kaupungin tavoitteiden kannalta oleellinen. Näiden tavoitteiden saavuttamiseksi kaupungin tuottamaa liikennedatapohjaa on laajennettava. Käytännössä tämä edellyttää uusien liikennedatata tuottavien lähteiden ja menetelmien kokeilemistä ja käyttöönottoa sekä nykyisten, olemassa olevien datalähteiden potentiaalin selvittämistä ja mahdollista uudenlaista hyödyntämistä.

6.2 Olemassa olevat liikennetiedon lähteet

Helsingin kaupungilla on laadittu liikennetiedon lähteiden inventaario, joka on päivitetty 31.8.2020. Inventaarioon on listattu olemassa olevat liikennetiedon lähteet tarkentavine tietoineen. Inventaariossa liikennetiedon lähteistä on mainittu

- tietolähteen nimi
- saatavan tiedon kuvaus
- tietolähteestä vastuussa oleva taho
- tietolähteiden lukumäärä tai laajuus
- saatavan tiedon dataformaatti
- datan aikaresoluutio, päivitystiheys ja -menetelmä
- onko data julkisesti saatavilla
- prioriteetti
- mahdolliset lisätiedot
- rajapintalinkki
- tieto lähteen viemisestä data-alustalle

(LIITE1_liikennedatat, 2020)

Seuraavissa alaluvuissa on esitetty sellaiset Helsingin kaupungin liikennetiedon lähteet, jotka on suunniteltu vietäväksi tavoiteohjelmassa kuvatulle data-alustalle. Tietolähteet on listattu tiivistetysti ja jaoteltuina inventaarion mukaisiin kategorioihin. Olemassa olevat tietolähteet on kuvattu tämän työn tavoitteen kannalta tarpeellisessa laajuudessa, eikä kuvaus sisällä kaikkia inventaariossa mainittuja tietoja.

6.2.1 Automaattiset laskentalaitteet

Liikennetietoa kerätään automaattisten laskentalaitteiden avulla erinäisistä lähteistä. Inventaarion mukaiset tietolähteet on esitetty taulukossa 1.

Taulukko 1. Inventaarion mukaiset automaattiset laskentalaitteet (mukaellen LIITE1_liikennedatat, 2020).

LIKENNETIEDON LÄHDE	SAATAVA LIKENNETIETO	DATA-FORMAATTI	DATAN AIKARESOLUUTIO	PÄIVITYSTIHEYS JA MENETELMÄ	AVOIMUUS JA LISÄTIEDOT
Pyöräliikenteen määrät (Eco-Counter)	Pyöräliikenteen määrä. Suuntia ei eroteltu. Induktioilmaisimet.	JSON, CSV, XLSX	Rajapinnan kautta 15 minuuttia.	Kerran vuorokaudessa, automaattisesti.	Osittain. Tunnistautumisen vaativa rajapinta.
Pyöräliikenteen määrät (Baana, Infracontrol)	Pyöräliikenteen määrä. Suunnat eroteltu. Tallentuu ajoneuvokohtaisesti.	XLSX, API	Tiedot kerätään pyörä-tarkkuudella.	Jatkuva, automaattisesti.	Avoin. Automaattiseen tiedonsiirtoon siirrytään kun laitteet uusitaan.
Pyöräliikenteen määrät (Marksman-laskimet)	Pyöräliikenteen määrä. Suuntia ei eroteltu. Induktioilmaisimet.	Tuntematon raakadata-formaatti.		1-2 kertaa vuodessa, manuaalisesti.	Ei. Laitteet poistuvat lähivuosina.
Autoliikenteen määrät ja nopeudet (M680)	Ajoneuvojen määrä kaistoittain. Osittain myös ajoneuvoluokka ja nopeus.	JSON, CSV, XLSX	Tiedot kerätään ajoneuvo-tarkkuudella.	5 minuuttia, automaattisesti.	Avoin. Kaksi eri rajapintaa, joita voidaan hyödyntää.
Autoliikenteen määrät ja nopeudet (vanhat laskennat)	Ajoneuvojen määrä ja nopeus suunnittain. Tieto haetaan useimmissa sijainneissa manuaalisesti.	Tuntematon teksti-formaatti.		1-2 kertaa vuodessa, manuaalisesti.	Osittain. Laitteet poistuvat lähivuosina.
Jalankulkijoiden määrät (Eco-Counter Byro-box)	Jalankulkijoiden määrä. Suunnat eroteltu.	JSON, CSV, XLSX	Rajapinnan kautta 15 minuuttia.	Kerran vuorokaudessa, automaattisesti.	Osittain. Tunnistautumisen vaativa rajapinta.
Nopeusnäyttötaulut	Ajoneuvojen nopeus. Epätarkka määrätieto.	Tuntematon.		Tuntematon, manuaalisesti.	Osittain. Ei rajapintaa. Tiedot saatavilla sähköpostitse.

Automaattisista laskentalaitteista saatava tieto on suurilta osin peräisin perinteisistä liikennetiedon lähteistä. Niiden tuottama liikennetieto on relevanttia ja se on oleellista myös asetettujen tavoitteiden kannalta, mutta näille tietolähteille ominaisesti ilmaisimien sijoittelu ja verrattain vähäinen peittoalue muodostavat tarpeen sekä ilmaisinjärjestelyjen laajentamisen, että muiden tietolähteiden käyttöönoton toteuttamiselle. (Gu ym., 2016, s.321; Leduc, 2008, s. 8) On myös huomionarvoista, että suuri osa liikennetietoa tuottavista ilmaisimista on induktioilmaisimia, joiden vikatilanteet ovat verrattain yleisiä ja jotka ovat herkkiä ulkoisille vaurioille. Tämä korostaa ylläpidon merkitystä ja muiden ilmaisintyyppien tarvetta. (Dagleish & Hoose, 2008, ss. 123-124; Houbraken ym., 2018, s. 2)

Automaattisten laskentalaitteiden heterogeenisen laitekannan aiheuttamat haasteet on nähtävissä dataformaattien ja tiedon avoimuuden vaihtelevuutena sekä useiden laskentalaitetekonaisuuksien poistumisena tai korvaamisena. Moninainen laitekanta hankaloittaa niistä saatavan tiedon jakamista (Pihlajamaa ym., 2020, ss. 60-62).

6.2.2 Manuaalilaskennat

Liikennetietoa kerätään myös manuaalisten käsinlaskentojen avulla. Inventaarion mukaiset tietolähteet on esitetty taulukossa 2.

Taulukko 2. Inventaarion mukaiset manuaalilaskennat (mukaellen LIITE1_liikennedatat, 2020).

LIIKENNETIEDON LÄHDE	SAATAVA LIIKENNETIETO	DATA-FORMAATTI	PÄIVITYSTIHEYS JA MENETELMÄ	AVOIMUUS JA LISÄTIEDOT
Jalankulkijoiden ja pyöräliikenteen käsinlaskennat	Manuaaliset käsinlaskennat. Tallennus minuutin välein. Suunnat eroteltu.	CSV	Kerran vuodessa, manuaalisesti.	Osittain. Aineisto vaatii manuaalista käsittelyä.
Jalankulun käsinlaskennat	Manuaaliset käsinlaskennat. Tallennus minuutin välein. Suunnat eroteltu.	CSV	Kerran vuodessa, manuaalisesti.	Osittain. Aineisto vaatii manuaalista käsittelyä.
Autoliikenteen käsinlaskennat	Ajoneuvoliikenteen määrät ajoneuvoluokittain ja suunnittain.	CSV	2 kertaa vuodessa, manuaalisesti.	Avoin, osittain. Aineisto vaatii manuaalista käsittelyä.
Matkustajalaskennat	Joukkoliikenteen matkustajalaskennat (seutu-, kauko- ja tilausliikenne).	CSV	Joka toinen vuosi, manuaalisesti.	Osittain. Aineisto vaatii manuaalista käsittelyä. Lasketaan laskentalinjoilla.
Liityntäpysäköintilaskennat	Pyörien ja ajoneuvojen liityntäpysäköintien inventointi ja käyttöaste.	XLSX	Joka toinen vuosi, manuaalisesti.	Rajoitetusti. Aineisto vaatii manuaalista käsittelyä. Lasketaan syksyisin juna- ja metroasemilla.
Autoliikenteen kehälaskennat	Ajoneuvoliikenteen määrät, ajoneuvoluokittain säännöllisesti seurattavilla kehäpistellä, poikkileikkauslaskentoja.	CSV	Kerran vuodessa, manuaalisesti.	Avoin, osittain. Aineisto vaatii manuaalista käsittelyä.

Inventaariosta on nähtävissä manuaalilaskennoille tyypillinen keräämisen jälkeisen manuaalisen käsittelyn tarve sekä harva päivityssykli. Niistä saatava tieto on kuitenkin usein hyvin kattavaa ja täten hyödyllistä, joskin ne aiheuttavat lisääntyvää työtä ja kustannuksia (Lee & Sener, 2019, s. 2).

Manuaalilaskennoista saatavaa tietoa voidaan kerätä myös muun muassa langattomiin signaaleihin tai GPS-paikkatietoon perustuvilla menetelmillä. Kelluvaan ajoneuvodataan perustuvissa menetelmissä liikennetietoa voidaan muodostaa myös muusta liikenteestä, kävelijöistä ja pyöräilijöistä muun muassa ihmisten mukana olevien älylaitteiden tai kelluvaan ajoneuvotarkkailuun perustuen. (Houbraken ym., 2018, s. 1; Leduc, 2008, s. 8; Naumann ym., 2010, ss. 19-20)

6.2.3 Muut liikennetiedon lähteet

Liikennetietoa kerätään automaattisten laskentalaitteiden ja manuaalisten laskentojen lisäksi liikennevaloista. Helsingin kaupunki toteuttaa myös säännöllisesti erinäisiä kokeellisia ja otosmaisia liikennetietoa muodostavia hankkeita ja projekteja.

Suuri osa Helsingissä sijaitsevista liikennevaloista kuuluu jonkin liikennevalojen keskusohjausjärjestelmän piiriin, joita ovat Helsingissä Omnia ja OmniVue. Tämä mahdollistaa muun muassa liikennevalojen ja niiden toiminnan seurannan ja hallinnan etäyhteydellä. Useimmat liikennevalot hyödyntävät toimintalogiikassaan fyysisiin ilmaisimiin, kuten induktioilmaisimiin tai erilaisiin tutka-ilmaisimiin perustuvia havaintoja liikenteestä. Tällaista havaintotietoa voidaan hyödyntää liikennevalojen toiminnan lisäksi myös liikennedatana muissa tarkoituksissa.

Omnia-keskusohjausjärjestelmästä saadaan liikennevalojen ilmaisimien tiedot, jotka päivittyvät jatkuvasti ja automaattisesti. Ilmaisimien tiedot ovat useimmiten kaistakohtaisia ja niiden dataformaatti on CSV, XLSX tai API. Järjestelmässä on vielä toistaiseksi suljettu rajapinta, josta ilmaisimien tiedot saadaan ilmaisinkohtaisesti. Ilmaisimien dokumentaatio ja sijaintitiedot ovat kuitenkin tällä hetkellä puutteelliset rajapinnan käytön kannalta. (LIITE1_liikennedatat, 2020)

OmniVue-keskusohjausjärjestelmä tuottaa liikennevalojen ilmaisimien tiedot, jotka päivittyvät epäsäännöllisesti ja manuaalisesti. Ilmaisimien tiedot ovat useimmiten kaistakohtaisia ja niiden dataformaatti on CSV tai XLSX. Tietojen saaminen järjestelmästä on vielä kehitysasteella, jonka aikataulusta ei ole selvyyttä. (LIITE1_liikennedatat, 2020)

Liikennevalojen ilmaisintietoon perustuvan liikennetiedon vahvuuksia ovat sen kattavuus ja keskittyminen liikenteellisesti oleellisiin sijainteihin, kuten suuriliikenteisille risteysalueille. Se soveltuu liikenteen reaaliaikaisen tilannekuvan muodostamisen ohella erityisesti liikenteenhallinnallisiin tarkoituksiin ja se on myös oleellinen osa autonomisten ajoneuvojen vaatimaa liikennetietoa. Liikennevaloihin liittyvät liikennedatata koskevat haasteet ovat tyypillisesti teknisiä, sillä liikennevalojärjestelmät ovat Helsingin kaupunginkin tapauksessa hyvin heterogeenisiä.

Liikennevalojen toimintaa ohjaavissa risteyskojeissa on suurta varianssia laitteiden tekniikan ja toimintalogiikan osalta ja tämä aiheuttaa haasteita niiden tuottaman liikennetiedon hyödyntämiseksi datalähteenä. On myös huomioitava, että ilmaisinjärjestelyt on usein toteutettu liikennevalojen toimintaa ajatellen, eikä tarvittavia ilmaisinjärjestelyjä esimerkiksi kaistakohtaisen liikennetiedon saamiseksi välttämättä ennestään ole. Risteyskojeiden heterogeenisen luonteen lisäksi keskusohjausjärjestelmät ovat tyypillisesti suljettuja järjestelmiä ja tästä syystä niiden sisältämän liikennedatan jakaminen avoimena datana on hankalaa.

Liikennevalojen ohella liikennetietoa muodostuu myös erinäisten kokeellisten ja otosmaisten hankkeiden ja projektien kautta. Tavoiteohjelman mukaiselle data-alustalle tällaisista on suunniteltu vietävän Länsisataman liikenteen ajantasainen tilannekuva sekä moottoriajoneuvoliikenteen sujuvuusaineisto.

Länsisataman liikenteen ajantasainen tilannekuva sisältää Länsisatamaan saapuvien laivojen ja niistä purkautuvan ajoneuvoliikenteen määrätiedon sataman sisällä erotellen raskaan liikenteen ja henkilöautot. Tieto saadaan automaattisesti yhden minuutin päivitystiheydellä ja sen dataformaatti on JSON. Tieto on avoimesti saatavilla rajapinnan kautta.

(LIITE1_liikennedatat, 2020)

Moottoriajoneuvoliikenteen sujuvuusaineisto kattaa lähes koko Helsingin päätie- ja katuverkon ja se perustuu ajoneuvojen navigaattorien keräämään dataan. Tieto on hankittu kaupungin käyttöön ulkopuoliselta toimittajalta raakadatan paikkatietomuodossa ja se päivitetään manuaalisesti noin vuoden välein. Tiedon muodosta ja hankintatavasta johtuen se edellyttää käsittelyä hyödyllisen informaation tuottamiseksi. Analysoimalla saatavaa tietoa ovat muun muassa ajoneuvojen keskimääräiset nopeudet sekä sujuvuustiedot. Raakadata ei ole itsessään julkista, mutta siitä saatavaa tietoa voidaan julkaista. Vuosittain toteutettava hankinta ei välttämättä kohdistu välttämättä samaan toimittajaan, joten aineiston laadulliset ominaisuudet voivat muuttua vuosittain. (LIITE1_liikennedatat, 2020)

6.3 Potentiaaliset tietolähteet ja datanmuodostusmenetelmät

Helsingin kaupunki on toteuttanut tai ollut osallisena lukuisissa hankkeissa ja projekteissa, joiden tavoitteena on ollut liikennedatan muodostaminen tai joiden kautta on syntynyt liikennedatata välillisesti.

Aiheeseen liittyen on aikaisemmin toteutettu muun muassa ”InnoTiku” -projekti vuonna 2016, jonka tavoitteeksi määriteltiin reaaliaikaisen tilannekuvan kehityksen edistäminen. Projektin lähtötilanteessa korostettiin tavoitteen edellyttämän reaaliaikaisen ja luotettavan liikennetiedon määrän merkitystä, jota pyrittiin lisäämään erinäisin perinteisistä liikennetiedon lähteistä eroavin innovatiivisin teknologioin ja menetelmin, kuten videoanalyysimenetelmillä. (Ramboll, 2016, s. 5)

Osana projektia toteutettiin videoanalytiikkaan perustuva kokeilu, jossa hyödynnettiin olemassa olevien liikennekameroiden kuvaamaa materiaalia liikennetiedon lähteenä. Menetelmän toiminta perustuu kiinteästi asennettuun kamerakuvaan ohjelmallisesti asetettavista ”ansalangoista”, joiden ylittävä liike voidaan tulkita havainnoksi esimerkiksi ajoneuvosta. Menetelmän avulla voidaan riittävän kattavilla kamerajärjestelyillä ja ohjelmallisilla konfiguraatioilla tuottaa liikennemäärätietoa eri kulkumuodoista. (Deloitte, 2016, s. 8)

Liikennekameroiden fyysisistä sijainneista, alkuperäisistä käyttötarkoituksista ja heterogeenisestä laitekannasta johtuen menetelmä ei sellaisenaan sovellu liikennedatan

muodostamiseen Helsingin kaupungissa ilman laajamittaisista laiteasennuksista ja -uusinoista aiheutuvia huomattavia investointikustannuksia. Laajamittaisessa skenaariossa tarvittavien kameralaitteiden lukumäärä vaihtelee useista sadoista tuhansiin, laitekustannusten ollessa useita tuhansia euroja kappaleelta. Yksittäisen risteyksen tapauksessa voidaan tarvita lukuisia kameroita kaista- ja kulkumuotokohtaiseen havainnointiin ja tämän ohella on otettava huomioon muun muassa tiedonsiirto- ja palvelinkustannukset sekä konfiguraatiotyön tarve. (Deloitte, 2016, ss. 48-50)

On huomionarvoista, että monien sinällään liikennedatan lähteiksi soveltuvien ja teknisesti toimivien liikenteen havainnointiin tarkoitettujen teknologioiden ja menetelmien käytön esteenä saattavat olla niistä aiheutuvat kustannukset. Suurin osa kustannuksista muodostuu tyypillisesti fyysisistä laitteista ja niiden edellyttämistä asennustöistä.

Siirtymisessä nykytilasta tavoitetilaan on huomioitava myös taloudelliset tekijät ja tämä asettaa rajoitteita suunniteltavien liikennedatan lähteiden käyttöön. Kokonaisuutena on löydettävä taloudellisesti kestäviä, juuri Helsingin kaupunkiin soveltuvia, sen nyky- ja tavoitetilan huomioivia ratkaisuja, jotka ovat myös taloudellisesti realistisesti toteutuskelpoisia järkevässä aikamääreessä.

Monien liikennedatata muodostavien menetelmien kustannusten muodostuessa pääasiallisesti laitteiden eli käytännössä fyysisesti asennettavien ilmaisimien hankinnasta ja asennuksesta, voidaan taloudellisestikin realistiseen tavoitetilaan pyrkiä käyttämällä sellaisia liikennedatata muodostavia menetelmiä, joissa erillisten ilmaisimien tarve on pieni. Tällaisissa menetelmissä esimerkiksi ihmisten mukana kulkevat älylaitteet toimivat itsessään sijaintinsa paljastavina ilmaisimina ja ne ovat myös valmiiksi olemassa, eikä fyysisten asennusten haittavaikutuksia ilmene.

InnoTiku -projektin osana toteutetussa liikkuvan videokuvan analysointiin perustuvassa kokeilussa liikennedatata muodostettiin ajoneuvoissa sijaitsevien älylaitteiden kameroiden kuvaamaa materiaalia analysoimalla (Roppo & Sutinen, 2016, s. 8). Menetelmä mahdollistaa koneoppimisen keinoin muun muassa eri kulkumuotojen tunnistamisen (Roppo & Sutinen, 2016, s. 3). Esimerkki menetelmän muodostamista havainnoista on esitetty kuvassa 10.

Kuva 10. Esimerkki kulkumuotojen tunnistamisesta koneoppimisen avulla (mukaellen Roppo & Sutinen, 2016, s. 15).



Kokeilussa havaittiin, että menetelmä soveltuu sellaisenaan hyvin liikenteen reaaliaikaisen tilannekuvan edellyttämän liikennedatan muodostamiseen. Suurimmat kokeilussa havaitut haasteet liittyivät reaaliaikaisen videoinformaation edellyttämiin tiedonsiirtoteknisiin nopeusvaatimukseen sekä videomateriaalin analysoinnissa tarvittavaan tietotekniseen laskentatehoon. (Roppo & Sutinen, 2016, ss. 25-26)

Kokeilun päättymisen jälkeen mobiilissa tiedonsiirrossa on tapahtunut edistysaskeleita ja Suomessa siirrytään lähivuosien aikana 5G -verkkojen mahdollistamiin aikaisempaa huomattavasti suurempiin tiedonsiirtonopeuksiin langattomissa verkoissa. Tämä mahdollistaa langattomasta tiedonsiirtonopeudesta riippuvaisten menetelmien aikaisempaa huomattavasti suuremman potentiaalin realistisina ja verrattain kustannustehokkaina tapoina muodostaa liikennedataa. Vastaava havainto tehtiin myös toteutetun kokeilun aikana (Roppo & Sutinen, 2016, s. 10).

Tietoteknisen laskentatehon vaatimukseen voidaan pyrkiä käyttämällä älylaitteita itsessään osana analytiikan edellyttämää käsittelyä (Roppo & Sutinen, 2016, s. 35). On myös huomioitava, että keskimääräinen älylaitteiden ja tietokoneiden laskentateho kasvaa teknologian kehittyessä, joka vähentää tämän haasteen vaikutusta ajan kuluessa. Samankaltainen kehitys on havaittavissa myös älylaitteiden sisäisissä kameroissa.

Kokeilussa havaittiin sen potentiaali myös osana joukkoistettua liikennedatata (Roppo & Sutinen, 2016, s. 36). Tällöin syntyvät kustannukset muodostuvat pääosin tietoteknisistä laitteista, kuten palvelimista, sillä erillisiä ilmaisimia ei tarvita. Kokonaisuutena kokeilun havainnot tukevat älylaitteisiin perustuvien liikennedatan muodostusmenetelmien potentiaalia.

Pyöräilijöiden havaitsemiseen soveltuvien datalähteiden osalta puutteita on havaittavissa, sillä nykyiset mittauspisteet kattavat vain verrattain pienen osan pyöräilyyn soveltuvista reiteistä. Tähän tarkoitukseen on olemassa vain polkupyörät havaitsevia induktioilmaisimia, jotka erottelevat moottoroidut ajoneuvot ja moottoripyörät. Ilmaisimen läheisyyteen on myös mahdollista toteuttaa informatiivisia laskentatauluja, jotka laskevat pyöräliikennettä visuaalisesti reaaliajassa. Menetelmää käytetään Suomen lisäksi muun muassa Yhdysvalloissa, Kanadassa ja Turkissa. (Eco-Counter, 2019)

Esimerkkitoiteutuksessa laskentataulu sijaitsee kävelykadulla, induktioilmaisimen ollessa asennettuna pyöräkaistalle. Ilmaisinkokonaisuus havaitsee myös jalankulkijat erillisen ilmaisimen avulla. Tekniikan vahvuutena on soveltuvuus sekaliikenteeseen sekä tarkkuus erityisesti pyöräkaistoilla, tämän lisäksi se soveltuu Suomen olosuhteisiin johtuen käytettävien ilmaisimien ominaisuuksista. (Eco-Counter, 2019)

Ilmaisinkokonaisuuteen on mahdollista toteuttaa kaapeloitu tai langaton tiedonsiirto, jolloin se toimii myös reaaliaikaisen liikennedatan lähteenä ja mittaustietoa voidaan hyödyntää osana muita järjestelmiä, kuten tilannekuvaa. Informatiivisella näyttötaululla voidaan lisäksi esittää reaaliaikaisesti pyöräilijöiden ja kävelijöiden päivä- ja vuosikohtainen lukumäärä ja täten vaikuttaa rohkaisevasti suosimaan pyöräilyä ja kävelyä ajoneuvon sijaan Helsingin tavoitteiden mukaisesti. Esimerkki toteutustavasta on esitetty kuvassa 11.

Kuva 11. Esimerkkitoteutus Eco-Counterin kävely- ja pyöräilyilmaisimen käytöstä (Eco-Counter, 2019).



Jalankulkijoiden ilmaisinerustaiseen havaitsemiseen liittyy haasteita, sillä ajoneuvoihin verrattaessa jalankulkijat ovat fyysiseltä profiililtaan pienempiä, liikkuvat moniulotteisemmin ja sattumanvaraisemmin ja kokoontuvat tiiviimpiin ryhmiin. Tyypillinen jalankulkijota havaitseva ilmaisimenetelmä perustuu kameratekniikkaan ja siihen yhdistettyihin koneoppimisen menetelmiin. Jalankulkijoiden havaitsemiseen soveltuu myös muun muassa LiDAR, jonka havaitsee liikkuvat kohteet myös huonossa valaistuksessa, toisin kuin kameratekniikkaan perustuvat ilmaisimet. (Liu ym., 2019, ss. 1-2)

Suomalaiset käyttävät yleisesti internetiä ja älypuhelimia, joten niiden sijaintitiedon käyttöön liikennetiedon lähteenä liittyy runsaasti potentiaalia. Tilastokeskuksen ”Väestön tieto- ja viestintätekniikan käyttö” -tutkimuksen (2019) mukaan 83 %:lla suomalaisista oli käytössään älypuhelin vuonna 2019 ja yleisimmin internetiä käytettiin älypuhelimella, käyttöosuuden ollessa 80 %. Täten lähes kaikki älypuhelimien omistavat suomalaiset käyttivät usein mukana olevalla laitteellaan internetiä. (Tilastokeskus, 2019).

Helsingin väkiluku oli 1.1.2019 noin 650 000 henkilöä ja väestön odotetaan kasvavan tasaisesti tulevaisuudessa. Väestötiheys koko kaupungin alueella oli vuonna 2018 noin 3000 henkilöä maa-neliökilometrillä. (Helsingin kaupunki, 2019d, ss. 4-9) Suuresta väestöpohjasta, väestötiheydestä ja älypuhelinien yleisyydestä johtuen sekä langattomiin signaaleihin, että kelluvaan ajoneuvodataan perustuvat liikennedatan keräämismenetelmät omaavat suurta potentiaalia Helsingin kaltaisella alueella.

6.3.1 Helsingin kaupungin rooli

Langattomiin signaaleihin perustuvat liikennedatan keräämismenetelmät soveltuvat erityisen hyvin Helsinkiin, mutta niihin liittyy myös haasteita. Suuren väestöpohjan ja väestötiheyden omaavalla alueella tällaisia signaaleja hyödyntäviä laitteita on runsaasti ja tiheästi, jolloin dataa on mahdollista saada määrällisesti paljon. Henkilön sijaintitietoa hyödyntävissä menetelmissä on kuitenkin huomioitava yksityisyyden suoja sekä GDPR, jolloin tällaisten menetelmien käyttö edellyttää riittävää tietosuojan toteutumista datan keräämis- ja hyödyntämävaiheessa. Erinäiset datan anonymisointimenetelmät ovat osa sijaintitietojen oikeanlaista käsittelyä. (Espoon kaupunki, 2019, ss. 7-9; Tietosuojavaltuutetun toimisto, 2020; Wolford, 2020)

Maailmalla on laajasti käytössä joukkoistettuun dataan perustuvia navigointisovelluksia, kuten esimerkiksi *Waze*. Tämän tyyppisissä sovelluksissa käyttäjät toimivat liikennedatan tuottajina ja hyödyntäjinä. (Amin-Naseri, 2018, s. 5) Tämän kaltaisten sovellusten ohella henkilön sijaintitietoa voidaan saada langattomien verkkojen ja -teknologioiden käyttöön perustuvien menetelmien avulla, joiden käyttöön Helsingin kaupunki voi omalla toiminnallaan rohkaista.

Mikäli kaupungissa liikkuvilla ihmisillä on tarjolla ilmaisia, nopeita ja kattavia langattomia verkkoja, liikkumista verkkojen alueella voidaan seurata sijaintitiedon lähteenä. Eräs potentiaalinen tapa lisätä niiden käyttöä on tarjota ilmaista langatonta WLAN-yhteyttä älylaitteiden käyttäjille. Potentiaalia on myös tulevassa 5G-teknologiassa, joiden tukiasemien sijoittelu vaikuttaa sijaintitiedon saamiseen. 5G mahdollistaa myös muiden teknologioiden vaatiman pienen latenssin ja suuren tiedonsiirtokapasiteetin esimerkiksi kelluvaa ajoneuvodataa ja autonomisia ajoneuvoja varten. (Carlson, 2020, s. 725; Traficom, 2020)

Suomessa on otettu laajasti käyttöön ”Koronavilkku” -sovellus, joka havaitsee muita sovelluksen sisältäviä älylaitteita hyödyntäen Bluetooth-tekniikkaa (THL, 2020a). Sovelluksella on yli 2,5 miljoonaa latauskertaa kahden kuukauden ajanjaksolta (THL, 2020b). Tästä voidaan muodostaa johtopäätös, että mikäli jokin sovellus tai käytännössä sen tuottama informaatio koetaan yleishyödylliseksi, niin siihen suhtaudutaan positiivisesti ja sen käyttö yleistyy. Vastaavan kaltainen yleinen tietoisuus liikennedatan keräämisen ja jakamisen hyödyllisyydestä voi luoda kiinnostusta ja uutta tarvetta liikennedatalle ja tämä mahdollistaa liikennedatan keräämisen hyvin kattavasti vastaavia menetelmiä hyödyntäen (Espoon kaupunki, 2019, s. 6). Tällaisen hyödyllisyyuskuvan luomisessa julkisen sektorin toimijoilla on oleellinen vaikutus, kuten ”Koronavilkku” -sovelluksenkin tapauksessa.

Sovelluspohjaisten menetelmien etuna ovat myös niiden helppous käyttäjän kannalta sekä olemattomat asennus- ja ylläpitokustannukset verrattuna perinteisiin liikennedatan lähteisiin. Ihmisten mukanaan kuljettamien älylaitteiden toimiessa datalähteinä ”ilmaisimet” ovat jo valmiiksi olemassa. Laitteiden lukumäärä ja maantieteellinen kattavuus on suurta ja menetelmä soveltuu myös kulkumuodon tunnistamiseen nopeuteen perustuen. (Leduc, 2008, s. 8)

Liikennedatan kontekstissa pienikin penetraatio eli käyttäjien tai toisin sanoen sijaintitietoa tuottavien yksiköiden lukumäärä on riittävä ja siten hyödyllinen liikennedatan lähde (Houbraken ym., 2018, ss. 2-3). Helsingin kaupungin on mahdollista toteuttaa oma sovelluksensa, jonka käytöstä voidaan mahdollisesti tarjota jonkinlainen etu käyttäjilleen sen suosion kasvattamiseksi. Sovellus voi olla ulkoisesti tai yhteistyönä toteutettu, kuten esimerkiksi julkiseen liikenteeseen tai muihin julkisiin palveluihin liittyen. Tällöin myös sovelluksen käyttäjilleen tuoma hyöty voi olla informaation ohella helpommin konkreettinen, kuten esimerkiksi alennettu joukkoliikenteen lippuhinta.

Riippumatta sovelluksen toteutustavasta, on olemassa lukuisia esimerkkejä potentiaalisesti hyödyllisistä asioista, joita kukaan ei koe tarpeellisiksi niiden näennäisestä hyödyllisyydestä huolimatta. Samanlainen ilmiö voidaan havaita monilla elämän osa-alueilla. Oleellisessa asemassa on koettu hyöty, joka on subjektiivinen tunne. Tärkein sovelluksen menestykseen vaikuttava yksittäinen tekijä on tyytyväinen käyttäjä (Berntsson-Svensson & Aurum, 2006, s. 149). Tästä syystä jokaisen tulisi voida halutessaan osallistua myös liikennedatan

tapauksessa lisäarvon muodostamiseen, esimerkiksi tiedon luoja tai jakajana, mielipiteen tai palautteen antajana tai vaikkapa osallistamalla sovelluskehitykseen.

Aidosti avoimen datan ja avoimen lähdekoodin yhdistelmä edistää ihmisten välistä yhteistyötä. Tällöin yhteistyön tuloksena syntyvä lopputulos on mahdollisesti myös enemmän ihmisten itsensä näköinen ja tarpeiden mukainen eli todennäköisemmin myös yleisesti paremmin vastaanotettu ja hyödyllisemmäksi koettu. Tämä on nähtävissä esimerkiksi Koronavilkku -sovelluksen menestyksessä, joka on seurausta sen hyödyllisyydestä ja avoimien teknologioiden käytön tuomista eduista (Lehto, 2020; Solita, 2020).

Liikennedatan potentiaalia, tarpeellisuutta ja hyötyjä voidaan korostaa esimerkiksi Helsinki Loves Developers -tyyppisillä tilaisuuksilla sekä erilaisissa sosiaalisen median kanavissa konkreettisin esimerkein. Yleisen datan jakamisprosessin mukaisesti datasta on hyödyllistä viestiä eri alustoilla ja samalla on myös mahdollista tuoda esiin jokaisen mahdollisuus vaikuttaa positiivisesti kaikille hyötyä tuovaan konseptiin (Espoon kaupunki, 2019, s. 10).

Laajasta ja avoimesta liikennedatapohjasta on mahdollista luoda monenlaisia, yleishyödyllisiä sovelluksia jokaisen käytettäväksi. Lyhentyneet matka-ajat, uudet teknologiset innovaatiot, ympäristön tilan kohentuminen päästöjen pienentyessä sekä näitä seuraavat fyysinen ja taloudellinen hyvinvointi ovat esimerkkejä asioista, joista jokainen voi hyötyä ja joita voidaan tavoitella myös liikennedatan avulla.

6.4 Liikenteen reaaliaikainen tilannekuva

Eräs Helsingin kaupungin oleellisimmista liikennedatata koskevista tavoitteista on liikenteen reaaliaikaisen tilannekuvan luominen (Helsingin kaupunki, 2019b, s. 15; Partanen, 2019). Tosiasiallisesti reaaliaikaisen liikenteen tilannekuvan muodostaminen edellyttää toimiakseen suuria, mahdollisesti erimuotoisia ja epäyhtenäisiä heterogeenisiä liikennedatamääriä useista liikennetiedon lähteistä (Korpela, 2019, s. 12). Kerättyä dataa on jalostettava ja yhdistettävä muuhun tietoon hyödyllisen informaation muodostamiseksi (Sundquist & Oesch, 2014, s. 16-18). Kerätyn raakamuotoisen liikenne datan laadun parantamiseksi hyödynnetään datan esikäsittelyä (Ramírez-Gallego ym., 2017, s. 39). Datalähteiden

tuottaman liikennetiedon ja muun tiedon, kuten paikkatiedon yhdistämiseen voidaan hyödyntää tekoälyyn perustuvia menetelmiä (Korpela, 2019, s. 12).

Toimivan liikenteen tilannekuvapalvelun toteuttamiseksi tarvitaan ajantasaista, mahdollisimman reaaliaikaista liikennedatata tieverkolta monista lähteistä - ajoneuvoista, ilmaisimista, liikennevaloista ja esimerkiksi liikennekameroista, ilmanlaadusta ja päästöistä. Pelkkä reaaliaikaisuus ei ole myöskään riittävää, vaan liikennedatan on lisäksi oltava riittävän kattavalta maantieteelliseltä alueelta ja datalähteestä riippuen siihen voi liittyä muitakin vaatimuksia, kuten tietosuojan huomioiminen. Datan laajempaa hyötykäyttöä ja sovelluskehitystä ajatellen liikennedatan tulisi myös täyttää avoimen datan kriteerit.

Esimerkiksi risteyksistä saatavan ilmaisintiedon, kuten induktioilmaisimien ja tutkien havaintojen hyödyntäminen edellyttää ilmaisimien fyysisen sijainnin assosioimista paikkatietoon, jolloin havainnoista voidaan laskea käyttökelpoisia suureita, kuten esimerkiksi ajoneuvon nopeus kahden ilmaisimen välisten havaintojen aikaeron perusteella. Käyttökelpoisen informaation saaminen edellyttää myös muun muassa riittäviä ilmaisinjärjestelyjä, jotka ovat monesti riittämättömiä, sillä ilmaisinjärjestelyt suunnitellaan usein risteysten toiminnan näkökulmasta.

Risteyksistä saatavan liikennedatan kontekstissa ilmaisimia ohjaavien laitteiden moninaisuus niin teknisen toteutustavan, valmistajan kuin laitteistojen iänkin suhteen saattaa edellyttää toimia yhteensopivuuden varmistamiseksi. Haasteita ovat vanhan ja uuden tekniikan toiminnan yhteensovittaminen, eri valmistajien laitteiden tekninen yhteensopivuus ja tiedonsiirron edellyttämien lisälaitteiden käyttö sekä myöskin laitetoimittajien kyky toteuttaa tarvittavia, eri valmistajien laitteiden välillä yhteensopivia ratkaisuja.

Tarkasteltaessa liikenteen tilannekuvapalveluita Suomessa, sisältää esimerkiksi Oulussa toteutettu ”Oulun liikenne” -palvelu reittiopastuksen lisäksi muun muassa liikennedatata hyödyntävää tilannekuvaa liikenneolosuhteista, indikaattoreita tieverkon sujuvuudesta, häiriötiedotteita tietöistä sekä mahdollisuuden tarkkailla liikenteen senhetkistä tilaa liikennekameroiden pienellä viiveellä kuvaaman liikennetilannemateriaalin avulla. Palvelu käyttää liikennetiedon lähteinään muun muassa Traffic Management Finland:n avointa liikennedatata. (Oulun kaupunki, 2020)

Vastaavan kaltaisia tilannekuvapalveluita on käytössä laajasti, muun muassa Ruotsissa Tukholmassa. Palvelun käyttämää liikennedatata kerätään alueen ilmaisimista sekä alueella liikkuviin ajoneuvoihin perustuvien menetelmien avulla (Trafikverket, 2020b). Esimerkki liikenteen tilannekuvapalvelusta on esitetty kuvassa 12.

Kuva 12. Tukholman liikenteen tilannekuvapalvelu (Trafikverket, 2020a).



Euroopassa ja maailmalla vastaavia toteutuksia on olemassa lukemattomia ja niitä yhdistävät tietyt elementit. Tyypillinen liikenteen reaaliaikaista tilannekuvaa ilmaiseva palvelu on karttapohjainen ja selaimella tai sovelluksena käytettävä, sisältäen ainakin suurimpien teiden ajantasaisen sujuvuustiedon värikoodein ilmaistuna. Tämän lisäksi palveluihin sisältyy usein väliaikaisia häiriöitä ilmaisevia informatiivisia merkkejä, joista on saatavilla lisätietoa esimerkiksi häiriötilanteen luonteesta ja arvioidusta kestosta. Alueen liikennekameroiden kuvaa on myös usein mahdollista tarkastella. Julkisen liikenteen pysäkkien sijainti- ja linjatiedot sekä mahdollisuus suunnitella reitti kohteesta toiseen sisältyvät myös useisiin tällaisiin palveluihin. Visuaalisia toteutusmahdollisuuksia on olemassa lähes rajattomasti.

Tilannekuvapalveluihin sisältyvien tai erillisten reittiopastuspalveluiden tyyppisenä käyttötarkoituksena on löytää paras reitti sijainnista toiseen senhetkisessä liikennetilanteessa. Reittivalintaa voidaan usein painottaa erilaisten tavoitteiden, kuten lyhimmän vaadittavan ajan tai kuljettavan matkan perusteella. Reitin muodostamiseen voidaan käyttää erilaisia menetelmiä.

Mikäli reitin arvioimiseen ja muodostamiseen käytetään ainoastaan staattista tietoa, vallitsevaa liikennetilannetta ei voida ottaa huomioon reittiä muodostettaessa. Dynaamista tietoa hyödynnettäessä reitti on mahdollista muodostaa vain lähtöhetkellä vallitsevan liikennetilanteen perusteella tai uudelleenreitityksien avulla, jolloin reittiä voidaan tarvittaessa muuttaa liikennetilanteen muutosten mukaisesti. Liikennetilanteen muutoksia voidaan pyrkiä ennustamaan ennen niiden tapahtumista nopean ja keskeytyksettömän reitin muodostamiseksi. (Falek ym., 2020, s. 5)

Kehittyneempien reitinmuodostusmenetelmien käyttäessä lukuisia erilaisia liikennedatan lähteitä sekä monimutkaisia algoritmeja, reitin muodostamisen tiedonsiirto- ja laskentatehovaatimukset kasvavat (Falek ym., 2020, s. 2). Perusmuodossaan lyhimmän reitin muodostamiseen voidaan käyttää esimerkiksi Dijkstran algoritmia, joka ei kuitenkaan huomioi negatiivisia arvoja eli esimerkiksi reitin sisällä olevia liikennetilanteesta johtuvia keskeytyksiä, kuten ruuhkia (Sembiring ym., 2018, s. 1). Toinen reitin muodostamisessa yleisesti käytettävä menetelmä on Floyd-Warshall -algoritmi (Risald ym., 2017, s. 155).

On huomionarvoista, että mikäli riittävän suuri määrä ajoneuvoja käyttää samaa, tietyllä ajanhetkellä parhaaksi arvioitua reittiä samanaikaisesti, muuttuu se ruuhkautumisen vaikutuksesta tavoitteen kannalta heikommaksi reittivaihtoehdoksi kapasiteetin ylittyessä. Tällöin on tarpeellista reitittää tietty osa ajoneuvoista joko lähtötilanteessa tai kesken matkan senhetkisen tilanteen kannalta huonommalle reitille keskimääräisesti parhaimman lopputuloksen saamiseksi suurimmalle osalle ajoneuvoista. Esimerkiksi ajallisesti pitkäkestoisen ruuhkan tai liikenteen häiriötilanteen, kuten onnettomuuden aiheuttama reitin merkittävä hidastuminen voi käytännössä tehdä näennäisesti huonosta reitistä tavoitteen kannalta optimaalisimman vaihtoehdon. Reitin valinnan kokonaisuus monimutkaistuu myös, mikäli tietyt kulkumuotoja, kuten joukkoliikennettä halutaan priorisoida.

Kokonaisuutena liikenteen reaaliaikaista tilannekuvaa tuottavat palvelut sekä niiden liikennetiedon lähteinä olevat ilmais- ja havainnointimenetelmät ovat usein selkeästi kokeiluasteella. Perehdyttäessä eri liikkumismuotojen havaitsemiseen tarkoitettuihin menetelmiin on havaittavissa, että uusia innovaatioita syntyy ja kokeiluja toteutetaan jatkuvasti. Teknologiset edistysaskeleet, kuten 5G ja koneoppiminen, eivät ole vielä saavuttaneet täyttä potentiaaliaan liikenteenkään kontekstissa. Maailmalla suoritetaan jatkuvasti uutta tutkimusta ja pilottihankkeita, joissa selvitetään uusien algoritmien ja potentiaalisten ratkaisujen soveltuvuutta alati kasvavan liikenteen tuomiin haasteisiin.

Helsingin kaupungilla on selkeät tavoitteet ja visio tulevaisuuden liikenteestä alueellaan, joihin kuuluu oleellisesti myös liikennedatan hyödyntäminen. Tällä hetkellä kokeilut, yhteistyöprojektit ja erinäiset hankkeet tuovat lisää tietoa asetettujen tavoitteiden saavuttamiseksi ja niitä tarvitaan myös jatkossa soveltuvimpien toimintatapojen löytämiseksi.

Lähteet

Amin-Naseri, M. (2018). *Adopting and incorporating crowdsourced traffic data in advanced transportation management systems*. [Väitöskirja, Iowa State University]. Haettu 6.12.2020 osoitteesta

<https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=7548&context=etd>

Avoinrajapinta.fi. (2014). Avoimen rajapinnan määritelmä. Haettu 4.12.2020 osoitteesta

<http://avoinrajapinta.fi/>

Berntsson-Svensson, R. & Aurum, A. (2006). Successful software project and products: An empirical investigation. *ISESE '06: Proceedings of the 2006 ACM/IEEE international symposium on Empirical software engineering*, 144-153. Haettu 10.12.2020

osoitteesta <https://doi.org/10.1145/1159733.1159757>

BITRE. Bureau of Infrastructure and Transport Research Economics. (2014). New traffic data sources - An overview. Haettu 23.11.2020 osoitteesta

<https://www.bitre.gov.au/sites/default/files/2019-12/NewDataSources-BackgroundPaper-April%202014.pdf>

Carlson, E. (2020). What Will 5G Bring?. Haettu 2.12.2020 osoitteesta

<https://reader.elsevier.com/reader/sd/pii/S2095809920301351>

Choi, W., Min, J., Lee, T., Hyun, K., Lim, T. & Jung, S. (2018). Deep Representation of Raw Traffic Data: An Embed-and-Aggregate Framework for High-Level Traffic Analysis.

Haettu 29.11.2020 osoitteesta https://doi.org/10.1007/978-981-10-7605-3_220

Dalgleish, M. & Hoose, N. (2008). *Highway Traffic Monitoring and Data Quality*. Norwood: Artech House Publishers.

- De Beer, J. (2016). Ownership of Open Data: Governance Options for Agriculture and Nutrition. Haettu 27.11.2020 osoitteesta https://www.researchgate.net/publication/319018708_Ownership_of_Open_Data_Governance_Options_for_Agriculture_and_Nutrition
- Deloitte. (2016). InnoTiku A Loppuraportti. Helsingin liikenteen reaaliaikainen tilannekuva – Liikennekameroiden videokuvan analyysipilotti. Haettu 22.12.2020 osoitteesta https://www.hel.fi/static/kanslia/Innovaatorahasto/2018/InnoTiku_Loppuraportti_Deloitte.pdf
- Diamond Traffic Products. (2020). Infrared Sensors. Haettu 5.12.2020 osoitteesta <https://diamondtraffic.com/technicaldescription/126>
- Eco-Counter. (2019). Real-time Displays. Haettu 5.12.2020 osoitteesta <https://www.eco-compteur.com/en/produits/real-time-displays/eco-display-classic-2/>
- Espoon kaupunki. (2019). Datan tietosuoja ja tietoturvallinen avaaminen. Haettu 30.11.2020 osoitteesta https://www.espoo.fi/materiaalit/espoo_kaupunki/verkkolehti/Datan-tietosuoja-ja-tietoturvallinen-avaaminen/html5/index.html
- ESRI. (2020). What is traffic data? Haettu 27.9.2020 osoitteesta <https://desktop.arcgis.com/en/arcmap/latest/extensions/network-analyst/traffic-what-is-traffic-data.htm>
- EU:n tietosuojaryhmä. (2014). Lausunto 5/2014 anonymisointitekniikoista. Haettu 2.12.2020 osoitteesta https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_fi.pdf
- Eyupoglu, C., Aydin, M., Zaim, A. & Sertbas, A. (2018). An Efficient Big Data Anonymization Algorithm Based on Chaos and Perturbation Techniques. *Entropy*, 20(5), 373. Haettu 2.12.2020 osoitteesta <https://doaj.org/article/7d8f3dfa8cf04a7cb737a370757a2611>

- Falek, A., Gallais, A., Pelsser, C., Julien, S. & Theoleyre, F. (2020). To Re-Route, or not to Re-Route: Impact of Real-Time Re-Routing in Urban Road Networks. Haettu 25.12.2020 osoitteesta <https://www.theoleyre.eu/uploads/Publis/ITS-dynamic-rerouting.pdf>
- Fedorov, A., Nikolskaia, K., Ivanov, S., Shepelev, V. & Minbaleev, A. (2019). Traffic flow estimation with data from a video surveillance camera. *Journal of Big Data*, 6(73). Haettu 5.12.2020 osoitteesta <https://doi.org/10.1186/s40537-019-0234-z>
- Feng, Y., Hourdos, J. & Davis, G. (2014). Probe vehicle based real-time traffic monitoring on urban roadways. *Transportation Research Part C: Emerging Technologies*, 40/2014, 160-178. Haettu 28.11.2020 osoitteesta <https://doi.org/10.1016/j.trc.2014.01.010>
- Finlex. (2015). Tietosuojalautakunta 4/2015. Haettu 3.12.2020 osoitteesta <https://www.finlex.fi/fi/viranomaiset/ftie/2015/20150004>
- Gu, Y., Qian, Z. & Chen, F. (2016). From Twitter to detector: Real-time traffic incident detection using social media data. *Transportation Research Part C: Emerging Technologies*, 67/2016, 321-342. Haettu 4.12.2020 osoitteesta <https://doi.org/10.1016/j.trc.2016.02.011>
- Helsingin kaupunki. (2019a). Helsingin datastrategia. Haettu 26.11.2020 osoitteesta https://digi.hel.fi/documents/248/Helsingin_kaupungin_datastrategia_LOPULLINEN.pdf
- Helsingin kaupunki. (2019b). Helsingin älyliikenteen kehittämisohjelma 2030. Haettu 25.11.2020 osoitteesta <https://dev.hel.fi/paatokset/media/att/6f/6fc1dd80054c8cdc56ae8e973858df19c6bc6383.pdf>
- Helsingin kaupunki. (2019c). Helsingin liikkumistottumukset 2018. Haettu 30.11.2020 osoitteesta <https://www.hel.fi/static/liitteet/kaupunkiymparisto/julkaisut/julkaisut/julkaisu-07-19.pdf>

Helsingin kaupunki. (2019d). Tilastotietoja Helsingistä 2019. Haettu 7.12.2020 osoitteesta https://www.hel.fi/hel2/tietokeskus/julkaisut/pdf/19_06_14_HKI-taskutilasto2019_suomi_w.pdf

Helsinki Region Infoshare. (2017). Datan avaajalle. Haettu 12.12.2020 osoitteesta <https://hri.fi/fi/ohjeet/datan-avaajalle/>

Hokajärvi, J. & Nieminen, I. (2018). Mestarintunnelin liikennevirran analysointi LiDAR-tutkalla. Haettu 5.12.2020 osoitteesta https://vayla.fi/documents/25230764/0/Unikie_loppuraportti/8ddf0089-a5a1-4727-887c-3b049cf35f9a

Houbraken, M., Logghe, S., Audenaert, P., Colle, D. & Pickavet, M. (2018). Examining the potential of Floating Car Data for Dynamic Traffic Management. *IET Intelligent Transport Systems*, 12(5), 335. Haettu 30.11.2020 osoitteesta <http://dx.doi.org/10.1049/iet-its.2016.0230>

HSL. Helsingin seudun liikenne (2020). Avoin data. Haettu 6.12.2020 osoitteesta <https://classic.hsl.fi/avoindata>

IBM. (2020). Application Programming Interface (API). Haettu 4.12.2020 osoitteesta <https://www.ibm.com/cloud/learn/api>

Isotalo, K. (2017). Liikennedatua pontevalle PK-yritykselle. Haettu 3.12.2020 osoitteesta https://www.maanmittauslaitos.fi/sites/maanmittauslaitos.fi/files/attachments/2017/08/Liikennedatua_pontevalle_PK-yritykselle.pdf

Korpela, E. (2019). Väylänpito tehostuu tekoälyn ja paikkatiedon avulla. Haettu 6.12.2020 osoitteesta <https://www.maanmittauslaitos.fi/sites/maanmittauslaitos.fi/files/attachments/2019/03/Vaylanpito-tehostuu-tekoalyn-ja-paikkatiedon-avulla.pdf>

- Kulmala, M. (2019). Kaupunkinäkökulma: Kannattaako datatalous? Mitä hyötyä liikenteen datan jakaminen ja alustauudistus voi tuoda?. Haettu 4.12.2020 osoitteesta https://www.hankintakeino.fi/sites/default/files/media/file/%C3%84liikenteen%20kehitt%C3%A4j%C3%A4ryhm%C3%A4_Mika%20Kulmala_1.pdf
- Lehto, L. (2020). Koronavilkku ennätysellisen suosittu, kiitos avoimen lähdekoodin. Haettu 10.12.2020 osoitteesta <https://coss.fi/uutiset/koronavilkku-ennatysellisen-suositu-kiitos-avoimen-lahdekoodin/>
- Leskinen, A., Jokinen, O. & Rintamäki, J. (2016). Joukkoliikenteen matkustajainformaation koontipalvelu. Esiselvitys. Haettu 1.12.2020 osoitteesta https://julkaisut.vayla.fi/pdf8/lts_2016-37_joukkoliikenteen_matkustajainformaation_web.pdf
- Leurent, F. & Nguyen, T. (2010). Network optimization with dynamic traffic information and tolling: A probabilistic model and economic analysis. Haettu 29.11.2020 osoitteesta https://www.researchgate.net/publication/229034930_Network_optimization_with_dynamic_traffic_information_and_tolling_a_probabilistic_model_and_economic_analysis
- Leduc, G. (2008). Road Traffic Data: Collection Methods and Applications. Haettu 22.11.2020 osoitteesta https://www.researchgate.net/publication/254424803_Road_Traffic_Data_Collection_Methods_and_Applications
- Lee, K. & Sener, I. (2019). Emerging data for pedestrian and bicycle monitoring: Sources and applications. *Transportation Research Interdisciplinary Perspectives*, 4/2020. Haettu 18.11.2020 osoitteesta <https://doi.org/10.1016/j.trip.2020.100095>
- Liikenne- ja viestintävirasto Traficom. (2020). 5G Muuttaa liikennettä. Haettu 2.12.2020 osoitteesta <https://www.traficom.fi/fi/ajankohtaista/podcastit/5g-muuttaa-liikennetta>

- LIITE1_liikennedatat. (2020). MS Excel-tiedosto. Helsingin kaupungin liikennedatan lähteet 31.8.2020. Saatu sähköpostin liitteenä Helsingin kaupungin liikenneinsinööri M. Lehtoselta 23.9.2020.
- Linturi, R. & Kuittinen, O. (2014). Liikennetiedon visiot. Liikenneviraston tutkimuksia ja selvityksiä 42/2014. Haettu 9.11.2020 osoitteesta https://julkaisut.vayla.fi/pdf8/lts_2014-42_liikennetiedon_visiot_web.pdf
- Liu, K., Wang, W. & Wang, J. (2019). Pedestrian Detection with Lidar Point Clouds Based on Single Template Matching. *Electronics*, 8(7), 780. Haettu 10.12.2020 osoitteesta <https://doi.org/10.3390/electronics8070780>
- Lopes, J., Huang, E., Antoniou, C. & Ben-Akiva, M. (2010). Traffic and Mobility Data Collection for Real-time Applications. *13th International IEEE Conference on Intelligent Transportation Systems*. Haettu 17.11.2020 osoitteesta <https://doi.org/10.1109/ITSC.2010.5625282>
- Majchrakova, D. (2016). Real-time Traffic Information - How Does it Work?. Haettu 28.11.2020 osoitteesta <https://www.sygi.com/blog/2016/real-time-traffic-information-how-does-it-work>
- Mamdouh, R. (2007). *Data Preparation for Data Mining Using SAS*. San Francisco: Elsevier.
- MGDS. Marine Geoscience Data System. (2020). Frequently-Asked Questions about data terminology. Haettu 4.12.2020 osoitteesta https://www.marine-geo.org/help/data_FAQ.php
- Mikrobitti. (2020) 5G-verkko kattaa koko Suomen viiden vuoden päästä. Haettu 2.12.2020 osoitteesta <https://www.mikrobitti.fi/uutiset/5g-verkko-kattaa-koko-suomen-viiden-vuoden-paasta/6ac8ac8a-42c9-44e6-82ca-b9a4772f90cd>

- Mohsin, M., Wong, W. & Bhatt, Y. (2001). Support for Real-Time Traffic in the Internet, and QoS Issues. Haettu 1.12.2020 osoitteesta <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.10.8075&rep=rep1&type=pdf>
- Narh, A., Thorpe, N., Bell, M. & Hill, G. (2016). Do new sources of traffic data make the application of Chaos Theory to traffic management a realistic possibility?. *Transport Reviews*, 36(5), 635-658. Haettu 19.11.2020 osoitteesta <https://doi.org/10.1080/01441647.2016.1140687>
- Naumann, S., Wolf, F. & Schönrock, R. (2010). Floating Car Observer - An innovative vehicle-sensor for urban and highway scenarios. *IFAC Proceedings Volumes*, 43(23), 19-24. Haettu 10.12.2020 osoitteesta <http://dx.doi.org/10.3182/20101005-4-RO-2018.00008>
- Niinikoski, M. & Laine, T. (2005). Liikennetietojen käyttökohteet ja kehittämistarpeet. Tiehallinnon selvityksiä 3/2005. Haettu 21.11.2020 osoitteesta https://julkaisut.vayla.fi/pdf/3200913-vliikennetietojen_kayttokohteet.pdf
- NPRA. The Norwegian Public Roads Administration. (2004). Traffic Data Collection and Analysis. Haettu 27.11.2020 osoitteesta <https://www.vegvesen.no/attachment/336339/binary/585485>
- Open Knowledge Foundation. (2020a). How to Open up Data. Haettu 3.12.2020 osoitteesta <https://opendatahandbook.org/guide/en/how-to-open-up-data/>
- Open Knowledge Foundation. (2020b). What is open?. Haettu 3.12.2020 osoitteesta <https://okfn.org/opendata/>
- Oulun kaupunki. (2020). Oulunliikenne.fi -palvelu. Tietoja palvelusta. Haettu 9.12.2020 osoitteesta <https://oulunliikenne.fi/tietoja-palvelusta>

Partanen, P. (2020). Datasta helpotusta Helsingin ruuhkiin. Haettu 4.12.2020 osoitteesta

<https://hri.fi/fi/datasta-helptusta-helsingin-ruuhkiin/>

Pastinen, V. (2012). Matkakertomus: International Conference on Transport Survey Methods, ISCTSC. WSP Finland Oy. Haettu 20.11.2020 osoitteesta

<https://www.traficom.fi/sites/default/files/media/file/Matkakertomus-lyhennelma-ISCTSC-2011.pdf>

Pihlajamaa, O., Lahti, J., Heino, I. & Lusikka, T. (2020). Joukkoliikenteen

matkatietopalveluiden digitaalinen infrastruktuuri. Haettu 28.11.2020 osoitteesta

https://cris.vtt.fi/files/35580168/VTT_MJDI_Tutkimusraportti_20201019_signed.pdf

Poikola, A., Kuikkaniemi, K., Kuittinen, O., Honko, H. & Knuutila, A. (2018). MyData - johdatus ihmiskeskeiseen henkilötiedon hyödyntämiseen. Haettu 27.11.2020 osoitteesta

<https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160954/MyData%202018.pdf>

Polzin, S. (2016). Implications to Public Transportation of Emerging Technologies. Haettu

27.11.2020 osoitteesta <https://doi.org/10.5038/CUTR-NCTR-RR-2016-10>

Ramboll. (2016). Helsingin liikenteen reaaliaikainen tilannekuva - osa-alue B loppuraportti

21.12.2016. Haettu 9.12.2020 osoitteesta

https://www.hel.fi/static/kanslia/Innovaatorahasto/2018/InnoTiku_Loppuraportti_Ramboll.pdf

Ramírez-Gallego, S., Krawczyk, B., García, S., Wozniak, M. & Herrera, F. (2017). A survey on data preprocessing for data stream mining: Current status and future directions.

Neurocomputing, 239, 39-57. Haettu 4.12.2020 osoitteesta

<https://doi.org/10.1016/j.neucom.2017.01.078>

- Risald, R., Mirino, A. & Suyoto, S. (2017). Best routes selection using Dijkstra and Floyd-Warshall algorithm. *International Conference on Information, Communication Technology and System (ICTS)*, 155-158. Haettu 25.12.2020 osoitteesta <http://dx.doi.org/10.1109/ICTS.2017.8265662>
- Roppo, R. & Sutinen M. (2016). Innotiku Pilotti C. Liikkuvan videokuvan analysointi. Loppuraportti. Haettu 22.12.2020 osoitteesta https://www.hel.fi/static/kanslia/Innovaatorahasto/2018/InnoTiku_Loppuraportti_Tieto.pdf
- Ruppe, S., Junghans, M., Haberjahn, M. & Troppenz, C. (2012). Augmenting the Floating Car Data Approach by Dynamic Indirect Traffic Detection. *Procedia - Social and Behavioral Sciences*, 48/2020, 1525-1534. Haettu 10.12.2020 osoitteesta <http://dx.doi.org/10.1016/j.sbspro.2012.06.1128>
- Saarinen, J. & Partanen, A. (2020). Avoimen datan mahdollisuuksia liikenteessä. Teoksessa J. Kukkamäki & M. Tarkkala (toim.), *Avoim Häme*. HAMK Unlimited Professional 4.6.2020. Haettu 2.12.2020 osoitteesta <https://unlimited.hamk.fi/uncategorized/avoimen-datan-mahdollisuuksia-liikenteessa>
- Sembring, P., Harahap, A. & Zalukhu, K. (2018). Implementation of Dijkstra's algorithm to find an effective route to avoid traffic jam on a busy hour. *Journal of Physics: Conference Series*, 1116(2), 1-4. Haettu 25.12.2020 osoitteesta <https://doi.org/10.1088/1742-6596/1116/2/022042>
- Solita. (2020). THL ja Solita julkaisivat Koronavilkun lähdekoodin - avoin koodi mahdollistaa sovelluksen riippumattoman arvioinnin. Haettu 10.12.2020 osoitteesta <https://www.solita.fi/thl-ja-solita-julkaisivat-koronavilkun-lahdekoodin/>
- Strom, E. (2019). How everyone can benefit from data sharing of real-time traffic information. Haettu 3.12.2020 osoitteesta <https://www.asmag.com/showpost/28147.aspx>

- Suh, W., Anderson, J., Guin, A. & Hunter, M. (2015). Evaluation of video detection system as a traffic data collection method. *Scientia Iranica*, 22(6), 2092-2102. Haettu 27.11.2020 osoitteesta http://scientiairanica.sharif.edu/article_3757.html
- Sundquist, H. & Oesch, K. (2014). DATAjalostus. Digitaalisten palvelujen tuottamisen perusta. Haettu 29.11.2020 osoitteesta http://fi.okfn.org/files/2014/05/Datajalostus_White-Paper.pdf
- Suomen virallinen tilasto (SVT). (2019). Väestön tieto- ja viestintätekniikan käyttö. Haettu 30.11.2020 osoitteesta <https://www.stat.fi/til/sutivi/index.html>
- Tang, Y., Zhang, C., Gu, R., Li, P. & Yang, B. (2017). Vehicle detection and recognition for intelligent traffic surveillance system. *Multimedia Tools and Applications*, 76(4), 5817-5832. Haettu 6.12.2020 osoitteesta <https://doi.org/10.1007/s11042-015-2520-x>
- tekoäly.info. (2020). Mitä tekoäly on?. Haettu 6.12.2020 osoitteesta https://tekoaly.info/mita_tekoaly_on/
- THL. Terveyden ja hyvinvoinnin laitos. (2020a). Koronavilkku-sovellus. Haettu 7.12.2020 osoitteesta <https://koronavilkku.fi/>
- THL. Terveyden ja hyvinvoinnin laitos. (2020b). Koronavilkku on ladattu yli 2,5 miljoonaa kertaa - laaja käyttö lisää sovelluksen vaikuttavuutta. Haettu 7.12.2020 osoitteesta <https://thl.fi/fi/-/koronavilkku-on-ladattu-yli-2-5-miljoonaa-kertaa-laaja-kaytto-lisaa-sovelluksen-vaikuttavuutta>
- Tietosuojavaltuutetun toimisto. (2020). Henkilötietojen minimointi tieteellisessä tutkimuksessa. Haettu 2.12.2020 osoitteesta <https://tietosuoja.fi/henkilotietojen-minimointi-tieteellisessa-tutkimuksessa>
- Trafikverket. (2020a). Trafiken.nu - Stockholm. Haettu 22.12.2020 osoitteesta <https://trafiken.nu/stockholm/>

Trafikverket. (2020b). Trafiken.nu - Vanliga frågor. Haettu 22.12.2020 osoitteesta

<https://trafiken.nu/stockholm/sidfot/vanliga-fragor/>

Tung, T. & Walrand, J. (2003). Providing QoS for Real-time Applications. Haettu 26.11.2020 osoitteesta

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.67.4171&rep=rep1&type=pdf>

Turán, J., Ovsenik, L. & Harasthy, T. (2012). Traffic Sign Recognition System based on Cambridge Correlator Image Comparator. *Carpathian Journal of Electronic and Computer Engineering*, 5(1), 127-132. Haettu 6.12.2020 osoitteesta

<https://doaj.org/article/6affa4e615d9424daf4deaec50253ed4>

Urbiotica. (2020). Wireless magnetometer sensor for traffic management. Haettu 5.12.2020 osoitteesta [https://www.urbiotica.com/en/traffic-monitoring/wireless-](https://www.urbiotica.com/en/traffic-monitoring/wireless-magnetometer-sensor-for-traffic-management/)

[magnetometer-sensor-for-traffic-management/](https://www.urbiotica.com/en/traffic-monitoring/wireless-magnetometer-sensor-for-traffic-management/)

U.S. Department of Transportation. (2019). Considerations of Current and Emerging Transportation Management Center Data. Haettu 16.11.2020 osoitteesta

<https://ops.fhwa.dot.gov/publications/fhwahop18084/fhwahop18084.pdf>

Väylävirasto. (2020). LAM-tiedot. Haettu 5.12.2020 osoitteesta

<https://vayla.fi/vaylista/aineistot/avoindata/tiestotiedot/lam-tiedot>

Wang, H., Ouyang, M., Meng, Q. & Kong, Q. (2020) A traffic data collection and analysis method based on wireless sensor network. *EURASIP Journal on Wireless Communications and Networking*, 2/2020. Haettu 17.11.2020 osoitteesta

<https://doi.org/10.1186/s13638-019-1628-5>

Wasser, L. (2020). The Basics of LiDAR - Light Detection and Ranging - Remote Sensing.

Haettu 5.12.2020 osoitteesta <https://www.neonscience.org/resources/learning-hub/tutorials/lidar-basics>

Wolford, B. (2020). What is GDPR, the EU's new data protection law?. Haettu 3.12.2020 osoitteesta <https://gdpr.eu/what-is-gdpr/>

Wu, J., Xu, H., Tian, Y., Pi, R. & Yue. R. (2020). Vehicle Detection under Adverse Weather from Roadside LiDAR Data. *Sensors* 2020, 20(12). Haettu 17.11.2020 osoitteesta <https://doi.org/10.3390/s20123433>

Yakimov, M. (2018). Internet technologies for traffic and pedestrian data collection. *Transportation Research Interdisciplinary Perspectives*, 4/2020. Haettu 25.11.2020 osoitteesta <http://dx.doi.org/10.1016/j.trip.2020.100095>

Ylitepsa, J. (2018). Työelämän tietosuojalaki. Haettu 3.12.2020 osoitteesta https://vm.fi/documents/10623/10947434/Juhta_ty%C3%B6el%C3%A4m%C3%A4n_tietosuoja_19112018_Ylitepsa.pdf

Zins, C. (2007). Conceptual Approaches for Defining Data, Information and Knowledge. *Journal of the American Society for Information Science and Technology*, 58(4), 479-493. Haettu 26.11.2020 osoitteesta <https://doi.org/10.1002/asi.20508>

