

# Avoimen datan avaamisprosessi ja sen haasteet

Case: Lahden kaupunki

LAHDEN  
AMMATTIKORKEAKOULU  
Liiketalouden ala  
Tietojenkäsittelyn koulutusohjelma  
Opinnäytetyö  
Syksy 2017  
Juha Korhonen

Lahden ammattikorkeakoulu  
Tietojenkäsittelyn koulutusohjelma

KORHONEN, JUHA:

Avoimen datan avaamisprosessi ja  
sen haasteet  
Case: Lahden kaupunki

Tietojenkäsittelyn opinnäytetyö, 44 sivua, 1 liitesivu

Syksy 2017

TIIVISTELMÄ

---

Opinnäytetyö käsittelee avointa dataa ja sen avaamisprosessin haasteita. Työ tehtiin toimeksiantona Lahden kaupungille ja sen tavoitteena oli edistää Lahden kaupungin eri yksiköiden tuottaman datan avaamista avoimeksi dataksi löytämällä avaamisprosessin keskeisimmät haasteet ja ratkaisut haasteille. Opinnäytetyö toteutettiin kvalitatiivisena tutkimuksena, jossa hyödynnettiin deduktiivista päättelytapaa. Työn tietoperusta koostuu tieteellisistä julkaisuista sekä asiantuntijoiden tuottamista materiaaleista. Tutkimusaineiston keräämiseen käytettiin teemahaastatteluja.

Tutkimuksen tuloksena havaittiin, että Lahden kaupungin datan avaamisprosessin haasteita ovat tiedon vähäinen kysyntä ja tietämättömyys siitä, millaista tietoa kehittäjät tarvitsevat. Tämän takia rajallisia resursseja ei ehkä haluta kuluttaa datan avaamiseen, koska varmuutta datan avaamisen hyödyistä ei ole. Muita esiin nousseita haasteita olivat myös teknisen osaamisen puute, taloudelliset näkökulmat sekä vaikeasti avattavissa oleva arkaluontoinen tieto.

Tulosten perusteella Lahden kaupungin kannattaisi tehdä yhteistyötä yritysten, kehittäjien ja datayhteisöjen kanssa, selvittääkseen minkälaiselle datalle löytyy kysyntää. Myös datan löydettävyyteen ja markkinointiin tulisi panostaa, jotta yritykset ja kehittäjät olisivat tietoisia Lahden kaupungin tarjoamista avoimista tietoaaineistoista ja alkaisivat hyödyntää niitä omassa toiminnassaan. Kaupungin toimintamalleja tulisi myös yhtenäistää siten, että ne tukisivat paremmin datan avaamista ja tulevaisuuden tietojärjestelmähankinnoissa kannattaisi kiinnittää huomiota datan avaamisen mahdollistaviin ratkaisuihin.

Asiasanat: avoin data, tietoaaineisto, tietovaranto

Lahti University of Applied Sciences  
Degree Programme in Information Technology

KORHONEN, JUHA: Open data and the challenges in the  
process of opening up data  
Case: The city of Lahti

Bachelor's Thesis in Information Technology, 44 pages, 1 page of  
appendix

Autumn 2017

ABSTRACT

---

This thesis discusses open data and the challenges in the process of opening up data. The thesis was commissioned by the city of Lahti. The main goal was to improve the city's process of opening up its data by finding the related key challenges and defining solutions to these challenges.

The thesis was conducted as a qualitative study. Empirical data was collected through semi-structured interviews. The theoretical background materials about open data were used as sources to analyze the gathered empirical data. The theoretical material was gathered from research publications and from open data experts' publications.

The two main challenges in opening up of data in the city of Lahti were the lack of demand for data and the lack of knowledge on what kind of data developers need. In addition, there were challenges regarding technical know-how, financial matters and confidential information.

Based on the results of the thesis, the city of Lahti should cooperate with companies, developers and open data communities to find out the demand of its data. The city of Lahti should invest in the better findability of data as well as marketing. After findability has been improved, companies and developers can find and use more of the city's data in their projects. The operational models of the city should be standardized to better support the opening up data. Finally, information systems should be improved to make the opening up of data easier in the future.

Keywords: open data, dataset, data resource

## SISÄLLYS

1	JOHDANTO	1
2	TUTKIMUKSEN TAUSTA	3
2.1	Tutkimustehtävä ja tutkimuskysymys	3
2.2	Tietoperusta ja keskeiset käsitteet	4
3	AVOIN DATA	5
3.1	Mitä data on?	5
3.2	Avoimen datan määritelmä	6
3.3	Dataformaatit	7
3.4	Avoimen datan laadun arvioiminen	9
3.5	Lisenssit	10
3.6	Datan avaamisen hyödyt	11
3.7	Haasteet datan avaamiselle	12
4	DATAN AVAAMISPROSESSI	15
4.1	Kartoittaminen ja selvittäminen	16
4.2	Arviointi ja valitseminen	17
4.3	Valmistelu ja avaaminen	20
4.4	Seuraaminen ja kehittäminen	22
5	TUTKIMUSMENETELMÄT	23
5.1	Tutkimusaineiston keruu- ja analyysimenetelmät	23
5.2	Tutkimusprosessi	24
6	TUTKIMUSAINEISTO JA SEN ANALYYSI	26
6.1	Haasteet maankäytön yksikössä	27
6.2	Haasteet ympäristövalvonnan yksikössä	29
6.3	Haasteet ympäristöterveyden yksikössä	32
6.4	Ilmenneiden haasteiden vertailu	36
7	JOHTOPÄÄTÖKSET	38
8	POHDINTA	40
	LÄHTEET	41
	LIITTEET	45

## 1 JOHDANTO

Julkishallinto tuottaa Suomessa suuren määrän dataa, joka hyödyttäisi paljon myös hallinnon ulkopuolisia toimijoita. Valitettavasti suuri osa näistä tietovarannoista on vielä vain harvojen ulottuvilla. Näiden tietovarantojen avaamisella avoimeksi dataksi voitaisiin saavuttaa parhaimmillaan suuria yhteiskunnallisia hyötyjä. Avoin data tarjoaa uusia mahdollisuuksia esimerkiksi sovelluskehittäjille sekä yrityksille, jotka voivat hyödyntää dataa omien palveluidensa kehittämiseen. Näin syntyy uusia innovaatioita, uutta yritystoimintaa sekä kansalaisten jokapäiväistä elämää helpottavia palveluja. (Valtiovarainministeriö 2017.)

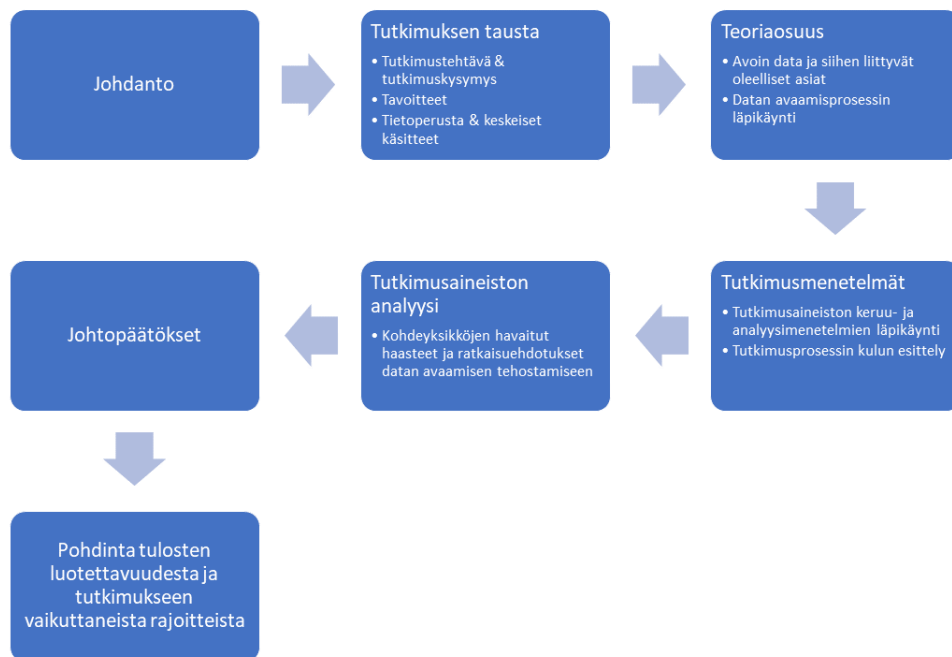
Tässä opinnäytetyössä käsitellään avointa dataa ja sen avaamisprosessia. Työ on tehty toimeksiantona Lahden kaupungille ja sen tavoitteena on edistää kaupungin eri yksiköiden tuottamien data-aineistojen avaamista. Työn tutkimustehtävä on rajattu koskemaan Lahden kaupungin datan avaamisprosessin haasteita. Tutkimustehtävä on rajattu näin, jotta Lahden kaupunki tunnistaisi datan avaamiseen liittyvät haasteet sekä löytäisi niihin sopivat ratkaisut. Näiden ratkaisujen avulla Lahden kaupunki voisi tehostaa data-aineistojensa avaamista, mikä taas hyödyttäisi monia kaupungin ulkopuolisia toimijoita kuten yrityksiä, kehittäjiä sekä tavallisia kansalaisia. Suomessa osa kaupungeista on jo avannut tietovarantojaan avoimeksi dataksi. Esimerkiksi niin sanottujen kuutoskaupunkien yhteisessä Avoin data ja rajapinnat -kärkihankkeessa Helsinki, Espoo, Vantaa, Turku, Tampere sekä Oulu ovat näin jo tehneet, käyttäen ja kehittämällä yhteisiä toimintamalleja. (Gaika 2017.)

Opinnäytetyö toteutettiin kvalitatiivisena eli laadullisena tutkimuksena, jossa hyödynnettiin deduktiivista eli teorialähtöistä päättelytapaa. Tutkimusaineiston keräämiseen käytettiin teemahaastatteluja, jotka tehtiin Lahden kaupungin teknisen ja ympäristötoimialan eri yksiköiden vastuuhenkilöille.

Opinnäytetyön rakenne on seuraavanlainen (kuvio 1): aluksi esitellään tutkimuksen tausta, jonka jälkeen käydään läpi tutkimustehtävä,

selvennetään tutkimuskysymys ja tutkimuksen tavoitteet sekä esitellään tietoperusta ja keskeiset käsitteet. Tätä seuraa teoriaosuus, jossa tutustutaan avoimeen dataan käsitteenä sekä käydään läpi datan avaamisprosessia. Teoriaosuuden tarkoituksena on perehdyttää lukija avoimeen dataan, antaa esimerkkejä avoimen datan hyödyistä ja haasteista sekä kuvata kuinka datan avaamisprosessin voisi toteuttaa. Lähteenä käytetään tieteellisiä julkaisuja sekä asiantuntijoiden tuottamaa materiaalia.

Teoriaosuuden jälkeen esitellään tutkimusmenetelmät, jonka jälkeen analysoidaan tutkimukseen kerättyä aineistoa ja vertaillaan saatuja tuloksia keskenään. Viimeiset kaksi kappaletta pitävät sisällään johtopäätökset tutkimuksen tuloksista sekä pohdintaosuuden tulosten luotettavuudesta sekä tutkimukseen vaikuttaneista rajoitteista.



KUVIO 1. Opinnäytetyön rakenne

## 2 TUTKIMUKSEN TAUSTA

Tämän opinnäytetyön aiheeksi valikoitui avoin data ja datan avaamisprosessin haasteet. Aihe työlle tuli Lahden kaupungilta, jolla oli tarvetta avoimeen dataan liittyvään tutkimukseen. Lahden kaupunki on jo avannut joitain yksittäisiä tietoaaineistojaan julkiseen käyttöön ja tulevaisuudessa tavoitteena olisi avata dataa vieläkin laajemmin. Datan avaamisprosessin haasteisiin aihe tarkentui kirjoittajan oman mielenkiinnon, toimeksiantajan tarpeiden ja aiheen ajankohtaisuuden takia, sillä monet organisaatiot ovat jo avanneet tietovarantojaan julkisiksi tai suunnittelevat sitä, ja kuuluuhan pääministeri Juha Sipilän hallitusohjelmaankin tavoite luoda avoimella datalla ja tietovarantojen tehokkaalla hyödyntämisellä uutta liiketoimintaa Suomeen (Valtiovarainministeriö 2017).

### 2.1 Tutkimustehtävä ja tutkimuskysymys

Työn tutkimustehtävä rajattiin koskemaan Lahden kaupungin eri yksiköiden tuottaman datan avaamisen haasteita, jotta datan avaamista vaikeuttavat asiat tunnistettaisiin sekä löydetäisiin niihin sopivat ratkaisut. Tutkimuskysymyksenä käytettiin kysymystä:

- Mitä haasteita Lahden kaupungin eri yksiköiden tuottaman datan avaamisessa on ja miten avaamista voitaisiin edistää?

Kysymys on luonteeltaan tutkiva kysymys, jonka avulla pyritään saamaan käsitys Lahden kaupungin yksiköiden datan avaamista vaikeuttavista asioista ja löytämään ratkaisut datan avaamisen helpottamiseksi. Tässä työssä ei tehdä tarkkaa suunnitelmaa Lahden kaupungin tuottaman datan avaamiseksi, vaan tarkoitus on käydä asioita yleisellä tasolla läpi ja antaa esimerkkien avulla ehdotuksia Lahden kaupungille datan avaamisprosessin tehostamiseen. Datan tehokkaammasta avaamisesta olisi hyötyä niin yrityksille, kehittäjille kuin tavallisille kansalaisillekin. Tutkimuksen päätavoitteena onkin edistää Lahden kaupungin eri yksiköiden tuottamien data-aineistojen avaamista avoimeksi dataksi.

## 2.2 Tietoperusta ja keskeiset käsitteet

Tämän opinnäytetyön tietoperusta sisältää avoimen datan määritelmän, datan avaamiseen liittyvien oleellisten asioiden esittelyn, datan avaamisen tunnettujen hyötyjen ja haasteiden läpikäynnin sekä avaamisprosessin eri vaiheet. Tietoperustaan valikoitui tutkittavan ilmiön eli avoimen datan ja sen avaamisprosessin kannalta keskeisimmät asiat. Tutkimus toteutettiin datan avaamisprosessin ja sen haasteiden näkökulmasta, mutta jotta avaamisprosessi on helpommin ymmärrettävissä, on tietoperustaan otettu mukaan myös avoimeen dataan yleisesti liittyviä asioita. Tietoperusta pohjautuu tieteellisiin julkaisuihin sekä avoimen datan asiantuntijoiden tuottamiin materiaaleihin.

Tämän opinnäytetyön kannalta keskeisimpiä ja tekstissä toistuvia käsitteitä ovat seuraavat:

- Tutkimuskysymyksessä ja usein tässä tutkimuksessakin esiintyvä termi data tarkoittaa digitaalista raaka-ainetta, joka voi olla esimerkiksi tilastoja, dokumentteja, kuvia, karttoja tai videoita (Poikola, Kola & Hintikka 2010, 14).
- Tietoaineisto ja data-aineisto tarkoittavat talletettujen tietojen muodostamaa kokonaisuutta eli vaikkapa yksittäistä taulukkoa. Tietovaranto taas tarkoittaa loogista tietojen kokonaisuutta, esimerkiksi rekisteriä, joka on jonkin tietyn toimijan hallinnassa. (Toikkanen, Kalliala, Poikola & Sillanpää 2014a.)
- Dataportaali ja datakatalogi viittaavat datan julkaisuun ja löydettävyyteen. Dataportaali on datan jakelualusta, jossa organisaatiot voivat julkaista tuottamaansa dataa. Datakatalogi taas on luettelo esimerkiksi jonkin organisaation avoimena olevista tietoaineistoista ja se sisältää myös kuvaukset aineistojen sisällöstä. (Vakkari 2013, 9.)

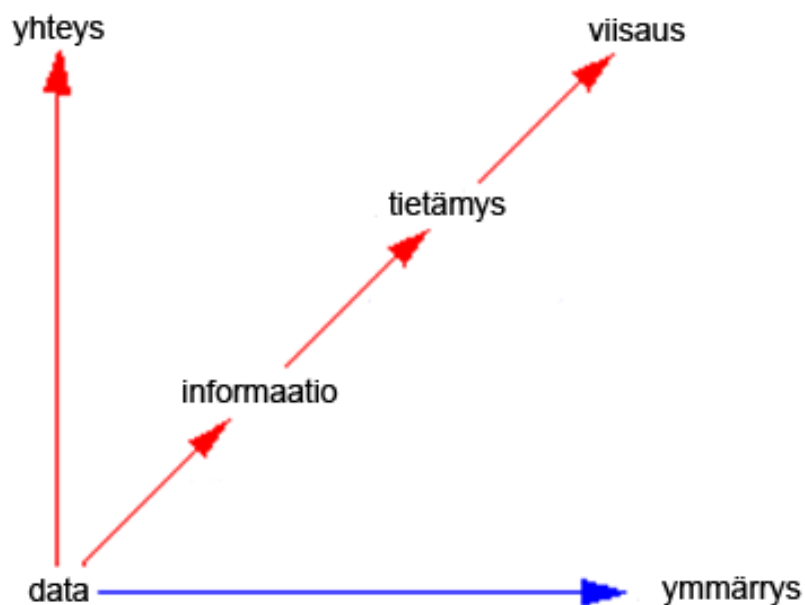


### 3 AVOIN DATA

Avoin data on ilmiönä vielä melko uusi asia. Kansainvälisesti datan avaaminen vauhdittui vuonna 2009 ja edelläkävijöinä kulkivat Yhdysvallat sekä Iso-Britannia, joiden esimerkkiä seuraten muutkin maat, mukaan lukien Suomi, alkoivat avata tietovarantojaan julkiseen käyttöön. Avointa dataa pidetään ajatusmallina, joka lisää ja automatisoi julkishallinnon, kansalaisten sekä kehittäjien välistä vuorovaikutusta ja tuottaa kaikkia osapuolia hyödyttäviä ratkaisuja. (Toikkanen ym. 2014a.)

#### 3.1 Mitä data on?

Poikolan ym. (2010, 13-14) mukaan suomen kielessä termit data, informaatio ja tieto merkitsevät lähes samaa asiaa. Tämän takia näiden termien ymmärtäminen ja merkityserojen havaitseminen voi olla hankalaa. Bellinger, Castro ja Mills (2004, 1-2) taas käsittelevät tietoa yhtenä isona kokonaisuutena, joka jakautuu neljään eri tasoon kuten kuviossa 2 on kuvattu. Nämä osat ovat data, informaatio, tietämys ja viisaus.



KUVIO 2. Tiedon eri tasot (Bellinger ym. 2004, 3)

Bellinger ym. (2004, 2-3) kuvaavat dataa raakatiedoksi, joka voi olla esimerkiksi jokin merkki, merkkijono tai luku. Sellaisenaan datalla ei ole merkitystä vaan se vaatii tulkintaa ja jatkokäsittelyä, jotta se voidaan ymmärtää. Datasta muodostuu jatkokäsittelyn jälkeen informaatiota, joka voidaan välittää eteenpäin, ja jonka vastaanottaja voi ymmärtää ja omaksua osaksi tietämystään. Kun vastaanottaja käyttää tätä tietämystä hyväkseen esimerkiksi arkipäivän valinnoissaan, voidaan sitä sanoa viisaudeksi.

Poikola ym. (2010, 14) taas määrittelevät datan digitaaliseen muotoon talletetuksi informaatioksi. Se voi olla esimerkiksi kuvia, ääni- tai videotiedostoja, dokumentteja tai tietokantoja. Arkipuheessa datalla viitataan usein tietojärjestelmissä olevaan informaatioon, jota voidaan käsitellä tietokoneiden avulla. Data voidaan ymmärtää hieman eri tavoin käyttötarkoituksesta riippuen, esimerkiksi avoimen datan yhteydessä data on raaka-ainetta, josta voidaan jalostaa uusia sovelluksia ja innovaatioita kansalaisten käyttöön.

### 3.2 Avoimen datan määritelmä

Kun puhutaan avoimesta datasta, on ymmärrettävä miten se eroaa julkisesta tiedosta. Julkinen tieto on tietoa esimerkiksi PDF-muodossa organisaation kotisivuilla, jota käyttäjät pääsevät lukemaan, lataamaan ja jakamaan. Julkinen tieto ei ole koneluettavassa muodossa, eikä sitä ole lisensoitu niin että sitä voisi vapaasti käyttää, muokata ja jakaa muokattuna eteenpäin. Avoin data taas on data-aineisto, jota kuka tahansa voi maksutta käyttää, muokata ja jakaa ilman käyttöä rajoittavia tekijöitä. (Helsinki Region Infoshare 2010.)

Open Knowledge Definition -määritelmä on kehitetty arvioimaan yksittäisen julkisen data-aineiston avoimuutta sekä uudelleenkäytettävyyttä. Se sisältää kriteerit joiden avulla tunnistetaan, voidaanko aineisto luokitella avoimeksi dataksi vai ei. Dataa voidaan avata, jos sitä eivät koske erilaiset rajoitteet kuten esimerkiksi yksityisyydensuoja. Avattava data-aineisto ei saa siis sisältää

henkilötietoja tai muuta arkaluontoista tietoa. Avoin data-aineisto on myös helposti löydettävissä ja saatavilla kokonaisuudessaan kellonajasta riippumatta ja kenellä tahansa tulee olla oikeus vapaasti ja maksutta käyttää aineistoa omiin tarkoituksiinsa. Datan tulee myös olla koneluettavassa muodossa, jotta se on käyttö- ja muokkauskelpoista. Ainoa datan käyttäjältä edellytettävä asia on korkeintaan se, että datan lähde mainitaan lisenssiehdoista selviävällä tavalla. Näin varmistetaan, että datan tuottaja saa nimensä esille ja käyttäjä saa tiedon datan alkuperästä. (Poikola ym. 2010, 34.)

Jotta julkaistun avoimen datan luotettavuus ja merkitys voidaan osoittaa datan käyttäjälle, on data-aineistoon liitettävä kuvaus sen sisällöstä. Tämä onnistuu dataan liitettävän kuvailutiedon, metadatan avulla. Metadata kertoo datan hyödyntäjälle esimerkiksi mitä tietoa data sisältää, mihin aikaväliin tieto pohjautuu, kuka on tuottanut ja julkaissut tiedon, minkä laatuista tieto on ja miten tietoa saa hyödyntää. (Toikkanen ym. 2014a.)

Kuvailutieto voidaan tehdä datan tavoin koneluettavaan muotoon metadatastandardeilla, mikä tekee tietojärjestelmien välisen tiedonsiirron helpommaksi, mahdollistaa eri sisältöjen tehokkaamman yhdistelyn sekä parantaa tietoaineiston löydettävyyttä. Kun metadata on pitkälle standardoitu, voidaan puhua linkitetystä avoimesta datasta, jonka avulla pystytään esittämään eri aiheita koskevia tietoja ja kehittämään entistä innovatiivisempia sovelluksia. (Toikkanen ym. 2014a; Euroopan Unionin avoimen datan portaali 2017.)

### 3.3 Dataformaatit

Dataa voidaan julkaista monissa eri tiedostomuodoissa, mutta kaikki formaatit eivät kuitenkaan sovi avoimen datan julkaisuun. Esimerkiksi PDF-, Word- tai HTML-dokumentit eivät ole suositeltavia formaatteja, sillä ne esittävät dataa lähinnä visuaalisessa muodossa ja niiden ohjelmallinen käsittely on hankalaa. (Ylärinne 2013, 10.)

Taulukkoon 1 on kerätty Poikolan ym. (2010), Ylärinteen (2013) sekä Kotkasen (2016) antamia esimerkkejä avoimen datan julkaisuun käytetyistä tiedostomuodoista.

TAULUKKO 1. Esimerkkejä avoimen datan tiedostomuodoista

<b>XML</b>	Yleiskäyttöinen, käyttöjärjestelmäriippumaton merkintäkieli, jolla kuvataan datan sisältöä eli metadataa (Poikola ym. 2010, 64; Ylärinne 2013, 28).
<b>CSV</b>	Avoin tiedostomuoto, jolla tallennetaan taulukkomuotoista tietoa tekstitiedostoon. Taulukon eri kentät erotellaan pilkkujen ja rivinvaihtojen avulla (Poikola ym. 2010, 64; Ylärinne 2013, 33).
<b>JSON</b>	Avoimen standardin tekstipohjainen tiedostomuoto, jolla välitetään tietoa (Poikola ym. 2010, 64; Kotkanen 2016, 33).
<b>RDF</b>	Standardoitu tietomalli ja tapa esittää metadataa. Mahdollistaa sovellusten välisen tiedon vaihdon etenkin web-ympäristössä (Poikola ym. 2010, 64).
<b>XSL</b>	XML-dokumenttien muotoilustandardi, jolla voidaan muuntaa XML-dokumentti toiseen formaattiin (Ylärinne 2013, 35).
<b>KML</b>	Googlen kehittämä XML-pohjainen tiedostomuoto, jolla kuvataan karttoja ja niihin liittyviä paikkatietoja (Ylärinne 2013, 40).
<b>GML</b>	XML-pohjainen formaatti paikkatietojen esittämiseen (Ylärinne 2013, 41).
<b>GeoJSON</b>	JSON-formaattiin pohjautuva tiedostomuoto paikkatietojen kuvaamiseen (Kotkanen 2016, 34).
<b>GTFS</b>	Joukkoliikenteeseen liittyvien tietojen jakoon käytettävä formaatti (Kotkanen 2016, 34).
<b>Excel</b>	Tunnettu taulukkolaskentaohjelma, jonka avulla datan muotoilu on helppoa. Excel-tiedostomuoto (XLS) ei itsessään täytä avoimen datan vaatimuksia, mutta Excelistä tieto voidaan tuoda ulos CSV-tiedostona. (Ylärinne 2013, 25.)

### 3.4 Avoimen datan laadun arvioiminen

Arvioitaessa avoimen data-aineiston laadukkuutta, kiinnitetään huomiota aineiston dataformaattiin ja siihen kuinka pitkälle metadata on standardoitu. Datan laadukkuuden arviointiin on olemassa eri tapoja, joista yksi on World Wide Webin kehittäjän Tim Berners-Leen viiden tähden arviointimalli (kuvio 3). Jotta datan voidaan sanoa olevan avointa dataa, tulee sen saada viidestä mahdollisesta tähdestä vähintään kolme. (Koski, Honkanen, Luukkonen, Pajarinen, Ropponen 2017, 57.)

Tähdet	Kuvaus
★☆☆☆☆	Data on saatavissa missä tahansa muodossa avoimella lisenssillä (esimerkiksi PDF-tiedosto tai skannattu kuva)
★★☆☆☆	Data on rakenteellisessa muodossa (esimerkiksi excel-taulukko)
★★★☆☆	Data on saatavilla avoimessa tiedostoformaattissa (esimerkiksi CSV, JSON, XML)
★★★★☆	Datassa käytetään yksilöllisiä tunnisteita, joihin voi viitata (URI)
★★★★★	Data sisältää linkkejä toisiin datoihin (sisällöllisiä linkityksiä)

KUVIO 3. Avoimen datan viiden tähden arviointimalli (Koski ym. 2017, 57)

Yhden ja kahden tähden data-aineistot ovat saatavilla internetistä avoimella lisenssillä ja niitä voidaan vapaasti lukea, ladata sekä jakaa. Yhden tähden aineistot voivat olla esimerkiksi PDF- tai HTML-dokumentteja ja kahden tähden aineistot rakenteellisessa muodossa olevia Excel-tiedostoja. Jotta data luokiteltaisiin kolmella tähdellä, tulee sen olla koneluettavassa muodossa ja sitä voi helposti muokata ja käyttää omiin tarkoituksiinsa. Sopiva tiedostomuoto on esimerkiksi CSV, sillä se ei ole sidottu mihinkään ohjelmistoon. Neljän ja viiden tähden aineistoissa tietojen yhdistely eri lähteistä on mahdollista. Neljän tähden aineistoksi luokiteltavan datan tietoalkiot sisältävät URI:n eli merkkijonon joka kuvaa tiedon sijainnin. Näin tiedon sijaintiin voidaan viitata sekä data-aineiston osiin pystytään linkittämään suoraan myös muualta. Viiden tähden aineisto taas on linkitettyä dataa, jossa tietoalkiot on liitetty tiettyyn tietomalliin. Tieto pystytään ymmärtämään koneellisesti ja sen yhdistely internetistä

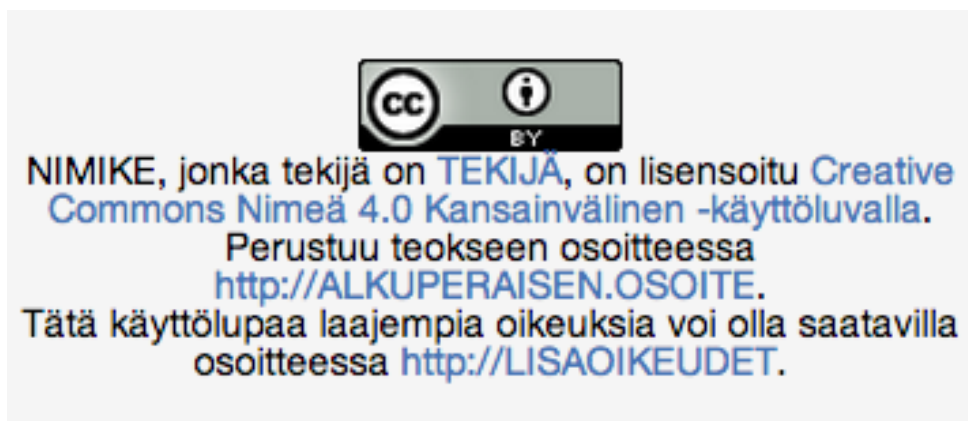
löytyviin ja samoja tietomalleja hyödyntäviin aineistoihin on helppoa. (Toikkanen ym. 2014a.)

### 3.5 Lisenssit

Suomessa julkishallinnon avoimen datan lisenssisuositus (JHS 189 Avoimen tietoaineiston käyttöluupa) on Creative Commons Nimeä 4.0 (engl. CC BY 4.0). Sen lisäksi tilanteissa joissa ei nähdä tarpeelliseksi ilmoittaa datan tuottajaa tai yksilöidä dataa tarkemmin, voidaan käyttää myös CC0-lisenssiä. (Avoindata.fi 2016a.)

CC-lisenssit ovat maailman yleisimmin käytetyt avoimen sisällön lisenssit. Creative Commons Nimeä 4.0 lisenssillä julkaistua dataa saa kuka tahansa ladata, kopioida, muokata ja jakaa kaupalliseen tai ei-kaupalliseen käyttöön, mutta datan tuottajan nimi on aina mainittava alkuperäisenä tekijänä. Jos datan tuottaja on julkaissut datan CC0 -lisenssillä, tarkoittaa se sitä, että hän on luovuttanut sen vapaaseen yleiseen käyttöön ja luopunut lain sallimissa rajoissa kaikista oikeuksistaan julkaistuu dataan. Tällöin dataa voi vapaasti ladata, muokata ja jakaa, eikä käyttäjän tarvitse mainita edes datan lähdettä. CC0-lisenssiä suositellaan käytettäväksi erityisesti metadatalle. Näiden kahden lisenssin lisäksi CC-lisenssejä on olemassa myös muita, mutta niiden käsittely ei tässä tutkimuksessa ole oleellista, sillä ne eivät täytä avoimen datan vaatimuksia. (Creative Commons Suomi 2017a.)

CC-lisenssin merkintään kannattaa käyttää Creative Commonsin luomaa lisenssivalitsinta, jonka avulla lisenssin muotoseikat saadaan varmimmin oikein. Valitsin tarjoaa lisenssimerkinnät verkkosivuille liitettävään merkintään, painetun teoksen merkintään sekä XML-tiedoston lisenssin upottamiseen data-aineiston metatietoihin. Kuvassa 1 on esimerkki siitä, millainen on data-aineiston sisältämälle verkkosivulle tehty lisenssimerkintä. (Creative Commons Suomi 2017b.)



KUVA 1. Esimerkki lisenssin merkinnästä data-aineiston sisältämällä verkkosivulla (Creative Commons Suomi 2017b)

### 3.6 Datan avaamisen hyödyt

Datan avaamisella voidaan saavuttaa hyötyjä, jotka vaikuttavat niin datan avaajaan, käyttäjään kuin koko yhteiskuntaankin. Toikkanen ym. (2014b) sanovat julkishallinnon datan avaamisen vaikuttavan demokratian ja hallinnon läpinäkyvyyteen, kansalaisten aktiivisuuteen, uusien innovaatioiden ja palveluiden syntymiseen sekä organisaatioiden toiminnan tehostumiseen. Schrier (2014, 19-21) on myös samaa mieltä avoimen datan hyödyistä ja nostaa lisäksi esiin tehokkaamman tiedon jakamisen, kansalaisten osallistamisen, hallinnon suorituskyvyn parantamisen sekä liiketoiminnan ja talouden kehittymisen.

Datan avaaminen lisää esimerkiksi datajournalismia, joka vahvistaa demokratiaa ja edistää tietoyhteiskunnan kehittymistä. Datajournalismin avulla kansalaisten tietoisuus ja ymmärrys paranevat ja kiinnostus demokratiaa kohtaan lisääntyy. Datan avaaminen lisää myös hallinnon läpinäkyvyyttä, kun hallinnon toimintaan liittyvä data on kaikkien saatavilla. Tämä taas vähentää korruptiota ja väärinkäytöksiä. Kun hallinnon toiminta on läpinäkyvää, toiminta tehostuu ja tulosvastuullisuus kasvaa. Tällöin myös mahdolliset kehityskohteet tulevat paremmin esiin ja pystytään tuottamaan entistä parempia palveluita. (Toikkanen ym. 2014b.)

Avoin data vaikuttaa myös positiivisesti kansalaisten aktiivisuuteen. Avoimen datan avulla kansalaiset voivat pelkän tiedon vastaanottamisen sijaan hyödyntää tietoa myös omiin tarkoituksiinsa. Tämä taas edistää asioiden ymmärtämistä ja parantaa kansalaisten osallistumismahdollisuuksia. (Toikkanen ym. 2014b.) Schrier (2014, 19-21) nostaa esimerkiksi myös sen, että datan avaamisella voidaan osallistaa kansalaisia kehittämään esimerkiksi asuinalueitaan ja varautumaan asuinympäristöä koskeviin uhkiin.

Toikkasen ym. (2014b) mukaan avoin data luo uusia innovaatioita ja kehittää sitä kautta liiketoimintaa. Tätä tukee myös valtioneuvoston selvitys- ja tutkimustoiminnan julkaisema raportti, jonka mukaan avointa dataa palveluiden luomiseen tai kehittämiseen hyödyntäneet informaatio- ja viestintäalan yritykset kasvattivat liikevaihtoaan vuosina 2012-2014 keskimäärin yli 17 prosenttia enemmän kuin vastaavan alan yritykset, jotka eivät dataa toiminnassaan hyödyntäneet. Tutkimuksen mukaan avointa dataa käyttäneet yritykset myös kehittivät uusia innovaatioita huomattavasti useammin kuin dataa käyttämättömät yritykset. (Koski ym. 2017, 2.)

Avoimella datalla on vaikutuksensa myös organisaatioiden sisäiseen toimintaan. Kun organisaation toiminnasta syntyvää tietoa avataan avoimeksi dataksi, kasvattaa se työntekijöiden ymmärrystä organisaation toiminnasta ja siihen liittyvistä prosesseista sekä tietojärjestelmistä. Tämä taas kehittää palveluiden laatua ja auttaa myös suuntaamaan resursseja tarkemmin tuottavampiin työtehtäviin. (Toikkanen ym. 2014b.)

### 3.7 Haasteet datan avaamiselle

Organisaation datan avaamiseen liittyy monia haasteita, joista useimmat johtuvat tiedon puutteesta. Ei esimerkiksi tiedetä mitä avoin data tarkoittaa vaan sekoitetaan se julkiseen tietoon ja luullaan, että pelkkä PDF-tiedosto kaikkien luettavilla on sama asia kuin avoin data. Toisekseen kuvitellaan, että datan avaaminen vaatii paljon taloudellisia sekä henkilöresursseja, vaikka alkuun voitaisiin avata pieniä tietoaineistoja jotka eivät vaadi suuria



uhrauksia. Monet eivät myöskään ymmärrä sitä, että datan avaamiseen käytetty aika ja vaiva tuovat paljon hyötyä niin itselle osaamisen kasvamisena kuin organisaatiolle toiminnan kehittymisenä. (Toikkanen ym. 2014d.)

Martinin (2015) tekemässä tutkimuksessa kerrotaan, että hallinnon tuottaman datan avaamisen haasteita ovat myös kysyntä- ja tarjontapuolen ongelmat. Hallinto ei kykene tuottamaan riittävän laadukasta ja ymmärrettävää dataa ja tämä taas vähentää käyttäjien kiinnostusta ja datan kysyntää. Organisaatioiden olisikin kyettävä todistamaan tuottamansa datan arvo käyttäjille, jotta nämä saataisiin kiinnostumaan datan hyödyntämisestä yhä enemmän.

Teknisen puolen haasteena voi olla mielikuva siitä, että datan avaaminen vaatii paljon osaamista. Pienten yksittäisten aineistojen kohdalla osaamista ei välttämättä tarvita kovinkaan paljon vaan dataa voidaan muokata esimerkiksi Excelissä, tallentaa CSV-muotoon ja ladata tiedosto jakoon organisaation verkkosivuille, datakatalogiin tai dataportaaliin. Jos organisaatio on avaamassa suurempia aineistoja, tällöin kannattaa miettiä rajapintojen luomista ja avaamista, jotka sitten vaativatkin jo hieman enemmän teknistä osaamista. Jos osaamista ei organisaation sisältä löydy, täytyy se hankkia siinä tapauksessa ulkopuolelta. Tästä taas saattaa aiheutua enemmän kustannuksia, jolloin taloudelliset resurssit voivat rajoittaa datan avaamista. (Toikkanen ym. 2014d.)

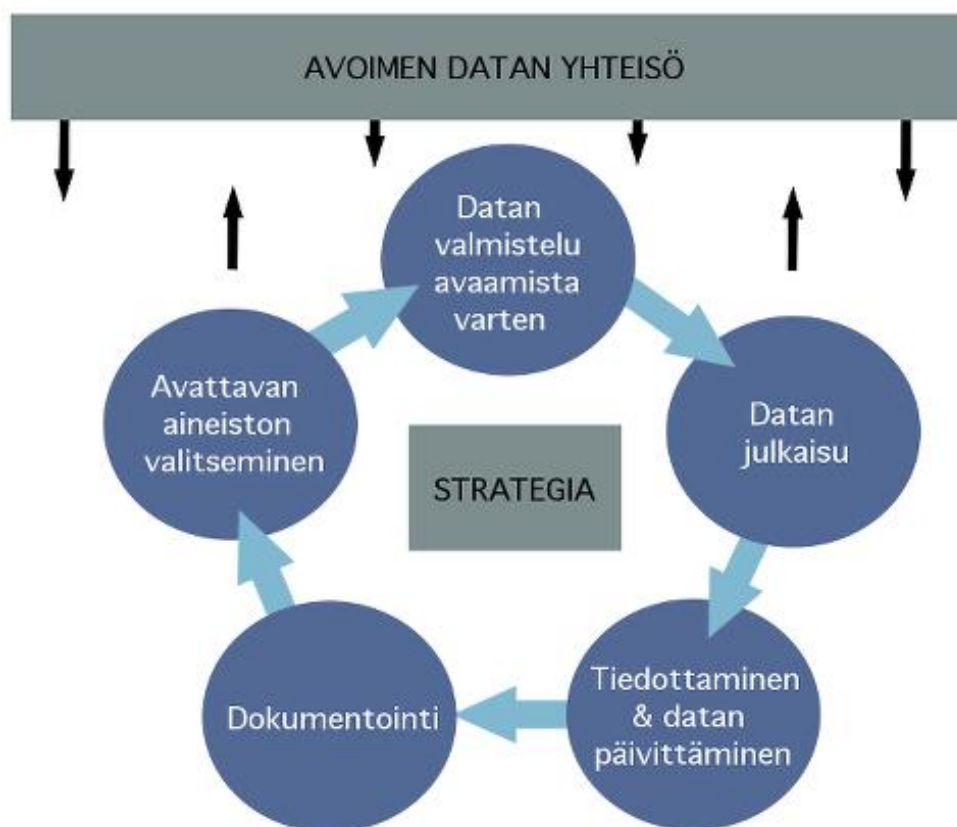
Datan saaminen ulos tietojärjestelmistä ei välttämättä ole aina helppoa. Tietoa voi olla hajallaan monessa eri paikassa ja eri formaatissa, jolloin tiedon yhdistely ja muokkaaminen käyttökelpoiseksi vie aikaa. Joskus tietojärjestelmään voi olla tarpeen luoda lisäosa, jonka avulla data saadaan ulos järjestelmästä oikeassa formaatissa. (Lahti, Parkkinen & Lehtomäki 2012.) Myös Poikola ym. (2010, 43) nostavat esiin tietoarkkitehtuurista syntyvät haasteet eli sen, että tietoa tuotetaan todella paljon ja epäyhtenäisesti. Pahimmassa tapauksessa organisaatiossa ei tiedetä mitä tietoa missäkin tietojärjestelmässä on ja samaa tietoa voidaan kerätä ja päivittää monessa eri paikassa.

Hallinnon näkökulmasta haasteena voi olla esimerkiksi tietämättömyys siitä, voiko ja kannattaako dataa avata. Lisäksi ei välttämättä myöskään tiedetä, kuka tai ketkä päättävät datan avaamisesta ja ketkä kaikki osallistuvat avaamisprosessiin. Laki asettaa myös omat haasteensa avaamiselle, sillä on otettava huomioon, ettei data sisällä henkilötietoja tai muuta salassapidettävää tietoa. Myös tekijänoikeudet ja lisenssit tulee selvittää, tarvittaessa lakimiehen avulla. Suomessa lähioikeudet suojaavat tietokantoja, mikä tarkoittaa, että esimerkiksi tietokannassa olevien kuvien oikeudet on selvitettävä ja lupa avaamiseen on hankittava. (Toikkanen ym. 2014d.)

Myös pelko seurauksista voi olla haasteena datan avaamiselle. Pelot voivat liittyä esimerkiksi datan sisältämiin virheisiin tai väärinkäyttöön. Väärinkäyttöä voidaan kuitenkin ehkäistä kattavalla metadatalalla. Virheiden pelko on myös turhaa, ja niiden havaitsemista varten kannattaakin luoda palautealusta. Palautteen pohjalta virheet ja vanhat tiedot voidaan korjata tai päivittää. Myöskään suurta palautemäärää ei tule nähdä negatiivisena asiana, vaan palaute kertoo, että data todella kiinnostaa ja sille on käyttöä. (Toikkanen ym. 2014d.)

#### 4 DATAN AVAAMISPROSESSI

Kun organisaatio harkitsee tuottamansa datan avaamista julkiseksi, on datan avaamisprosessi suunniteltava huolella ja läpikäytävä datan avaamiseen liittyvät rajoitteet ja haasteet. Jokaisen tietoaiteiston kohdalla avaamisprosessi on jatkuva, sillä se ei pääty datan avaamiseen kuten kuvioista 4 voidaan huomata. (Toikkanen ym. 2014c.)



KUVIO 4. Datan avaaminen on jatkuva prosessi (Toikkanen ym. 2014c)

Avattaessa dataa ensimmäisen kerran, on hyvä aloittaa avaamalla vain hyvin rajattu tietovarantojoukko. Näin havaitaan mitä korjattavaa prosessissa on ja onko organisaatio riittävän kyvykäs ylipäätään avaamaan dataa. Saadun palautteen pohjalta on ongelmakohdat helpompi havaita ja korjata, jonka jälkeen tietovarantoja voidaan avata vieläkin laajemmin. Datan avaamisprosessi voidaan jakaa neljään eri vaiheeseen,

jotka ovat kartoittaminen ja selvittäminen, arviointi ja valitseminen, valmistelu ja avaaminen sekä seuraaminen ja kehittäminen. (Avoindata.fi 2016f.)

#### 4.1 Kartoittaminen ja selvittäminen

Datan avaamisprosessissa on hyvä lähteä liikkeelle kartoittamalla mitä tietojärjestelmiä ja millaisia tietovarantoja organisaatiolla on.

Tietojärjestelmien kartoittaminen on helpompaa, jos organisaatiolla on olemassa tietojärjestelmäluettelo josta kaikki käytössä olevat järjestelmät selviävät. Kun käytettävät järjestelmät ovat tiedossa, on tietovarantojenkin paikantaminen helpompaa. (Toikkanen ym. 2014c.)

Tietovarantojen rajoitteet ja velvoitteet kannattaa ottaa selville, jotta varmistutaan että tieto on ylipäättään avattavissa. Tällaisia rajoitteita ovat esimerkiksi henkilötietolaki, tekijänoikeuteen liittyvät rajoitukset, tekniset rajoitteet tai tiedon maksullisuus. Lisäksi on tärkeää ymmärtää, onko kaikki data tuotettu itse vai onko se tai osa siitä saatu muualta. Jos tietoa on saatu muualta, on varmistuttava, että sen avaamiseen on olemassa lupa. Tietovarantojen kartoituksen lisäksi on hyvä selvittää, ketkä kaikki tulevat osallistumaan datan avaamiseen ja mistä asioista he ovat vastuussa. (Avoindata.fi 2016b.)

Tietovarantojen kartoituksen jälkeen on hyvä jakaa ne avattuihin, avattavissa oleviin, rajoitteita sisältäviin sekä sellaisiin tietovarantoihin, joita ei voi avata. Tällä jaottelulla varmistetaan, ettei salassa pidettävää materiaalia pääse julkaistavaksi. (Toikkanen ym. 2014c.) Rajoitteita sisältäviä tietoja voidaan muokata niin, että niistä saadaan julkaisukelpoisia. Esimerkiksi henkilötietoja voidaan anonymisoida, jolloin niiden julkaisua ei enää estä henkilötietolaki. (Mäkinen 2012, 6.)

Ensimmäisen vaiheen jälkeen organisaatiolla tulisi olla selkeä käsitys sen käytössä olevista tietojärjestelmistä, tietovarannoista ja niiden rajoituksista sekä avaamisprosessista vastuussa olevista henkilöistä. (Avoindata.fi 2016b.)

## 4.2 Arviointi ja valitseminen

Tietovarantojen ja niiden avattavuuden ollessa selvillä, on hyvä pohtia strategisia tavoitteita datan avaamiselle. On tärkeää miettiä, miksi halutaan avata dataa, mitä tavoitteita datan avaamisella halutaan saavuttaa sekä mitä seurauksia datan avaaminen mahdollisesti aiheuttaa. On hyvä myös miettiä mitä datan avaamisella halutaan välttää, jotta ei saada aikaan organisaation toimintaa uhkaavaa vahinkoa. (Kolehmainen 2014.)

Datan avaaminen voi olla innovaatio- tai liiketoimintalähtöistä tai sitten pakon sanelemaa. Innovaatiolähtöisessä mallissa tiedon tarpeesta ja käyttäjistä ei ole alussa tietoa. Organisaatiolla on halu avata dataa ja nähdä minkälaisia sovelluksia sen myötä syntyy. Alkuun kannattaakin lähteä julkaisemaan dataa maltillisesti ja uusien innovaatioiden syntyessä kehittää tiedonjakoa ja vaihtaa sitten liiketoimintalähtöiseen malliin. Liiketoimintalähtöisessä mallissa organisaatiolla on tieto, että sen avaamalle datalle on käyttöä ja myös datan avaamisen muoto on selvillä. Tällöin tiedonjulkaisuun panostaminen on perusteltua. Jos datan avaaminen taas on pakon sanelemaa, kannattaa organisaation tehdä se mahdollisimman yksinkertaisesti ja pienillä resursseilla. (Avoindata.fi 2016c.)

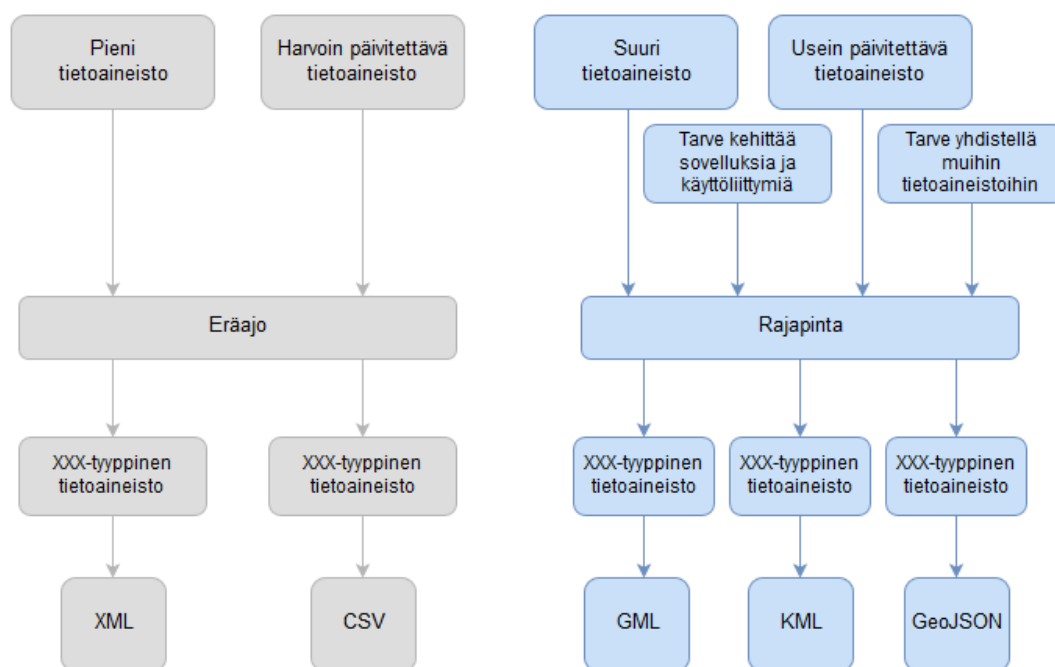
Avattavasta datasta on tunnistettava, mikä tieto on arvokasta ja mikä taas ei. Tiedon arvoa määrittäessä tärkeimmät asiat ovat tiedon uudelleenkäytettävyys sekä tiedon arvo sen tuottajalle ja käyttäjälle. Uudelleenkäytettävyyteen hyvä mittari on jo edellä mainittu Tim Berners-Leen kehittämä viiden tähden arviointimalli (kuvio 3). Tiedon tuottajan puolesta tiedon julkaisu-arvoa voidaan pohtia seuraavien asioiden kautta, joista yhdenkin kohdan toteutuminen on merkki tiedon arvokkuudesta:

- Tiedon julkaisu lisää läpinäkyvyyttä.
- Laki velvoittaa tiedon julkaisuun.
- Tiedolla on merkitys tiedon tuottajan julkiseen tehtävään.
- Kulut vähenevät tiedon julkaisun myötä. (Avoindata.fi 2016b.)

Tiedon mahdolliset käyttäjät tai kehittäjäyhteisöt kannattaa pitää mukana datan avaamisprosessissa heti alusta alkaen. Näin saadaan selvyys siitä, minkälaisille data-aineistoille löytyy kysyntää ja hyödyllisten data-aineistojen valitseminen helpottuu. Hyvä tapa saada palautetta tiedon arvokkuudesta, on julkaista tietovarannoista katsaus, josta käyttäjät selvästi ymmärtävät millaista dataa organisaatio tuottaa. Julkaisukanavaa valittaessa on myös hyvä ottaa huomioon kohderyhmä eli tiedon käyttäjät, jotta julkaistu avoin data tavoittaa siitä kiinnostuneet. Julkaisukanavana voi olla esimerkiksi yleinen dataportaali tai sitten organisaation oma datakatalogi. (Toikkanen ym. 2014c.)

Teknisiin valintoihin vaikuttavat tiedon koko ja ominaisuudet.

Tietovarannoista kannattaa selvittää, onko tieto staattista eli pysyvää tietoa vaiko dynaamista eli reaaliaikaista tietoa. Staattinen tieto ei päivity kovinkaan usein, kun taas dynaaminen tieto muuttuu jatkuvasti. Tästä syystä staattinen tieto vaatii pienemmät tekniset vaatimukset kuin dynaaminen tieto. Riskit ovat myös erilaiset staattisen ja dynaamisen tiedon välillä. Jos salassa pidettävää staattista tietoa julkaistaan vahingossa, on julkaisua vaikea perua, sillä tieto säilyy tiedon ladanneilla. Vahingossa julkaistun dynaamisen tiedon peruminen voidaan suorittaa sulkemalla tiedon rajapinta. Teknisiä valintoja pohtiessa kannattaa siis miettiä, kuinka isoja ja kuinka usein päivittyviä data-aineistoja ollaan avaamassa (kuvio 5). Pienille ja staattisille aineistoille riittää eräajo tietojärjestelmästä, kun suuremmat ja jatkuvasti muuttuvat aineistot taas kannattaa julkaista suoraan rajapinnasta ja lisätä rajapinnan linkki esimerkiksi datakatalogiin. (Avoindata.fi 2016b.)



KUVIO 5. Tietoaineiston koon ja päivitystiheyden vaikutus teknisiin valintoihin (Avoindata.fi 2016b)

Eräajo on vaivattomin tapa julkaista avointa dataa. Siinä tietojärjestelmästä tuodaan eräajotiedosto, joka julkaistaan esimerkiksi organisaation kotisivuilla tai dataportaalissa. Eräajotiedosto on hyvä valinta, kun julkaistaan harvoin päivittyvää staattista dataa. Eräajotiedosto on mahdollista julkaista monessa eri dataformaatissa ja se on käyttäjäystävällinen, sillä käyttäjä saa kaiken tiedon käyttöönsä yhdessä tiedostossa. (Ylärinne 2013, 16.) Rajapinta taas on hyvä ratkaisu datan julkaisulle, jos data-aineisto on suuri ja päivittyy usein. Rajapinnan avulla tietojen yhdistely toisiin data-aineistoihin on helppoa ja tämä tukee esimerkiksi sovelluskehitystä. Rajapinnasta tieto on myös saatavilla yhtä ajantasaisena kuin se on tietojärjestelmässäkin. (Kotkanen 2016, 12.)

Tiedon avaamisen vaikutukset organisaation sisäisiin prosesseihin on hyvä myös arvioida ennen datan avaamista, samoin kuin vaadittavat resurssit ja kulut. Tiedon avaamisen myötä organisaatiolle tulee uusia sisäisiä prosesseja, jotka vaativat henkilöresursseja. Mikäli organisaatiolla ei ole omasta takaa osaamista tiedon avaamiseen, tarvitaan siihen resursseja ulkopuolelta. Tiedon avaamiseen vaadittavat tekniset

toteutukset taas tuottavat datan ominaisuuksista ja koosta riippuen erinäisiä kustannuksia. (Avoindata.fi 2016c.)

Lopulta organisaatiolla tulisi olla tiedossa, millaiselle tiedolle löytyy kysyntää, mitä sisältöä tullaan avaamaan ja missä, suunnitelma tietovarantojen avaamisesta ja niiden kustannuksista sekä tavoitteet datan avaamiselle ja käsitys siitä miten avaamisprosessia seurataan ja kehitetään. (Avoindata.fi 2016c.)

#### 4.3 Valmistelu ja avaaminen

Kun avattavat tietoaaineistot ovat selvillä, on kannattavaa määritellä vastuut tietovarannon avaamisen, päivittämisen ja seurannan suhteen. Näin jokainen prosessiin osallistuva pysyy selvillä omista vastuualueistaan ja toiminta on tehokkaampaa. Datan päivystiheys ja päivityksistä tiedottaminen tulee myös olla selvillä, jotta käyttäjät pysyvät tietoisina datan ajantasaisuudesta. (Toikkanen ym. 2014c.)

Toikkasen ym. (2014c) mukaan avattavan aineiston laatu ja oikeellisuus on myös hyvä varmistaa ja mahdolliset virheet korjata. Aineistoihin tulisi myös laatia metatiedot, jotta käyttäjät voivat varmistua tiedon oikeellisuudesta. Suositeltavaa olisi, että metatiedot olisivat mahdollisimman kattavat, ajantasaiset sekä standardien mukaiset. Hyvät metatiedot edistävät myös tietoaaineiston löydettävyyttä hakukoneilla.

Tietoaaineistolle on valittava myös lisenssi, jolla annetaan oikeus tiedon käyttäjille hyödyntää dataa. Avoimen datan lisenssinä suositellaan Suomessa Creative Commons Nimeä 4.0 (CC BY 4.0) -lisenssiä. Lisenssi sallii vapaan jatkokäytön, kunhan datan lähde mainitaan. Toinen, erityisesti metatiedon käyttöluvaksi suositeltava lisenssi on CC0, jossa käytölle ei aseteta mitään rajoitteita. (Avoindata.fi 2016d.)

On suositeltavaa määritellä myös palautekanavat ja tarjota mahdollisesti käyttäjille tukea datan käyttöön. Palautetta voidaan kerätä ulkoapäin esimerkiksi verkkosivujen palautelomakkeen kautta tai vaikkapa sosiaalisen median välityksellä. Organisaation sisäistä palautetta



kannattaa kerätä myös, jotta avaamisprosessia voidaan kehittää. Käyttäjille voidaan tarjota tukea esimerkiksi organisaation verkkosivuilla esitettävien käyttöesimerkkien avulla. Myös toistuvista kysymyksistä laadittu lista voi olla hyödyllinen datan käyttäjän näkökulmasta. Palautteen ja sen pohjalta tehtävien kehitystoimenpiteiden avulla varmistetaan, että tiedon hyödyntäminen on jatkuvaa. (Avoindata.fi 2016d.)

Vielä ennen datan lopullista avaamista on hyvä käydä läpi kaikki oleellinen vaikkapa taulukon 2 kaltaisella tarkistuslistalla ja varmistaa ettei mitään tärkeää jäänyt suorittamatta. Tämän jälkeen data voidaan julkaista ja tiedottaa avauksesta.

TAULUKKO 2. Tarkistuslista datan avaajalle

<input type="checkbox"/> Tietoaineistolle on annettu nimi
<input type="checkbox"/> Tietoaineistolle on luotu kuvailutiedot eli metatiedot
<input type="checkbox"/> Datat avattavuuden rajoitteet ja velvoitteet on selvitetty
<input type="checkbox"/> Datat avattavuus joidenkin muokkauksen jälkeen on selvitetty
<input type="checkbox"/> Strategiset tavoitteet ovat selvät, esim. se mitä avaamisella tavoitellaan
<input type="checkbox"/> Data on lisensoitu uudelleenkäytön sallivasti, esim. CC Nimeä 4.0 lisenssillä
<input type="checkbox"/> Data on julkaistu avoimessa ja koneluettavassa formaatissa
<input type="checkbox"/> Datat laatu on tarkistettu ja virheet korjattu sekä arkaluontoinen tieto on piilotettu
<input type="checkbox"/> Datat julkaisukanava(t) on päätetty
<input type="checkbox"/> Datat päivitystiheys ja päivityksien aikataulu on suunniteltu
<input type="checkbox"/> Viestintä ja markkinointi datat avaamisen jälkeen on suunniteltu
<input type="checkbox"/> Data avattu

#### 4.4 Seuraaminen ja kehittäminen

Datan avaamisen jälkeen kannattaa heti aloittaa seuraamaan sen käyttöä ja kehittää dataa käyttäjien palautteen pohjalta. Palautteen avulla saadaan selville, millaiselle datalle on kysyntää ja millaiselle ei. Näin taas tiedetään mitä tietoaaineistoja ylipäättään kannattaa avata ja ylläpitää. Toinen hyvä mittari tiedon kysynnälle ovat tilastot datan käytöstä sekä latauskerrat. Tiedon vaikuttavuudesta saa kuvaa tutkimalla, kuinka moni sovellus hyödyntää avattua tietoaaineistoa. (Avoindata.fi 2016e.)

Tietoaaineistosta löytyvät virheet tai vanhentunut tieto kannattaa korjata mahdollisimman pian, jotta tieto olisi mahdollisimman käyttökelpoista ja oikeaa. Jos tiedon käytössä ilmenee väärinkäytöksiä, on niihin myös syytä puuttua. Viestiminen ja tiedottaminen ovat myös tärkeä osa datan avaamisprosessia, jotta käyttäjät pysyvät ajan tasalla ja tietoisina siitä mitä tietoaaineistoja julkaistaan. Palautteen, tietoaaineiston käytön ja vaikuttavuuden pohjalta tiedon avaamisprosessia on helpompi lähteä kehittämään, jotta tiedon jakaminen ja sitä kautta sen käyttö tehostuisivat. (Avoindata.fi 2016e.) Kuviossa 6 on kuvattu datan avaamisprosessi vielä tiivistettynä.



KUVIO 6. Datan avaamisprosessi tiivistettynä

## 5 TUTKIMUSMENETELMÄT

Tämän opinnäytetyön tutkimusmenetelmäksi valittiin kvalitatiivinen eli laadullinen tutkimusmenetelmä. Kvalitatiivisella tutkimusmenetelmällä kartoitetaan valitun kohdehenkilön omia kokemuksia tutkittavasta asiasta (Hirsjärvi & Hurme 2008, 27), ja sen takia se nähtiin tähän tutkimukseen sopivimpana menetelmänä. Tutkimuksessa haluttiin selvittää Lahden kaupungin eri yksiköiden vastuuhenkilöiden näkemyksiä ja kokemuksia avoimeen dataan ja sen avaamiseen liittyen sekä saada selville mitä haasteita kohdehenkilön edustaman yksikön datan avaamiselle on. Kohdeyksiköt valittiin tarkoituksenmukaisesti, sillä haluttiin saada näkemyksiä erilaista tietoa tuottavilta ja datan avaamisessa eri vaiheissa olevilta yksiköiltä.

### 5.1 Tutkimusaineiston keruu- ja analyysimenetelmät

Kohdehenkilöiden näkemyksien ja kokemusten selvittämiseksi heille suoritettiin haastattelut. Haastattelutyypinä käytettiin teemahaastattelua, jonka avulla pystyttiin selvittämään haastateltavan omat mielipiteet ja kokemukset mahdollisimman hyvin. Teemahaastattelussa keskustelu kulkee valittujen teemojen avulla eteenpäin ja teemat ovat kaikille haastateltaville samat. Kysymykset eivät ole yksityiskohtaisia, vaan pyrkivät johdattelemaan haastattelutilannetta eteenpäin, jotta teemaan liittyvistä aiheista saataisiin mahdollisimman kattavasti tietoa. (Hirsjärvi & Hurme 2008, 48.) Tämän tutkimuksen teemahaastattelun teemoja oli kolme:

- Ensimmäinen teema oli tietojärjestelmät ja niissä oleva tieto. Pyrittiin selvittämään, millaiset tietojärjestelmät yksiköllä on käytössään, millaista tietoa yksikön toiminta tuottaa, missä tieto sijaitsee ja missä formaateissa se on, kuinka usein tieto päivittyy ja kuinka paljon sitä on, sisältääkö tieto arkaluontoista materiaalia, onko tieto säädetty maksulliseksi ja mitkä asiat vaikuttavat tiedon julkaisuun.

- Toinen teema oli avoin data yksikön toiminnassa. Tässä pyrittiin selvittämään, onko yksikkö jo avannut dataa. Jos yksikkö oli jo näin tehnyt, haluttiin selvittää kokemuksia avaamisprosessista ja datan avaamisen hyödyistä yksikölle. Jos taas dataa ei oltu avattu, pyrittiin löytämään keskeisimmät syyt sille miksi näin ei ole vielä tehty.
- Kolmas teema pyrki selvittämään yksikön osaamisen ja resurssit datan avaamiselle. Haluttiin saada selville, onko avoin data käsitteenä tuttu yksikön henkilökunnalle, miten datan avaamiseen suhtaudutaan, liittyykö datan avaamiseen joitain pelkoja, onko yksikössä riittävästi teknistä osaamista ja riittävätkö yksikön henkilö- ja taloudelliset resurssit datan avaamiseen.

Nämä teemat valittiin teemahaastatteluun, koska niiden avulla kyettiin selvittämään toteutuvatko jo teorian pohjalta tunnetut haasteet ja ongelmakohdat datan avaamisprosessissa Lahden kaupungin osalta.

Haastatteluista kerätyn tutkimusaineiston analyysiin käytettiin deduktiivista eli teorialähtöistä päättelytapaa. Tälle päättelytavalle on olennaista, että lähtökohtana on teoria, jonka pohjalta muodostetaan hypoteeseja eli olettamuksia. Nämä olettamukset todennetaan empirian, kuten esimerkiksi haastatteluiden avulla. (Grönfors 2011, 14.) Tässä tutkimuksessa kerätyn aineiston analyysiin hyödynnettiin tieteellisiä julkaisuja ja asiantuntijoiden tuottamaa materiaalia, johon teemahaastatteluissa esiin nousseita asioita peilattiin. Näin pystyttiin todentamaan teorian pohjalta syntyneet olettamukset oikeiksi. Deduktiivinen päättelytapa valittiin tähän tutkimukseen juurikin siitä syystä, että avoimesta datasta ja datan avaamisesta löytyi hyvin asiantuntijoiden tuottamaa materiaalia, jonka pohjalta voitiin testata teorian paikkaansa pitävyyttä Lahden kaupungin tapauksessa.

## 5.2 Tutkimusprosessi

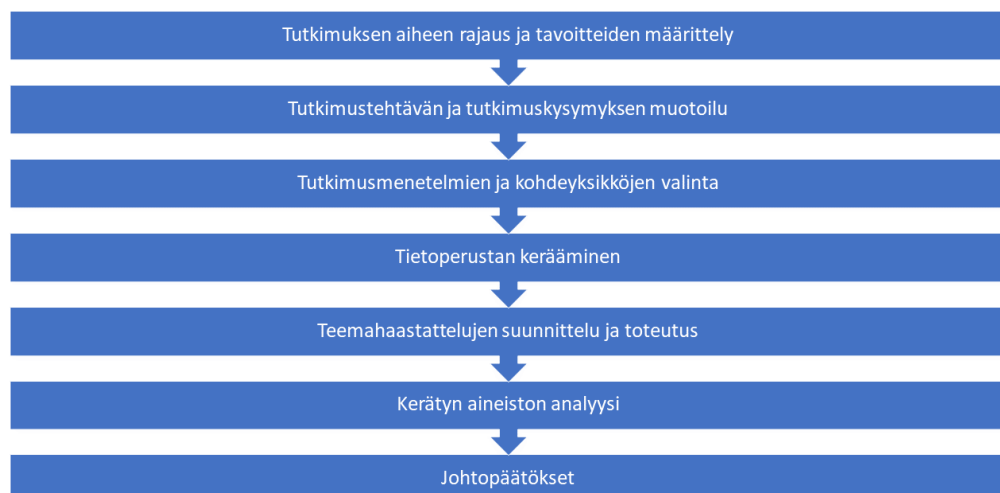
Tämän opinnäytetyön tutkimusprosessi (kuvio 7) käynnistyi palaverilla toimeksiantajan eli Lahden kaupungin edustajien kanssa. Palaverissa

käytiin läpi tutkimuksen aikataulus, rajattiin tutkimuksen aihe datan avaamisprosessiin ja sen haasteisiin sekä määriteltiin tutkimuksen tavoitteet.

Seuraavassa vaiheessa määriteltiin tutkimustehtävä ja tutkimuskysymys sekä valittiin tutkimusmenetelmät ja kohdehenkilöt, joille haastattelut tehtiin. Tämän jälkeen kerättiin tietoperustaa, jotta aihealueesta saataisiin tarpeeksi kattava tietopaketti, jonka pohjalta taas empiriaosuuteen saataisiin syvyyttä.

Tietoperustan keräämisen jälkeen valittiin haastatteluun sopivat teemat ja suunniteltiin niihin liittyvät kysymykset, joiden avulla kyettiin selvittämään Lahden kaupungin organisaatioiden datan avaamisprosessin haasteet. Tämän jälkeen tutkimusaineisto kerättiin teemahaastatteluilla, joissa haastateltiin Lahden kaupungin eri yksiköiden vastuuhenkilöitä. Haastatteluista ilmenneet asiat kirjattiin ylös analysointia varten.

Kun kaikki tutkimusaineisto oli kerätty, aloitettiin saatujen tulosten analysointi. Analyysin pohjalta tehtiin johtopäätökset siitä, mitkä ovat suurimmat haasteet Lahden kaupungin organisaatioiden datan avaamiselle sekä miten nämä haasteet voitaisiin ratkaista, jotta datan avaaminen helpottuisi.



KUVIO 7. Tutkimusprosessi ja sen kulku

## 6 TUTKIMUSAINEISTO JA SEN ANALYYSI

Tätä tutkimusta varten kerättiin tutkimusaineisto, jonka sisältö koostuu kolmesta eri teemahaastattelusta. Teemahaastattelut suoritettiin Lahden kaupungin eri yksiköiden vastuuhenkilöille ja tavoitteena oli kartoittaa datan avaamisen haasteita eri yksikköjen näkökulmasta.

Teemahaastatteluissa käytettiin kolmea eri teemaa ja niihin liittyviä kysymyksiä, jotka on kuvattu kappaleessa 5.1. Haastattelut olivat kestoltaan 40-70 minuuttisia ja ne suoritettiin Lahden kaupungin tiloissa, joko kokoushuoneessa tai haastateltavan omassa työhuoneessa. Valittujen kohdehenkilöiden edustamat yksiköt olivat maankäyttö, ympäristövalvonta sekä ympäristöterveys. Valitut yksiköt kuuluvat teknisen ja ympäristötoimialan alle, joka työllistää Lahdessa reilut 200 työntekijää (Lahden kaupunki 2016).

Teemahaastatteluissa esiin nousseet oleelliset asiat kirjattiin ylös paperille, jotta mitään tärkeää ei unohtuisi, kun aineistoa aletaan analysoida. Myöhemmin haastattelun jälkeen havainnot kirjoitettiin vielä puhtaaksi Word-dokumenttiin ja jäsenneltiin teemoittain järjestykseen, jotta analysointi olisi helpompaa suorittaa. Haastatteluista kerätyn tiedon lisäksi tutkimusaineiston tukena käytettiin Lahden kaupungin tuottamaa materiaalia, kuten PowerPoint -esityksiä yksiköiden toiminnasta.

Kun tutkimusaineisto oli kerätty, aloitettiin aineiston analysoiminen. Teemahaastatteluihin osallistuneet yksiköt analysoitiin kukin erikseen, koska ne ovat datan avaamisessa eri vaiheissa. Näin saatiin selville tiettyssä vaiheessa olevan yksikön haasteet ja löydettyä ratkaisut, jotka olisivat sovellettavissa myös muihin samassa tilanteessa oleviin yksiköihin. Maankäytön ja ympäristövalvonnan yksiköt ovat jo avanneet tietoaaineistojaan avoimeksi dataksi, kun taas ympäristöterveys ei vielä näin ole tehnyt. Seuraavissa kappaleissa tarkempi analyysi kunkin yksikön datan avaamisesta ja siihen liittyvistä haasteista sekä ehdotukset haasteiden selvittämiseksi. Lopuksi vielä vertailua havaittujen haasteiden välillä.

## 6.1 Haasteet maankäytön yksikössä

Maankäytön yksikkö oli avannut dataa jo aikaisemmin. Yksikön tietojärjestelmässä on 4 kunnan paikkatiedot sekä 12 kunnan väestötiedot. Näistä tiedoista paikkatietoaineistoja on julkaistu avoimena datana Lahden kaupungin kotisivuilla. Paikkatiedot sisältävät sijaintitietoja ja ominaisuuksia kuvaavia tietoja kuten tunnisteita, mittaushavaintoja, kuvailuja sekä luokituksia. Tiedot ovat yhdessä paikkatietojärjestelmässä, joka on yhteydessä moniin muihin järjestelmiin. Tietoa on paljon ja se päivittyy automaattisesti.

Maankäytön yksikön avoimena datana julkaistut paikkatietoaineistot ovat saatavilla WMS- ja WFS-rajapintapalveluiden kautta. Näistä rajapinnoista paikkatiedot on mahdollista hakea suoraan tiedon tallennuspaikasta. WFS-rajapinnan GML-formaatti on todettu yksikön käytössä vaikeakäyttöiseksi ja tämä on tuottanut haasteita datan käytölle.

Maankäytön yksikön tuottamista tiedoista suurin osa olisi mahdollista avata avoimeksi dataksi. Kaikkia tietoja ei kuitenkaan voida tai nähdä tarpeelliseksi avata. Osa tiedosta, kuten väestötiedot ovat salassapidettävää tietoa jota ei voida sellaisenaan julkaista. Osa taas, kuten kantakartat, ovat sellaisia tietoja jotka on säädetty maksulliseksi ja joista yksikkö saa toimintansa kannalta merkittäviä tuloja.

Avoimeen dataan maankäytön yksikössä suhtaudutaan positiivisesti, vaikka vielä ei datan avaamisella ole saavutettukkaan suuria hyötyjä. Yksikössä on riittävästi tietoa ja teknistä osaamista sekä taloudellisia että henkilöresursseja datan avaamiseen. Mitään erityisiä pelkoja datan avaamista kohtaan ei ole, mutta koska Lahden markkina-alue on melko pieni, on mietittävä kuinka paljon datan avaamiseen kannattaa panostaa. Kysyntää avoimena datana julkaistulle paikkatiedolle on ollut vielä vähän, johon osasyynä saattaa olla tiedon huono löydettävyyys. Tällä hetkellä linkit maankäytön avoimena datana julkaistuihin paikkatietorajapintoihin ovat löydettävissä monen mutkan kautta Lahden kaupungin sivuilta eikä avoimen datan markkinointiin ole vielä juuri panostettu.

Suurimpina haasteina maankäytön yksikön datan avaamisessa voidaan siis nähdä tiedon vähäinen kysyntä ja pieni markkina-alue. Kuten Martin (2015) toteaa tutkimuksessaan, datan käyttäjille olisi kyettävä todistamaan datan arvokkuus, jotta nämä saataisiin hyödyntämään dataa enemmän. Tämä vaatisi datan parempaa löydettävyyttä ja tehokkaampaa markkinointia, jotta käyttäjät olisivat ylipäätään tietoisia maankäytön yksikön julkaisemien tietoaaineistojen olemassaolosta. Myös yhteistyö yritysten, kehittäjäyhteisöjen tai vaikkapa alueen oppilaitosten kanssa voisi edistää avoimen datan käyttöä, sillä näin saataisiin tietoa millaiselle datalle olisi tarvetta käyttäjien näkökulmasta ja missä formaatissa käyttäjät dataa haluaisivat. Koska datan avaaminen on jatkuva prosessi, kuten Toikkanen ym. (2014c) kuvaavat, kannattaisi maankäytön yksikön kerätä palautetta avoimesta datasta ja sen pohjalta parantaa tietoaaineistojensa laatua sekä kehittää avaamisprosessiaan. Vaikeakäyttöisiin dataformaatteihin olisi hyvä antaa ohjeistuksia tai vaihtoehtoisesti jakaa linkkejä ohjeita sisältäviin julkaisuihin. Tämä tukisi käyttäjien kehitystyötä ja tekisi mahdollisesti datan hyödyntämisestä mielekkäämpää, kun tarvittavat tiedot ja tuki olisivat helposti saatavilla.

Datan löydettävyyteen voitaisiin vaikuttaa esimerkiksi yhtenäistämällä kaikkien Lahden kaupungin yksikköjen toimintamalleja siten, että kaikki avoimena datana julkaistut tietoaaineistot julkaistaisiin samassa paikassa. Tämä paikka voisi olla esimerkiksi kaupungin sivuilta helposti löydettävissä oleva datakatalogi, jossa kaikki tietoaaineistot ja linkit rajapintoihin olisivat selkeästi esillä ja niitä voisi hakea nimellä. Datan käyttöön voisi olla myös ohjeistuksia ja eri yksiköt voisivat antaa vinkkejä minkälaisiin sovelluksiin dataa voisi käyttää. Datakatalogissa käyttäjien olisi myös mahdollista esittää toiveita avattavista tietoaaineistoista sekä antaa palautetta jo julkaistusta datasta. Ominaisuutena voisi olla myös esimerkiksi postituslistalle liittyminen, jolloin käyttäjä saisi aina tiedon sähköpostiinsa, kun uusia tietoaaineistoja julkaistaan. Toinen mahdollinen paikka datan julkaisulle datakatalogin rinnalle olisi dataportaali, kuten avoindata.fi - palvelu. Dataportaalin kautta tavoitettaisiin potentiaalista käyttäjäkuntaa ja saataisiin avatut tietoaaineistot esimerkiksi kehittäjien tietoon. Mikäli dataa



päädytään julkaisemaan useammassa kuin yhdessä paikassa, olisi suositeltavaa, että esimerkiksi dataportalissa olisi vain linkit tietoaisteistoihin sekä rajapintoihin. Näin datan ylläpitäjän työ määrä ei kasva liian suureksi, kun tietoaisteistoa ei tarvitse päivittää useaan eri paikkaan ja samalla tämä lisää varmuutta siitä, että avattu data on jokaisessa julkaisukanavassa ajan tasalla.

Myös avoimen datan markkinoinnissa Lahden kaupungin kannattaisi käyttää yhteistä linjaa ja markkinointia voisi toteuttaa aktiivisesti esimerkiksi kaupungin sosiaalisen median kanavissa.

Hakukoneoptimointiin kannattaisi panostaa, jotta data olisi helposti löydettävissä myös hakukoneiden avulla. Hakukonelöydettävyyteen vaikuttavat myös tietoaisteistojen tarkat datan metatiedot.

## 6.2 Haasteet ympäristövalvonnan yksikössä

Ympäristövalvonnan yksikkö on avannut myös jossakin määrin tietovarantojaan avoimena datana. Yksikön tehtäviin kuuluu ympäristönsuojeluun liittyvien lakien ja asetusten sekä määräysten ja päätösten valvonta. Yksikön toiminnassa syntyy esimerkiksi tietoa vedenlaadusta, pohjavesistä, luontopoluista sekä uhanalaisten eläimien elinpaikoista. Avoimena datana on avattu esimerkiksi vedenlaatutietoja Vesijärveltä sekä Kymijärveltä. Lisäksi Lahden kaupungin TILDA-tilastotietokannasta on löydettävissä ympäristökatsauksia vuodesta 2000 lähtien. Vedenlaatutiedot ovat saattavilla Luode -datapalvelun kautta, jonne tieto päivittyy kerran päivässä. Mittaus suoritetaan automaattisesti mittauslautan avulla, joka mittaa vedenlaatua kerran tunnissa. Saatavia tietoja ovat veden lämpötila ja happipitoisuus sekä a-klorofylli- ja sinilevöpitoisuus. Näitä tietoja mitataan eri syvyyksistä, kuten yhden, neljän ja kahdeksan metrin syvyyksistä. Tiedot ovat nähtävissä Luode -datapalvelun internetsivuilla erilaisina graafisina kuvaajina sekä taulukkona. Käyttäjä voi ladata myös tiedot itselleen CSV-tiedostona. TILDA-tilastotietokannan tiedot ovat myöskin saatavilla vastaavassa CSV muodossa. Ympäristövalvonnan tuottamia tietoja on näiden lisäksi

saatavilla avoimena datana myös pintavesien tilan tietojärjestelmän vedenlaatuosiossa Suomen ympäristökeskuksen ylläpitämässä avoin tieto -palvelussa. Siellä tietoihin pääsee käsiksi rekisteröitymällä palveluun.

Ympäristövalvonnan tuottamista tietovarannoista suurin osa olisi avattavissa avoimena datana. Jotkin tiedot saattavat sisältää henkilöihin liittyvää salassa pidettävää tietoa, jota tulisi muokata avaamisen mahdollistamiseksi. Myös tiedot uhanalaisten eläimien elinpaikoista on pidettävä salassa. Yksikkö tuottaa myös palveluita esimerkiksi Hollolan kunnalle, joista se saa tuloja. Näistä palveluista syntyvät tiedot ovat kuitenkin myös avoimia samalla tapaa kuin Lahden alueenkin tiedot.

Ympäristövalvonnan yksikölle avoin data on vielä varsin uusi asia, mutta asiasta ollaan yksikön sisällä tietoisia. Datan avaamiseen liittyy jonkin verran pelkoja esimerkiksi toimintatapojen muutoksesta ja työmäärän kasvusta, joka olisi taas muista tehtävistä pois. Henkilöresurssit kuitenkin koetaan riittäviksi, jotta dataa voidaan avata. Myös taloudelliset edellytykset entistä paremmin datan avaamisen mahdollistaviin ratkaisuihin on olemassa, mutta tässäkin tapauksessa on mietittävä millaisiin ja kuinka kalliisiin ratkaisuihin kannattaa panostaa. Tällä hetkellä ympäristövalvonnan tuottamille tiedoille ei ole ollut vielä suurta kysyntää ja täyttä varmuutta ei ole myöskään siitä, millaista tietoa käyttäjät haluaisivat. Resursseja ei myöskään haluta haaskata jakamalla ja ylläpitämällä avointa dataa, jolla ei ole minkäänlaista käyttöä.

Ympäristövalvonnan yksikön ongelmana vaikuttaisi olevan maankäytön yksikön tapaan tiedon löydettävyyks. Avoimena datana julkaistut tietoaineistot ovat monessa eri paikassa ja monen klikkauksen takana. Tietoaineistojen markkinointiin ei ole myöskään panostettu. Huono löydettävyyks saattaa osaltaan vaikuttaa myös tiedon käyttöasteeseen, kun tiedon hyödyntäjät eivät ole tietoisia datasta tai eivät vain löydä julkaistuja tietoaineistoja. Löydettävyyteen tulisi panostaa esimerkiksi edellisessä kappaleessa mainituin keinoin. Datakatalogista tai dataportaalista tieto olisi helpommin löydettävissä ja markkinointi sosiaalisessa mediassa sekä

hakukoneoptimointi nostaisivat käyttäjäpuolen tietoisuutta avoimena olevasta datasta.

Koska ympäristövalvonnalla ei ole tarkkaa tietoa käyttäjien datatarpeista, tulisi yksikön myöskin tehdä yhteistyötä kehittäjäpuolen kanssa selvittääkseen, minkälaista dataa käyttäjät haluaisivat. Yhteistyötä voitaisiin tehdä tiiviimmin myös Lahden kaupungin sisällä, sillä ympäristövalvonnan tuottamat tiedot pitävät sisällään myöskin paljon paikkatietoja. Olisikin hyvä pohtia, voisivatko esimerkiksi ympäristövalvonnan ja maankäytön tuottamat tiedot toimia yhteen ja tuottaa toisilleen lisäarvoa. Näin molempien yksiköiden tuottaman tiedon kysyntää ja käytettävyyttä voitaisiin nostaa.

Luode -datapalvelussa olevien veden laatutietojen käyttökelpoisuus ei myöskään käyttäjän näkökulmasta ole helposti hyödynnettävää. Tieto on kyllä reaaliaikaista ja päivittyy kerran päivässä, mutta tiedot voi ladata itselleen CSV-tiedostona joka antaa vain lataushetkellä olevat tulokset. Käyttäjän tulisi käydä lataamassa joka päivä uusi tiedosto, jotta esimerkiksi dataa hyödyntävien sovellusten tiedot pysyisivät ajan tasalla. Tässä kannattaisikin arvioida tietoaineiston laadukkuutta kuviossa 3 esitetyn avoimen datan viiden tähden arviointimallin mukaan. Jotta tiedot saataisiin aina ajantasaisina esimerkiksi sovellusten käyttöön, olisi tietojärjestelmään luotava rajapinta josta tiedot saisi ladattua automaattisesti suoraan tiedon tallennuspaikasta. Ongelmana tässä on tosin myös se, että Luode -datapalvelua ylläpitää yritys, joka ei ole sidoksissa Lahden kaupunkiin ja näin ollen tilanteen korjaaminen vaatisi toimia palvelua ylläpitävältä yritykseltä.

Koska datan avaaminen tuntuu aiheuttavan pelkoja toimintatapojen muutoksista ja työmäärän kasvusta, olisi ympäristövalvonnan ja miksei kaikkien muidenkin Lahden kaupungin yksiköiden kasvatettava tietouttaan avoimesta datasta ja sen hyödyistä. Yksiköiden sisällä voitaisiin keskustella asioista ja pohtia, miten mahdollisia uusia sisäisiä prosesseja voitaisiin jakaa työntekijöiden kesken. Lisäksi kaupunki voisi järjestää henkilöstölleen koulutuksia ja infotilaisuuksia avoimeen dataan liittyen.

Kun tietoisuus asioista lisääntyisi, voisi se pienentää pelkoja ja näin opittaisiin hyödyntämään dataa myös omassa jokapäiväisessä toiminnassa. Tämä taas voisi omalta osaltaan tehostaa työskentelytapoja ja auttaa kaupunkia ja sen yksiköitä tuottamaan parempia palveluita.

### 6.3 Haasteet ympäristöterveyden yksikössä

Ympäristöterveyden yksikkö ei ole vielä avannut tietoaineistojaan avoimena datana. Yksikön tehtäviin kuuluu terveydensuojelun, elintarvikkeiden, tupakkalain sekä eläinten terveyden ja hyvinvoinnin valvominen. Yksikön käytössä on kuntatietojärjestelmä, johon sen toiminnassa syntyvät tiedot tallennetaan. Tällaisia tietoja ovat esimerkiksi valvontatiedot sekä valvontakohtetiedot. Valvontakohteita on satoja ja niiden tiedot päivittyvät automaattisesti, kun joku lisää tietoa järjestelmään. Yksikön käyttämä tietojärjestelmä on yhteydessä myös valtakunnallisiin tietojärjestelmiin. Tieto tallentuu järjestelmään taulukkomuodossa, mutta sen saa ulos vain Word-dokumenttina RTF-muodossa olevana raporttina. Tämä onkin ensimmäinen haaste datan avaamiselle, sillä RTF-formaatti ei täytä avoimen datan vaatimuksia ja sen ohjelmallinen käsittely on hankalaa. Elintarvikevalvonnan tietoja on saatavilla vapaasti Oivahymy - palvelusta, jossa elintarvikealan yritysten valvontatulokset ovat julkisesti näkyvillä. Tässäkään tapauksessa palvelusta saatava raportti ei kuitenkaan täytä avoimen datan vaatimuksia, sillä raportti on saatavilla ainoastaan PDF-dokumenttina.

Vaikka ympäristöterveyden tarkastustoiminta on itsessään julkista, tulee haasteita myös tiedon salassapidettävyyden ja arkaluontoisuuden takia. Tarkastustoiminnasta syntyvät tiedot sisältävät esimerkiksi terveydentilaan ja uskontoon liittyviä tietoja, joita ei voida julkaista avoimena datana. Tieto saattaa sisältää myös liikesalaisuuksia, jotka voivat julkiseksi tullessaan vaikuttaa yritysten liiketoimintaan negatiivisesti.

Ympäristöterveyden yksikkö ei ole vielä avannut tietoaineistojaan avoimena datana juurikin siitä syystä, että viranomaistyössä syntyvän datan seulominen on vaikeaa ja eikä esimerkiksi yritysten liiketoimintaa

haluta vaarantaa. Yksikössä nähdään, että datan avaaminen ei oikein tue viranomaistyötä. Myöskään datan avaamisen hyötyjä omalle organisaatiolle kuin kehittäjillekin ei ole vielä löydetty. Yksikössä ei osata sanoa tarkkaan, millaiselle ympäristöterveyden tuottamalle datalle olisi kysyntää, mutta esimerkiksi uimarantojen vesistä saatava tieto voisi olla sellaista, joka voisi hyödyttää sekä kehittäjiä että kansalaisia. Tiedon tuottaminen ja avaaminen vaatisi kuitenkin yhteistyötä uimarantojen omistajien kanssa, jotta käyttäjille saataisiin aina ajantasaista tietoa.

Avoin data on ympäristöterveyden yksikölle vielä varsin uusi asia, josta ei ole yksikön sisällä keskusteltu. Tiedonhankinta aiheesta on ollut lähinnä jokaisen työntekijän omalla vastuulla, kiinnostuksesta riippuen.

Epäilyksenä on, että avoin data lisää liikaa työn julkisuutta ja työn vaativuus kasvaisi, koska jouduttaisiin miettimään tarkemmin mitä tietoa mistäkin kirjataan. Olisi siis varottava virheitä, ettei salassapidettävää materiaalia pääse julkisuuteen.

Yksikön resurssit ovat myös rajalliset ja ajan ja energian riittäminen datan avaamiseen mietityttää. Myöskään teknistä osaamista yksiköllä ei omasta takaa ole, jolloin tekniset toteutukset olisi hankittava ulkopuoliselta toimijalta. Tämä luonnollisesti lisää kustannuksia ja siksi onkin mietittävä, onko rajallisten taloudellisten resurssien sijoittaminen datan avaamiseen kannattavaa vai olisiko parempi investoida rahat johonkin toiseen käyttötarkoitukseen.

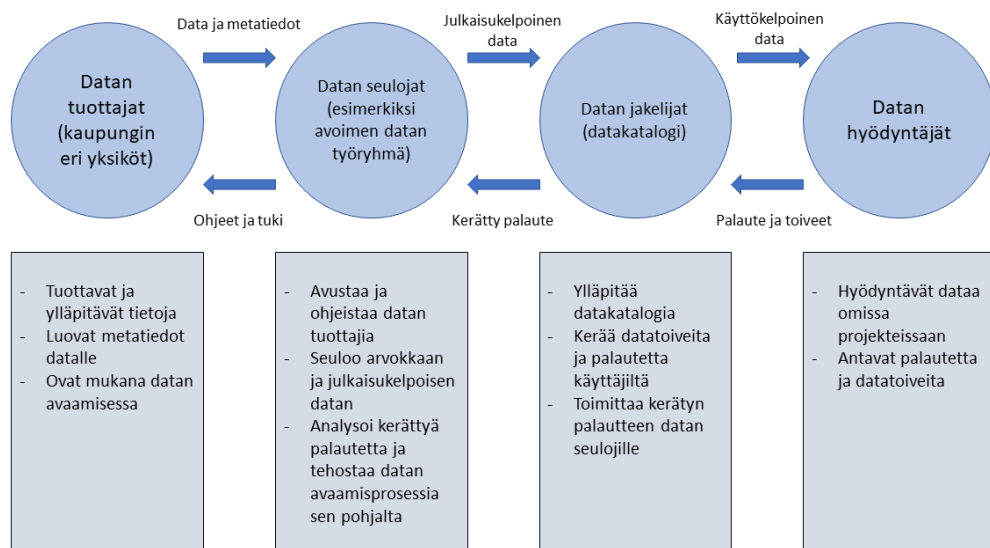
Ympäristöterveyden yksikön suurimmat haasteet datan avaamiselle ovat selvästi viranomaistyössä syntyvä arkaluontoinen tieto, tietämättömyys avoimen datan hyödyistä ja kysynnästä sekä rajalliset resurssit. Yksikön kannattaisikin lähteä liikkeelle luomalla tietovarannoistaan katsaus niin kuin Toikkanen ym. (2014c.) ehdottavat ja tehdä yhteistyötä datan hyödyntäjien kanssa. Näin saadaan selville, onko yksiköllä tietoa joka kiinnostaisi yrityksiä tai kehittäjiä. Yksikön tietovarannot kannattaa myös jakaa avattavissa oleviin, rajoitteita sisältäviin sekä sellaisiin tietovarantoihin, joita ei voi avata. Näin havaitaan helpommin mitä tietoa

voidaan avata sellaisenaan, mitä voidaan muokata avaamiskelpoiseksi ja mikä tieto ei ole avattavissa.

Kun tietovarannoista saadaan selkeä käsitys mitä tietoa on mahdollista avata ja mille tiedolle olisi kysyntää, on hyvä lähteä pohtimaan strategisia tavoitteita. Yksikön olisi esimerkiksi hyvä pohtia yhdessä, miksi dataa kannattaisi avata, mitä avaamisella tavoitellaan ja onko avaamisella mahdollisesti negatiivisia seuraamuksia. Myös avaamisprosessin rooleja olisi hyvä miettiä, jotta jokaisella olisi selkeä käsitys siitä mitä tehdä ja löydettäisiin kuhunkin tehtävään osaavimmat sekä motivoituneimmat tekijät. Tekniset valinnat olisi myös pohdittava, sillä tällä hetkellä ympäristöterveyden yksikön käyttämässä tietojärjestelmässä ei ole mahdollista tuoda dataa ulos avointa dataa tukevassa muodossa. Yksi vaihtoehto tähän olisi luoda järjestelmään datan ulostuonnin oikeassa formaatissa mahdollistava lisäosa, niin kuin Lahti ym. (2012) tekstissään kirjoittavat. Tämä tietysti aiheuttaisi kuluja, joten on mietittävä myös taloudellisia edellytyksiä. Toinen vaihtoehto olisi julkaista esimerkiksi Oivahymy -palvelussa olevia elintarvikevalvonnan kaltaisia tietoja jossakin taulukkomuotoisessa formaatissa. Tiedot voitaisiin muokata esimerkiksi Excelissä, tallentaa CSV-formaattiin ja ladata jakelukanaviin. Tämä ei ehkä olisi se tehokkain tapa avata dataa, mutta tähän ei vaadittaisi suuria resursseja eikä ohjelmointiosaamista. Alkuun olisikin hyvä lähteä pienin askelin eteenpäin, seurata datan avaamisen tuloksia ja kehittää avaamisprosessia sekä tietoaaineistoja palautteiden pohjalta.

Tulevaisuudessa Lahden kaupunki voisi ottaa yhä enemmän osaksi strategiaansa datan avaamisen, jotta ympäristöterveyden kaltaisten pienempien yksiköidenkin datan avaaminen helpottuisi. Kaupunki on jo ottanut askeleen kohti datan tehokkaampaa avaamista perustamalla avoimen datan työryhmän, joka käsittelee avoimeen dataan liittyviä asioita ja pohtii datan avaamisprosessia ja jakelukanavia. Ryhmän jäsenet myös toimivat tukihenkilöinä omassa yksikössään ja avustavat datan tuottajia datan avaamisprosessissa eteenpäin. Seuraava askel olisi saada kaikki yksiköt mukaan datan avaamiseen ja tehdä datan avaamisesta osa normaalia toimintaa. Kaupungin toimintamalleja voisi yhtenäistää

esimerkiksi siten, että yksiköt tuottaisivat ja ylläpitäisivät tietojaan sekä olisivat mukana datan avaamisprosessissa (kuvio 8). Yksiköt myös loisivat tietoaaineistoistaan metatiedot, jotta saadaan selkeä käsitys mitä aineisto sisältää. Avoimen datan työryhmän kaltainen asiantuntijoiden ryhmä antaisi yksiköille ohjeistuksia ja apua datan avaamiseen sekä seuloisi tiedot, joilla näkee olevan arvoa avoimena datana, hyväksyisi ne julkaisukelpoisiksi ja lisensoisi ne. Datan jakelu tapahtuisi esimerkiksi kappaleessa 6.1 kuvatun kaltaisessa datakatalogissa, jota ylläpitäisi oma ryhmänsä. Tiedonjakelijan tehtäviin voisi kuulua datakatalogin ylläpito sekä datatoiveiden ja käyttäjien palautteen kerääminen. Saatu palaute ja toiveet toimitettaisiin seulojille analysoitavaksi, jonka pohjalta datan avaamisprosessia voitaisiin kehittää.



KUVIO 8. Ehdotus yhteisestä toimintamallista, jota Lahden kaupunki voisi hyödyntää datan avaamisessaan

Lahden kaupungin kannattaisi myös kiinnittää huomiota uusissa tietojärjestelmähankinnoissa datan avaamisen paremmin mahdollistaviin järjestelmiin. Uusissa järjestelmissä tulisi olla mahdollisuus avoimen rajapinnan käyttöön. Datan formaatti päätettäisiin sen mukaan, kuinka usein data päivittyy. Harvoin päivittyvän datan saisi järjestelmästä ulos

eräajotiedostona, joka ladattaisiin jakelukanavaan. Usein päivittyvän tiedon käyttäjät saisivat taas suoraan tietojärjestelmään luodusta rajapinnasta. Datan avaamista tukisi myös yhä suurempi palvelujen ja asioiden muuntaminen sähköiseksi, jolloin saataisiin enemmän valmiskäyttöistä dataa julkaistavaksi. Olisi myös hyvä miettiä, voisivatko eri yksiköiden tuottamat tiedot olla yhteentoimivia, jolloin myös pienempien yksikköjen, kuten ympäristöterveyden tuottamalle datalle voisi löytyä enemmän kysyntää.

#### 6.4 Ilmenneiden haasteiden vertailu

Tutkimustuloksista nousee selkeästi esiin jokaisen haastatellun yksikön kohdalla yksi asia, riippumatta siitä ovatko yksiköt avanneet jo dataa vai eivät. Tiedolle ei ole missään yksikössä vielä nähty suurta kysyntää ja tämä osaltaan pistää miettimään kuinka paljon datan avaamiseen kannattaa panostaa. Tosin Lahden kaupungin datan avaaminen on vielä vasta alkuvaiheessa ja panostukset datan löydettävyyteen ja markkinointiin ovat kaikilla haastatelluilla yksiköillä olleet vähäiset. Tämä on voinut vaikuttaa merkittävästi käyttäjien tietoisuuteen avoinna olevasta datasta. Myös tietoisuus siitä minkälaista dataa käyttäjät haluaisivat, on vielä kaikilla yksiköillä heikko.

Dataa jo avanneiden yksiköiden resurssit ja tekninen osaaminen taas vaikuttaisivat olevan korkeammalla tasolla kuin dataa vielä avaamattoman yksikön. Tämä lienee jo yksi syy, miksi maankäyttö ja ympäristövalvonta ovat jo dataa avanneet ja ympäristöterveys taas ei. Dataa avanneiden yksiköiden sisällä myös tietoisuus avoimesta datasta on parempaa ja asioista on yksikön sisällä keskusteltu.

Tiedon avaamattomuuteen on voinut vaikuttaa myös se, että ympäristöterveys tuottaa arkaluontoisempaa tietoa, jota on vaikeampi lähteä julkaisemaan kuin taas maankäytön tai ympäristövalvonnan tuottamia tietoja. Tämä on osaltaan voinut johtaa siihen, että ympäristöterveyden yksikössä ei ole koettu avointa dataa yksikön



toimintaa hyödyttävänä asiana ja näin ollen siihen ei ole nähty tarvetta perehtyä.

Tekniset ratkaisut datan ulostuontiin ja julkaisuun ovat kaikilla haastatelluilla yksiköillä hieman erilaiset. Tähän vaikuttaa tietysti myös yksikköjen eroavat toiminta-alueet. Maankäytön data on saatavissa suoraan tiedon tallennuspaikasta rajapinnan kautta, ympäristövalvonnan data taas eräajotiedostoina ja ympäristöterveyden tietojärjestelmästä dataa ei vielä saa ulos avoimen datan kriteerien mukaan.

Tietojärjestelmissäkin on eroja ja edistyneempiä järjestelmiä käyttävät yksiköt näyttäisivät olevan pidemmällä datan avaamisessaan.

Julkaisukanavat ovat tällä hetkellä dataa jo avanneiden yksiköiden kohdalla varsin hajanaiset. Maankäytön paikkatietorajapinnan linkit ovat vaikeasti löydettävissä Lahden kaupungin kotisivuilta, kun taas ympäristövalvonnan avoimen datan tietoaineistot ovat hajallaan useammassa eri paikassa. Tämä vaikuttanee tiedon käyttöasteeseen, sillä tällä hetkellä käyttäjän tulee nähdä paljon vaivaa löytääkseen julkaistut tietoaineistot tai linkit rajapintoihin.

Ainoastaan maankäytön yksikössä ei datan avaaminen herättänyt suurempia pelkoja tai huolia. Ympäristövalvonnan ja ympäristöterveyden yksikköjen huolena olivat toimintatapojen muutokset sekä työmäärän kasvu. Lisäksi ympäristöterveyden yksikössä pelkona oli se, että työn luonne muuttuisi liiankin julkiseksi ja jouduttaisiin kiinnittämään liikaa huomiota siihen, mitä mistäkin valvontakohteesta voidaan kirjata. Näihin pelkoihin vaikuttaa osaltaan ympäristöterveyden tekemä viranomaistyö, jossa syntyy arkaluontoista materiaalia, joka voisi julkiseksi tullessaan vaikuttaa negatiivisesti esimerkiksi valvottavien yritysten liiketoimintaan.

## 7 JOHTOPÄÄTÖKSET

Tämän opinnäytetyön tarkoituksena oli havaita Lahden kaupungin eri yksiköiden tuottaman datan avaamiseen vaikuttavia haasteita ja löytää ratkaisuja, joilla avaamista voitaisiin tehostaa. Tutkimus pyrki vastaamaan tutkimuskysymykseen, joka oli:

- Mitä haasteita Lahden kaupungin eri yksiköiden tuottaman datan avaamisessa on ja miten avaamista voitaisiin edistää?

Tutkimuksen tuloksissa on havaittavissa, että suurimpina haasteina Lahden kaupungin datan avaamiselle nähdään datan vähäinen kysyntä ja tietämättömyys siitä, millaista dataa kehittäjät tarvitsevat. Esiin nousivat myös joidenkin dataformaattien vaikeakäyttöisyys, teknisen osaamisen puute, rajalliset resurssit, taloudelliset näkökulmat sekä vaikeasti avattavissa oleva arkaluontoinen tieto. Myös datan avaamisen vaikutukset oman työn luonteeseen herättävät epäilyksiä.

Teknisen puolen haasteena ovat tietojärjestelmät. Kaupungin käyttämistä tietojärjestelmistä osa on sen verran vanhanaikaisia, että ne eivät tue datan avaamista. Dataa on vaikea saada ulos järjestelmästä oikeassa formaatissa eikä järjestelmissä ole mahdollisuutta rajapinnan käyttöön.

Rajalliset resurssit sekä tietotaito hidastavat myös datan avaamista. Pienimmissä yksiköissä datan avaamisesta ei ole vielä edes keskusteltu ja aiheeseen perehtyminen on ollut lähinnä jokaisen siitä kiinnostuneen omalla vastuulla. Näissä yksiköissä resurssit ovat myös niukat ja ajan sekä energian riittäminen mietityttää.

Tehostaakseen datan avaamista, Lahden kaupungin tulisi yhtenäistää omia toimintamallejaan sekä tehdä yhteistyötä yritysten, kehittäjien ja datayhteisöjen kanssa. Yhteistyön avulla datan hyödyntäjille arvokkaan tiedon löytäminen helpottuu ja datalle saadaan hyödyntäjien kannalta oikeat julkaisukanavat sekä dataformaatit.

Toimintamallejaan yhtenäistämällä Lahden kaupunki edistäisi myös pienempien yksiköiden datan avaamista. Yhteisten toimintamallien avulla

sitoutettaisiin Lahden kaupungin eri yksiköt datan avaamiseen ja siitä tehtäisiin osa normaalia toimintaa. Yksiköt sitoutuisivat tuottamaan ja ylläpitämään dataansa sekä luomaan datalle metatiedot. Datan avaamisessa olisi mukana erillinen asiantuntijoiden ryhmä, kuten esimerkiksi avoimen datan työryhmä. Asiantuntijat olisivat yksiköiden tukena ja hoitasivat julkaisukelpoisen datan seulomisen ja lisensoinnin.

Data tulisi julkaista yhteisessä julkaisukanavassa, josta löytyisivät kaikki Lahden kaupungin avoimet tietoaineistot. Näin Lahden kaupungin asioista kiinnostuneet käyttäjät löytäisivät kaiken oleellisen tiedon helposti samasta paikasta. Julkaisukanava voisi olla esimerkiksi yhteinen datakatalogi, jossa myös palautteen ja toiveiden kerääminen olisi helppoa. Palautteen pohjalta asiantuntijat voisivat kehittää datan avaamisprosessia taas eteenpäin. Datan markkinointiin olisi myös syytä panostaa, sillä sen avulla datan löydettävyys paranee ja yhä useampi olisi tietoinen Lahden kaupungin avoimena datana julkaistuista tietoaineistoista. Markkinointia voitaisiin toteuttaa kaupungin sosiaalisen median kanavissa sekä myös uutiskirjeinä, jotka voisi tilata datakatalogin kautta.

Tulevaisuuden tietojärjestelmähankinnoissa Lahden kaupungin kannattaisi ottaa huomioon datan avaamisen mahdollistavat ratkaisut, jotta avaamisprosessi helpottuisi. Yksiköille olisi hyvä myös järjestää koulutuksia avoimeen dataan liittyen, jotta avoin data ja sen tuottamat hyödyt ymmärrettäisiin paremmin. Kun dataa opitaan hyödyntämään tehokkaammin, voi se parhaimmillaan kehittää yksiköiden sisäistä toimintaa ja auttaa yksiköitä tuottamaan parempia palveluita.

Yhteenvetona voidaan todeta, että Lahden kaupungin datan avaaminen on toistaiseksi vielä alkuvaiheessa, eikä siihen ole vielä panostettu kovinkaan paljon. Kaupungin tulisikin ottaa osaksi yhteistä strategiaansa datan avaaminen ja lähteä pienin askelin kehittämään avaamisprosessia eteenpäin. Yhteistyö datan hyödyntäjien kanssa sekä palautteen kerääminen datan avaamisesta ovat avainasemassa. Näin saadaan tietoa, jonka avulla avaamisprosessia voidaan tehostaa ja löydetään käyttäjien sekä kaupungin kannalta arvokkaat tietoaineistot julkaistavaksi.

## 8 POHDINTA

Kvalitatiivisen tutkimuksen luotettavuutta voidaan arvioida tulosten yleistettävyydellä tai siirrettävyydellä. Tämä tarkoittaa sitä, että pohditaan, ovatko tutkimustulokset yleistettävissä tai siirrettävissä muihin tilanteisiin. (Jyväskylän yliopisto 2010.) Tämän tutkimuksen tulokset voidaan nähdä tässä mielessä luotettavina, sillä haastateltavat kohdeyksiköt valittiin tarkoituksenmukaisesti, jotta saatiin selvyttä eri tilanteissa datan avaamisessa olevilta yksiköiltä. Näin ollen tutkimuksen tuloksia ja ratkaisuehdotuksia voidaan hyödyntää myös kohdeyksikköjen kanssa samassa tilanteessa oleviin yksiköihin. Eri yksiköt tietysti tuottavat erilaista dataa, mutta erilaiset toimintamallit ovat siirrettävissä kaikkien käyttöön.

Haasteena tutkimukselle oli se, että avoin data on vielä varsin uusi asia Lahden kaupungille. Tästä johtuen haastateltavien yksiköiden määrä jäi pieneksi, sillä ei nähty järkevänä haastatella montaa asiaan vielä perehtymätöntä yksikköä. Toisaalta tietyt koetut haasteet, kuten datan vähäinen kysyntä, toistuivat haastateltavien keskuudessa riippumatta siitä, olivatko näiden yksiköt jo avanneet dataa vai eivät. Toisena haasteena tutkimukselle oli sen ajankohta. Koska haastatteluita suoritettiin kesän aikana, ei kaikkia kohdehenkilöitä tavoittanut lomien takia helposti. Tämä toi haasteita aikataulutukseen ja työn sujuvaan etenemiseen.

Työn tarkoituksena oli nostaa esiin datan avaamisen kannalta olennaisia haasteita ja antaa Lahden kaupungille ehdotuksia datan avaamisprosessin tehostamiseen. Työn vaikutuksia on vielä vaikea arvioida, mutta tutkimustulokset nostivat esiin muutamia selkeitä kehityskohteita, joihin datan avaamisprosessissa olisi syytä panostaa. Tämän opinnäytetyön ja sen sisältämien ehdotusten pohjalta Lahden kaupunki voisi lähteä kehittämään itselleen sopivaa avaamisprosessia. Kun Lahden kaupunki saa tulevaisuudessa avattua dataa nykyistä laajemmin, voisi se tarjota tälle tutkimukselle jatkotutkimuksen aiheen esimerkiksi siitä, kuinka hyvin käyttäjät ovat lähteneet Lahden kaupungin avointa dataa hyödyntämään. Tämänkaltaisesta tutkimuksesta saataisiin tietoa siitä, kuinka kaupunki on onnistunut avaamisprosessissaan ja mitä asioita pitäisi vielä jatkokehittää.

## LÄHTEET

6aika 2017. Avoin data ja rajapinnat [viitattu 31.5.2017]. Saatavissa:  
<https://6aika.fi/avoin-data-ja-rajapinnat-karkihanke/>

Avoindata.fi 2016a. Lisenssit. Väestökisterikeskus, avoindata.fi -palvelu [viitattu 7.6.2017]. Saatavissa: <https://www.avoindata.fi/fi/content/lisenssit>

Avoindata.fi 2016b. Kartoita ja selvitä. Väestökisterikeskus, avoindata.fi -palvelu [viitattu 8.6.2017]. Saatavissa:  
<https://www.avoindata.fi/fi/opas/kartoita>

Avoindata.fi 2016c. Arvioi ja valitse. Väestökisterikeskus, avoindata.fi -palvelu [viitattu 9.6.2017]. Saatavissa:  
<https://www.avoindata.fi/fi/opas/analysoi>

Avoindata.fi 2016d. Valmistele ja avaa. Väestökisterikeskus, avoindata.fi -palvelu [viitattu 9.6.2017]. Saatavissa:  
<https://www.avoindata.fi/fi/opas/valmistele>

Avoindata.fi 2016e. Seuraa ja kehitä. Väestökisterikeskus, avoindata.fi -palvelu [viitattu 9.6.2017]. Saatavissa:  
<https://www.avoindata.fi/fi/opas/seuraa>

Avoindata.fi 2016f. Tiedon avaamisen vaiheet. Väestökisterikeskus, avoindata.fi -palvelu [viitattu 12.6.2017]. Saatavissa:  
<https://www.avoindata.fi/fi/opas/jatko>

Bellinger, G., Castro, D., Mills, A. 2004. Data, Information, Knowledge and Wisdom [viitattu 2.6.2017]. Saatavissa:  
<http://geoffreyanderson.net/capstone/export/37/trunk/research/ackoffDiscussion.pdf>

Creative Commons Suomi 2017a. Tietoa lisensseistä [viitattu 7.6.2017]. Saatavissa: <https://creativecommons.fi/lisenssit/>

Creative Commons Suomi 2017b. CC-lisenssin merkitsemisohjeet [viitattu 13.6.2017]. Saatavissa: <https://creativecommons.fi/lisenssien-kayttoohje/cc-lisenssin-merkitsemisohjeet/>

Euroopan Unionin avoimen datan portaali 2017. Tietoa linkitetystä datasta [viitattu 7.6.2017]. Saatavissa: <http://data.europa.eu/euodp/fi/linked-data>

Grönfors, M. 2011. Laadullisen tutkimuksen kenttätömenetelmät. Hämeenlinna: SoFia-Sosiologi-Filosofiapu [viitattu 2.7.2017]. Saatavissa: [https://books.google.fi/books?hl=fi&lr=&id=F3oaCwAAQBAJ&oi=fnd&pg=PA4&dq=deduktiivinen+p%C3%A4%C3%A4ttely&ots=Yf3xOszyoW&sig=rt\\_Fctq37qKNAncVGXK883rP3WU&redir\\_esc=y#v=onepage&q=deduktiivinen%20p%C3%A4%C3%A4ttely&f=false](https://books.google.fi/books?hl=fi&lr=&id=F3oaCwAAQBAJ&oi=fnd&pg=PA4&dq=deduktiivinen+p%C3%A4%C3%A4ttely&ots=Yf3xOszyoW&sig=rt_Fctq37qKNAncVGXK883rP3WU&redir_esc=y#v=onepage&q=deduktiivinen%20p%C3%A4%C3%A4ttely&f=false)

Helsinki Region Infoshare 2010. Mitä on avoin data? [viitattu 1.6.2017]. Saatavissa: <http://www.hri.fi/fi/mita-on-avoin-data/>

Hirsjärvi, S. & Hurme, H. 2008. Tutkimushaastattelu - Teemahaastattelun teoria ja käytäntö. E-kirja. Helsinki: Gaudeamus [viitattu 2.7.2017]. Saatavissa: <https://masto.finna.fi/Record/masto.142614>

Jyväskylän yliopisto 2010. Tutkimuksen toteuttaminen [viitattu 31.8.2017]. Saatavissa: <https://koppa.jyu.fi/avoimet/hum/menetelmapolkuja/tutkimusprosessi/tutkimuksen-toteuttaminen>

Kolehmainen, M. 2014. Julkishallinnon omistaman datan avaaminen: mitä tulisi varmistaa ennen teknistä toteutusta? Gofore Oy -blogi [viitattu 8.6.2017]. Saatavissa: <https://gofore.com/julkishallinnon-omistaman-datan-avaaminen-mita-tulisi-varmistaa-ennen-teknista-toteutusta/>

Koski, H., Honkanen, M., Luukkonen, J., Pajarinen, M., Ropponen, T. 2017. Avoimen datan hyödyntäminen ja vaikuttavuus. Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 40/2017 [viitattu 6.6.2017]. Saatavissa: [http://tietokayttoon.fi/documents/10616/3866814/40\\_avoimen+datan+16032017.pdf/0444467d-5400-4f0c-8728-2447cef039ad?version=1.0](http://tietokayttoon.fi/documents/10616/3866814/40_avoimen+datan+16032017.pdf/0444467d-5400-4f0c-8728-2447cef039ad?version=1.0)

- Kotkanen, H. 2016. Mikä on (avoin) rajapinta? Helsingin kaupunki, Näin avaat dataa -koulutus [viitattu 9.6.2017]. Saatavissa: [https://www.hel.fi/hel2/tietokeskus/data/dokumentit/koulutusmateriaali/Hyodyntamiskoulutus\\_Henri\\_Kotkanen.pdf](https://www.hel.fi/hel2/tietokeskus/data/dokumentit/koulutusmateriaali/Hyodyntamiskoulutus_Henri_Kotkanen.pdf)
- Lahden kaupunki 2016. Henkilöstö lukuina [viitattu 10.7.2017]. Saatavissa: <http://lahdenvuosi.fi/henkilostokertomus-2016/henkilosto-lukuina>
- Lahti, L., Parkkinen, J., Lehtomäki, J. 2012. Avoin suomalainen kaupunkidata. Haasteet. Kaupunkitutkimuksen päivät 3.5.2012 – PowerPoint-esitys [viitattu 12.6.2017]. Saatavissa: [https://louhos.github.io/static/ktp\\_esitys.js/deck/index.html#sorvi-intro-1](https://louhos.github.io/static/ktp_esitys.js/deck/index.html#sorvi-intro-1)
- Martin, C. 2015. Why does the Open Government Data agenda face such barriers? The Policy and Internet Blog [viitattu 19.6.2017]. Saatavissa: <http://blogs.oii.ox.ac.uk/policy/why-does-the-open-government-data-agenda-face-such-barriers/>
- Mäkinen, K. 2012. Avoin data ja lainsäädäntö—mitä voin julkaista? Helsingin kaupunki, hallintokeskus [viitattu 8.6.2017]. Saatavissa: [http://www.hri.fi/fi/files/2012/10/HRI\\_Julkisuus-ja-henkil%C3%B6tiedot.pdf](http://www.hri.fi/fi/files/2012/10/HRI_Julkisuus-ja-henkil%C3%B6tiedot.pdf)
- Poikola, A., Kola, P. & Hintikka, K.A. 2010. Julkinen data, johdatus tietovarantojen avaamiseen. Liikenne- ja viestintäministeriö [viitattu 1.6.2017]. Saatavissa: <http://www.lvm.fi/documents/20181/815557/Julkinen+data/467e7da9-3038-46a1-b47e-994d7cd102d2?version=1.0>
- Schrier, B. 2014. Government Open Data: Benefits, Strategies, and Use [viitattu 19.6.2017]. Saatavissa: <https://depts.washington.edu/esreview/wordpress/wp-content/uploads/2014/07/2014-Government-Open-Data.pdf>
- Toikkanen, T., Kalliala, E., Poikola, A. & Sillanpää, L. 2014a. Mitä on avoin data? Datan avaajan kurssi. Peer 2 Peer University [viitattu 1.6.2017]. Saatavissa: <https://courses.p2pu.org/en/courses/2486/content/5069/>

Toikkanen, T., Kalliala, E., Poikola, A. & Sillanpää, L. 2014b. Miksi julkaista avointa dataa? Datan avaajan kurssi. Peer 2 Peer University [viitattu 6.6.2017]. Saatavissa:

<https://courses.p2pu.org/en/courses/2486/content/5068/>

Toikkanen, T., Kalliala, E., Poikola, A. & Sillanpää, L. 2014c. Näin alkuun datan avaamisessa. Datan avaajan kurssi. Peer 2 Peer University [viitattu 8.6.2017]. Saatavissa:

<https://courses.p2pu.org/en/courses/2486/content/5073/>

Toikkanen, T., Kalliala, E., Poikola, A. & Sillanpää, L. 2014d. Esteiden raivaaminen. Datan avaajan kurssi. Peer 2 Peer University [viitattu 12.6.2017]. Saatavissa:

<https://courses.p2pu.org/en/courses/2486/content/5071/>

Vakkari, M. 2013. JulkICT Lab ja Dataportaali, avoin data ja palvelukokeilut [viitattu 20.6.2017]. Saatavissa:

<http://slideplayer.fi/slide/11290659/>

Valtiovarainministeriö 2017. Avoin tieto [viitattu 31.5.2017]. Saatavissa:

<http://vm.fi/avoin-tieto>

Ylärinne, T. 2013. Avoimen datan julkaisutekniikat [viitattu 7.6.2017].

Saatavissa:

[https://www.hel.fi/hel2/tietokeskus/data/dokumentit/koulutusmateriaali/Avoimen\\_datan\\_julkaisutekniikat.pdf](https://www.hel.fi/hel2/tietokeskus/data/dokumentit/koulutusmateriaali/Avoimen_datan_julkaisutekniikat.pdf)



## LIITTEET

### LIITE 1. Teemahaastattelun runko

#### Teema 1. Tietojärjestelmät ja niissä oleva tieto

- Mitä tietojärjestelmiä yksiköllä on käytössään?
- Minkälaista tietoa järjestelmissä on?
- Missä formaateissa tieto on järjestelmissä?
- Kuinka usein päivittyvää ja minkä kokoista tietoa järjestelmissä on?
- Onko tieto saatavilla helposti ulos järjestelmistä?
- Sisältääkö tieto henkilötietoja, muuta salassa pidettävää tai arkaluontoista tietoa?
- Onko tiedon luovutus säädetty maksulliseksi?
- Onko tieto ulkopuolisen toimijan tuottamaa ja sitooko sitä toimeksiantosopimus joka estää/rajoittaa tiedon julkaisemista?

#### Teema 2. Avoin data yksikön toiminnassa

- Onko yksikkönne avannut jo dataa julkiseen käyttöön?
- Jos ei ole avattu, niin miksi?
- Jos dataa on jo avattu, niin minkälaisia kokemuksia datan avaamisesta on saatu?
- Jos dataa on jo avattu, onko avattu data mielestänne helposti löydettävissä?
- Onko tiedossa, minkälaiselle datalle on kysyntää?

#### Teema 3. Osaaminen ja resurssit

- Onko avoimen datan käsite yksikössänne tuttu?
- Miten avaamiseen suhtaudutaan?
- Liittykö datan avaamiseen mahdollisesti jotain pelkoja?
- Onko yksikössänne riittävästi teknistä osaamista datan avaamiseen?
- Onko yksikkönne resurssit riittävät datan avaamiselle ja ylläpidolle?